CALCUL STOCHASTIQUE ET FINANCE

Peter TANKOV
peter.tankov@ensae.fr

Nizar TOUZI
nizar.touzi@polytechnique.edu

ECOLE POLYTECHNIQUE
DÉPARTEMENT DE MATHÉMATIQUES APPLIQUÉES

Last update: Septembre 2018

Contents

Chapter 1

Introduction: discrete time derivatives pricing

Financial mathematics is a young field of applications of mathematics which experienced a huge growth during the last thirty years. It is by now considered as one of the most challenging fields of applied mathematics by the diversity of the questions which are raised, and the high technical skills that it requires.

These lecture notes provide an introduction to stochastic finance for the students of third year of Ecole Polytechnique. Our objective is to cover the basic Black-Scholes theory from the modern martingale approach. This requires the development of the necessary tools from stochastic calculus and their connection with partial differential equations.

Modeling financial markets by continuous-time stochastic processes was initiated by Louis Bachelier (1900) in his thesis dissertation under the supervision of Henri Poincaré. Bachelier's work was not recognized until the recent history. Sixty years later, Samuelson (Nobel Prize in economics 1970) came back to this idea, suggesting a Brownian motion with constant drift as a model for stock prices. However, the real success of Brownian motion in the financial applications was realized by Fisher Black, Myron Scholes, et Robert Merton (Nobel Prize in economics 1997) who founded between 1969 and 1973 the modern theory of financial mathematics by introducing the portfolio theory and the no-arbitrage pricing arguments. Since then, this theory gained an important amount of rigor and precision, essentially thanks to the martingale theory developed in the eighties.

Although continuous-time models are more demanding from the technical viewpoint, they are widely used in the financial industry because of the simplicity of the resulting formulae for pricing and hedging. This is related to the powerful tools of differential calculus which are available only in continuous-time. We shall first provide a self-contained introduction of the main concept from stochastic analysis: Brownian motion, stochastic integration with respect to the Brownian motion, Itô's formula, Girsanov change of measure Theorem, connection with the heat equation, and stochastic differential equations. We then consider the Black-Scholes continuous-time financial market where the no-arbitrage concept is sufficient for the determination of market prices of derivative securities. Prices are expressed in terms of the unique risk-neutral measure, and can be expressed in closed form for a large set of relevant derivative securities. The final chapter provides the main concepts in interest rates models in the gaussian case.

In order to motivate the remaining content of theses lecture notes, we would like to draw the reader about the following major difference between financial engineering and more familiar applied sciences. Mechanical engineering is based on the fundamental Newton's law. Electrical engineering is based on the Maxwell equations. Fluid mechanics are governed by the Navier-Stokes and the Bernoulli equations. Thermodynamics rest on the fundamental laws of conservation of energy, and entropy increase. These principles and laws are derived by empirical observation, and are sufficiently robust for the future development of the corresponding theory.

In contrast, financial markets do not obey to any fundamental law except the simplest *no-dominance* principle which states that valuation obeys to a trivial monotonicity rule, see Section 1.2 below.

Consequently, *there is no universally accurate model in finance.* Financial modeling is instead based upon comparison between assets. The Black-Scholes model derives the price of an option by comparison to the underlying asset price. But in practice, more information is available and one has to incorporate the relevant information in the model. For this purpose, the Black-Scholes model is combined with convenient calibration techniques to the available relevant information. Notice that information is different from one market to the other, and the relevance criterion depends on the objective for which the model is built (prediction, hedging, risk management...). Therefore, the nature of the model depends on the corresponding market and its final objective.

So again, *there is no universal model, and any proposed model is wrong.* Financial engineering is about building convenient tools in order to make these wrong models less wrong. This is achieved by accounting for all relevant information, and using the only *no-dominance* law, or its stronger version of *no-arbitrage.* An introduction to this important aspect is contained in Chapter 10.

Given this major limitation of financial modeling, a most important issue is to develop tools which measure the underlying risk in any financial position, and also the risk induced by any model used in its management. The importance of this activity was highlighted by the past financial crisis, and even more emphasized during the recent subprime financial crisis. Chapter 14 provides the main tools and ideas in this area.

In the remaining of this introduction, we introduce the reader to the main notions in derivative securities markets. We shall focus on some popular examples of derivative assets, and we provide some properties that their prices must satisfy independently of the distribution of the primitive assets prices. The only ingredient which will be used in order to derive these properties is the *no* *dominance principle* introduced in Section1.2 below.

1.1 Basic derivative products

1.1.1 European and American options

The most popular examples of derivative securities are European and American call and put options. More examples of contingent claims are listed in Section 1.5 below.

A *European call option* on the asset S^i is a contract where the seller promises to deliver the risky asset S^i at the maturity T for some given *exercise price*, or *strike*, $K > 0$. At time T, the buyer has the possibility (and not the obligation) to exercise the option, i.e. to buy the risky asset from the seller at strike K. Of course, the buyer would exercise the option only if the price which prevails at time T is larger than K. Therefore, the gain of the buyer out of this contract is

$$B \;=\; (S^i_T - K)^+ \quad = \quad \max\{S^i_T - K, 0\}\,,$$

i.e. if the time T price of the asset S^i is larger than the strike K, then the buyer receives the payoff $S^i_T - K$ which corresponds to the benefit from buying the asset from the seller of the contract rather than on the financial market. If the time T price of the asset S^i is smaller than the strike K, the contract is worthless for the buyer.

A *European put option* on the asset S^i is a contract where the seller promises to purchase the risky asset S^i at the maturity T for some given *exercise price*, or *strike*, $K > 0$. At time T, the buyer has the possibility, and not the obligation, to exercise the option, i.e. to sell the risky asset to the seller at strike K. Of course, the buyer would exercise the option only if the price which prevails at time T is smaller than K. Therefore, the gain of the buyer out of this contract is

$$B \;=\; (K - S^i_T)^+ \quad = \quad \max\{K - S^i_T, 0\}\,,$$

i.e. if the time T price of the asset S^i is smaller than the strike K, then the buyer receives the payoff $K - S^i_T$ which corresponds to the benefit from selling the asset to the seller of the contract rather than on the financial market. If the time T price of the asset S^i is larger than the strike K, the contract is worthless for the buyer, as he can sell the risky asset for a larger price on the financial market.

An *American call (resp. put) option* with maturity T and strike $K > 0$ differs from the corresponding European contract in that it offers the possibility to be exercised at any time before maturity (and not only at the maturity).

The seller of a derivative security requires a compensation for the risk that he is bearing. In other words, the buyer must pay the price or the premium for the benefit of the contrcat. The main interest of this course is to determine this price. In the subsequent sections of this introduction, we introduce the no dominance principle which already allows to obtain some model-free properties of options which hold both in discrete and continuous-time models.

In the subsequent sections, we shall consider call and put options with exercise price (or strike) K, maturity T, and written on a single risky asset with price S. At every time $t \leq T$, the American and the European call option price are respectively denoted by

$$\mathrm{Call}(t, S_t, T, K) \quad \text{and} \quad \mathrm{call}(t, S_t, T, K).$$

Similarly, the prices of the American and the Eurpoean put options are respectively denoted by

$$\mathrm{Put}(t, S_t, T, K) \quad \text{and} \quad \mathrm{put}(t, S_t, T, K).$$

The intrinsic value of the call and the put options are respectively:

$$\begin{array}{rcccl} \mathrm{Call}(t, S_t, t, K) & = & \mathrm{call}(t, S_t, t, K) & = & (S_t - K)^+ . \\ \mathrm{Put}(t, S_t, t, K) & = & \mathrm{put}(t, S_t, t, K) & = & (K - S_t)^+ , \end{array}$$

i.e. the value received upon immediate exercing the option. An option is said to be *in-the-money* (resp. *out-of-the-money*) if its intrinsic value is positive (resp. negative). If $K = S_t$, the option is said to be *at-the-money*. Thus a call option is in-the-money if $S_t > K$, while a put option is in-the-money if $S_t < K$.

1.1.2 Bonds and term structure of interest rates

A *zero-coupon bon* (ZCB) is a contrat stipulating that the seller delivers a fixed unitary amont at some fixed maturity (expressed in the corresponding currency), and receives a premium from the buyer at the signature of the contract. For instance, a $T-$maturity ZCB, denominated in Euros, is defined by the payment of 1 Euro at the maturity T.

The singularity of ZCB's is that they are defined so as to isolate the effect of interest rates, avoiding any other randomness in the definition of the contract. We shall denote by

$$P_t(T)$$

the time $t-$price of the $T-$maturity ZCB. The corresponding yield to maturity $R_t(T)$ is defined by

$$e^{-(T-t)R_t(T)} := P_t(T)$$

and allows to compare different maturities interest rates after convenient normalization. The graph of the map:

$$T \in (t, \infty) \quad \longmapsto \quad R_t(T)$$

is called the *term structure of interest rates*. By extrapolating this graph to the origin $T = t$, we may define *the instantaneous interest rate*

$$r_t := R_t(t),$$

which is not more than a limiting value, un-observable from the market data, and allowing to simplify the mathematical analysis by passage to the continuous-time setting.

A bond is a contract defined by a deterministic stream of payments $F_1, \ldots, F_n$ at given maturities $T_1 < \ldots < T_n$, respectively. Given the term structure of interest rates, the price of such a bon at any time $t \leq T_1$ is given by:

$$F_1 P_t(T_1) + \ldots + F_n P_t(T_n).$$

In practice, bonds available on the financial market are all defined by means of annual stream of payments called *coupons*, which are payables at fixed annual dates. Consequently, only bonds with less than one year remaining maturity are free of coupons and can be considered as ZCB. The prices of the remaining ZCB (with more than one year remaining maturties) are computed from the market prices of bonds by solving a linear system. This method of resolution is called the *Bootstrapping* technique.

The last financial crisis evidenced the importance of the counterparty risk, and highlighted the crucial importance of liquidity and funding risks. In particular, the assumption stating the existence of a "default-free" reference counterparty, whose emission bonds serve as a reference for the term structure of interest rates, can not be admitted any more. Such a role was played by government bounds, and can not be accepted in view of the current sovereign debt crisis... An important consequence is the co-existence of various term structures of interest rates $\{R_t^i(T), T \geq t\}$, $i = 1, \ldots, N$, corresponding to different underlying debtors.

1.1.3 Forward contrats

A forward contract stipulates that the seller delivers one unit of some underlying asset S at some given maturity T for an exercise price K_t fixed at the date t of signature of the contract, with the two following major differences with european calls:

- the buyer has the obligation to pay the price K_t for the risky asset at the maturity T, inducing a linear payoff $S_T^i - K_t$ at the maturity T which may be negative,

- the forward contract is set so as there is no premium transfer between the buyer and the seller at inception of the contract.

From the last feature of forward contracts, it is clear that the exercice price K_t is the key-quantity on which the buyer and the seller need to agree in order for the trade to happen. For this reason, K_t is referred to as the price of the forward contract at time t.

We shall see in Section 1.2.1 that the time $t-$price of the forward contract $S_t^{(\mathrm{fw})}(T)$ in a financial market **without frictions** is computed from the underlying spot price and the ZCB by:

$$S_t^{(\mathrm{fw})}(T) \quad = \quad \frac{S_t}{P_t(T)}.$$

This relation will be seen to hold regardless of any choice of modeling, i.e. model-free.

1.2 No dominance principle and first properties

We shall assume that there are no market imperfections as transaction costs, taxes, or portfolio constraints, and we will make use of the following concept.

No dominance principle *Let X be the gain from a portfolio strategy with initial cost x. If $X \geq 0$ in every state of the world, Then $x \geq 0$.*

1.2.1 Valuation of forward Contracts

Recall that a forward contract F on the underlying S, traded at time $t$for the maturity T, sets an exercise price K_t (the price of the forward contract) so as to generate a payoff at maturity T:

$$F_T \quad = \quad S_T - K_t,$$

with zero initial value of the contract at time t.

Assume that the underlying asset is **without dividends**. It is clear that F_T is the time $T-$value of the portfolio consisting of a long position of one unit of the underlying and a short position of a quantity K_t of $T-$maturity ZCBs. By the domination principle, applied both to F and $-F$, it follows that $F_t = S_t - K_t P_t(T)$. Consequently, the price of the forward contract is defined by solving $0 = F_t = S_t - K_t P_t(T)$. This shows that the forward contract price of completely determined by the no-domination principle:

$$K_t \quad = \quad \frac{S_t}{P_t(T)},$$

regardless of any choice of modeling.

1.2.2 Some properties of options prices

1 Notice that, choosing to exercise the American option at the maturity T provides the same payoff as the European counterpart. Then the portfolio consisting of a long position in the American option and a short position in the European counterpart has at least a zero payoff at the maturity T. It then follows from the dominance principle that *American calls and puts are at least as valuable as their European counterparts:*

$$\mathrm{Call}(t, S_t, T, K) \geq \mathrm{call}(t, S_t, T, K) \quad \text{and} \quad \mathrm{Put}(t, S_t, T, K) \geq \mathrm{put}(t, S_t, T, K)$$

2 By a similar easy argument, we now show that *American and European call (resp. put) options prices are decreasing (resp. increasing) in the exercise price,* i.e. for $K_1 \geq K_2$:

$$\begin{array}{rcl} \mathrm{Call}(t, S_t, T, K_1) \leq \mathrm{Call}(t, S_t, T, K_2) & \text{and} & \mathrm{call}(t, S_t, T, K_1) \leq \mathrm{call}(t, S_t, T, K_2) \\ \mathrm{Put}(t, S_t, T, K_1) \geq \mathrm{Put}(t, S_t, T, K_2) & \text{and} & \mathrm{put}(t, S_t, T, K_1) \geq \mathrm{put}(t, S_t, T, K_2) \end{array}$$

Let us justify this for the case of American call options. If the holder of the low exercise price call adopts the optimal exercise strategy of the high exercise price call, the payoff of the low exercise price call will be higher in all states of the world. Hence, the value of the low exercise price call must be no less than the price of the high exercise price call.

3 *American/European Call and put prices are convex in K.* Let us justify this property for the case of American call options. For an arbitrary time instant $u \in [t, T]$ and $\lambda \in [0, 1]$, it follows from the convexity of the intrinsic value that

$$\lambda \left(S_u - K_1\right)^+ + (1 - \lambda) \left(S_u - K_2\right)^+ - \left(S_u - \lambda K_1 + (1 - \lambda) K_2\right)^+ \quad \geq \quad 0\,.$$

We then consider a portfolio X consisting of a long position of λ calls with strike K_1, a long position of $(1 - \lambda)$ calls with strike K_2, and a short position of a call with strike $\lambda K_1 + (1 - \lambda) K_2$. If the two first options are exercised on the optimal exercise date of the third option, the resulting payoff is non-negative by the above convexity inequality. Hence, the value at time t of the portfolio is non-negative.

4 We next show the following result for the sensitivity of European call options with respect to the exercise price:

$$-P_t(T) \;\leq\; \frac{\mathrm{call}\,(t, S_t, T, K_2) - \mathrm{call}\,(t, S_t, T, K_1)}{K_2 - K_1} \;\leq\; 0$$

The right hand-side inequality follows from the decrease of the European call option c in K. To see that the left hand-side inequality holds, consider the portfolio X consisting of a short position of the European call with exercise price K_1, a long position of the European call with exercise price K_2, and a long position of $K_2 - K_1$ zero-coupon bonds. The value of this portfolio at the maturity T is

$$X_T \quad = \quad -(S_T - K_1)^+ + (S_T - K_2)^+ + (K_2 - K_1) \;\geq\; 0\,.$$

By the dominance principle, this implies that $-\mathrm{call}\,(S_t, \tau, K_1) + \mathrm{call}\,(S_t, \tau, K_2) + P_t(\tau)(K_2 - K_1) \geq 0$, which is the required inequality.

5 *American call and put prices are increasing in maturity*, i.e. for $T_1 \;\geq\; T_2$:

$$\mathrm{Call}(t, S_t, T_1, K) \geq \mathrm{Call}(t, S_t, T_2, K) \text{and } \mathrm{Put}(t, S_t, T_1, K_1) \geq \mathrm{Put}(t, S_t, T_2, K_2)$$

This is a direct consequence of the fact tat all stopping strategies of the shorter maturity option are allowed for the longer maturity one. Notice that this argument is specific to the American case.

1.3 Put-Call Parity

When the underlying security pays no income before the maturity of the options, the prices of calls and puts are related by

$$\mathrm{put}(t, S_t, T, K) \quad = \quad \mathrm{call}(t, S_t, T, K) - S_t + KP_t(T)\,.$$

Indeed, Let X be the portfolio consisting of a long position of a European put option and one unit of the underlying security, and a short position of a European call option and K zero-coupon bonds. The value of this portfolio at the maturity T is

$$X_T \;=\; (K - S_T)^+ + S_T - (S_T - K)^+ - K \quad = \quad 0\,.$$

Applying the dominance principle to X_T and $-X_T$, we see that the initial value of this portfolio is both non-negative and non-positive, which provides the required identity.

Notice that this argument is specific to European options. We shall see in fact that the corresponding result does not hold for American options.

Finally, if the underlying asset pays out some dividends then, the above argument breaks down because one should account for the dividends received by holding the underlying asset S. If we assume that the dividends are known in advance, i.e. non-random, then it is an easy exercise to adapt the put-call parity to this context. However, if the dividends are subject to uncertainty as in real life, there is no direct way to adapt the put-call parity.

1.4 Bounds on call prices and early exercise of American calls

1. From the monotonicity of American calls in terms of the exercise price, we see that

$$\mathrm{call}(t, S_t, \tau, K) \;\le\; \mathrm{Call}(t, S_t, \tau, K) \quad \le \quad S_t$$

Moreover, when the underlying security pays no dividends before maturity, we have the following lower bound on call options prices:

$$\mathrm{Call}(t, S_t, T, K) \;\ge\; \mathrm{call}(t, S_t, T, K) \quad \ge \quad (S_t - KP_t(T))^+\,.$$

Indeed, consider the portfolio X consisting of a long position of a European call, a long position of K $T-$maturity zero-coupon bonds, and a short position of one share of the underlying security. The required result follows from the observation that the final value at the maturity of the portfolio is non-negative, and the application of the dominance principle.

2. *Assume that interest rates are positive. Then, an American call on a security that pays no dividend before the maturity of the call will never be exercised early.*

Indeed, let u be an arbitrary instant in $[t, T)$,
- the American call pays $S_u - K$ if exercised at time u,
- but $S_u - K < S - KP_u(T)$ because interest rates are positive.
- Since Call$(S_u, u, K) \geq S_u - KP_u(T)$, by the lower bound, the American option is always worth more than its exercise value, so early exercise is never optimal.

3. *Assume that the security price takes values as close as possible to zero. Then, early exercise of American put options may be optimal before maturity.*

Indeed, suppose the security price at some time u falls so deeply that $S_u < K - KP_u(T)$.
- Observe that the maximum value that the American put can deliver when if exercised at maturity is K.
- The immediate exercise value at time u is $K - S_u > K - [K - KP_u(T)] = KP_u(T) \equiv$ the discounted value of the maximum amount that the put could pay if held to maturity,

Hence, in this case waiting until maturity to exercise is never optimal.

1.4.1 Risk effect on options prices

1 *The value of a portfolio of European/american call/put options, with common strike and maturity, always exceeds the value of the corresponding basket option.*

Indeed, let $S^1, \ldots, S^n$ be the prices of n security, and consider the portfolio composition $\lambda^1, \ldots, \lambda^n \geq 0$. By sublinearity of the maximum,

$$\sum_{i=1}^{n} \lambda^i \max\left\{S_u^i - K, 0\right\} \quad \geq \quad \max\left\{\sum_{i=1}^{n} \lambda^i S_u^i - K, 0\right\}$$

i.e. if the portfolio of options is exercised on the optimal exercise date of the option on the portfolio, the payoff on the former is never less than that on the latter. By the dominance principle, this shows that the portfolio of options is more valuable than the corresponding basket option.

2 For a security with spot price S_t and price at maturity S_T, we denote its return by

$$R_t(T) \quad := \quad \frac{S_T}{S_t}\,.$$

Definition Let $R_t^i(T)$, $i = 1, 2$ be the return of two securities. We say that security 2 is more risky than security 1 if

$$R_t^2(T) \;=\; R_t^1(T) + \varepsilon \quad \text{with} \quad \mathbb{E}\left[\varepsilon | R_t^1(T)\right] = 0\,.$$

As a consequence, if security 2 is more risky than security 1, the above definition implies that

$$\begin{aligned} \mathbb{V}ar\left[R_t^2(T)\right] &= \mathbb{V}ar\left[R_t^1(T)\right] + \mathbb{V}ar[\varepsilon] + 2\mathbb{C}ov[R_t^1(T), \varepsilon] \\ &= \mathbb{V}ar\left[R_t^1(T)\right] + \mathbb{V}ar[\varepsilon] \;\geq\; \mathbb{V}ar\left[R_t^1(T)\right] \end{aligned}$$

3 We now assume that the pricing functional is continuous in some sense to be precised below, and we show that *the value of an European/American call/put is increasing in its riskiness.*

To see this, let $R := R_t(T)$ be the return of the security, and consider the set of riskier securities with returns $R^i := R_t^i(T)$ defined by

$$R^i = R + \varepsilon_i \quad \text{where} \quad \varepsilon_i \text{ are iid and } \mathbb{E}[\varepsilon_i|R] = 0\,.$$

Let $\mathrm{Call}^i(t, S_t, T, K)$ be the price of the American call option with payoff $\left(S_t R^i - K\right)^+$, and $\overline{\mathrm{Call}}_n(t, S_t, T, K)$ be the price of the basket option defined by the payoff $\left(\frac{1}{n}\sum_{i=1}^n S_t R^i - K\right)^+ = \left(S_T + \frac{1}{n}\sum_{i=1}^n S_t\varepsilon_i - K\right)^+$.

We have previously seen that the portfolio of options with common maturity and strike is worth more than the corresponding basket option:

$$\mathrm{Call}^1(t, S_t, T, K) \;=\; \frac{1}{n}\sum_{i=1}^n \mathrm{Call}^i(t, S_t, T, K) \;\geq\; \overline{\mathrm{Call}}_n(t, S_t, T, K).$$

Observe that the final payoff of the basket option $\overline{\mathrm{Call}}_n(T, S_T, T, K) \longrightarrow (S_T - K)^+$ a.s. as $n \to \infty$ by the law of large numbers. Then assuming that the pricing functional is continuous, it follows that $\overline{\mathrm{Call}}_n(t, S_t, T, K) \longrightarrow \mathrm{Call}(t, S_t, T, K)$, and therefore: that

$$\mathrm{Call}^1(t, S_t, T, K) \;\geq\; \mathrm{Call}(t, S_t, T, K).$$

Notice that the result holds due to the convexity of European/American call/put options payoffs.

1.5 Some popular examples of contingent claims

Example 1.1. *(Basket call and put options)* Given a subset I of indices in $\{1, \ldots, n\}$ and a family of positive coefficients $(a_i)_{i\in I}$, the payoff of a Basket call (resp. put) option is defined by

$$B \;=\; \left(\sum_{i\in I} a_i S_T^i - K\right)^+ \qquad \text{resp.} \quad \left(K - \sum_{i\in I} a_i S_T^i\right)^+ .$$

$\diamondsuit$

Example 1.2. *(Option on a non-tradable underlying variable)* Let $U_t(\omega)$ be the time t realization of some observable state variable. Then the payoff of a call (resp. put) option on U is defined by

$$B \;=\; (U_T - K)^+ \qquad \text{resp.} \;\; (K - U_T)^+ \;.$$

For instance, a *Temperature call option* corresponds to the case where U_t is the temperature at time t observed at some location (defined in the contract).

Example 1.3. *(Asian option)*An Asian call option on the asset S^i with maturity $T > 0$ and strike $K > 0$ is defined by the payoff at maturity:

$$\left(\overline{S}_T^i - K\right)^+ ,$$

where $\overline{S}_T^i$ is the average price process on the time period $[0, T]$. With this definition, there is still choice for the type of Asian option in hand. One can define $\overline{S}_T^i$ to be the arithmetic mean over of given finite set of dates (outlined in the contract), or the continuous arithmetic mean...

Example 1.4. *(Barrier call options)* Let $B, K > 0$ be two given parameters, and $T > 0$ given maturity. There are four types of barrier call options on the asset S^i with stike K, barrier B and maturity T:

- When $B > S_0$:
  - an *Up and Out Call option* is defined by the payoff at the maturity T:

    $$\mathrm{UOC}_T \quad = \quad (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t \leq B\}}.$$

    The payoff is that of a European call option if the price process of the underlying asset never reaches the barrier B before maturity. Otherwise it is zero (the contract knocks out).
  - an *Up and In Call option* is defined by the payoff at the maturity T:

    $$\mathrm{UIC}_T \quad = \quad (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t > B\}}.$$

    The payoff is that of a European call option if the price process of the underlying asset crosses the barrier B before maturity. Otherwise it is zero (the contract knocks out). Clearly,

    $$\mathrm{UOC}_T + \mathrm{UIC}_T = \mathrm{call}_T$$

    is the payoff of the corresponding European call option.
- When $B < S_0$:
  - an *Down and In Call option* is defined by the payoff at the maturity T:

    $$\mathrm{DIC}_T \quad = \quad (S_T - K)^+ \mathbf{1}_{\{\min_{[0,T]} S_t \geq B\}}.$$

    The payoff is that of a European call option if the price process of the underlying asset never reaches the barrier B before maturity. Otherwise it is zero (the contract knocks out).

– an *Down and Out Call option* is defined by the payoff at the maturity T:

$$\mathrm{DOC}_T = (S_T - K)^+ \mathbf{1}_{\{\min_{[0,T]} S_t < B\}}.$$

The payoff is that of a European call option if the price process of the underlying asset crosses the barrier B before maturity. Otherwise it is zero (the contract knocks out). Clearly,

$$\mathrm{DOC}_T + \mathrm{DIC}_T = \mathrm{call}_T$$

is the payoff of the corresponding European call option.

$\diamondsuit$

Example 1.5. *(Barrier put options)* Replace calls by puts in the previous example

Chapter 2

A first approach to the Black-Scholes formula

2.1 The single period binomial model

We first study the simplest one-period financial market $T = 1$. Let $\Omega = \{\omega_u, \omega_d\}$, $\mathcal{F}$ the σ-algebra consisting of all subsets of Ω, and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$ such that $0 < \mathbb{P}(\omega_u) < 1$.

The financial market contains a non-risky asset with price process

$$S_0^0 \ = \ 1 \ , \qquad S_1^0(\omega_u) \ = \ S_1^0(\omega_d) \ = \ e^r \ ,$$

and one risky asset ($d = 1$) with price process

$$S_0 \ = \ s \ , \qquad S_1(\omega_u) \ = \ su \ , \ \ S_1(\omega_d) \ = \ sd \ ,$$

where s, r, u and d are given strictly positive parameters with $u > d$. Such a financial market can be represented by the binomial tree :

$$\begin{tikzcd}
 & \text{time } 0 & \text{time } 1 \\
 & & Su \\
\text{Risky asset} & S_0 = s \arrow[ur] \arrow[dr] & \\
 & & Sd \\
\text{Non-risky asset} & S_0^0 = 1 \arrow[r] & R = e^r
\end{tikzcd}$$

In the terminology of the Introduction Section 1, the above model is the simplest *wrong model* which illustrates the main features of the valuation theory in financial mathematics.

The discounted prices are defined by the value of the prices relative to the nonrisky asset price, and are given by

$$\tilde{S}_0 := S_0, \ \ \tilde{S}_0^0 := 1, \quad \text{and} \quad \tilde{S}_1 := \frac{S_1}{R}, \ \ \tilde{S}_1^0 := 1.$$

A self-financing trading strategy is a pair $(x, \theta) \in \mathbb{R}^2$ where x is an initial capital and θ is the number of shares of risky asset that the investor chooses to hold over the time period $[0, 1]$. The corresponding wealth process at time 1 is given by :

$$X_1^{x,\theta} \quad := \quad (x - \theta S_0)R + \theta S_1,$$

or, in terms of discounted value

$$\tilde{X}_1^{x,\theta} \quad := \quad x + \theta(\tilde{S}_1 - \tilde{S}_0).$$

(i) The No-Arbitrage condition : An arbitrage opportunity is a portfolio strategy $\theta \in \mathbb{R}$ such that

$$X_1^{0,\theta}(\omega_i) \geq 0, \ i \in \{u, d\}, \quad \text{and} \quad \mathbb{P}[X_1^{0,\theta} > 0] > 0.$$

It can be shown that excluding all arbitrage opportunities is equivalent to the condition

$$d < R < u. \tag{2.1}$$

Exercise 2.1. *In the context of the present one-period binomial model, prove that the no-arbitrage condition is equivalent to* (2.1)*.*

Under the no-arbitrage condition (2.1), we ay introduce the equivalent[1] probability measure $\mathbb{Q}$ defined by

$$\mathbb{Q}[S_1 = uS_0] \ = \ 1 - \mathbb{Q}[S_1 = dS_0] \ = \ q \quad := \quad \frac{R-d}{u-d}, \tag{2.2}$$

Then, we see that the discounted price process satisfies

$$\tilde{S} \quad \text{is a martingale under } \mathbb{Q}\text{, i.e.} \quad \mathbb{E}^{\mathbb{Q}}[\tilde{S}_1] = S_0. \tag{2.3}$$

The probability measure $\mathbb{Q}$ is called *risk-neutral measure*, or *equivalent martingale measure*.

(ii) Hedging contingent claims : A contingent claim is defined by its payoff $B_u := B(\omega_u)$ and $B_d := B(\omega_d)$ at time 1.

In the context of the binomial model, it turns out that there exists a pair $(x^0, \theta^0) \in \mathbb{R} \times \mathcal{A}$ such that $X_T^{x^0,\theta^0} = B$. Indeed, the equality $X_1^{x^0,\theta^0} = B$ is a system of two (linear) equations with two unknowns which can be solved straightforwardly :

$$x^0(B) \ = \ q\frac{B_u}{R} + (1-q)\frac{B_d}{R} \ = \ \mathbb{E}^{\mathbb{Q}}[\tilde{B}] \quad \text{and} \quad \theta^0(B) \ = \ \frac{B_u - B_d}{su - sd}.$$

The portfolio $(x^0(B), \theta^0(B))$ satisfies $X_T^{x^0,\theta^0} = B$, and is therefore called a *perfect replication* strategy for B.

[1]Two probability measures are equivalent if they have the same zero-measure events.

(iii) No arbitrage valuation : Suppose that the contingent claim B is available for trading at time 0 with market price $p(B)$, and let us show that, under the no-arbitrage condition, the price $p(B)$ of the contingent claim contract is necessarily given by

$$p(B) \quad = \quad x^0(B) \; = \; \mathbb{E}^{\mathbb{Q}}[\tilde{B}].$$

(iii-a) Indeed, suppose that $p(B) < x^0(B)$, and consider the following portfolio strategy :

- at time 0, pay $p(B)$ to buy the contingent claim, so as to receive the payoff B at time 1,

- perform the self-financing strategy $(-x^0, -\theta)$, this leads to paying $-x^0$ at time 0, and receiving $-B$ at time 1.

The initial capital needed to perform this portfolio strategy is $p(B) - x^0 < 0$. At time 1, the terminal wealth induced by the self-financing strategy exactly compensates the payoff of the contingent claim. We have then built an arbitrage opportunity in the financial market augmented with the contingent claim contract, thus violating the no-arbitrage condition on this enlarged financial market.

(iii-b) If $p(B) > x^0(B)$, we consider the following portfolio strategy :

- at time 0, receive $p(B)$ by selling the contingent claim, so as to pay the payoff B at time 1,

- perform the self-financing strategy (x^0, θ), this leads to paying x^0 at time 0, and receiving B at time 1.

The initial capital needed to perform this portfolio strategy is $-p(B) + x^0 < 0$. At time 1, the terminal wealth induced by the self-financing strategy exactly compensates the payoff of the contingent claim. This again defines an arbitrage opportunity in the financial market augmented with the contingent claim contract, thus violating the no-arbitrage condition on this enlarged financial market.

2.2 The Cox-Ross-Rubinstein model

In this section, we present a dynamic version of the previous binomial model.

Let $\Omega = \{-1, 1\}^{\mathbb{N}}$, and let $\mathcal{F}$ be the Borel σ-algebra on Ω. Let $(Z_k)_{k \geq 0}$ be a sequence of independent random variables with distribution $\mathbb{P}[Z_k = 1] = \mathbb{P}[Z_k = -1] = 1/2$. We shall see later that we may replace the value 1/2 by any parameter in $(0, 1)$, see Remark 2.4). We consider the trivial filtration $\mathcal{F}_0 = \{\emptyset, \mathcal{F}\}$, $\mathcal{F}_k = \sigma(Z_0, \ldots, Z_k)$ and $\mathbb{F}^n = \{\mathcal{F}_0, \ldots, \mathcal{F}_n\}$.

Let $T > 0$ be some fixed finite horizon, and $(b_n, \sigma_n)_{n \geq 1}$ the sequence defined by :

$$b_n \; = \; b \, \frac{T}{n} \quad \text{and} \quad \sigma_n \; = \; \sigma \, \left(\frac{T}{n} \right)^{1/2} ,$$

where b and σ are two given strictly positive parameters.

Remark 2.2. All the results of this section hold true with a sequence (b_n, σ_n) satisfying :

$$nb_n \longrightarrow bT \quad \text{and} \quad \sqrt{n}\sigma_n \longrightarrow \sigma\sqrt{T} \quad \text{whenever } n \to \infty .$$

For $n \geq 1$, we consider the price process of a single risky asset $S^n = \{S^n_k, k = 0, \ldots, n\}$ defined by :

$$S^n_0 = s \quad \text{and} \quad S^n_k = s \exp\left(kb_n + \sigma_n \sum_{i=1}^{k} Z_i\right), \quad k = 1, \ldots, n .$$

The non-risky asset is defined by a constant interest rate parameter r, so that the return from a unit investment in the bank during a period of length T/n is

$$R_n := e^{r(T/n)} .$$

For each $n \geq 1$, we have then defined a financial market with time step T/n.

In order to ensure that these financial markets satisfy the no-arbitrage condition, we assume that :

$$d_n < R_n < u_n \quad \text{where} \quad u_n = e^{b_n+\sigma_n}, \quad d_n = e^{b_n-\sigma_n}, \quad n \geq 1 . \tag{2.4}$$

Under this condition, the risk-neutral measure $\mathbb{Q}_n$ defined by :

$$\mathbb{Q}_n[Z_i = 1] = q_n := \frac{R_n - d_n}{u_n - d_n} .$$

2.3 Valuation and hedging in the Cox-Ross-Rubinstein model

Consider the contingent claims

$$B^n := g(S^n_n) \quad \text{where} \quad g(s) = (s - K)^+ \text{ and } K > 0 .$$

At time $n-1$, we are facing a binomial model, and we can therefore conclude from our previous discussions that the no-arbitrage market price of this contingent claim at time $n-1$ and the corresponding perfect hedging strategy are given by

$$B^n_{n-1} := \mathbb{E}^{\mathbb{Q}_n}_{n-1}[\tilde{B}^n] \quad \text{and} \quad \theta^n_{n-1}(\omega_{n-1}) = \frac{B^n(\omega_{n-1}, u_n) - B^n(\omega_{n-1}, d_n)}{u_n S^n_{n-1}(\omega_{n-1}) - d_n S^n_{n-1}(\omega_{n-1})},$$

where $\omega_{n-1} \in \{d_n, u_n\}^{n-1}$, and $\mathbb{E}^{\mathbb{Q}_n}_{n-1}$ denotes the expectation operator under $\mathbb{Q}_n$ conditional on the information at time $n-1$. Arguying similarly step by step, backward in time, we may define the contingent claim B^n_k at each time step k as the no-arbitrage market price of the contingent claim B^n_{k+1} and the corresponding perfect hedging strategy:

$$B^n_k := \mathbb{E}^{\mathbb{Q}_n}_k[\tilde{B}^n_{k+1}] \quad \text{and} \quad \theta^n_k(\omega_k) = \frac{B^n_{k+1}(\omega_k, u_n) - B^n_{k+1}(\omega_k, d_n)}{u_n S^n_k(\omega_k) - d_n S^n_k(\omega_k)}, \quad k = 0, \ldots, n-1.$$

Remark 2.3. The hedging strategy is the finite differences approximation (on the binomial tree) of the partial derivative of the price of the contingent claim with respect to the spot price of the underlying asset. This observation will be confirmed in the continuous-time framework.

By the law of iterated expectations, we conclude that the no-arbitrage price of the European call option is :

$$p^n(B^n) \quad := \quad e^{-rT}\mathbb{E}^{\mathbb{Q}_n}\left[(S_n^n - K)^+\right] .$$

Under the probability measure $\mathbb{Q}_n$, the random variables $(1+Z_i)/2$ are independent and identically distributed as a Bernoulli with parameter q_n. Then :

$$\mathbb{Q}_n\left[\sum_{i=1}^{n}\frac{1+Z_i}{2} \;=\; j\right] \;=\; C_n^j q_n^j(1-q_n)^{n-j} \quad \text{for} \quad j=0,\ldots,n .$$

This provides

$$p^n(B^n) \quad = \quad e^{-rT}\sum_{j=0}^{n} g\left(su_n^j d_n^{n-j}\right) C_n^j q_n^j (1-q_n)^{n-j} .$$

Remark 2.4. The reference measure $\mathbb{P}$ is not involved neither in the valuation formula, nor in the hedging formula. This is due to the fact that the no-arbitrage price in the present framework coincides with the perfect replication cost, which in turn depends on the reference measure only through the corresponding zero-measure sets.

2.4 Continuous-time limit

In this paragraph, we examine the asymptotic behavior of the Cox-Ross-Rubinstein model when the time step T/n tends to zero, i.e. when $n \longrightarrow \infty$. Our final goal is to show that the limit of the discrete-time valuation formulae coincides with the Black-Scholes formula which was originally derived in [7] in the continuous-time setting.

Although the following computations are performed in the case of European call options, the convergence argument holds for a large class of contingent claims.

Introduce the sequence :

$$\eta_n \quad := \quad \inf\{j=0,\ldots,n \;:\; su_n^j d_n^{n-j} \geq K\} ,$$

and let

$$B(n,p,\eta) \quad := \quad \text{Prob}\left[\text{Bin}(n,p) \;\geq\; \eta\right] ,$$

where $\text{Bin}(n,p)$ is a Binomial random variable with parameters (n,p).

The following Lemma provides an interesting reduction of our problem.

Lemma 2.5. *For $n \geq 1$, we have :*

$$p^n(B^n) \quad = \quad sB\left(n, \frac{q_n u_n}{R_n}, \eta_n\right) - Ke^{-rT}B(n, q_n, \eta_n)\ .$$

Proof. Using the expression of $p^n(B^n)$ obtained in the previous paragraph, we see that

$$\begin{aligned} p^n(B^n) \quad &= \quad R_n^{-n} \sum_{j=\eta_n}^{n} \left(su_n^j d_n^{n-j} - K\right) C_n^j q_n^j (1-q_n)^{n-j} \\ &= \quad s \sum_{j=\eta_n}^{n} C_n^j \left(\frac{q_n u_n}{R_n}\right)^j \left(\frac{(1-q_n)d_n}{R_n}\right)^{n-j} - \frac{K}{R_n^n} \sum_{j=\eta_n}^{n} C_n^j q_n^j (1-q_n)^{n-j}\ . \end{aligned}$$

The required result follows by noting that $q_n u_n + (1-q_n)d_n = R_n$. ♢

Hence, in order to derive the limit of $p^n(B^n)$ when $n \longrightarrow \infty$, we have to determine the limit of the terms $B\left(n, q_n u_n/R_n, \eta_n\right)$ et $B(n, q_n, \eta_n)$. We only provide a detailed exposition for the second term ; the first one is treated similarly.

The main technical tool in order to obtain these limits is the following.

Lemma 2.6. *Let $(X_{k,n})_{1\leq k\leq n}$ be a triangular sequence of iid Bernoulli ramdom variables with parameter π_n:*

$$\mathbb{P}[X_{k,n} = 1] \ = \ 1 - \mathbb{P}[X_{k,n} = 0] \quad = \quad \pi_n\ .$$

Then:

$$\frac{\sum_{k=1}^n X_{k,n} \ - \ n\pi_n}{\sqrt{n\pi_n(1-\pi_n)}} \quad \longrightarrow \quad \mathcal{N}(0,1) \qquad \textit{in distribution}\ .$$

The proof of this lemma is reported at the end of this paragraph.

Exercise 2.7. Use Lemma 2.6 to show that

$$\ln\left(\frac{S_n^n}{s}\right) \quad \longrightarrow \quad \mathcal{N}(bT, \sigma^2 T) \qquad \text{in distribution under } \mathbb{P}\ .$$

This shows that the Cox-Ross-Rubinstein model is a discrete-time approximation of a continuous-time model where the risky asset price has a log-normal distribution.

Theorem 2.8. *In the context of the Cox-Ross-Rubinstein model, the no-arbitrage price $p^n(B^n)$ of a European call option converges, as $n \to \infty$, to the Black-Scholes price :*

$$p(B) \quad = \quad s\,\mathbf{N}\left(\mathbf{d}_+(s, \tilde{K}, \sigma^2 T)\right) \ - \ \tilde{K}\,\mathbf{N}(\mathbf{d}_-\left(s, \tilde{K}, \sigma^2 T)\right)$$

where

$$\tilde{K} := Ke^{-rT}, \qquad \mathbf{d}_{\pm}(s,k,v) := \frac{\ln(s/k)}{\sqrt{v}} \pm \frac{\sqrt{v}}{2},$$

and $\mathbf{N}(x) = \int_{-\infty}^{x} e^{-v^2/2}dv/\sqrt{2\pi}$ *is the cumulative distribution function of standard Gaussian* $\mathcal{N}(0,1)$.

Proof. Under the probability measure $\mathbb{Q}_n$, notice that $B_i := (Z_i+1)/2$, $i \geq 0$, defines a sequence of iid Bernoulli random variables with parameter q_n. Then

$$B(n,q_n,\eta_n) \quad = \quad \mathbb{Q}_n\left[\sum_{j=1}^{n} B_i \ \geq \ \eta_n\right].$$

We shall only develop the calculations for this term.
1. By definition of η_n, we have

$$su_n^{\eta_n-1}d_n^{n-\eta_n+1} \ \leq \ K \ \leq \ su_n^{\eta_n}d_n^{n-\eta_n}.$$

Then,

$$2\eta_n\sigma\sqrt{\frac{T}{n}} + n\left(b\frac{T}{n} - \sigma\sqrt{\frac{T}{n}}\right) \quad = \quad \ln\left(\frac{K}{s}\right) + O\left(n^{-1/2}\right),$$

which provides, by direct calculation that

$$\eta_n \quad = \quad \frac{n}{2} \ + \ \sqrt{n}\,\frac{\ln(K/s) - bT}{2\sigma\sqrt{T}} \ + \ \circ(\sqrt{n}). \tag{2.5}$$

We also compute that

$$nq_n \quad = \quad \frac{1}{2} + \frac{\left(r - b - \frac{\sigma^2}{2}\right)T}{2\sigma\sqrt{T}}\,\sqrt{n} + \circ(\sqrt{n}). \tag{2.6}$$

By (2.5) and (2.6), it follows that

$$\lim_{n\to\infty} \frac{\eta_n - nq_n}{\sqrt{nq_n(1-q_n)}} \quad = \quad -\mathbf{d}_-(s,\tilde{K},\sigma^2T).$$

2. Applying Lemma 2.6 to the sequence $(Z_1,\ldots,Z_n)$, we see that :

$$\mathcal{L}^{\mathbb{Q}_n}\left(\frac{\frac{1}{2}\sum_{k=1}^{n}(1+Z_j) \ - \ nq_n}{\sqrt{nq_n(1-q_n)}}\right) \quad \longrightarrow \quad \mathcal{N}(0,1),$$

where $\mathcal{L}^{\mathbb{Q}_n}(X)$ denotes the distribution under $\mathbb{Q}_n$ of the random variable X. Then :

$$\begin{aligned}
\lim_{n\to\infty} B(n,q_n,\eta_n) &= \lim_{n\to\infty} \mathbb{Q}_n\left[\frac{\frac{1}{2}\sum_{k=1}^{n}(1+Z_j) - nq_n}{\sqrt{nq_n(1-q_n)}} \geq \frac{\eta_n - nq_n}{\sqrt{nq_n(1-q_n)}}\right] \\
&= 1 - \mathbf{N}\left(-\mathbf{d}_-(s,\tilde{K},\sigma^2T)\right) \ = \ \mathbf{N}\left(\mathbf{d}_-(s,\tilde{K},\sigma^2T)\right).
\end{aligned}$$

$\diamondsuit$

Proof of Lemma 2.6. (i) We start by recalling a well-known result on characteristic functions (see Exercise B.9). Let X be a random variable with $\mathbb{E}[X^n] < \infty$. Then :

$$\Phi_X(t) \;:=\; \mathbb{E}\left[e^{itX}\right] \;=\; \sum_{k=0}^{n} \frac{(it)^k}{k!}\mathbb{E}[X^k] \;+\; \circ(t^n)\;. \qquad (2.7)$$

To prove this result, we denote $F(t,x) := e^{itx}$ and $f(t) = \mathbb{E}[F(t,X)]$. The function $t \longmapsto F(t,x)$ is differentiable with respect to the t variable. Since $|F_t(t,X)| = |iXF(t,X)| \le |X| \in \mathbb{L}^1$, it follows from the dominated convergence theorem that the function f is differentiable with $f'(t) = E[iXe^{itX}]$. In particular, $f'(0) = i\mathbb{E}[X]$. Iterating this argument, we see that the function f is n times differentiable with $n-$th order derivative at zero given by :

$$f^{(n)}(0) \;=\; i^n\mathbb{E}[X^n]\;.$$

The expansion (2.7) is an immediate consequence of the Taylor-Young formula.
(ii) We now proceed to the proof of Lemma 2.6. Let

$$Y_j \;:=\; \frac{X_{j,n}-\pi_n}{\sqrt{n\pi_n(1-\pi_n)}} \quad \text{and} \quad \Sigma Y_n \;:=\; \sum_{k=1}^{n} Y_j\;.$$

Since the random variables Y_j are independent and identically distributed, the characteristic function $\Phi_{\Sigma Y_n}$ of ΣY_n factors in terms of the common characteristic function Φ_{Y_1} of the Y_i's as :

$$\Phi_{\Sigma Y_n}(t) \;=\; \left(\Phi_{Y_1}(t)\right)^n\;.$$

Moreover, we compute directly that $E[Y_j] = 0$ and $E[Y_j^2] = 1/n$. Then, it follows from (2.7) that :

$$\Phi_{Y_1}(t) \;=\; 1 - \frac{t^2}{2n} + \circ\left(\frac{1}{n}\right)\;.$$

Sending n to ∞, this provides :

$$\lim_{n\to\infty}\Phi_{\Sigma Y_n}(t) \;=\; e^{-t^2/2} \;=\; \Phi_{\mathcal{N}(0,1)}(t)\;.$$

This shows the convergence in distribution of ΣY_n towards the standard normal distribution. $\diamondsuit$

Chapter 3

Some preliminaries on continuous-time processes

3.1 Filtration and stopping times

Throughout this chapter, $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space.

A *stochastic process* with values in a set E is a map

$$\begin{array}{rcl} V \ : \ \mathbb{R}_+ \times \Omega & \longrightarrow & E \\ (t,\omega) & \longmapsto & V_t(\omega) \end{array}$$

The index t is conveniently interpreted as the time variable. In the context of these lectures, the *state space* E will be a subset of a finite dimensional space, and we shall denote by $\mathcal{B}(E)$ the corresponding Borel $\sigma-$field. The process V is said to be measurable if the mapping

$$\begin{array}{rcl} V \ : \ (\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}) & \longrightarrow & (E, \mathcal{B}(E)) \\ (t,\omega) & \longmapsto & V_t(\omega) \end{array}$$

is measurable. For a fixed $\omega \in \Omega$, the function $t \in \mathbb{R}_+ \longmapsto V_t(\omega)$ is the sample path (or trajectory) of V corresponding to ω.

3.1.1 Filtration

A *filtration* $\mathbb{F} = \{\mathcal{F}_t,\ t \geq 0\}$ is an increasing family of sub-$\sigma-$algebras of $\mathcal{F}$. Similar to the discrete-time context, $\mathcal{F}_t$ is intuitively understood as the information available up to time t. The increasing feature of the filtration, $\mathcal{F}_s \subset \mathcal{F}_t$ for $0 \leq s \leq t$, means that information can only increase as time goes on.

Definition 3.1. *A stochastic process V is said to be*
(i) *adapted to the filtration $\mathbb{F}$ if the random variable V_t is $\mathcal{F}_t-$measurable for*

every $t \in \mathbb{R}_+$,
(ii) *progressively measurable with respect to the filtration* $\mathbb{F}$ *if the mapping*

$$\begin{array}{rcl} V \ : \ ([0,t] \times \Omega, \mathcal{B}([0,t]) \otimes \mathcal{F}_t) & \longrightarrow & (E, \mathcal{B}(E)) \\ (s,\omega) & \longmapsto & V_s(\omega) \end{array}$$

is measurable for every $t \in \mathbb{R}_+$.

Given a stochastic process V, we define its *canonical filtration* by $\mathcal{F}_t^V := \sigma(V_s, s \le t)$, $t \in \mathbb{R}_+$. This is the smallest filtration to which the process V is adapted.

Obviously, any progressively measurable stochastic process is measurable and adapted. The following result states that these two notions are equivalent for processes which are either right-continuous or left-continuous.

Proposition 3.2. *Let* V *be a stochastic process with right-continuous sample paths or else left-continuous sample paths. Then, if* V *is adapted to a filtration* $\mathbb{F}$*, it is also progressively measurable with respect to* $\mathbb{F}$.

Proof. Assume that every sample path of V is right-continuous (the case of left-continuous sample paths is treated similarly), and fix an arbitraty $t \ge 0$. Observe that $V_s(\omega) = \lim_{n\to\infty} V_s^n(\omega)$ for every $s \in [0,t]$, where V^n is defined by

$$V_s^n(\omega) = V_{kt/n}(\omega) \quad \text{for} \quad (k-1)t < sn \le kt \text{ and } k = 1, \ldots, n\,.$$

Since the restriction of the map V^n to $[0,t]\times\Omega$ is obviously $\mathcal{B}([0,t])\otimes\mathcal{F}_t-$measurable, we deduce the measurability of the limit map V defined on $[0,t] \times \Omega$. ♢

3.1.2 Stopping times

A *random time* is a random variable τ with values in $[0,\infty]$. It is called
- a *stopping time* if the event set $\{\tau \le t\}$ is in $\mathcal{F}_t$ for every $t \in \mathbb{R}_+$,
- an *optional time* if the event set $\{\tau < t\}$ is in $\mathcal{F}_t$ for every $t \in \mathbb{R}_+$.

Obviously, any stopping time is an optional time. It is an easy exercise to show that these two notions are in fact identical whenever the filtration $\mathbb{F}$ is right-continuous, i.e.

$$\mathcal{F}_{t+} \ := \ \cap_{s>t}\mathcal{F}_s \ = \ \mathcal{F}_t \quad \text{for every} \quad t \ge 0\,.$$

This will be the case in all of the financial applications of this course. An important example of a stopping time is:

Exercise 3.3. *(first exit time)* Let V be a stochastic process with continuous paths adapted to $\mathbb{F}$, and consider a closed subset $\Gamma \in \mathcal{B}(E)$ together with the random time

$$T_\Gamma \ := \ \inf\{t \ge 0 \ : \ X_t \notin \Gamma\}\,,$$

with the convention $\inf\emptyset = \infty$. Show that if Γ is closed (resp. open), then T_Γ is a stopping time (resp. optional time).

Proposition 3.4. *Let τ_1 and τ_2 be two stopping times. Then so are $\tau_1 \wedge \tau_2$, $\tau_1 \vee \tau_2$, and $\tau_1 + \tau_2$, .*

Proof. For all $t \geq 0$, we have $\{\tau_1 \wedge \tau_2 \leq t\} = \{\tau_1 \leq t\} \cup \{\tau_2 \leq t\} \in \mathcal{F}_t$, and $\{\tau_1 \vee \tau_2 \leq t\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\} \in \mathcal{F}_t$. This proves that $\tau_1 \wedge \tau_2$, $\tau_1 \vee \tau_2$ are stopping times. Finally:

$$\begin{aligned}
\{\tau_1 + \tau_2 > t\} \quad = \quad & \big(\{\tau_1 = 0\} \cap \{\tau_2 > t\}\big) \cup \big(\{\tau_1 > t\} \cap \{\tau_2 = 0\}\big) \\
& \cup \big(\{\tau_1 \geq t\} \cap \{\tau_2 > 0\}\big) \\
& \cup \big(\{0 < \tau_1 < t\} \cap \{\tau_1 + \tau_2 > t\}\big).
\end{aligned}$$

Notice that the first three events sets are obviously in $\mathcal{F}_t$. As for the fourth one, we rewrite it as

$$\{0 < \tau_1 < t\} \cap \{\tau_1 + \tau_2 > t\} = \bigcup_{r \in (0,t) \cap \mathbb{Q}} \big(\{r < \tau_1 < t\} \cap \{\tau_2 > t - r\}\big) \in \mathcal{F}_t.$$

$\diamondsuit$

Exercise 3.5. (i) If $(\tau_n)_{n \geq 1}$ is a sequence of optional times, then so are $\sup_{n \geq 1} \tau_n$, $\inf_{n \geq 1} \tau_n$, $\limsup_{n \to \infty} \tau_n$, $\liminf_{n \to \infty} \tau_n$.
(ii) If $(\tau_n)_{n \geq 1}$ is a sequence of stopping times, then so is $\sup_{n \geq 1} \tau_n$.

Given a stopping time τ with values in $[0, \infty]$, we shall frequently use the approximating sequence

$$\tau_n \ := \ \frac{\lfloor n\tau \rfloor + 1}{n} \mathbf{1}_{\{\tau < \infty\}} + \infty \mathbf{1}_{\{\tau = \infty\}}, \quad n \geq 1, \tag{3.1}$$

which defines a decreasing sequence of stopping times converging a.s. to τ. Here $\lfloor t \rfloor$ denotes the largest integer less than or equal to t. Notice that the random time $\frac{\lfloor n\tau \rfloor}{n}$ is not a stopping time in general.

The following example is a complement to Exercise (3.5).

Exercise 3.6. Let τ be a finite optional time, and consider the sequence $(\tau_n)_{n \geq 1}$ defined by (3.1). Show that τ_n is a stopping time for all $n \geq 1$.

As in the discrete-time framework, we provide a precise definition of the information available up to some stopping time τ of a filtration $\mathbb{F}$:

$$\mathcal{F}_\tau \quad := \quad \{A \in \mathcal{F} \ : \ A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for every } t \in \mathbb{R}_+\}.$$

Definition 3.7. *Let τ be a stopping time, and V a stochastic process. We denote*

$$V_\tau(\omega) := V_{\tau(\omega)}(\omega), \quad V_t^\tau := V_{t \wedge \tau} \quad \textit{for all } t \geq 0,$$

and we call V^τ the process V stopped at τ.

Proposition 3.8. *Let $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ be a filtration, τ an $\mathbb{F}-$stopping time, and V a progressively measurable stochastic process. Then*
(i) *$\mathcal{F}_\tau$ is a $\sigma-$algebra,*
(ii) *V_τ is $\mathcal{F}_\tau-$measurable, and the stopped process $\{V_t^\tau, t \geq 0\}$ is progressively measurable.*

Proof. (i) First, for all $t \geq 0$, $\Omega \cap \{\tau \leq t\} = \{\tau \leq t\} \in \mathcal{F}_t$ proving that $\Omega \in \mathcal{F}_\tau$.

Next, for any $A \in \mathcal{F}_\tau$, we have $A^c \cap \{\tau \leq t\} = \{\tau \leq t\} \cap (A \cap \{\tau \leq t\})^c$. Since τ is a stopping time, $\{\tau \leq t\} \in \mathcal{F}_t$. Since $A \cap \{\tau \leq t\} \in \mathcal{F}_t$, its complement is in $\mathcal{F}_t$, and we deduce that the intersection $\{\tau \leq t\} \cap (A \cap \{\tau \leq t\})^c$ is in $\mathcal{F}_t$. Since this holds true for any $t \geq 0$, this shows that $A^c \in \mathcal{F}_\tau$.

Finally, for any countable family $(A_i)_{i \geq 1} \subset \mathcal{F}_\tau$, we have $(\cup_{i\geq 1} A_i) \cap \{\tau \leq t\} = \cup_{i \geq 1}(A_i \cap \{\tau \leq t\}) \in \mathcal{F}_t$ proving that $\cup_{i\geq 1} A_i \in \mathcal{F}_\tau$.
(ii) We first prove that the stopped process V^τ is progressively measurable. To see this, observe that the map $(s, \omega) \longmapsto X_s^\tau(\omega)$ is the composition of the $\mathcal{B}([0,t]) \otimes \mathcal{F}_t-$measurable maps

$$f : (s, \omega) \longmapsto (\tau(\omega), \omega) \quad \text{and} \quad V : (s, \omega) \longmapsto V_s(\omega),$$

where the measurability of the second map is is exactly the progressive measurability assumption on V, and that of the first one follows from the fact that for all $u \leq t$, $A \in \mathcal{F}_t$:

$$f|_{[0,t]}^{-1}([0,s) \times A) = \{(u,\omega) \in [0,t] \times \Omega : u < s, u < \tau(\omega), \omega \in A\} \in \mathcal{B}[0,t] \times \mathcal{F}_t,$$

as a consequence of $\{\tau(\omega) > u\} \in \mathcal{F}_u \subset \mathcal{F}_t$.

Finally, For every $t \geq 0$ and $B \in \mathcal{B}(E)$, we write $\{X_\tau \in B\} \cap \{\tau \leq t\} = \{X_t^\tau \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t$ by the progressive measurability of the stopped process X^τ. ♢

Proposition 3.9. *Let τ_1 and τ_2 be two $\mathbb{F}-$stopping times. Then the events sets $\{\tau_1 < \tau_2\}$ and $\{\tau_1 = \tau_2\}$ are in $\mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$.*

Proof. (i) We first prove that $\{\tau_1 > \tau_2\} \in \mathcal{F}_{\tau_2}$. For an arbitrary $t \geq 0$, we have $\{\tau_1 \leq \tau_2\} \cap \{\tau_2 \leq t\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\} \cap \{\tau_1 \wedge t \leq \tau_2 \wedge t\}$. Notice that $\{\tau_i \leq t\} \in \mathcal{F}_t$, by the definition of τ_i as stopping times, and $\{\tau_1 \wedge t \leq \tau_2 \wedge t\} \in \mathcal{F}_t$ because both $\tau_1 \wedge t$ and $\tau_2 \wedge t$ are in $\mathcal{F}_t$. Consequently, $\{\tau_1 \leq \tau_2\} \cap \{\tau_2 \leq t\} \in \mathcal{F}_t$, and therefore $\{\tau_1 \leq \tau_2\} \in \mathcal{F}_{\tau_2}$ by the arbitrariness of $t \geq 0$ and the definition of $\mathcal{F}_{\tau_2}$. Since $\mathcal{F}_{\tau_2}$ is a $\sigma-$algebra, this shows that $\{\tau_1 > \tau_2\} \in \mathcal{F}_{\tau_2}$.
(ii) On the other hand, $\{\tau_1 > \tau_2\} = \{\tau_1 \wedge \tau_2 < \tau_1\} \in \mathcal{F}_{\tau_1}$, since $\tau_1 \wedge \tau_2$ and τ_1 are $\mathcal{F}_{\tau_1}-$measurable.
(iii) Finally $\{\tau_1 = \tau_2\} = \{\tau_1 > \tau_2\}^c \cap \{\tau_2 > \tau_1\}^c \in \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$ by the first part of this proof and the fact that $\mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$ is a $\sigma-$algebra. ♢

3.2 Martingales and optional sampling

In this section, we shall consider real-valued adapted stochastic processes V on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. The notion of martingales is defined similarly as in the discrete-time case.

Definition 3.10. *Let V be an $\mathbb{F}-$adapted stochastic process with $\mathbb{E}|V_t| < \infty$ for every $t \in \mathbb{R}_+$.*
(i) *V is a submartingale if $\mathbb{E}[V_t|\mathcal{F}_s] \geq V_s$ for $0 \leq s \leq t$,*
(ii) *V is a supermartingale if $\mathbb{E}[V_t|\mathcal{F}_s] \leq V_s$ for $0 \leq s \leq t$,*
(iii) *V is a martingale if $\mathbb{E}[V_t|\mathcal{F}_s] = V_s$ for $0 \leq s \leq t$.*

The following Doob's optional sampling theorem states that submartingales and supermartingales satisfy the same inequalities when sampled along random times, under convenient conditions.

Theorem 3.11. (Optional sampling) *Let $V = \{V_t, 0 \leq t \leq \infty\}$ be a right-continuous submartingale where the last element $V_\infty := \lim_{t\to\infty} V_t$ exists for almost every $\omega \in \Omega$ (see Remark 3.12). If $\tau_1 \leq \tau_2$ are two stopping times, then*

$$\mathbb{E}\left[V_{\tau_2}|\mathcal{F}_{\tau_1}\right] \;\geq\; V_{\tau_1} \qquad \mathbb{P}-a.s. \tag{3.2}$$

Proof. For stopping times τ_1 and τ_2 taking values in a finite set, the result reduces to the context of discrete-time martingales, and is addressed in Section 3.5.1 below. In order to extend the result to general stopping times, we approximate the stopping times τ_i by the decreasing sequences $(\tau_i^n)_{n\geq 1}$ of (3.1). Then by the discrete-time optional sampling theorem,

$$\mathbb{E}\left[V_{\tau_2^n}|\mathcal{F}_{\tau_1^n}\right] \;\geq\; V_{\tau_1^n} \qquad \mathbb{P}-\text{a.s.}$$

- By definition of the conditional expectation, this means that $\int V_{\tau_2^n}\mathbf{1}_A \leq \int V_{\tau_1^m}\mathbf{1}_A$ for all $A \in \mathcal{F}_{\tau_1^m}$. Since $\tau_1 \leq \tau_1^m$, we have $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_1^m}$, and therefore

$$\mathbb{E}\left[V_{\tau_2^n}\mathbf{1}_A\right] \geq \mathbb{E}\left[V_{\tau_1^m}\mathbf{1}_A\right] \quad \text{for all} \quad A \in \mathcal{F}_{\tau_1}. \tag{3.3}$$

- For all $i = 1, 2$, the sequence $\{V_{\tau_i^n}, n \geq 1\}$ is an $\{\mathcal{F}_{\tau_i^n}, n \geq 1\}-$backward submartingale in the sense that $\mathbb{E}|V_{\tau_i^n}| < \infty$ and $\mathbb{E}\left[V_{\tau_i^n}|\mathcal{F}_{\tau_i^{n+1}}\right] \geq V_{\tau_i^{n+1}}$. Moreover, $\mathbb{E}\left[V_{\tau_i^n}\right] \geq \mathbb{E}[V_0]$. Then it follows from Lemma 3.14 below that it is uniformly integrable. Therefore, taking limits in (3.3) and using the right-continuity of V, we obtain that

$$\mathbb{E}\left[V_{\tau_2}\mathbf{1}_A\right] \geq \mathbb{E}\left[V_{\tau_1}\mathbf{1}_A\right] \quad \text{for all} \quad A \in \mathcal{F}_{\tau_1},$$

completing the proof. $\diamondsuit$

Remark 3.12. The previous optional sampling theorem requires the existence of a *last element* $V_\infty \in \mathbb{L}^1$, as defined in the statement of the theorem, with $\mathbb{E}[V_\infty|\mathcal{F}_t] \geq V_t$. For completeness, we observe that the existence of a last element $V_\infty \in \mathbb{L}^1$ is verified for right-continuous submartingales V with $\sup_{t\geq 0} \mathbb{E}\left[V_t^+\right] < \infty$. This is the so-called submartingale convergence theorem, see e.g. Karatzas and Shreve [30] Theorem 1.3.15. Notice however that this does not guarantee that $\mathbb{E}[V_\infty|\mathcal{F}_t] \geq V_t$.

In the context of these lectures, we shall simply apply the following consequence of the optional sampling theorem.

Exercise 3.13. *For a right-continuous submartingale V and two stopping times $\tau_1 \leq \tau_2$, the optional sampling theorem holds under either of the following conditions:*
(i) *$\tau_2 \leq a$ for some constant $a > 0$,*
(ii) *there exists an integrable r.v. Y such that $V_t \leq \mathbb{E}[Y|\mathcal{F}_t]$ $\mathbb{P}-a.s.$ for every $t \geq 0$.* *Hint:* under this condition, the existence of V_∞ is guaranteed by the submartingale convergence theorem (see Remark 3.12), and the submartingale property at infinity is a consequence of Fatou's lemma.

We conclude this section by proving a uniform integrability result for *backward submartingales* which was used in the above proof of Theorem 3.11.

Lemma 3.14. *Let $\{\mathcal{D}_n, n \geq 1\}$ be a sequence of sub-$\sigma-$algebras of $\mathcal{F}$ with $\mathcal{D}_n \supset \mathcal{D}_{n+1}$ for all $n \geq 1$. Let $\{X_n, n \geq 1\}$ be an integrable stochastic process with*

$$X_n \ \mathcal{D}_n - \text{measurable and } \mathbb{E}[X_n|\mathcal{D}_{n+1}] \geq X_{n+1} \quad \text{for all} \quad n \geq 1. \tag{3.4}$$

Suppose that the sequence $(\mathbb{E}[X_n])_{n\geq 1}$ is bounded from below, then $\{X_n, n \geq 1\}$ is uniformly integrable.

Proof. We organize the proof in three steps.
Step 1: By the Jensen inequality, it follows from (3.4) that $\mathbb{E}[X_n^+|\mathcal{D}_{n+1}] \geq X_{n+1}^+$ for all $n \geq 1$. Then $\mathbb{E}[X_{n+1}^+] \leq \mathbb{E}[\mathbb{E}\{X_n^+|\mathcal{F}_{n+1}\}] = \mathbb{E}[X_n^+]$

$$\mathbb{P}[|X_n| > \lambda] \leq \frac{1}{\lambda}\mathbb{E}|X_n| = \frac{1}{\lambda}\left(2\mathbb{E}[X_n^+] - \mathbb{E}[X_n]\right) \leq \frac{1}{\lambda}\left(2\mathbb{E}[X_1^+] - \mathbb{E}[X_n]\right).$$

Since the sequence $(\mathbb{E}[X_n])_{n\geq 1}$ is bounded from below, this shows that, as $\lambda \to \infty$, $\sup_{n\geq 1} \mathbb{P}[|X_n| > \lambda] \longrightarrow 0$. In other words:

$$\text{for all } \delta > 0, \quad \text{there exists } \lambda^* > 0 \text{ s.t.} \quad \mathbb{P}[|X_n| > \lambda] \leq \delta \text{ for all } n \geq 1. \tag{3.5}$$

Step 2: We finally prove that both $\{X_n^-, n \geq 1\}$ is uniformly integrable. By (3.4) and the Jensen inequality, we directly estimate for $\lambda > 0$ that

$$\mathbb{E}\left[X_n^+ \mathbf{1}_{\{X_n^+ > \lambda\}}\right] \leq \mathbb{E}\left[X_1^+ \mathbf{1}_{\{X_n > \lambda\}}\right].$$

Let $\mathcal{A}^\delta$ be the set of all events $A \in \mathcal{F}$ such that $\mathbb{P}[A^\delta] \le \delta$. Let $\varepsilon > 0$ be given. Since X_1^+ is integrable, there exists $\delta > 0$ such that $\mathbb{E}[X_1^+ \mathbf{1}_A] \le \varepsilon$ for all $A \in \mathcal{A}^\delta$. By (3.5), there exists λ^* such that $\{\{X_n > \lambda\} \in \mathcal{A}^\delta$ for all $n \ge 1$, and therefore

$$\mathbb{E}\left[X_n^+ \mathbf{1}_{\{X_n^+ > \lambda\}}\right] \le \varepsilon \quad \text{for all} \quad \lambda \ge \lambda^* \text{ and } n \ge 1.$$

Step 3: We now prove that both $\{X_n^+, n \ge 1\}$ is uniformly integrable.Similarly, for $\lambda > 0$, it follows from (3.4) that $\mathbb{E}\left[X_n^- \mathbf{1}_{\{X_n^- \ge \lambda\}}\right] \le \mathbb{E}\left[X_k^- \mathbf{1}_{\{X_n^- \ge \lambda\}}\right]$ for all $k \le n$. Then

$$0 \le \mathbb{E}\left[X_n^- \mathbf{1}_{\{X_n^- < \lambda\}}\right] \le u_n - u_m + \mathbb{E}\left[X_n \mathbf{1}_{\{X_n < -\lambda\}}\right], \tag{3.6}$$

where $u_n := \mathbb{E}[X_n]$, $n \ge 1$, is bounded from below and decreasing by (3.4). Then for all $\varepsilon > 0$, there exists $k^* > 0$ such that $|u_n - u_{k^*}| \le \varepsilon$ for all $n \ge k^*$. Arguying as in Step 2, it follows from the integrability of X_{k^*} that

$$\sup_{n \ge k^*} \mathbb{E}\left[|X_{k^*}| \mathbf{1}_{\{X_n < -\lambda\}}\right] \le \varepsilon,$$

which concludes the proof by (3.6). $\diamondsuit$

3.3 Maximal inequalities for submartingales

In this section, we recall the Doob's maximal inequality for discrete-time martingales, and extend it to continuous-time martingales.

Theorem 3.15. *(Doob's maximal inequality)*
(i) *Let $\{M_n, n \in \mathbb{N}\}$ be a nonnegative submartingale, and set $M_n^* := \sup_{k \le n} M_k$. Then for all $n \ge 0$:*

$$c\mathbb{P}\big[M_n^* \ge c\big] \le \mathbb{E}\big[M_n \mathbf{1}_{\{M_n^* \ge c\}}\big] \quad \textit{and} \quad \|M_n^*\|_p \le \frac{p}{p-1}\|M_n\|_p \textit{ for all } p > 1.$$

(ii) *Let $\{M_t, t \ge 0\}$ be a nonnegative continuous submartingale, and set $M_t^* := \sup_{s \in [0,t]} M_s$. Then for all $t \ge 0$:*

$$c\mathbb{P}\big[M_t^* \ge c\big] \le \mathbb{E}\big[M_t \mathbf{1}_{\{M_t^* \ge c\}}\big] \quad \textit{and} \quad \|M_t^*\|_p \le \frac{p}{p-1}\|M_t\|_p \textit{ for all } p > 1.$$

Proof. 1- We first prove that

$$c\mathbb{P}[M_n^* \ge c] \le \mathbb{E}\left[M_n \mathbf{1}_{\{M_n^* \ge c\}}\right] \le \mathbb{E}[M_n] \quad \text{for all} \quad c > 0 \text{ and } n \in \mathbb{N}. \tag{3.7}$$

To see this, observe that

$$\{M_n^* \ge c\} = \bigcup_{k \le n} F_k, \;\; F_0 := \{M_0 \ge c\}, \;\; F_k := \left(\bigcap_{i \le k-1} \{M_i < c\}\right) \cap \{M_k \ge c\}. \tag{3.8}$$

Since $F_k \in \mathcal{F}_k$, $M_k \geq c$ sur F_k, it follows from the submartingale property of M that

$$\mathbb{E}[M_n \mathbf{1}_{F_k}] \geq \mathbb{E}[M_k \mathbf{1}_{F_k}] \geq c\mathbb{P}[F_k], \qquad k \geq 0.$$

Summing over k, this provides

$$\mathbb{E}[M_n \mathbf{1}_{\cup_k F_k}] \geq c \sum_{k \geq 0} \mathbb{P}[F_k] \geq c\mathbb{P}[\cup_k F_k]$$

and (3.7) follows from (3.8).
2- Soit $p > 0$ et $q := p/(p-1)$. It follows from (3.7) that

$$L := \int_0^\infty pc^{p-1}\mathbb{P}[M_n^* \geq c]dc \quad \leq \quad R := \int_0^\infty pc^{p-2}\mathbb{E}\left[M_n \mathbf{1}_{\{M_n^* \geq c\}}\right] dc.$$

since $M \geq 0$, it follows from Fubini's theorem that

$$L = \mathbb{E}\left[\int_0^{M_n^*} pc^{p-1}dc\right] = \mathbb{E}\left[(M_n^*)^p\right]$$

and

$$R = \mathbb{E}\left[\int_0^{M_n^*} pc^{p-2}dc\right] = q\mathbb{E}\left[[M_n(M_n^*)^{p-1}\right] \leq q\|M_n\|_p \|(M_n^*)^{p-1}\|_q,$$

by Hölder inequality. Hence $\|(M_n^*)^p\|_p^p \leq q\|M_n\|_p\|(M_n^*)^{p-1}\|_q$, and the required inequality follows.
3- The extension of the inequality to continuous-time martingales which are pathwise continuous follows from an obvious discretization and the monotone convergence theorem. ♢

3.4 Submartingales with a.s. càd-làg versions

In this section, we show that submartingales have nice trajectorial properties, up to the passage to a convenient version. We first show that almost all sample paths of submartingales have left and right limits at any point.

Lemma 3.16. *Let $X = \{X_t, t \geq 0\}$ be a submartingale. Then,*

$$X_{t+} := \lim_{\mathbb{Q} \ni s \downarrow t} X_s \quad \text{exists for all} \quad t \geq 0, \quad \mathbb{P}-a.s.$$

Moreover, the process $\{X_{t+}, t \geq 0\}$ is a $\{\mathcal{F}_{t+}\}_{t \geq 0}-$submartingale with càd-làg sample paths satisfying $\mathbb{E}[X_{t+}|\mathcal{F}_t] \geq X_t$, $\mathbb{P}-$a.s. for all $t \geq 0$.

A similar statement holds for the left-hand side limits.

Proof. *Step 1* We first prove that,

$$X_{t+} \quad := \quad \lim_{s\searrow t, s\in\mathbb{Q}} X_s \;\; \text{exists for all } t \in [0,T),\ \mathbb{P}-\text{a.s.} \tag{3.9}$$

and we will next prove, in Step 2 below, that the process $\{X_{t+}, t \geq 0\}$ is our required version.

For $\alpha < \beta$ and $T > 0$, we denote by $U_T^X(\alpha,\beta)$ the number of up-crossings of the interval $[\alpha,\beta]$ by the discrete-time process $\{X_t, t \in [0,T]\cap\mathbb{Q}\}$. Then, by a classical result on discrete-time martingales, see the review in Subsection 3.5.2 below, we have $\mathbb{E}\big[U_T^X(\alpha,\beta)\big] < \infty$. Consequently the event set

$$\Omega_T^X(\alpha,\beta) \quad := \quad \big\{\omega\in\Omega : U_T^{X(\omega)}(\alpha,\beta) = \infty\big\}$$

is negligible under $\mathbb{P}$. Since

$$\big\{\omega\in\Omega : \liminf_{s\downarrow t} X_s(\omega) < \limsup_{s\downarrow t} X_s(\omega),\ \text{for some } t\in[0,T]\big\} \subset \cup_{\alpha,\beta\in\mathbb{Q}}\Omega_T^X(\alpha,\beta),$$

and $\mathbb{P}\big[\cup_{\alpha,\beta\in\mathbb{Q}}\,\Omega_T^X(\alpha,\beta)\big] = 0$, we conclude that (3.9) holds true.
Step 2 Clearly, the process $\{X_{t+}, t \geq 0\}$ is càd-làg, and for all sequence of rationals $t_n \downarrow t$, we have $\mathbb{E}[X_{t+}|\mathcal{F}_t] = \mathbb{E}[\lim_{n\to\infty} X_{t_n}|\mathcal{F}_t]$, $\mathbb{P}-$a.s. Since X is a submartingale, it follows from Lemma 3.14, that $(X_{t_n})_n$ is uniformly integrable. Then,

$$\mathbb{E}[X_{t+}|\mathcal{F}_t] = \lim_{n\to\infty}\mathbb{E}[X_{t_n}|\mathcal{F}_t] \geq X_t, \quad \mathbb{P}-\text{a.s.}$$

Similarly, for all sequence of rationals $s_n \downarrow s < t$, we have $\mathbb{E}[X_{t+}|\mathcal{F}_{s_n}] \geq X_{s_n}$, and therefore $\mathbb{E}[X_{t+}|\mathcal{F}_{s+}] \geq X_{s+}$. $\diamondsuit$

The main result of this section provides a characterization of submartingales which have a pathwise càd-làg version. We say that a filtration $\mathbb{F} = \{\mathcal{F}_t, t\geq 0\}$ satisfies the usual properties if it is right continuous, and $\mathcal{F}_0$ contains all null sets of $\mathcal{F}$.

Theorem 3.17. *Let $\mathbb{F}$ be a filtration satisfying the usual conditions, and X an $\mathbb{F}-$submartingale. Then, X has a càd-làg version if and only if the map $t \longmapsto \mathbb{E}[X_t]$ is right-continuous.*

Moreover, such a càd-làg version is a submartingale.

Proof. Since $\mathbb{F}$ satisfies the usual conditions, we see that the càd-làg process $\{X_{t+}, t\geq 0\}$, introduced in Lemma 3.16, is a $\mathbb{F}-$submartingale satisfying $X_{t+} = \mathbb{E}[X_{t+}|\mathcal{F}_t] \geq X_t$, $\mathbb{P}-$a.s.

For any sequence of rationals $t_n \downarrow t$, it follows from the submartingale property of X, together with Lemma 3.14, that $(X_{t_n})_n$ is uniformly integrable. Then, $\mathbb{E}[X_{t+} - X_t] = \lim_{n\to\infty}\mathbb{E}[X_{t_n} - X_t] = 0$. hence, $X_{t+} = X_t$, $\mathbb{P}-$a.s. if and only if $\mathbb{E}[X_{t_n}] \longrightarrow \mathbb{E}[X_t]$ as $n\to\infty$. By arbitrariness of the sequence $t_n \downarrow t$, this provides the required equivalence. $\diamondsuit$

3.5 Appendix: on discrete-time martingales

3.5.1 Doob's optional sampling for discrete martingales

This section reports the proof of the optional sampling theorem for discrete-time martingales which was the starting point for the proof of the continuous-time extension of Theorem 3.11.

Lemma 3.18. *Let $\{X_n, n \geq 0\}$ be a supermartingale (resp. submartingale, martingale) and ν a stopping time on $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$. Then, the stopped process X^ν is a supermartingale (resp. submartingale, martingale).*

Proof. We only prove the result in the martingale case. We first observe that X^ν is $\mathbb{F}-$adapted since, for all $n \geq 0$ and $B \in \mathcal{E}$,

$$(X_n^\nu)^{-1}(B) = \left[\cup_{k \leq n-1}\{\nu = k\} \cap (X_k)^{-1}(B)\right] \cup \left[\{\nu \leq n-1\}^c \cap (X_n)^{-1}(B)\right].$$

For $n \geq 1$, $|X_n^\nu| \leq \sum_{k \leq n} |X_k| \in \mathcal{L}^1(\Omega, \mathcal{F}_n, \mathbb{P})$, and we directly compute that

$$\begin{aligned}
\mathbb{E}[X_n^\nu|\mathcal{F}_{n-1}] &= \mathbb{E}\left[X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + X_n \mathbf{1}_{\{\nu > n-1\}}|\mathcal{F}_{n-1}\right] \\
&= X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + \mathbb{E}\left[X_n \mathbf{1}_{\{\nu > n-1\}}|\mathcal{F}_{n-1}\right].
\end{aligned}$$

Since ν is a stopping time, the event $\{\nu > n-1\} = \{\nu \leq n-1\}^c \in \mathcal{F}_{n-1}$. Since X_n and $X_n \mathbf{1}_{\{\nu > n-1\}}$ are integrable, we deduce that

$$\begin{aligned}
\mathbb{E}[X_n^\nu|\mathcal{F}_{n-1}] &= X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + \mathbf{1}_{\{\nu > n-1\}}\mathbb{E}\left[X_n|\mathcal{F}_{n-1}\right] \\
&\leq X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + \mathbf{1}_{\{\nu > n-1\}} X_{n-1} = X_{n-1}^\nu.
\end{aligned}$$

♢

Theorem 3.19. (Optional sampling, Doob). *Let $\{X_n, n \geq 0\}$ be a martingale (resp. supermartingale) and $\underline{\nu}$, $\bar{\nu}$ two* <u>bounded</u> *stopping times satisfying $\underline{\nu} \leq \bar{\nu}$ a.s. Then :*

$$\mathbb{E}\left[X_{\bar{\nu}}|\mathcal{F}_{\underline{\nu}}\right] = X_{\underline{\nu}} \quad \textit{(resp.} \leq X_{\underline{\nu}}\textit{).}$$

Proof. We only prove the result for the martingale case ; the corresponding result for supermartingales is proved by the same argument. Let $N \in \mathbb{N}$ be a bound on $\overline{\nu}$.

(i) We first show that $E[X_N|\mathcal{F}_\nu] = X_\nu$ for all stopping time ν. For an arbitrary event $A \in \mathcal{F}_\nu$, we have $A \cap \{\nu = n\} \in \mathcal{F}_n$ and therefore:

$$\mathbb{E}\left[(X_N - X_\nu)\mathbf{1}_{A \cap \{\nu = n\}}\right] = \mathbb{E}\left[(X_N - X_n)\mathbf{1}_{A \cap \{\nu = n\}}\right] = 0$$

since X is a martingale. Summing up over n, we get:

$$0 = \sum_{n=0}^{N} \mathbb{E}\left[(X_N - X_\nu)\mathbf{1}_{A \cap \{\nu = n\}}\right] = \mathbb{E}\left[(X_N - X_\nu)\mathbf{1}_A\right].$$

By the arbitrariness of A in $\mathcal{F}_\nu$, this proves that $E[X_N - X_\nu|\mathcal{F}_\nu] = 0$.
(ii) It follows from Lemma 3.18 that the stopped process $X^{\bar\nu}$ is a martingale. Applying the result established in (i), we see that:

$$\mathbb{E}\left[X_{\bar\nu}|\mathcal{F}_{\underline{\nu}}\right] = \mathbb{E}\left[X_T^{\bar\nu}|\mathcal{F}_{\underline{\nu}}\right] = X_{\underline{\nu}}^{\bar\nu} = X_{\underline{\nu}}$$

since $\underline{\nu} \le \bar\nu$. $\diamondsuit$

3.5.2 Upcrossings of discrete-time submartingales

In this subsection, we consider a submartingale $\{X_n, n \ge 0\}$. For all $\alpha < \beta$, we define the sequence of stopping times

$$\tau_0 = 0,\ \theta_{n+1} := \inf\{i \ge \tau_n \ : X_i \le \alpha\}, \quad \text{and} \quad \tau_{n+1} := \inf\{i \ge \theta_{n+1} \ : X_i \ge \beta\}.$$

Then, for all $n \ge 0$, the random variable

$$U_n^{\alpha,\beta} := \max\{j \ : \ \tau_j \le n\} \tag{3.10}$$

represents of crossings of the level β starting below the level α on the time interval $[0, n]$. We call U_n the number of upcrossings of the interval $[\alpha, \beta]$ before time n.

Lemma 3.20. *For a submartingale $\{X_n, n \ge 0\}$, and two scalars $\alpha < \beta$. we have:*

$$\mathbb{E}\left[U_n^{\alpha,\beta}\right] \le \frac{1}{\beta-\alpha}\mathbb{E}\left[(X_n-\alpha)^+\right].$$

Proof. Denote $Y_n := (X_n - a)^+, n \ge 0$. By the Jensen inequality, we immediately verify that $\{Y_n, n \ge 0\}$ inherits the submartingale property of $\{X_n, n \ge 0\}$. Since $\theta_{n+1} > n$, we have

$$\begin{aligned} Y_n &= Y_{n\wedge\theta_1} + \sum_{i=1}^n \left(Y_{n\wedge\tau_i} - Y_{n\wedge\theta_i}\right) + \sum_{i=1}^n \left(Y_{n\wedge\theta_{i+1}} - Y_{n\wedge\tau_i}\right) \\ &\ge \sum_{i=1}^n \left(Y_{n\wedge\tau_i} - Y_{n\wedge\theta_i}\right) + \sum_{i=1}^n \left(Y_{n\wedge\theta_{i+1}} - Y_{n\wedge\tau_i}\right). \end{aligned}$$

By definition of upcrossings, we have $Y_{\theta_i} = 0$, $Y_{\tau_i} \ge b - a$ on $\{\tau_i \le n\}$, and $Y_{n\wedge\tau_i} - Y_{n\wedge\theta_i} \ge 0$, a.s. Then:

$$(\beta-\alpha)\mathbb{E}\left[U_n^{\alpha,\beta}\right] \le \mathbb{E}[Y_n] - \sum_{i=1}^n \mathbb{E}\left[Y_{n\wedge\theta_{i+1}} - Y_{n\wedge\tau_i}\right] \le \mathbb{E}[Y_n],$$

where the second inequality follows from the optional sampling theorem, together with the submartingale property of the process $\{Y_n, n \ge 0\}$. $\diamondsuit$

Chapter 4

The Brownian Motion

The Brownian motion was introduced by the scottish botanist Robert Brown in 1828 to describe the movement of pollen suspended in water. Since then it has been widely used to model various irregular movements in physics, economics, finance and biology. In 1905, Albert Einstein (1879-1955) introduced a model for the trajectory of atoms subject to shocks, and obtained a Gaussian density. Louis Bachelier (1870-1946) was the very first to use the Brownian motion as a model for stock prices in his thesis in 1900, but his work was not recognized until the recent history. It is only sixty years later that Samuelson (1915-2009, Nobel Prize in economics 1970) suggested the Brownian motion as a model for stock prices. The real success of Brownian motion in the financial application was however realized by Fisher Black (1938-1995), Myron Scholes (1941-), and Robert Merton (1944-) who received the Nobel Prize in economics 1997 for their seminal work between 1969 and 1973 founding the modern theory of financial mathematics by introducing the portfolio theory and the no-arbitrage pricing argument.

The first rigorous construction of the Brownian motion was achieved by Norbert Wiener (1894-1964) in 1923, who provided many applications in signal theory and telecommunications. Paul Lévy, (1886-1971, Alumni X1904 from Ecole Polytechnique, and Professor at Ecole Polytechnique from 1920 to 1959) contributed to the mathematical study of the Brownian motion and proved many surprising properties. Kyioshi Itô (1915-2008) developed the stochastic differential calculus. The theory benefitted from the considerable activity on martingales theory, in particular in France around Paul-André Meyer. (1934-2003).

The purpose of this chapter is to introduce the Brownian motion and to derive its main properties.

4.1 Definition of the Brownian motion

Definition 4.1. *Let $W = \{W_t,\ t \in \mathbb{R}_+\}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{F}$ a filtration. W is an $\mathbb{F}-$standard Brownian motion if*
(i) W *is $\mathbb{F}-$adapted.*
(ii) $W_0 = 0$ *and the sample paths $W_.(\omega)$ are continuous for a.e. $\omega \in \Omega$,*
(iii) *independent increments: $W_t - W_s$ is independent of $\mathcal{F}_s$ for all $s \le t$,*
(iv) *the distribution of $W_t - W_s$ is $\mathcal{N}(0, t-s)$ for all $t > s \ge 0$,*

An interesting consequence of (iii) is that:

(iii') the increments $\left(W_{t_i} - W_{t_{i-1}}\right)_{i \le n}$ are independent for all $n \in \mathbb{N}$ and all $0 \le t_0 \le \cdots \le t_n$.

Let us observe that, for any given filtration $\mathbb{F}$, an $\mathbb{F}-$standard Brownian motion is also an $\mathbb{F}^W-$standard Brownian motion, where $\mathbb{F}^W$ is the canonical filtration of W. This justifies the consistency of the above definition with the following one which does not refer to any filtration:

Definition 4.2. *Let $W = \{W_t,\ t \in \mathbb{R}_+\}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. W is a standard Brownian motion if it satisfies conditions* (ii), (iii') *and* (iv) *of Definition 4.1.*

We observe that the pathwise continuity condition in the above Property (ii) can be seen to be redundant. This is a consequence of the Kolmogorov-Čentsov Theorem, that we recall (without proof) for completeness, and which shows that the pathwise continuity follows from Property (iv).

Theorem 4.3. *Let $\{X_t, t \in [0,T]\}$ be a process satisfying*

$$\mathbb{E}|X_t - X_s|^r \le C|t-s|^{1+\gamma r}, \quad 0 \le s, t \le T, \quad \textit{for some} \ \ r, \gamma, C \ge 0.$$

Then there exists a modification $\{\tilde{X}_t, t \in [0,T]\}$ of X ($\mathbb{P}[X_t = \tilde{X}_t] = 1$ for all $t \in [0,T]$) which is a.s. $\alpha-$Hölder continuous for every $\alpha \in (0, \gamma)$.

Exercise 4.4. *Use Theorem 4.3 to prove that, $\mathbb{P}-$a.s., the Brownian motion is $(\frac{1}{2} - \varepsilon)-$Hölder continuous for any $\varepsilon > 0$ (Theorem 4.22 below shows that this result does not hold for $\varepsilon = 0$.*

We conclude this section by extending the definition of the Brownian motion to the vector case.

Definition 4.5. *Let $W = \{W_t,\ t \in \mathbb{R}_+\}$ be an $\mathbb{R}^n-$valued stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{F}$ a filtration. W is an $\mathbb{F}-$standard Brownian motion if the components $W^i, i = 1, \ldots, n$, are independent $\mathbb{F}-$standard Brownian motions (i.e.* (i)-(ii)-(iv) *of Definition 4.1 hold), and W satisfies the following Property* (iii)$_n$*:*

Figure 4.1: Approximation of a sample path of a Brownian motion

(iii)$_n$ *the distribution of* $W_t - W_s$ *is* $\mathcal{N}(0, (t-s)\mathrm{I}_n)$ *for all* $t > s \geq 0$, *where* I_n *is the identity matrix of* $\mathbb{R}^n$.

4.2 The Brownian motion as a limit of a random walk

Before discussing the properties of the Brownian motion, let us comment on its existence as a continuous-time limit of a random walk. Given a family $\{Y_i,\ i = 1, \ldots, n\}$ of n independent random variables defined by the distribution

$$\mathbb{P}[Y_i = 1] \;=\; 1 - \mathbb{P}[Y_i = -1] \quad = \quad \frac{1}{2}\,, \tag{4.1}$$

we define the symmetric random walk

$$M_0 = 0 \text{ and } M_k \;=\; \sum_{j=1}^{k} Y_j \quad \text{for} \quad k = 0, \ldots, n\,.$$

A continuous-time process can be obtained from the sequence $\{M_k,\ k = 0, \ldots, n\}$ by linear interpolation:

$$M_t \;:=\; M_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\, Y_{\lfloor t \rfloor + 1} \quad \text{for} \quad t \geq 0\,,$$

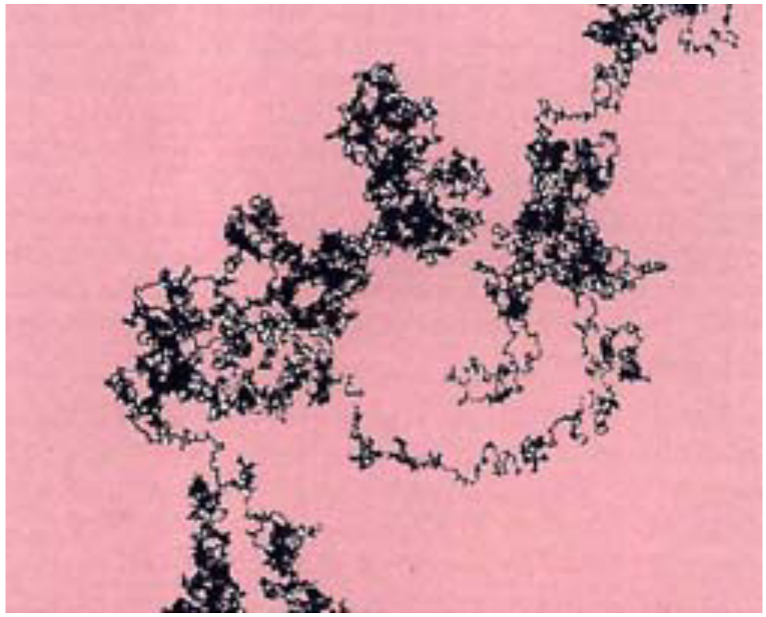

Figure 4.2: A sample path of the two-dimensional Brownian motion

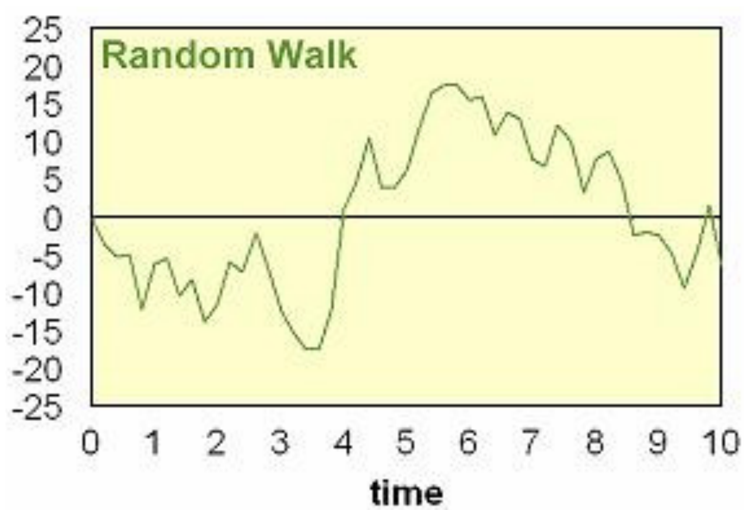

Figure 4.3: Sample path of a random walk

where $\lfloor t \rfloor$ denotes the largest integer less than or equal to t. Figure 4.3 shows a typical sample path of the process M.

We next define a stochastic process W^n from the previous process by speeding up time and conveniently scaling:

$$W^n_t \;:=\; \frac{1}{\sqrt{n}} M_{nt}\,, \quad t \geq 0\,.$$

In the above definition, the normalization by $\sqrt{n}$ is suggested by the Central Limit Theorem. We next set

$$t_k \;:=\; \frac{k}{n} \quad \text{for} \quad k \in \mathbb{N}$$

and we list some obvious properties of the process W^n:

• for $0 \leq i \leq j \leq k \leq \ell \leq n$, the increments $W^n_{t_\ell} - W^n_{t_k}$ and $W^n_{t_j} - W^n_{t_i}$ are independent,

• for $0 \leq i \leq k$, the two first moments of the increment $W^n_{t_k} - W^n_{t_i}$ are given by

$$\mathbb{E}[W^n_{t_k} - W^n_{t_i}] \;=\; 0 \quad \text{and} \quad \mathbb{V}ar[W^n_{t_k} - W^n_{t_i}] \;=\; t_k - t_i\,,$$

which shows in particular that the normalization by $n^{-1/2}$ in the definition of W^n prevents the variance of the increments from blowing up,

• with $\mathcal{F}^n_t := \sigma\left(Y_j,\; j \leq \lfloor nt \rfloor\right)$, $t \geq 0$, the sequence $\left\{W^n_{t_k},\; k \in \mathbb{N}\right\}$ is a discrete $\left\{\mathcal{F}^n_{t_k},\; k \in \mathbb{N}\right\}$ −martingale:

$$\mathbb{E}\left[W^n_{t_k} | \mathcal{F}^n_{t_i}\right] \;=\; W^n_{t_i} \quad \text{for} \quad 0 \leq i \leq k\,.$$

Hence, except for the Gaussian feature of the increments, the discrete-time process $\left\{W^n_{t_k},\; k \in \mathbb{N}\right\}$ is approximately a Brownian motion. One could even obtain Gaussian increments with the required mean and variance by replacing the distribution (4.1) by a convenient normal distribution. However, since our objective is to imitate the Brownian motion in the asymptotics $n \to \infty$, the Gaussian distribution of the increments is expected to hold in the limit by a central limit type of argument.

Figure 4.4 represents a typical sample path of the process W^n. Another interesting property of the rescaled random walk, which will be inherited by the Brownian motion, is the following quadratic variation result:

$$[W^n, W^n]_{t_k} \;:=\; \sum_{j=1}^{k} \left(W^n_{t_j} - W^n_{t_{j-1}}\right)^2 \;=\; t_k \quad \text{for} \quad k \in \mathbb{N}\,.$$

A possible proof of the existence of the Brownian motion consists in proving the convergence in distribution of the sequence W^n toward a Brownian motion, i.e. a process with the properties listed in Definition 4.2. This is the so-called Donsker's invariance principle. The interested reader may consult a rigorous treatment of this limiting argument in Karatzas and Shreve [30] Theorem 2.4.20.

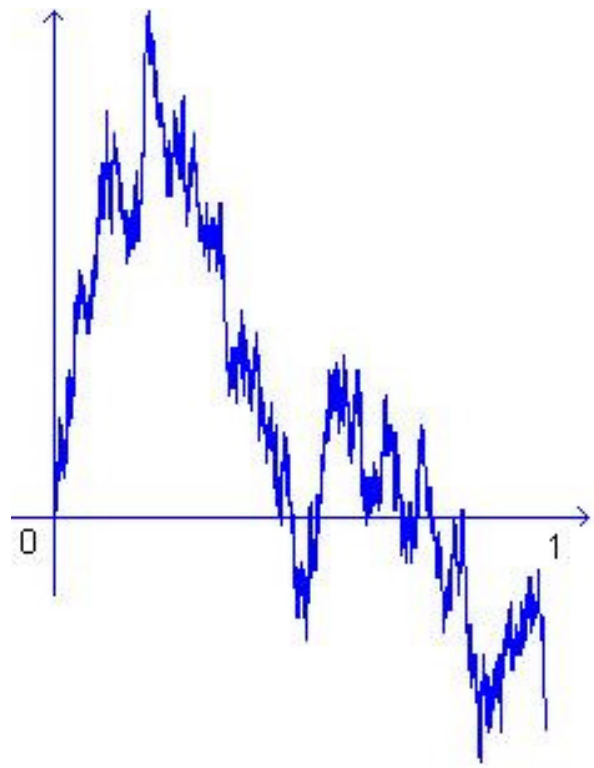

Figure 4.4: The rescaled random walk

4.3 Distribution of the Brownian motion

Let W be a standard real Brownian motion. In this section, we list some properties of W which are directly implied by its distribution.

• *The Brownian motion is a martingale:*

$$\mathbb{E}[W_t|\mathcal{F}_s] = W_s \quad \text{for} \quad 0 \le s \le t\,,$$

where $\mathbb{F}$ is any filtration containing the canonical filtration $\mathbb{F}^W$ of the Brownian motion. From the Jensen inequality, it follows that the squared Brownian motion W^2 is a submartingale:

$$\mathbb{E}\left[W_t^2|\mathcal{F}_s\right] \ge W_s^2\,, \quad \text{for} \quad 0 \le s < t\,.$$

The precise departure from a martingale can be explicitly calculated

$$\mathbb{E}\left[W_t^2|\mathcal{F}_s\right] = W_s^2 + (t-s)\,, \quad \text{for} \quad 0 \le s < t\,,$$

which means that the process $\left\{W_t^2 - t,\ t \geq 0\right\}$ is a martingale. This is an example of the very general Doob-Meyer decomposition of submartingales which extends to the continuous-time setting under some regularity conditions, see Chapter 12.

• *The Brownian motion is a Markov process*, i.e.

$$\mathbb{E}\left[\phi\left(W_s, s \geq t\right)|\mathcal{F}_t\right] \quad = \quad \mathbb{E}\left[\phi\left(W_s, s \geq t\right)|W_t\right]$$

for every $t \geq 0$ and every bounded continuous function $\phi : C^0(\mathbb{R}_+) \longrightarrow \mathbb{R}$, where $C^0(\mathbb{R}_+)$ is the set of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$. This follows immediately from the fact that $W_s - W_t$ is independent of $\mathcal{F}_t$ for every $s \geq t$. We shall see later in Corollary 4.12 that the Markov property holds in a stronger sense by replacing the deterministic time t by an arbitrary stopping time τ.

• *The Brownian motion is a centered Gaussian process:* A process $\{X_t, t \geq 0\}$ is gaussian if $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian vector for all $n \geq 1$ and $0 \leq t_1 < \ldots < t_n$.

By definition of the Brownian motion, the random vector $(W_{t_1}, \ldots, W_{t_n})$ is Gaussian for every $0 \leq t_1 < \ldots < t_n$. Centered Gaussian processes can be characterized in terms of their covariance function, see Exercise 4.7 below. A direct calculation provides the covariance function of the Brownian motion

$$\mathbb{C}ov\left(W_t, W_s\right) \quad = \quad \mathbb{E}\left[W_t W_s\right] \ = \ t \wedge s \ = \ \min\{t, s\}$$

The Kolmogorov theorem provides an alternative construction of the Brownian motion as a centered Gaussian process with the above covariance function, we will not elaborate more on this and we send the interested reader to Karatzas and Shreve [30], Section 2.2.

We conclude this section by the following property which is very useful for the purpose of simulating the Brownian motion.

Exercise 4.6. For $0 \leq t_1 < \hat{t} < t_2$, show that the conditional distribution of $W_{\hat{t}}$ given $(W_{t_1}, W_{t_2}) = (x_1, x_2)$ is Gaussian, and provided its mean and variance in closed form. ♢

Exercise 4.7. *1. Prove that a Gaussian process X is characterized by the mean and covariances functions:*

$$m(t) := \mathbb{E}[X_t] \quad \textit{and} \quad c(s,t) := \mathbb{C}ov[X_s, X_t], \quad s, t \geq 0.$$

2. Let $(X_t)_{t \in \mathbb{R}^+}$ be a Gaussian process with continuous sample paths, a.s. and $X_0 = 0$. Prove that X is a Brownian motion if and only if the corresponding mean and covariances functions are given by $m(t) = 0$, and $c(s,t) = s \wedge t$, $s, t \geq 0$.

Solution of Exercise 4.7 1. For $Y := (X_{t_1}, \ldots, X_{t_n})$, we have $\mathbb{E}[Y] = \big(m(t_1), \ldots, m(t_n)\big)^{\mathrm{T}}$ and $\mathbb{V}ar[Y] = \big(c(t_i, t_j)\big)_{1\le i\le j\le n}$. This proves that the functions m and c characterize completely any Gaussian process.
2. First, the mean and the covariances functions of the Brownian motion are indeed given by $m(t) = 0$, and $c(s,t) = s\wedge t$, $s,t \ge 0$. Now let X be a Gaussian process with a.s. continuous sample paths, $X_0 = 0$, and mean and covariances functions $m(t) = 0$, and $c(s,t) = s \wedge t$, $s,t \ge 0$. We need to prove that X has independent increments and that $X_{t+h} - X_t$ is $\mathcal{N}(0,h)$ for all $t, h \ge 0$:

– For $t_1 \le \cdots \le t_n$, we have $\mathbb{C}ov\big[X_{t_{i+2}} - X_{t_{i+1}}, X_{t_i} - X_{t_{i-1}}\big] = c(t_{i+2}, t_i) - c(t_{i+2}, t_{i-1}) - c(t_{i+1}, t_i) + c(t_{i+1}, t_{i-1}) = 0$. As the vector $(X_{t_1}, \ldots, X_{t_n})$ is Gaussian, it follows that the vector $(X_{t_{i+2}} - X_{t_{i+1}}, X_{t_i} - X_{t_{i-1}})_i$ is also Gaussian, so that zero covariances is equivalent to independence.

– As the random vector (X_t, X_{t+h}) is Gaussian, the difference $X_{t+h} - X_t$ is Gaussian, and we directly compute that $\mathbb{E}\big[X_{t+h} - X_t\big] = m(t+h) - m(t) = 0$, and $\mathbb{V}ar\big[X_{t+h} - X_t\big] = c(t+h, t+h) + c(t,t) - 2c(t, t+h) = (t+h) + t - 2t = h$.

• *Distribution*: By definition of the Brownian motion, for $0 \le t < T$, the conditional distribution of the random variable W_T given $W_t = x$ is a $\mathcal{N}(x, T-t)$:

$$p(t,x,T,y)dy \;:=\; \mathbb{P}\left[W_T \in [y, y+dy] | W_t = x\right] \;=\; \frac{1}{\sqrt{2\pi(T-t)}}\, e^{-\frac{(y-x)^2}{2(T-t)}}\, dy$$

An important observation is that this density function satisfies the heat equation for every fixed (t,x):

$$\frac{\partial p}{\partial T} \;=\; \frac{1}{2}\,\frac{\partial^2 p}{\partial y^2}\,,$$

as it can be checked by direct calculation. One can also fix (T,y) and express the heat equation in terms of the variables (t,x):

$$\frac{\partial p}{\partial t} + \frac{1}{2}\,\frac{\partial^2 p}{\partial x^2} \;=\; 0\,. \tag{4.2}$$

We next consider a function g with polynomial growth, say, and we define the conditional expectation:

$$V(t,x) \;=\; \mathbb{E}\left[g\left(W_T\right)|W_t = x\right] \;=\; \int g(y)p(t,x,T,y)dy\,. \tag{4.3}$$

Remark 4.8. Since p is C^∞, it follows from the dominated convergence theorem that V is also C^∞.

By direct differentiation inside the integral sign, it follows that the function V is a solution of

$$\frac{\partial V}{\partial t} + \frac{1}{2}\,\frac{\partial^2 V}{\partial x^2} = 0 \quad \text{and} \quad V(T,.) = g\,. \tag{4.4}$$

We shall see later in Section 8.5 that the function V defined in (4.3) is the unique solution of the above linear partial differential equation in the class of polynomially growing functions.

4.4 Scaling, symmetry, and time reversal

The following easy properties follow from the properties of the centered Gaussian distribution.

Proposition 4.9. *Let W be a standard Brownian motion, $t_0 > 0$, and $c > 0$. Then, so are the processes*
- $\{-W_t,\ t \geq 0\}$ *(symmetry),*
- $\{c^{-1/2} W_{ct},\ t \geq 0\}$ *(scaling),*
- $\{W_{t_0+t} - W_{t_0},\ t \geq 0\}$ *(time translation),*
- $\{W_{T-t} - W_T,\ 0 \leq t \leq T\}$ *(time reversal).*

Proof. Properties (ii), (iii') and (iv) of Definition 4.2 are immediately checked. ♢

Remark 4.10. For a Brownian motion W in $\mathbb{R}^n$, the symmetry property of the Brownian motion extends as follows: *for any $(n \times n)$ matrix A, with $AA^{\mathrm{T}} = I_n$, the process $\{AW_t,\ t \geq 0\}$ is a Brownian motion.*

Another invariance property for the Brownian motion will be obtained by time inversion in subsection 4.6 below. Indeed, the process B defined by $B_0 := 0$ and $B_t := tW_{1/t}$, $t > 0$, obviously satisfies properties (iii') and (iv); property (ii) will be obtained as a consequence of the law of large numbers.

We next investigate whether the translation property of the Brownian motion can be extended to the case where the deterministic time t_0 is replaced by some random time. The following result states that this is indeed the case when the random time is a stopping time.

Proposition 4.11. *Let W be a Brownian motion, and consider some finite stopping time τ. Then, the process B defined by*

$$B_t := W_{t+\tau} - W_\tau\,, \qquad t \geq 0\,,$$

is a Brownian motion independent of $\mathcal{F}_\tau$.

Proof. Clearly $B_0 = 0$ and B has a.s. continuous sample paths. In the rest of this proof, we show that, for $0 \leq t_1 < t_2 < t_3 < t_4$, $s > 0$, and bounded continuous functions ϕ, ψ and f:

$$\begin{aligned} & \mathbb{E}\left[\phi\left(B_{t_4} - B_{t_3}\right)\psi\left(B_{t_2} - B_{t_1}\right) f\left(W_s\right) \mathbf{1}_{s \leq \tau}\right] \\ = \ & \mathbb{E}\left[\phi\left(W_{t_4} - W_{t_3}\right)\right] \mathbb{E}\left[\psi\left(W_{t_2} - W_{t_1}\right)\right] \mathbb{E}\left[f\left(\underline{W}_u\right) \mathbf{1}_{u<\tau}\right], \end{aligned} \tag{4.5}$$

where we denote $\underline{W}_u = (W_r, r \leq u)$. The extension to arbitrary increments $t_1 \leq \cdots \leq t_n$ is immediate, so that this would imply that B has independent increments with the required Gaussian distribution.

Observe that we may restrict our attention to the case where τ has a finite support $\{s_1, \ldots, s_n\}$. Indeed, given that (4.5) holds for such stopping times, one may approximate any stopping time τ by a sequence of bounded stopping times $(\tau^N := \tau \wedge N)_{N \geq 1}$, and then approximate each τ^N by the decreasing sequence of stopping times $\tau^{N,n} := \left(\lfloor n\tau^N \rfloor + 1\right)/n$ of (3.1), apply (4.5) for each $n \geq 1$, and pass to the limit by the dominated convergence theorem thus proving that (4.5) holds for τ.

For a stopping time τ with finite support $\{s_1, \ldots, s_n\}$, we have:

$$
\begin{aligned}
& \mathbb{E}\left[\phi\left(B_{t_4}-B_{t_3}\right)\psi\left(B_{t_2}-B_{t_1}\right) f\left(W_s\right)\mathbf{1}_{\{s\leq\tau\}}\right] \\
= & \sum_{i=1}^{n} \mathbb{E}\left[\phi\left(B_{t_4}-B_{t_3}\right)\psi\left(B_{t_2}-B_{t_1}\right) f\left(\underline{W}_u\right)\mathbf{1}_{\{u\leq\tau\}}\mathbf{1}_{\{\tau=s_i\}}\right] \\
= & \sum_{i=1}^{n} \mathbb{E}\left[\phi\left(W_{t_4^i}-W_{t_3^i}\right)\psi\left(W_{t_2^i}-W_{t_1^i}\right) f\left(\underline{W}_u\right)\mathbf{1}_{\{\tau=s_i\geq u\}}\right]
\end{aligned}
$$

where we denoted $t_k^i := s_i + t_k$ for $i = 1, \ldots, n$ and $k = 1, \ldots, 4$. We next condition upon $\mathcal{F}_{s_i}$ for each term inside the sum, and recall that $\mathbf{1}_{\{\tau=s_i\}}$ is $\mathcal{F}_{s_i}-$measurable as τ is a stopping time. This provides

$$
\begin{aligned}
& \mathbb{E}\left[\phi\left(B_{t_4}-B_{t_3}\right)\psi\left(B_{t_2}-B_{t_1}\right) f\left(\underline{W}_u\right)\mathbf{1}_{\{u\leq\tau\}}\right] \\
= & \sum_{i=1}^{n} \mathbb{E}\left\{\mathbb{E}\left[\phi\left(W_{t_4^i}-W_{t_3^i}\right)\psi\left(W_{t_2^i}-W_{t_1^i}\right)\Big|\mathcal{F}_{s_i}\right] f\left(\underline{W}_u\right)\mathbf{1}_{\{\tau=s_i\geq u\}}\right\} \\
= & \sum_{i=1}^{n} \mathbb{E}\left\{\mathbb{E}\left[\phi\left(W_{t_4}-W_{t_3}\right)\right]\mathbb{E}\left[\psi\left(W_{t_2}-W_{t_1}\right)\right] f\left(\underline{W}_u\right)\mathbf{1}_{\{\tau=s_i\geq u\}}\right\}
\end{aligned}
$$

where the last equality follows from the independence of the increments of the Brownian motion and the symmetry of the Gaussian distribution. Hence

$$
\begin{aligned}
& \mathbb{E}\left[\phi\left(B_{t_4}-B_{t_3}\right)\psi\left(B_{t_2}-B_{t_1}\right) f\left(\underline{W}_u\right)\mathbf{1}_{\{u\leq\tau\}}\right] \\
= & \mathbb{E}\left[\phi\left(W_{t_4}-W_{t_3}\right)\right]\mathbb{E}\left[\psi\left(W_{t_2}-W_{t_1}\right)\right]\sum_{i=1}^{n}\mathbb{E}\left[f\left(\underline{W}_u\right)\mathbf{1}_{\{\tau=s_i\geq u\}}\right]
\end{aligned}
$$

which is exactly (4.5). ♢

An immediate consequence of Proposition 4.11 is the strong Markov property of the Brownian motion.

Corollary 4.12. *The Brownian motion satisfies the strong Markov property:*

$$\mathbb{E}\left[\phi\left(W_{s+\tau}, s\geq 0\right)|\mathcal{F}_\tau\right] = \mathbb{E}\left[\phi\left(W_s, s\geq\tau\right)|W_\tau\right]$$

for every stopping time τ, and every bounded function $\phi : C^0(\mathbb{R}_+) \longrightarrow \mathbb{R}$.

Proof. Since $B_s := W_{s+\tau} - W_\tau$ is independent of $\mathbb{F}_\tau$ for every $s \geq 0$, we have

$$
\begin{aligned}
\mathbb{E}\left[\phi\left(W_s, s\geq\tau\right)|\mathcal{F}_\tau\right] &= \mathbb{E}\left[\phi\left(B_s+W_\tau, s\geq\tau\right)|\mathcal{F}_\tau\right] \\
&= \mathbb{E}\left[\phi\left(B_s+W_\tau, s\geq\tau\right)|W_\tau\right].
\end{aligned}
$$

♢

We next use the symmetry property of Proposition 4.11 in order to provide explicitly the joint distribution of the Brownian motion W and the corresponding running maximum process:

$$W_t^* := \sup_{0\le s\le t} W_s, \qquad t \ge 0.$$

The key-idea for this result is to make use of the Brownian motion started at the first hitting time of some level y:

$$T_y \quad := \quad \inf\{t > 0 \; : \; W_t > y\} \; .$$

Observe that

$$\{W_t^* \ge y\} \quad = \quad \{T_y \le t\},$$

which implies in particular a connection between the distributions of the running maximum W^* and the first hitting time T_y.

Proposition 4.13. *Let W be a Brownian motion and W^* the corresponding running maximum process. Then, for $t > 0$, the random variables W_t^* and $|W_t|$ have the same distribution, i.e.*

$$\mathbb{P}[W_t^* \ge y] \quad = \quad \mathbb{P}[|W_t| \ge y] \; .$$

Furthermore, the joint distribution of the Brownian motion and the corresponding running maximum is characterized by

$$\mathbb{P}[W_t \le x \, , \, W_t^* \ge y] \; = \; \mathbb{P}[W_t \ge 2y - x] \quad \text{for} \quad y > 0 \text{ and } x \le y \, .$$

Proof. From Exercise 3.3 and Proposition 4.11, the first hitting time T_y of the level y is a stopping time, and the process

$$B_t := \left(W_{t+T_y} - W_{T_y}\right) \, , \qquad t \ge 0 \, ,$$

is a Brownian motion independent of $\mathcal{F}_{T_y}$. Since B_t and $-B_t$ have the same distribution and $W_{T_y} = y$, we compute that

$$\begin{aligned}
\mathbb{P}[W_t \le x \, , \, W_t^* \ge y] &= \mathbb{P}\left[y + B_{t-T_y} \le x \, , \, T_y \le t\right] \\
&= \mathbb{E}\Big[\mathbf{1}_{\{T_y \le t\}} \mathbb{P}\{B_{t-T_y} \le x - y | \mathcal{F}_{T_y}\}\Big] \\
&= \mathbb{E}\Big[\mathbf{1}_{\{T_y \le t\}} \mathbb{P}\{-B_{t-T_y} \le x - y | \mathcal{F}_{T_y}\} \\
&= \mathbb{P}[W_t \ge 2y - x \, , \, W_t^* \ge y] \; = \; \mathbb{P}[W_t \ge 2y - x] \; ,
\end{aligned}$$

where the last equality follows from the fact that $\{W_t \ge 2y - x\} \subset \{W_t^* \ge y\}$ as $y \ge x$. As for the marginal distribution of the running maximum, we decompose:

$$\begin{aligned}
\mathbb{P}[W_t^* \ge y] &= \mathbb{P}[W_t < y \, , \, W_t^* \ge y] + \mathbb{P}[W_t \ge y \, , \, W_t^* \ge y] \\
&= \mathbb{P}[W_t < y \, , \, W_t^* \ge y] + \mathbb{P}[W_t \ge y] \\
&= 2\mathbb{P}[W_t \ge y] \; = \; \mathbb{P}[|W_t| \ge y]
\end{aligned}$$

where the two last equalities follow from the first part of this proof together with the symmetry of the Gaussian distribution. ♢

Exercise 4.14. For a Brownian motion W and $t > 0$, show that

$$\mathbb{P}\left[W_t^* \geq y | W_t = x\right] \;=\; e^{\frac{-2}{t} y(y-x)} \quad \text{for} \quad y \geq x^+ .$$

4.5 Brownian filtration and the Zero-One law

Because the Brownian motion has a.s. continuous sample paths, the corresponding canonical filtration $\mathbb{F}^W := \{\mathcal{F}_t^W, t \geq 0\}$ is left-continuous, i.e. $\cup_{s<t}\mathcal{F}_s^W = \mathcal{F}_t^W$. However, $\mathbb{F}^W$ is not right-continuous. To see this, observe that the event set $\{W$ has a local maximum at $t\}$ is in $\mathcal{F}_{t+}^W := \cap_{s>t}\mathcal{F}_s^W$, but is not in $\mathcal{F}_t^W$.

This difficulty can be overcome by slightly enlarging the canonical filtration by the collection of zero-measure sets:

$$\overline{\mathcal{F}}_t^W \;:=\; \mathcal{F}_t^W \vee \mathcal{N}(\mathcal{F}) := \sigma\big(\mathcal{F}_t \cup \mathcal{N}(\mathcal{F})\big), \quad t \geq 0,$$

where

$$\mathcal{N}(\mathcal{F}) \;:=\; \big\{A \in \Omega : \text{there exists } \; \tilde{A} \in \mathcal{F} \text{ s.t. } A \subset \tilde{A} \text{ and } \mathbb{P}[\tilde{A}] = 0\big\}.$$

The resulting filtration $\overline{\mathbb{F}}^W := \{\overline{\mathcal{F}}_t^W, t \geq 0\}$ is called the *augmented canonical filtration* which will now be shown to be continuous.

We first start by the Blumenthal Zero-One Law.

Theorem 4.15. *For any $A \in \mathcal{F}_{0+}^W$, we have $\mathbb{P}[A] \in \{0,1\}$.*

Proof. Since the increments of the Brownian motion are independent, it follows that $\mathcal{F}_{0+}^W \subset \mathcal{F}_\varepsilon$ is independent of $\mathcal{G}_\varepsilon := \sigma\big(W_s - W_\varepsilon, \varepsilon \leq s \leq 1\big)$, for all $\varepsilon > 0$. Then, for $A \in \mathcal{F}_{0+}^W$, we have $\mathbb{P}[A|\mathcal{G}_\varepsilon] = \mathbb{P}[A]$, a.s.

On the other hand, since $W_\varepsilon \longrightarrow W_0$, $\mathbb{P}-$a.s. we see that, for all $t > 0$, $W_t = \lim_{\varepsilon\to 0}(W_t - W_\varepsilon)$, a.s. so that W_t is measurable with respect to the $\sigma-$algebra $\mathcal{G} := (\vee_n \mathcal{G}_{1/n}) \vee \mathcal{N}(\mathcal{F})$. Then $\mathcal{F}_{0+}^W \subset \mathcal{G}$, and therefore by the monotonicity of $\mathcal{G}_\varepsilon$:

$$\mathbf{1}_A = \mathbb{P}[A|\mathcal{G}] = \lim_{\varepsilon\to 0}\mathbb{P}[A|\mathcal{G}_\varepsilon] = \mathbb{P}[A].$$

♢

Theorem 4.16. *Let W be a Brownian motion. Then the augmented filtration $\overline{\mathbb{F}}^W$ is continuous and W is an $\overline{\mathbb{F}}^W-$Brownian motion.*

Proof. The left-continuity of $\overline{\mathbb{F}}^W$ is a direct consequence of the path continuity of W. The inclusion $\overline{\mathcal{F}}_0^W \subset \overline{\mathcal{F}}_{0+}^W$ is trivial and, by Theorem 4.15, we have $\mathcal{F}_{0+}^W \subset \sigma(\mathcal{N}(\mathcal{F})) \subset \overline{\mathcal{F}}_0^W$. Similarly, by the independent increments property, $\mathcal{F}_{t+}^W = \overline{\mathcal{F}}_t^W$ for all $t \geq 0$. Finally, W is an $\overline{\mathbb{F}}^W-$Brownian motion as it satisfies all the required properties of Definition 4.1. ♢

In the rest of these notes, we will always work with the augmented filtration, and we still denote it as $\mathbb{F}^W := \overline{\mathbb{F}}^W$.

4.6 Small/large time behavior of the Brownian sample paths

The discrete-time approximation of the Brownian motion suggests that $\frac{W_t}{t}$ tends to zero at least along natural numbers, by the law of large numbers. With a little effort, we obtain the following strong law of large numbers for the Brownian motion.

Theorem 4.17. *For a Brownian motion W, we have*

$$\frac{W_t}{t} \longrightarrow 0 \quad \mathbb{P}-a.s. \;\; as \quad t\to\infty\,.$$

Proof. We first decompose

$$\frac{W_t}{t} \;=\; \frac{W_t - W_{\lfloor t\rfloor}}{t} + \frac{\lfloor t\rfloor}{t}\frac{W_{\lfloor t\rfloor}}{\lfloor t\rfloor}$$

By the law of large numbers, we have

$$\frac{W_{\lfloor t\rfloor}}{\lfloor t\rfloor} \;=\; \frac{1}{\lfloor t\rfloor}\sum_{i=1}^{\lfloor t\rfloor}(W_i - W_{i-1}) \quad\longrightarrow\quad 0 \;\; \mathbb{P}-\text{a.s.}$$

We next estimate that

$$\frac{\left|W_t - W_{\lfloor t\rfloor}\right|}{t} \;\le\; \frac{\lfloor t\rfloor}{t}\frac{\left|\Delta_{\lfloor t\rfloor}\right|}{\lfloor t\rfloor}, \quad \text{where} \quad \Delta_n := \sup_{n-1<t\le n}(W_t - W_{n-1})\;,\; n\ge 1\,.$$

Clearly, $\{\Delta_n, n\ge 1\}$ is a sequence of independent identically distributed random variables. The distribution of Δ_n is explicitly given by Proposition 4.13. In particular, by a direct application of the Chebychev inequality, it is easily seen that $\sum_{n\ge 1}\mathbb{P}[\Delta_n \ge n\varepsilon] = \sum_{n\ge 1}\mathbb{P}[\Delta_1 \ge n\varepsilon] < \infty$. By the Borel Cantelli Theorem, this implies that $\Delta_n/n \longrightarrow 0$ $\mathbb{P}-$a.s.

$$\frac{W_t - W_{\lfloor t\rfloor}}{\lfloor t\rfloor} \longrightarrow 0 \quad \mathbb{P}-\text{a.s. as} \quad t\to\infty\,,$$

and the required result follows from the fact that $t/\lfloor t\rfloor \longrightarrow 1$ as $t\to\infty$. ♢

As an immediate consequence of the law of large numbers for the Brownian motion, we obtain the invariance property of the Brownian motion by time inversion:

Proposition 4.18. *Let W be a standard Brownian motion. Then the process*

$$B_0 \;=\; 0 \quad and \quad B_t \;:=\; tW_{\frac{1}{t}} \;\, for\; t>0$$

is a Brownian motion.

Proof. This result relies on the characterization of the Brownian motion as a centered gaussian process with appropriate covariance function $c(s,t) = s \wedge t$ and a.s. continuous sample paths, see Exercice 4.7. Notice that the a.s. continuity of the sample paths at the origin is equivalent to the law of large numbers stated in Proposition 4.17. $\diamondsuit$

The following result shows the path irregularity of the Brownian motion.

Proposition 4.19. *Let W be a Brownian motion in $\mathbb{R}$. Then, $\mathbb{P}-$a.s. W changes sign infinitely many times in any time interval $[0,t]$, $t > 0$.*

Proof. Observe that the random times

$$\tau^+ := \inf\{t > 0 : \ W_t > 0\} \quad \text{and} \quad \tau^- := \inf\{t > 0 : \ W_t < 0\}$$

are stopping times with respect to the augmented filtration $\mathbb{F}^W$. Since this filtration is continuous, it follows that the event sets $\{\tau^+ = 0\}$ and $\{\tau^- = 0\}$ are in $\mathcal{F}_0^W$. By the symmetry of the Brownian motion, its non degeneracy on any interval $[0,t]$, $t > 0$, and the fact that $\mathcal{F}_0^W$ is trivial, it follows that $\mathbb{P}[\tau^+ = 0] = \mathbb{P}[\tau^- = 0] = 1$. Hence for a.e. $\omega \in \Omega$, there are sequences of random times $\tau_n^+ \searrow 0$ and $\tau_n^- \nearrow 0$ with $W_{\tau_n^+} > 0$ and $W_{\tau_n^-} < 0$ for $n \geq 1$. $\diamondsuit$

We next state that the sample path of the Brownian motion is not bounded and oscillating at infinity, $\mathbb{P}-$a.s.

Proposition 4.20. *For a standard Brownian motion W, we have*

$$\limsup_{t\to\infty} W_t \ = \ \infty \quad \textit{and} \quad \liminf_{t\to\infty} W_t \ = \ -\infty, \quad \mathbb{P} - a.s.$$

Proof. By symmetry of the Brownian motion, we only have to prove the limsup result. The invariance of the Brownian motion by time inversion of Proposition 4.18 implies that

$$\limsup_{t\to\infty} W_t = \limsup_{u\to 0} \frac{1}{u} B_u \quad \text{where} \quad B_u := uW_{1/u}\mathbf{1}_{\{u\neq 0\}}$$

defines a Brownian motion. Then, it follows from the Zero-One law of Theorem 4.15 that $C_0 := \limsup_{t\to\infty} W_t$ is deterministic. By the symmetry of the Brownian motion, we see that $C_0 \in \mathbb{R}_+ \cup \{\infty\}$.

By the translation invariance of the Brownian motion, we see that

$$C_0 = \limsup_{t\to\infty}(W_t - W_s) \quad \text{in distribution for every} \quad s \geq 0.$$

Then, if $C_0 < \infty$, it follows that

$$e^{-\lambda C_0} = \mathbb{E}\left[e^{-\lambda C_0 + \lambda W_s}\right] = e^{-\lambda C_0 + \lambda^2 s/2}$$

which can not happen. Hence $C_0 = \infty$. $\diamondsuit$

Another consequence is the following result which shows the complexity of the sample paths of the Brownian motion.

Proposition 4.21. *For any $t_0 \geq 0$, we have*

$$\liminf_{t \searrow t_0} \frac{W_t - W_{t_0}}{t - t_0} \;=\; -\infty \quad \text{and} \quad \limsup_{t \searrow t_0} \frac{W_t - W_{t_0}}{t - t_0} \;=\; \infty\,.$$

Proof. From the invariance of the Brownian motion by time translation, it is sufficient to consider $t_0 = 0$. From Proposition 4.18, $B_t := tW_{1/t}$ defines a Brownian motion. Since $W_t/t = B_{1/t}$, it follows that the behavior of W_t/t for $t \searrow 0$ corresponds to the behavior of B_u for $u \nearrow \infty$. The required limit result is then a restatement of Proposition 4.20. ♢

We conclude this section by the law of the iterated logarithm for the Brownian motion. This result will not be used in our applications to finance, and is only reported for completeness. We shall organize its proof in the subsequent problem set.

Theorem 4.22. *For a Brownian motion W, we have*

$$\limsup_{t\to 0} \frac{W_t}{\sqrt{2t\ln(\ln\frac{1}{t})}} \;=\; 1 \quad \text{and} \quad \liminf_{t\to 0} \frac{W_t}{\sqrt{2t\ln(\ln\frac{1}{t})}} \;=\; -1, \quad \mathbb{P}-a.s.$$

In particular, this result shows that the Brownian motion is nowhere $\frac{1}{2}$−Hölder continuous, see Exercise 4.4.

Exercise 4.23. *(Law of Iterated Logarithm)*
Let W be a Brownian motion, and $h(t) := 2t\ln(\ln(1/t))$. We want to proove the Law of Iterated Logarithm :

$$\limsup_{t\searrow 0} \frac{W_t}{\sqrt{h(t)}} \;=\; 1 \quad a.s.$$

1. (a) *For $\lambda, T > 0$ with $2\lambda T < 1$, prove that*

$$\mathbb{P}\big[\max_{0\leq t\leq T}\{W_t^2 - t\} \geq \alpha\big] \leq e^{-\lambda\alpha}\mathbb{E}\big[e^{\lambda(W_T^2 - T)}\big].$$

   (b) *For $\theta, \eta \in (0,1)$, and $\lambda_n := \big[2\theta^n(1+\eta)\big]^{-1}$, deduce that:*

$$\mathbb{P}\left[\max_{0\leq t\leq \theta^n}\{W_t^2 - t\} \;\geq\; (1+\eta)^2 h(\theta^n)\right] \;\leq\; e^{-1/2(1+\eta)}(1+\eta^{-1})^{1/2}|n\ln\theta|^{-(1+\eta)}.$$

   (c) *By the Borel Cantelli Lemma, justify that $\limsup_{t\searrow 0}\frac{W_t^2}{h(t)} \leq \frac{(1+\eta)^2}{\theta}$, $\mathbb{P}-a.s.$*

   (d) *Conclude that $\limsup_{t\searrow 0}\frac{W_t}{\sqrt{h(t)}} \leq 1$, $\mathbb{P}-a.s.$*

2. *For $\theta \in (0,1)$, consider the event sets*

$$A_n := \left\{ W_{\theta^n} - W_{\theta^{n+1}} \geq \sqrt{1-\theta}\sqrt{h(\theta^n)} \right\}, \quad n \geq 1.$$

(a) *Using the inequality $\int_x^\infty e^{-u^2/2}du \geq \frac{xe^{-x^2/2}}{1+x^2}$, show that for some constant C*

$$\mathbb{P}[A_n] \geq \frac{C}{n\sqrt{\ln(n)}} \quad \textit{for} \quad n \quad \textit{sufficiently large.}$$

(b) *By the Borel-Cantelli Lemma, deduce that $W_{\theta^n} - W_{\theta^{n+1}} \geq \sqrt{1-\theta}\sqrt{h(\theta^n)}$, $\mathbb{P}-$a.s.*

(c) *Combining with question (1d), show that $\limsup_{t\searrow 0} \frac{W_t}{\sqrt{h(t)}} \geq \sqrt{1-\theta} - 4\theta$, $\mathbb{P}-$a.s.*

(d) *Deduce that $\limsup_{t\searrow 0} \frac{W_t}{\sqrt{h(t)}} = 1$, $\mathbb{P}-$a.s.*

Solution of Exercise 4.23
1. We first show that

$$\limsup_{t\searrow 0} \frac{W_t}{\sqrt{h(t)}} \;\leq\; 1 \quad \text{a.s.} \tag{4.6}$$

Let $T > 0$ and $\lambda > 0$ be such that $2\lambda T < 1$. Notice that $\{W_t^2 - t, t \leq T\}$ is a martingale. Then $\{e^{\lambda(W_t^2 - t)}, t \leq T\}$ is a nonnegative submartingale, by the Jensen inequality. It follows from the Doob maximal inequality for submartingales that for all $\alpha \geq 0$,

$$\begin{aligned}
\mathbb{P}\Big[\max_{0\leq t\leq T}\{W_t^2 - t\} \;\geq\; \alpha\Big] \;&=\; \mathbb{P}\big[\max_{0\leq t\leq T} e^{\lambda(W_t^2-t)} \;\geq\; e^{\lambda\alpha}\big] \\
&\leq\; e^{-\lambda\alpha}\mathbb{E}\big[e^{\lambda(W_T^2-T)}\big] \;=\; \frac{e^{-\lambda(\alpha+T)}}{\sqrt{1-2\lambda T}}.
\end{aligned}$$

Then, for $\theta, \eta \in (0,1)$, and

$$\alpha_k \;:=\; (1+\eta)^2 h(\theta^k), \quad \lambda_k \;:=\; [2\theta^k(1+\eta)]^{-1}, \quad k \in \mathbb{N},$$

we have

$$\mathbb{P}\left[\max_{0\leq t\leq \theta^k}(W_t^2 - t) \;\geq\; (1+\eta)^2 h(\theta^k)\right] \;\leq\; e^{-1/2(1+\eta)}\left(1+\eta^{-1}\right)^{\frac{1}{2}}(-k\log\theta)^{-(1+\eta)}.$$

Since $\sum_{k\geq 0} k^{-(1+\eta)} < \infty$, it follows from the Borel-Cantelli lemma that, for almost all $\omega \in \Omega$, there exists a natural number $K^{\theta,\eta}(\omega)$ such that for all $k \geq K^{\theta,\eta}(\omega)$,

$$\max_{0\leq t\leq\theta^k}(W_t^2(\omega) - t) \quad < \quad (1+\eta)^2 h(\theta^k).$$

In particular, for all $t \in (\theta^{k+1}, \theta^k]$, $W_t^2(\omega) - t < (1+\eta)^2 h(\theta^k) \leq (1+\eta)^2 h(t)/\theta$, and therefore:

$$\limsup_{t \searrow 0} \frac{W_t^2}{h(t)} \;=\; \limsup_{t \searrow 0} \frac{W_t^2 - t}{h(t)} \;<\; \frac{(1+\eta)^2}{\theta} \quad \text{a.s.}$$

and the required result follows by letting θ tend to 1 and η to 0 along the rationals.

2. We now show the converse inequality. Let $\theta \in (0,1)$ be fixed, and define:

$$A_n := \left\{ W_{\theta^n} - W_{\theta^{n+1}} \geq \sqrt{1-\theta}\sqrt{h(\theta^n)} \right\}, \quad n \geq 1.$$

Using the inequality $\int_x^\infty e^{-u^2/2} du \geq \frac{x e^{-x^2/2}}{1+x^2}$ with $x = x_n := \sqrt{\frac{h(\theta^n)}{\theta^n}}$, we see that

$$\mathbb{P}[A_n] = \mathbb{P}\Big[\frac{W_{\theta^n} - W_{\theta^{n+1}}}{\sqrt{\theta^n - \theta^{n+1}}} \geq x_n \Big] \geq \frac{e^{-x_n^2/2}}{\sqrt{2\pi}(x_n + \frac{1}{x_n})} \geq \frac{C}{n\sqrt{\ln(n)}}$$

for $n > -1/\ln\theta$, and some constant $C > 0$. Then $\sum_n \mathbb{P}[A_n] = \infty$. Since the events A_n's are independent, it follows from the Borel-Cantelli lemma that $W_{\theta^n} - W_{\theta^{n+1}} \geq \sqrt{1-\theta}\sqrt{h(\theta^n)}$, $\mathbb{P}-$a.s. Since $(-W)$ is a Brownian motion, it satisfies (4.6). Then

$$\frac{W_{\theta^n}}{\sqrt{h(\theta^n)}} \geq \sqrt{1-\theta} - 4\theta, \quad \text{and therefore} \quad \limsup_{n\to\infty} \frac{W_{\theta^n}}{\sqrt{h(\theta^n)}} \geq \sqrt{1-\theta} - 4\theta, \mathbb{P} - \text{a.s.}$$

We finally send $\theta \searrow 0$ along the rationals to conclude that $\limsup_{t \searrow 0} \frac{W_t}{\sqrt{h(t)}} \geq 1$.

♢

4.7 Quadratic variation

In this section, we consider a sequence of partitions $\pi^n = (t_i^n)_{i \geq 1} \subset \mathbb{R}^+$, $n \geq 1$, such that

$$\Delta t_i^n := t_i^n - t_{i-1}^n \geq 0 \quad \text{and} \quad |\pi^n| := \sup_{i \geq 1} |\Delta t_i^n| \longrightarrow 0 \;\; \text{as} \;\; n \to \infty, \tag{4.7}$$

where we set $t_0^n := 0$, and we define the discrete quadratic variation:

$$\mathrm{QV}_t^{\pi^n}(W) := \sum_{i \geq 1} \left| W_{t_i^n \wedge t} - W_{t_{i-1}^n \wedge t} \right|^2 \quad \text{for all} \quad n \geq 1. \tag{4.8}$$

As we shall see shortly, the Brownian motion has infinite total variation, see (4.9) below. In particular, this implies that classical integration theories are not suitable for the case of the Brownian motion. The key-idea in order to define an integration theory with respect to the Brownian motion is the following result which states that the quadratic variation defined as the $\mathbb{L}^2-$limit of (4.8) is finite.

Before stating the main result of this section, we observe that the quadratic variation (along any subdivision) of a continuously differentiable function f converges to zero. Indeed, $\sum_{t_i\le t}|f(t_{i+1})-f(t_i)|^2 \le \|f'\|^2_{\mathbb{L}^\infty([0,t])}\sum_{t_i\le t}|t_{i+1}-t_i|^2 \longrightarrow 0$. Because of the non-differentiability property stated in Proposition 4.21, this result does not hold for the Brownian motion.

Proposition 4.24. *Let W be a standard Brownian motion in $\mathbb{R}$, and $(\pi^n)_{n\ge1}$ a partition as in* (4.7). *Then the quadratic variation of the Brownian motion is finite and given by:*

$$\langle W\rangle_t \;:=\; \mathbb{L}^2-\lim_{n\to\infty} QV_t^{\pi^n}(W) \;=\; t \;\; \textit{for all} \;\; t\ge 0.$$

Proof. We directly compute that:

$$\begin{aligned}
\mathbb{E}\Big[\big(\mathrm{QV}_t^{\pi^n}(W)-t\big)^2\Big] &= \mathbb{E}\Big[\Big(\sum_{i\ge1}\big|W_{t_i^n\wedge t}-W_{t_{i-1}^n\wedge t}\big|^2-(t_i^n\wedge t-t_{i-1}^n\wedge t)\Big)^2\Big]\\
&= \sum_{i\ge1}\mathbb{E}\Big[\Big(\big|W_{t_i^n\wedge t}-W_{t_{i-1}^n\wedge t}\big|^2-(t_i^n\wedge t-t_{i-1}^n\wedge t)\Big)^2\Big]\\
&= 2\sum_{i\ge1}(t_i^n\wedge t-t_{i-1}^n\wedge t)^2 \;\le\; 2t|\pi^n|
\end{aligned}$$

by the independence of the increments of the Brownian motion and the fact that $W_{t_i^n\wedge t}-W_{t_{i-1}^n\wedge t}\sim\mathcal{N}(0,t_i^n\wedge t-t_{i-1}^n\wedge t)$. ♢

Remark 4.25. Proposition 4.24 has a natural direct extension to the multi-dimensional setting. Let W be a standard Brownian motion in $\mathbb{R}^d$, then:

$$\sum_{t_i^n\le t}\big(W_{t_{i+1}^n}-W_{t_i^n}\big)\big(W_{t_{i+1}^n}-W_{t_i^n}\big)^{\mathrm{T}} \;\longrightarrow\; t\, I_d, \quad \text{in } \;\mathbb{L}^2, \quad \text{for all } \; t\ge 0,$$

where I_d is the identity matrix of $\mathbb{R}^d$. We leave the verification of this result as an exercise.

The convergence result of Proposition 4.24 can be improved for partition π^n whose mesh $|\pi^n|$ satisfies a fast convergence to zero. As a complement, the following result considers the dyadic partition $\delta^n=(\delta_i^n)_{i\ge1}$ defined by:

$$\delta_i^n := i2^{-n}\,, \quad \text{for integers} \quad i\ge 0 \text{ and } n\ge 1.$$

Proposition 4.26. *Let W be a standard Brownian motion in $\mathbb{R}$. Then:*

$$\mathbb{P}\Big[\lim_{n\to\infty} V_t^{\delta^n}(W)=t\,, \; \textit{for every } t\ge 0\Big] \;=\; 1.$$

Before proceeding to the proof of this result, we make some important comments.

Remark 4.27. Inspecting the proof of Proposition 4.26, we see that the quadratic variation along any subdivision $0 = s_0^n < \ldots < s_n^n = t$ satisfies:

$$\sum_{i=1}^{n} \left| W_{s_{i+1}^n} - W_{s_i^n} \right|^2 \longrightarrow t \quad \mathbb{P}-\text{a.s. whenever} \quad \sum_{n \geq 1} \sup_{1 \leq i \leq n} |s_{i+1}^n - s_i^n| < 0.$$

Remark 4.28. We finally observe that Proposition 4.24 implies that the total variation of the Brownian motion infinite:

$$\mathbb{L}^2 - \lim_{n \to \infty} \sum_{i \geq 1} \left| W_{t_i^n \wedge t} - W_{t_{i-1}^n \wedge t} \right| \quad = \quad \infty. \tag{4.9}$$

This follows from the inequality

$$\mathrm{QV}_t^{\pi^n}(W) \quad \leq \quad \max_{i \geq 1} \left| W_{t_i^n \wedge t} - W_{t_{i-1}^n \wedge t} \right| \sum_{i \geq 1} \left| W_{t_i^n \wedge t} - W_{t_{i-1}^n \wedge t} \right|,$$

together with the fact that $\max_{i \geq 1} \left| W_{t_i^n \wedge t} - W_{t_{i-1}^n \wedge t} \right| \longrightarrow 0$, $\mathbb{P}-$a.s., due to the continuity of the Brownian motion. For this reason, the Stieltjes theory of integration does not apply to the Brownian motion.

Proof of Proposition 4.26 We shall simply denote $V_t^n := V_t^{\delta^n}(W)$.
(i) We first fix $t > 0$ and show that $V_t^n \longrightarrow t$ $\mathbb{P}-$a.s. as $n \to \infty$, or equivalently:

$$\sum_{t_i^n \leq t} Z_i \stackrel{n \to \infty}{\longrightarrow} 0 \quad \mathbb{P}-\text{a.s. where} \quad Z_i := \Big(\big(W_{t_{i+1}^n} - W_{t_i^n} \big)^2 - 2^{-n} \Big).$$

Observe that $\mathbb{E}[Z_i Z_j] = 0$ for $i \neq j$, and $\mathbb{E}[Z_j^2] = C 2^{-2n}$ for some constant $C > 0$. Then

$$\sum_{n \leq N} \mathbb{E}\Big[\Big(\sum_{t_i^n \leq t} Z_i \Big)^2 \Big] = \sum_{n \leq N} \sum_{t_i^n \leq t} \mathbb{E}[Z_i^2] = C \sum_{n \leq N} 2^{-n} \sum_{t_i^n \leq t} 2^{-n} = C \sum_{n \leq N} 2^{-n}.$$

Then, it follows from the monotone convergence theorem that

$$\mathbb{E}\Big[\sum_{n \geq 1} \Big(\sum_{t_i^n \leq t} Z_i \Big)^2 \Big] \quad \leq \quad \liminf_{N \to \infty} \sum_{n \leq N} \mathbb{E}\Big[\Big(\sum_{t_i^n \leq t} Z_i \Big)^2 \Big] \ < \ \infty.$$

In particular, this shows that the series $\sum_{n \geq 1} \Big(\sum_{t_i^n \leq t} Z_i \Big)^2$ is a.s. finite, and therefore $\sum_{t_i^n \leq t} Z_i \longrightarrow 0$ $\mathbb{P}-$a.s. as $n \to \infty$.
(ii) From the first step of this proof, we can find a zero measure set N_s for each rational number $s \in \mathbb{Q}$. For an arbitrary $t \geq 0$, let (s_p) and (s'_p) be two monotonic sequences of rational numbers with $s_p \nearrow t$ and $s'_p \searrow t$. Then, except on the zero-measure set $N := \cup_{s \in \mathbb{Q}} N_s$, it follows from the monotonicity of the quadratic variation that

$$\begin{aligned} s_p \ = \ \lim_{n \to \infty} V_{s_p}^n \ &\leq \ \liminf_{n \to \infty} V_t^n \\ &\leq \ \limsup_{n \to \infty} V_t^n \ \leq \ \lim_{n \to \infty} V_{s'_p}^n \ = \ s'_p. \end{aligned}$$

Sending $p \to \infty$ shows that $V_t^n \longrightarrow t$ as $n \to \infty$ for every ω outside the zero-measure set N. $\diamondsuit$

Chapter 5

Stochastic integration with respect to the Brownian motion

Recall from (4.9) that the total variation of the Brownian motion is infinite:

$$\lim_{n\to\infty} \sum_{t_i^n \le t} \left| W_{t_{i+1}^n} - W_{t_i^n} \right| \quad = \quad \infty, \quad \mathbb{P}-\text{a.s.}$$

Because of this property, one can not hope to define the stochastic integral with respect to the Brownian motion pathwise. To understand this, let us forget for a moment about stochastic processes. Let $\varphi, f : [0,1] \longrightarrow \mathbb{R}$ be continuous functions, and consider the Riemann sum:

$$S_n \quad := \quad \sum_{t_i^n \le 1} \varphi\left(t_{i-1}^n\right) \left[f\left(t_i^n\right) - f\left(t_{i-1}^n\right) \right].$$

Then, if the total variation of f is infinite, one can not guarantee that the above sum converges for every continuous function φ.

In order to circumvent this limitation, we shall make use of the finiteness of the quadratic variation of the Brownian motion, which allows to obtain an $\mathbb{L}^2-$definition of stochastic integration.

5.1 Stochastic integrals of simple processes

Throughout this section, we fix a final time $T > 0$. A process ϕ is called *simple* if there exists a strictly increasing sequence $(t_n)_{n\ge0}$ in $\mathbb{R}$ and a sequence of random variables $(\varphi_n)_{n\ge0}$ such that

$$\phi_t = \varphi_0 \mathbf{1}_{\{0\}}(t) + \sum_{n=0}^{\infty} \varphi_n \mathbf{1}_{(t_n, t_{n+1}]}(t), \qquad t \ge 0,$$

and

$$\varphi_n \text{ is } \mathcal{F}_{t_n} - \text{measurable for every } n \geq 0 \quad \text{and} \quad \sup_{n\geq 0} \|\varphi_n\|_\infty < \infty.$$

We shall denote by $\mathcal{S}$ the collection of all simple processes. For $\phi \in \mathcal{S}$, we define its stochastic integral with respect to the Brownian motion by:

$$I_t^0(\phi) \;:=\; \sum_{n\geq 0} \varphi_n \left(W_{t\wedge t_{n+1}} - W_{t\wedge t_n}\right), \qquad 0 \leq t \leq T. \tag{5.1}$$

By this definition, we immediately see that:

$$\mathbb{E}\left[I_t^0(\phi)|\mathcal{F}_s\right] = I_s(\phi) \quad \text{for} \quad 0 \leq s \leq t, \tag{5.2}$$

i.e. $\{I_t^0(\phi), t \geq 0\}$ is a martingale. We also calculate that

$$\mathbb{E}\left[I_t^0(\phi)^2\right] = \mathbb{E}\left[\int_0^t |\phi_s|^2 ds\right] \quad \text{for} \quad t \geq 0. \tag{5.3}$$

Exercise 5.1. *Prove properties* (5.2) *and* (5.3).

Our objective is to extend I^0 to a stochastic integral operator I acting on the larger set

$$\mathbb{H}^2 := \Big\{\phi : \text{ measurable, } \mathbb{F} - \text{adapted processes with } \mathbb{E}\Big[\int_0^T |\phi_t|^2 dt\Big] < \infty\Big\},$$

which is a Hilbert space when equipped with the norm

$$\|\phi\|_{\mathbb{H}^2} \quad := \quad \Big(\mathbb{E}\Big[\int_0^T |\phi_t|^2 dt\Big]\Big)^{1/2}.$$

The extension of I^0 to $\mathbb{H}^2$ is crucially based on the following density result.

Proposition 5.2. *The set of simple processes $\mathcal{S}$ is dense in $\mathbb{H}^2$, i.e. for every $\phi \in \mathbb{H}^2$, there is a sequence $\left(\phi^{(n)}\right)_{n\geq 0}$ of processes in $\mathcal{S}$ such that $\|\phi - \phi^{(n)}\|_{\mathbb{H}^2} \longrightarrow 0$ as $n \to \infty$.*

The proof of this result is reported in the Complements section 5.4.

5.2 Stochastic integrals of processes in $\mathbb{H}^2$

5.2.1 Construction

We now consider a process $\phi \in \mathbb{H}^2$, and we intend to define the stochastic integral $I_T(\phi)$ for every $T \geq 0$ by using the density of simple processes.

a. From Proposition 5.2, there is a sequence $\left(\phi^{(n)}\right)_{n\geq 0}$ which approximates ϕ in the sense that $\|\phi-\phi^{(n)}\|_{\mathbb{H}^2} \longrightarrow 0$ as $n\to\infty$. We next observe from (5.3) that, for every $t\geq 0$:

$$\begin{aligned}\left\|I_t^0\left(\phi^{(n)}\right)-I_t^0\left(\phi^{(m)}\right)\right\|_{\mathbb{L}^2}^2 &= \mathbb{E}\left[I_t^0\left(\phi^{(n)}-\phi^{(m)}\right)^2\right]\\ &= \mathbb{E}\left[\int_0^t |\phi_s^{(n)}-\phi_s^{(m)}|^2 ds\right]\\ &= \left\|\phi^{(n)}-\phi^{(m)}\right\|_{\mathbb{H}^2}^2\end{aligned}$$

converges to zero as $n,m\to\infty$. This shows that the sequence $\left(I_t^0\left(\phi^{(n)}\right)\right)_{n\geq 0}$ is a Cauchy sequence in $\mathbb{L}^2$, and therefore

$$I_t^0\left(\phi^{(n)}\right)\longrightarrow I_t(\phi) \quad \text{in } \mathbb{L}^2 \text{ for some random variable} \quad I_t(\phi).$$

b. We next show that the limit $I_t(\phi)$ does not depend on the choice of the approximating sequence $\left(\phi^{(n)}\right)_n$. Indeed, for another approximating sequence $\left(\psi^{(n)}\right)_n$ of ϕ, we have

$$\begin{aligned}\left\|I_t^0\left(\psi^{(n)}\right)-I_t(\phi)\right\|_{\mathbb{L}^2} &\leq \left\|I_t^0\left(\psi^{(n)}\right)-I_t^0\left(\phi^{(n)}\right)\right\|_{\mathbb{L}^2}+\left\|I_t^0\left(\phi^{(n)}\right)-I_t(\phi)\right\|_{\mathbb{L}^2}\\ &\leq \left\|\psi^{(n)}-\phi^{(n)}\right\|_{\mathbb{H}^2}+\left\|I_t^0\left(\phi^{(n)}\right)-I_t(\phi)\right\|_{\mathbb{L}^2}\\ &\longrightarrow\ 0 \text{ as } n\to\infty.\end{aligned}$$

c. Observe that the above construction applies without any difficulty if the time index $t\leq T$ is replaced by a stopping time τ with values in $[0,T]$. The only ingredient needed for this is the Doob's optional sampling theorem 3.11. The notation of the stochastic integral is naturally extended to $I_\tau(\phi)$ for all such stopping time.

The following result summarizes the above construction.

Theorem 5.3. *For $\phi\in\mathbb{H}^2$ and a stopping time τ with values in $[0,T]$, the stochastic integral denoted*

$$I_\tau(\phi) \quad := \quad \int_0^\tau \phi_s dW_s$$

is the unique limit in $\mathbb{L}^2$ of the sequence $\left(I_\tau^0(\phi^{(n)})\right)_n$ for every choice of an approximating sequence $\left(\phi^{(n)}\right)_n$ of ϕ in $\mathbb{H}^2$.

5.2.2 The stochastic integral as a continuous process

For every $\phi\in\mathbb{H}^2$, the previous theorem defined a family $\{I_\tau(\phi),\tau\}$ where τ ranges in the set of all stopping times with values in $[0,T]$. We now aim at aggregating this family into a process $\{I_t(\phi),t\in[0,T]\}$ so that

$$I_\tau(\phi)(\omega) \quad = \quad I_{\tau(\omega)}(\phi)(\omega) \ = \ I_T(\phi\mathbf{1}_{[0,\tau]})(\omega). \tag{5.4}$$

The meaning of (5.4) is the following. For a stopping time τ with values in $[0,T]$, and a process $\phi \in \mathbb{H}^2$, we may compute the stochastic integral of ϕ with respect to the Brownian motion on $[0,\tau]$ either by $I_\tau(\phi)$ or by $I_T(\phi \mathbf{1}_{[0,\tau]})$. Therefore, for the consistency of the stochastic integral operator, we have to verify that $I_\tau(\phi) = I_T(\phi \mathbf{1}_{[0,\tau]})$.

Proposition 5.4. *For a process $\phi \in \mathbb{H}^2$ and a stopping time τ with values in $[0,T]$, we have $I_\tau(\phi) = I_T(\phi \mathbf{1}_{[0,\tau]})$.*

Proof. Consider the approximation of τ by the decreasing sequence of stopping times $\tau_n := (\lfloor n\tau \rfloor + 1)/n$. Let $t_i := i/n$, and observe that $\mathbf{1}_{[0,\tau_n]} = \sum_i \mathbf{1}_{[t_i,t_{i+1})}(\tau)\mathbf{1}_{(0,t_{i+1}]}$ is a simple process. Then, the equality

$$I_{\tau_n}(\phi) = I_T(\phi \mathbf{1}_{[0,\tau_n]}) \quad \text{for simple processes} \quad \phi \in \mathcal{S} \tag{5.5}$$

is trivial. Since $\phi \mathbf{1}_{[0,\tau_n]} \longrightarrow \phi \mathbf{1}_{[0,\tau]}$ in $\mathbb{H}^2$, it follows that $I_T(\phi \mathbf{1}_{[0,\tau_n]}) \longrightarrow I_T(\phi \mathbf{1}_{[0,\tau]})$ in $\mathbb{L}^2$. Then, by the pathwise continuity of $\{I_t(\phi), t \in [0,T]\}$, we deduce that the proposition holds true for simple processes.

Now for $\phi \in \mathbb{H}^2$ with an approximating sequence $\phi^n \longrightarrow \phi$ in $\mathbb{H}^2$, we have $I_\tau(\phi^n) = I_T(\phi^n \mathbf{1}_{[0,\tau]})$ for all n, and we obtain the required result by sending n to infinity. $\diamondsuit$

In the previous proof we used the fact that, for a simple process $\phi \in \mathcal{S}$, the process $\{I_t(\phi), t \in [0,T]\}$ is pathwise continuous. The next result extends this property to any $\phi \in \mathbb{H}^2$.

Proposition 5.5. *Let ϕ be a process in $\mathbb{H}^2$. Then the process $\{I_t(\phi), t \in [0,T]\}$ has continuous sample paths a.s.*

Proof. Denote $M_t := I_t(\phi)$. By definition of the stochastic integral, M_t is the $\mathbb{L}^2-$limit of $M_t^n := I_t^0(\phi^n)$ for some sequence $(\phi^n)_n$ of simple processes converging to ϕ in $\mathbb{H}^2$. By definition of the stochastic integral of simple integrands in (5.1), notice that the process $\{M_t^n - M_t^m = I_t^0(\phi^n) - I_t^0(\phi^m), t \geq 0\}$ is a.s. continuous and the process $(|M_t^n - M_t^m|)_{m \geq n}$ is a non-negative submartingale. We then deduce from the Doobs maximal inequality of Theorem 3.15 that:

$$\mathbb{E}\left[(|M^n - M^m|_t^*)^2\right] \quad \leq \quad 4\,\mathbb{E}\left[\int_0^t |\phi_s^n - \phi_s^m|^2 ds\right].$$

This shows that the sequence $(M^n)_n$ is a Cauchy sequence in the Banach space of continuous processes endowed with the norm $\mathbb{E}\left[\sup_{[0,T]}|X_s|^2\right]$. Then M^n converges towards a continuous process $\bar{M}$ in the sense of this norm. We know however that $M_t^n \longrightarrow M_t := I_t(\phi)$ in $\mathbb{L}^2$ for all $t \in [0,T]$. By passing to subequences we may deduce that $\bar{M} = M$ is continuous. $\diamondsuit$

5.2.3 Martingale property and the Itô isometry

We finally show that the uniquely defined limit $I_t(\phi)$ satisfies the analogue properties of (5.2)-(5.3).

Proposition 5.6. *For $\phi \in \mathbb{H}^2$ and $t \le T$, we have:*

- *Martingale property:* $\mathbb{E}\left[I_T(\phi)|\mathcal{F}_t\right] = I_t(\phi)$,
- *Itô isometry:* $\mathbb{E}\left[I_T(\phi)^2\right] = \|\phi\|^2_{\mathbb{H}^2}$.

Proof. To see that the martingale property holds, we directly compute with $\left(\phi^{(n)}\right)_n$ an approximating sequence of ϕ in $\mathbb{H}^2$ that

$$\begin{aligned}
\left\|\mathbb{E}\left[I_T(\phi)|\mathcal{F}_t\right] - I_t(\phi)\right\|_{\mathbb{L}^2} &\le \left\|\mathbb{E}\left[I_T(\phi)|\mathcal{F}_t\right] - \mathbb{E}\left[I^0_t(\phi^{(n)})|\mathcal{F}_t\right]\right\|_{\mathbb{L}^2} \\
&\quad + \left\|\mathbb{E}\left[I^0_T(\phi^{(n)})|\mathcal{F}_t\right] - I_t(\phi)\right\|_{\mathbb{L}^2} \\
&= \left\|\mathbb{E}\left[I_T(\phi)|\mathcal{F}_t\right] - \mathbb{E}\left[I^0_T(\phi^{(n)})|\mathcal{F}_t\right]\right\|_{\mathbb{L}^2} \\
&\quad + \left\|I^0_t(\phi^{(n)}) - I_t(\phi)\right\|_{\mathbb{L}^2}
\end{aligned}$$

by (5.2). By the Jensen inequality and the law of iterated expectations, this provides

$$\left\|\mathbb{E}\left[I_T(\phi)|\mathcal{F}_t\right] - I_t(\phi)\right\|_{\mathbb{L}^2} \le \left\|I_T(\phi) - I^0_T(\phi^{(n)})\right\|_{\mathbb{L}^2} + \left\|I^0_t(\phi^{(n)}) - I_t(\phi)\right\|_{\mathbb{L}^2}$$

which implies the required result by sending n to infinity.

As for the Itô isometry, it follows from the $\mathbb{H}^2-$ convergence of $\phi^{(n)}$ towards ϕ and the $\mathbb{L}^2-$convergence of $I^0\left(\phi^{(n)}\right)$ towards $I(\phi)$, together with (5.3), that:

$$\mathbb{E}\left[\int_0^T |\phi_s|^2 dt\right] = \lim_{n\to\infty} \mathbb{E}\left[\int_0^T |\phi^{(n)}_t|^2 dt\right] = \lim_{n\to\infty} \mathbb{E}\left[I^0_T\left(\phi^{(n)}\right)^2\right] \quad = \quad \mathbb{E}\left[I_T(\phi)^2\right].$$

$\diamondsuit$

5.2.4 Deterministic integrands

We report the main message of this subsection in the following exercise.

Exercise 5.7. *Let $f : [0,T] \longrightarrow \mathbb{R}^d$ be a deterministic function with $\int_0^T |f(t)|^2 dt < \infty$.*

1. *Prove that*

$$\int_0^T f(t)\cdot dW_t \quad \textit{has distribution} \quad \mathcal{N}\left(0, \int_0^T |f(t)|^2 dt\right).$$

Hint: Use the closeness of the Gaussian space.

2. Prove that the process

$$\exp\left(\int_0^t f(s)\cdot dW_s - \frac{1}{2}\int_0^t |f(s)|^2 ds\right), \qquad t \geq 0$$

is a martingale.

5.3 Stochastic integration beyond $\mathbb{H}^2$ and Itô processes

Our next task is to extend the stochastic integration to integrands in the set

$$\mathbb{H}^2_{\text{loc}} := \left\{\phi \text{ measurable, } \mathbb{F}-\text{adapted } \int_0^T |\phi_s|^2 ds < \infty \text{ a.s.}\right\}. \tag{5.6}$$

To do this, we consider for every $\phi \in \mathbb{H}^2_{\text{loc}}$ the sequence of stopping times

$$\tau_n := \inf\left\{t > 0 : \int_0^t |\phi_u|^2 du \geq n\right\}.$$

Clearly, $(\tau_n)_n$ is non-decreasing sequence of stopping times and

$$\tau_n \longrightarrow \infty \quad \mathbb{P}-\text{a.s.} \quad \text{when } n \to \infty.$$

For fixed $n > 0$, the process $\phi^n := \phi_. \mathbf{1}_{.\wedge\tau_n}$ is in $\mathbb{H}^2$. Then the stochastic integral $I_t(\phi^n)$ is well-defined by Theorem 5.3. Since $(\tau_n)_n$ is a non-decreasing and $\mathbb{P}[\tau_n \geq t \text{ for some } n \geq 1] = 1$, it follows that the limit

$$I_t(\phi) \quad := \quad \text{(a.s.)} \lim_{n\to\infty} I_t(\phi^n) \tag{5.7}$$

exists (in fact $I_t(\phi) = I_t(\phi^n)$ for n sufficiently large, a.s.).

Remark 5.8. The above extension of the stochastic integral to integrands in $\mathbb{H}^2_{\text{loc}}$ does not imply that $I_t(\phi)$ satisfies the martingale property and the Itô isometry of Proposition 5.6. This issue will be further developed in the next subsection. However the continuity property of the stochastic integral is conserved because it is a pathwise property which is consistent with the pathwise definition of (5.7).

As a consequence of this remark, when the integrand ϕ is in $\mathbb{H}^2_{\text{loc}}$ but is not in $\mathbb{H}^2$, the stochastic integral fails to be a martingale, in general. This leads us to the notion of local martingale.

Definition 5.9. *An $\mathbb{F}-$adapted process $M = \{M_t,\ t \geq 0\}$ is a local martingale if there exists a sequence of stopping times $(\tau_n)_{n\geq 0}$ (called a localizing sequence) such that $\tau_n \longrightarrow \infty$ $\mathbb{P}-$a.s. as $n \to \infty$, and the stopped process $M^{\tau_n} = \{M_{t\wedge\tau_n},\ t \geq 0\}$ is a martingale for every $n \geq 0$.*

Proposition 5.10. *Let ϕ be a process in $\mathbb{H}^2_{\text{loc}}$. Then, for every $T > 0$, the process $\{I_t(\phi),\ 0 \le t \le T\}$ is a local martingale.*

Proof. The above defined sequence $(\tau_n)_n$ is easily shown to be a localizing sequence. The result is then a direct consequence of the martingale property of the stochastic integral of a process in $\mathbb{H}^2$. $\diamondsuit$

An example of local martingale which fails to be a martingale will be given in the next chapter.

Definition 5.11. *An Itô process X is a continuous-time process defined by:*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s \cdot dW_s, \qquad t \ge 0,$$

where μ and σ are measurable, $\mathbb{F}-$adapted processes with $\int_0^t (|\mu_s| + |\sigma_s|^2) ds < \infty$ a.s.

Remark 5.12. Observe that the process μ above is only assumed to be measurable and $\mathbb{F}-$adapted, so we may ask whether the process $\{\int_0^t \mu_s ds, t \le T\}$ is adapted. This is indeed true as a consequence of the density result of Proposition 5.2: Let $(\mu^n)_{n\ge1}$ be an approximation of μ in $\mathbb{H}^2$. This implies that $\int_0^t \mu_s^n ds \longrightarrow \int_0^t \mu_s ds$ a.s. along some subsequence. Since μ^n is a simple process, $\int_0^t \mu_s^n ds$ is $\mathcal{F}_t-$adapted, and so is the limit $\int_0^t \mu_s ds$.

We conclude this section by the following easy result:

Lemma 5.13. *Let $M = \{M_t,\ 0 \le t \le T\}$ be a local martingale bounded from below by some constant m, i.e. $M_t \ge m$ for all $t \in [0, T]$ a.s. Then M is a supermartingale.*

Proof. Let $(T_n)_n$ be a localizing sequence of stopping times for the local martingale M, i.e. $T_n \longrightarrow \infty$ a.s. and $\{M_{t\wedge T_n},\ 0 \le t \le T\}$ is a martingale for every n. Then :

$$\mathbb{E}\left[M_{t\wedge T_n} | \mathcal{F}_u\right] \;=\; M_{u\wedge T_n}, \qquad 0 \le u \le t \le T,$$

for every fixed n. We next send n to infinity. By the lower bound on M, we can use Fatou's lemma, and we deduce that :

$$\mathbb{E}\left[M_t | \mathcal{F}_u\right] \;\le\; M_u, \qquad 0 \le u \le t \le T\,,$$

which is the required inequality. $\diamondsuit$

Exercise 5.14. *Show that the conclusion of Lemma 5.13 holds true under the weaker condition that the local martingale M is bounded from below by a martingale.*

5.4 Complement: density of simple processes in $\mathbb{H}^2$

The proof of Proposition 5.2 is a consequence of the following Lemmas. Throughout this section, $t_i^n := i2^{-n}, i \geq 0$ is the sequence of dyadic numbers.

Lemma 5.15. *Let ϕ be a bounded $\mathbb{F}-$adapted process with continuous sample paths. Then ϕ can be approximated by a sequence of simple processes in $\mathbb{H}^2$.*

Proof. Define the sequence

$$\phi_t^{(n)} := \phi_0 \mathbf{1}_{\{0\}}(t) + \sum_{t_i^n \leq T} \phi_{t_i^n} \mathbf{1}_{(t_i^n, t_{i+1}^n]}(t), \qquad t \leq T.$$

Then, $\phi^{(n)}$ is a simple process for each $n \geq 1$. By the dominated convergence theorem, $\mathbb{E}\left[\int_0^T |\phi_t^{(n)} - \phi_t|^2\right] \longrightarrow 0$ as $n \to \infty$. ♢

Lemma 5.16. *Let ϕ be a bounded $\mathbb{F}-$progressively measurable process. Then ϕ can be approximated by a sequence of simple processes in $\mathbb{H}^2$.*

Proof. Notice that the process

$$\psi_t^{(k)} := k \int_{0 \vee (t - \frac{1}{k})}^t \phi_s ds, \qquad 0 \leq t \leq T,$$

is progressively measurable as the difference of two adapted continuous processes, see Proposition 3.2, and satisfies

$$\left\| \psi^{(k)} - \phi \right\|_{\mathbb{H}^2} \longrightarrow 0 \quad \text{as} \quad k \to \infty, \tag{5.8}$$

by the dominated convergence theorm. For each $k \geq 1$, we can find by Lemma 5.15 a sequence $\left(\psi^{(k,n)}\right)_{n \geq 0}$ of simple processes such that $\|\psi^{(k,n)} - \psi^{(k)}\|_{\mathbb{H}^2} \longrightarrow 0$ as $n \to \infty$. Then, for each $k \geq 0$, we can find n_k such that

$$\text{the process } \phi^{(k)} := \psi^{(k,n_k)} \quad \text{satisfies} \quad \left\| \phi^{(k)} - \phi \right\|_{\mathbb{H}^2} \longrightarrow 0 \text{ as } k \to \infty.$$

♢

Lemma 5.17. *Let ϕ be a bounded measurable and $\mathbb{F}-$adapted process. Then ϕ can be approximated by a sequence of simple processes in $\mathbb{H}^2$.*

Proof. In the present setting, the process $\psi^{(k)}$, defined in the proof of the previous Lemma 5.16, is measurable but is not known to be adapted. For each $\varepsilon > 0$, there is an integer $k \geq 1$ such that $\psi^\varepsilon := \psi^{(k)}$ satisfies $\|\psi^\varepsilon - \phi\|_{\mathbb{H}^2} \leq \varepsilon$. Then, with $\phi_t = \phi_0$ for $t \leq 0$:

$$\begin{aligned}
\|\phi - \phi_{.-h}\|_{\mathbb{H}^2} &\leq \|\phi - \psi^\varepsilon\|_{\mathbb{H}^2} + \left\|\psi^\varepsilon - \psi^\varepsilon_{.-h}\right\|_{\mathbb{H}^2} + \left\|\psi^\varepsilon_{.-h} - \phi_{.-h}\right\|_{\mathbb{H}^2} \\
&\leq 2\varepsilon + \left\|\psi^\varepsilon - \psi^\varepsilon_{.-h}\right\|_{\mathbb{H}^2}.
\end{aligned}$$

By the continuity of ψ^ε, this implies that

$$\limsup_{h \searrow 0} \|\phi - \phi_{.-h}\|_{\mathbb{H}^2} \quad \le \quad 4\varepsilon^2. \tag{5.9}$$

We now introduce

$$\varphi_n(t) \quad := \quad \mathbf{1}_{\{0\}}(t) + \sum_{i \ge 1} t_i^n \mathbf{1}_{(t_{i-1}^n, t_i^n]},$$

and

$$\phi_t^{(n,s)} \;:=\; \phi_{\varphi_n(t-s)+s}, \qquad t \ge 0, s \in (0,1].$$

Clearly $\phi^{n,s}$ is a simple adapted process, and

$$\begin{aligned}
\mathbb{E}\left[\int_0^T \int_0^1 |\phi_t^{(n,s)} - \phi_t|^2 ds dt\right] &= 2^n \mathbb{E}\left[\int_0^T \int_0^{2^{-n}} |\phi_t - \phi_{t-h}|^2 dh dt\right] \\
&= 2^n \int_0^{2^{-n}} \mathbb{E}\left[\int_0^T |\phi_t - \phi_{t-h}|^2 dt\right] dh \\
&\le \max_{0 \le h \le 2^{-n}} \mathbb{E}\left[\int_0^T |\phi_t - \phi_{t-h}|^2 dt\right]
\end{aligned}$$

which converges to zero as $n \to \infty$ by (5.9). Hence

$$\phi_t^{(n,s)}(\omega) \;\longrightarrow\; \phi_t(\omega) \quad \text{for almost every} \quad (s,t,\omega) \in [0,1] \times [0,T] \times \Omega,$$

and the required result follows from the dominated convergence theorem. $\diamondsuit$

Lemma 5.18. *The set of simple processes $\mathcal{S}$ is dense in $\mathbb{H}^2$.*

Proof. We only have to extend Lemma 5.17 to the case where ϕ is not necessarily bounded. This is easily achieved by applying Lemma 5.17 to the bounded process $\phi \wedge n$, for each $n \ge 1$, and passing to the limit as $n \to \infty$. $\diamondsuit$

Chapter 6

Itô Differential Calculus

In this chapter, we focus on the differential properties of the Brownian motion. To introduce the discussion, recall that $\mathbb{E}\left[W_t^2\right] = t$ for all $t \geq 0$. If standard differential calculus were valid in the present context, then one would expect that W_t^2 be equal to $M_t = 2\int_0^t W_s dW_s$. But the process M is a square integrable martingale on every finite interval $[0,T]$, and therefore $E[M_t] = M_0 = 0 \neq \mathbb{E}[W_t^2]$!

So, the standard differential calculus is not valid in our context. We should not be puzzled by this small calculation, as we already observed that the Brownian motion sample path has very poor regularity properties, has infinite total variation, and finite quadratic variation.

We can elaborate more on the above example by considering a discrete-time approximation of the stochastic integral M_t:

$$\begin{aligned} \sum_{i=1}^n 2W_{t_{i-1}}\left(W_{t_i} - W_{t_{i-1}}\right) &= -\sum_{i=1}^n \left(W_{t_i} - W_{t_{i-1}}\right)^2 + \sum_{i=1}^n \left(W_{t_i}^2 - W_{t_{i-1}}^2\right) \\ &= W_t^2 - \sum_{i=1}^n \left(W_{t_i} - W_{t_{i-1}}\right)^2, \end{aligned}$$

where $0 = t_0 < t_1 < \ldots, t_n = t$. We know that the latter sum converges in $\mathbb{L}^2$ towards t, the quadratic variation of the Brownian motion at time t (the convergence holds even $\mathbb{P}-$a.s. if one takes the dyadics as $(t_i)_i$). Then, by sending n to infinity, this shows that

$$\int_0^t 2W_s dW_s = W_t^2 - t, \quad t \geq 0. \tag{6.1}$$

In particular, by taking expectations on both sides, there is no contradiction anymore.

6.1 Itô's formula for the Brownian motion

The purpose of this section is to prove the Itô formula for the change of variable. Given a smooth function $f(t,x)$, we will denote by f_t, Df and $D^2 f$, the partial gradients with respect to t, to x, and the partial Hessian matrix with respect to x.

Theorem 6.1. *Let $f : \mathbb{R}_+ \times \mathbb{R}^d \longrightarrow \mathbb{R}$ be $C^{1,2}\left([0,T],\mathbb{R}^d\right)$. Then, with probability 1, we have:*

$$f(T,W_T) \quad = \quad f(0,0) + \int_0^T Df\,(t,W_t)\cdot dW_t + \int_0^T \left(f_t + \frac{1}{2}\mathrm{Tr}[D^2 f]\right)(t,W_t)dt$$

for every $T \geq 0$.

Proof. **1** We first fix $T > 0$, and we show that the above Itô's formula holds with probability 1. Also, without loss of generality, we may assume that f and its derivatives f_t, Df and D^2f are bounded by some constant $C > 0$. Indeed, we may reduce the analysis to the stochastic interval $[0,\tau_N]$, with $\tau_N := \inf\{t \geq 0 : |W_t| \geq N\}$, so that by the $C^{1,2}$ regularity of f, it follows that f, f_t, Df, D^2f are bounded in $[0,\tau_N]$, and we obtain the required result by sending $N \to \infty$.Finally, by possibly adding a constant to f, we may assume that $f(0,0) = 0$.

Let $\pi^n = (t_i^n)_{i\geq 0}$ be a partition of $\mathbb{R}_+$ with $t_0^n = 0$, and denote $n(T) := \sup\{i \geq 0 : t_i^n \leq T\}$, so that $t^n_{n(T)} \leq T < t^n_{n(T)+1}$. We also denote

$$\Delta_i^n W := W_{t_{i+1}^n} - W_{t_i^n} \quad \text{and} \quad \Delta_i^n t := t_{i+1}^n - t_i^n.$$

1.a We first decompose

$$\begin{aligned} f\left(t^n_{n(T)+1}, W_{t^n_{n(T)+1}}\right) \quad &= \quad \sum_{t_i^n \leq T}\left[f\left(t_{i+1}^n, W_{t_{i+1}^n}\right) - f\left(t_i^n, W_{t_{i+1}^n}\right)\right] \\ &\quad + \sum_{t_i^n \leq T}\left[f\left(t_i^n, W_{t_{i+1}^n}\right) - f\left(t_i^n, W_{t_i^n}\right)\right]. \end{aligned}$$

By a Taylor expansion, this provides:

$$\begin{aligned} I_T^n(Df) \quad &:= \quad \sum_{t_i^n \leq T} Df\left(t_i^n, W_{t_i^n}\right)\cdot \Delta_i^n W \\ &= \quad f\left(t^n_{n(T)+1}, W_{t^n_{n(T)+1}}\right) - \sum_{t_i^n\leq T} f_t\left(\tau_i^n, W_{t_{i+1}^n}\right)\Delta_i^n t \\ &\quad -\frac{1}{2}\sum_{t_i^n\leq T}\mathrm{Tr}\left[\left(D^2f\left(t_i^n,\xi_i^n\right) - D^2 f\left(t_i^n, W_{t_i^n}\right)\right)\Delta_i^n W \Delta_i^n W^{\mathrm{T}}\right] \\ &\quad -\frac{1}{2}\sum_{t_i^n\leq T}\mathrm{Tr}\left[D^2 f\left(t_i^n, W_{t_i^n}\right)\Delta_i^n W\Delta_i^n W^{\mathrm{T}}\right], \end{aligned} \tag{6.2}$$

where τ_i^n is a random variable with values in $\left[t_i^n, t_{i+1}^n\right]$, and $\xi_i^n = \lambda_i^n W_{t_i^n} + (1-\lambda_i^n)\,W_{t_{i+1}^n}$ for some random variable λ_i^n with values in $[0,1]$.

1.b Since a.e. sample path of the Brownian motion is continuous, and therefore uniformly continuous on the compact interval $[0, T+1]$, it follows that

$$f\left(t^n_{n(T)+1}, W_{t^n_{n(T)+1}}\right) \longrightarrow f(T, W_T), \quad \mathbb{P}-\text{a.s.}$$
$$\textstyle\sum_{t^n_i \le T} f_t\left(\tau^n_i, W_{t^n_{i+1}}\right)\Delta^n_i t \longrightarrow \int_0^T f_t(t, W_t)dt \quad \mathbb{P}-\text{a.s.}$$

Next, using again the above uniform continuity together with Proposition 4.24 and the fact that the $\mathbb{L}^2-$convergence implies the a.s. convergence along some subsequence, we see that:

$$\sum_{t^n_i \le T} \mathrm{Tr}\left[\left(D^2 f\left(t^n_i, \xi^n_i\right) - D^2 f\left(t^n_i, W_{t^n_i}\right)\right)\Delta^n_i W \Delta^n_i W^{\mathrm{T}}\right] \longrightarrow 0 \qquad \mathbb{P}-\text{a.s.}$$

1.c. For the last term in the decomposition (6.2), we estimate:

$$\begin{aligned}\Big|\sum_{t^n_i \le T} \mathrm{Tr}\Big[D^2 f\left(t^n_i, W_{t^n_i}\right)\Delta^n_i W \Delta^n_i W^{\mathrm{T}}\Big]\Big| &\le \Big|\sum_{t^n_i \le T} \mathrm{Tr}\left[D^2 f\left(t^n_i, W_{t^n_i}\right)\right]\Delta^n_i t\Big| \\ &+ \Big|\sum_{t^n_i \le T} \mathrm{Tr}\Big[D^2 f\left(t^n_i, W_{t^n_i}\right)\left(\Delta^n_i W \Delta^n_i W^{\mathrm{T}} - \Delta^n_i t I_d\right)\Big]\Big|\end{aligned} \tag{6.3}$$

Denote $\varphi_i := D^2 f\left(t^n_i, W_{t^n_i}\right)$ Since $|\varphi_i| \le C$ and φ_i is $\mathcal{F}_{t_{i-1}}-$measurable for all i, we have

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{t^n_i \le T} \mathrm{Tr}\Big[\varphi_i\big(\Delta^n_i W \Delta^n_i W^{\mathrm{T}} - \Delta^n_i t I_d\big)\Big]\Big)^2\Big] &\le C^2 \sum_{t^n_i \le T} \mathbb{E}\Big[\big||\Delta^n_i W|^2 - d\Delta^n_i t\big|^2\Big] \\ &= 2C^2 d^2 \sum_{t^n_i \le T} \left|\Delta^n_i t\right|^2 \\ &\le 2C^2 d^2 |\pi^n| T \longrightarrow 0.\end{aligned}$$

This $\mathbb{L}^2-$convergence implies that $\sum_{t^n_i \le T} \mathrm{Tr}\Big[\varphi_i\big(\Delta^n_i W \Delta^n_i W^{\mathrm{T}} - \Delta^n_i t I_d\big)\Big] \longrightarrow 0$, $\mathbb{P}-$a.s. along some subsequence. Then, it follows from (6.3) that, along some subsequence,

$$\sum_{t^n_i \le T} \mathrm{Tr}\left[D^2 f\left(t^n_i, W_{t^n_i}\right)\Delta^n_i W \Delta^n_i W^{\mathrm{T}}\right] \longrightarrow \int_0^T \mathrm{Tr}\left[D^2 f(t, W_t)\right] dt, \quad \mathbb{P}-\text{a.s.}$$

1.d In order to complete the proof of Itô's formula for fixed $T > 0$, it remains to prove that

$$I^n_T(Df) \longrightarrow \int_0^T Df(t, W_t)\cdot dW_t \ \mathbb{P}-\text{a.s. along some subsequence.} \tag{6.4}$$

Notice that $I^n_T(Df) = I^0_T\left(\phi^{(n)}\right)$ where $\phi^{(n)}$ is the simple process defined by

$$\phi^{(n)}_t = \sum_{t^n_i \le T} Df\left(t^n_i, W_{t^n_i}\right)\mathbf{1}_{[t^n_i, t^n_{i+1})}(t), \qquad t \ge 0.$$

Since Df is continuous, it follows from the proof of Proposition 5.2 that $\phi^{(n)} \longrightarrow \phi$ in $\mathbb{H}^2$ with $\phi_t := Df(t, W_t)$. Then $I_T^n(Df) \longrightarrow I_T(\phi)$ in $\mathbb{L}^2$, by the definition of the stochastic integral in Theorem 5.3, and (6.4) follows from the fact that the $\mathbb{L}^2$ convergence implies the a.s. convergence along some subsequence.

2. From the first step, we have the existence of subsets $N_t \subset \mathcal{F}$ for every $t \geq 0$ such that $\mathbb{P}[N_t] = 0$ and the Itô's formula holds on N_t^c, the complement of N_t. Of course, this implies that the Itô's formula holds on the complement of the set $N := \cup_{t\geq 0} N_t$. But this does not complete the proof of the theorem as this set is a non-countable union of zero measure sets, and is therefore not known to have zero measure. We therefore appeal to the continuity of the Brownian motion and the stochastic integral, see Proposition 3.15. By usual approximation along rational numbers, it is easy to see that, with probability 1, the Itô formula holds for every $T \geq 0$. ♢

Remark 6.2. Since, with probability 1, the Itô formula holds for every $T \geq 0$, it follows that the Itô's formula holds when the deterministic time T is replaced by a random time τ.

Remark 6.3. *(Itô's formula with generalized derivatives)* Let $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ be $C^2(\mathbb{R}^d \setminus K)$ for some compact subset K of $\mathbb{R}^d$. Assume that $f \in W^2(K)$, i.e. there is a sequence of functions $(f^n)_{n\geq 1}$ such that

$$f^n = f \text{ on } \mathbb{R}^d \setminus K, \quad f^n \in C^2(K) \quad \text{and } \|f_x^n - f_x^m\|_{\mathbb{L}^2(K)} + \|f_{xx}^n - f_{xx}^m\|_{\mathbb{L}^2(K)} \longrightarrow 0.$$

Then, Itô's formula holds true:

$$f(W_t) = f(0) + \int_0^t Df(W_s)dW_s + \frac{1}{2}\int_0^t D^2 f(W_s)ds,$$

where Df and $D^2 f$ are the generalized derivatives of f. Indeed, Itô's formula holds for f^n, $n \geq 1$, and we obtain the required result by sending $n \to \infty$.

A similar statement holds for a function $f(t, x)$.

Exercise 6.4. *Let W be a Brownian motion in $\mathbb{R}^d$ and consider the process*

$$X_t := X_0 + bt + \sigma W_t, \quad t \geq 0,$$

where b is a vector in $\mathbb{R}^d$ and σ is an $(d \times d)-$matrix. Let f be a $C^{1,2}(\mathbb{R}_+, \mathbb{R}^d)$ function. Show that

$$df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t) \cdot dX_t + \frac{1}{2}\, Tr\left[\frac{\partial^2 f}{\partial x \partial x^{\mathrm{T}}}(t, X_t)\sigma\sigma^{\mathrm{T}}\right].$$

Exercise 6.5. *Let W be a Brownian motion in $\mathbb{R}^d$, and consider the process*

$$S_t := S_0 \exp\left(bt + \sigma \cdot W_t\right), \quad t \geq 0,$$

where $b \in \mathbb{R}$ and $\sigma \in \mathbb{R}^d$ are given.

1. *For a $C^{1,2}$ function $f : \mathbb{R}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}$, show that $\{f(t, S_t), t \geq 0\}$ is an Itô process, and provide its dynamics.*

2. *Find a function f so that the process $\{f(t, S_t), t \geq 0\}$ is a local martingale.*

Exercise 6.6. *Let $f : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ be a $C^{1,2}$ function such that, for some constant $C > 0$,*

$$|f(t,x)| + |f'_t(t,x)| + |f'_x(t,x)| + |f''_{x,x}(t,x)| \leq C\exp(C|x|)$$

for all $(t,x) \in \mathbb{R}^+ \times \mathbb{R}$.

1. *If T is a bounded stopping time, show that*

$$\mathbb{E}(f(T,W_T)) = f(0,0) + \mathbb{E}\left[\int_0^T [f'_t(s,W_s) + \frac{1}{2}f''_{x,x}(s,W_s)]\,ds\right]$$

2. *When T is a bounded stopping time, compute $\mathbb{E}(W_T)$ and $\mathbb{E}(W_T^2)$.*

3. *Show that if T is a stopping time such that $\mathbb{E}(T) < +\infty$, then $\mathbb{E}(W_T) = 0$.*

4. *For every real number $a \neq 0$, we define*

$$\tau_a = \inf\{t \geq 0 : W_t = a\}.$$

   Is it a finite stopping time? Is it bounded? Deduce from question 3) that $\mathbb{E}(\tau_a) = +\infty$.

5. *Show that the law of τ_a is characterized by its Laplace transform :*

$$\mathbb{E}(\exp\{-\lambda\tau_a\}) = \exp(-\sqrt{2\lambda}|a|), \qquad \lambda \geq 0.$$

6. *From question 2), deduce the value of $\mathbb{P}(\tau_a < \tau_b)$, for $a > 0$ and $b < 0$.*

6.2 Extension to Itô processes

We next provide Itô's formula for a general Itô process with values in $\mathbb{R}^n$:

$$X_t := X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \qquad t \geq 0,$$

where μ and σ are adapted processes with values in $\mathbb{R}^n$ and $\mathcal{M}_{\mathbb{R}}(n,d)$, respectively and satisfying

$$\int_0^T |\mu_s| ds + \int_0^T |\sigma_s|^2 ds < \infty \qquad \text{a.s.}$$

Observe that stochastic integration with respect to the Itô process X reduces to the stochastic integration with respect to W: for any $\mathbb{F}-$adapted $\mathbb{R}^n-$valued process ϕ with $\int_0^T |\sigma_t^{\mathrm{T}}\phi_t|^2 dt + \int_0^T |\phi_t \cdot \mu_t| dt < \infty$,a.s.

$$\begin{aligned}\int_0^T \phi_t \cdot dX_t &= \int_0^T \phi_t \cdot \mu_t dt + \int_0^T \phi_t \cdot \sigma_t dW_t \\ &= \int_0^T \phi_t \cdot \mu_t dt + \int_0^T \sigma_t^{\mathrm{T}}\phi_t \cdot dW_t.\end{aligned}$$

Theorem 6.7. *Let $f : \mathbb{R}_+ \times \mathbb{R}^n \longrightarrow \mathbb{R}$ be $C^{1,2}([0,T],\mathbb{R}^n)$. Then, with probability 1, we have:*

$$\begin{aligned}f(T,X_T) &= f(0,0) + \int_0^T Df(t,X_t) \cdot dX_t \\ &\quad + \int_0^T \left(f_t(t,X_t) + \frac{1}{2}\mathrm{Tr}[D^2 f(t,X_t)\sigma_t\sigma_t^{\mathrm{T}}]\right) dt\end{aligned}$$

for every $T \geq 0$.

Proof. Let $\tau_N := \inf\left\{t : \ \max\left(|X_t - X_0|, \int_0^t \sigma_s^2 ds, \int_0^t |\mu_s| ds\right) \geq N\right\}$. Obviously, $\tau_N \to \infty$ a.s. when $N \to \infty$, and it is sufficient to prove Itô's formula on $[0,\tau_N]$, since any $t \geq 0$ can be reached by sending N to infinity. In view of this, we may assume without loss of generality that X, $\int_0^t \mu_s ds$, $\int_0^t \sigma_s^2 ds$ are bounded and that f has compact support. We next consider an approximation of the integrals defining X_t by step functions which are constant on intervals of time $(t_{i-1}^n, t_i^n]$ for $i = 1,\ldots,n$, and we denote by X^n the resulting simple approximating process. Notice that Itô's formula holds true for X^n on each interval $(t_{i-1}^n, t_i^n]$ as a direct consequence of Theorem 6.1. The proof of the theorem is then concluded by sending n to infinity, and using as much as needed the dominated convergence theorem. ♢

Exercise 6.8. *Let W be a Brownian motion in $\mathbb{R}^d$, and consider the process*

$$S_t \quad := \quad S_0 \exp\left(\int_0^t b_u du + \int_0^t \sigma_u \cdot dW_u\right), \quad t \geq 0,$$

where b and σ are measurable and $\mathbb{F}-$adapted processes with values in $\mathbb{R}$ and $\mathbb{R}^d$, repectively, with $\int_0^T |b_u| du + \int_0^T |\sigma_u|^2 du < \infty$, a.s. for all $T > 0$.

1. *For a function $f : \mathbb{R}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}$, show that $\{f(t,S_t), t \geq 0\}$ is an Itô process, and provide its dynamics.*

2. *Let γ be a measurable $\mathbb{F}-$adapted process with values in $\mathbb{R}$ with $\int_0^T |\gamma_u| du < \infty$, a.s. Show that the process $\{X_t := e^{-\int_0^t \gamma_u du} S_t, t \geq 0\}$ is an Itô process, and provide its dynamics.*

3. *Find a process γ such that $\{X_t, t \geq 0\}$ is a local martingale.*

Exercise 6.9. *Let W be a Brownian motion in $\mathbb{R}^d$, and X and Y be the Itô processes defined for all $t \geq 0$ by:*

$$X_t = X_0 + \int_0^t \mu_u^X du + \int_0^t \sigma_u^X \cdot dW_u$$
$$Y_t = Y_0 + \int_0^t \mu_u^Y du + \int_0^t \sigma_u^Y \cdot dW_u,$$

where $\mu^X, \mu^Y, \sigma^X, \sigma^Y$ are measurable and $\mathbb{F}-$adapted processes with appropriate dimension, satisfying $\int_0^T (|\mu_u^X| + |\mu_u^Y| + |\sigma_u^X|^2 + |\sigma_u^Y|^2) du < \infty$, a.s.

Provide the dynamics of the process $Z := f(X, Y)$ for the following functions f:

1. *$f(x, y) = xy$, $x, y \in \mathbb{R}$,*
2. *$f(x, y) = \frac{x}{y}$, $x, y \in \mathbb{R}$, (assuming that Y_0, μ^Y and σ^Y are such that Y is a positive process).*

Exercise 6.10. *Let a and σ be two measurable and $\mathbb{F}-$adapted processes such that $\int_0^t |a_s| ds + \int_0^t |\sigma_s|^2 ds < \infty$.*

1. *Prove the integration by parts formula:*

$$\int_0^t \int_0^s \sigma_u a_s dW_u ds = \left(\int_0^t \sigma_u dW_u \right) \left(\int_0^t a_s ds \right) - \int_0^t \int_0^u a_s \sigma_u ds dW_u.$$

2. *We now take the processes $a_s = a(s)$ and $b_s = b(s)$ to be deterministic functions. Show that the random variable $\int_0^t \int_0^s \sigma_u a_s dW_u ds$ has a Gaussian distribution, and compute the corresponding mean and variance.*

We conclude this section by providing an example of local martingale which fails to be a martingale, although positive and uniformly integrable.

Example 6.11. *(A strict local martingale) Let W be a Brownian motion in $\mathbb{R}^d$.*

• *In the one-dimensional case $d = 1$, we have $\mathbb{P}[|W_t| > 0, \ \forall \ t > 0] = 0$, see also Proposition 4.19.*

• *When $d \geq 2$, the situation is drastically different as it can be proved that $\mathbb{P}[|W_t| > 0$, for every $t > 0] = 1$, see e.g. Karatzas and Shreve [30] Proposition 3.22 p161. In words, this means that the Brownian motion never returns to the origin $\mathbb{P}-$a.s. This is a well-known result for random walks on $\mathbb{Z}^d$. Then, for a fixed $t_0 > 0$, the process*

$$X_t := |W_{t_0+t}|^{-1}, \qquad t \geq 0,$$

is well-defined, and it follows from Itô's formula that

$$dX_t = X_t^3 \left(\frac{1}{2}(3 - d) dt - W_t \cdot dW_t \right).$$

• *We now consider the special case $d = 3$. By the previous Proposition 5.10, it follows from Itô's formula that X is a local martingale. However, by the scaling property of the Brownian motion, we have*

$$\mathbb{E}[X_t] = \sqrt{\frac{t_0}{t+t_0}}\,\mathbb{E}[X_0]\,, \qquad t \geq 0\,,$$

so that X has a non-constant expectation and can not be a martingale.
• *Passing to the polar coordinates, we calculate directly that*

$$\begin{aligned}
\mathbb{E}\left[X_t^2\right] &= (2\pi(t_0+t))^{-3/2}\int |x|^{-2}e^{-|x|^2/2(t_0+t)}dx \\
&= (2\pi(t_0+t))^{-3/2}\int r^{-2}e^{-r^2/2(t_0+t)}4\pi r^2 dr \\
&= \frac{1}{t_0+t} \leq \frac{1}{t_0} \text{ for every } t \geq 0.
\end{aligned}$$

This shows that $\sup_{t\geq 0}\mathbb{E}\left[X_t^2\right] < \infty$. In particular, X is uniformly integrable.

6.3 Lévy's characterization of Brownian motion

The next result is valid in a larger generality than the present framework.

Theorem 6.12. *Let W be a Brownian motion in $\mathbb{R}^d$, and ϕ an $\mathcal{M}_{\mathbb{R}}(n,d)-$valued process with components in $\mathbb{H}^2$, and such that $\int_0^t \phi_s\phi_s^T\,ds = t\,\mathrm{I}_n$ for all $t \geq 0$. Then, the process X defined by:*

$$X_t^j \;:=\; X_0^j + \sum_{k=1}^d \int_0^t \phi_t^{jk}dW_t^k, \qquad j = 1,\ldots,n$$

is a Brownian motion on $\mathbb{R}^n$.

Proof. Clearly $X_0 = 0$ and X has continuous sample paths, a.s. and is $\mathbb{F}-$adapted. To complete the proof, we show that $X_t - X_s$ is independent of $\mathcal{F}_s$, and is distributed as a $\mathbf{N}\left(0,(t-s)\mathrm{I}_n\right)$. By using the characteristic function, this is equivalent to show

$$\mathbb{E}\left[e^{iu\cdot(X_t-X_s)}\middle|\mathcal{F}_s\right] = e^{-|u|^2(t-s)/2} \quad \text{for all} \quad u \in \mathbb{R}^n, 0 \leq s \leq t. \qquad (6.5)$$

For fixed s, we apply Itô's formula to the function $f(x) := e^{iu\cdot x}$:

$$e^{iu\cdot(X_t-X_s)} \;=\; 1 + i\sum_{j=1}^d u_j\int_s^t e^{iu\cdot(X_r-X_s)}dX_r^j - \frac{1}{2}|u|^2\int_s^t e^{iu\cdot(X_r-X_s)}dr$$

Since $|f| \le 1$ and $\phi_s \phi_s^T = I_n$, $dt \times d\mathbb{P}-$a.s., we have $\mathbb{E}\left[\int_s^t e^{iu\cdot(X_r-X_s)}dX_r^j \,\middle|\, \mathcal{F}_s\right] = 0$. Then, the function $h(t) := \mathbb{E}\left[e^{iu\cdot(X_t-X_s)} \,\middle|\, \mathcal{F}_s\right]$ satisfies the ordinary differential equation:

$$h(t) \quad = \quad 1 - \frac{1}{2}|u|^2 \int_s^t h(r)dr,$$

and therefore $h(t) = e^{-|u|^2(t-s)/2}$, which is the required result (6.5). ♢

6.4 A verification approach to the Black-Scholes model

Our first contact with the Black-Scholes model was by means of the continuous-time limit of the binomial model. We shall have a cleaner presentation in the larger class of models later on when we will have access to the change of measure tool. In this paragraph, we provide a continuous-time presentation of the Black-Scholes model which only appeals to Itô's formula.

Consider a financial market consisting of a nonrisky asset with constant interest rate $r \ge 0$, and a risky asset defined by

$$S_t \quad = \quad S_0 \exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\right), \quad t \ge 0. \tag{6.6}$$

An immediate application of Itô's formula shows that the dynamics of this process are given by:

$$\frac{dS_t}{S_t} \quad = \quad \mu dt + \sigma dW_t, \quad t \ge 0.$$

A portfolio strategy is a measurable and $\mathbb{F}-$adapted process $\{\theta_t, t \in [0,T]\}$ with $\int_0^T |\theta_t|^2 dt < \infty$, a.s. where θ_t represents the amount invested in the risky asset at time t, corresponding to a number of shares θ_t/S_t. Denoting by X_t the value of the portfolio at time t, we see that the investment in the nonrisky asset is given by $X_t - \theta_t$, and the variation of the portfolio value under the self-financing condition is given by

$$dX_t^\theta \quad = \quad \theta_t \frac{dS_t}{S_t} + (X_t - \theta_t)rdt, \ t \ge 0.$$

We say that the portfolio θ is admissible if in addition the corresponding portfolio value X^θ is bounded from below. The admissibility condition means that the investor is limited by a credit line below which he is considered bankrupt. We denote by $\mathcal{A}$ the collection of all admissible portfolios.

Using again Itô's formula, we see that the discounted portfolio value process $\tilde{X}_t := X_t e^{-rt}$ satisfies the dynamics:

$$d\tilde{X}_t^\theta = e^{-rt}\theta_t \frac{d\tilde{S}_t}{\tilde{S}_t} \quad \text{where} \quad \tilde{S}_t := S_t e^{-rt}, \tag{6.7}$$

and then

$$d\tilde{S}_t = e^{-rt}(-rS_t dt + dS_t) = \tilde{S}_t\big((\mu - r)dt + \sigma dW_t\big). \tag{6.8}$$

We recall the no-arbitrage principle which says that if a portfolio strategy on the financial market produces an a.s. nonegative final portfolio value, starting from a zero intial capital, then the portfolio value is zero a.s.

A contingent claim is an $\mathcal{F}_T-$measurable random variable which describes the random payoff of the contract at time T. The following result is specific to the case where the contingent claim is $g(S_T)$ for some deterministic function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$. Such contingent claims are called *Vanilla options*.

In preparation of the main result of this section, we start by

Proposition 6.13. *Suppose that the function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ has polynomial growth, i.e. $|g(s)| \le \alpha(1 + s^\beta)$ for some $\alpha, \beta \ge 0$. Then, the linear partial differential equation on $[0, T) \times (0, \infty)$:*

$$\mathcal{L}v := \frac{\partial v}{\partial t} + rs\frac{\partial v}{\partial s} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 v}{\partial s^2} - rv = 0 \quad \text{and} \quad v(T, .) = g, \tag{6.9}$$

has a unique solution $v \in C^0([0, T] \times (0, \infty)) \cap C^{1,2}([0, T), (0, \infty))$ in the class of polynomially growing functions, and given by

$$v(t, s) = \mathbb{E}\left[e^{-r(T-t)} g\left(\hat{S}_T^{t,s}\right)\right],\ (t, s) \in [0, T] \times (0, \infty)$$

where

$$\hat{S}_T^{t,s} := se^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(W_T - W_t)}.$$

Proof. We denote $V(t, s) := \mathbb{E}\left[e^{-r(T-t)} g\left(\hat{S}_T^{t,s}\right)\right]$.
1- We first observe that $V \in C^0([0, T] \times (0, \infty)) \cap C^{1,2}([0, T), (0, \infty))$. To see this, we simply write

$$V(t, s) = e^{-r(T-t)} \int_{\mathbb{R}} g(e^x) \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} e^{-\frac{1}{2}\left(\frac{x - ln(s)(r-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right)^2} dx,$$

and see that the claimed regularity holds true by the dominated convergence theorem.
2- Immediate calculation reveals that V inherits the polynomial growth of g. Let

$$\hat{S}_u := e^{(r-\mu)t} S_u,\ u \ge t, \quad \text{so that} \quad \frac{d\hat{S}_u}{\hat{S}_u} = rdu + \sigma dW_u,$$

an consider the stopping time $\tau := \inf\left\{u > t : |\ln(\hat{S}_u/s)| > 1\right\}$. Then, it follows from the law of iterated expectations that

$$V(t, s) = \mathbb{E}_{t,s}\left[e^{-r(\tau\wedge h-t)} V\left(\tau \wedge h, \hat{S}_{\tau\wedge h}\right)\right]$$

for every $h > t$, where we denoted by $\mathbb{E}_{t,s}$ the expectation conditional on $\{\hat{S}_t = s\}$. It then follows from Itô's formula that

$$\begin{aligned} 0 &= \mathbb{E}_{t,s}\left[\int_t^{\tau\wedge h} e^{-ru}\mathcal{L}V(u,\hat{S}_u)du + \int_t^{\tau\wedge h} e^{-ru}\frac{\partial V}{\partial s}(u,\hat{S}_u)\sigma\hat{S}_u dW_u\right] \\ &= \mathbb{E}_{t,s}\left[\int_t^{\tau\wedge h} e^{-ru}\mathcal{L}V(u,\hat{S}_u)du\right]. \end{aligned}$$

Normalizing by $(h-t)$ and sending $h \searrow t$, il follows from the dominated convergence theorem that V is a solution of (6.9).

3- We next prove the uniqueness among functions of polynomial growth. Let $\tau_n := \inf\left\{u > t : |\ln(\hat{S}_u/s)| > n\right\}$, and consider an arbitrary solution v of (6.9) with $|v(t,s)| \le \alpha(1+s^\beta)$ for some $\alpha, \beta \ge 0$. Then, it follows from Itô's formula together with the fact that v solves (6.9) that:

$$e^{-r(T\wedge\tau_n)}v\left(T\wedge\tau_n, \hat{S}_{T\wedge\tau_n}\right) = e^{-rt}v(t,s) + \int_t^{T\wedge\tau_n} e^{-rt}\frac{\partial V}{\partial s}(u,\hat{S}_u)\sigma\hat{S}_u dW_u.$$

Since the integrand in the latter stochastic integral is bounded on $[t, T\wedge\tau_n]$, this provides

$$e^{-rt}v(t,s) = \mathbb{E}_{t,s}\left[e^{-r(T\wedge\tau_n)}v\left(T\wedge\tau_n, \hat{S}_{T\wedge\tau_n}\right)\right].$$

By the continuity of v, we see that $e^{-r(T\wedge\tau_n)}v\left(T\wedge\tau_n, S_{T\wedge\tau_n}\right) \longrightarrow e^{-rT}v(T,S_T) = e^{-rT}g(S_T)$, a.s. We next observe from the polynomial growth of v that

$$\left|e^{-r(T\wedge\tau_n)}v\left(T\wedge\tau_n, \hat{S}_{T\wedge\tau_n}\right)\right| \le \alpha\left(1+\beta e^{rT+\sigma\max_{u\le T}W_u}\right) \in \mathbb{L}^1.$$

We then deduce from the dominated convergence theorem that $v(t,s) = V(t,s)$.

$\diamondsuit$

We now have the tools to obtain the Black-Scholes formula in the continuous-time framework of this section.

Proposition 6.14. *Let $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ be a polynomially growing function bounded from below, i.e. $-c \le g(s) \le \alpha(1+s^\beta)$ for some $\alpha, \beta \ge 0$. Assume that the contingent claim $g(S_T)$ is available for trading, and that the financial market satisfies the no-arbitrage condition. Then:*
(i) the market price at time 0 of the contingent claim $g(S_T)$ is given by $V(0,S_0) = \mathbb{E}\left[e^{-rT}g(\hat{S}_T^{0,S_0})\right]$,
(ii) there exists a replicating portfolio $\theta^* \in \mathcal{A}$ for $g(S_T)$, i.e. $X_0^{\theta^*} = V(0,S_0)$ and $X_T^{\theta^*} = g(S_T)$, a.s. given by

$$\theta_t^* = \frac{\partial V}{\partial s}(t,S_t), \quad \textit{where} \quad V(t,s) := \mathbb{E}\left[e^{-r(T-t)}g(\hat{S}_T^{t,s})\right].$$

Proof. By Itô's formula, we compute that:

$$e^{-rT}V(T,S_T) = V(0,S_0) + \int_0^T e^{-rt}\left(-rV + \frac{\partial V}{\partial t} + \mu s\frac{\partial V}{\partial s} + \frac{1}{2}\sigma^2 s^2\frac{\partial^2 V}{\partial s^2}\right)(t,S_t)dt$$
$$+\int_0^T e^{-rt}\frac{\partial V}{\partial s}(t,S_t)\sigma S_t dW_t.$$

Since V solves the PDE (6.9), this provides:

$$\begin{aligned} e^{-rT}g(S_T) &= V(0,S_0) + \int_0^T e^{-rt}\frac{\partial V}{\partial s}(t,S_t)\left(-rS_t dt + dS_t\right) \\ &= V(0,S_0) + \int_0^T \tilde{S}_t\frac{\partial V}{\partial s}(t,S_t)\frac{d\tilde{S}_t}{\tilde{S}_t}, \end{aligned}$$

which, in view of (6.8), can be written in:

$$e^{-rT}g(S_T) = V(0,S_0) + \int_0^T e^{-rt}\theta_t^*\frac{d\tilde{S}_t}{\tilde{S}_t} \quad \text{where} \quad \theta_t^* := S_t\frac{\partial v}{\partial s}(t,S_t).$$

By (6.7), we see that $e^{-rT}g(S_T) = \tilde{X}_T^{\theta^*}$ with $\tilde{X}_0^{\theta^*} = V(0,S_0)$, and therefore

$$X_T^{\theta^*} = g(S_T), \quad \text{a.s.}$$

Notice that $\tilde{X}_t^{\theta^*} = \mathbb{E}[e^{-rT}g(S_T)|\mathcal{F}_t]$, $t \in [0,T]$. Then, the process X inherits the lower bound of g, implying that $\theta^* \in \mathcal{A}$.

We finally conclude the proof by using the no-arbitrage property of the market consisting of the nonrisky asset, the risky one, and the contingent claim, arguying exactly as in Section 2.1 (iii). ♢

Exercise 6.15. *Let $g(s) := (s-K)^+$ for some $K > 0$. Show that the no-arbitrage price of the last proposition coincides with the Black-Scholes formula* (2.8).

6.5 The Ornstein-Uhlenbeck process

We consider the Itô process defined on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ by

$$X_t := b + (X_0 - b)e^{-at} + \sigma\int_0^t e^{-a(t-s)}dW_s, \qquad t \geq 0$$

where X_0 is an $\mathcal{F}_0-$measurable square integrable r.v.

6.5.1 Distribution

Conditional on X_0, X is a continuous gaussian process whose covariance function can be explicitly computed:

$$\mathbb{C}ov\left(X_s, X_t|X_0\right) = e^{-a(t-s)}\mathbb{V}ar\left[X_s|X_0\right] \quad \text{for} \quad 0 \leq s \leq t.$$

The conditional mean and variance ar given by:

$$\mathbb{E}[X_t|X_0] = b + (X_0 - b)\,e^{-at}, \qquad \mathbb{V}ar[X_t|X_0] = \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right).$$

We then deduce the unconditional mean:

$$\mathbb{E}[X_t] = b + (\mathbb{E}[X_0] - b)\,e^{-at}$$

and the variance

$$\begin{aligned}\mathbb{V}ar[X_t] &= \mathbb{E}\{\mathbb{V}ar[X_t|X_0]\} + \mathbb{V}ar\{\mathbb{E}[X_t|X_0]\} \\ &= e^{-2at}\mathbb{V}ar[X_0] + \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)\end{aligned}$$

Since $X|X_0$ is Gaussian, we deduce that

- if X_0 is distributed as $\mathcal{N}\left(b, \frac{\sigma^2}{2a}\right)$, then $X_t =_d X_0$ for any $t \geq 0$, where $=_d$ denotes equality in distribution,
- if X_0 is an $\mathcal{F}_0-$Gaussian r.v. then X_0 is independent of W, and

$$X_t \longrightarrow \mathcal{N}\left(b, \frac{\sigma^2}{2a}\right) \quad \text{in law as} \quad t \to \infty.$$

We say that $\mathcal{N}\left(b, \frac{\sigma^2}{2a}\right)$ is the invariant (or stationary) distribution of the process Y.

Finally, when X_t models the instantaneous interest rate, the discount factor for a payoff at time T is defined by $\exp\left(-\int_0^T X_t dt\right)$ plays an important role in the theory of interest rates. The distribution of $\int_0^T X_t dt$ can be characterized as follows:

$$\int_0^T X_t dt = bT + (X_0 - b)A(T) + \sigma \int_0^T M_t dA(t),$$

where $M_t := \displaystyle\int_0^t e^{as} dW_s dt$ and $A(t) := \displaystyle\int_0^t e^{-as} ds = \frac{1 - e^{-at}}{a}$. Recall that the integration by parts formula is valid, as a consequence of Itô's formula. Then,

$$\begin{aligned}\int_0^T X_t dt &= bT + (X_0 - b)A(T) + \sigma \int_0^T \left(\int_t^T e^{-as} ds\right) e^{at} dW_t \\ &= bT + (X_0 - b)A(T) + \sigma \int_0^T A(T - t) dW_t.\end{aligned}$$

Hence, conditional on X_0, the r.v. $\displaystyle\int_0^T X_t dt$ is distributed as Gaussian with mean $bT + (X_0 - b)A(T)$ and variance $\sigma^2 \int_0^T A(t)^2 dt$. We can even conclude that, conditional on X_0, the joint distribution of the pair $Z_T := \left(X_T, \int_0^T X_t dt\right)$ is Gaussian with mean

$$\mathbb{E}[Z_T] \quad = \quad \left(\; b + (X_0 - b)\, e^{-aT} bT + (X_0 - b) A(T) \; \right)$$

and variance

$$\mathbb{V}[Z_T] \quad = \quad \sigma^2 \begin{pmatrix} \int_0^T e^{-2at} dt & \int_0^T e^{-at} A(t) dt \\ \int_0^T e^{-at} A(t) dt & \int_0^T A(t)^2 dt \end{pmatrix}.$$

6.5.2 Differential representation

Let $Y_t := e^{at} X_t$, $t \geq 0$. Then, by direct calculation, we get:

$$Y_t \quad = \quad X_0 + b\left(e^{at} - 1\right) + \sigma \int_0^t e^{as} dW_s,$$

or, in differential form,

$$dY_t \quad = \quad abe^{at} dt + \sigma e^{at} dW_t.$$

By direct application of Itô's formula, we then obtain the process $X_t = e^{-at} Y_t$ in differential form:

$$dX_t \quad = \quad a\,(b - X_t) + \sigma dW_t.$$

This is an example of stochastic differential equation, see Chapter 8 for a systematic treatment. The last differential form shows that, whenever $a > 0$, the dynamics of the process X exhibit a mean reversion effect in the sense that

- if $X_t > b$, then the drift is pushing the process down towards b,
- similarly, if $X_t < b$, then the drift is pushing the process up towards b.

The mean-reversion of this gaussian process is responsible for its stationarity property, and its popularity in many application. In particular, in finance this process is commonly used for the modelling of interest rates.

Exercise 6.16. *Let B be a Brownian motion, and consider the processes*

$$X_t := e^{-t} B_{e^{2t}}, \quad Y_t := X_t - X_0 + \int_0^t X_s ds, \quad t \geq 0.$$

be an Ornstein-Uhlenbeck process defined by the dynamics $dY_t = -Y_t dt + \sqrt{2} dW_t$.

1. *Prove that $Y_t = \int_1^{e^{2t}} u^{-1/2} dB_u$, $t \geq 0$.*
2. *Deduce that $\{X_t, t \geq 0\}$ is an Ornstein-Uhlenbeck process.*

6.6 The Merton optimal portfolio allocation

6.6.1 Problem formulation

Consider a financial market consisting of a nonrisky asset, with constant interest rate r, and a risky asset with price process defined by (6.6), so that

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t.$$

A portfolio strategy is an $\mathbb{F}-$adapted process π such that $\int_0^T |\pi_t|^2 dt < \infty$ $\mathbb{P}-$a.s. representing the proportion of wealth invested in the risky asset at time t. Let $\mathcal{A}$ denote the set of all portfolio strategies.

Under the self-financing condition, the portfolio value at time t is defined by:

$$X_t = e^{rt}\left(x + \int_0^t \pi_u X_u e^{-ru} \frac{d\left(S_u e^{-ru}\right)}{S_u e^{-ru}}\right).$$

Then,

$$dX_t = rX_t dt + \pi_t X_t \left((\mu - r)dt + \sigma dW_t\right),$$

and it follows from a direct application of Itô's formula to the function $\ln X_t$ that

$$d\ln X_t = \left(r + (\mu - r)\pi_t - \frac{1}{2}\sigma^2\pi_t^2\right) dt + \sigma\pi_t dW_t.$$

This provides the expression of X_t in terms of the portfolio strategy π and the initial capital X_0:

$$X_t = X_0 \exp\left(\int_0^t \left(r + (\mu - r)\pi_u - \frac{1}{2}\sigma^2\pi_u^2\right) du + \int_0^t \sigma\pi_u dW_u\right). \tag{6.10}$$

The continuous-time optimal portfolio allocation problem, as formulated by Merton (1969), is defined by:

$$V_0(x) := \sup_{\pi\in\mathcal{A}} \mathbb{E}\left[U\left(X_T^{x,\pi}\right)\right], \tag{6.11}$$

where U is an increasing strictly concave function representing the investor utility, i.e. describing his preferences and attitude towards risk. We assume that U is bounded from below, which guarantees that he expectation in (6.11) is well-defined.

6.6.2 The dynamic programming equation

The dynamic programming technique is a powerful approach to stochatsic control problems of the type (6.11). This method was developed by Richard Bellman in the fifties, while at the same time the russian school was exploring the stochastic extension of the Pontryagin maximum principle. Our objective is to provide an intuitive introduction to the dynamic programming approach in order to motivate our solution approach used in the subsequent section.

The main idea is to define a dynamic version $V(t,x)$ of the problem (6.11) by moving the time origin from 0 to t. For simplicity, let us restrict the set of portfolio strategies to those so-called Markov ones, i.e. $\pi_t = \pi(t, X_t)$ Then, it follows from the law of iterated expectations that

$$\begin{aligned} V(t,x) &= \sup_{\pi\in\mathcal{A}} \mathbb{E}_{t,x}\left[U\left(X_T^\pi\right)\right] \\ &= \sup_{\pi\in\mathcal{A}} \mathbb{E}_{t,x}\left[\mathbb{E}_{t+h,X^\pi_{t+h}}\left\{U\left(X_T^\pi\right)\right\}\right], \end{aligned}$$

where we denoted by $\mathbb{E}_{t,x}$ the expectation operator conditional on $X_t^\pi = x$. Then, we formally expect that the following dynamic programming principle

$$V(t,x) = \sup_{\pi\in\mathcal{A}} \mathbb{E}_{t,x}\left[V\left(t+h, X^\pi_{t+h}\right)\right]$$

holds true. A rigorous proof of the claim is far from obvious, and is not needed in these notes, as this paragraph is only aiming at developing a good intuition for the subsequent solution approach of the Merton problem.

We next assume that V is known to be sufficiently smooth so as to allow for the use of Itô's formula. Then:

$$\begin{aligned} 0 &= \sup_{\pi\in\mathcal{A}} \mathbb{E}_{t,x}\left[V\left(t+h, X^\pi_{t+h}\right) - V(t,x)\right] \\ &= \sup_{\pi\in\mathcal{A}} \mathbb{E}_{t,x}\left[\int_t^{t+h} \mathcal{L}^{\pi_t} V\left(u, X_u^\pi\right) du + \int_t^{t+h} V_x\left(u, X_u^\pi\right)\sigma\pi_u X_u dW_u\right], \end{aligned}$$

where, for a function $v \in C^{1,2}([0,T),\mathbb{R})$, we denote

$$\mathcal{L}^\alpha v := \frac{\partial v}{\partial t} + x\left(r + \alpha(\mu - r)\right)\frac{\partial v}{\partial x} + \frac{1}{2}x^2\alpha^2\sigma^2\frac{\partial^2 v}{\partial x^2}. \tag{6.12}$$

We continue our intuitive presentation by forgetting about any difficulties related to some strict local martingale feature of the stochastic integral inside the expectation. Then

$$0 = \sup_{\pi\in\mathcal{A}} \mathbb{E}_{t,x}\left[\frac{1}{h}\int_t^{t+h} \mathcal{L}^{\pi_t} V\left(u, X_u^\pi\right) du\right],$$

and by sending $h \to 0$, we expect from the mean value theorem that V solves the nonlinear partial differential equation

$$\sup_{\alpha\in\mathbb{R}} \mathcal{L}^\alpha V(t,x) = 0 \text{ on } [0,T)\times\mathbb{R} \quad \text{and} \quad V(T,.) = U.$$

The latter partial differential equation is the so-called dynamic programming equation, also referred to as the Hamilton-Jacobi-Bellman equation.

Our solution approach for the Merton problem will be the following. suppose that one is able to derive a solution v of the dynamic programming equation, then use a *verification argument* to prove that the candidate v is indeed coinciding with the value function V of the optimal portfolio allocation problem.

6.6.3 Solving the Merton problem

We recall that for a $C^{1,2}\left([0,T],\mathbb{R}\right)$ function v, it follows from Itô's formula that

$$v(t,X_t^\pi) \quad = \quad v(0,X_0)+\int_0^T \mathcal{L}^{\pi_t}v(s,X_s)ds+\int_0^t v_x(s,X_s)\pi_t X_t\sigma dW_t,$$

where the operator $\mathcal{L}^\alpha$ is defined in (6.12).

Proposition 6.17. *Let $v \in C^0\left([0,T]\times\mathbb{R}\right)\cap C^{1,2}\left([0,T),\mathbb{R}\right)$ be a nonnegative function satisfying*

$$v(T,.)\geq U \quad \textit{and} \quad -\mathcal{L}^\alpha v(t,x)\geq 0 \textit{ for all } (t,x)\in[0,T)\times\mathbb{R},\ \alpha\in\mathbb{R}.$$

Then $v(0,x)\geq V_0(x)$.

Proof. For every portfolio strategy $\pi \in \mathcal{A}$ and $t \leq T$, it follows from Itô's formula that:

$$\begin{aligned} M_t \quad &:= \quad \int_0^t \frac{\partial v}{\partial x}\left(u,X_u^{x,\pi}\right)X_u^{x,\pi}\pi_u\sigma dW_u \\ &= \quad v\left(t,X_t^{x,\pi}\right)-v(0,x)-\int_0^t \mathcal{L}^{\pi_u}v\left(u,X_u^{x,\pi}\right)du \\ &\geq \quad v\left(t,X_t^{x,\pi}\right)-v(0,x). \end{aligned}$$

Since $v \geq 0$, the process M is a supermartingale, as a local martingale bounded from below by a constant. Then:

$$0 \;\; \geq \;\; \mathbb{E}\left[M_T\right] \; \geq \; \mathbb{E}\left[v\left(T,X_T^{x,\pi}\right)\right]-v(0,x) \; = \; \mathbb{E}\left[U\left(X_T^{x,\pi}\right)\right]-v(0,x),$$

and the required inequality follows from the arbitrariness of $\pi \in \mathcal{A}$. ♢

We continue the discussion of the optimal portfolio allocation problem in the context of the power utility function:

$$U(x)=x^p, \qquad 0<p<1. \tag{6.13}$$

This induces an important simplification as we immediately verify from (6.10) that $V_0(x)=x^pV_0(1)$. We then search for a solution of the partial differential equation

$$\sup_{\alpha\in\mathbb{R}} \mathcal{L}^\alpha v \;=\; 0 \quad \text{and} \quad v(T,x)=x^p,$$

of the form $v(t,x) = x^p h(t)$ for some function h. Plugging this form in the above nonlinear partial differential equation leads to an ordinary differential equation for the function h, and provides the candidate solution:

$$v(t,x) := x^p e^{p(T-t)\left(r+\frac{(\mu-r)^2}{2(1-p)\sigma^2}\right)}, \qquad t \in [0,T], x \geq 0.$$

By candidate solution, we mean that v satisfies:

$$\sup_{\alpha\in\mathbb{R}} \mathcal{L}^\alpha v \;=\; \mathcal{L}^{\hat{\alpha}} v \;=\; 0 \quad \text{and} \quad v(T,x) = x^p\,, \text{ where } \hat{\alpha} := \frac{\mu - r}{(1-p)\sigma^2}.$$

Proposition 6.18. *In the context of the power utility function* (6.13)*, the value function of the optimal portfolio allocation problem is given by:*

$$V_0(x) \;=\; v(0,x) \;\;=\;\; x^p e^{pT\left(r+\frac{(\mu-r)^2}{2(1-p)\sigma^2}\right)},$$

and the constant portfolio strategy $\hat{\pi}_u := \hat{\alpha}$ is an optimal portfolio allocation.

Proof. Let $\hat{X} := X^{x,\hat{\pi}}$, $\tau_n := T \wedge \inf\left\{t > 0 \;:\; \hat{X}_t \geq n\right\}$, and

$$\hat{M}_t \;\;:=\;\; \int_0^t \frac{\partial v}{\partial x}\left(u, \hat{X}_u\right) \hat{X}_u \hat{\pi}_u \sigma dW_u,\; t \in [0,T).$$

Since the integrand in the expression of $\hat{M}$ is bounded on $[0,\tau_n]$, we see that the stopped process $\{M_{t\wedge\tau_n}, t \in [0,T]\}$ is a martingale. We then deduce from Itô's formula together with the fact that $\mathcal{L}^{\hat{\alpha}} v = 0$ that:

$$0 = \mathbb{E}\left[\hat{M}_{\tau_n}\right] = \mathbb{E}\left[v\left(\tau_n, \hat{X}_{\tau_n}\right)\right] - v(0,x). \tag{6.14}$$

We next observe that, for some constant $C > 0$,

$$0 \leq v\left(\tau_n, \hat{X}_{\tau_n}\right) \leq Ce^{C\max_{t\leq T} W_t} \in \mathbb{L}^1,$$

recall that $\max_{t\leq T} W_t =_d |W_t|$. We can then use the dominated convergence theorem to pass to the limit $n \to \infty$ in (6.14):

$$\lim_{n\to\infty} \mathbb{E}\left[v\left(\tau_n, \hat{X}_{\tau_n}\right)\right] \;\;=\;\; \mathbb{E}\left[v\left(T, \hat{X}_T\right)\right] \;=\; \mathbb{E}\left[U\left(\hat{X}_T\right)\right],$$

where the last equality is due to $v(T,.) = U$. Hence, $V_0(x) = v(0,x)$ and $\hat{\pi}$ is an optimal portfolio strategy. ♢

Chapter 7

Martingale representation and change of measure

In this chapter, we develop two essential tools in financial mathematics. The martingale representation is the mathematical counterpart of the hedging portfolio. Change of measure is a crucial tool for the representation and the calculation of valuation formulae in the everyday life of the financial industry oriented towards derivative securities. The intuition behind these two tools can be easily understood in the context of the one-period binomial model of Section 2.1 of Chapter 2:

- Perfect hedging of derivative securities in the simple one-period model is always possible, and reduces to a linear system of two equations with two unknowns. It turns out that this property is valid in the framework of the Brownian filtration, and this is exactly what the martingale representation is about. But, we should be aware that this result is specific to the Brownian filtration, and fails in more general models... this topic is outside the scope of the present lectures notes.

- The hedging cost in the simple one-period model, which is equal to the no-arbitrage price of the derivative secutity, can be expressed as an expected value of the discounted payoff under the risk-neutral measure. This representation is very convenient for the calculations, and builds a strong intuition for the next developments of the theory. The Girsanov theorem provides the rigorous way to express expectation under alternative measures than the initially given one.

7.1 Martingale representation

Let $\mathbb{F}$ be the canonical filtration of the Brownian motion completed with the null sets, and consider a random variable $F \in \mathbb{L}^1(\mathcal{F}_T, \mathbb{P})$. The goal of this section is to show that any such random variable or, in other words, any path-dependent functional of the Brownian motion, can be represented as a stochastic integral of some process with respect to the Brownian motion (and hence, a martingale). This has a natural application in finance where one is interested in replicating contingent claims with hedging portfolios. We start with square integrable random variables.

Theorem 7.1. *For any $F \in \mathbb{L}^2(\mathcal{F}_T, \mathbb{P})$ there exists a unique adapted process $H \in \mathbb{H}^2$ such that*

$$F = E[F] + \int_0^T H_s dW_s, \quad \mathbb{P} - a.s. \tag{7.1}$$

Proof. The uniqueness is an immediate consequence of the Itô isometry. We prove the existence in the two following steps.
Step 1: We start by proving the claim in the special case $F = f(W_T)$, for some bounded measurable function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$. By standard molification, we may find a sequence f_n of bounded C^2 functions such that $f_n(W_T) \longrightarrow f(W_T)$ in $\mathbb{L}^2$. Then, the map $(t,x) \longmapsto u(t,x) := \mathbb{E}[f_n(W_T)|W_t = x]$ is $C^\infty([0,T] \times \mathbb{R}^d)$, and it follows from Itô's formula that

$$f_n(W_T) = u(T, W_T) = \mathbb{E}[f_n(W_T)] + \int_0^T H_t^n \cdot dW_t, \text{ with } H_t^n := \frac{\partial u}{\partial x}(t, W_t), \tag{7.2}$$

see Remark 4.8. Then, it follows for $n, m \geq 1$ that

$$f_n(W_T) - f_m(W_T) \quad = \quad \mathbb{E}\big[f_n(W_T) - f_m(W_T)\big] + \int_0^T (H_t^n - H_t^m) \cdot dW_t.$$

By the Itô isometry, this implies that $\|H_n - H_m\|_{\mathbb{H}^2} \leq \mathbb{E}\big|f_n(W_T) - f_m(W_T)\big| + \big\|(f_n - f_m)(W_T)\big\|_{\mathbb{L}^2}$. As $f_n(W_T) \longrightarrow f(W_T)$ in $\mathbb{L}^2$, this shows that $(H_n)_n$ is a Cauchy sequence in $\mathbb{H}^2$, and therefore $H_n \longrightarrow H$ in $\mathbb{H}^2$. We may then pass to limits in (7.2), and get $f(W_T) = \mathbb{E}[f(W_T)] + \int_0^T H_t \cdot dW_t$, as required.
Step 2: We next fix an integer $n \geq 1$, $0 \leq t_1 < \ldots < t_n$, and prove that for every bounded function $f : (\mathbb{R}^d)^n \longrightarrow \mathbb{R}$, the representation (7.1) holds with $\xi := f(W_{t_1}, \ldots, W_{t_n})$ for some $H \in \mathbb{H}^2$. To see this, denote $\underline{x}_i := (x_1, \ldots, x_i)$ and set

$$u_n(t, \underline{x}_n) := \mathbb{E}\left[f(\underline{x}_{n-1}, W_{t_n})|W_t = x_n\right], \quad t_{n-1} \leq t \leq t_n.$$

Then, for all $(\underline{x}_{n-1})$ fixed, we may apply the result of Step 1 to the mapping $(t, x_n) \longmapsto u_n(t, \underline{x}_{n-1}, x_n)$, and get

$$f(\underline{x}_{n-1}, W_{t_n}) = f_{n-1}(\underline{x}_{n-1}) + \int_{t_{n-1}}^{t_n} H_s \cdot dW_s, \tag{7.3}$$

where $f_{n-1}(\underline{x}_{n-1}) := u_n(t_{n-1}, \underline{x}_{n-1}, x_{n-1})$. Similarly, we fix $(\underline{x}_{n-2})$, and apply the result of Step 1 to the function of $(t, x_{n-1}) \in [t_{n-2}, t_{n-1}) \times \mathbb{R}^d$:

$$u_{n-1}(t, \underline{x}_{n-2}, x_{n-1}) \quad := \quad \mathbb{E}\left[f_{n-1}(\underline{x}_{n-2}, W_{t_{n-1}})|W_t = x_{n-1}\right],$$

so that

$$f_{n-1}(\underline{x}_{n-2}, W_{t_{n-1}}) = f_{n-2}(\underline{x}_{n-2}) + \int_{t_{n-2}}^{t_{n-1}} H_s \cdot dW_s, \tag{7.4}$$

where $f_{n-2}(\underline{x}_{n-2}) := u_{n-1}(t_{n-2}, \underline{x}_{n-2}, x_{n-2})$. Combining (7.3) and (7.4), we obtain:

$$f(\underline{x}_{n-2}, W_{t_{n-1}}, W_{t_n}) \quad = \quad f_{n-2}(\underline{x}_{n-2}) + \int_{t_{n-2}}^{t_n} H_s \cdot dW_s.$$

Repeating this argument, it follows that

$$f(W_{t_1}, \ldots, W_{t_n}) \quad = \quad f_0(0) + \int_0^{t_n} H_s \cdot dW_s.$$

Since f is bounded, it follows that $H \in \mathbb{H}^2$.
Step 3: The subset $\mathcal{R}$ of $\mathbb{L}^2$ for which (7.1) holds is a closed linear subspace of the Hilbert space $\mathbb{L}^2$. To complete the proof, we now prove that its $\mathbb{L}^2-$orthogonal $\mathcal{R}^\perp$ is reduced to $\{0\}$.

Let $\mathcal{I}$ denote the set of all events $E = \{(W_{t_1}, \ldots, W_{t_n}) \in A\}$ for some $n \geq 1$, $0 \leq t_1 < \ldots, t_n$, and $A \in \mathcal{B}\left((\mathbb{R}^d)^n\right)$. Then $\mathcal{I}$ is a $\pi-$system (i.e. stable by intersection), and by the previous step $\mathbf{1}_A(W_{t_1}, \ldots, W_{t_n}) \in \mathcal{R}$ for all $0 \leq t_1 < \ldots, t_n$. Then, for all $\xi \in \mathcal{R}^\perp$, we have $\mathbb{E}[\xi \mathbf{1}_A(W_{t_1}, \ldots, W_{t_n})] = 0$ or, equivalently

$$\mathbb{E}[\xi^+ \mathbf{1}_A(W_{t_1}, \ldots, W_{t_n})] \quad = \quad \mathbb{E}[\xi^- \mathbf{1}_A(W_{t_1}, \ldots, W_{t_n})].$$

In other words, the measures defined by the densities ξ^+ and ξ^- agree on the $\pi-$system $\mathcal{I}$. Since $\sigma(\mathcal{I}) = \mathcal{F}_T$, it follows from Proposition A.5 that $\xi^+ = \xi^-$ a.s. ♢

The last theorem can be stated equivalently in terms of square integrable martingales.

Theorem 7.2. *Let $\{M_t, 0 \leq t \leq T\}$ be a square integrable martingale. Then, there exists a unique process $H \in \mathbb{H}^2$ such that*

$$M_t = M_0 + \int_0^t H_s \cdot dW_s, \quad 0 \leq t \leq T, \quad \mathbb{P} - a.s.$$

In particular, M has continuous sample paths, a.s.

Proof. Apply Theorem 7.1 to the square integrable r.v. M_T, and take conditional expectations. ♢

We next extend the representation result to $\mathbb{L}^1$.

Theorem 7.3. *For any $F \in \mathbb{L}^1(\mathcal{F}_T, \mathbb{P})$ there exists a process $H \in \mathbb{H}^2_{\text{loc}}$ such that*

$$F = E[F] + \int_0^T H_s dW_s, \quad \mathbb{P} - a.s. \tag{7.5}$$

Proof. Let $M_t = E[F|\mathcal{F}_t]$. The first step is to show that M is continuous. Since bounded functions are dense in $\mathbb{L}^1$, there exists a sequence of bounded random variables (F_n) such that $\|F_n - F\|_{\mathbb{L}^1} \leq 3^{-n}$. We define $M_t^n = E[F_n|\mathcal{F}_t]$. From Theorem 3.11 (ii),

$$\mathbb{P}[\sup_{0\leq t\leq T} |M_t^n - M_t| > 2^{-n}] \leq 2^n E[|F - F_n|] \leq \left(\frac{2}{3}\right)^n.$$

Therefore, by Borel-Cantelli lemma, the martingales M^n converge uniformly to M, but since $F^n \in \mathbb{L}^2(\mathcal{F}_T, \mathbb{P})$, M^n is continuous for each n (by theorem 7.1), so the uniform limit M is also continous.

The second step is to show that M can be represented as stochastic integral. From continuity of M it follows that $|M_{t\wedge\tau_n}| \leq n$ for all t with $\tau_n = \inf\{t : |M_t| \geq n\}$. Therefore, by theorem 7.1, for each n, there exists a process $H^n \in \mathbb{H}^2$ with

$$M_{t\wedge\tau_n} = E[M_{t\wedge\tau_n}] + \int_0^{t\wedge\tau_n} H_s^n dW_s = E[F] + \int_0^{t\wedge\tau_n} H_s^n dW_s.$$

Moreover, Itô isometry shows that the processes H^n and H^m must coincide for $t \leq \tau_n \wedge \tau_m$. Define the process

$$H_t := \sum_{n\geq 1} H_t^n \mathbf{1}_{t\in]\tau_{n-1},\tau_n]}.$$

Since M is continuous, it is uniformly continuous on $[0, T]$, and therefore bounded a.s., which means that almost surely, starting from some n, $H_t = H_t^n$ for all $t \in [0, T]$, and so

$$\int_0^T H_s^2 ds < \infty \quad a.s.$$

On the other hand, for every n,

$$M_{t\wedge\tau_n} = E[F] + \int_0^{t\wedge\tau_n} H_s dW_s.$$

By passing to the almost-sure limit on each side of this equation, the proof is complete. ♢

With the last result, we can now extend the representation Theorem 7.2 to local martingales. Notice that uniqueness of the integrand process is lost.

Theorem 7.4. *Let $\{M_t, 0 \leq t \leq T\}$ be a local martingale. Then, there exists a process $H \in \mathbb{H}^2_{\text{loc}}$ such that*

$$M_t = M_0 + \int_0^t H_s \cdot dW_s, \quad 0 \leq t \leq T.$$

In particular, M has continuous sample paths, a.s.

Proof. There exists an increasing sequence of stopping times $(\tau_n)_{n\geq 1}$ with $\tau_n \longrightarrow \infty$, a.s., such that the stopped process $\{M_t^n := M_{t\wedge\tau_n}, t\in[0,T]\}$ is a bounded martingale for all n. Applying Theorem 7.3 to the bounded r.v. M_T^n, we obtain the representation $M_T^n = M_0 + \int_0^T H_s^n dW_s$, for some $H^n \in \mathbb{H}^2$. By direct conditioning, it follows that $M_t^n = M_0 + \int_0^t H_s^n dW_s$ for all $t \in [0,T]$. Similar to the proof of the previous Theorem 7.3, it follows from the Itô isometry that H^n and H^m coincide $dt \otimes d\mathbb{P}-$a.s. on $[0, \tau_n \wedge \tau_m]$, and we may define $H_t := \sum_{n\geq 1} H_t^n \mathbf{1}_{]\tau_{n-1},\tau_n]}$ so that $M_t^n = M_0 + \int_0^t H_s dW_s$, and we conclude by taking the almost sure limit in n. ♢

7.2 The Cameron-Martin change of measure

Let N be a Gaussian random variable with mean zero and unit variance. The corresponding probability density function is

$$\frac{\partial}{\partial x}\mathbb{P}[N \leq x] \;=\; f(x) \;=\; \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

For any constant $a \in \mathbb{R}$, the random variable $N + a$ is Gaussian with mean a and unit variance with probability density function

$$\frac{\partial}{\partial x}\mathbb{P}[N + a \leq x] \;=\; f(x-a) \;=\; f(x)e^{ax-\frac{a^2}{2}}, \qquad x \in \mathbb{R}.$$

Then, for every (at least bounded) function ψ, we have

$$\mathbb{E}\left[\psi(N+a)\right] = \int \psi(x) f(x) e^{ax-\frac{a^2}{2}} dx \;\;=\;\; \mathbb{E}\left[e^{aN-\frac{a^2}{2}}\psi(N)\right].$$

This easy result can be translated in terms of a change of measure. Indeed, since the random variable $e^{aN-\frac{a^2}{2}}$ is positive and integrates to 1, we may introduce the equivalent measure $\mathbb{Q} := e^{aN-\frac{a^2}{2}} \cdot \mathbb{P}$. Then, the above equality says that the $\mathbb{Q}-$distribution of N coincides with the $\mathbb{P}-$distribution of $N + a$, i.e.

$$\text{under } \mathbb{Q},\; N - a \quad \text{is distributed as} \quad \mathcal{N}(0,1).$$

The purpose of this subsection is to extend this result to a Brownian motion W in $\mathbb{R}^d$.

Let $h : [0,T] \longrightarrow \mathbb{R}^d$ be a deterministic function in $\mathbb{L}^2$, i.e. $\int_0^T |h(t)|^2 dt < \infty$. From Theorem 5.3, the stochastic integral

$$N \;\;:=\;\; \int_0^T h(t)\cdot dW_t$$

is well-defined as the $\mathbb{L}^2-$limit of the stochastic integral of some $\mathbb{H}^2-$approximating simple function. In particular, since the space of Gaussian random variables is closed, it follows that

$$N \quad \text{is distributed as} \quad \mathcal{N}\left(0, \int_0^T |h(t)|^2 dt\right),$$

and we may define an equivalent probability measure $\mathbb{Q}$ by:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \quad := \quad e^{\int_0^T h(t)\cdot dW_t - \frac{1}{2}\int_0^T |h(t)|^2 dt}. \tag{7.6}$$

Theorem 7.5 (Cameron-Martin formula)**.** *For a Brownian motion W in $\mathbb{R}^d$, let $\mathbb{Q}$ be the probability measure equivalent to $\mathbb{P}$ defined by* (7.6)*. Then, the process*

$$B_t := W_t - \int_0^t h(u)du\,, \qquad t \in [0,T]\,,$$

is a Brownian motion under $\mathbb{Q}$.

Proof. We first observe that $B_0 = 0$ and B has a.s. continuous sample paths. It remains to prove that, for $0 \le s < t$, $B_t - B_s$ is independent of $\mathcal{F}_s$ and distributed as a centered Gaussian with variance $t-s$. To do this, we compute the $\mathbb{Q}-$Laplace transform

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[e^{\lambda\cdot(W_t-W_s)}\middle|\mathcal{F}_s\right] &= \mathbb{E}\left[\frac{\mathbb{E}\left\{\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\mathcal{F}_t\right\}}{\mathbb{E}\left\{\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\mathcal{F}_s\right\}}e^{\lambda\cdot(W_t-W_s)}\middle|\mathcal{F}_s\right] \\
&= \mathbb{E}\left[e^{\int_s^t h(u)\cdot dW_u - \frac{1}{2}\int_s^t |h(u)|^2 du}e^{\lambda\cdot(W_t-W_s)}\middle|\mathcal{F}_s\right] \\
&= e^{-\frac{1}{2}\int_s^t |h(u)|^2 du}\mathbb{E}\left[e^{\int_s^t (h(u)+\lambda)\cdot dW_u}\middle|\mathcal{F}_s\right]
\end{aligned}$$

Since the random variable $\int_s^t (h(u)+\lambda)\cdot dW_u$ is a centered Gaussian with variance $\int_s^t |h(u)+\lambda|^2 du$, independent of $\mathcal{F}_s$, this provides:

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[e^{\lambda\cdot(W_t-W_s)}\middle|\mathcal{F}_s\right] &= e^{-\frac{1}{2}\int_s^t |h(u)|^2 du}e^{\frac{1}{2}\int_s^t |h(u)+\lambda|^2} \\
&= e^{\frac{1}{2}\lambda^2(t-s)+\lambda\int_s^t h(u)du}.
\end{aligned}$$

This shows that $W_t - W_s$ is independent of $\mathcal{F}_s$ and is distributed as a Gaussian with mean $\int_s^t h(u)du$ and variance $t-s$, i.e. $B_t - B_s$ is independent of $\mathcal{F}_s$ and is distributed as a centered Gaussian with variance $t-s$. ♢

7.3 The Girsanov's theorem

The Cameron-Martin change of measure formula of the preceding section can be extended to adapted stochastic processes satisfying suitable integrability conditions. Let W be a d-dimensional Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Given a $\mathcal{F}_T$-measurable positive random variable Z such that $E^{\mathbb{P}}[Z] = 1$, we define a new probability $\mathbb{Q}$ via $\mathbb{Q} := Z\mathbb{P}$:

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

Z is called the density of $\mathbb{Q}$ with respect to $\mathbb{P}$. For every $A \in \mathcal{F}_t$,

$$\mathbb{Q}(A) = E^{\mathbb{P}}[Z\mathbf{1}_A] = E^{\mathbb{P}}[\mathbf{1}_A E[Z|\mathcal{F}_t]].$$

Therefore, the martingale $Z_t := E[Z|\mathcal{F}_t]$ plays the role of the density of $\mathbb{Q}$ with respect to $\mathbb{P}$ on $\mathcal{F}_t$. The following lemma shows how to compute conditional expectations under $\mathbb{Q}$.

Lemma 7.6 (Bayes rule)**.** *Let $Y \in \mathcal{F}_T$ with $E^{\mathbb{Q}}[|Y|] < \infty$. Then*

$$E^{\mathbb{Q}}[Y|\mathcal{F}_t] = \frac{E[ZY|\mathcal{F}_t]}{E[Z|\mathcal{F}_t]} = \frac{1}{Z_t}E[ZY|\mathcal{F}_t].$$

Proof. Let $A \in \mathcal{F}_t$.

$$\begin{aligned} E^{\mathbb{Q}}[Y\mathbf{1}_A] &= E^{\mathbb{P}}[ZY\mathbf{1}_A] = E^{\mathbb{P}}[E^{\mathbb{P}}[ZY|\mathcal{F}_t]\mathbf{1}_A] \\ &= E^{\mathbb{P}}\left[\frac{Z}{E^{\mathbb{P}}[Z|\mathcal{F}_t]}E^{\mathbb{P}}[ZY|\mathcal{F}_t]\mathbf{1}_A\right] = E^{\mathbb{Q}}\left[\frac{E^{\mathbb{P}}[ZY|\mathcal{F}_t]}{E^{\mathbb{P}}[Z|\mathcal{F}_t]}\mathbf{1}_A\right]. \end{aligned}$$

Since the above is true for any $A \in \mathcal{F}$, this finishes the proof. □

Let ϕ be a process in $\mathbb{H}^2_{\text{loc}}$. Inspired by equation (7.6), we define a candidate for the martingale density:

$$Z_t = \exp\left(\int_0^t \phi_s \cdot dW_s - \frac{1}{2}\int_0^t |\phi_s|^2 ds\right), \quad 0 \le t \le T. \tag{7.7}$$

An application of Itô formula gives the dynamics of Z:

$$dZ_t = Z_t\phi_t \cdot dW_t,$$

which shows that Z is a local martingale (take $\tau_n = \inf\{t : \int_0^t Z_s^2\phi_s^2 ds \le n\}$ as localizing sequence). As shown by the following lemma, Z is also a supermartingale and therefore satisfies $E[Z_t] \le 1$ for all t.

Lemma 7.7. *For any stopping time $s \le t \le T$, $E[Z_t|\mathcal{F}_s] \le Z_s$.*

Proof. Let (τ_n) be a localizing sequence for Z. Then $Z_{t\wedge\tau_n}$ is a true martingale and by Fatou's lemma, which can be applied to conditional expectations,

$$E[Z_t|\mathcal{F}_s] = E[\lim_n Z_{t\wedge\tau_n}|\mathcal{F}_s] \le \liminf_n E[Z_{t\wedge\tau_n}|\mathcal{F}_s] = \liminf_n Z_{s\wedge\tau_n} = Z_s.$$

□

However, Z may sometimes fail to be a true martingale. A sufficient condition for Z to be a true martingale for given in the next section; for now let us assume that this is the case.

Theorem 7.8 (Girsanov)**.** *Let Z be given by (7.7) and suppose that $E[Z_T] = 1$. Then the process*

$$\tilde{W}_t := W_t - \int_0^t \phi_s ds, \quad t \le T$$

is a Brownian motion under the probability $\mathbb{Q} := Z_T \mathbb{P}$ on $\mathcal{F}_T$.

Proof. Step 1: Let $Y_t := \int_0^t b_s \cdot d\tilde{W}_s$, where b is an adapted process with $\int_0^T |b_s|^2 ds$, and let $X_t := Z_t Y_t$. Applying Itô's formula to X, we get

$$dX_t = Z_t(b_t \cdot d\tilde{W}_t) + Y_t(\phi_t Z_t \cdot dW_t) + b_t \cdot Z_t \phi_t dt = Z_t(b_t + Y_t \phi_t) \cdot dW_t,$$

which shows that X is a local martingale under $\mathbb{P}$. Let (τ_n) be a localizing sequence for X under $\mathbb{P}$. By lemma 7.6, for $t \ge s$,

$$\begin{aligned}
E^{\mathbb{Q}}[Y_{t\wedge\tau_n}|\mathcal{F}_s] &= \frac{1}{Z_s} E^{\mathbb{P}}[Z_T Y_{t\wedge\tau_n}|\mathcal{F}_s] \\
&= \frac{1}{Z_s} E^{\mathbb{P}}[E^{\mathbb{P}}[Z_T|\mathcal{F}_{s\vee\tau_n\wedge t}] Y_{t\wedge\tau_n}|\mathcal{F}_s] = \frac{1}{Z_s} E^{\mathbb{P}}[Z_{s\vee\tau_n\wedge t} Y_{t\wedge\tau_n}|\mathcal{F}_s] \\
&= \frac{1}{Z_s} E^{\mathbb{P}}[X_{\tau_n\wedge t} 1_{\tau_n \ge s} + Z_s Y_{\tau_n} 1_{\tau_n < s}|\mathcal{F}_s] \\
&= \frac{1}{Z_s} \{X_s 1_{\tau_n \ge s} + Z_s Y_{\tau_n} 1_{\tau_n < s}\} = Y_{s\wedge\tau_n},
\end{aligned}$$

which shows that Y is a local martignale under $\mathbb{Q}$ with (τ_n) as localizing sequence.

Step 2: For a fixed $u \in \mathbb{R}^d$, applying the Itô formula to $e^{iu\tilde{W}}$ between s and t and multiplying the result by $e^{-iu\cdot\tilde{W}_s}$, we get:

$$e^{iu\cdot(\tilde{W}_t - \tilde{W}_s)} = 1 + i\int_s^t e^{iu\cdot(\tilde{W}_r - \tilde{W}_s)} \cdot u d\tilde{W}_r - \frac{1}{2}|u|^2 \int_s^t e^{iu\cdot(\tilde{W}_r - \tilde{W}_s)} dr. \tag{7.8}$$

By step 1, the stochastic integral

$$\int_s^t e^{iu\cdot(\tilde{W}_r - \tilde{W}_s)} u \cdot d\tilde{W}_r$$

is a local martingale (as a function of the parameter t), and since it is also bounded (because all the other terms in the equation (7.8) are bounded), the dominated convergence theorem implies that it is a true martingale and satisfies

$$E^{\mathbb{Q}}\left[\int_s^t e^{iu\cdot(\tilde{W}_r - \tilde{W}_s)} u \cdot d\tilde{W}_r \middle| \mathcal{F}_s\right] = 0.$$

Let $h^{\mathbb{Q}}(t) := E^{\mathbb{Q}}[e^{iu\cdot(\tilde{W}_r - \tilde{W}_s)} \big| \mathcal{F}_s]$. Taking $\mathbb{Q}$-expectations on both sides of (7.8) leads to an integral equation for $h^{\mathbb{Q}}$:

$$h^{\mathbb{Q}}(t) = 1 - \frac{1}{2}|u|^2 \int_s^t h^{\mathbb{Q}}(r) dr \quad \Rightarrow \quad h^{\mathbb{Q}}(t) = e^{-|u|^2(t-s)/2}.$$

This proves that $\tilde{W}$ has independent stationary and normally distributed increments under $\mathbb{Q}$. Since it is also continuous and $\tilde{W}_0$, $\tilde{W}$ is a Brownian motion under $\mathbb{Q}$. $\square$

7.3.1 The Novikov's criterion

Theorem 7.9 (Novikov)**.** *Suppose that*

$$E[e^{\frac{1}{2}\int_0^T |\phi_s|^2 ds}] < \infty.$$

Then $E[Z_T] = 1$ and the process $\{Z_t, 0 \le t \le T\}$ is a martingale.

In the following lemma and below, we use the notation

$$Z_t^{(a)} = \exp\left(\int_0^t a\phi_s \cdot dW_s - \frac{a^2}{2}\int_0^t |\phi_s|^2 ds\right), \quad 0 \le t \le T.$$

Lemma 7.10. *Assume that*

$$\sup_\tau E[e^{\frac{1}{2}\int_0^\tau \phi_s \cdot dW_s}] < \infty,$$

where the sup *is taken over all stopping times τ with $\tau \le T$. Then for all $a \in (0,1)$ and all $t \le T$, $E[Z_t^{(a)}] = 1$.*

Proof. Let $q = \frac{1}{a(2-a)} > 1$ et $r = \frac{2-a}{a} > 1$, and let τ be a stopping time with $\tau \le T$. Then, applying the Hölder inequality with $\frac{1}{s} + \frac{1}{r} = 1$ and observing that $s(aq - a\sqrt{\frac{q}{r}}) = \frac{1}{2}$,

$$\begin{aligned}
\mathbb{E}[(Z_\tau^{(a)})^q] &= \mathbb{E}[e^{aq\int_0^\tau \phi_s \cdot dW_s - \frac{a^2 q}{2}\int_0^\tau |\phi_s|^2 ds}] \\
&= \mathbb{E}[e^{a\sqrt{\frac{q}{r}}\int_0^\tau \phi_s \cdot dW_s - \frac{a^2 qr}{2}\int_0^\tau |\phi_s|^2 ds} e^{(aq - a\sqrt{\frac{q}{r}})\int_0^\tau \phi_s \cdot dW_s}] \\
&\le \mathbb{E}[Z_\tau^{(a\sqrt{qr})}]^{\frac{1}{r}} E[e^{\frac{1}{2}\int_0^\tau \phi_s \cdot dW_s}]^{\frac{1}{s}} \le \mathbb{E}[e^{\frac{1}{2}\int_0^\tau \phi_s \cdot dW_s}]^{\frac{1}{s}}
\end{aligned}$$

by Lemma 7.7. Therefore, $\sup_\tau \mathbb{E}[(Z_\tau^{(a)})^q]$ is bounded. With the sequence $\{\tau_n\}$ defined by $\tau_n = \inf\{t : \int_0^t a^2 (Z_s^{(a)})^2 \phi_s^2 ds \le n\}$, we now have by Doob's maximal inequality of Theorem 3.15,

$$\mathbb{E}[\sup_{t\le T} Z_{t\wedge\tau_n}^{(a)}] \le \mathbb{E}\left[\left(\sup_{t\le T} Z_{t\wedge\tau_n}^{(a)}\right)^q\right]^{\frac{1}{q}} \le \left(\frac{q}{q-1}\sup_{t\le T}\mathbb{E}[(Z_{t\wedge\tau_n}^{(a)})^q]\right)^{\frac{1}{q}} < \infty.$$

By monotone convergence, this implies

$$\mathbb{E}[\sup_{t\le T} Z_t^{(a)}] < \infty,$$

and by dominated convergence we then conclude that $E[Z_t^{(a)}] = 1$ for all $t \le T$. $\square$

Proof of theorem 7.9. For any stopping time $\tau \leq T$,

$$e^{\frac{1}{2}\int_0^\tau \phi_s \cdot dW_s} = Z_\tau^{\frac{1}{2}} \left(e^{\frac{1}{2}\int_0^\tau |\phi_s|^2 ds} \right)^{\frac{1}{2}},$$

and by an application of the Cauchy-Schwartz inequality, using lemma 7.7 and the assumption of the theorem it follows that

$$\sup_\tau \mathbb{E}[e^{\frac{1}{2}\int_0^\tau \phi_s \cdot dW_s}] < \infty. \tag{7.9}$$

Fixing $a < 1$ and $t \leq T$ and observing that

$$Z_t^{(a)} = e^{a^2 \int_0^t \phi_s \cdot dW_s - \frac{a^2}{2}\int_0^t |\phi_s|^2 ds} e^{a(1-a)\int_0^t \phi_s \cdot dW_s} = Z_t^{a^2} e^{a(1-a)\int_0^t \phi_s \cdot dW_s},$$

we get, by Hölder's inequality and Lemma 7.10,

$$1 = \mathbb{E}[Z_t^{(a)}] \leq \mathbb{E}[Z_t]^{a^2} \mathbb{E}[e^{\frac{a}{1+a}\int_0^t \phi_s \cdot dW_s}]^{1-a^2} \leq \mathbb{E}[Z_t]^{a^2} \mathbb{E}[e^{\frac{1}{2}\int_0^t \phi_s \cdot dW_s}]^{2a(1-a)}.$$

Sending a to 1 and using (7.9) yields $\mathbb{E}[Z_t] \geq 1$, it follows from Lemma 7.7 that $\mathbb{E}[Z_t] = 1$.

Now, let $s \leq t \leq T$. In the same way as in Lemma 7.7, one can show that $\mathbb{E}[Z_t|\mathcal{F}_s] \leq Z_s$, and taking the expectation of both sides shows that $\mathbb{E}[Z_t] = 1$ can hold only if $\mathbb{E}[Z_t|\mathcal{F}_s] = Z_s$, $\mathbb{P}$-a.s. □

7.4 Application: the martingale approach to the Black-Scholes model

This section contains the modern approach to the Black-Scholes valuation and hedging theory. The prices will be modelled by Itô processes, and the results obtained by the previous approaches will be derived by the elegant martingale approach.

7.4.1 The continuous-time financial market

Let T be a finite horizon, and $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space supporting a Brownian motion $W = \{(W_t^1, \ldots, W_t^d), 0 \leq t \leq T\}$ with values in $\mathbb{R}^d$. We denote by $\mathbb{F} = \mathbb{F}^W = \{\mathcal{F}_t, 0 \leq t \leq T\}$ the canonical augmented filtration of W, i.e. the canonical filtration augmented by zero measure sets of $\mathcal{F}_T$.

The financial market consists in $d + 1$ assets :

(i) The first asset S^0 is non-risky, and is defined by

$$S_t^0 = \exp\left(\int_0^t r_u du \right), \qquad 0 \leq t \leq T,$$

where $\{r_t, t \in [0, T]\}$ is a non-negative measurable and adapted processes with $\int_0^T r_t dt < \infty$ a.s., and represents the instantaneous interest rate.

(ii) The d remaining assets S^i, $i = 1, \ldots, d$, are risky assets with price processes defined by the equations

$$S_t^i = S_0^i \exp\left(\int_0^t \left(\mu_u^i - \frac{1}{2}\sum_{j=1}^d \left|\sigma_u^{ij}\right|^2\right) du + \sum_{j=1}^d \int_0^t \sigma_u^{ij} dW_u^j\right), \qquad t \geq 0,$$

where μ, σ are measurable and $\mathbb{F}-$adapted processes with $\int_0^T |\mu_t^i| dt + \int_0^T |\sigma^{i,j}|^2 dt < \infty$ for all $i, j = 1, \ldots, d$. Applying Itô's formula, we see that

$$\frac{dS_t^i}{S_t^i} \quad = \quad \mu_t^i dt + \sum_{j=1}^d \sigma_t^{i,j} dW_t^j, \ t \in [0,T],$$

for $1 \leq i \leq d$. It is convenient to use the matrix notations to represent the dynamics of the price vector $S = (S^1, \ldots, S^d)$:

$$dS_t \quad = \quad \mathrm{diag}[S_t] \left(\mu_t dt + \sigma_t dW_t\right), \ t \in [0,T],$$

where $\mathrm{diag}[S_t]$ is the $d \times d-$diagonal matrix with diagonal $i-$th component given by S_t^i, and μ, σ are the $\mathbb{R}^d-$vector with components μ^i's, and the $\mathcal{M}_{\mathbb{R}}(d,d)-$matrix with entries $\sigma^{i,j}$.

We assume that the $\mathcal{M}_{\mathbb{R}}(d,d)-$matrix σ_t is invertible for every $t \in [0,T]$ a.s., and we introduce the process

$$\lambda_t \ := \ \sigma_t^{-1}\left(\mu_t - r_t \mathbf{1}\right), \qquad 0 \leq t \leq T,$$

called the *risk premium process*. Here $\mathbf{1}$ is the vector of ones in $\mathbb{R}^d$. We shall frequently make use of the discounted processes

$$\tilde{S}_t \ := \ \frac{S_t}{S_t^0} \quad = \quad S_t \exp\left(-\int_0^t r_u du\right),$$

Using the above matrix notations, the dynamics of the process $\tilde{S}$ are given by

$$d\tilde{S}_t \quad = \quad \mathrm{diag}[\tilde{S}_t]\left\{(\mu_t - r_t\mathbf{1})dt + \sigma_t dW_t\right\} \ = \ \mathrm{diag}[\tilde{S}_t]\sigma_t\left(\lambda_t dt + dW_t\right).$$

7.4.2 Portfolio and wealth process

A portfolio strategy is an $\mathbb{F}-$adapted process $\theta = \{\theta_t, 0 \leq t \leq T\}$ with values in $\mathbb{R}^d$. For $1 \leq i \leq n$ and $0 \leq t \leq T$, θ_t^i is the amount (in Euros) invested in the risky asset S^i.

We next recall the self-financing condition in the present framework. Let X_t^θ denote the portfolio value, or wealth, process at time t induced by the portfolio strategy θ. Then, the amount invested in the non-risky asset is $X_t^\theta - \sum_{i=1}^n \theta_t^i$ $= X_t - \theta_t \cdot \mathbf{1}$.

Under the self-financing condition, the dynamics of the wealth process is given by

$$dX_t^\theta \quad = \quad \sum_{i=1}^n \frac{\theta_t^i}{S_t^i} \, dS_t^i \ + \ \frac{X_t - \theta_t \cdot \mathbf{1}}{S_t^0} \, dS_t^0.$$

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

$$\tilde{X}_t \ := \ X_t \exp\left(-\int_0^t r(u)du\right), \qquad 0 \le t \le T\,.$$

Then, by an immediate application of Itô's formula, we see that

$$\begin{aligned} d\tilde{X}_t &= \tilde{\theta}_t \cdot \mathrm{diag}[\tilde{S}_t]^{-1} d\tilde{S}_t && (7.10)\\ &= \tilde{\theta}_t \cdot \sigma_t \left(\lambda_t dt + dW_t\right), \ 0 \le t \le T\,. && (7.11) \end{aligned}$$

We still need to place further technical conditions on θ, at least in order for the above wealth process to be well-defined as a stochastic integral.

Before this, let us observe that, assuming that the risk premium process satisfies the Novikov condition:

$$\mathbb{E}\left[e^{\frac{1}{2}\int_0^T |\lambda_t|^2 dt}\right] \ < \ \infty,$$

it follows from the Girsanov theorem that the process

$$B_t \ := \ W_t + \int_0^t \lambda_u du\,, \qquad 0 \le t \le T\,, \tag{7.12}$$

is a Brownian motion under the equivalent probability measure

$$\mathbb{Q} := Z_T \cdot \mathbb{P} \text{ on } \mathcal{F}_T \quad \text{where} \quad Z_T := \exp\left(-\int_0^T \lambda_u \cdot dW_u - \frac{1}{2}\int_0^T |\lambda_u|^2 du\right).$$

In terms of the $\mathbb{Q}$ Brownian motion B, the discounted price process satisfies

$$d\tilde{S}_t \ = \ \mathrm{diag}[\tilde{S}_t]\sigma_t dB_t, \quad t \in [0,T],$$

and the discounted wealth process induced by an initial capital X_0 and a portfolio strategy θ can be written in

$$\tilde{X}^\theta_t \ = \ \tilde{X}_0 + \int_0^t \tilde{\theta}_u \cdot \sigma_u dB_u, \quad \text{for} \quad 0 \le t \le T. \tag{7.13}$$

Definition 7.11. *An admissible portfolio process $\theta = \{\theta_t, t \in [0,T]\}$ is a measurable and $\mathbb{F}-$adapted process such that $\int_0^T |\sigma_t^{\mathrm{T}} \tilde{\theta}_t|^2 dt < \infty$, a.s. and the corresponding discounted wealth process is bounded from below by a $\mathbb{Q}-$martingale*

$$\tilde{X}^\theta_t \ge M^\theta_t, \ 0 \le t \le T, \quad \textit{for some } \mathbb{Q}-\textit{martingale} \quad M^\theta > 0.$$

The collection of all admissible portfolio processes will be denoted by $\mathcal{A}$.

The lower bound M^θ, which may depend on the portfolio θ, has the interpretation of a finite credit line imposed on the investor. This natural generalization of the more usual constant credit line corresponds to the situation where the total credit available to an investor is indexed by some financial holding, such as the physical assets of the company or the personal home of the investor, used as collateral. From the mathematical viewpoint, this condition is needed in order to exclude any arbitrage opportunity, and will be justified in the subsequent subsection.

7.4.3 Admissible portfolios and no-arbitrage

We first define precisely the notion of no-arbitrage.

Definition 7.12. *We say that the financial market contains no arbitrage opportunities if for any admissible portfolio process $\theta \in \mathcal{A}$,*

$$X_0 = 0 \text{ and } X_T^\theta \geq 0 \ \mathbb{P}-a.s. \quad \text{implies} \quad X_T^\theta \ = \ 0 \ \mathbb{P}-a.s.$$

The purpose of this section is to show that the financial market described above contains no arbitrage opportunities. Our first observation is that, by the very definition of the probability measure $\mathbb{Q}$, the discounted price process $\tilde{S}$ satisfies:

$$\text{the process} \quad \left\{\tilde{S}_t, \ 0 \leq t \leq T\right\} \quad \text{is a } \mathbb{Q}-\text{local martingale.} \tag{7.14}$$

For this reason, $\mathbb{Q}$ is called a *risk neutral measure*, or an *equivalent local martingale measure*, for the price process S.

We also observe that the discounted wealth process satisfies:

$$\tilde{X}^\theta \quad \text{is a } \mathbb{Q}-\text{local martingale for every} \quad \theta \in \mathcal{A}, \tag{7.15}$$

as a stochastic integral with respect to the $\mathbb{Q}-$Brownian motion B.

Theorem 7.13. *The continuous-time financial market described above contains no arbitrage opportunities.*

Proof. For $\theta \in \mathcal{A}$, the discounted wealth process $\tilde{X}^\theta$ is a $\mathbb{Q}-$local martingale bounded from below by a $\mathbb{Q}-$martingale. From Lemma 5.13 and Exercise 5.14, we deduce that $\tilde{X}^\theta$ is a $\mathbb{Q}-$super-martingale. Then $\mathbb{E}^{\mathbb{Q}}\left[\tilde{X}_T^\theta\right] \leq \tilde{X}_0 = X_0$. Recall that $\mathbb{Q}$ is equivalent to $\mathbb{P}$ and S^0 is strictly positive. Then, this inequality shows that, whenever $X_0^\theta = 0$ and $X_T^\theta \geq 0$ $\mathbb{P}-$a.s. (or equivalently $\mathbb{Q}-$a.s.), we have $\tilde{X}_T^\theta = 0$ $\mathbb{Q}-$a.s. and therefore $X_T^\theta = 0$ $\mathbb{P}-$a.s. $\diamondsuit$

7.4.4 Super-hedging and no-arbitrage bounds

Let G be an $\mathcal{F}_T-$measurable random variable representing the payoff of a derivative security with given maturity $T > 0$. The *super-hedging* problem consists in finding the minimal initial cost so as to be able to face the payment G without risk at the maturity of the contract T:

$$V(G) \quad := \quad \inf\left\{X_0 \in \mathbb{R} \ : \ X_T^\theta \geq G \ \mathbb{P}-\text{a.s. for some } \theta \in \mathcal{A}\right\} .$$

Remark 7.14. Notice that $V(G)$ depends on the reference measure $\mathbb{P}$ only by means of the corresponding null sets. Therefore, the super-hedging problem is not changed if $\mathbb{P}$ is replaced by any equivalent probability measure.

The following properties of the super-hedging problem are easy to prove.

Proposition 7.15. *The function $G \longmapsto V(G)$ is*

1. *monotonically increasing, i.e.* $V(G_1) \geq V(G_2)$ *for every* $G_1, G_2 \in \mathbb{L}^0(T)$ *with* $G_1 \geq G_2$ $\mathbb{P}-a.s.$
2. *sublinear, i.e.* $V(G_1 + G_2) \leq V(G_1) + V(G_2)$ *for* $G_1, G_2 \in \mathbb{L}^0(T)$*,*
3. *positively homogeneous, i.e.* $V(\lambda G) = \lambda V(G)$ *for* $\lambda > 0$ *and* $G \in \mathbb{L}^0(T)$*,*
4. $V(0) = 0$ *and* $V(G) \geq -V(-G)$ *for every contingent claim* G*.*
5. *Let* G *be a contingent claim, and suppose that* $X_T^{\theta^0} = G$ $\mathbb{P}-a.s.$ *for some* X_0 *and* $\theta^0 \in \mathcal{A}$*. Then*

$$V(G) \quad = \quad \inf\left\{X_0 \in \mathbb{R} \ : \ X_T^\theta = G \ \mathbb{P} - a.s. \ for\ some\ \theta \in \mathcal{A}\right\}.$$

Exercise 7.16. *Prove Proposition 7.15.*

We now show that, under the no-arbitrage condition, the super-hedging problem provides *no-arbitrage bounds* on the market price of the derivative security.

Assume that the buyer of the contingent claim G has the same access to the financial market than the seller. Then $V(G)$ *is the maximal amount that the buyer of the contingent claim contract is willing to pay.* Indeed, if the seller requires a premium of $V(G) + 2\varepsilon$, for some $\varepsilon > 0$, then the buyer would not accept to pay this amount as he can obtain at least G be trading on the financial market with initial capital $V(G) + \varepsilon$.

Now, since selling of the contingent claim G is the same as buying the contingent claim $-G$, we deduce from the previous argument that

$$-V(-G) \quad \leq \ \text{market price of } G \ \leq \quad V(G)\,. \tag{7.16}$$

Observe that this defines a non-empty interval for the market price of B under the no-arbitrage condition, by Proposition 7.15.

7.4.5 Heuristics from linear programming

In this subsection, we present some heuristics which justify that the occurrence of the risk-neutral measure $\mathbb{Q}$ as a crucial tool for the superhedging problem $V(G)$. We first re-write this optimization problem as:

$$V(G) \quad = \quad \inf_{(X_0,\theta)\in\mathcal{S}(G)} X_0,$$

where $\mathcal{S}(G)$ is the collection of all superhedging strategies for the payoff G at maturity T:

$$\mathcal{S}(G) \quad := \quad \left\{(X_0,\theta) \in \mathbb{R}\times\mathcal{A} : \tilde{X}_T^\theta \geq \tilde{G}, \ \mathbb{P} - \text{a.s.}\right\}.$$

Under this form, we see that $V(G)$ is a linear optimization problem under linear inequality constraints. The standard approach for such an optimization problem is to use the Kuhn-Tucker theorem. However, one then faces the main difficulty that $\mathcal{S}(G)$ consists of an infinite number of constraints, i.e. for $\mathbb{P}-$almost all $\omega \in \Omega$, we have to satisfy the constraint $\tilde{X}^\theta_T(\omega) \ge \tilde{G}(\omega)$.

Let us ignore this difficulty, and introduce Lagrange multipliers $Z(\omega) \ge 0$, $\omega \in \Omega$ for each inequality constraint, so that

$$\begin{aligned} V(G) &= \inf_{(X_0,\theta)\in\mathbb{R}\times\mathcal{A}} \sup_{Z\in\mathbb{L}^0(\mathbb{R}_+)} \big\{X_0 - \mathbb{E}\big[Z(\tilde{X}^\theta_T - \tilde{G})\big]\big\} \\ &= \inf_{(X_0,\theta)\in\mathbb{R}\times\mathcal{A}} \sup_{Z\in\mathbb{L}^0(\mathbb{R}_+)} \Big\{X_0(1-\mathbb{E}[Z]) - \mathbb{E}\Big[Z\Big(\int_0^T \Delta_t\cdot d\tilde{S}_t - \tilde{G}\Big)\Big]\Big\}, \end{aligned}$$

where the process Δ is defined by $\Delta_t := \mathrm{diag}[S_t]^{-1}\theta_t$, $t \in [0,T]$.

The next step is to justify the commutation of the inf and sup operators by the min-max theorem. This is the main difficulty of this approach in our infinite dimensional optimization problem. In the present heuristic argument, we ignore this difficulty and we assume that we may indeed justify that

$$V(G) = \sup_{Z\in\mathbb{L}^0(\mathbb{R}_+)} \inf_{(X_0,\theta)\in\mathbb{R}\times\mathcal{A}} \Big\{X_0(1-\mathbb{E}[Z]) - \mathbb{E}\Big[Z\Big(\int_0^T \Delta_t\cdot d\tilde{S}_t - \tilde{G}\Big)\Big]\Big\}.$$

Under this form, we see that in order to guarantee a finite value, it is necessary to restrict the minimization to those Lagrange multipliers $Z \in \mathbb{L}^0(\mathbb{R}_+)$ such that $\mathbb{E}[Z] = 1$. Then, Z can be identified to the probability measure $\tilde{\mathbb{P}}$ absolutely continuous with respect to $\mathbb{P}$, defined by the density $d\tilde{\mathbb{P}} = Zd\mathbb{P}$ on $\mathcal{F}_T$. The problem is thus reduced to:

$$V(G) = \sup_{\tilde{\mathbb{P}}\sim\mathbb{P}} \inf_{\theta\in\mathcal{A}} \mathbb{E}^{\tilde{\mathbb{P}}}\Big[-\int_0^T \Delta_t\cdot d\tilde{S}_t + \tilde{G}\Big].$$

We next consider some special hedging strategies $\theta \in \mathcal{A}$. Consider the buy-and-hold strategies

$$\Delta^{t,t',H}_u := H_t \mathbf{1}_{[t,t']}(u),\ u\in[0,T], \quad \text{for all} \quad t \le t',\ \text{and } H \in \mathbb{L}^\infty(\mathcal{F}_t).$$

Then, we directly compute that

$$\mathbb{E}^{\tilde{\mathbb{P}}}\Big[-\int_0^T \Delta^{t,t',H}_u\cdot d\tilde{S}_u\Big] = \mathbb{E}^{\tilde{\mathbb{P}}}\big[-H_t\cdot(\tilde{S}_{t'} - \tilde{S}_t)\big],$$

and it follows from the arbitrariness of $H \in \mathbb{L}^\infty(\mathcal{F}_t)$ that, in order to guarantee that the above inf is finite, it is necessary to restrict the maximization to those probability measures $\tilde{\mathbb{P}} \sim \mathbb{P}$ such that $\mathbb{E}^{\tilde{\mathbb{P}}}[-H_t\cdot(\tilde{S}_{t'} - \tilde{S}_t)] = 0$ for all $H \in \mathbb{L}^\infty(\mathcal{F}_t)$. By the definition of the conditional expectation, this is equivalent to $\mathbb{E}^{\tilde{\mathbb{P}}}\big[\tilde{S}_{t'}|\mathcal{F}_t\big] = \tilde{S}_t$, a.s. Since $t \le t'$ are arbitrary, this means that $\tilde{\mathbb{P}}$ must be chosen so that $\tilde{S}$ is $\tilde{\mathbb{P}}-$martingale. Then ignoring the local martingale issue of the stochastic integral $\int \Delta_t d\tilde{S}_t$, we obtain our final result

$$V(G) = \sup_{\tilde{\mathbb{P}}\in\mathcal{M}(\mathbb{P})} \mathbb{E}^{\tilde{\mathbb{P}}}\big[\tilde{G}\big] \quad \text{where} \quad \mathcal{M}(\mathbb{P}) := \big\{\tilde{\mathbb{P}}\sim\mathbb{P} : \tilde{S} \text{ is a } \tilde{\mathbb{P}} - \text{martingale}\big\}.$$

In the context of the financial market considered throughout this section, we observe that the set of equivalent martingale measures $\mathcal{M}(\mathbb{P}) = \{\mathbb{Q}\}$ contains one single element, namely the risk-neutral measure $\mathbb{Q}$. Then, the expected value of the superhedging cost is:

$$V(G) \quad = \quad \mathbb{E}^{\mathbb{Q}}[\tilde{G}].$$

This will be proved rigorously in Theorem 7.17 below.

7.4.6 The no-arbitrage valuation formula

We denote by $p(G)$ the market price of a derivative security G.

Theorem 7.17. *Let G be an $\mathcal{F}_T-$measurabel random variable representing the payoff of a derivative security at the maturity $T > 0$, and recall the notation $\tilde{G} := G \exp\left(-\int_0^T r_t dt\right)$. Assume that $\mathbb{E}^{\mathbb{Q}}[|\tilde{G}|] < \infty$. Then*

$$p(G) \ = \ V(G) \quad = \quad \mathbb{E}^{\mathbb{Q}}[\tilde{G}].$$

Moreover, there exists a portfolio $\theta^* \in \mathcal{A}$ such that $X_0^{\theta^*} = p(G)$ and $X_T^{\theta^*} = G$, a.s., that is θ^* is a perfect replication strategy.

Proof. 1- We first prove that $V(G) \geq \mathbb{E}^{\mathbb{Q}}[\tilde{G}]$. Let X_0 and $\theta \in \mathcal{A}$ be such that $X_T^\theta \geq G$, a.s. or, equivalently, $\tilde{X}_T^\theta \geq \tilde{G}$ a.s. Since $\tilde{X}^\theta$ is a $\mathbb{Q}-$local martingale bounded from below by a $\mathbb{Q}-$martingale, we deduce from Lemma 5.13 and Exercise 5.14 that $\tilde{X}^\theta$ is a $\mathbb{Q}-$super-martingale. Then $X_0 = \tilde{X}_0 \geq \mathbb{E}^{\mathbb{Q}}[\tilde{X}_T^\theta] \geq \mathbb{E}^{\mathbb{Q}}[\tilde{G}]$.
2- We next prove that $V(G) \leq \mathbb{E}^{\mathbb{Q}}[\tilde{G}]$. Define the $\mathbb{Q}-$martingale $Y_t := \mathbb{E}^{\mathbb{Q}}[\tilde{G}|\mathcal{F}_t]$ and observe that $\mathbb{F}^W = \mathbb{F}^B$. Then, it follows from the martingale representation theorem 7.3 that $Y_t = Y_0 + \int_0^t \phi_u \cdot dB_u$ for some $\phi \in \mathbb{H}^2_{\text{loc}}$. Setting $\theta^* := (\sigma^{\mathrm{T}})^{-1}\phi$, we see that

$$\theta^* \in \mathcal{A} \quad \text{and} \quad Y_0 + \int_0^T \theta^* \cdot \sigma_t dB_t = \tilde{G} \ \mathbb{P} - \text{a.s.}$$

which implies that $Y_0 \geq V(G)$ and θ^* is a perfect hedging stratgey for G, starting from the initial capital Y_0.
3- From the previous steps, we have $V(G) = \mathbb{E}^{\mathbb{Q}}[\tilde{G}]$. Applying this result to $-G$, we see that $V(-G) = -V(G)$, so that the no-arbitrage bounds (7.16) imply that the no-arbitrage market price of G is given by $V(G)$. ♢

7.5 The continuous time Kalman-Bucy filter

In this section, we provide the continuous time analogue of Section C.4.

7.5.1 Formulation

Let $f, g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ be deterministic continuous functions, $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t\ge 0}, \mathbb{P})$ a filtered probability space supporting a standard Brownian motion $W := (W^X, W^Y)$ in $\mathbb{R}^2$, and consider the processes X and Y defined by the linear stochastic differential equations

$$X_t = \int_0^t f(s) X_s ds + W_t^X, \qquad Y_t = \int_0^t g(s) X_s ds + W_t^Y, \quad t \ge 0.$$

The process X models a non-observable signal, and the process Y represents a noisy observation of X.

Notice that the process X is a time-dependent Ornstein-Uhlenbeck process. Similarly to the constant coefficients case, we may obtain it in explicit form by direct application of Itô's formula to the product $e^{-\int_0^t f(s)ds} X_s$:

$$X_t = \int_0^t e^{\int_s^t f(r)dr} dW_s^X, \quad \text{for all} \quad t \ge 0,$$

Then, X_t is a centered Gaussian with variance $\int_0^t e^{2\int_s^t f(r)dr} ds$.

In fact the pair (X, Y) is a vector Ornstein-Uhlenbeck process with time-dependent (matrix) coefficients, and we may therefore show similarly that (X, Y) is a Gaussian process.

The main objective of the present problem is to characterize the best estimate of X given the observation of Y:

$$\hat{X}_t := \mathbb{E}\big[X_t | \mathcal{F}_t^Y\big], \quad \text{where} \quad \mathcal{F}_t^Y := \sigma(Y_s, s \le t), \quad t \ge 0.$$

We denote by $\mathbb{F}^Y := \{\mathcal{F}_t^Y, t \ge 0\}$ the canonical filtration of Y, and by $\mathbb{L}^2(\mathcal{F}_t^Y)$ the space of square integrable $\mathcal{F}_t^Y-$measurable random variables. By definition, the conditional expectation $\mathbb{E}[.|\mathcal{F}_t^Y]$ is the orthogonal projection on the vector space $\mathbb{L}^2(\mathcal{F}_t^Y)$. Then $X_t - \hat{X}_t$ is orthogonal to $\mathbb{L}^2(\mathcal{F}_t^Y)$, and

$$\Gamma(t) \quad := \quad \mathbb{E}\big[(X_t - \hat{X}_t)^2\big] \ \le \ \mathbb{E}\big[X_t^2\big] < \infty. \tag{7.17}$$

In addition, as the process (X, Y) is Gaussian, we deduce that

$$X_t - \hat{X}_t \quad \text{is independent of} \quad \mathbb{L}^2(\mathcal{F}_t^Y), \quad \text{for all} \quad t \ge 0. \tag{7.18}$$

See Section C.4.1 for a review on conditional distributions of Gaussian vectors.

7.5.2 Main result

Definition 7.18. *The innovation process $\mathfrak{I} = \{\mathfrak{I}_t, t \ge 0\}$ is defined by the $\mathbb{F}^Y-$adapted process $\mathfrak{I}_t := Y_t - \int_0^t g(s)\hat{X}_s ds$, $t \ge 0$, so that*

$$dY_t \quad = \quad g(t)\hat{X}_t dt + d\mathfrak{I}_t, \quad t \ge 0.$$

Theorem 7.19 (The Kalman-Bucy filter). *The innovation process $\mathfrak{I}$ is an $(\mathbb{P}, \mathbb{F}^Y)-$Brownian motion, and $\hat{X}$ is an Ornstein-Uhlenbeck process with dynamics:*

$$\hat{X}_t \;=\; \int_0^t f(s)\hat{X}_s ds + \int_0^t g(s)\Gamma(s)d\mathfrak{I}_s, \quad \textit{for all} \quad t \geq 0, \tag{7.19}$$

where Γ is the unique solution of the ODE:

$$\Gamma(t) = \int_0^t \big(2f(s)\Gamma(s) - g(t)^2\Gamma(t)^2 + 1\big)ds, \quad t \geq 0. \tag{7.20}$$

The system (7.19)-(7.20) defined by the joint dynamics of $\hat{X}$ and Γ is called the continuous time Kalman-Bucy filter.

7.5.3 The innovation process

Lemma 7.20. *The process $\mathfrak{I}$ is an $(\mathbb{P}, \mathbb{F}^Y)-$Brownian motion.*

Proof. The process $\mathfrak{I}$ is $\mathbb{F}^Y-$adapted, with a.s. continuous sample paths and $\mathfrak{I}_0 = 0$. It remains to prove that the increment $\mathfrak{I}_t - \mathfrak{I}_s$ is independent of $\mathcal{F}_s^Y$ and distributed as a centered gaussian with variance $t-s$, for all $0 \leq s \leq t$. To prove this, we shall now verify that $h(t) := \mathbb{E}\big[e^{iu(\mathfrak{I}_t - \mathfrak{I}_s)}\big|\mathcal{F}_s^Y\big] = e^{-\frac{1}{2}u^2(t-s)}$. Indeed, it follows by Itô's formula that

$$\begin{aligned} e^{iu(\mathfrak{I}_t - \mathfrak{I}_s)} &= 1 + iu\int_s^t e^{iu(\mathfrak{I}_r - \mathfrak{I}_s)}d\mathfrak{I}_r - \frac{1}{2}u^2\int_s^t e^{iu(\mathfrak{I}_r - \mathfrak{I}_s)}dr \\ &= 1 + iu\int_s^t e^{iu(\mathfrak{I}_r - \mathfrak{I}_s)}dW_r^Y \\ &\quad + iu\int_s^t e^{iu(\mathfrak{I}_r - \mathfrak{I}_s)}g(r)(X_r - \hat{X}_r)dr - \frac{1}{2}u^2\int_s^t e^{iu(\mathfrak{I}_r - \mathfrak{I}_s)}dt. \end{aligned}$$

Taking conditional expectations with respect to $\mathcal{F}_s^Y$, observing that the integrand in the stochastic integral with respect to W^Y has unit modulus, and using the tower property with $\mathcal{F}_s^Y \subset \mathcal{F}_s$, this provides:

$$\begin{aligned} h(t) &= 1 + iu\int_s^t \mathbb{E}\big[e^{iu(\mathfrak{I}_r - \mathfrak{I}_s)}g(r)\mathbb{E}\{(X_r - \hat{X}_r)|\mathcal{F}_r^Y\}\big]dr - \frac{1}{2}u^2\int_s^t h(r)dt \\ &= 1 + -\frac{1}{2}u^2\int_s^t h(r)dt. \end{aligned}$$

This requires the required result as the unique solution of the last ODE is given by $h(t) = e^{-\frac{1}{2}u^2(t-s)}$ for all $t \geq 0$. ♢

7.5.4 Dynamics of the best estimate

Lemma 7.21. *The process $\hat{X}$ is an Ornstein-Uhlenbeck process with dynamics* (7.19) *of Theorem 7.19.*

Proof. We proceed in several steps.
1. For all $n \in \mathbb{N}$, we introduce the $\mathbb{F}^Y-$stopping times $\tau_n := \inf\{t > 0 : |\hat{X}_t| \geq n\}$, so that the process $\hat{X}$ is bounded by n on $[0, \tau_n]$, and $\int_0^{t\wedge\tau_n} g(s)^2 \hat{X}_s^2 ds \leq n^2 \int_0^t g(s)^2 ds < \infty$. We may then introduce a probability measure $\mathbb{Q}^n$ equivalent to $\mathbb{P}$ on $[0, \tau_n]$, with density process

$$Z_t^n = \frac{d\mathbb{Q}^n}{d\mathbb{P}}\Big|_{\mathcal{F}_t^Y} := e^{-\int_0^{t\wedge\tau_n} g(s)\hat{X}_s d\mathfrak{I}_s - \frac{1}{2}\int_0^{t\wedge\tau_n} g(s)^2 \hat{X}_s^2 ds}, \quad t \geq 0.$$

By direct application of the Girsanov theorem, we see that the process

$$Y^n_{\cdot\wedge\tau_n} := \mathfrak{I}_{\wedge\tau_n} + \int_0^{\cdot\wedge\tau_n} g(s)\hat{X}_s ds, \text{ is a Brownian motion, with } Y_t^n = Y_t \text{ for } t \leq \tau_n.$$

2. Denote $\hat{W}_t^X := \mathbb{E}\big[W_t^X | \mathcal{F}_t^Y\big]$, $t \geq 0$. In this step, we show that $\hat{W}^X$ is an $(\mathbb{P}, \mathbb{F}^Y)-$martingale, and that $M^n := \big(\frac{\hat{W}^X}{Z^n}\big)_{\cdot\wedge\tau_n}$ is an $(\mathbb{Q}^n, \mathbb{F}^Y)-$martingale.

First $\hat{W}^X$ is an integrable process as the conditional expectation of the integrable process W. Next, for $0 \leq s \leq t$, we use the tower property, the inclusion $\mathcal{F}_t^Y \subset \mathcal{F}_t$, and the $(\mathbb{P}, \mathbb{F})-$martingale property of W^X, to compute that

$$\begin{aligned}
\mathbb{E}\big[\hat{W}_t^X \big| \mathcal{F}_s^Y\big] = \mathbb{E}\big[\mathbb{E}\big\{\hat{W}_t^X \big| \mathcal{F}_t^Y\big\} \big| \mathcal{F}_s^Y\big] &= \mathbb{E}\big[W_t^X \big| \mathcal{F}_s^Y\big] \\
&= \mathbb{E}\big[\mathbb{E}\big\{W_t^X \big| \mathcal{F}_s\big\} \big| \mathcal{F}_s^Y\big] = \mathbb{E}\big[W_s^X \big| \mathcal{F}_s^Y\big] = \hat{W}_s^X.
\end{aligned}$$

This shows that $\hat{W}^X$ is a $(\mathbb{P}, \mathbb{F}^Y)-$martingale.

We next verify that M^n defines a $(\mathbb{Q}^n, \mathbb{F}^Y)-$martingale on $[0, \tau_n]$. Indeed, $\mathbb{E}^{\mathbb{Q}^n}\big[|M^n_{t\wedge\tau_n}|\big] = \mathbb{E}\big[|W^X_{t\wedge\tau_n}|\big] < \infty$, and for all $0 \leq s \leq t$, it follows from the Bayes rule together with the $(\mathbb{P}, \mathbb{F}^Y)-$martingale property of W^X and $Z^n_{\cdot\wedge\tau_n}$ that:

$$\mathbb{E}^{\mathbb{Q}^n}\big[M^n_{t\wedge\tau_n} \big| \mathcal{F}_s^Y\big] \;=\; \frac{\mathbb{E}\big[W^X_{t\wedge\tau_n} \big| \mathcal{F}_s^Y\big]}{\mathbb{E}\big[Z^n_{t\wedge\tau_n} \big| \mathcal{F}_s^Y\big]} \;=\; \frac{W^X_{s\wedge\tau_n}}{Z^n_{s\wedge\tau_n}} \;=\; M^n_{s\wedge\tau_n}.$$

3. As $\mathbb{F}^Y$ is the canonical filtration of the $\mathbb{Q}^n-$Brownian motion Y, and $M^n := \frac{\hat{W}}{Z^n}$ is a $(\mathbb{Q}^n, \mathbb{F}^Y)-$martingale on $[0, \tau_n]$ starting from zero, it follows from the martingale representation theorem that $M^n_{t\wedge\tau_n} = \int_0^{t\wedge\tau_n} H_s^n dY_s$, $t \geq 0$, for some $\mathbb{F}^Y-$adapted process $H^n \in \mathbb{H}^2$. We now show that, for all $n \in \mathbb{N}$, we may find a process $\lambda^n \in \mathbb{H}^2$ such that

$$\hat{W}^X_{t\wedge\tau_n} = \int_0^{t\wedge\tau_n} \lambda_s^n d\mathfrak{I}_s, \; t \geq 0, \; \mathbb{P} - a.s. \quad \text{and} \quad \lambda^n \text{ is } \mathbb{F}^Y-\text{adapted.} \quad (7.21)$$

Indeed, as Z^n satisfies the dynamics $dZ^n_t = -Z^n_t g(t)\hat{X}_t d\Im_t$ for $t \le \tau_n$, it follows from Itô's formula that

$$\begin{aligned}
d\hat{W}^X_t = d(M^n_t Z^n_t) &= M^n_t dZ^n_t + Z^n_t dM^n_t - Z^n_t g(t)\hat{X}_t H^n_t dt \\
&= -M^n_t Z^n_t g(t)\hat{X}_t d\Im_t + Z^n_t H^n_t dY_t - Z^n_t g(t)\hat{X}_t H^n_t dt \\
&= -\hat{W}^X_t g(t)\hat{X}_t d\Im_t + Z^n_t H^n_t\big(g(t)\hat{X}_t dt + d\Im_t\big) - Z^n_t g(t)\hat{X}_t H^n_t dt \\
&= \big(Z^n_t H^n_t - g(t)\hat{W}^X_t \hat{X}_t\big)d\Im_t.
\end{aligned}$$

The required result follows by setting $\lambda^n := Z^n H^n - g\hat{W}^X\hat{X}$ which inherits the adaptability to $\mathbb{F}^Y$ from that of the process ϕ.

4. For an arbitrary $\mathbb{F}^Y-$progressively measurable bounded process ϕ, set $V_t := \int_0^t \phi_s d\Im_s$. By direct application of Itô's formula to the product XV, we see that for $t \ge 0$:

$$X_t V_t - \int_0^t f(s)X_s V_s ds = \int_0^t \phi_s g(s) X_s\big(X_s - \hat{X}_s\big)ds + \int_0^t V_s dW^X_s + \int_0^t \phi_s X_s dW^Y_s.$$

We now show that for all $t \ge 0$ and $n \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{E}\Big[\int_0^{t\wedge\tau_n} g(s)\phi_s X_s\big(X_s - \hat{X}_s\big)ds\Big] &= \mathbb{E}\Big[\hat{W}^X_{t\wedge\tau_n} V_{t\wedge\tau_n}\Big] \qquad (7.22)\\
&= \mathbb{E}\Big[\Big(\hat{X}_{t\wedge\tau_n} - \int_0^{t\wedge\tau_n} f(s)\hat{X}_s ds\Big)V_{t\wedge\tau_n}\Big].
\end{aligned}$$

Clearly, $X \in \mathbb{H}^2$, and by the boundedness of ϕ, we have $V \in \mathbb{H}^2$ on $[0,t]$ by the the Itô isometry. Then, by taking expectation on both hands of the last equality, we get:

$$\begin{aligned}
\mathbb{E}\Big[\int_0^{t\wedge\tau_n} g(s)\phi_s X_s\big(X_s - \hat{X}_s\big)ds\Big] &= \mathbb{E}\Big[X_{t\wedge\tau_n} V_{t\wedge\tau_n} - \int_0^{t\wedge\tau_n} f(s)X_s V_s ds\Big] \qquad (7.23)\\
&= \mathbb{E}\Big[V_{t\wedge\tau_n}\Big(X_{t\wedge\tau_n} - \int_0^{t\wedge\tau_n} f(s)X_s ds\Big)\Big] \\
&= \mathbb{E}\Big[V_{t\wedge\tau_n} W^X_{t\wedge\tau_n}\Big] \\
&= \mathbb{E}\big[V_{t\wedge\tau_n}\mathbb{E}\{W^X_{t\wedge\tau_n} | \mathcal{F}^Y_t\}\big] \\
&= \mathbb{E}\Big[V_{t\wedge\tau_n}\hat{W}^X_{t\wedge\tau_n}\Big],
\end{aligned}$$

where the second equality follows from the martingale property of V (as a stochastic integral with respect to the Brownian motion $\Im$, with bounded integrand), and the fourth inequality uses the inclusion $\mathcal{F}^Y_t \subset \mathcal{F}_0$ together with the fact that V is $\mathbb{F}^Y-$adapted.

To prove the second equality, we return to (7.23), and observe that the the right hand side may be rearranged by using the tower property combined with the fact that V is $\mathbb{F}^Y-$adapted and τ_n is an $\mathbb{F}^Y-$stopping time:

$$\begin{aligned}
& \mathbb{E}\Big[\mathbb{E}\{X_{t\wedge\tau_n}|\mathcal{F}^Y_{t\wedge\tau_n}\}V_{t\wedge\tau_n} - \int_0^{t\wedge\tau_n} f(s)\mathbb{E}\{X_s|\mathcal{F}^Y_s\}V_s ds\Big] \\
= \;& \mathbb{E}\Big[\hat{X}_{t\wedge\tau_n}V_{t\wedge\tau_n} - \int_0^{t\wedge\tau_n} f(s)\hat{X}_s V_s ds\Big] \\
= \;& \mathbb{E}\Big[\Big(\hat{X}_{t\wedge\tau_n} - \int_0^{t\wedge\tau_n} f(s)\hat{X}_s ds\Big)V_{t\wedge\tau_n}\Big];
\end{aligned}$$

the last equality holds as V is a martingale.
5. As $\mathbb{E}[X_s|\mathcal{F}^Y_s] = \hat{X}_s$, we have $\mathbb{E}\big[X_s(X_s-\hat{X}_s)\big|\mathcal{F}^Y_s\big] = \mathbb{E}\big[(X_s-\hat{X}_s)^2\big|\mathcal{F}^Y_s\big]$. By the independence of $X_s - \hat{X}_s$ and $\mathbb{L}^2(\mathcal{F}^Y_s)$, we see that

$$\mathbb{E}\big[X_s(X_s-\hat{X}_s)\big|\mathcal{F}^Y_s\big] \;=\; \mathbb{E}\big[(X_s-\hat{X}_s)^2\big] \;=:\; \Gamma(s),$$

as defined in (7.17). Substituting in the equality of the first equality in (7.22), and using the tower property together with the $\mathbb{F}^Y-$adaptability of the process V, we get

$$\mathbb{E}\Big[V_{t\wedge\tau_n}\hat{W}^X_{t\wedge\tau_n}\Big] = \mathbb{E}\Big[\int_0^{t\wedge\tau_n} g(s)\phi_s\mathbb{E}\big\{X_s(X_s-\hat{X}_s)\big|\mathcal{F}^Y_s\big\}ds\Big] = \mathbb{E}\Big[\int_0^{t\wedge\tau_n} g(s)\phi_s\Gamma(s)ds\Big].$$

6. As $V_t = \int_0^t \phi_s dW_s$, and $\hat{W}^X_{t\wedge\tau_n} = \int_0^{t\wedge\tau_n}\lambda^n_s d\mathfrak{I}_s$ by (7.21), it follows from Lemma 7.20 together with the Itô isometry that $\mathbb{E}\big[V_{t\wedge\tau_n}\hat{W}^X_{t\wedge\tau_n}\big] = \mathbb{E}\big[\int_0^{t\wedge\tau_n}\phi_s\lambda^n_s ds\big]$. Then, we obtain from the last equality of the last step that

$$\mathbb{E}\Big[\int_0^{t\wedge\tau_n}\phi_s\big(\lambda^n_s - g(s)\Gamma(s)\big)ds\Big] = 0, \quad t \geq 0.$$

Since ϕ is an arbitrary bounded $\mathbb{F}^Y-$adapted process, this implies that $\lambda^n_s = g(s)\Gamma(s)$, $ds\otimes d\mathbb{P}-$a.e. on $[0,\tau_n]$. In particular, λ^n is independent of n, and the representation of $\hat{W}^X$ in (7.21) holds on $\mathbb{R}_+$, i.e.

$$\hat{W}^X_t = \int_0^t g(s)\Gamma(s)d\mathfrak{I}_s, \quad t \geq 0.$$

On the other hand, by sending $n \nearrow \infty$ in the second equality of (7.22), we obtain by the dominated convergence theorem that:

$$\mathbb{E}\Big[V_t\Big(\hat{X}_t - \int_0^t f(s)\hat{X}_s ds - \hat{W}^X_t\Big)\Big] = 0, \quad \text{for all} \quad t \geq 0.$$

By the martingale representation theorem, we see by standard approximation that the set $\{c + V_t : \; c \in \mathbb{R},\; \phi \text{ bounded process in } \mathbb{H}^2\}$ is dense in $\mathbb{L}^2(\mathcal{F}^Y_t)$. Then, it follows from the arbitrariness of ϕ that $t \longmapsto \hat{X}_t - \int_0^t f(s)\hat{X}_s ds - \hat{W}^X_t$ is a (deterministic) constant, which is necessarily zero by considering its value at time zero. Hence

$$\hat{X}_t = \int_0^t f(s)\hat{X}_s ds + \hat{W}^X_t = \int_0^t f(s)\hat{X}_s ds + \int_0^t g(s)\Gamma(s)d\mathfrak{I}_s$$

by the previous identification of the process λ^n. ♢

7.5.5 ODE characterization of the variance

We finally derive the characterization of the function Γ.

Lemma 7.22. *The function* Γ *is the unique solution of the the ODE* (7.20) *of Theorem 7.19.*

Proof. By direct calculation, we see that the error estimate $X-\hat{X}$ is an Ornstein-Uhlenbeck process with dynamics

$$d\big(X_t - \hat{X}_t\big) \quad = \quad \big(f(t) - g(t)^2\Gamma(t)\big)(X_t - \hat{X}_t)dt + dW_t^X - g(t)\Gamma(t)dW_t^Y.$$

Then, $X - \hat{X}$ is a centered Gaussian process with variance $\Gamma(t)$. In order to show that Γ solves the ODE of Lemma 7.22, we may use Itô's formula to obtain the dynamics of $(X - \hat{X})^2$ and then take expectations while being careful with the local martingale components...

Alternatively, we may compute explicitly, denoting $\gamma := f - g^2\Gamma$, and $\delta_t := X - \hat{X}$, that

$$d\Big(e^{-\int_0^t \gamma(s)ds}\delta_t\Big) \quad = \quad e^{-\int_0^t \gamma(s)ds}\Big(dW_t^X - g(t)\Gamma(t)dW_t^Y\Big).$$

As $\delta_0 = 0$, this provides

$$\delta_t = \int_0^t e^{\int_s^t \gamma(r)dr}\Big(dW_s^X - g(s)\Gamma(s)dW_s^Y\Big) \quad \text{with distribution} \quad \mathcal{N}\big(0, \Gamma(t)\big),$$

which implies that $\Gamma(t) = \mathbb{E}[\delta_t^2]$, and by the Itô isometry:

$$\Gamma(t) \quad = \quad \int_0^t e^{2\int_s^t \gamma(r)dr}\Big(1 + g(s)^2\Gamma(s)^2\Big)ds$$

Differentiating both sides with respect to t, we see that

$$\Gamma'(t) = 2\gamma\Gamma + 1 + g^2\Gamma^2 = 2(f - g^2\Gamma)\Gamma + 1 + g^2\Gamma^2 = 2f\Gamma - g^2\Gamma^2 + 1,$$

which provides the required expression, given that $\Gamma(0) = 0$.

We finally justify that the ODE of Lemma (7.22) has a unique solution in $\mathbb{R}_+$. Notice that any solution of this ODE lies in $[0, M]$ where M is the unique (explicit) solution of the ODE $M(t) = \int_0^t (2f(s)M(s) + 1)ds$. We next solve locally the ODE of Lemma (7.22) by the Cauchy Lipschitz theorem, and we prove by standard arguments that the explosion time can not be finite, thanks to the bounds 0 and M. ♢

Chapter 8

Stochastic differential equations

8.1 First examples

In the previous sections, we have handled the geometric Brownian motion defined by

$$S_t \quad := \quad S_0 \exp\left[(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\right], \quad t \geq 0, \tag{8.1}$$

where W is a scalar Brownian motion, and $X_0 > 0, \mu, \sigma \in \mathbb{R}$ are given constants. An immediate application of Itô's formula shows that X satisfies the dynamics

$$dS_t \quad = \quad \mu S_t dt + \sigma S_t dW_t, \quad t \geq 0. \tag{8.2}$$

This is our first example of stochastic differential equations since S appears on both sides of the equation. Of course, the geometric Brownian motion (8.1) is a solution of the stochastic differential equation. A natural question is whether this solution is unique. In this simple model, the answer to this squestion is easy:

- Since $S_0 > 0$, and any solution S of (8.2) has a.s. continuous sample paths, as a consequence of the contituity of the stochastic integral $t \longmapsto \int_0^t \sigma S_s dW_s$, we see that S hits 0 before any negative real number.
- Let $\theta := \inf\{t : \ S_t = 0\}$ be the first hitting time of 0, and set $L_t := \ln S_t$ for $t < \theta$; then, it follows from Itô's formula that

  $$dL_t \quad = \quad \big(\mu - \frac{1}{2}\sigma^2\big)dt + \sigma dW_t, \quad t < \theta,$$

  which leads uniquely to $L_t = L_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$, $t < \theta$, which correspond exactly to the solution (8.1). In particular $\theta = +\infty$ a.s. and the above arguments holds for any $t \geq 0$.

Our next example of stochastic differential equations is the so-called Ornstein-Uhlenbeck process, which is widely used for the modelling of the term structure of interest rates:

$$dX_t = k(m - X_t)dt + \sigma dW_t, \quad t \geq 0. \tag{8.3}$$

where $k, m, \sigma \in \mathbb{R}$ are given constants. From the modelling viewpoint, the motivation is essentially in the case $k > 0$ which induces the so-called mean reversion: when $X_t > m$, the drift points downwards, while for $X_t < m$ the drift coefficient pushes the solution upwards, hence the solution (if exists !) is attracted to the mean level m. Again, the issue in (8.3) is that X appears on both sides of the equation.

This example can also be handled explicitly by using the analogy with the deterministic case (corresponding to $\sigma = 0$) which suggests the change of variable $Y_t := e^{kt}X_t$. Then, it follows from Itô's formula that

$$dY_t = mke^{kt}dt + \sigma e^{kt}dW_t, \quad t \geq 0,$$

and we obtain as a unique solution

$$Y_t = Y_0 + m\left(e^{kt} - 1\right) + \sigma \int_0^t e^{ks}dW_s, \quad t \geq 0,$$

or, back to X:

$$X_t = X_0 e^{-kt} + m\left(1 - e^{-kt}\right) + \sigma \int_0^t e^{-k(t-s)}dW_s, \quad t \geq 0.$$

The above two examples are solved by a (lucky) specific change of variable. For more general stochastic differential equations, it is clear that, as in the deterministic framework, a systematic analysis of the existence and uniqueness issues is needed, without any access to a specific change of variable. In this section, we show that existence and uniqueness hold true under general Lipschitz conditions, which of course reminds the situation in the deterministic case. More will be obtained outside the Lipschitz world in the one-dimensional case.

Finally, let us observe that the above solutions S and X can be expressed starting from an initial condition at time t as:

$$\begin{aligned} S_u &= S_t \exp{(\mu - \frac{1}{2}\sigma^2)(u-t) + \sigma(W_u - W_t)} \\ X_u &= X_t e^{-k(u-t)} + m\left(1 - e^{-k(u-t)}\right) + \sigma \int_t^u e^{-k(u-s)}dW_s, \end{aligned}$$

for all $u \geq t \leq 0$. In particular, we see that:

- the distribution of S_u conditional on the past values $\{S_s, s \leq t\}$ of S up to time t equals to the distribution of S_u conditional on the current value S_t at time t; this is the so-called Markov property,

- let $S_u^{t,s}$ and $X_u^{t,x}$ be given by the above expressions with initial condition at time t frozen to $S_t = s$ and $X_t = x$, respectiveley, then the random functions $s \longmapsto S_u^{t,s}$ and $x \longmapsto X_u^{t,x}$ are strictly increasing for all u; this is the so-called increase of the flow property.

These results which are well known for deterministic differential equations will be shown to hold true in the more general stochastic framework.

8.2 Strong solution of a stochastic differential equation

8.2.1 Existence and uniqueness

Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := \{\mathcal{F}_t\}_t, \mathbb{P})$ supporting a $d-$dimensional Brownian motion W, we consider the stochastic differential equation

$$dX_t \quad = \quad b(t, X_t)dt + \sigma(t, X_t)dW_t, \;\; t \in [0, T], \tag{8.4}$$

for some $T \in \mathbb{R}$. Here, b and σ are function defined on $[0, T] \times \mathbb{R}^n$ taking values respectively in $\mathbb{R}^n$ and $\mathcal{M}_{\mathbb{R}}(n, d)$.

Definition 8.1. *A strong solution of* (8.4) *is an* $\mathbb{F}-$*adapted process* X *such that* $\int_0^T (|b(t, X_t)| + |\sigma(t, X_t)|^2)dt < \infty$, *a.s. and*

$$X_t \quad = \quad X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad t \in [0, T].$$

Let us mention that there is a notion of weak solutions which relaxes some conditions from the above definition in order to allow for more general stochastic differential equations. Weak solutions, as opposed to strong solutions, are defined on some probabillistic structure (which becomes part of the solution), and not necessarilly on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W)$. Thus, for a weak solution we search for a probability structure $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{W})$ and a process $\tilde{X}$ such that the requirement of the above definition holds true. Obviously, any strong solution is a weak solution, but the opposite claim is false.

Clearly, one should not expect that the stochastic differential equation (8.4) has a unique solution without any condition on the coefficients b and σ. In the deterministic case $\sigma \equiv 0$, (8.4) reduces to an ordinary differential equation for which existence and uniqueness requires Lipschitz conditions on b. The following is an example of non-uniqueness.

Exercise 8.2. *Consider the stochastic differential equation:*

$$dX_t \quad = \quad 3X_t^{1/3}dt + 3X_t^{2/3}dW_t$$

with initial condition $X_0 = 0$. *Show that* $X_t = W_t^3$ *is a solution (in addition to the solution* $X = 0$*).*

Our main existence and uniqueness result is the following.

Theorem 8.3. *Let $X_0 \in \mathbb{L}^2$ be a r.v. independent of W, and assume that the functions $|b(t,0)|$, $|\sigma(t,0)| \in \mathbb{L}^2(\mathbb{R}_+)$, and that for some $K>0$:*

$$|b(t,x)-b(t,y)|+|\sigma(t,x)-\sigma(t,y)| \le K|x-y| \quad \text{for all} \quad t\in[0,T],\ x,y\in\mathbb{R}^n.$$

Then, for all $T>0$, there exists a unique strong solution of (8.4) *in* $\mathbb{H}^2$. *Moreover,*

$$\mathbb{E}\left[\sup_{t\le T}|X_t|^2\right] \quad \le \quad C\left(1+\mathbb{E}|X_0|^2\right)e^{CT}, \tag{8.5}$$

for some constant $C=C(T,K)$ depending on T and K.

Proof. We first establish the existence and uniqueness result, then we prove the estimate (8.5).
<u>Step 1</u> For a constant $c>0$, to be fixed later, we introduce the norm

$$\|\phi\|_{\mathbb{H}^2_c} := \mathbb{E}\left[\int_0^T e^{-ct}|\phi_t|^2dt\right]^{1/2} \qquad \text{for every} \quad \phi\in\mathbb{H}^2.$$

Clearly $e^{-cT}\|\phi\|_{\mathbb{H}^2} \le \|\phi\|_{\mathbb{H}^2_c} \le \|\phi\|_{\mathbb{H}^2}$. So the norm $\|.\|_{\mathbb{H}^2_c}$ is equivalent to the standard norm $\|.\|_{\mathbb{H}^2}$ on the Hilbert space $\mathbb{H}^2$.

We define a map U on $\mathbb{H}^2([0,T]\times\Omega)$ by:

$$U(X)_t \ := \ X_0+\int_0^t b(s,X_s)ds+\int_0^t \sigma(s,X_s)dW_s, \qquad 0\le t\le T.$$

This map is well defined, as the processes $\{b(t,X_t),\sigma(t,X_t), t\in[0,T]\}$ are immediately checked to be in $\mathbb{H}^2$. In order to prove existence and uniqueness of a solution for (8.4), we shall prove that $U(X)\in\mathbb{H}^2$ for all $X\in\mathbb{H}^2$ and that U is a contracting mapping with respect to the norm $\|.\|_{\mathbb{H}^2_c}$ for a convenient choice of the constant $c>0$.
1- We first prove that $U(X)\in\mathbb{H}^2$ for all $X\in\mathbb{H}^2$. To see this, we decompose:

$$\begin{aligned}\|U(X)\|^2_{\mathbb{H}^2} \quad &\le \quad 3T\|X_0\|^2_{\mathbb{L}^2}+3\mathbb{E}\left[\int_0^T\left|\int_0^t b(s,X_s)ds\right|^2dt\right]\\ &\qquad +3\mathbb{E}\left[\int_0^T\left|\int_0^t \sigma(s,X_s)dW_s\right|^2dt\right]\end{aligned}$$

By the Lipschitz-continuity of b and σ in x, uniformly in t, we have $|b(t,x)|^2+|\sigma(t,x)|^2\le K(1+|b(t,0)|^2+|x|^2)$ for some constant K. We then estimate the second term by:

$$\mathbb{E}\left[\int_0^T\left|\int_0^t b(s,X_s)ds\right|^2dt\right] \quad \le \quad KT\mathbb{E}\left[\int_0^T(1+|b(t,0)|^2+|X_s|^2)ds\right]<\infty,$$

since $X \in \mathbb{H}^2$, and $b(.,0) \in \mathbb{L}^2([0,T])$.

As, for the third term, we first use the Itô isometry:

$$\begin{aligned}
\mathbb{E}\left[\int_0^T \left|\int_0^t \sigma(s,X_s)dW_s\right|^2 dt\right] &\leq T\mathbb{E}\left[\int_0^T |\sigma(s,X_s)|^2 ds\right] \\
&\leq TK\mathbb{E}\left[\int_0^T (1+|\sigma(t,0)|^2+|X_s|^2)ds\right] < \infty.
\end{aligned}$$

2- We next show that U is a contracting mapping for the norm $\|.\|_{\mathbb{H}^2_c}$ for some convenient choice of $c > 0$. For $X, Y \in \mathbb{H}^2$ with $X_0 = Y_0 = 0$, we have

$$\begin{aligned}
&\mathbb{E}\left|U(X)_t - U(Y)_t\right|^2 \\
\leq\;& 2\mathbb{E}\left|\int_0^t (b(s,X_s)-b(s,Y_s))\,ds\right|^2 + 2\mathbb{E}\left|\int_0^t (\sigma(s,X_s)-\sigma(s,Y_s))\,dW_s\right|^2 \\
=\;& 2\mathbb{E}\left|\int_0^t (b(s,X_s)-b(s,Y_s))\,ds\right|^2 + 2\mathbb{E}\int_0^t \left|\sigma(s,X_s)-\sigma(s,Y_s)\right|^2 ds \\
\leq\;& 2t\mathbb{E}\int_0^t \left|b(s,X_s)-b(s,Y_s)\right|^2 ds + 2\mathbb{E}\int_0^t \left|\sigma(s,X_s)-\sigma(s,Y_s)\right|^2 ds \\
\leq\;& 2(T+1)K \int_0^t \mathbb{E}\left|X_s - Y_s\right|^2 ds.
\end{aligned}$$

Then,

$$\begin{aligned}
\|U(X)-U(Y)\|_{\mathbb{H}^2_c} &\leq 2K(T+1)\int_0^T e^{-ct}\int_0^t \mathbb{E}|X_s-Y_s|^2 ds\,dt \\
&= \frac{2K(T+1)}{c}\int_0^T e^{-cs}\mathbb{E}|X_s-Y_s|^2(1-e^{-c(T-s)})ds \\
&\leq \frac{2K(T+1)}{c}\|X-Y\|_{\mathbb{H}^2_c}.
\end{aligned}$$

Hence, U is a contracting mapping for sufficiently large $c > 1$.

Step 2 We next prove the estimate (8.5). We shall alleviate the notation writing $b_s := b(s,X_s)$ and $\sigma_s := \sigma(s,X_s)$. We directly estimate:

$$\begin{aligned}
\mathbb{E}\left[\sup_{u\leq t}|X_u|^2\right] &= \mathbb{E}\left[\sup_{u\leq t}\left|X_0+\int_0^u b_s ds + \int_0^u \sigma_s dW_s\right|^2\right] \\
&\leq 3\left(\mathbb{E}|X_0|^2 + t\mathbb{E}\left[\int_0^t |b_s|^2 ds\right] + \mathbb{E}\left[\sup_{u\leq t}\left|\int_0^u \sigma_s dW_s\right|^2\right]\right) \\
&\leq 3\left(\mathbb{E}|X_0|^2 + t\mathbb{E}\left[\int_0^t |b_s|^2 ds\right] + 4\mathbb{E}\left[\int_0^t |\sigma_s|^2 ds\right]\right)
\end{aligned}$$

where we used the Doob's maximal inequality of Proposition 3.15. Since b and

σ are Lipschitz-continuous in x, uniformly in t, this provides:

$$\mathbb{E}\left[\sup_{u\le t}|X_u|^2\right] \quad\le\quad C(K,T)\left(1+\mathbb{E}|X_0|^2+\int_0^t \mathbb{E}\left[\sup_{u\le s}|X_u|^2\right]ds\right)$$

and we conclude by using the Gronwall lemma. $\diamondsuit$

The following exercise shows that the Lipschitz-continuity condition on the coefficients b and σ can be relaxed. Further relaxation of this assumption is possible in the one-dimensional case, see Section 8.3.

Exercise 8.4. *In the context of this section, assume that the coefficients μ and σ are locally Lipschitz with linear growth. By a localization argument, prove that strong existence and uniqueness holds for the stochastic differential equation (8.4).*

8.2.2 The Markov property

Let $X^{t,x}_{\cdot}$ denote the solution of the stochastic differential equation

$$X_s = x+\int_t^s b(u,X_u)du+\int_t^s \sigma(u,X_u)dW_u \quad s\ge t$$

The two following properties are obvious:

- Clearly, $X^{t,x}_s = F\left(t,x,s,(W_\cdot - W_t)_{t\le u\le s}\right)$ for some deterministic function F.
- For $t\le u\le s$: $X^{t,x}_s = X^{u,X^{t,x}_u}_s$. This follows from the pathwise uniqueness, and holds also when u is a stopping time.

With these observations, we have the following Markov property for the solutions of stochastic differential equations.

Proposition 8.5. *(Markov property) For all $0\le t\le s$:*

$$\mathbb{E}\left[\Phi\left(X_u, t\le u\le s\right)|\mathcal{F}_t\right] \quad=\quad \mathbb{E}\left[\Phi\left(X_u,t\le u\le s\right)|X_t\right]$$

for all bounded function $\Phi:\ C[t,s]\longrightarrow\mathbb{R}$.

8.3 More results for scalar stochastic differential equations

We first start by proving a uniqueness result for scalar stochastic differential equation under weaker conditions than the general $n-$dimensional result of Theorem 8.3. For example, the following extension allows to consider the so-called Cox-Ingersol-Ross square root model for interest rates:

$$dr_t \quad=\quad k(b-r_t)dt+\sigma\sqrt{r_t}dW_t,$$

or the *Stochastic Volatility Constant Elasticity of Variance* (SV-CEV) models which are widely used to account for the dynamics of implied volatility, see Chapter 10,

$$\frac{dS_t}{S_t} = \mu dt + S_t^\alpha \sigma_t dW_t,$$

where the process $(\sigma_t)_{t\geq 0}$ is generated by another autonomous stochastic differential equation.

The question of existence will be skipped in these notes as it requires to develop the theory of weak solutions of stochastic differential equations, which we would like to avoid. Let us just mention that the main result, from Yamada and Watanabe (1971), is that *the existence of weak solutions for a stochastic differential equation for which strong uniqueness holds implies existence and uniqueness of a strong solution.*

Theorem 8.6. *Let $b, \sigma : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ be two functions satisfying for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$:*

$$|\mu(t,x) - \mu(t,y)| \leq K|x-y| \quad \text{and} \quad |\sigma(t,x) - \sigma(t,y)| \leq h(|x-y|), \tag{8.6}$$

where $h : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is strictly increasing, $h(0) = 0$, and $\int_{(0,\varepsilon)} h(u)^{-2}du = \infty$ for all $\varepsilon > 0$. Then, there exists at most one strong solution for the stochastic differential equation

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad t \geq 0.$$

Proof. The conditions imposed on the function h imply the existence of a strictly decreasing sequence $(a_n)_{n\geq 0} \subset (0,1]$, with $a_0 = 1$, $a_n \longrightarrow 0$, and $\int_{a_n}^{a_{n-1}} h(u)^{-2}du = n$ for all $n \geq 1$. Then, for all $n \geq 0$, there exists a continuous function $\rho_n : \mathbb{R} \longrightarrow \mathbb{R}$ such that

$$h = 0 \text{ outside of } [a_n, a_{n-1}], \quad 0 \leq \rho_n \leq \tfrac{2}{nh^2} \quad \text{and} \quad \int_{a_n}^{a_{n-1}} \rho_n(x)dx = 1.$$

We now introduce the C^2-function $\psi_n(x) := \int_0^{|x|} \int_0^y \rho_n(u)dudy$, $x \in \mathbb{R}$, and we observe that

$$|\psi_n'| \leq 1, \ |\psi_n''| \leq \frac{2}{nh^2}\mathbf{1}_{[a_n,a_{n-1}]}, \quad \text{and} \quad \psi_n(x) \longrightarrow |x|, \ n \to \infty. \tag{8.7}$$

Let X and Y be two solutions of (8.6) with $X_0 = Y_0$, define the stopping time $\tau_n := \inf\{t > 0 : (\int_0^t \sigma(s,X_s)^2 ds) \vee (\int_0^t \sigma(s,Y_s)^2 ds) \geq n\}$, and set $\delta_t := X_t - Y_t$. Then, it follows from Itô's formula that:

$$\begin{aligned}
\psi_n\left(\delta_{t\wedge\tau_n}\right) &= \int_0^{t\wedge\tau_n} \psi_n'(\delta_s)\left(\mu(s,X_s) - \mu(s,Y_s)\right)ds \\
&\quad + \frac{1}{2}\int_0^{t\wedge\tau_n} \psi_n''(\delta_s)\left(\sigma(s,X_s) - \sigma(s,Y_s)\right)^2 ds \\
&\quad + \int_0^{t\wedge\tau_n} \psi_n'(\delta_s)\left(\sigma(s,X_s) - \sigma(s,Y_s)\right)dW_s.
\end{aligned} \tag{8.8}$$

Then, by the definition of τ_n and the boundedness of ψ_n', the stochastic integral term has zero expectation. Then, using the Lipschitz-continuity of μ and the definition of the function h:

$$\begin{aligned}
\mathbb{E}[\psi_n\left(\delta_{t\wedge\tau_n}\right)] &\leq K\mathbb{E}\left[\int_0^{t\wedge\tau_n}|\delta_s|ds\right]+\mathbb{E}\left[\int_0^{t\wedge\tau_n}\psi_n''(\delta_s)h^2(|\delta_s|)ds\right]\\
&\leq K\mathbb{E}\left[\int_0^{t}|\delta_s|ds\right]+\mathbb{E}\left[\int_0^{t}\psi_n''(\delta_s)h^2(|\delta_s|)ds\right]\\
&\leq K\int_0^t\mathbb{E}[|\delta_s|]ds+\frac{t}{n}.
\end{aligned}$$

By sending $n\to\infty$, we see that $\mathbb{E}[|\delta_t|]\leq K\int_0^t\mathbb{E}[|\delta_s|]ds+\frac{t}{n}$, $t\geq 0$, and we deduce from the Gronwall inequality that $\delta_t=0$ for all $t\geq 0$. ♢

Remark 8.7. It is well-known that the deterministic differential equations $X_t = X_0+\int_0^t b(s,X_s)$, has a unique solution for sufficiently small $t>0$ when b is locally Lipschitz in x uniformly in t, and bounded on compact subsets of $\mathbb{R}^+\times\mathbb{R}$. In the absence of these conditions, we may go into problems of existence and uniqueness. For example, for $\alpha\in(0,1)$, the equation $X_t=\int_0^t|X_s|^\alpha ds$ has a continuum of solutions $X^\theta,\theta\geq 0$, defined by:

$$X_t^\theta := [(1-\alpha)(t-\theta)]^{1/(1-\alpha)}\mathbf{1}_{[\theta,\infty)}(t), \qquad t\geq 0.$$

The situation in the case of stochastic differential equations is different, as the previous theorem 8.6 shows that strong uniqueness holds for the stochastic differential equation $X_t=\int_0^t|X_s|^\alpha dW_s$ when $\alpha\geq 1/2$, and therefore $X=0$ is the unique strong solution.

We next use the methodology of proof of the previous theorem 8.6 to prove a monotonicity of the solution of a scalar stochastic differential equation with respect to the drift coefficient and the initial condition.

Proposition 8.8. *Let X and Y be two $\mathbb{F}-$adapted processes with continuous sample paths, satisfying for $t\in\mathbb{R}_+$:*

$$\begin{aligned}
X_t &= X_0+\int_0^t\mu(t,X_t)dt+\int_0^t\sigma(t,X_t)dW_t,\\
Y_t &= Y_0+\int_0^t\eta(t,X_t)dt+\int_0^t\sigma(t,X_t)dW_t.
\end{aligned}$$

for some continuous functions μ, η, and σ. Assume further that either μ or η is Lipschitz-continuous, and that σ satisfies Condition (8.6) *from theorem 8.6. Then*

$$\mu(.)\leq\eta(.) \text{ and } X_0\leq Y_0 \text{ a.s.} \quad \text{implies that} \quad X_.\leq Y_. \text{ a.s.}$$

Proof. Define $\varphi_n(x) := \psi_n(x)\mathbf{1}_{(0,\infty)}(x)$, where ψ_n is as defined in the proof of Theorem 8.6. With $\delta_t := X_t-Y_t$, and $\tau_n := \inf\{t>0 : (\int_0^t\sigma(s,X_s)^2ds)\vee$

$(\int_0^t \sigma(s,Y_s)^2 ds) \geq n\}$, we apply Itô's formula, and we deduce from the analogue of (8.8), with φ_n replacing ψ_n, that

$$\mathbb{E}[\varphi_n(\delta_t)] - \frac{t}{n} \quad \leq \quad \mathbb{E}\left[\int_0^{t\wedge\tau_n} \varphi_n'(\delta_s)\left(\mu(s,X_s) - \eta(s,Y_s)\right)\right] ds$$

Since $\varphi_n \geq 0$ and $\mu \leq \eta$, this provides

$$\mathbb{E}[\varphi_n(\delta_t)] - \frac{t}{n} \quad \leq \quad \mathbb{E}\left[\int_0^{t\wedge\tau_n} \varphi_n'(\delta_s)\left(\mu(s,X_s) - \mu(s,Y_s)\right)\right] ds \tag{8.9}$$

$$\mathbb{E}[\varphi_n(\delta_t)] - \frac{t}{n} \quad \leq \quad \mathbb{E}\left[\int_0^{t\wedge\tau_n} \varphi_n'(\delta_s)\left(\eta(s,X_s) - \eta(s,Y_s)\right)\right] ds. \tag{8.10}$$

The following inequality then follows from (8.9) if μ is Lipschitz-continuous, or (8.10) if η is Lipschitz-continuous:

$$\mathbb{E}[\varphi_n(\delta_t)] - \frac{t}{n} \quad \leq \quad K\int_0^t \mathbb{E}[\delta_s^+]ds,$$

By sending $n \to \infty$, this provides $\mathbb{E}[\delta_t^+] \leq K\int_0^t \mathbb{E}[\delta_s^+]ds$, and we conclude by Gronwall lemma that $\mathbb{E}[\delta_t^+] = 0$ for all t. Hence, $\delta_t = 0$, a.s. for all $t \geq 0$, and by the pathwise continuity of the process δ, we deduce that $\delta_. = 0$, a.s. $\diamondsuit$

We conclude this section by the following exercise which proves the existence of a solution for the so-called Cox-Ingersol-Ross process (or, more exactly, the Feller process) under the condition of non-attainability of the origin.

Exercise 8.9. *(Exam, December 2002) Given a scalar Brownian motion W, we consider the stochastic differential equation:*

$$X_t \quad = \quad x + \int_0^t (\delta - 2\lambda X_s)ds + 2\int_0^t \sqrt{X_s}dW_s, \tag{8.11}$$

where λ and the initial condition x are given positive parameters, and $\delta \geq 2$. Denote $T_a := \inf\{t \geq 0 : X_t = a\}$.

1. *Let $u(x) := \int_1^x y^{-\delta/2}e^{\lambda y}dy$. Show that $u(X_t)$ is an Itô process and provide its dynamics.*

2. *For $0 < \varepsilon \leq x \leq a$, prove that $\mathbb{E}\left[u(X_{t\wedge T_\varepsilon\wedge T_a})\right] = u(x)$ for all $t \geq 0$.*

3. *For $0 < \varepsilon \leq x \leq a$, show that there exists a scalar $\alpha > 0$ such that the function $v(x) := e^{\alpha x^2}$ satisfies*

$$(\delta - 2\lambda x)v'(x) + 2xv''(x) \geq 1 \quad \textit{for} \quad x \in [\varepsilon, a].$$

Deduce that $\mathbb{E}\left[v(X_{t\wedge T_\varepsilon\wedge T_a})\right] \geq v(x) + \mathbb{E}[t\wedge T_\varepsilon\wedge T_a]$ for all $t \geq 0$, and then $\mathbb{E}[T_\varepsilon \wedge T_a] < \infty$.

4. *Prove that* $\mathbb{E}[u(X_{T_\varepsilon \wedge T_a})] = u(x)$, *and deduce that*

$$\mathbb{P}[T_a \le T_\varepsilon] \quad = \quad \frac{u(x) - u(\varepsilon)}{u(a) - u(\varepsilon)}.$$

5. *Assume* $\delta \ge 0$. *Prove that* $\mathbb{P}[T_a \le T_0] = 1$ *for all* $a \ge x$. *Deduce that* $\mathbb{P}[T_0 = \infty] = 1$ *(you may use without proof that* $T_a \longrightarrow \infty$, *a.s. when* $a \to \infty$.

8.4 Linear stochastic differential equations

8.4.1 An explicit representation

In this paragraph, we focus on linear stochastic differential equations:

$$X_t \quad = \quad \xi + \int_0^t [A(s)X_s + a(s)]\, ds + \int_0^t \sigma(s) dW_s, \ t \ge 0, \tag{8.12}$$

where W is a $d-$dimensional Brownian motion, ξ is a r.v. in $\mathbb{R}^d$ independent of W, and $A : \mathbb{R}_+ \longrightarrow \mathcal{M}_\mathbb{R}(n,n)$, $a : \mathbb{R}_+ \longrightarrow \mathbb{R}^n$, et $\sigma : \mathbb{R}_+ \longrightarrow \mathcal{M}_\mathbb{R}(n,d)$ are deterministic Borel measurable functions with A bounded, and $\int_0^T |a(s)|ds + \int_0^T |\sigma(s)|^2 ds < \infty$ for all $T > 0$.

The existence and uniqueness of a strong solution for the above linear stochastic differential equation is a consequence of Theorem 8.3.

We next use the analogy with deterministic linear differential systems to provide a general representation of the solution of (8.12). Let $H : \mathbb{R}_+ \longmapsto \mathcal{M}_\mathbb{R}(d,d)$ be the unique solution of the linear ordinary differential equation:

$$H(t) = I_d + \int_0^t A(s)H(s)ds \quad \text{for} \quad t \ge 0. \tag{8.13}$$

Observe that $H(t)$ is invertible for every $t \ge 0$ for otherwise there would exist a vector λ such that $H(t_0)\lambda = 0$ for some $t_0 > 0$; then since $x := H\lambda$ solves the ordinary differential equation $\dot{x} = A(t)x$ on $\mathbb{R}^d$, it follows that $x = 0$, contradicting the fact that $H(0) = I_d$.

Given the invertible matrix solution H of the fundamental equation (8.13), it follows that the unique solution of the deterministic ordinary differential equation $x(t) = x(0) + \int_0^t [A(s)x(s) + a(s)]ds$ is given by

$$x(t) = H(t)\left(x(0) + \int_0^t H(s)^{-1} a(s) ds\right) \quad \text{for} \quad t \ge 0.$$

Proposition 8.10. *The unique solution of the linear stochastic differential equation* (8.12) *is given by*

$$X_t := H(t)\left(\xi + \int_0^t H(s)^{-1} a(s) ds + \int_0^t H(s)^{-1}\sigma(s) dW_s\right) \quad \textit{for} \quad t \ge 0.$$

Proof. This is an immediate application of Itô's formula, which is left as an exercise. ♢

Exercise 8.11. *In the above context, show that for all $s, t \in \mathbb{R}_+$:*

$$\begin{aligned}
\mathbb{E}[X_t] &= H(t)\left(\mathbb{E}[X_0] + \int_0^t H(s)^{-1}a(s)ds\right), \\
\mathbb{C}ov[X_t, X_s] &= H(t)\left(\mathbb{V}ar[X_0] + \int_0^{s\wedge t} H(s)^{-1}\sigma(s)[H(s)^{-1}\sigma(s)]^{\mathrm{T}}ds\right)H(t)^{\mathrm{T}}.
\end{aligned}$$

8.4.2 The Brownian bridge

We now consider the one-dimensional linear stochastic differential equation:

$$X_t = a + \int_0^t \frac{b - X_s}{T - s}ds + W_t, \quad \text{for} \quad t \in [0, T), \tag{8.14}$$

where $T > 0$ and $a, b \in \mathbb{R}$ are given. This equation can be solved by the method of the previous paragraph on each interval $[0, T - \varepsilon]$ for every $\varepsilon > 0$, with fundamental solution of the corresponding linear equation

$$H(t) = 1 - \frac{t}{T} \quad \text{for all} \quad t \in [0, T). \tag{8.15}$$

This provides the natural unique solution on $[0, T)$:

$$X_t = a\left(1 - \frac{t}{T}\right) + b\frac{t}{T} + (T - t)\int_0^t \frac{dW_s}{T - s}, \quad \text{pour} \quad 0 \le t < T. \tag{8.16}$$

Proposition 8.12. *Let $\{X_t, t \in [0, T]\}$ be the process defined by the unique solution* (8.16) *of* (8.14) *on $[0, T)$, and $X_T = b$. Then X has a.s. continuous sample paths, and is a gaussian process with*

$$\begin{aligned}
\mathbb{E}[X_t] &= a\left(1 - \frac{t}{T}\right) + b\frac{t}{T}, \ t \in [0, T], \\
\mathbb{C}ov(X_t, X_s) &= (s \wedge t) - \frac{st}{T}, \ s, t \in [0, T].
\end{aligned}$$

Proof. Consider the processes

$$M_t := \int_0^t \frac{dW_s}{T - s}, \ t \ge 0, \ B_u := M_{h^{-1}(u)}, \ u \ge 0, \text{ where } h(t) := \frac{1}{T - t} - \frac{1}{T}.$$

Then, $B_0 = 0$, B has a.s. continuous sample paths, and has independent increments, and we directly see that for $0 \le s < t$, the distribution of the increment $B_t - B_s$ is gaussian with zero mean and variance

$$\mathbb{V}ar[B_t - B_s] = \int_{h^{-1}(s)}^{h^{-1}(t)} \frac{dr}{(T - r)^2} = \left[\frac{-1}{T - r}\right]_{h^{-1}(s)}^{h^{-1}(t)} = \big[h(r)\big]_{h^{-1}(s)}^{h^{-1}(t)} = t - s.$$

Then, $m(t) := \mathbb{E}[B_t] = 0$, and for $s \leq t$, $c(s,t) := \mathbb{C}ov[B_t, B_s] = \mathbb{V}ar[B_s] + \mathbb{C}ov[B_t - B_s, B_s] = \mathbb{V}ar[B_s] = s$. By Exercise 4.7, we may then conclude that B is a Brownian motion, and therefore

$$\lim_{t \nearrow T} (T-t) M_t \quad = \quad \lim_{u \nearrow \infty} \frac{B_u}{u + T^{-1}} \;=\; 0, \text{ a.s.}$$

by the law of large numbers for the Brownian motion. Hence $X_t \longrightarrow b$ a.s. when $t \nearrow T$. The expressions of the mean and the variance are obtained by direct calculation. ♢

8.5 Connection with linear partial differential equations

8.5.1 Generator

Let $\{X_s^{t,x}, s \geq t\}$ be the unique strong solution of

$$X_s^{t,x} \quad = \quad x + \int_t^s \mu(u, X_u^{t,x}) du + \int_t^s \sigma(u, X_u^{t,x}) dW_u, \;\; s \geq t,$$

where μ and σ satisfy the required condition for existence and uniqueness of a strong solution.

For a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, we define the function $\mathcal{A}f$ by

$$\mathcal{A}f(t,x) = \lim_{h \to 0} \frac{\mathbb{E}[f(X_{t+h}^{t,x})] - f(x)}{h} \quad \text{if the limit exists}$$

Clearly, $\mathcal{A}f$ is well-defined for all bounded C^2- function with bounded derivatives and

$$\mathcal{A}f \quad = \quad \mu \cdot \frac{\partial f}{\partial x} + \frac{1}{2} \mathrm{Tr} \left[\sigma \sigma^{\mathrm{T}} \frac{\partial^2 f}{\partial x \partial x^{\mathrm{T}}} \right], \tag{8.17}$$

(Exercise !). The linear differential operator $\mathcal{A}$ is called the *generator* of X. It turns out that the process X can be completely characterized by its generator or, more precisely, by the generator and the corresponding domain of definition...

As the following result shows, the generator provides an intimate connection between conditional expectations and linear partial differential equations.

Proposition 8.13. *Assume that the function $(t,x) \longmapsto v(t,x) := \mathbb{E}\left[g(X_T^{t,x})\right]$ is $C^{1,2}\left([0,T) \times \mathbb{R}^n\right)$. Then v solves the partial differential equation:*

$$\frac{\partial v}{\partial t} + \mathcal{A}v = 0 \quad \textit{and} \quad v(T,.) = g.$$

Proof. Given (t,x), let $\tau_1 := T \wedge \inf\{s > t : |X_s^{t,x} - x| \geq 1\}$. By the law of iterated expectation, it follows that

$$V(t,x) \quad = \quad \mathbb{E}\left[V\left(s \wedge \tau_1, X_{s \wedge \tau_1}^{t,x}\right)\right].$$

Since $V \in C^{1,2}([0,T),\mathbb{R}^n)$, we may apply Itô's formula, and we obtain by taking expectations:

$$\begin{aligned}
0 &= \mathbb{E}\left[\int_t^{s\wedge\tau_1}\left(\frac{\partial v}{\partial t}+\mathcal{A}v\right)(u,X_u^{t,x})du\right]\\
&\quad +\mathbb{E}\left[\int_t^{s\wedge\tau_1}\frac{\partial v}{\partial x}(u,X_s^{t,x})\cdot\sigma(u,X_u^{t,x})dW_u\right]\\
&= \mathbb{E}\left[\int_t^{s\wedge\tau_1}\left(\frac{\partial v}{\partial t}+\mathcal{A}v\right)(u,X_u^{t,x})du\right],
\end{aligned}$$

where the last equality follows from the boundedness of $(u, X_u^{t,x})$ on $[t, s\wedge\tau_1]$. We now send $s \searrow t$, and the required result follows from the dominated convergence theorem. $\diamondsuit$

8.5.2 Cauchy problem and the Feynman-Kac representation

In this section, we consider the following linear partial differential equation

$$\begin{array}{l}\frac{\partial v}{\partial t}+\mathcal{A}v-k(t,x)v+f(t,x)=0, \quad (t,x)\in[0,T)\times\mathbb{R}^d\\ v(T,.)=g\end{array} \tag{8.18}$$

where $\mathcal{A}$ is the generator (8.17), g is a given function from $\mathbb{R}^d$ to $\mathbb{R}$, k and f are functions from $[0,T]\times\mathbb{R}^d$ to $\mathbb{R}$, b and σ are functions from $[0,T]\times\mathbb{R}^d$ to $\mathbb{R}^d$ and $\mathcal{M}_{\mathbb{R}}(d,d)$, respectively. This is the so-called Cauchy problem.

For example, when $k = f \equiv 0$, $b \equiv 0$, and σ is the identity matrix, the above partial differential equation reduces to the heat equation.

Our objective is to provide a representation of this purely deterministic problem by means of stochastic differential equations. We then assume that μ and σ satisfy the conditions of Theorem 8.3, namely that

$$\mu,\sigma \text{ Lipschitz in } x \text{ uniformly in } t, \quad \int_0^T\left(|\mu(t,0)|^2+|\sigma(t,0)|^2\right)dt<\infty. \tag{8.19}$$

Theorem 8.14. *Let the coefficients μ, σ be continuous and satisfy* (8.19). *Assume further that the function k is uniformly bounded from below, and f has quadratic growth in x uniformly in t. Let v be a $C^{1,2}\left([0,T),\mathbb{R}^d\right)$ solution of* (8.18) *with quadratic growth in x uniformly in t. Then*

$$v(t,x) \;=\; \mathbb{E}\left[\int_t^T \beta_s^{t,x}f(s,X_s^{t,x})ds+\beta_T^{t,x}g\left(X_T^{t,x}\right)\right], \qquad t\le T,\ x\in\mathbb{R}^d,$$

where $X_s^{t,x} := x+\int_t^s \mu(u,X_u^{t,x})du+\int_t^s \sigma(u,X_u^{t,x})dW_u$ and $\beta_s^{t,x} := e^{-\int_t^s k(u,X_u^{t,x})du}$ for $t \le s \le T$.

Proof. We first introduce the sequence of stopping times

$$\tau_n \quad := \quad (T-n^{-1}) \wedge \inf\left\{s>t \ : \ \left|X_s^{t,x}-x\right| \geq n\right\},$$

and we oberve that $\tau_n \longrightarrow T$ $\mathbb{P}-$a.s. Since v is smooth, it follows from Itô's formula that for $t \leq s < T$:

$$\begin{aligned}
d\left(\beta_s^{t,x} v\left(s, X_s^{t,x}\right)\right) &= \beta_s^{t,x}\left(-kv+\frac{\partial v}{\partial t}+\mathcal{A}v\right)\left(s, X_s^{t,x}\right) ds \\
&\quad +\beta_s^{t,x}\frac{\partial v}{\partial x}\left(s, X_s^{t,x}\right)\cdot\sigma\left(s, X_s^{t,x}\right) dW_s \\
&= \beta_s^{t,x}\left(-f(s, X_s^{t,x})ds+\frac{\partial v}{\partial x}\left(s, X_s^{t,x}\right)\cdot\sigma\left(s, X_s^{t,x}\right) dW_s\right),
\end{aligned}$$

by the PDE satisfied by v in (8.18). Then:

$$\begin{aligned}
&\mathbb{E}\left[\beta_{\tau_n}^{t,x} v\left(\tau_n, X_{\tau_n}^{t,x}\right)\right]-v(t,x) \\
&= \mathbb{E}\left[\int_t^{\tau_n} \beta_s^{t,x}\left(-f(s, X_s)ds+\frac{\partial v}{\partial x}\left(s, X_s^{t,x}\right)\cdot\sigma\left(s, X_s^{t,x}\right) dW_s\right)\right].
\end{aligned}$$

Now observe that the integrands in the stochastic integral is bounded by definition of the stopping time τ_n, the smoothness of v, and the continuity of σ. Then the stochastic integral has zero mean, and we deduce that

$$v(t,x) \quad = \quad \mathbb{E}\left[\int_t^{\tau_n} \beta_s^{t,x} f\left(s, X_s^{t,x}\right) ds+\beta_{\tau_n}^{t,x} v\left(\tau_n, X_{\tau_n}^{t,x}\right)\right]. \tag{8.20}$$

Since $\tau_n \longrightarrow T$ and the Brownian motion has continuous sample paths $\mathbb{P}-$a.s. it follows from the continuity of v that, $\mathbb{P}-$a.s.

$$\begin{aligned}
&\int_t^{\tau_n} \beta_s^{t,x} f\left(s, X_s^{t,x}\right) ds+\beta_{\tau_n}^{t,x} v\left(\tau_n, X_{\tau_n}^{t,x}\right) \\
&\quad \overset{n\to\infty}{\longrightarrow} \int_t^{T} \beta_s^{t,x} f\left(s, X_s^{t,x}\right) ds+\beta_T^{t,x} v\left(T, X_T^{t,x}\right) \\
&\quad = \int_t^{T} \beta_s^{t,x} f\left(s, X_s^{t,x}\right) ds+\beta_T^{t,x} g\left(X_T^{t,x}\right)
\end{aligned} \tag{8.21}$$

by the terminal condition satisfied by v in (8.18). Moreover, since k is bounded from below and the functions f and v have quadratic growth in x uniformly in t, we have

$$\left|\int_t^{\tau_n} \beta_s^{t,x} f\left(s, X_s^{t,x}\right) ds+\beta_{\tau_n}^{t,x} v\left(\tau_n, X_{\tau_n}^{t,x}\right)\right| \quad \leq \quad C\left(1+\max_{t\leq T}|X_t|^2\right).$$

By the estimate stated in the existence and uniqueness theorem 8.3, the latter bound is integrable, and we deduce from the dominated convergence theorem that the convergence in (8.21) holds in $\mathbb{L}^1(\mathbb{P})$, proving the required result by taking limits in (8.20). ♢

The above Feynman-Kac representation formula has an important numerical implication. Indeed it opens the door to the use of Monte Carlo methods in order to obtain a numerical approximation of the solution of the partial differential equation (8.18). For sake of simplicity, we provide the main idea in the case $f = k = 0$. Let $\left(X^{(1)}, \ldots, X^{(k)}\right)$ be an iid sample drawn in the distribution of $X_T^{t,x}$, and compute the mean:

$$\hat{v}_k(t,x) \quad := \quad \frac{1}{k}\sum_{i=1}^{k} g\left(X^{(i)}\right).$$

By the Law of Large Numbers, it follows that $\hat{v}_k(t,x) \longrightarrow v(t,x)$ $\mathbb{P}-$a.s. Moreover the error estimate is provided by the Central Limit Theorem:

$$\sqrt{k}\left(\hat{v}_k(t,x) - v(t,x)\right) \overset{k\to\infty}{\longrightarrow} \mathcal{N}\left(0, \mathbb{V}ar\left[g\left(X_T^{t,x}\right)\right]\right) \qquad \text{in distribution,}$$

and is remarkably independent of the dimension d of the variable X !

8.5.3 Representation of the Dirichlet problem

Let D be an open subset of $\mathbb{R}^d$. The *Dirichlet problem* is to find a function u solving:

$$\mathcal{A}u - ku + f = 0 \text{ on } D \quad \text{and} \quad u = g \text{ on } \partial D, \tag{8.22}$$

where ∂D denotes the boundary of D, and $\mathcal{A}$ is the generator of the process X^{0,X_0} defined as the unique strong solution of the stochastic differential equation

$$X_t^{0,X_0} = X_0 + \int_0^t \mu(s, X_s^{0,X_0})ds + \int_0^t \sigma(s, X_s^{0,X_0})dW_s, \; t \geq 0.$$

Similarly to the the representation result of the Cauchy problem obtained in Theorem 8.14, we have the following representation result for the Dirichlet problem.

Theorem 8.15. *Let u be a C^2-solution of the Dirichlet problem* (8.22)*. Assume that k is bounded from below, and*

$$\mathbb{E}[\tau_D^x] < \infty, \; x \in \mathbb{R}^d, \quad \textit{where} \quad \tau_D^x := \inf\left\{t \geq 0 : \; X_t^{0,x} \notin D\right\}.$$

Then, we have the representation:

$$u(x) \quad = \quad \mathbb{E}\left[g\left(X_{\tau_D}^{0,x}\right)e^{-\int_0^{\tau_D} k(X_s)ds} + \int_0^{\tau_D} f\left(X_t^{0,x}\right)e^{-\int_0^t k(X_s)ds}dt\right].$$

Exercise 8.16. *Provide a proof of Theorem 8.15 by imitating the arguments in the proof of Theorem 8.14.*

8.6 The hedging portfolio in a Markov financial market

In this paragraph, we return to the context of Section 7.4, and we assume further that the Itô process S is defined by a stochastic differential equation, i.e.

$$\mu_t = \mu(t, S_t), \;\; \sigma_t = \sigma(t, S_t),$$

and the interest rate process $r_t = r(t, S_t)$. In particular, the risk premium process is also a deterministic function of the form $\lambda_t = \lambda(t, S_t)$, and the dynamics of the process S under the risk-neutral measure $\mathbb{Q}$ is given by:

$$dS_t \;\; = \;\; \mathrm{diag}[S_t]\left(r(t, S_t)\mathbf{1}dt + \sigma(t, S_t)dB_t\right),$$

where we recall the B is a $\mathbb{Q}-$Brownian motion. We assume that these coefficients are subject to all required conditions so that existence and uniqueness of the Itô processes, together with the conditions of Section 7.4 are satisfied.

Finally, we assume that the derivative security is defined by a Vanilla contract, i.e. $G = g(S_T)$ for some function g with quadratic growth.

Then, from Theorem 7.17, the no-arbitrage market price of the derivative security $g(S_T)$ is given by

$$p(G) \;\; = \;\; V(0, S_0) := \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r_u du} g(S_T)\right].$$

Moreover, a careful inspection of the proof shows that the perfect replicating strategy θ^* is obtained by means of the martingale representations of the martingale $Y_t := \mathbb{E}^{\mathbb{Q}}[\tilde{G}|\mathcal{F}_t]$. In order to identify the optimal portfolio, we introduce the derivative security's price at each time $t \in [0, T]$:

$$\begin{aligned} V(t, S_t) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_u du} g(S_T)\Big| \mathcal{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_u du} g(S_T)\Big| S_t\right], \quad t \in [0, T], \end{aligned}$$

and we observe that $Y_t = e^{-\int_0^t r_u du} V(t, S_t)$, $t \in [0, T]$.

Proposition 8.17. *In the above context, assume that the function $(t, s) \longmapsto V(t, s)$ is $C^{1,2}\left([0, T), (0, \infty)^d\right)$. Then the perfect replicating strategy of the derivative security $G = g(S_T)$ is given by*

$$\theta^*_t = \mathrm{diag}[S_t]\, \frac{\partial V}{\partial s}(t, S_t), \;\; t \in [0, T).$$

In other words the perfect replicating strategy requires that the investor holds a hedging portfolio consisting of

$$\Delta^i_t := \frac{\partial V}{\partial s^i}(t, S_t) \quad \textit{shares of } S^i \textit{ at each time} \quad t \in [0, T).$$

Proof. From the discussion preceeding the statement of the proposition, the perfect hedging strategy is obtained from the martingale representation of the process $Y_t = e^{-\int_0^t r_u du}V(t,S_t)$, $t \in [0,T]$. Since V has the required regularity for the application of Itô's formula, we obtain:

$$dY_t \quad = \quad e^{-\int_0^t r_u du}\left(...dt + \frac{\partial V}{\partial s}(t,S_t)\cdot dS_t\right) \;=\; \frac{\partial V}{\partial s}(t,S_t)\cdot d\tilde{S}_t,$$

where, in the last equality, the "dt" coefficient is determined from the fact that $\tilde{S}$ and Y are $\mathbb{Q}-$martingales. The expression of θ^* is then easily obtained by identifying the latter expression with that of a portfolio value process. ♢

8.7 Application to importance sampling

Importance sampling is a popular variance reduction technique in Monte Carlo simulation. In this section, we recall the basic features of this technique in the simple context of simulating a random variable. Then, we show how it can be extended to stochastic differential equations.

8.7.1 Importance sampling for random variables

Let X be a square integrable r.v. on $\mathbb{R}^n$. We assume that its distribution is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^n$ with probability density function f_X. Our task is to provide an approximation of

$$\theta \;:=\; \mathbb{E}[X]$$

by Monte Carlo simulation. To do this, we assume that independent copies $(X_i)_{i\geq 1}$ of the r.v. X are available. Then, from the law of large numbers, we have:

$$\hat{\theta}_N \;:=\; \frac{1}{N}\sum_{i=1}^N X_i \quad\longrightarrow\quad \theta \;\; \mathbb{P}-\text{a.s.}$$

Moreover, the approximation error is given by the central limit theorem:

$$\sqrt{N}\left(\hat{\theta}_N - \theta\right) \quad\longrightarrow\quad \mathcal{N}\left(0, \mathbb{V}ar[X_1]\right) \;\; \text{in distribution.}$$

We call the approximation $\hat{\theta}_N$ the *naive Monte Carlo estimator* of θ, in the sense that it is the most natural. Indeed, one can devise many other Monte Carlo estimators as follows. Let Y be any other r.v. absolutely continuous with respect to the Lebesgue measure with density f_Y satisfying the support restriction

$$f_Y > 0 \quad \text{on} \quad \{f_X > 0\}.$$

Then, one can re-write θ as:

$$\theta \quad = \quad \mathbb{E}\left[\frac{f_X(Y)}{f_Y(Y)}\, Y\right].$$

Then, assuming that independent copies $(Y_i)_{i\geq 1}$ of the r.v. Y are available, this suggests an alternative Monte Carlo estimator:

$$\hat{\theta}_N(Y) \quad := \quad \frac{1}{N}\sum_{i=1}^{N}\frac{f(Y_i)}{g(Y_i)}\, Y_i.$$

By the law of large numbers and the central limit theorem, we have:

$$\hat{\theta}_N(Y) \quad \longrightarrow \quad \theta, \quad \text{a.s.}$$

and

$$\sqrt{N}\left(\hat{\theta}_N(Y)-\mu\right) \quad \longrightarrow \quad \mathcal{N}\left(0, \mathbb{V}ar\left[\frac{f_X(Y)}{f_Y(Y)}\, Y\right]\right) \text{ in distribution}$$

Hence, for every choice of a probability density function f_Y satisfying the above support restriction, one may build a corresponding Monte Calo estimator $\hat{\theta}_N(Y)$ which is consistent, but differs from the naive Monte Carlo estimator by the asymptotic variance of the error. It is then natural to wonder whether one can find an optimal density in the sense of minimization of the asymptotic variance of the error:

$$\min_{f_Y} \mathbb{V}ar\left[\frac{f_X(Y)}{f_Y(Y)}\, Y\right].$$

This minimization problem turns out to be very easy to solve. Indeed, since $\mathbb{E}\left[\frac{f_X(Y)}{f_Y(Y)}\, Y\right]=\mathbb{E}[X]$ and $\mathbb{E}\left[\frac{f_X(Y)}{f_Y(Y)}\, |Y|\right]=\mathbb{E}[|X|]$ do not depend on f_Y, we have the equivalence between the following minimization problems:

$$\min_{f_Y} \mathbb{V}ar\left[\frac{f_X(Y)}{f_Y(Y)}\, Y\right] \quad \equiv \quad \min_{f_Y} \mathbb{E}\left[\frac{f_X(Y)^2}{f_Y(Y)^2}\, Y^2\right] \;\equiv\; \min_{f_Y} \mathbb{V}ar\left[\frac{f_X(Y)}{f_Y(Y)}\, |Y|\right],$$

and the solution of the latter problem is given by

$$f_Y^*(y) := \frac{|y| f_X(y)}{\mathbb{E}[|X|]},$$

Moreover, when $X\geq 0$ a.s. the minimum variance is zero ! this means that, by simulating one single copy Y_1 according to the optimal density f_Y^*, one can calculate the required expected value θ without error !

Of course, this must not be feasible, and the problem here is that the calculation of the optimal probability density function f_Y^* involves the computation of the unknown expectation θ.

However, this minimization is useful, and can be used as follows:

- Start from an initial (poor) estimation of θ, from the naive Monte Carlo estimator for instance. Deduce an estimator $\hat{f}^*_Y$ of the optimal probability density f^*_Y. Simulate independent copies $(Y_i)_{i\geq 1}$ according to $\hat{f}^*_Y$, and compute a second stage Monte Carlo estimator.

- Another application is to perform a Hastings-Metropolis algorithm and take advantage of the property that the normalizing factor $\mathbb{E}[|X|]$ is not needed... (MAP 432).

8.7.2 Importance sampling for stochastic differential equations

We aim at approximating

$$u(0,x) := \mathbb{E}\left[g\left(X_T^{0,x}\right)\right]$$

where $X_\cdot^{0,x}$ is the unique strong solution of

$$X_t^{0,x} = x + \int_0^t \mu\left(t, X_t^{0,x}\right)dt + \int_0^t \sigma\left(t, X_t^{0,x}\right)dW_t,\ t \geq 0.$$

For $\mathbb{Q} := Z_T \cdot \mathbb{P}$, a probability measure equivalent to $\mathbb{P}$ on $\mathcal{F}_T$, we have

$$u(0,x) := \mathbb{E}^{\mathbb{Q}}\Big[\frac{1}{Z_T} g\big(X_T^{0,x}\big)\Big].$$

Then, assuming that one can produce independent copies $\left(\hat{X}_T^i, \hat{Z}_T^i\right)$ under $\mathbb{Q}$ of the r.v. $\left(X_T^{0,x}, Z_T\right)$ (in practice, one can only generate a discrete-time approximation...), we see that each choice of density Z suggests a Monte Carlo approximation:

$$\hat{u}_N^Z(0,x) \ := \ \frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{Z}_T^i} g\left(\hat{X}_T^i\right) \quad \longrightarrow \quad u(0,x) \ \ \mathbb{P}-\text{a.s.}$$

and the central limit theorem says that an optimal choice of Z consists in minimizing the asymptotic variance

$$\min_{Z_T} \mathbb{V}ar^{\mathbb{Q}}\left[\frac{1}{Z_T} g\left(X_T^{0,x}\right)\right].$$

To do this, we restrict our attention to those densities defined by

$$Z_0^h = 1 \quad \text{and} \quad dZ_t^h = Z_t^h h_t \cdot dW_t,\ t \in [0,T],$$

for some $h_t = h(t, X_t)$ satisfying $\int_0^T |h_t|^2 dt < \infty$ $\mathbb{P}-$a.s. and $\mathbb{E}[Z_T^h] = Z_0^h = 1$. We denote by $\mathcal{H}$ the collection of all such processes.

Under the probability measure $\mathbb{Q}^h := Z_T^h \cdot \mathbb{P}$ on $\mathcal{F}_T$, it follows from the Girsanov theorem that

$$\left\{ W_t^h = W_t - \int_0^t h_u du, \; 0 \le t \le T \right\} \qquad \text{is a } \mathbb{Q}^h - \text{Brownian motion}$$

The dynamics of X and $M^h := \left(Z^h\right)^{-1}$ under $\mathbb{Q}^h$ are given by

$$\begin{aligned} dX_t &= (\mu + \sigma h)dt + \sigma(t, X_t)dW_t^h, \\ dM_t^h &= -M_t^h h(t, X_t) \cdot dW_t^h. \end{aligned}$$

We now solve the minimization problem

$$V_0 \;:=\; \min_{h \in \mathcal{H}} \mathbb{V}ar^{\mathbb{Q}^h}\left[M_T^h g\left(X_T^{t,x}\right)\right]$$

The subsequent calculation uses the fact that the function $u(t,x) := \mathbb{E}\left[g\left(X_T^{t,x}\right)\right] = \mathbb{E}^{\mathbb{Q}^h}\left[M_T^h g\left(X_T^{t,x}\right)\right]$ solves the partial differential equation

$$\frac{\partial u}{\partial t} + \mathcal{L}_t^0 u \;=\; 0,$$

by Proposition 8.13, where $\mathcal{L}^0$ is the generator of X under the original probability measure $\mathbb{P}$.

Applying Itô's formula to the product $M_t^u u(t, X_t)$, we see that

$$\begin{aligned} M_T^h g(X_T) &= M_T^h u(T, X_T) \\ &= u(0,x) + \int_0^T M_t^h \big\{ -(hu)(t,X_t) \cdot dW_t^h + du(t,X_t) - (h\sigma)(t,X_t)dt \big\} \\ &= u(0,x) + \int_0^T M_t^h \big\{ -(hu)(t,X_t) \cdot dW_t^h + (\partial_t + \mathcal{L}^0)u(t,X_t)dt \\ &\qquad\qquad\qquad\qquad\qquad\qquad + (Du \cdot \sigma)(t,X_t)dW_t^h \big\} \\ &= u(0,x) + \int_0^T M_t^h \big\{ \sigma^{\mathrm{T}} Du - uh \big\}(t,X_t) \cdot dW_t^h. \end{aligned}$$

Then, if $h^*(t,x) := \sigma^{\mathrm{T}}(\partial \ln u/\partial x)(t,x)$ induces a process in $\mathcal{H}$, this shows that $M_T^{h^*} g(X_T) = u(0,x)$, implying that $\mathbb{V}ar^{\mathbb{Q}^{h^*}}\left[M_T^{h^*} g\left(X_T^{t,x}\right)\right] = 0$. Since the variance is non-negative, this shows that

$$V_0 = 0 \quad \text{and a solution is} \quad h^* := \sigma^{\mathrm{T}}(\partial \ln u/\partial x)(., X_.)$$

Similarly to the case of random variables, this result can be used either to devise a two-stage Monte Carlo method, or to combine with a Hastings-Metropolis algorithm.

Chapter 9

The Black-Scholes model and its extensions

In the previous chapters, we have seen the Black-Sholes formula proved from three different approaches: continuous-time limit of the Cox-Ingersol-Ross binomial model, verification from the solution of a partial differential equation, and the elegant martingale approach. However, none of these approaches was originally used by Black and Scholes. The first section of this chapter presents the original intuitive argument contained in the seminal paper by Black and Scholes. The next section reviews the Black-Scholes formula and shows various extensions which are needed in the every-day practive of the model within the financial industry. The final section provides some calculations for barrier options.

9.1 The Black-Scholes approach for the Black-Scholes formula

In this section, we derive a formal argument in order to obtain the valuation PDE (6.9) from Chapter 6. The following steps have been employed by Black and Scholes in their pioneering work [7].
1. Let $p(t, S_t)$ denote the time$-t$ market price of a contingent claim defined by the payoff $B = g(S_T)$ for some function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$. Notice that we are accepting without proof that p is a deterministic function of time and the spot price, this has been in fact proved in the previous section.

2. The holder of the contingent claim completes his portfolio by some investment in the risky assets. At time t, he decides to holds $-\Delta^i$ shares of the risky asset S^i. Therefore, the total value of the portfolio at time t is

$$P_t := p(t, S_t) - \Delta \cdot S_t \,, \qquad 0 \le t < T \,.$$

3. Considering *delta* as a constant vector in the time interval $[t, t+dt)$, and assuming that the function p is of class $C^{1,2}$, the variation of the portfolio value is given by :

$$dP_t \;=\; \mathcal{L}p(t,S_t)dt + \frac{\partial p}{\partial s}(t,S_t)\cdot dS_t - \Delta\cdot dS_t\,.$$

where $\mathcal{L}p = p_t + \frac{1}{2}\mathrm{Tr}[\mathrm{diag}[s]\sigma\sigma^T\mathrm{diag}[s]\frac{\partial^2 p}{\partial s\partial s^{\mathrm{T}}}]$. In particular, by setting

$$\Delta \;=\; \frac{\partial p}{\partial s}\,,$$

we obtain a portfolio value with finite quadratic variation

$$dP_t \;=\; \mathcal{L}p(t,S_t)dt\,. \tag{9.1}$$

4. The portfolio P_t is non-risky since the variation of its value in the time interval $[t, t+dt)$ is known in advance at time t. Then, by the no-arbitrage argument, we must have

$$\begin{aligned} dP_t &= r(t,S_t)P_t dt \;=\; r(t,S_t)[p(t,S_t) - \Delta\cdot S_t]dt \\ &= r(t,S_t)\left[p(t,S_t) - \frac{\partial p}{\partial s}\cdot S_t\right]dt \end{aligned} \tag{9.2}$$

By equating (9.1) and (9.2), we see that the function p satisfies the PDE

$$\frac{\partial p}{\partial t} + rs\cdot\frac{\partial p}{\partial s} + \frac{1}{2}\mathrm{Tr}\left[\mathrm{diag}[s]\sigma\sigma^T\mathrm{diag}[s]\frac{\partial^2 p}{\partial s\partial s^{\mathrm{T}}}\right] - rp \;=\; 0\,,$$

which is exactly the PDE obtained in the previous section.

9.2 The Black and Scholes model for European call options

9.2.1 The Black-Scholes formula

In this section, we consider the one-dimensional Black-Scholes model $d = 1$ so that the price process S of the single risky asset is given in terms of the $\mathbb{Q}-$Brownian motion B :

$$S_t \;=\; S_0\exp\left[\left(r - \frac{\sigma^2}{2}\right)t + \sigma B_t\right]\,, \qquad 0 \le t \le T\,. \tag{9.3}$$

Observe that the random variable S_t is log-normal for every fixed t. This is the key-ingredient for the next explicit result.

Proposition 9.1. *Let $G = (S_T - K)^+$ for some $K > 0$. Then the no-arbitrage price of the contingent claim G is given by the so-called Black-Scholes formula :*

$$p_0(G) \;=\; S_0\,\mathbf{N}\left(\mathbf{d}_+(S_0,\tilde{K},\sigma^2 T)\right) \;-\; \tilde{K}\,\mathbf{N}\left(\mathbf{d}_-(S_0,\tilde{K},\sigma^2 T)\right)\,, \tag{9.4}$$

where

$$\tilde{K} := Ke^{-rT}\,, \qquad \mathbf{d}_{\pm}(s,k,v) := \frac{\ln(s/k)}{\sqrt{v}} \pm \frac{1}{2}\sqrt{v}\,, \tag{9.5}$$

and the optimal hedging strategy is given by

$$\hat{\pi}_t \;=\; S_t\mathbf{N}\left(\mathbf{d}_+(S_t,\tilde{K},\sigma^2(T-t))\right)\,, \qquad 0 \le t \le T\,. \tag{9.6}$$

Proof. This formula can be derived by various methods. One can just calculate directly the expected value by exploiting the explicit probability density function of the random variable S_T. One can also guess a solution for the valuation PDE corresponding to the call option. We shall present another method which relies on the technique of change of measure and reduces considerably the computational effort. We first decompose

$$p_0(G) \;=\; \mathbb{E}^{\mathbb{Q}}\left[\tilde{S}_T\,\mathbf{1}_{\{\tilde{S}_T \ge \tilde{K}\}}\right] \;-\; \tilde{K}\,\mathbb{Q}\left[\tilde{S}_T \ge \tilde{K}\right] \tag{9.7}$$

where as usual, the *tilda* notation corresponds to discounting, i.e. multiplication by e^{-rT} in the present context.

1. The second term is directly computed by exploiting the knowledge of the distribution of $\tilde{S}_T$:

$$\begin{aligned}
\mathbb{Q}\left[\tilde{S}_T \ge \tilde{K}\right] &= \mathbb{Q}\left[\frac{\ln(\tilde{S}_T/S_0) + (\sigma^2/2)T}{\sigma\sqrt{T}} \ge \frac{\ln(\tilde{K}/S_0) + (\sigma^2/2)T}{\sigma\sqrt{T}}\right] \\
&= 1 - \mathbf{N}\left(\frac{\ln(\tilde{K}/S_0) + (\sigma^2/2)T}{\sigma\sqrt{T}}\right) \\
&= \mathbf{N}\left(\mathbf{d}_-(S_0,\tilde{K},\sigma^2 T)\right)\,.
\end{aligned}$$

2. As for the first expected value, we define the new measure $\mathbb{P}^1 := Z_T^1 \cdot \mathbb{Q}$ on $\mathcal{F}_T$, where

$$Z_T^1 \;:=\; \exp\left(\sigma B_T - \frac{\sigma^2}{2}T\right) \;=\; \frac{\tilde{S}_T}{S_0}\,.$$

By the Girsanov theorem, the process $W_t^1 := B_t - \sigma t$, $0 \le t \le T$, defines a Brownian motion under $\mathbb{P}^1$, and the random variable

$$\frac{\ln(\tilde{S}_T/S_0) - (\sigma^2/2)T}{\sigma\sqrt{T}} \quad \text{is distributed as} \quad \mathcal{N}(0,1) \text{ under } \mathbb{P}^1\,.$$

We now re-write the first term in (9.7) as

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[\tilde{S}_T\,\mathbf{1}_{\{\tilde{S}_T \ge \tilde{K}\}}\right] &= S_0\mathbb{P}^1\left[\tilde{S}_T \ge \tilde{K}\right] \\
&= S_0\mathbb{P}\mathrm{rob}\left[\mathcal{N}(0,1) \ge \frac{\ln(\tilde{K}/S_0) - (\sigma^2/2)T}{\sigma\sqrt{T}}\right] \\
&= S_0\mathbf{N}\left(\mathbf{d}_+(S_0,\tilde{K},\sigma^2 T)\right)\,.
\end{aligned}$$

3. The optimal hedging strategy is obtained by directly differentiating the price formula with respect to the underlying risky asset price, see Proposition 8.17. $\diamondsuit$

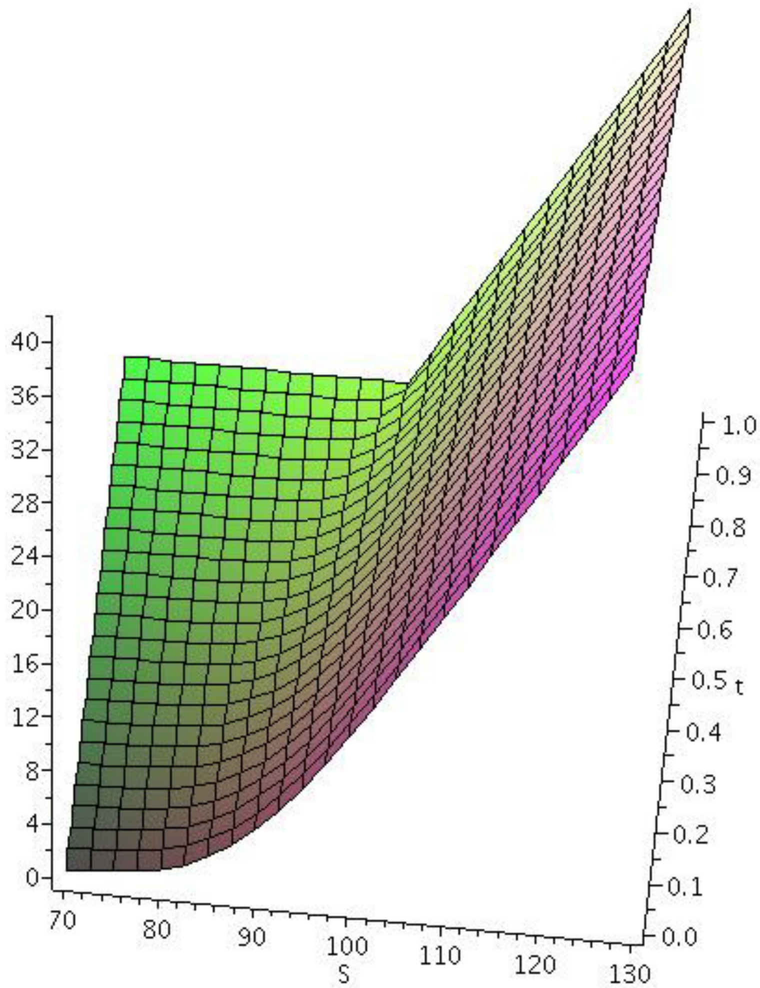

Figure 9.1: The Black-Scholes formula as a function of S and t

Exercise 9.2 (Black-Scholes model with time-dependent coefficients)**.** Consider the case where the interest rate is a deterministic function $r(t)$, and the risky asset price process is defined by the time dependent coefficients $b(t)$ and $\sigma(t)$. Show that the European call option price is given by the extended Black-Sholes formula:

$$p_0(G) \quad = \quad S_0 \, \mathbf{N}\left(\mathbf{d}_+(S_0, \tilde{K}, v(T))\right) - \tilde{K} \, \mathbf{N}\left(\mathbf{d}_-(S_0, \tilde{K}, v(T))\right) \qquad (9.8)$$

where

$$\tilde{K} := Ke^{-\int_0^T r(t)dt}, \qquad v(T) := \int_0^T \sigma^2(t)dt \,. \qquad (9.9)$$

What is the optimal hedging strategy. $\diamondsuit$

9.2.2 The Black's formula

We again assume that the financial market contains one single risky asset with price process defined by the constant coefficients Black-Scholes model. Let $\{F_t, t \geq 0\}$ be the price process of the forward contract on the risky asset with maturity $T' > 0$. Since the interest rates are deterministic, we have

$$F_t \;=\; S_t e^{r(T'-t)} \quad = \quad F_0 e^{-\frac{1}{2}\sigma^2 t + \sigma B_t}, \; 0 \leq t \leq T \,.$$

In particular, we observe that the process $\{F_t, \; t \in [0, T']\}$ is a martingale under the risk neutral measure $\mathbb{Q}$. As we shall see in next chapter, this property is specific to the case of deterministic interest rates, and the corresponding result in a stochastic interest rates framework requires to introduce the so-called forward neutral measure.

We now consider the European call option on the forward contract F with maturity $T \in (0, T']$ and strike price $K > 0$. The corresponding payoff at the maturity T is $G := (F_T - K)^+$. By the previous theory, its price at time zero is given by

$$p_0(G) \quad = \quad \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\left(F_T - K\right)^+\right] \,.$$

In order to compute explicitly the above expectation, we shall take advantage of the previous computations, and we observe that $e^{rT} p_0(G)$ corresponds to the Black-Scholes formula for a zero interest rate. Hence:

$$p_0(G) \quad = \quad e^{-rT}\left[F_0 \mathbf{N}\left(\mathbf{d}_+(F_0, K, \sigma^2 T)\right) - K \mathbf{N}\left(\mathbf{d}_-(F_0, K, \sigma^2 T)\right)\right] . \qquad (9.10)$$

This is the so-called Black's formula.

9.2.3 Option on a dividend paying stock

When the risky asset S pays out some dividend, the previous theory requires some modifications. We shall first consider the case where the risky asset pays a lump sum of dividend at some pre-specified dates, assuming that the process S is defined by the Black-Scholes dynamics between two successive dates of dividend payment. This implies a downward jump of the price process upon the payment of the dividend. We next consider the case where the risky asset pays a continuous dividend defined by some constant rate. The latter case can be viewed as a model simplification for a risky asset composed by a basket of a large number of dividend paying assets.

Lump payment of dividends Consider a European call option with maturity $T > 0$, and suppose that the underlying security pays out a lump of dividend at the pre-specified dates $t_1, \ldots, t_n \in (0, T)$. At each time t_j, $j = 1, \ldots, n$, the amount of dividend payment is

$$\delta_j S_{t_j -}$$

where $\delta_1, \ldots, \delta_n \in (0, 1)$ are some given constants. In other words, the dividends are defined as known fractions of the security price at the pre-specified dividend payment dates. After the dividend payment, the security price jumps down immediately by the amount of the dividend:

$$S_{t_j} \;=\; (1-\delta_j) S_{t_j -}\,, \qquad j = 1, \ldots, n\,.$$

Between two successives dates of dividend payment, we are reduced to the previous situation where the asset pays no dividend. Therefore, the discounted security price process must be a martingale under the risk neutral measure $\mathbb{Q}$, i.e. in terms of the Brownian motion B, we have

$$S_t \;=\; S_{t_{j-1}} e^{\left(r - \frac{\sigma^2}{2}\right)(t - t_{j-1}) + \sigma\left(B_t - B_{t_{j-1}}\right)}\,, \qquad t \in [t_{j-1}, t_j)\;,$$

for $j = 1, \ldots, n$ with $t_0 := 0$. Hence

$$S_T = \hat{S}_0 e^{\left(r - \frac{\sigma^2}{2}\right)T + \sigma B_T} \quad \text{where} \quad \hat{S}_0 := S_0 \prod_{j=1}^{n} (1 - \delta_j)\,,$$

and the no-arbitrage European call option price is given by

$$\mathbb{E}^{\mathbb{Q}}\left[e^{rT}(S_T - K)^+\right] \;=\; \hat{S}_0\, \mathbf{N}\left(\mathbf{d}_+(\hat{S}_0, \tilde{K}, \sigma^2 T)\right) \;-\; \tilde{K}\, \mathbf{N}\left(\mathbf{d}_-(\hat{S}_0, \tilde{K}, \sigma^2 T)\right)\;,$$

with $\tilde{K} = K e^{-rT}$, i.e. the Black-Scholes formula with modified spot price from S_0 to $\hat{S}_0$.

Continuous dividend payment We now suppose that the underlying security pays a continuous stream of dividend $\{\delta S_t,\ t \geq 0\}$ for some given constant rate $\delta > 0$. This requires to adapt the no-arbitrage condition so as to account for the dividend payment. From the financial viewpoint, the holder of the option can immediately re-invest the dividend paid in cash into the asset at any time $t \geq 0$. By doing so, the position of the security holder at time t is

$$S_t^{(\delta)} \;:=\; S_t e^{\delta t}\,, \qquad t \geq 0\,.$$

In other words, we can reduce the problem the non-dividend paying security case by increasing the value of the security. By the no-arbitrage theory, the discounted process $\left\{e^{-rt}S_t^{(\delta)},\ t \geq 0\right\}$ must be a martingale under the risk neutral measure $\mathbb{Q}$:

$$S_t^{(\delta)} \;=\; S_0^{(\delta)} e^{\left(r-\frac{\sigma^2}{2}\right)t+\sigma B_t} \;=\; S_0 e^{\left(r-\frac{\sigma^2}{2}\right)t+\sigma B_t}, \quad t \geq 0\,,$$

where B is a Brownian motion under $\mathbb{Q}$. By a direct application of Itô's formula, this provides the expression of the security price process in terms of the Brownian motion B:

$$S_t \;=\; S_0 e^{\left(r-\delta-\frac{\sigma^2}{2}\right)t+\sigma B_t}, \qquad t \geq 0\,. \tag{9.11}$$

We are now in a position to provide the call option price in closed form:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\left(S_T-K\right)^+\right] &= e^{-\delta T}\,\mathbb{E}^{\mathbb{Q}}\left[e^{-(r-\delta)T}\left(S_T-K\right)^+\right] \\ &= e^{-\delta T}\left[S_0\,\mathbf{N}\left(\mathbf{d}_+(S_0,\tilde{K}^{(\delta)},\sigma^2T)\right)\right. \\ &\qquad \left. -\,\tilde{K}^{(\delta)}\,\mathbf{N}\left(\mathbf{d}_-(S_0,\tilde{K}^{(\delta)},\sigma^2T)\right)\right],\end{aligned} \tag{9.12}$$

where

$$\tilde{K}^{(\delta)} \;:=\; Ke^{-(r-\delta)T}\,.$$

9.2.4 The Garman-Kohlhagen model for exchange rate options

We now consider a *domestic* country and a *foreign* country with different currencies. The instantaneous interest rates prevailing in the domestic country and the forign one are assumed to be constant, and will be denoted respectively by r^{d} and r^{f}.

The exchange rate from the domestic currency to the foreign one is denoted by $\mathcal{E}_t^{\mathrm{d}}$ at every time $t \geq 0$. This is the price at time t, expressed in the domestic currency, of one unit of the foreig currency. For instance, if the domestic currency is the Euro, and the foreign currency is the Dollar, then $\mathcal{E}_t^{\mathrm{d}}$ is the time$-t$ value in Euros of one Dollar; this is the Euro/Dollar exchange rate.

Similarly, one can introduce the exchange rate from the foreign currency to the domestic one $\mathcal{E}_t^{\mathrm{f}}$. Assuming that all exchange rates are positive, and that the international financial market has no frictions, it follows from a simple no-arbitrage argument that

$$\mathcal{E}_t^{\mathrm{f}} \;=\; \frac{1}{\mathcal{E}_t^{\mathrm{d}}} \quad \text{for every} \quad t \geq 0\,. \tag{9.13}$$

We postulate that the exchange rate process $\mathcal{E}^{\mathrm{d}}$ is defined by the Black-Scholes model

$$\mathcal{E}_t^{\mathrm{d}} \;=\; \mathcal{E}_0^{\mathrm{d}} e^{\left(\mu^{\mathrm{d}}-\frac{|\sigma^{\mathrm{d}}|^2}{2}\right)t+\sigma^{\mathrm{d}}W_t}, \qquad t \geq 0\,. \tag{9.14}$$

Our objective is to derive the no-arbitrage price of the exchange rate call option

$$G^{\mathrm{d}} \quad := \quad \left(\mathcal{E}_T^{\mathrm{d}} - K\right)^+ ,$$

for some $K > 0$, where the payoff G^{d} is expressed in domestic currency.

To do this, we will apply our results from the no-arbitrage valuation theory. We will first identify a risky asset in the domestic country which will isolate a unique risk neutral measure $\mathbb{P}^{\mathrm{d}}$, so that the no arbitrage price of the contingent claim G^{d} will be easily obtained once the distribution of $\mathcal{E}^{\mathrm{d}}$ under $\mathbb{P}^{\mathrm{d}}$ is determined.

1. In order to relate the exchange rate to a financial asset of the domestic country, consider the following strategy, for an investor of the domestic country, consisting of investing in the non-risky asset of the foreign country. Let the initial capital at time t be $P_t := 1$ Euro or, after immediate conversion $\left(\mathcal{E}^{\mathrm{d}}\right)^{-1}$ Dollars. Investing this amount in the foreign country non-risky asset, the investor collects the amount $\left(\mathcal{E}^{\mathrm{d}}\right)^{-1}(1 + r^{\mathrm{f}} dt)$ Dollars after a small time period dt. Finally, converting back this amount to the domestic currency provides the amount in Euros

$$P_{t+dt} = \frac{\mathcal{E}_{t+dt}^{\mathrm{d}}(1 + r^{\mathrm{f}} dt)}{\mathcal{E}_t^{\mathrm{d}}} \quad \text{and therefore} \quad dP_t = \frac{d\mathcal{E}_t^{\mathrm{d}}}{\mathcal{E}_t^{\mathrm{d}}} + r^{\mathrm{f}} dt .$$

2. Given the expression (9.14) of the exchange rate, it follows from a direct application of Itô's formula that

$$dP_t \quad = \quad \left(\mu^{\mathrm{d}} + r^{\mathrm{f}}\right) dt + \sigma^{\mathrm{d}} dW_t \ = \ r^{\mathrm{d}} dt + \sigma^{\mathrm{d}} dB_t^{\mathrm{d}}$$

where

$$B_t^{\mathrm{d}} := W_t + \lambda^{\mathrm{d}} t\,, \ t \geq 0\,, \quad \text{where} \quad \lambda^{\mathrm{d}} := \frac{\mu^{\mathrm{d}} + r^{\mathrm{f}} - r^{\mathrm{d}}}{\sigma^{\mathrm{d}}} .$$

Since P_t is the value of a portfolio of the domestic country, the unique risk neutral measure $\mathbb{P}^{\mathrm{d}}$ is identified by the property that the processes W^{d} is a Brownian motion under $\mathbb{P}^{\mathrm{d}}$, which provides by the Girsanov theorem:

$$\frac{d\mathbb{Q}^{\mathrm{d}}}{d\mathbb{P}} \quad = \ e^{-\lambda^{\mathrm{d}} W_T - \frac{1}{2}|\lambda^{\mathrm{d}}|^2 T} \quad \text{on} \quad \mathcal{F}_T .$$

3. We now can rewrite the expression of the exchange rate (9.14) in terms of the Brownian motion B^{d} of the risk neutral measure $\mathbb{Q}^{\mathrm{d}}$:

$$\mathcal{E}_t^{\mathrm{d}} \ = \ \mathcal{E}_0^{\mathrm{d}} e^{\left(r^{\mathrm{d}} - r^{\mathrm{f}} - \frac{|\sigma^{\mathrm{d}}|^2}{2}\right) t + \sigma^{\mathrm{d}} B_t^{\mathrm{d}}} , \qquad t \geq 0 .$$

Comparing with (9.11), we obtain the following

Interpretation *The exchange rate $\mathcal{E}^d$ is equivalent to an asset of the domestic currency with continuous dividend payment at the rate r^{f}.*

4. We can now compute the call option price by directly applying (9.12):

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}^{\mathrm{d}}}\left[e^{-r^{\mathrm{d}}T}G^{\mathrm{d}}\right] \quad := \quad e^{-r^{\mathrm{f}}T}\Big[&\mathcal{E}_0^{\mathrm{d}}\,\mathbf{N}\left(\mathbf{d}_+(\mathcal{E}_0^{\mathrm{d}},\tilde{K}^{(r^{\mathrm{f}})},|\sigma^{\mathrm{d}}|^2T)\right)\\ &-\tilde{K}^{(r^{\mathrm{f}})}\,\mathbf{N}\left(\mathbf{d}_-(\mathcal{E}_0^{\mathrm{d}},\tilde{K}^{(r^{\mathrm{f}})},|\sigma^{\mathrm{d}}|^2T)\right)\Big]\,,\end{aligned}$$

where $\tilde{K}^{(r^{\mathrm{f}})} := Ke^{-(r^{\mathrm{d}}-r^{\mathrm{f}})T}$.

5. Of course the previous analysis may be performed symmetrically from the point of view of the foreign country. By (9.13) and (9.14), we have:

$$\mathcal{E}_t^{\mathrm{f}} \;=\; \mathcal{E}_0^{\mathrm{f}}e^{\left(\mu^{\mathrm{f}}-\frac{|\sigma^{\mathrm{f}}|^2}{2}\right)t+\sigma^{\mathrm{f}}W_t}\,, \qquad t\geq 0\,,$$

where

$$\mu^{\mathrm{f}} := -\mu^{\mathrm{d}}+|\sigma^{\mathrm{d}}|^2 \quad \text{and} \quad \sigma^{\mathrm{f}} := -\sigma^{\mathrm{d}}\,. \tag{9.15}$$

The foreign financial market risk neutral measure together with the corresponding Brownian motion are definied by

$$\frac{d\mathbb{Q}^{\mathrm{f}}}{d\mathbb{P}} \;=\; e^{-\lambda^{\mathrm{f}}W_T-\frac{1}{2}|\lambda^{\mathrm{f}}|^2T} \text{ on } \mathcal{F}_T \quad \text{and} \quad B_t^{\mathrm{f}} := W_t+\lambda^{\mathrm{f}}t\,,$$

where

$$\lambda^{\mathrm{f}} \quad := \quad \frac{\mu^{\mathrm{f}}+r^{\mathrm{d}}-r^{\mathrm{f}}}{\sigma^{\mathrm{f}}} \;=\; \lambda^{\mathrm{d}}-\sigma^{\mathrm{d}}$$

by (9.15).

9.2.5 The practice of the Black-Scholes model

The Black-Scholes model is used allover the industry of derivative securities. Its practical implementation requires the determination of the coefficients involved in the Black-Scholes formula. As we already observed the drift parameter μ is not needed for the pricing and hedging purposes. This surprising feature is easily understood by the fact that the perfect replication procedure involves the underlying probability measure only through its zero-measure sets, and therefore the problem is not changed by passage to any equivalent probability measure; by the Girsanov theorem this means that the problem is not modified by changing the drift μ in the dynamics of the risky asset.

Since the interest rate is observed, only the volatility parameter σ needs to be determined in order to implement the Black-Scholes formula. After discussing this important issue, we will focus on the different control variables which are carefully scrutinized by derivatives traders in order to account for the departure of real life financial markets from the simple Black-Scholes model.

Volatility: statistical estimation versus calibration

1. According to the Black-Scholes model, given the observation of the risky asset prices at times $t_i := ih$, $i = 1, \ldots, n$ for some time step $h > 0$, the returns

$$R_{t_i} := \ln\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right) \quad \text{are iid distributed as} \quad \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right)h\,,\,\sigma^2 h\right).$$

Then the sample variance

$$\hat{\sigma}_n^2 \;:=\; \frac{1}{n}\sum_{i=1}^{n}\left(R_{t_i} - \bar{R}_n\right)^2\,, \quad \text{where} \quad \bar{R}_n := \frac{1}{n}\sum_{i=1}^{n} R_{t_i}\,,$$

is the maximum likelihood estimator for the parameter σ^2. The estimator $\hat{\sigma}_n$ is called *the historical volatility* parameter.

The natural way to implement the Black-Scholes model is to plug the historical volatility into the Black-Scholes formula

$$\mathrm{BS}\left(S_t, \sigma, K, T\right) := S_t\mathbf{N}\left(\mathbf{d}_+\left(S_t, \tilde{K}, \sigma^2 T\right)\right) - \tilde{K}\mathbf{N}\left(\mathbf{d}_-\left(S_t, \tilde{K}, \sigma^2 T\right)\right) \quad (9.16)$$

to compute an estimate of the option price, and into the optimal hedge ratio

$$\Delta\left(S_t, \sigma, K, T\right) \;:=\; \mathbf{N}\left(\mathbf{d}_+\left(S_t, \tilde{K}, \sigma^2 T\right)\right) \quad (9.17)$$

in order to implement the optimal hedging strategy.

Unfortunately, the options prices estimates provided by this method performs very poorly in terms of fitting the observed data on options prices. Also, the use of the historical volatility for the hedging purpose leads to a very poor hedging strategy, as it can be verified by a back-testing procedure on observed data.

2. This anomaly is of course due to the simplicity of the Black-Scholes model which assumes that the log-returns are gaussian independent random variables. The empirical analysis of financial data reveals that securities prices exhibit fat tails which are by far under-estimated by the gaussian distribution. This is the so-called leptokurtic effect. It is also documented that financial data exhibits an important skewness, i.e. asymmetry of the distribution, which is not allowed by the gaussian distribution.

Many alternative statistical models have suggested in the literature in order to account for the empirical evidence (see e.g. the extensive literature on ARCH models). But none of them is used by the practioners on (liquid) options markets. The simple and by far imperfect Black-Scholes models is still used allover the financial industry. It is however the statistical estimation procedure that practitioners have gave up very early...

3. On liquid options markets, prices are given to the practitioners and are determined by the confrontation of demand and supply on the market. Therefore, their main concern is to implement the corresponding hedging strategy. To do this, they use the so-called *calibration* technique, which in the present context reduce to the calculation of the *implied volatility* parameter.

It is very easily checked that the Black-Scholes formula (9.16) is a one-to-one function of the volatility parameter, see (9.22) below. Then, given the observation of the call option price $C_t^*(K,T)$ on the financial market, there exists a unique parameter σ^{imp} which equates the observed option price to the corresponding Black-Scholes formula:

$$\text{BS}\left(S_t, \sigma_t^{\text{imp}}(K,T), K, T\right) \quad = \quad C_t^*(K,T)\,, \tag{9.18}$$

provided that C_t^* satisfies the no-arbirage bounds of Subsection 1.4. This defines, for each time $t \geq 0$, a map $(K,T) \longmapsto \sigma_t^{\text{imp}}(K,T)$ called the *implied volatility surface*. For their hedging purpose, the option trader then computes the hedge ratio

$$\Delta_t^{\text{imp}}(T,K) \quad := \quad \Delta\left(S_t, \sigma_t^{\text{imp}}(K,T), K, T\right)\,.$$

If the constant volatility condition were satisfied on the financial data, then the implied volatility surface would be expected to be flat. But this is not the case on real life financial markets. For instance, for a fixed maturity T, it is usually observed that that the implied volatility is U-shaped as a function of the strike price. Because of this empirical observation, this curve is called the *volatility smile*. It is also frequently argued that the smile is not symmetric but *skewed* in the direction of large strikes.

From the conceptual point of view, this practice of options traders is in contradiction with the basics of the Black-Scholes model: while the Black-Scholes formula is established under the condition that the volatility parameter is constant, the practical use via the implied volatility allows for a stochastic variation of the volatility. In fact, by doing this, the practioners are determining a *wrong volatility parameter out of a wrong formula* !

Despite all the criticism against this practice, it is the standard on the derivatives markets, and it does perform by far better than the statistical method. It has been widely extended to more complex derivatives markets as the fixed income derivatives, defaultable securities and related derivatives...

More details can be found in Chapter 10 is dedicated to the topic of implied volatility.

Risk control variables: the Greeks

With the above definition of the implied volatility, all the parameters needed for the implementation of the Black-Scholes model are available. For the purpose of controlling the risk of their position, the practitioners of the options markets various sensitivities, commonly called *Greeks*, of the Black-Scholes formula to the different variables and parameters of the model. The following picture shows a typical software of an option trader, and the objective of the following discussion is to understand its content.

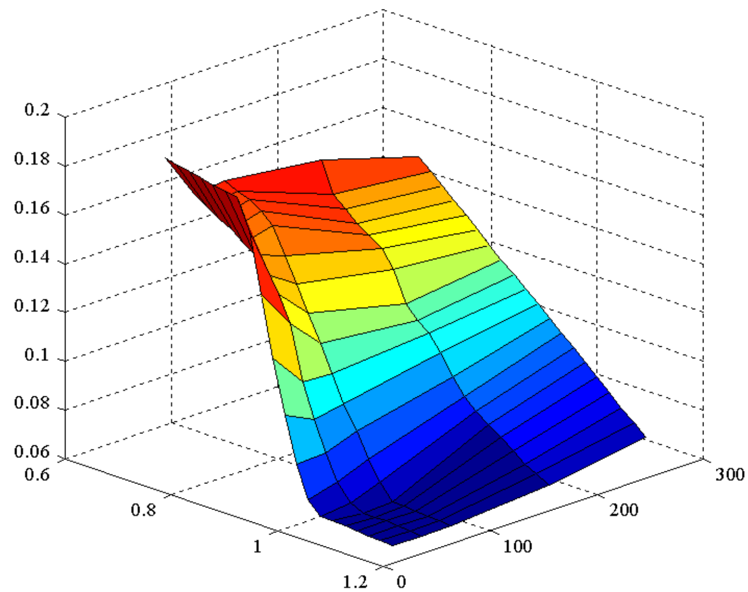

Figure 9.2: an example of implied volatility surface

1. *Delta*: This control variable is the most important one as it represents the number of shares to be held at each time in order to perform a perfect (dynamic) hedge of the option. The expression of the Delta is given in (9.17). An interesting observation for the calculation of this control variables and the subsequent ones is that

$$s\mathbf{N}'\left(\mathbf{d}_+(s,k,v)\right) \quad = \quad k\mathbf{N}'\left(\mathbf{d}_-(s,k,v)\right) ,$$

where $\mathbf{N}'(x) = (2\pi)^{-1/2}e^{-x^2/2}$.

2. *Gamma*: is defined by

$$\begin{aligned} \Gamma\left(S_t,\sigma,K,T\right) \quad &:= \quad \frac{\partial^2 \mathrm{BS}}{\partial s^2}\left(S_t,\sigma,K,T\right) \\ &= \quad \frac{1}{S_t\sigma\sqrt{T-t}}\mathbf{N}'\left(\mathbf{d}_+\left(S_t,\tilde{K},\sigma^2 T\right)\right) . \end{aligned} \tag{9.19}$$

The interpretation of this risk control coefficient is the following. While the simple Black-Scholes model assumes that the underlying asset price process is continuous, practitioners believe that large movemements of the prices, or jumps, are possible. A *stress* scenario consists in a sudden jump of the underlying asset price. Then the Gamma coefficient represent the change in the hedging strategy induced by such a stress scenario. In other words, if the underlying asset jumps

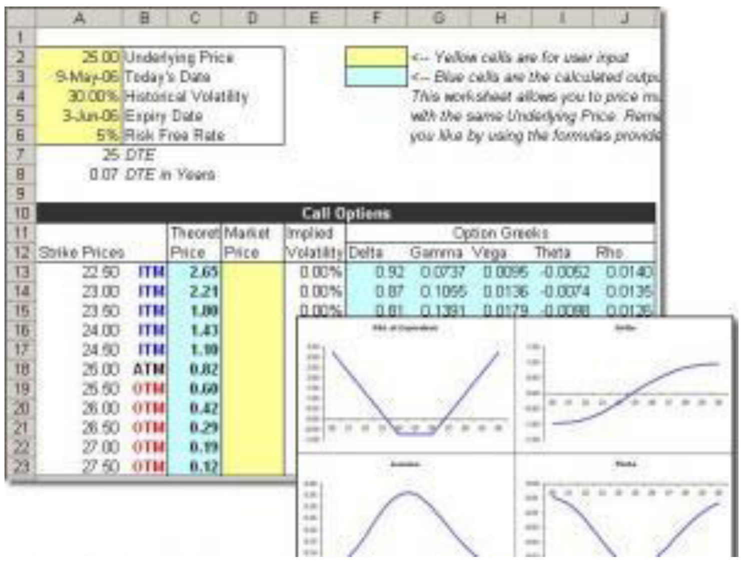

Figure 9.3: A typical option trader software

immediately from S_t to $S_t+\xi$, then the option hedger must immediately modify his position in the risky asset by buying $\Gamma_t \xi$ shares (or selling if $\Gamma_t \xi < 0$).

Given this interpretation, a position with a large Gamma is very risky, as it would require a large adjustment in case of a stress scenario.

3. *Rho*: is defined by

$$\begin{aligned} \rho\left(S_t, \sigma, K, T\right) &:= \frac{\partial \mathrm{BS}}{\partial r}\left(S_t, \sigma, K, T\right) \\ &= \tilde{K}(T-t)\mathbf{N}\left(\mathbf{d}_-\left(S_t, \tilde{K}, \sigma^2 T\right)\right), \end{aligned} \tag{9.20}$$

and represents the sensitivity of the Black-Scholes formula to a change of the instantaneous interest rate.

4. *Theta*: is defined by

$$\begin{aligned} \theta\left(S_t, \sigma, K, T\right) &:= \frac{\partial \mathrm{BS}}{\partial T}\left(S_t, \sigma, K, T\right) \\ &= \frac{1}{2} S_t \sigma \sqrt{T-t}\mathbf{N}'\left(\mathbf{d}_-\left(S_t, \tilde{K}, \sigma^2 T\right)\right), \end{aligned} \tag{9.21}$$

is also called the time value of the call option. This coefficient isolates the depreciation of the option when time goes on due to the maturity shortening.

5. *Vega*: is one of the most important Greeks (although it is not a Greek letter !), and is defined by

$$\begin{aligned}\mathcal{V}(S_t, \sigma, K, T) &:= \frac{\partial \mathrm{BS}}{\partial \sigma}(S_t, \sigma, K, T) \\ &= S_t\sqrt{T-t}\mathbf{N}'\left(\mathbf{d}_-\left(S_t, \tilde{K}, \sigma^2 T\right)\right). \end{aligned} \tag{9.22}$$

This control variable provides the exposition of the call option price to the volatility risk. Practitioners are of course aware of the stochastic nature of the volatility process (recall the smile surface above), and are therefore seeking a position with the smallest possible Vega in absolute value.

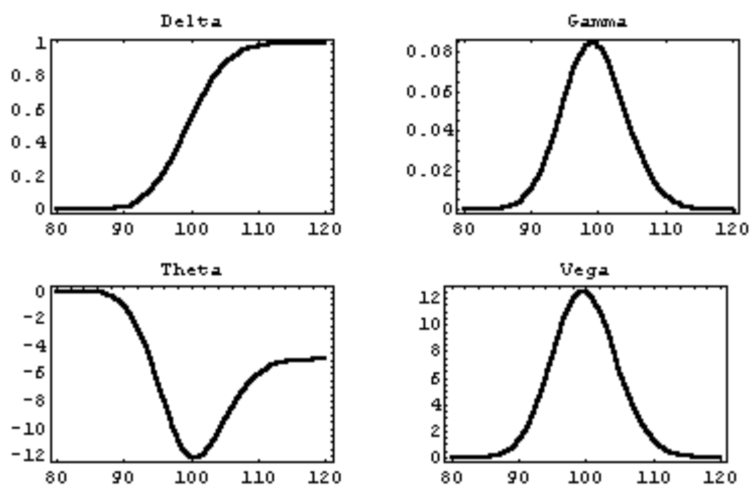

Figure 9.4: Representation of the Greeks

9.2.6 Hedging with constant volatility: robustness of the Black-Scholes model

In this subsection, we analyze the impact of a hedging strategy based on a constant volatility parameter Σ in a model where the volatility is stochastic:

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma_t dW_t. \tag{9.23}$$

Here, the volatility σ is a process in $\mathbb{H}^2$, the drift process μ is measurable adapted with $\int_0^T |\mu_u| du < \infty$, and B is the Brownian motion under the risk-neutral measure. We denote by r the instantaneous interest rate assumed to be constant.

Consider the position of a seller of the option who hedges the promised payoff $(S_T - K)^+$ by means of a self-financing portfolio based on the Black-Scholes hedging strategy with constant volatility Σ. Then, the discounted final value of the portfolio is:

$$\tilde{X}_T^{\Delta^{\mathrm{BS}}} := \mathrm{BS}(t, S_t, \Sigma)e^{-rt} + \int_t^T \Delta^{\mathrm{BS}}(u, S_u, \Sigma)d\tilde{S}_u$$

where $\mathrm{BS}(t, S_t, \Sigma)$ is the Black-Scholes formula parameterized by the relevant parameters for the present analysis, and $\Delta^{\mathrm{BS}} := \frac{\partial \mathrm{BS}}{\partial s}$. We recall that:

$$\left(\frac{\partial}{\partial t} + rs\frac{\partial}{\partial s} + \frac{1}{2}\Sigma^2 s^2 \frac{\partial^2}{\partial s^2} - r\cdot\right)\mathrm{BS}(t, s, \Sigma) = 0. \tag{9.24}$$

The *Profit and Loss* is defined by

$$\mathrm{P\&L}_T(\Sigma) \quad := \quad X_T^{\Delta^{\mathrm{BS}}} - (S_T - K)^+,$$

Since $\mathrm{BS}(T, s, \Sigma) = (s - K)^+$ independently of Σ:

$$e^{-rT}\ \mathrm{P\&L}_T(\Sigma) = \int_t^T \Delta^{\mathrm{BS}}(u, S_u, \Sigma)d\tilde{S}_u - \int_t^T d\{e^{-ru}\ \mathrm{BS}(u, S_u, \Sigma)\} \tag{9.25}$$

By the smoothness of the Black-Scholes formula, it follows from the Itô's formula and the (true) dynamics of the underlying security price process (9.23) that:

$$\begin{aligned}
d\,\mathrm{BS}(u, S_u, \Sigma) &= \Delta^{\mathrm{BS}}(u, S_u, \Sigma)dS_u + \left(\frac{\partial}{\partial t} + \frac{1}{2}s^2\sigma_u^2\frac{\partial^2}{\partial s^2}\right)\mathrm{BS}(u, S_u, \Sigma)du \\
&= \Delta^{\mathrm{BS}}(u, S_u, \Sigma)dS_u + \frac{1}{2}(\sigma_u^2 - \Sigma^2)S_u^2\Gamma^{\mathrm{BS}}(u, S_u, \Sigma)du \\
&\quad +(r\mathrm{BS} - rs\Delta^{\mathrm{BS}})(u, S_u, \Sigma)du,
\end{aligned}$$

where $\Gamma^{\mathrm{BS}} = \frac{\partial^2 \mathrm{BS}}{\partial s^2}$, and the last equality follows from (9.24). Plugging this expression in (9.25), we obtain:

$$\mathrm{P\&L}_T(\Sigma) \quad = \quad \frac{1}{2}\int_t^T e^{r(T-u)}(\Sigma^2 - \sigma_u^2)S_u^2\Gamma^{\mathrm{BS}}(u, S_u, \Sigma)du. \tag{9.26}$$

An interesting consequence of the latter beautiful formula is the following robustness property of the Black-Scholes model which holds true in the very general setting of the model (9.23).

Proposition 9.3. *Assume that $\sigma_t \le \Sigma$, $0 \le t \le T$, a.s. Then $\mathrm{P\&L}_T(\Sigma) \ge 0$, a.s., i.e. hedging the European call option within the (wrong) Black-Scholes model induces a super-hedging strategy for the seller of the option.*

Proof. It suffices to observe that $\Gamma^{\mathrm{BS}} \ge 0$. ♢

9.3 Complement: barrier options in the Black-Scholes model

So far, we have developed the pricing and hedging theory for the so-called *plain vanilla options* defined by payoffs $g(S_T)$ depending on the final value of the security at maturity. We now examine the example of barrier options which are the simplest representatives of the so-called *path-dependent options.*

A European barrier call (resp. put) option is a European call (resp. put) option which appears or disappears upon passage from some barrier. To simplify the presentation, we will only concentrate on European barrier call options. The corresponding definition for European barrier call options follow by replacing calls by puts.

The main technical tool for the derivation of explicit formulae for the no arbitrage prices of barrier options is the explicit form of the joint distribution of the Brownian motion W_t and its running maximum $W_t^* := \max_{s\le t} W_s$, see Proposition 4.13:

$$f_{W_t^*,W_t}(m,w) = \frac{2(2m-w)}{t\sqrt{2\pi t}} \exp\left(-\frac{(2m-w)^2}{2t}\right) \mathbf{1}_{\{m\ge 0\}}\mathbf{1}_{\{w\le m\}}$$

In our context, the risky asset price process is defined as an exponential of a drifted Brownian motion, i.e. the Black and Scholes model. For this reason, we need the following result.

Proposition 9.4. *For a given constant $a \in \mathbb{R}$, let $X_t = W_t + at$ and $X_t^* = \max_{s\le t} X_s$ the corresponding running maximum process. Then, the joint distribution of (X_t^*, X_t) is characterized by the density:*

$$f_{X_t^*,X_t}(y,x) = \frac{2(2y-x)}{t\sqrt{2\pi t}} \exp\left(ax - \frac{a^2}{2}t - \frac{(2y-x)^2}{2t}\right) \mathbf{1}_{\{y\ge 0\}}\mathbf{1}_{\{x\le y\}}$$

Proof. By the Cameron-Martin theorem, X is a Brownian motion under the probability measure $\mathbb{Q}$ with density

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-aW_t-\frac{1}{2}a^2t} = e^{-aX_t+\frac{1}{2}a^2t}.$$

Then,

$$\mathbb{P}\left[X_t^* \le y,\ X_t \le x\right] = \mathbb{E}^{\mathbb{Q}}\left[e^{aX_t-\frac{1}{2}a^2t}\mathbf{1}_{\{X_t^*\le y\}}\mathbf{1}_{\{X_t\le x\}}\right].$$

Differentiating, we see that

$$f_{X_t^*,X_t}(y,x) = e^{ax-\frac{1}{2}a^2t} f^{\mathbb{Q}}_{X_t^*,X_t}(y,x) = e^{ax-\frac{1}{2}a^2t} f_{W_t^*,W_t}(y,x),$$

where we denoted by $f^{\mathbb{Q}}_{X_t^*,X_t}$ the joint density under $\mathbb{Q}$ of the pair (X_t^*, X_t). $\diamondsuit$

9.3.1 Barrier options prices

We Consider a financial market with a non-risky asset S^0 defined by

$$S_t^0 = e^{rt}, \qquad t \geq 0,$$

and a risky security with price process defined by the Black and Scholes model

$$S_t = S_0 e^{(r-\frac{\sigma^2}{2})t+\sigma B_t}, \qquad t \geq 0,$$

where B is a Brownian motion under the risk neutral measure $\mathbb{Q}$.

An up-and-out call option is defined by the payoff at maturity T:

$$\mathrm{UOC}_T := (S_T - K)^+ \mathbf{1}_{\{\max_{0\leq t\leq T} S_t \leq B\}}$$

Introducing the parameters

$$a := \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right) \quad \text{and} \quad b = \frac{1}{\sigma}\log\left(\frac{B}{S_0}\right)$$

we may re-write the payoff of the up-and-out call option in:

$$\mathrm{UOC}_T = \left(S_0 e^{\sigma X_T} - K\right)^+ \mathbf{1}_{\{X_T^* \leq b\}} \quad \text{where} \quad X_t := W_t + at,\ t \geq 0,$$

and $X_t^* = \max_{0\leq u\leq t} X_u$, $t \geq 0$. The no-arbitrage price at time 0 of the up-and-out call is

$$\mathrm{UOC}_0 \;=\; \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\left(S_0 e^{\sigma X_T} - K\right)^+ \mathbf{1}_{\{X_T^* \leq b\}}\right].$$

We now show how to obtain an explicit formula for the up-and-out call option price in the present Black and Scholes framework.

a. By our general no-arbitrage valuation theory, together with the change of measure, it follows that

$$\begin{aligned}\mathrm{UOC}_0 &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\mathrm{UOC}_T\right] \\ &= S_0\hat{\mathbb{P}}\left[X_T \geq k\ ,\ X_T^* \leq b\right] - Ke^{-rT}\mathbb{Q}\left[X_T \geq k\ ,\ X_T^* \leq b\right] \quad (9.27)\end{aligned}$$

where we set:

$$k \;:=\; \frac{1}{\sigma}\log\left(\frac{K}{S_0}\right)$$

and $\hat{\mathbb{P}}$ is an equivalent probability measure defined by the density:

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{Q}} \;=\; e^{-rT}e^{\sigma X_T} = e^{-rT}\frac{S_T}{S_0}.$$

By the Cameron-Martin theorem, the process

$$\hat{W}_t = B_t - \sigma t, \qquad t \geq 0,$$

defines a Brownian motion under $\hat{\mathbb{P}}$. We introduce one more notation

$$\hat{a} := a + \sigma \quad \text{so that} \quad X_t \;=\; B_t + at \;=\; \hat{W}_T + \hat{a}t,\; t \geq 0. \tag{9.28}$$

b. Using the explicit joint distribution of (X_T, X_T^*) derived in Proposition 9.4, we compute that:

$$\begin{aligned}
\mathbb{Q}\left[X_T \leq k, X_T^* \leq b\right] &= \int_0^b \int_{-\infty}^{y \wedge k} \frac{2(2y-x)}{T\sqrt{2\pi T}} \exp\left(ax - \frac{1}{2}a^2T - \frac{(2y-x)^2}{2T}\right) dxdy \\
&= \int_{-\infty}^{k} \frac{e^{ax-\frac{1}{2}a^2T}}{\sqrt{2\pi T}} \int_{x^+}^{b} \frac{4(2y-x)}{2T} e^{-\frac{(2y-x)^2}{2T}} dy\, dx \\
&= \int_{-\infty}^{k} \frac{e^{ax-\frac{1}{2}a^2T}}{\sqrt{2\pi T}} \left[-e^{-(2b-x)^2/2T} + e^{-x^2/2T}\right] dx \\
&= \phi\left(\frac{k-aT}{\sqrt{T}}\right) - e^{2ab}\phi\left(\frac{k-at-2b}{\sqrt{T}}\right)
\end{aligned} \tag{9.29}$$

By (9.28), it follows that the second probability in (9.27) can be immediately deduced from (9.29) by substituting $\hat{a}$ to a:

$$\hat{\mathbb{P}}\left[X_T \leq k\;,\; X_T^* \leq b\right] \;=\; \phi\left(\frac{k-(a+\sigma)T}{\sqrt{T}}\right) - e^{2(a+\sigma)b}\phi\left(\frac{k-(a+\sigma)T-2b}{\sqrt{T}}\right)$$

c. We next compute

$$\begin{aligned}
\mathbb{Q}\left[X_T^* \leq b\right] &= 1 - \mathbb{Q}\left[X_T^* > b\right] \\
&= 1 - \mathbb{Q}\left[X_T^* > b, X_T < b\right] - \mathbb{Q}\left[X_T^* > b, X_T \geq b\right] \\
&= 1 - \mathbb{Q}\left[X_T^* > b, X_T < b\right] - \mathbb{Q}\left[X_T \geq b\right] \\
&= \mathbb{Q}\left[X_T^* \leq b, X_T \leq b\right] - \mathbb{Q}\left[X_T \geq b\right].
\end{aligned}$$

Then,

$$\begin{aligned}
\mathbb{Q}\left[X_T \geq k, X_T^* \leq b\right] &= \mathbb{Q}\left[X_T^* \leq b\right] - \mathbb{Q}\left[X_T \leq k, X_T^* \leq b\right] \\
&= \mathbf{N}\left(\frac{b-aT}{\sqrt{T}}\right) - e^{2ab}\mathbf{N}\left(-\frac{b+aT}{\sqrt{T}}\right) \\
&\quad -\mathbf{N}\left(\frac{k-aT}{\sqrt{T}}\right) + e^{2ab}\mathbf{N}\left(\frac{k-aT-2b}{\sqrt{T}}\right)
\end{aligned} \tag{9.30}$$

Similarly:

$$\begin{aligned}
\hat{\mathbb{P}}\left[X_T \geq k, X_T^* \leq b\right] &= \mathbf{N}\left(\frac{b-\hat{a}T}{\sqrt{T}}\right) - e^{2\hat{a}b}\mathbf{N}\left(-\frac{b+\hat{a}T}{\sqrt{T}}\right) \\
&\quad -\mathbf{N}\left(\frac{k-\hat{a}T}{\sqrt{T}}\right) + e^{2\hat{a}b}\mathbf{N}\left(\frac{k-\hat{a}T-2b}{\sqrt{T}}\right).
\end{aligned} \tag{9.31}$$

d. The explicit formula for the price of the up-and-out call option is then obtained by combining (9.27), (9.30) and (9.31).

e. An Up-and-in call option is defined by the payoff at maturity T:

$$\mathrm{UIC}_T := (S_T - K)^+ \mathbf{1}_{\{\max_{0\le t\le T} S_t \ge B\}}$$

The no-arbitrage price at time 0 of the up-and-in call is easily deduced from the explicit formula of the up-and-out call price:

$$\begin{aligned} \mathrm{UIC}_0 &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\left(S_0 e^{\sigma X_T} - K\right)^+ \mathbf{1}_{\{Y_T \ge b\}}\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\left(S_0 e^{\sigma X_T} - K\right)^+\right] - \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\left(S_0 e^{\sigma X_T} - K\right)^+ \mathbf{1}_{\{Y_T \le b\}}\right] \\ &= c_0 - \mathrm{U0C}_0, \end{aligned}$$

where c_0 is the Black-Scholes price of the corresponding European call option.

f. A down-and-out call option is defined by the payoff at maturity T:

$$\begin{aligned} \mathrm{DOC}_T &:= (S_T - K)^+ \mathbf{1}_{\{\min_{0\le t\le T} S_t \ge B\}} \\ &= \left(S_0 e^{\sigma X_T} - K\right)^+ \mathbf{1}_{\{\min_{0\le t\le T} X_t \ge b\}} \end{aligned}$$

Observe that the process $\{X_t = B_t + at, t \ge 0\}$ has the same distribution as the process $\{-x_t, t \ge 0\}$, where

$$x_t := B_t - at, \qquad t \ge 0.$$

Moreover $\min_{0\le t\le T} X_t$ has the same distribution as $-\max\limits_{0\le t\le T} x_t$. Then the no-arbitrage price at time 0 of the down-and-out call is

$$\mathrm{DOC}_0 = \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\left(S_0 e^{-\sigma x_T} - K\right)^+ \mathbf{1}_{\{\max_{0\le t\le T} x_t \le -b\}}\right].$$

We then can exploit the formula established above for the up-and-out call option after substituting $(-\sigma, -a, -b)$ to (σ, a, b).

g. A down-and-in call option is defined by the payoff at maturity T:

$$DIC_T := (S_T - K)^+ \mathbf{1}_{\{\min_{0\le t\le T} S_t \le B\}}$$

The problem of pricing the down-and-in call option reduces to that of the down-and-out call option:

$$\begin{aligned} \mathrm{DIC}_0 &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(S_T - K)^+ \mathbf{1}_{\{\min_{0\le t\le T} X_t \le b\}}\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(S_T - K)^+\right] - \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(S_T - K)^+ \mathbf{1}_{\{\min_{0\le t\le T} X_t \ge b\}}\right] \\ &= c_0 - \mathrm{DOC}_0, \end{aligned}$$

where c_0 is the Black-Scholes price of the corresponding European call option.

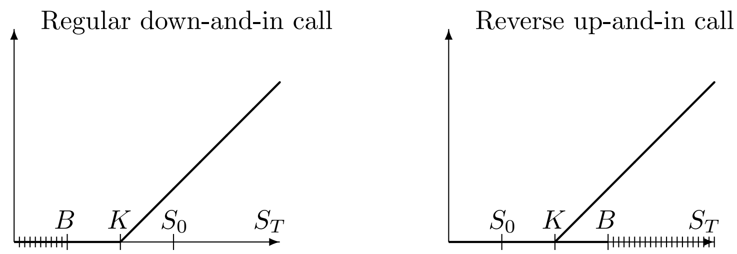

Figure 9.5: Types of barrier options.

9.3.2 Dynamic hedging of barrier options

We only indicate how the Black-Scholes hedging theory extends to the case of barrier options. We leave the technical details for the reader. If the barrier is hit before maturity, the barrier option value at that time is known to be either zero, or the price of the corresponding European option. Hence, it is sufficient to find the hedge before hitting the barrier $T_B \wedge T$ with

$$T_B := \inf \{t \geq 0 : \; S_t = B\} .$$

Prices of barrier options are smooth functions of the underlying asset price in the *in-region*, so Itô's formula may be applied up to the stopping time $T \wedge T_B$. By following the same line of argument as in the case of plain vanilla options, it then follows that perfect replicating strategy consists in:

$$\text{holding } \frac{\partial f}{\partial s}(t, S_t) \quad \text{shares of the underlying asset for} \quad t \leq T_B \wedge T$$

where $f(t, S_t)$ is price of the barrier option at time t.

9.3.3 Static hedging of barrier options

In contrast with European calls and puts, the delta of barrier options is not bounded, which makes these options difficult to hedge dynamically. We conclude this section by presenting a hedging strategy for barrier options, due to P. Carr et al. [9], which uses only static positions in European products.

A barrier option is said to be regular if its pay-off function is zero at and beyond the barrier, and reverse otherwise (see Figure 9.5 for an illustration).

In the following, we will treat barrier options with arbitrary pay-off functions (not necessarily calls or puts). The price of an Up and In barrier option which pays $f(S_T)$ at date T if the barrier B has been crossed before T will be denoted by $\mathrm{UI}_t(S_t, B, f(S_T), T)$, where t is the current date and S_t is the current stock price. In the same way, UO denotes the price of an Up and Out option and $\mathrm{EUR}_t(S_t, f(S_T), T)$ is the price of a European option with pay-off $f(S_T)$. These functions satisfy the following straightforward parity relations:

$$\begin{aligned}
&\mathrm{UI}_t + \mathrm{UO}_t = \mathrm{EUR}_t\\
&\mathrm{UI}_t(S_t, B, f(S_T), T) = \mathrm{EUR}_t(S_t, f(S_T), T) \quad \text{if } f(z) = 0 \text{ for } z < B\\
&\mathrm{UI}_t(S_t, B, f(S_T), T) = \mathrm{UI}_t(S_t, B, f(S_T)1_{S_T<B}, T)\\
&\qquad\qquad + \mathrm{EUR}_t(S_t, f(S_T)1_{S_T\geq B}, T) \quad \text{in general.}
\end{aligned}$$

This means that in order to hedge an arbitrary barrier option, it is sufficient to study options of type In Regular. In addition, Up and Down options can be treated in the same manner, so we shall concentrate on Up and In regular options.

The method is based on the following symmetry relationship:

$$\mathrm{EUR}_t(S_t, f(S_T), T) = \mathrm{EUR}_t\left(S_t, \left(\frac{S_T}{S_t}\right)^\gamma f\left(\frac{S_t^2}{S_T}\right), T\right), \tag{9.32}$$

with $\gamma = 1 - \frac{2(r-q)}{\sigma^2}$ where r is the interest rate, q the dividend rate and σ the volatility. It is easy to check that this relation holds in the Black-Scholes model, but the method also applies to other models which possess a similar symmetry property.

Replication of regular options Let f be the pay-off function of an Up and In regular option. This means that $f(z) = 0$ for $z \geq B$. We denote by T_B the first passage time by the price process above the level B. Consider the following static hedging strategy:

- At date t, buy the European option $\mathrm{EUR}_t\left(S_t, (\frac{S_T}{B})^\gamma f\left(\frac{B^2}{S_T}\right), T\right)$.
- When and if the barrier is hit, sell $\mathrm{EUR}_{T_B}\left(B, (\frac{S_T}{B})^\gamma f\left(\frac{B^2}{S_T}\right), T\right)$ and buy $\mathrm{EUR}_{T_B}(B, f(S_T), T)$. This transaction is costless by the symmetry relationship (9.32).

It is easy to check that this strategy replicates the option $\mathrm{UI}_t(S_t, f(S_T), T)$. As a by-product, we obtain the pricing formula:

$$\begin{aligned}
\mathrm{UI}_t(S_t, B, f(S_T), T) &= \mathrm{EUR}_t\left(S_t, \left(\frac{S_T}{B}\right)^\gamma f\left(\frac{B^2}{S_T}\right), T\right)\\
&= \left(\frac{S_t}{B}\right)^\gamma \mathrm{EUR}_t\left(S_t, f\left(\frac{B^2}{S_t^2}S_T\right), T\right).
\end{aligned} \tag{9.33}$$

The case of calls and puts Equation (9.33) shows that the price of a regular In option can be expressed via the price of the corresponding European option, for example,

$$\mathrm{UIP}_t(S_t, B, K, T) = \left(\frac{S_t}{B}\right)^{\gamma-2} p_t\left(S_t, \frac{KS_t^2}{B^2}, T\right).$$

However, unless $\gamma = 1$, the replication strategies will generally involve European payoffs other than calls or puts. If $\gamma = 1$ (that is, the dividend yield equals the risk-free rate), then regular In options can be statically replicated with a single call / put option. For example,

$$\begin{aligned}\mathrm{EUR}_t\left(S_t, \left(\frac{S_T}{B}\right)^{\gamma}\left(K - \frac{B^2}{S_T}\right)^+, T\right) &= \mathrm{EUR}_t\left(S_t, \left(\frac{KS_T}{B} - B\right)^+, T\right) \\ &= \frac{K}{B} c_t\left(S_t, \frac{B^2}{K}, T\right).\end{aligned}$$

The replication of reverse options will involve payoffs other than calls or puts even if $\gamma = 1$.

Chapter 10

Local volatility models and Dupire's formula

10.1 Implied volatility

In the Black-Scholes model the only unobservable parameter is the volatility. We therefore focus on the dependence of the Black-Scholes formula in the volatility parameter, and we denote:

$$C^{BS}(\sigma) \quad := \quad s\mathbf{N}\big(\mathbf{d}_+(s,\tilde{K},\sigma^2T)\big) - \tilde{K}\mathbf{N}\big(\mathbf{d}_-(s,\tilde{K},\sigma^2T)\big),$$

where $\mathbf{N}$ is the cumulative distribution function of the $\mathcal{N}(0,1)$ distribution, T is the time to maturity, s is the spot price of the underlying asset, and $\tilde{K}$, $\mathbf{d}_\pm$ are given in (9.5).

In this section, we provide more quantitative results on the volatility calibration discussed in Section 9.2.5. First, observe that the model can be calibrated from a single option price because the Black-Scholes price function is strictly increasing in volatility:

$$\lim_{\sigma\downarrow 0} C^{BS}(\sigma) = (s-\tilde{K})^+, \;\; \lim_{\sigma\uparrow\infty} C^{BS}(\sigma) = s, \text{ and } \frac{\partial C^{BS}}{\partial\sigma} = s\mathbf{N}'(\mathbf{d}_+)\sqrt{T} > 0 \tag{10.1}$$

不提前执行期权的另一个视角

Then, whenever the observed market price C of the call option lies within the no-arbitrage bounds:

$$(s-\tilde{K})^+ < C < s$$

there is a unique solution $I(C)$ to the equation

$$C^{BS}(\sigma) = C$$

called the *implied volatility* of this option. Direct calculation also shows that

$$\frac{\partial^2 C^{BS}}{\partial\sigma^2} = \frac{s\mathbf{N}'(\mathbf{d}_+)\sqrt{T}}{\sigma}\Big(\frac{m^2}{\sigma^2T} - \frac{\sigma^2T}{4}\Big) \quad \text{where} \quad m = \ln\big(\frac{s}{Ke^{-rT}}\big), \tag{10.2}$$

is the option moneyness. Equation (10.2) shows that the function $\sigma \mapsto C^{BS}(\sigma)$ is convex on the interval $(0, \sqrt{\frac{2|m|}{T-t}})$ and concave on $(\sqrt{\frac{2|m|}{T-t}}, \infty)$. Then the implied volatility can be approximated by means of the Newton's algorithm:

$$\sigma_0 = \sqrt{\frac{2m}{T}} \quad \text{and} \quad \sigma_n = \sigma_{n-1} + \frac{C - C^{BS}(\sigma_{n-1})}{\frac{\partial C^{BS}}{\partial \sigma}(\sigma_n)}$$

which produces a monotonic sequence of positive scalars $(\sigma_n)_{n\geq 0}$. However, in practice, when C is too close to the arbitrage bounds, the derivative $\frac{\partial C^{BS}}{\partial \sigma}(\sigma_n)$ becomes too small, leading to numerical instability. In this case, it is better to use the bisection method.

In the Black-Scholes model, the implied volatility of all options on the same underlying must be the same and equal to the historical volatility (standard deviation of annualized returns) of the underlying. However, when I is computed from market-quoted option prices, one observes that

- The implied volatility is always greater than the historical volatility of the underlying.

- The impied volatilities of different options on the same underlying depend on their strikes and maturity dates.

The left graph on Fig. 10.1 shows the implied volatilities of options on the S&P 500 index as function of their strike and maturity, observed on January 23, 2006. One can see that

- For almost all the strikes, the implied volatility is decreasing in strike (the *skew* phenomenon).

- For very large strikes, a slight increase of implied volatility can sometimes be observed (the *smile* phenomenon).

- The smile and skew are more pronounced for short maturity options; the implied volatility profile as function of strike flattens out for longer maturities.

The difference between implied volatility and historical volatility of the underlying can be explained by the fact that the cost of hedging an option in reality is actually higher than its Black-Scholes price, due, in particular to the transaction costs and the need to hedge the risk sources not captured by the Black-Scholes model (such as the volatility risk). The skew phenomenon is due to the fact that the Black-Scholes model underestimates the probability of a market crash or a large price movement in general. The traders correct this probability by increasing the implied volatilities of options far from the money. Finally, the smile can be explained by the liquidity premiums that are higher for far from the money options. The right graph in figure 10.1 shows that the implied volatilities of far from the money options are almost exclusively explained by the Bid prices that have higher premiums for these options because of a lower offer.

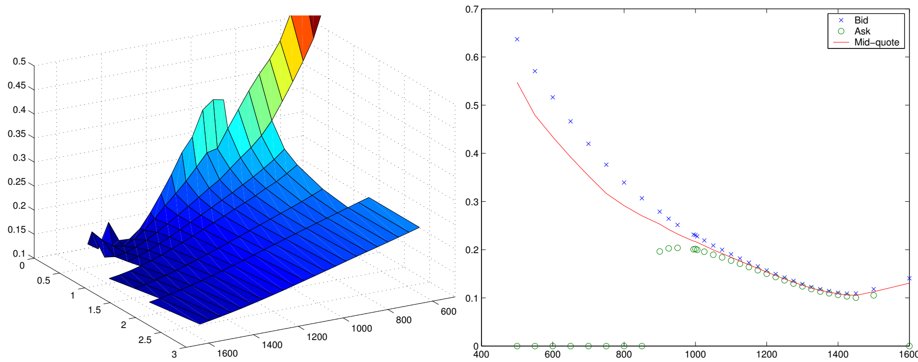

Figure 10.1: Left: Implied volatility surface of options on the S&P500 index on January 23, 2006. Right: Implied volatilities of Bid and Ask prices.

10.2 Local volatility models

In section 10.1 we saw that the Black-Scholes model with constant volatility cannot reproduce all the option prices observed in the market for a given underlying because their implied volatility varies with strike and maturity. To take into account the market implied volatility smile while staying within a Markovian and complete model (one risk factor), a natural solution is to model the volatility as a deterministic function of time and the value of the underlying:

$$\frac{dS_t}{S_t} = rdt + \sigma(t, S_t)dB_t, \tag{10.3}$$

where r is the interest rate, assumed to be constant, and B is the Brownian motion under the risk-neutral measure $\mathbb{Q}$. The SDE (10.3) defines a *local volatility model*.

We recall from Section 8.6 that the price of an option with payoff $h(S_T)$ at date T is given by

$$C(t,s) \quad = \quad \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}h(S_T)|S_t = s\right],$$

and is characterized by the partial differential equation:

$$rC = \frac{\partial C}{\partial t} + rs\frac{\partial C}{\partial s} + \frac{1}{2}\sigma(t,s)^2 s^2 \frac{\partial^2 C}{\partial s^2}, \quad C(T,s) = h(s). \tag{10.4}$$

The self-financing hedging portfolio contains $\delta_t = (\partial C/\partial s)(t, S_t)$ shares and the amount $\delta_t^0 = C(t, S_t) - \delta_t S_t$ in cash. The pricing equation has the same form as in the Black-Scholes model, but one can no longer deduce an explicit pricing formula, because the volatility is now a function of the underlying.

The naive way to use the model is to estimate the parameters, under the statistical measure $\mathbb{P}$, and to estimate the risk premium, typically by using historical data on options prices. This allows to specify completely the risk-neutral probability measure and the model (10.3).

The drawback of this approach is that option prices would then be completely determined by the historical model, and there is no hope for such an estimated model to produce exactly the observed prices of the quoted options. Consequently, the model can not be used under this form because it would immediately lead to arbitrage opportunities on the options market.

For this reason, practicioners have adopted a different approach which allows to use all observed prices of quoted options as an input for their pricing and hedging activities. This is the so-called *model calibration* approach. The parameters obtained by calibration are of course different from those which would be obtained by historical estimation. But this does not imply any problem related to the presence of arbitrage opportunities.

The model calibration approach is adopted in view of the fact that financial markets do not obey to any fundamental law except the simplest *no-dominance* or the slightly stronger no-arbitrage, see Section 1.2 below. There is no universally accurate model in finance, and any proposed model is wrong. Therefore, practitioners primarily base their strategies on comparison between assets, this is exactly what calibration does.

10.2.1 CEV model

A well studied example of a parametric local volatility model is provided by the CEV (Constant Elasticity of Variance) model [12]. In this model, the volatility is a power-law function of the level of the underlying. For simplicity, we formulate a CEV model on the forward price of the underlying $F_t = e^{r(T-t)}S_t$:

$$dF_t = \sigma_0 F_t^\alpha dB_t, \quad \text{for some} \quad \alpha \in (0,1], \tag{10.5}$$

together with the restriction that the left endpoint 0 is an absorbing boundary: if $F_t = 0$ for some t, $F_s \equiv 0$ for all $s \geq t$. The constraint $0 < \alpha \leq 1$ needs to be imposed to ensure that the above equation defines a martingale, see Lemma 10.2 below. We observe that one can show that for $\alpha > 1$, (F_t) is a strict local martingale, that is, not a true martingale. This can lead, for example, to Call-Put parity violation and other problems.

In the above CEV model, the volatility function $\sigma(f) := \sigma_0 f^\alpha$ has a constant elasticity:

$$\frac{f\sigma'(f)}{\sigma(f)} \quad = \quad \alpha.$$

The Black-Scholes model and the Gaussian model are particular cases of this formulation corresponding to $\alpha = 1$ and $\alpha = 0$ respectively. When $\alpha < 1$, the CEV model exhibits the so-called *leverage effect*, commonly observed in equity markets, where the volatility of a stock is decreasing in terms of the spot price of the stock.

For the equation (10.5) the existence and uniqueness of solution do not follow from the classical theory of strong solutions of stochastic differential equations, because of the non-Lipschitz nature of the coefficients. The following exercise shows that the existence of a weak solution can be shown by relating the CEV process with the so-called Bessel processes. We also refer to [17] for the proof of the existence of an equivalent martingale measure in this model.

Exercise 10.1. *Let W be a scalar Brownian motion. For $\delta \in \mathbb{R}$, $X_0 > 0$, we assume that there is a unique strong solution X to the SDE:*

$$dX_t \quad = \quad \delta dt + 2\sqrt{|X_t|}dW_t,$$

called the $\delta-$dimensional square Bessel process.

1. *Let $\hat{W}$ be a Brownian motion in $\mathbb{R}^d$, $d \geq 2$. Show that $\|\hat{W}\|^d$ is a $d-$dimensional square Bessel process.*

2. *Find a scalar power γ and a constant $a \in \mathbb{R}$ so that the process $Y_t := aX_t^\beta, t \geq 0$, is a CEV process satisfying (10.5), as long as X does not hit the origin.*

3. *Conversely, given a CEV process (10.5), define a Bessel process by an appripriate change of variable.*

Lemma 10.2. *For $0 < \alpha \leq 1$, let F be a solution of the SDE (10.5). Then F is a square-integrable martingale on $[0, T]$ for all $T < \infty$.*

Proof. It suffices to show that

$$\mathbb{E}^{\mathbb{Q}}\left\{\sigma_0^2 \int_0^T F_t^{2\alpha} dt\right\} < \infty. \tag{10.6}$$

Let $\tau_n = \inf\{t : F_t \geq n\}$. Then, $F_{T\wedge\tau_n}$ is square integrable and for all $0 < \alpha \leq 1$,

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[F^2_{\tau_n\wedge T}] &= \sigma_0^2 \mathbb{E}^{\mathbb{Q}}\left[\int_0^{\tau_n\wedge T} F_t^{2\alpha} dt\right] \\
&\leq \sigma_0^2 \mathbb{E}^{\mathbb{Q}}\left[\int_0^{\tau_n\wedge T} (1 + F_t^2) dt\right] \;\leq\; \sigma_0^2 \mathbb{E}^{\mathbb{Q}}\left[\int_0^T (1 + F^2_{t\wedge\tau_n}) dt\right]
\end{aligned}$$

By Gronwall's lemma we then get

$$\sigma_0^2 \mathbb{E}^{\mathbb{Q}}\left[\int_0^{\tau_n\wedge T} F_t^{2\alpha} dt\right] \quad = \quad \mathbb{E}^{\mathbb{Q}}[F^2_{\tau_n\wedge T}] \;\leq\; \sigma_0^2 T e^{\sigma_0^2 T},$$

and (10.6) now follows by monotone convergence. $\diamondsuit$

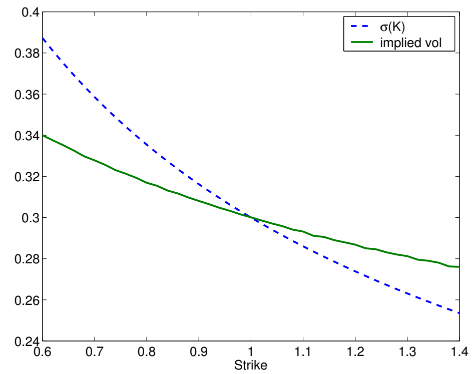

Figure 10.2: Skew (decreasing profile) of implied volatility in the CEV model with $\sigma = 0.3$, $\alpha = 0.5$, $S_0 = 1$ and $T = 1$

The shape of the implied volatility in the CEV model is known from the asymptotic approximation (small volatility) of Hagan and Woodward [25]:

$$\sigma^{imp}(K,T) \approx \frac{\sigma_0}{F_m^{1-\alpha}}\left\{1 + \frac{(1-\alpha)(2+\alpha)}{24}\left(\frac{F_0 - K}{F_m}\right)^2 + \frac{(1-\alpha)^2}{24}\frac{\sigma_0^2 T}{F_m^{2-2\alpha}} + \dots\right\},$$

$$F_m = \frac{1}{2}(F_0 + K).$$

To the first order, therefore, $\sigma^{imp}(K,T) \approx \frac{\sigma_0}{F_m^{1-\alpha}}$: the implied volatility has the same shape as local volatility but with an at the money slope which is two times smaller than that of the local volatility (see figure 10.2).

10.3 Dupire's formula

We now want to exploit the pricing partial differential equation (10.4) to deduce the local volatility function $\sigma(t,s)$ from observed call option prices for all strikes and all maturities. Unfortunately, equation (10.4) does not allow to reconstruct the local volatility from the formula

$$\sigma^2(t,s) \quad = \quad \frac{rC - \frac{\partial C}{\partial t} - rs\frac{\partial C}{\partial s}}{\frac{1}{2}s^2\frac{\partial^2 C}{\partial s^2}},$$

because at a given date, the values of t and s are fixed, and the corresponding partial derivatives cannot be evaluated. The solution to this problem was given by Bruno Dupire [21] who suggested a method for computing $\sigma(t,s)$ from the observed option prices for all strikes and maturities at a given date.

To derive the Dupire's equation, we need the following conditions on the local volatility model (10.3).

Assumption 10.3. *For all $x > 0$ and all small $\delta > 0$, there exists $\alpha > 0$ and a continuous function $c(t)$ such that:*

$$|x\sigma(t,x) - y\sigma(t,y)| \leq c(t)|x-y|^\alpha \quad \text{for} \quad |x-y| < \delta,\ t \in (t_0, \infty).$$

Theorem 10.4. *Let Assumption 10.3 hold true. Let $t_0 \geq 0$ be fixed and $(S_t)_{t_0 \leq t}$ be a square integrable solution of (10.3) with $\mathbb{E}^{\mathbb{Q}}\left[\int_{t_0}^t S_t^2 dt\right] < \infty$ for all $t \geq t_0$. Assume further that the random variable S_t has a continuous density $p(t,x)$ on $(t_0,\infty)\times(0,\infty)$. Then the call price function $C(T,K) = e^{-r(T-t_0)}\mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$ satisfies Dupire's equation*

$$\frac{\partial C}{\partial T} = \frac{1}{2}\sigma^2(T,K)K^2\frac{\partial^2 C}{\partial K^2} - rK\frac{\partial C}{\partial K}, \quad (T,K) \in [t_0,\infty)\times[0,\infty) \tag{10.7}$$

with the initial condition $C(t_0,K) = (S_{t_0} - K)^+$.

Proof. The proof is based on an application of Itô's formula to the process $e^{-rt}(S_t - K)^+$. Since the function $f(x) = x^+$ is not C^2, the usual Itô formula does not apply directly. A possible solution [22] is to use the Meyer-Itô formula for convex functions [35]. The approach used here, is instead to regularize the function f, making it suitable for the usual Itô formula. Introduce the function

$$f_\varepsilon(x) = \frac{(x+\varepsilon/2)^2}{2\varepsilon}\mathbf{1}_{-\varepsilon/2 \leq x \leq \varepsilon/2} + x\mathbf{1}_{x > \varepsilon/2}.$$

Notice that f_ε and f are equal outside the interval $[-\varepsilon/2, \varepsilon/2]$. Direct calculation provides:

$$f'_\varepsilon(x) = \frac{x+\varepsilon/2}{\varepsilon}\mathbf{1}_{-\varepsilon/2 \leq x \leq \varepsilon/2} + \mathbf{1}_{x>\varepsilon/2}, \quad \text{and for } 2|x| \neq \varepsilon, \quad f''_\varepsilon(x) = \frac{1}{\varepsilon}\mathbf{1}_{-\varepsilon/2 \leq x \leq \varepsilon/2}.$$

Then, we may apply Itô's formula (with generalized derivatives, see Remark 6.3) to $e^{-rt}f_\varepsilon(S_t - K)$ between T and $T + \theta$:

$$\begin{aligned} e^{-r(T+\theta)}f_\varepsilon(S_{T+\theta} - K) - e^{-rT}f_\varepsilon(S_T - K) &= -r\int_T^{T+\theta} e^{-rt}f_\varepsilon(S_t - K)dt \\ + \int_T^{T+\theta} e^{-rt}f'_\varepsilon(S_t - K)dS_t &+ \frac{1}{2}\int_T^{T+\theta} e^{-rt}f''_\varepsilon(S_t - K)\sigma^2(t,S_t)S_t^2 dt. \end{aligned} \tag{10.8}$$

The last term satisfies

$$\begin{aligned} &\int_T^{T+\theta} e^{-rt}f''_\varepsilon(S_t - K)\sigma^2(t,S_t)S_t^2 dt \\ &\qquad = \int_T^{T+\theta} dt e^{-rt}K^2\sigma^2(t,K)\frac{1}{\varepsilon}\mathbf{1}_{K-\varepsilon/2 \leq S_t \leq K+\varepsilon/2} \\ &\qquad\quad + \int_T^{T+\theta} dt e^{-rt}(S_t^2\sigma^2(t,S_t) - K^2\sigma^2(t,K))\frac{1}{\varepsilon}\mathbf{1}_{K-\varepsilon/2 \leq S_t \leq K+\varepsilon/2}. \end{aligned}$$

Using Assumption 10.3 above, the last term is dominated, up to a constant, by

$$\int_T^{T+\theta} dt e^{-rt} c(t) \frac{\varepsilon^\alpha}{\varepsilon} \mathbf{1}_{K-\varepsilon/2 \le S_t \le K+\varepsilon/2} \tag{10.9}$$

Taking the expectation of each term in (10.8) under the assumption 1, we find

$$\begin{aligned} e^{-r(T+\theta)} \mathbb{E}^{\mathbb{Q}}\big[f_\varepsilon(S_{T+\theta}-K)\big] - e^{-rT}\mathbb{E}^{\mathbb{Q}}\big[f_\varepsilon(S_T-K)\big] &= -r\int_T^{T+\theta} e^{-rt}\mathbb{E}^{\mathbb{Q}}\big[f_\varepsilon(S_t-K)\big]dt \\ &+ \int_T^{T+\theta} e^{-rt}\mathbb{E}^{\mathbb{Q}}\big[f'_\varepsilon(S_t-K)S_t\big] r dt \\ + \frac{1}{2}\int_T^{T+\theta} e^{-rt}K^2\sigma^2(t,K)\frac{1}{\varepsilon}\mathbb{E}^{\mathbb{Q}}\big[\mathbf{1}_{K-\varepsilon/2\le S_t\le K+\varepsilon/2}\big]dt + O(\varepsilon^\alpha), \end{aligned} \tag{10.10}$$

where the estimate $O(\varepsilon^\alpha)$ for the last term is obtained using (10.9) and the continuous density assumption. By the square integrability of S, we can pass to the limit $\varepsilon \to 0$:

$$\begin{aligned} &C(T+\theta,K) - C(T,K) \\ &= -r\int_T^{T+\theta} \mathbb{E}^{\mathbb{Q}}\big[(S_t-K)^+\big]dt + r\int_T^{T+\theta} e^{-rt}\mathbb{E}^{\mathbb{Q}}\big[S_t\mathbf{1}_{S_t\ge K}\big]dt \\ &\quad + \frac{1}{2}\int_T^{T+\theta} e^{-rt}\sigma^2(t,K)K^2 p(t,K)dt \\ &= rK\int_T^{T+\theta} e^{-rt}\mathbb{Q}\big[S_t \ge K\big]dt + \frac{1}{2}\int_T^{T+\theta} e^{-rt}\sigma^2(t,K)K^2p(t,K)dt. \end{aligned}$$

Dividing both sides by θ and passing to the limit $\theta \to 0$, this gives

$$\frac{\partial C}{\partial T} = rKe^{-rT}\mathbb{Q}\big[S_T \ge K\big] + \frac{1}{2}e^{-rT}\sigma^2(T,K)K^2p(T,K).$$

Finally, observing that

$$e^{-rT}\mathbb{Q}[S_T \ge K] = -\frac{\partial C}{\partial K} \quad \text{and} \quad e^{-rT}p(T,K) = \frac{\partial^2 C}{\partial K^2},$$

the proof of Dupire's equation is completed. ♢

The Dupire equation (10.7) can be used to deduce the volatility function $\sigma(\cdot,\cdot)$ from option prices. In a local volatility model, the volatility function σ can therefore be uniquely recovered via

$$\sigma(T,K) = \sqrt{2\frac{\frac{\partial C}{\partial T} + rK\frac{\partial C}{\partial K}}{K^2\frac{\partial^2 C}{\partial K^2}}} \tag{10.11}$$

Notice that the fact that one can find a unique continuous Markov process from European option prices does not imply that there are no other models (non-Markovian or discontinuous) that produce the same European option prices.

Knowledge of European option prices determines the marginal distributions of the process, but the law of the process is not limited to these marginal distributions.

Another feature of the Dupire representation is the following. Suppose that the true model (under the risk-neutral probability) can be written in the form

$$\frac{dS_t}{S_t} = rdt + \sigma_t dB_t,$$

where σ is a general adapted process, and not necessarily a deterministic function of the underlying. It can be shown that in this case the square of Dupire's local volatility given by equation (10.11) coincides with the expectation of the squared stochastic volatility conditioned by the value of the underlying:

$$\sigma^2(t,S) = \mathbb{E}^{\mathbb{Q}}[\sigma_t^2 | S_t = S].$$

Dupire's formula can therefore be used to find the Markovian diffusion which has the same marginal distributions as a given Itô martingale. In this sense, a local volatility surface can be seen as an arbitrage-free representation of a set of call prices for all strikes and all maturities just as the implied volatility represents the call price for a single strike and maturity.

Theorem 10.4 allows to recover the volatility coefficient starting from a complete set of call prices at a given date *if we know that these prices were produced by a local volatility model*. It does not directly allow to answer the following question: given a system of call option prices $(C(T,K))_{T\geq 0, K\geq 0}$, does there exist a continuous diffusion model reproducing these prices? To apply Dupire's formula (10.11), we need to at least assume that $\frac{\partial^2 C}{\partial K^2} > 0$ and $\frac{\partial C}{\partial T} + rK\frac{\partial C}{\partial K} \geq 0$. These constraints correspond to arbitrage constraints of the positivity of *butterfly spreads* and *calendar spreads* respectively.

A butterfly spread is a portfolio containing one call with strike $K-\Delta$, one call with strike $K+\Delta$ and a short position in two calls with strike K, where all the options have the same expiry date. Since the terminal pay-off of this portfolio is positive, its price must be positive at all dates: $C(K-\Delta)-2C(K)+C(K+\Delta) \geq 0$. This shows that the prices of call (and put) options are convex in strike, which implies $\frac{\partial^2 C}{\partial K^2} > 0$ if the price is twice differentiable and the second derivative remains strictly positive.

A (modified) calendar spread is a combination of a call option with strike K and maturity date $T+\Delta$ with a short position in a call option with strike $Ke^{-r\Delta}$ and maturity date T. The Call-Put parity implies that this portfolio has a positive value at date T; its value must therefore be positive at all dates before T: $C(K, T+\Delta) \geq C(Ke^{-r\Delta}, T)$. Passing to the limit $\Delta \downarrow 0$ we have, under differentiability assumption,

$$\lim_{\Delta\downarrow 0} \frac{C(K,T+\Delta) - C(Ke^{-r\Delta},T)}{\Delta} = \frac{\partial C}{\partial T} + rK\frac{\partial C}{\partial K} \geq 0.$$

Nevertheless, it may happen that the volatility $\sigma(t,s)$ exists but does not lead to a Markov process satisfying the three assumptions of theorem 10.4, for example, models with jumps in stock prices typically lead to explosive volatility surfaces, for which the SDE (10.3) does not have a solution.

10.3.1 Dupire's formula in practice

The figure 10.3 shows the results of applying Dupire's formula to artificially simulated data (left) and real prices of options on the S&P 500 index. While on the simulated data, the formula produces a smooth local volatility surface, its performance for real data is not satisfactory for several reasons:

- Market prices are not known for all strikes and all maturities. They must be interpolated and the final result is very sensitive to the interpolation method used.
- Because of the need to calculate the second derivative of the option price function $C(T, K)$, small data errors lead to very large errors in the solution (ill-posed problem).

Due to these two problems, in pracrice, Dupire's formula is not used directly on the market prices. To avoid solving the ill-posed problem, practitioners typically use one of two approaches:

- Start by a preliminary calibration of a parametric functional form to the implied volatility surface (for example, a function quadratic in strike and exponential in time may be used). With this smooth parametric function, recalculate option prices for all strikes, which are then used to calculate the local volatility by Dupire's formula.
- Reformulate Dupire's equation as an optimization problem by introducing a penalty term to limit the oscillations of the volatility surface. For example, Lagnado and Osher [31], Crepey [15] and other authors propose to minimize the functional

$$J(\sigma) \equiv \sum_{i=1}^{N} w_i (C(T_i, K_i, \sigma) - C_M(T_i, K_i))^2 + \alpha \|\nabla \sigma\|_2^2, \tag{10.12}$$

$$\|\nabla \sigma\|_2^2 \equiv \int_{K_{\min}}^{K_{\max}} dK \int_{T_{\min}}^{T_{\max}} dT \left\{ \left(\frac{\partial \sigma}{\partial K}\right)^2 + \left(\frac{\partial \sigma}{\partial T}\right)^2 \right\}, \tag{10.13}$$

where $C_M(T_i, K_i)$ is the market price of the option with strike K_i et expiry date T_i and $C(T_i, K_i, \sigma)$ corresponds to the price of the same option computed with the local volatility surface $\sigma(\cdot, \cdot)$.

10.3.2 Link between local and implied volatility

Dupire's formula (10.11) can be rewritten in terms of market implied volatilities, observing that for every option,

$$C(T, K) = C^{BS}(T, K, I(T, K)),$$

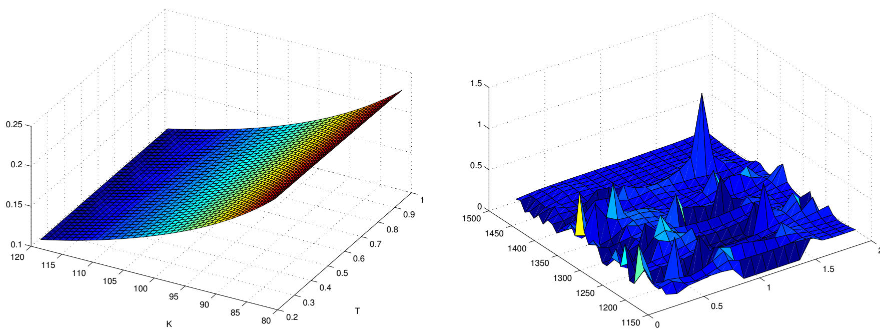

Figure 10.3: Examples of local volatility surface. Left: artificial data; the implied volatility is of the form $I(K) = 0.15 \times \frac{100}{K}$ for all maturities ($S_0 = 100$). Right: local volatility computed from S&P 500 option prices with spline interpolation.

where $C^{BS}(T, K, \sigma)$ denotes the Black-Scholes call price with volatility σ and $I(T, K)$ is the implied volatility for strike K and maturity date T.

Substituting this expression into Dupire's formula, we get

$$\begin{aligned}\sigma^2(T,K) &= 2\frac{\frac{\partial C^{BS}}{\partial T} + \frac{\partial C^{BS}}{\partial \sigma}\frac{\partial I}{\partial T} + rK\left(\frac{\partial C^{BS}}{\partial K} + \frac{\partial C^{BS}}{\partial \sigma}\frac{\partial I}{\partial K}\right)}{K^2\left(\frac{\partial^2 C^{BS}}{\partial K^2} + 2\frac{\partial^2 C^{BS}}{\partial K \partial \sigma}\frac{\partial I}{\partial K} + \frac{\partial^2 C^{BS}}{\partial \sigma^2}\left(\frac{\partial I}{\partial K}\right)^2 + \frac{\partial C^{BS}}{\partial \sigma}\frac{\partial^2 I}{\partial K^2}\right)} \\ &= \frac{\frac{I}{T} + 2\frac{\partial I}{\partial T} + 2rK\frac{\partial I}{\partial K}}{K^2\left(\frac{1}{K^2 IT} + 2\frac{\mathbf{d}_+}{KI\sqrt{T}}\frac{\partial I}{\partial K} + \frac{\mathbf{d}_+\mathbf{d}_-}{I}\left(\frac{\partial I}{\partial K}\right)^2 + \frac{\partial^2 I}{\partial K^2}\right)},\end{aligned} \tag{10.14}$$

with the usual notation

$$\mathbf{d}_\pm = \frac{\log\left(\frac{S}{Ke^{-rT}}\right) \pm \frac{1}{2}I^2 T}{I\sqrt{T}}.$$

Suppose first that the implied volatility does not depend on the strike (no smile). In this case, the local volatility is also independent from the strike and equation (10.14) is reduced to

$$\sigma^2(T) = I^2(T) + 2I(T)T\frac{\partial I}{\partial T},$$

and so

$$I^2(T) = \frac{\int_0^T \sigma^2(s)ds}{T}.$$

The implied volatility is thus equal to the root of mean squared local volatility over the lifetime of the option.

To continue the study of equation (10.14), let us make a change of variable to switch from the strike K to the log-moneyness variable $x = \log(S/\tilde{K})$, with $I(T,K) = J(T,x)$. The equation (10.14) becomes

$$2JT\frac{\partial J}{\partial T} + J^2 - \sigma^2\left(1 - \frac{x}{J}\frac{\partial J}{\partial x}\right)^2 - \sigma^2 JT\frac{\partial^2 J}{\partial x^2} + \frac{1}{4}\sigma^2 J^2 T^2\left(\frac{\partial J}{\partial x}\right)^2 = 0.$$

Assuming that I and its derivatives remain bounded when $T \to 0$, we obtain by sending T to 0:

$$J^2(0,x) = \sigma^2(0,x)\left(1 - \frac{x}{J}\frac{\partial J}{\partial x}\right)^2.$$

This differential equation can be solved explicitly:

$$J(0,x) = \left\{\int_0^1 \frac{dy}{\sigma(0,xy)}\right\}^{-1}. \tag{10.15}$$

We have thus shown that, *in the limit of very short time to maturity, the implied volatility is equal to the harmonic mean of local volatilities.* This result was established by Berestycki and Busca [8]. When the local volatility $\sigma(0,x)$ is differentiable at $x = 0$, equation (10.15) allows to prove that (the details are left to the reader)

$$\frac{\partial J(0,0)}{\partial x} = \frac{1}{2}\frac{\partial \sigma(0,0)}{\partial x}.$$

The slope of the local volatility at the money is equal, for short maturities, to twice the slope of the implied volatility.

This asymptotic makes it clear that the local volatility model, although it allows to calibrate the prices of all options on a given date, does not reproduce the dynamic behavior of these prices well enough. Indeed, the market implied volatility systematically flattens out for long maturities (see Figure 10.1), which results in the flattening of the local volatility surface computed from Dupire's formula. Assuming that the model is correct and that the local volatility surface remains constant over time, we therefore find that the ATM slope of the implied volatility for very short maturities should systematically decrease with time, a property which is not observed in the data. This implies that the local volatility surface cannot remain constant but must evolve with time: $\sigma(T,K) = \sigma_t(T,K)$, an observation which leads to *local stochastic volatility* models.

Chapter 11

Backward SDEs and funding problems

The recent financial crisis motivated important deviations from the frictionless market model developed in the previous chapter. An important aspect highlighted by the financial crisis is the importance of the funding needed for the implementation of the hedging strategy, and the liquidity of the underlying securities. In the simple frictionless model, the is one single interest rate which serves both for borrowing and lending the non-risky asset (i.e. the cash). In real financial markets, the situation is of course drastically different as the lending and the borrowing rates are significantly different. Also, the there is no cost for holding the underlying asset and, in particular, shorting the security is costless. In practice, shorting the security requires a financial contract which allows to borrow it for a "renting price" referred to as the *repo*.

The objective of this chapter is to account for the last market imperfections which gained a primary importance after the recent financial crisis. In order to develop the corresponding models, we need to develop a nonlinear pricing theory motivated by the underlying financial engineering techniques. This is in contrast to the linear Black-Scholes model derived in the previous chapters. As usual, we start by developing the required mathematical tools, and we return to the financial application in the last sections.

11.1 Preliminaries: the BDG inequality

In this section, we consider a one-dimensional Brownian motion W, and we introduce a local martingale

$$M_t := \int_0^t H_s dW_s, \ t \geq 0, \quad \text{for some} \quad H \in \mathbb{H}^2_{\text{loc}}. \tag{11.1}$$

The corresponding quadratic variation process is $\langle M \rangle_t := \int_0^t H_s^2 ds$, and we denote $M_t^* := \max_{u \leq t} |M_u|$, $t \geq 0$, the running maximum process of the non-negative local submartingale $|M|$.

If it happens that M is a martingale, then it follows from the Doob's martingale inequality of Theorem 3.15 that:

$$\mathbb{E}\big[(M_T^*)^p\big] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[M_T^p\big] \quad \text{for all} \quad p>1. \tag{11.2}$$

In particular, for $p=2$, we have $M_T^2 = 2\int_0^T M_s dM_s + \langle M\rangle_T$, and it follows from a direct localization argument that $\mathbb{E}\big[M_T^2\big] = \mathbb{E}\big[\langle M\rangle_T\big]$, Then the Doob's martingale inequality reduces to

$$\mathbb{E}\big[(M_T^*)^2\big] \quad \leq \quad 4\,\mathbb{E}\big[\langle M\rangle_T\big].$$

Our objective in this section is to extend the last inequality to an arbitrary power $p>0$, so as to obtain the so-called Burkholder-Davis-Gundy (BDG) inequality:

$$\mathbb{E}\big[(M_T^*)^p\big] \;\leq\; C_p\,\mathbb{E}\big[\langle M\rangle_T^{p/2}\big] \quad \text{for all} \quad p>0,$$

where $C_p>0$ is a universal constant which only depends on p. We observe that there are also universal constants $c_p>0$ such that

$$c_p\,\mathbb{E}\big[\langle M\rangle_T^{p/2}\big] \;\leq\; \mathbb{E}\big[(M_T^*)^p\big] \quad \text{for all} \quad p>0,$$

but we shall not prove the last inequality as it will not be used in our applications to finance.

11.1.1 The smooth power case

The power function $x \longmapsto |x|^p$ is C^2 for $p \geq 2$. Since the Doob's martingale inequality holds true in this case, the main ingredient for the following statement is to derive an upper bound for $\mathbb{E}\big[M_T^p\big]$ in terms of the quadratic variation process. This is naturally obtained by using the Itô differential calculus.

Proposition 11.1. *For M as in* (11.1)*, and $p \geq 2$, there exists a constant C_p only depending on p such that*

$$\mathbb{E}\big[(M_T^*)^p\big] \quad \leq \quad C_p\,\mathbb{E}\big[\langle M\rangle_T^{p/2}\big].$$

Proof. By direct localization, we may assume without loss of generality that M is bounded. Notice that the function $x \longmapsto f(x) := |x|^p$ is C^2. Then, it follows from Itô's formula that $M_T^p = \int_0^T f'(M_s)dM_s + \frac{1}{2}\int_0^T f''(M_s)d\langle M\rangle_s$. Taking expected values, this implies that:

$$\mathbb{E}\big[M_T^p\big] = \frac{1}{2}p(p-1)\mathbb{E}\Big[\int_0^T |M_s|^{p-2}d\langle M\rangle_s\Big] \leq \frac{1}{2}p(p-1)\mathbb{E}\Big[(M_T^*)^{p-2}\langle M\rangle_T\Big].$$

We next use the Hölder inequality to obtain

$$\mathbb{E}\big[M_T^p\big] \leq \frac{1}{2}p(p-1)\mathbb{E}\Big[(M_T^*)^p\Big]^{\frac{p-2}{p}}\,\mathbb{E}\Big[\langle M\rangle_T^{\frac{p}{2}}\Big]^{\frac{2}{p}},$$

and the required result follows from the Doob's martingale inequality (11.2). ♢

11.1.2 The case of an arbitrary power

Theorem 11.2. *For M as in* (11.1)*, and $p > 0$, there exists a constant C_p only depending on p such that*

$$\mathbb{E}\big[(M_T^*)^p\big] \quad \le \quad C_p\, \mathbb{E}\big[\langle M\rangle_T^{p/2}\big].$$

Proof. Given the result of Proposition 11.1, it only remain to address the case $p \in (0,2)$. Let $X := (M^*)^2$ and $A := C_2\langle M\rangle$. Our starting point is the observation that

$$A, X \text{ are continuous nondecreasing, and } \mathbb{E}[X_\tau] \le \mathbb{E}[A_\tau] \text{ for all } \tau \in \mathcal{T}_b, \tag{11.3}$$

where $\mathcal{T}_b$ is the collection of all bounded stopping times, and the last inequality is precisely the statement of Proposition 11.1 with $p = 2$. We organize the proof in two steps.

1. For $a, x > 0$, let $\tau := \inf\{t \ge 0 : A_t \ge a\}$, and we compute that

$$\begin{aligned}
\mathbb{P}\big[X_T \ge x, A_t < a\big] \quad &= \quad \mathbb{P}\big[X_T \ge x, T < \tau\big] \\
&\le \quad \mathbb{P}\big[X_{T\wedge\tau} \ge x\big] \le \frac{\mathbb{E}\big[X_{T\wedge\tau}\big]}{x} \le \frac{\mathbb{E}\big[A_{T\wedge\tau}\big]}{x} = \frac{\mathbb{E}\big[a \wedge A_T\big]}{x},
\end{aligned}$$

where we used the non-decrease of A and X as stated in (11.3).

2. For $0 < k < 1$, it follows from the trivial identity $y^k = \int_0^\infty \mathbf{1}_{\{y \ge x\}} k x^{k-1} dx$ that

$$\begin{aligned}
\mathbb{E}[(X_T)^k] \quad &= \quad k\int_0^\infty \mathbb{P}[X_T \ge x] x^{k-1} dx \\
&= \quad k\int_0^\infty \big(\mathbb{P}[X_T \ge x, A_T \ge x] + \mathbb{P}[X_T \ge x, A_T < x]\big) x^{k-1} dx \\
&\le \quad k\int_0^\infty \big(\mathbb{P}[A_T \ge x] + \mathbb{P}[X_T \ge x, A_T < x]\big) x^{k-1} dx \\
&\le \quad k\int_0^\infty \big(\mathbb{P}[A_T \ge x] + x^{-1}\mathbb{E}[x \wedge A_T]\big) x^{k-1} dx,
\end{aligned}$$

by the inequality derived in the first step. Since $\mathbb{E}[x \wedge A_T] = x\mathbb{P}[x \le A_T] + \mathbb{E}\big[A_T \mathbf{1}_{\{A_T < x\}}\big]$, this provides:

$$\mathbb{E}[(X_T)^k] \quad \le \quad k\int_0^\infty \Big(2\mathbb{P}[A_T \ge x] + \frac{1}{x}\mathbb{E}\big[A_T \mathbf{1}_{\{A_T < x\}}\big]\Big) x^{k-1} dx \;=\; \frac{2-k}{1-k}\mathbb{E}\big[(A_T)^k\big].$$

Returning to our original notations, the last inequality translates into:

$$\mathbb{E}[(M_T^*)^{2k}] \quad \le \quad \frac{2-k}{1-k}\mathbb{E}\big[(C_2\langle M_T\rangle)^k\big],$$

which is the required inequality for $p = 2k \in (0,2)$. $\Diamond$

11.2 Backward SDEs

Throughout this section, we consider a $d-$dimensional Brownian motion W on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we denote by $\mathbb{F} = \mathbb{F}^W$ the corresponding augmented filtration.

Given two integers $n, d \in \mathbb{N}$, we consider the mapping

$$f : [0, T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n \times d} \quad \longrightarrow \quad \mathbb{R}.$$

We assume that the process $\{f_t(y, z), t \in [0, T]\}$ is $\mathbb{F}-$progressively measurable, for every fixed $(y, z) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$.

Our objective in this section is to find a pair process (Y, Z) satisfying the following backward stochastic differential equation (BSDE):

$$dY_t = -f_t(Y_t, Z_t)dt + Z_t dW_t \quad \text{and} \quad Y_T = \xi, \ \mathbb{P}-\text{a.s.} \tag{11.4}$$

where ξ is some given $\mathcal{F}_T-$measurable r.v. with values in $\mathbb{R}^n$.

We will refer to (11.4) as BSDE(f, ξ). The map f is called the *generator*. We may also rewrite the BSDE (11.4) in the integrated form:

$$Y_t = \xi + \int_t^T f_s(Y_s, Z_s)ds - \int_t^T Z_s dW_s, \quad t \leq T, \quad , \mathbb{P}-\text{a.s.} \tag{11.5}$$

11.2.1 Martingale representation for zero generator

When the generator $f \equiv 0$, the BSDE problem reduces to the martingale representation theorem in the present Brownian filtration. More precisely, for every $\xi \in \mathbb{L}^2(\mathbb{R}^n, \mathcal{F}_T)$, there is a unique pair process (Y, Z) in $\mathbb{H}^2(\mathbb{R}^n \times \mathbb{R}^{n \times d})$ satisfying (**??**):

$$\begin{aligned} Y_t \ := \ \mathbb{E}[\xi|\mathcal{F}_t] \ &= \ \mathbb{E}[\xi] + \int_0^t Z_s dW_s \\ &= \ \xi - \int_t^T Z_s dW_s. \end{aligned}$$

Here, for a subset E of $\mathbb{R}^k$, $k \in \mathbb{N}$, we denoted by $\mathbb{H}^2(E)$ the collection of all $\mathbb{F}-$progressively measurable $\mathbb{L}^2([0, T] \times \Omega, \text{Leb} \otimes \mathbb{P})-$processes with values in E. We shall frequently simply write $\mathbb{H}^2$ keeping the reference to E implicit.

Let us notice that $\{Y_t, t \in [0, T]\}$ is a martingale. Moreover, by the Doob's maximal inequality, we have:

$$\|Y\|^2_{\mathcal{S}^2} \ := \ \mathbb{E}\left[\sup_{t \leq T} |Y_t|^2\right] \ \leq \ 4\mathbb{E}\left[|Y_T|^2\right] \ = \ 4\|Z\|^2_{\mathbb{H}^2}. \tag{11.6}$$

Hence, the process Y is in the space $\mathcal{S}^2-$ of continuous processes with finite $\mathcal{S}^2-$norm.

11.2.2 BSDEs with affine generator

We next consider a scalar BSDE ($n = 1$) with generator

$$f_t(y, z) \quad := \quad a_t + b_t y + c_t \cdot z, \tag{11.7}$$

where a, b, c are $\mathbb{F}-$progressively measurable processes with values in $\mathbb{R}$, $\mathbb{R}$ and $\mathbb{R}^d$, respectively. We also assume that b, c are bounded and $\mathbb{E}[\int_0^T |a_t|^2 dt] < \infty$. This case is easily handled by reducing to the zero generator case.

We first introduce the equivalent probability $\mathbb{Q} \sim \mathbb{P}$ defined by the density

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \quad = \quad \exp\left(\int_0^T c_t \cdot dW_t - \frac{1}{2}\int_0^T |c_t|^2 dt\right).$$

By the Girsanov theorem, the process $B_t := W_t - \int_0^t c_s ds$ defines a Brownian motion under $\mathbb{Q}$. Then, by viewing the BSDE under $\mathbb{Q}$, we are reduced to the case where the generator does not depend on z:

$$dY_t \quad = \quad -(a_t + b_t Y_t)dt + Z_t \cdot dB_t.$$

Similarly, the linear term in y can be easily by-passed by considering the change of variable:

$$\overline{Y}_t := Y_t e^{\int_0^t b_s ds} \quad \text{so that} \quad d\overline{Y}_t = -a_t e^{\int_0^t b_s ds} dt + Z_t e^{\int_0^t b_s ds} dB_t.$$

Finally, defining

$$\overline{\overline{Y}}_t := \overline{Y}_t + \int_0^t a_u e^{\int_0^u b_s ds} du,$$

we arrive at a BSDE with zero generator for $\overline{\overline{Y}}_t$ which can be solved by the martingale representation theorem under the equivalent probability measure $\mathbb{Q}$.

Of course, one can also express the solution under $\mathbb{P}$:

$$Y_t \quad = \quad \mathbb{E}\left[\Gamma_T^t \xi + \int_t^T \Gamma_s^t a_s ds \Big| \mathcal{F}_t\right], \ t \leq T,$$

where

$$\Gamma_s^t := \exp\left(\int_t^s b_u du - \frac{1}{2}\int_t^s |c_u|^2 du + \int_t^s c_u \cdot dW_u\right), \quad 0 \leq t \leq s \leq T. \tag{11.8}$$

11.2.3 The main existence and uniqueness result

Similar to the case of forward stochastic differential equation, we now prove an existence and uniqueness result for BSDEs by means of a fixed point argument for the Picard iteration procedure.

Theorem 11.3. *Assume that* $\{f_t(0,0), t \in [0,T]\} \in \mathbb{H}^2$ *and, for some constant* $C > 0$,

$$|f_t(y,z) - f_t(y',z')| \le C(|y-y'| + |z-z'|), \quad dt \otimes d\mathbb{P} - a.s.$$

for all $t \in [0,T]$ *and* $(y,z),(y',z') \in \mathbb{R}^n \times \mathbb{R}^{n\times d}$*. Then, for every* $\xi \in \mathbb{L}^2$*, there is a unique solution* $(Y,Z) \in \mathcal{S}^2 \times \mathbb{H}^2$ *to the BSDE*(f,ξ).

Proof. Denote $S = (Y,Z)$, and introduce the equivalent norm in the corresponding $\mathbb{H}^2$ space:

$$\|S\|_\alpha \quad := \quad \mathbb{E}\left[\int_0^T e^{\alpha t}(|Y_t|^2 + |Z_t|^2)dt\right].$$

where α will be fixed later. We consider the operator

$$\phi : s = (y,z) \in \mathbb{H}^2 \quad \longmapsto \quad S^s = (Y^s, Z^s)$$

defined by:

$$Y_t^s \quad = \quad \xi + \int_t^T f_u(y_u, z_u)du - \int_t^T Z_u^s \cdot dW_u, \quad t \le T.$$

This defines the Picard iteration operator which consists of a BSDE with constant (random) generator independent of the unknown variables (Y^s, Z^s).

1. First, since $|f_u(y_u,z_u)| \le |f_u(0,0)| + C(|y_u| + |z_u|)$, it follows from our assumption on the generator together with the fact that $s \in \mathbb{H}^2$ that the process $\{f_u(y_u,z_u), u \le T\}$ is in $\mathbb{H}^2$. Then $S^s = \phi(s) \in \mathbb{H}^2$ is well-defined by the martingale representation theorem, as outlined in the previous subsection. Moreover,

$$\begin{aligned}
\mathbb{E}\Big[\sup_{t\le T}|Y_t|^2\Big] \quad &\le \quad 3\mathbb{E}\Big[|\xi|^2 + \int_0^T |f_t(y_t,z_t)|^2 dt + \sup_{t\le T}\Big|\int_0^t Z_s dW_s\Big|^2\Big] \\
&\le \quad 3\Big(|\xi|^2_{\mathbb{L}^2} + \|f_t(0,0)\|^2_{\mathbb{H}^2} + C\|s\|^2_{\mathbb{H}^2} + \mathbb{E}\Big[\sup_{t\le T}\Big|\int_0^t Z_s dW_s\Big|^2\Big]\Big) \\
&\le \quad 3\Big(|\xi|^2_{\mathbb{L}^2} + \|f_t(0,0)\|^2_{\mathbb{H}^2} + C\|s\|^2_{\mathbb{H}^2} + 4\|Z\|_{\mathbb{H}^2}\Big)
\end{aligned}$$

by the Doob's martingale inequality, together with the Itô isometry. Hence $Y \in \mathcal{S}^2$.

2. For $s, s' \in \mathbb{H}^2$, denote $\delta s := s - s'$, $\delta S := S^s - S^{s'}$ and $\delta f := f_t(S^s) - f_t(S^{s'})$. Since $\delta Y_T = 0$, it follows from Itô's formula that:

$$\begin{aligned}
e^{\alpha t}|\delta Y_t|^2 + \int_t^T e^{\alpha u}|\delta Z_u|^2 du \quad &= \quad \int_t^T e^{\alpha u}\left(2\delta Y_u \cdot \delta f_u - \alpha|\delta Y_u|^2\right) du \\
&\quad -2\int_t^T e^{\alpha u}(\delta Z_u)^{\mathrm{T}}\delta Y_u \cdot dW_u.
\end{aligned}$$

In the remaining part of this step, we prove that

$$M_t := \int_0^t e^{\alpha u}(\delta Z_u)^{\mathrm{T}}\delta Y_u \cdot dW_u,\ t \in [0,T], \quad \text{is a martingale,} \tag{11.9}$$

so that we deduce from the previous equality that

$$\mathbb{E}\left[e^{\alpha t}|\delta Y_t|^2 + \int_t^T e^{\alpha u}|\delta Z_u|^2 du\right] = \mathbb{E}\left[\int_t^T e^{\alpha u}\left(2\delta Y_u \cdot \delta f_u - \alpha|\delta Y_u|^2\right) du\right]. \tag{11.10}$$

To prove (11.9), we verify that $\sup_{t\le T}|M_t| \in \mathbb{L}^1$. Indeed, by the Burkholder-Davis-Gundy inequality of Theorem 11.2, we have:

$$\begin{aligned}
\mathbb{E}\Big[\sup_{t\le T}|M_t|\Big] &\le C\mathbb{E}\left[\left(\int_0^T e^{2\alpha u}|\delta Y_u|^2|\delta Z_u|^2 du\right)^{1/2}\right] \\
&\le C'\mathbb{E}\left[\sup_{u\le T}|\delta Y_u|\left(\int_0^T |\delta Z_u|^2 du\right)^{1/2}\right] \\
&\le \frac{C'}{2}\left(\mathbb{E}\left[\sup_{u\le T}|\delta Y_u|^2\right] + \mathbb{E}\left[\int_0^T |\delta Z_u|^2 du\right]\right) < \infty.
\end{aligned}$$

3. We now continue estimating (11.10) by using the Lipschitz property of the generator, we see that for any arbitrary parameter $\varepsilon > 0$:

$$\begin{aligned}
\mathbb{E}\Big[e^{\alpha t}|\delta Y_t|^2 &+ \int_t^T e^{\alpha u}|\delta Z_u|^2 du\Big] \\
&\le \mathbb{E}\Big[\int_t^T e^{\alpha u}\left(-\alpha|\delta Y_u|^2 + C2|\delta Y_u|(|\delta y_u| + |\delta z_u|)\right) du\Big] \\
&\le \mathbb{E}\left[\int_t^T e^{\alpha u}\left(-\alpha|\delta Y_u|^2 + C\big(\varepsilon^2|\delta Y_u|^2 + \varepsilon^{-2}(|\delta y_u| + |\delta z_u|)^2\big)\right) du\right].
\end{aligned}$$

Choosing $C\varepsilon^2 = \alpha$, we obtain:

$$\begin{aligned}
\mathbb{E}\left[e^{\alpha t}|\delta Y_t|^2 + \int_t^T e^{\alpha u}|\delta Z_u|^2 du\right] &\le \mathbb{E}\left[\int_t^T e^{\alpha u}\frac{C^2}{\alpha}(|\delta y_u| + |\delta z_u|)^2 du\right] \\
&\le 2\frac{C^2}{\alpha}\|\delta s\|_\alpha^2.
\end{aligned}$$

This provides

$$\|\delta Z\|_\alpha^2 \le 2\frac{C^2}{\alpha}\|\delta s\|_\alpha^2 \quad \text{and} \quad \|\delta Y\|_\alpha^2 \le 2\frac{C^2 T}{\alpha}\|\delta s\|_\alpha^2$$

where we abused notation by writing $\|\delta Y\|_\alpha$ and $\|\delta Z\|_\alpha$ although these processes do not have the dimension required by the definition. Finally, these two estimates imply that

$$\|\delta S\|_\alpha \quad \leq \quad \sqrt{\frac{2C^2}{\alpha}(1+T)}\|\delta s\|_\alpha.$$

By choosing $\alpha > 2(1+T)C^2$, it follows that the map ϕ is a contraction on $\mathbb{H}^2$, and that there is a unique fixed point.

4. It remain to prove that $Y \in \mathcal{S}^2$. This is easily obtained as in the first step of this proof, by first estimating:

$$\mathbb{E}\left[\sup_{t\leq T}|Y_t|^2\right] \leq C\left(|Y_0|^2 + \mathbb{E}\left[\int_0^T |f_t(Y_t,Z_t)|^2 dt\right] + \mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t Z_s\cdot dW_s\right|^2\right]\right),$$

and then using the Lipschitz property of the generator and the Doob's martingale inequality. ♢

Remark 11.4. Consider the Picard iterations:

$$(Y^0,Z^0) = (0,0), \text{ and}$$
$$Y_t^{k+1} = \xi + \int_t^T f_s(Y_s^k, Z_s^k)ds + \int_t^T Z_s^{k+1}\cdot dW_s,$$

Given (Y^k,Z^k), the next step (Y^{k+1},Z^{k+1}) is defined by means of the martingale representation theorem. Then, $S^k = (Y^k,Z^k) \longrightarrow (Y,Z)$ in $\mathbb{H}^2$ as $k\to\infty$. Moreover, since

$$\|S^k\|_\alpha \quad \leq \quad \left(\frac{2C^2}{\alpha}(1+T)\right)^k,$$

it follows that $\sum_k \|S^k\|_\alpha < \infty$, and we conclude by the Borel-Cantelli lemma that the convergence $(Y^k,Z^k)\longrightarrow(Y,Z)$ also holds $dt\otimes d\mathbb{P}-$a.s.

11.2.4 Complementary properties

This subsection collects some further properties of BSDEs, which are not essential for the purpose of our financial application.

Comparison

We first prove a monotonicity of property of BSDE which is the analogue of the corresponding property for forward stochastic differential equations. We observe that the following comparison result provides another (stronger) argument for the uniqueness of the solution to a BSDE.

Theorem 11.5. *Let $n=1$, and let (Y^i,Z^i) be the solution of BSDE(f^i,ξ^i) for some pair (ξ^i,f^i) satisfying the conditions of Theorem 11.3, $i=0,1$. Assume that*

$$\xi^1 \geq \xi^0 \quad \textit{and} \quad f_t^1(Y_t^0,Z_t^0) \geq f_t^0(Y_t^0,Z_t^0),\ dt\otimes d\mathbb{P}-a.s. \qquad (11.11)$$

Then $Y^1_t \geq Y^0_t$, $t \in [0,T]$, $\mathbb{P}-a.s.$

Proof. We denote

$$\delta Y := Y^1 - Y^0,\ \delta Z := Z^1 - Z^0,\ \delta_0 f := f^1(Y^0, Z^0) - f^0(Y^0, Z^0),$$

and we compute that

$$d(\delta Y_t) \quad = \quad -\left(\alpha_t \delta Y_t + \beta_t \cdot \delta Z_t + \delta_0 f_t\right) dt + \delta Z_t \cdot dW_t, \qquad (11.12)$$

where

$$\alpha_t \quad := \quad \frac{f^1_t(Y^1_t, Z^1_t) - f^1_t(Y^0_t, Z^1_t)}{\delta Y_t}\, \mathbf{1}_{\{\delta Y_t \neq 0\}},$$

and, for $j = 1, \ldots, d$,

$$\beta^j_t \quad := \quad \frac{f^1_t\left(Y^0_t, Z^1_t \oplus_{j-1} Z^0_t\right) - f^1_t\left(Y^0_t, Z^1_t \oplus_j Z^0_t\right)}{\delta Z^{0,j}_t}\, \mathbf{1}_{\{\delta Z^{0,j}_t \neq 0\}},$$

where $\delta Z^{0,j}$ denotes the $j-$th component of δZ^0, and for every $z^0, z^1 \in \mathbb{R}^d$, $z^1 \oplus_j z^0 := \left(z^{1,1}, \ldots, z^{1,j}, z^{0,j+1}, \ldots, z^{0,d}\right)$ for $0 < j < d$, $z^1 \oplus_0 z^0 := z^0$, $z^1 \oplus_d z^0 := z^1$.

Since f^1 is Lipschitz-continuous, the processes α and β are bounded. Solving the linear BSDE (11.12) as in subsection 11.2.2, we get:

$$\delta Y_t \quad = \quad \mathbb{E}\left[\Gamma^t_T \delta Y_T + \int_t^T \Gamma^t_u \delta_0 f_u du \Big| \mathcal{F}_t\right],\ t \leq T,$$

where the process Γ^t is defined as in (11.8) with $(\delta_0 f, \alpha, \beta)$ substituted to (a, b, c). Then Condition (11.11) implies that $\delta Y \geq 0$, $\mathbb{P}-$a.s. ♢

Stability

Our next result compares the difference in absolute value between the solutions of the two BSDEs, and provides a bound which depends on the difference between the corresponding final datum and the generators. In particular, this bound provides a transparent information about the nature of conditions needed to pass to limits with BSDEs.

Theorem 11.6. *Let (Y^i, Z^i) be the solution of BSDE(f^i, ξ^i) for some pair (f^i, ξ^i) satisfying the conditions of Theorem 11.3, $i = 0, 1$. Then:*

$$\|Y^1 - Y^0\|^2_{\mathcal{S}^2} + \|Z^1 - Z^0\|^2_{\mathbb{H}^2} \quad \leq \quad C\left(\|\xi^1 - \xi^0\|^2_{\mathbb{L}^2} + \|(f^1 - f^0)(Y^0, Z^0)\|^2_{\mathbb{H}^2}\right),$$

where C is a constant depending only on T and the Lipschitz constant of f^1.

Proof. We denote $\delta\xi := \xi^1 - \xi^0$, $\delta Y := Y^1 - Y^0$, $\delta f := f^1(Y^1, Z^1) - f^0(Y^0, Z^0)$, and $\Delta f := f^1 - f^0$. Given a constant β to be fixed later, we compute by Itô's formula that:

$$\begin{aligned} e^{\beta t}|\delta Y_t|^2 &= e^{\beta T}|\delta\xi|^2 + \int_t^T e^{\beta u}\left(2\delta Y_u \cdot \delta f_u - |\delta Z_u|^2 - \beta|\delta Y_u|^2\right) du \\ &\quad +2\int_t^T e^{\beta u}\delta Z_u^{\mathrm{T}}\delta Y_u \cdot dW_u. \end{aligned}$$

By the same argument as in the proof of Theorem 11.3, we see that the stochastic integral term has zero expectation. Then

$$e^{\beta t}|\delta Y_t|^2 = \mathbb{E}_t\left[e^{\beta T}|\delta\xi|^2 + \int_t^T e^{\beta u}\left(2\delta Y_u \cdot \delta f_u - |\delta Z_u|^2 - \beta|\delta Y_u|^2\right) du\right], \tag{11.13}$$

where $\mathbb{E}_t := \mathbb{E}[.|\mathcal{F}_t]$. We now estimate that, for any $\varepsilon > 0$:

$$\begin{aligned} 2\delta Y_u \cdot \delta f_u &\le \varepsilon^{-1}|\delta Y_u|^2 + \varepsilon|\delta f_u|^2 \\ &\le \varepsilon^{-1}|\delta Y_u|^2 + \varepsilon\big(C(|\delta Y_u| + |\delta Z_u|) + |\Delta f_u(Y_u^0, Z_u^0)|\big)^2 \\ &\le \varepsilon^{-1}|\delta Y_u|^2 + 3\varepsilon\big(C^2(|\delta Y_u|^2 + |\delta Z_u|^2) + |\Delta f_u(Y_n^0, Z_u^0)|^2\big). \end{aligned}$$

We then choose $\varepsilon := 1/(6C^2)$ and $\beta := 3\varepsilon C^2 + \varepsilon^{-1}$, and plug the latter estimate in (11.13). This provides:

$$e^{\beta t}|\delta Y_t|^2 + \frac{1}{2}\mathbb{E}_t\left[\int_t^T |\delta Z_u|^2 du\right] \le \mathbb{E}_t\left[e^{\beta T}|\delta\xi|^2 + \frac{1}{2C^2}\int_0^T e^{\beta u}|\Delta f_u(Y_u^0, Z_u^0)|^2 du\right],$$

which implies the required inequality by taking the supremum over $t \in [0, T]$ and using the Doob's maximal inequality for the martingale $\{\mathbb{E}_t[e^{\beta T}|\delta\xi|^2], t \le T\}$.
$\diamondsuit$

BSDEs and semilinear PDEs

In this section, we specialize the discussion to the so-called Markov BSDEs in the one-dimensional case $n = 1$. This class of BSDEs corresponds to the case where

$$f_t(y, z) = F(t, X_t, y, z) \quad \text{and} \quad \xi = g(X_T),$$

where $F : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}$ and $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ are measurable, and X is a Markov diffusion process defined by some initial data X_0 and the SDE:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{11.14}$$

Here μ and σ are continuous and satisfy the usual Lipschitz and linear growth conditions in order to ensure existence and uniqueness of a strong solution to the SDE (11.14), and

$$\begin{gathered} F, g \text{ have polynomial growth in x} \\ \text{and } F \text{ is uniformly Lipschitz in } (y, z). \end{gathered}$$

Then, it follows from Theorem 11.3 that the above Markov BSDE has a unique solution.

We next move the time origin by considering the solution $\{X_s^{t,x}, s \ge t\}$ of (11.14) with initial data $X_t^{t,x} = x$. The corresponding solution of the BSDE

$$dY_s = -F(s, X_s^{t,x}, Y_s, Z_s)ds + Z_s dW_s, \quad Y_T = g\left(X_T^{t,x}\right) \tag{11.15}$$

will be denote by $(Y^{t,x}, Z^{t,x})$.

Proposition 11.7. *The process $\{(Y_s^{t,x}, Z_s^{t,x}), s \in [t,T]\}$ is adapted to the filtration*

$$\mathcal{F}_s^t := \sigma\left(W_u - W_t, u \in [t,s]\right), \quad s \in [t,T].$$

In particular, $u(t,x) := Y_t^{t,x}$ is a deterministic function and

$$Y_s^{t,x} = Y_s^{s,X_s^{t,x}} = u\left(s, X_s^{t,x}\right), \text{ for all } s \in [t,T], \ \mathbb{P}-a.s.$$

Proof. The first claim is obvious, and the second one follows from the fact that $X_r^{t,x} = X_r^{s,X_s^{t,x}}$. ♢

Proposition 11.8. *Let u be the function defined in Proposition 11.7, and assume that $u \in C^{1,2}([0,T), \mathbb{R}^d)$. Then:*

$$-\partial_t u - \mu \cdot Du - \frac{1}{2} Tr[\sigma\sigma^T D^2 u] - f(., u, \sigma^{\mathrm{T}} Du) = 0 \text{ on } [0,T) \times \mathbb{R}^d. \tag{11.16}$$

Proof. This an easy application of Itô's formula together with the usual localization technique. ♢

We conclude this chapter by an nonlinear version of the Feynman-Kac formula.

Theorem 11.9. *Let $v \in C^{1,2}([0,T), \mathbb{R}^d)$ be a solution of the semilinear PDE* (11.16) *with polynomially growing v and $\sigma^{\mathrm{T}} Dv$. Then*

$$v(t,x) = Y_t^{t,x} \quad \text{for all} \quad (t,x) \in [0,T] \times \mathbb{R}^d,$$

where $(Y^{t,x}, Z^{t,x})$ is the solution of the BSDE (11.15).

Proof. For fixed (t,x), denote $Y_s := v(s, X_s^{t,x})$ and $Z_s := \sigma^{\mathrm{T}}(s, X_s^{t,x})$. Then, it follows from Itô's formula that (Y,Z) soves (11.15). From the polynomial growth on v and Dv, we see that the processes Y and Z are both in $\mathbb{H}^2$. Then they coincide with the unique solution of (11.15). ♢

11.3 Application: Funding Value Adjustment of the Black-Scholes theory

As previously mentioned, the standard Black-Scholes theory is a extreme simplification of real world financial markets. One important focus of derivatives trading in the recent period is on the funding possibilities of the hedging strategy. The standard Black-Scholes model assumes that the cash component of the hedging strategy is borrowed or invested at the same interest rate r, regardless of any guarantee or collateral that the hedger may possess in portfolio. This assumption was accepted for a long time with the Libor (London interbank offered rate) as agreed interest rate. However, since the financial crisis started in 2007, the funding and the liquidity aspects became extremely important. As a consequence, revisiting the standard Black-Scholes model to account for this key aspect of the financial engineering activity is necessary.

In the following subsections, we show how to incorporate the funding value adjustment (FVA) into the Black-Scholes theory. We shall see that this leads to a nonlinear pricing rule. For Vanilla options with payoff defined by a deterministic function of some underlying primitive securities, the hedging and pricing values can be characterized by a nonlinear version of the Black-Scholes partial differential equation (PDE). However, general exotic derivatives exhibit a path-dependency in their defining payoff. This is precisely the place where backward SDEs are useful as they offer a substitute to the PDE characterization.

11.3.1 The BSDE point of view for the Black-Scholes model

In this subsection, we modify the standard notations in order to adapt to the BSDE framework. We return to the problem of pricing and hedging a derivative security defined by the payoff ξ at the maturity T. As in our previous analysis, ξ is an $\mathcal{F}_T-$measurable random variable, and in order to fit into the BSDE framework, we assume in addition that $\xi \in \mathbb{L}^2(\mathbb{Q}, \mathcal{F}_T)$ is square integrable under the risk-neutral measure $\mathbb{Q}$.

The financial market contains a nonrisky asset, defined by the $\mathbb{F}-$adapted bounded process of interest rates $\{r_t, t \geq 0\}$, and d risky primitive assets $S = (S^1, \ldots, S^d)$ defined by the dynamics

$$dS_t \quad = \quad \mathrm{diag}[S_t]\big(r_t \mathbf{1} dt + \sigma_t dB_t\big), \tag{11.17}$$

where B is a Brownian motion under the risk-neutral measure $\mathbb{Q}$, σ is an $\mathbb{F}-$adapted bounded process of volatility with values in the set of $d \times d$ invertible symmetric matrices satisfying the uniform ellipticity condition

$$\big|\sigma_t x\big| \geq \varepsilon \|x\| \quad \text{for all} \quad x \in \mathbb{R}^d,\ t \geq 0,\ \omega \in \Omega,$$

for some $\varepsilon > 0$. We also recall that $\mathbf{1} := (1, \ldots, 1) \in \mathbb{R}^d$, and that $\mathrm{diag}[s]$ is the diagonal $d \times d$ symmetric matrix with $i-$th diagonal entry s^i for all $s = (s^1, \ldots, s^d) \in \mathbb{R}^d$.

A portfolio strategy is an $\mathbb{F}-$adapted process θ with values in $\mathbb{R}^d$. For the sake of excluding doubling strategies, we restrict the portfolio process θ the set $\mathcal{A}$ of admissible strategies, as introduced in Definition 7.11 CHECK REFERENCE. For all $t \geq 0$, the scalar process θ_t^i represents the amount invested in the $i-$th risky asset. Then, denoting by X^θ the corresponding induced value process, it follows that the amount invested in the non-risky asset, i.e. the cash investment, is given by

$$\eta_t^\theta := X_t^\theta - \theta_t \cdot \mathbf{1} \quad \text{for all} \quad t \in [0,T]. \tag{11.18}$$

In the standard frictionless Black-Scholes model, the last cash amount η is invested in the non-risky asset whose return is defined by the interest rate r. Therefore, the self-financing condition determines the dynamics of the value process X:

$$dX_t^\theta = \eta_t^\theta r_t dt + \theta_t \cdot \mathrm{diag}[S_t]^{-1} dS_t, \quad t \in [0,T]. \tag{11.19}$$

Substituting the value of the cash investment η, we recover the dynamics of the value process

$$dX_t^\theta = r_t X_t^\theta dt + \sigma_t \theta_t \cdot dB_t, \quad t \in [0,T]. \tag{11.20}$$

Under this form, we now see that the problem of perfect hedging the derivative security ξ is reduced to the problem of backward SDE:

$$dY_t = r_t Y_t dt + Z_t \cdot dB_t, \quad t \in [0,T], \quad Y_T = \xi, \tag{11.21}$$

with $Y = Z$ and $Z = \sigma\theta$.

Applying the previous existence and uniqueness results, we obtain immediately the following connection between the problem of hedging/pricing and the BSDE.

Proposition 11.10. *Let r, σ be bounded $\mathbb{F}-$progressively measurable processes, and $\xi \in \mathbb{L}^2(\mathbb{Q})$.*
(i) *There is a unique solution $(Y,Z) \in \mathbb{H}^2(\mathbb{Q}) \times \mathbb{S}^2(\mathbb{Q})$ of the backward SDE* (11.21)*, with*

$$Y_t = \mathbb{E}_t^{\mathbb{Q}}\Big[\xi e^{-\int_t^T r_u du}\Big], \quad t \in [0,T].$$

(ii) *For all $t \in [0,T]$, $Y_t = X_t^\theta$, where $\theta := \sigma^{-1} Z \in \mathcal{A}$, and Y_t is the no-arbitrage price of the derivative security with $T-$maturity payoff ξ.*

11.3.2 Funding Value Adjustment (FVA)

The traditional derivatives valuation reviewed in the previous subsection starts by assuming the existence of a risk-free rate r, which serves as a discounting rate for all cash instruments.

Since the recent financial crisis, the existence of such a risk-free rate is not relevant anymore. Nothing in the current financial market looks like a theoretically-classic money market account. It is now necessary to distinguish between cash, in positive amount, lended to the bank, or in negative amount borrowed from the bank. Moreover, financial transactions are subject to a collateral which serves to secure the deal throughout its duration.

The subsequent model ignores deliberately further market realities related to the potential possible default of both parties of the transaction, and also the potential default of the funding entities holding the cash component of the portfolio, or lending it if negative.

Recall from the previous subsection that the cash component of the portfolio η is defined in (11.18). We denote by C_t the collateral deposited at time t. The precise definition of C_t is specified in the derivative contract, and we shall consider different specifications which are practiced in real financial markets. In order to capture the funding aspect of the hedging strategy, we decompose η into three components:

$$\eta_t^\theta := X_t^\theta - \theta_t \cdot \mathbf{1} = C_t + (\eta_t^\theta - C_t)^+ - (\eta_t^\theta - C_t)^-. \tag{11.22}$$

Now, each one of these three components are subject to a specific interest rate. The positive deposits in cash gain interest at the rate $\underline{r}_t$, while the negative deposits correspond to a borrowing transaction from the bank are subject to the interest rate $\overline{r}_t$. Of course, $\underline{r} \le \overline{r}$. The collateral part C is also deposited in cash and gains interest at the rate r_t^C. All interest rate processes $\underline{r}, \overline{r}$ and r^C are $\mathbb{F}-$adapted and bounded.

As a consequence of this modeling, the dynamics of the cash component of the portfolio is given by:

$$d\eta_t^\theta = \phi_t(\eta_t^\theta, C_t)dt, \quad \text{with } \phi_t(\alpha, \beta) := r_t^C \beta + (\alpha - \beta)^+ \underline{r}_t - (\alpha - \beta)^- \overline{r}_t. \tag{11.23}$$

This determines the dynamics of the value process X induced by the portfolio strategy θ:

$$\begin{aligned} dX_t^\theta &= d\eta_t + \theta_t \text{diag}[S_t]^{-1} dS_t \\ &= \big[r_t \theta_t \cdot \mathbf{1} + \phi_t(C_t, X_t^\theta - \theta_t \cdot \mathbf{1})\big] dt + \sigma_t \theta_t \cdot dB_t, t \in [0,T]. \end{aligned} \tag{11.24}$$

In the remaining part of this chapter, we distinguish different cases of collaterals.

Non-collaterized transaction

We first consider the case $C \equiv 0$, where no collateral is involved in the transaction. Then, the above dynamics reduces to

$$dX_t^\theta = \big[r_t\theta_t \cdot \mathbf{1} + \underline{r}_t(X_t^\theta - \theta_t \cdot \mathbf{1})^+ - \overline{r}_t(X_t^\theta - \theta_t \cdot \mathbf{1})^-\big]dt + \sigma_t\theta_t \cdot dB_t, t \in [0,T]. \tag{11.25}$$

In order to solve the hedging problem in the present context, we introduce the BSDE

$$\begin{aligned} &dY_t^0 = \big[r_t \sigma_t^{-1} Z_t^0 \cdot \mathbf{1} + \underline{r}_t (Y_t^0 - \sigma_t^{-1} Z_t^0 \cdot \mathbf{1})^+ - \overline{r}_t (Y_t^0 - \sigma_t^{-1} Z_t^0 \cdot \mathbf{1})^-\big] dt + Z_t^0 \cdot dB_t, \\ &Y_T^0 = \xi. \end{aligned} \tag{11.26}$$

Under our conditions on the coefficients, it is immediately seen that the generator of the last BSDE satisfies the wellposedness conditions required in Theorem 11.3. We may then deduce that, for all $\xi \in \mathbb{L}^2(\mathbb{Q})$, there is a unique solution (Y^0, Z^0) to the BSDE (11.30). Consequently, the value of the non-collaterized contract at time t is Y_t^0, and the corresponding perfect hedging strategy is given by:

$$\hat{\theta}_t^0 := \sigma_t^{-1} Z_t^0, \quad t \in [0,T]. \tag{11.27}$$

Fully collaterized transaction

The next relevant case is when the collateral is given by the value of the derivative security. Then, the problem reduces to the BSDE

$$\begin{aligned} dY_t^1 &= \big[r_t \sigma_t^{-1} Z_t^1 \cdot \mathbf{1} + \phi_t(Y_t^1, Y_t^1 - \sigma_t^{-1} Z_t^1 \cdot \mathbf{1})\big] dt + Z_t^1 \cdot dB_t \\ Y_T^1 &= \xi. \end{aligned} \tag{11.28}$$

We again verify immediately that the generator of the last BSDE satisfies the wellposedness conditions of Theorem 11.3, and we deduce that there is a unique solution (Y^1, Z^1) for all $\xi \in \mathbb{L}^2(\mathbb{Q})$. Consequently, the value of the fully-collaterized contract at time t is Y_t^1, and the corresponding perfect hedging strategy is given by:

$$\hat{\theta}_t^1 := \sigma_t^{-1} Z_t^1, \quad t \in [0,T]. \tag{11.29}$$

Partially collaterized transaction

We finally consider the case where the collateral is fixed to some given proportion $\alpha \in [0,1]$ of the value of the derivative security. Then, the problem reduces to the BSDE

$$\begin{aligned} dY_t^\alpha &= \big[r_t \sigma_t^{-1} Z_t^\alpha \cdot \mathbf{1} + \phi_t(\alpha Y_t^\alpha, Y_t^\alpha - \sigma_t^{-1} Z_t^\alpha \cdot \mathbf{1})\big] dt + Z_t^\alpha \cdot dB_t \\ Y_T^\alpha &= \xi. \end{aligned} \tag{11.30}$$

By the same argument as for the cases $\alpha \in \{0,1\}$, there is a unique solution (Y^α, Z^α) for all $\xi \in \mathbb{L}^2(\mathbb{Q})$, and we may conclude that the value of the partially collaterized contract at time t is Y_t^α, with corresponding perfect hedging strategy:

$$\hat{\theta}_t^\alpha := \sigma_t^{-1} Z_t^\alpha, \quad t \in [0,T]. \tag{11.31}$$

PDE characterization in the Markov case

Suppose that the diffusion coefficient σ of the SDE (11.17) driving the dynamics of the underlying risky assets is a deterministic function, i.e. $\sigma_t = \sigma(t, S_t)$. This is the so-called local volatility model. As usual, the function σ is needed to be continuous in (t,s), and Lipschitz in s uniformly in t. Assume further that the various interest rates $r, \underline{r}, \overline{r}, r^C$ are all constant.

Then, it follows from Proposition 11.7 that $Y_t^\alpha = u^\alpha(t, S_t)$ for some function $u^\alpha : [0, T] \times \mathbb{R}^d \longrightarrow \mathbb{R}$. Moreover, assuming that $u^\alpha \in C^{1,2}([0, T), \mathbb{R}_+^d)$, we deduce from Proposition 11.8 that the function u^α is a solution of the semilinear PDE

$$-\partial_t u^\alpha - \frac{1}{2}\mathrm{Tr}\big[\mathrm{diag}[s]\sigma(t,s)^2\mathrm{diag}[s]D^2u^\alpha\big] + \phi(\alpha u^\alpha, u^\alpha - s \cdot Du^\alpha) = 0,$$

where

$$\phi(\alpha, \beta) := r^C\beta + (\alpha - \beta)^+\underline{r} - (\alpha - \beta)^-\overline{r}.$$

Finally, the perfect hedging strategy is given by

$$\hat\theta_t^\alpha \quad = \quad \mathrm{diag}[S_t]Du^\alpha(t, S_t), \quad t \in [0, T].$$

Chapter 12

Doob-Meyer decomposition, optimal stopping and American options

In contrast with European call and put options, American options can be exercised at any time before the maturity. The holder of the American contract chooses the exercise time depending on his information set, so his exercising decision is adapted to his information. This leads to the notion of stopping times. At the exercise time, the payoff received by the holder is the current intrinsic value of the option. The problem of optimal choice of such an exercise time is the so-called optimal stopping problem, which we also review the main theory in this section.

Another important ingredient for our application to American options is the Doob-Meyer decomposition of submartingales, a crucial result for many questions in stochastic analysis. The first section of this chapter is dedicated to the proof of this result. The next section provides an overview of the theory of optimal stopping. The final section focuses on the application to the valuation and hedging of American derivatives. In the Markov case, we derive the valuation equation. This turns out to be an obstacle partial differential equation which has no explicit solution, in general. Finally, we specialize the discussion the perpetual American puts. Although these contract are not traded on real financial markets, their analysis is interesting as we will be able to obtain a closed-form pricing formula.

12.1 The Doob-Meyer decomposition

In this section, we prove a decomposition result for submartingales which plays a central role in the theory of stochastic processes. The discrete-time decomposition was introduced by Doob. The corresponding continuous-time result was later obtained by Meyer and is proved by technically involved limiting arguments.

The following statement uses the following notion. A process is said to be $\mathbb{F}-$predictable if it is measurable with respect to the filtration generated by all $\mathbb{F}-$adapted left-continuous processes.

Theorem 12.1. *Let X be a càd-làg submartingale. Then:*
(i) *there is a unique decomposition $X = M + A$ for some càdlàg local martingale M, and a predictable nondecreasing process A with $A_0 = 0$; moreover, A is integrable whenever $\lim_{t\to\infty} \mathbb{E}[X_t] < \infty$.*
(ii) *Assume further that X is of class D, i.e. the family $\{X_\tau, \tau \in \mathcal{T}^T\}$ is uniformly integrable.Then, the component M of the decomposition is a martingale.*

By a classical localization argument, we may obtain Claim (i) of the last theorem as a direct consequence of Claim (ii). Therefore, we shall only focus on the proof of Claim (ii) which is restated in Theorem 12.6 below. The proof reported in this section follows the simplified arguments in [6].

Remark 12.2. Our use of the Doob-Meyer decomposition in the application to the pricing and hedging of American options does not require the predictability of the process A.

12.1.1 The discrete-time Doob decomposition

We first recall the Doob decomposition of submartingales in discrete-time, which is a simple calculation.

Proposition 12.3. *Let $(X_n)_{n\geq 1}$ be a discrete-time process. Then, there exists a unique decomposition $X = M + A$, where M is a martingale, and A is a predictable process (i.e. A_n is $\mathcal{F}_{n-1}-$measurable), with $A_0 = 0$.*

Moreover, X is a submartingale if and only if A is nondecreasing.

Proof. To prove uniqueness, we consider another decomposition $X = M' + V'$, and we observe that $M - M' = V - V'$ is predictable. Then $M_n - M'_n = M_{n-1} - M'_{n-1} = \ldots = M_0 - M'_0 = 0$ and $V_n = V'_n$. Next, denote $\Delta X_n := X_n - X_{n-1}$, $\Delta M_n := M_n - M_{n-1}$, $\Delta A_n := A_n - A_{n-1}$, and notice that the desired decomposition satisfies $\Delta X_n = \Delta M_n + \Delta A_n$. Since A is predictable and M is a martingale, this determines $\Delta A_n = \mathbb{E}[\Delta X_n|\mathcal{F}_{n-1}]$, and suggests the following pair (A, M):

$$A_n := \sum_{i=1}^{n} \mathbb{E}[\Delta X_i|\mathcal{F}_{i-1}] \quad \text{and} \quad M_n := X_n - A_n, \; n \geq 1.$$

Immediate verification shows that the pair (M, A) satisfies all requirements. $\diamondsuit$

12.1.2 Convergence of uniformly integrable sequences

In order to extend the previous decomposition to the continuous-time case, we shall sample the process on a discrete-time grid, and pass to the limit with the discrete-time decomposition. This requires some techniques which are isolated in the present subsection.

The space of random variable being of infinite dimension, the structure of compact subsets requires an important care as the unit ball is not compact. Since the space $\mathbb{L}^2$ is a Hilbert reflexive space, we shall use the following well-known result in functional analysis.

Lemma 12.4. *Let $(X_n)_{n\geq 1}$ be a bounded sequence in $\mathbb{L}^2$. Then, there exist a family $(\lambda_i^n, i = n, \ldots, N_n)_{n\geq 1}$, with $\lambda_i^n \geq 0$ and $\sum_{i=1}^{N_n} \lambda_i^n = 1$, such that*

$$\hat{X}_n := \sum_{i=n}^{N_n} \lambda_i^n X_i \quad \textit{converges in} \quad \mathbb{L}^2.$$

Our convergence result will make use of the following more refined convergence result.

Lemma 12.5. *Let $(X_n)_{n\geq 1}$ be a uniformly integrable sequence in $\mathbb{L}^1$. Then, there exist a family $(\lambda_i^n, i = n, \ldots, N_n)_{n\geq 1}$, with $\lambda_i^n \geq 0$ and $\sum_{i=1}^{N_n} \lambda_i^n = 1$, such that*

$$\hat{X}_n := \sum_{i=n}^{N_n} \lambda_i^n X_i \quad \textit{converges in} \quad \mathbb{L}^1.$$

Proof. Define the truncated sequences $\big(X_n^k := X_n \mathbf{1}_{\{|X_n|\leq k\}}, n \geq 1\big)$, $k \geq 1$. By Lemma 12.4, we may find $(\mu_i^n, n \leq i \leq N_n)_{n\geq 1}$, with $\mu_i^n \geq 0$ and $\sum_{i=1}^{N_n} \mu_i^n = 1$, such that $\sum_{i=n}^{N_n} \mu_i^n X_i^1$ converges in $\mathbb{L}^2$ to some $X^1 \in \mathbb{L}^2$.

We next consider the sequence $\hat{X}_n^2 := \sum_{i=n}^{N_n} \mu_i^n X_i^2$, which inherits from the sequence $(X_n^2)_{n\geq 1}$ its $\mathbb{L}^2-$bound. Applying again Lemma 12.4, we may find $(\hat{\mu}_i^n, n \leq i \leq N_n)_{n\geq 1}$, with $\hat{\mu}_i^n \geq 0$ and $\sum_{i=1}^{N_n} \hat{\mu}_i^n = 1$, such that $\sum_{i=n}^{N_n} \hat{\mu}_i^n \hat{X}_i^2$ converges in $\mathbb{L}^2$ to some $X^2 \in \mathbb{L}^2$. Notice that $\sum_{i=n}^{N_n} \hat{\mu}_i^n \hat{X}_i^2 = \sum_{i=n}^{N_n} \tilde{\mu}_i^n X_i^2$, where $(\tilde{\mu}_i^n)_i$ are the coefficients of a convex combination defined as the composition of the two convex combinations, and that both $\sum_{i=n}^{N_n} \tilde{\mu}_i^n X_i^1$ and $\sum_{i=n}^{N_n} \tilde{\mu}_i^n X_i^2$ converge in $\mathbb{L}^2$.

Iterating this procedure, and applying a diagonalization argument, we obtain sequences of convex combinations $(\lambda_i^n)_{n\leq i\leq N_n}$ such that all sequences $\sum_{i=n}^{N_n} \lambda_i^n X_i^k$, $k \geq 1$, converge in $\mathbb{L}^2$.

We finally use the uniform integrability of the sequence $(X_n)_{n\geq 1}$ which ensures that $X_n^k \longrightarrow X_n$ in $\mathbb{L}^1$, as $k \to \infty$, uniformly in n. This implies that

$$\sum_{i=1}^{N_n} \lambda_i^n X_i^k \longrightarrow \sum_{i=1}^{N_n} \lambda_i^n X_i \quad \text{in} \quad \mathbb{L}^1 \quad \text{as } k \to \infty, \text{ uniformly in } n,$$

and we conclude that $\left(\sum_{i=1}^{N_n} \lambda_i^n X_i\right)_{n\geq 1}$ is a Cauchy sequence in $\mathbb{L}^1$. Since $\mathbb{L}^1$ is complete, this provides the required result. $\diamondsuit$

12.1.3 The continuous-time Doob-Meyer decomposition

We now proceed to the continuous-time setting of Theorem 12.1. We recall that Claim (i) therein follows from (ii) by a localization technique. The following statement corresponds to Claim (ii) which we shall prove in this section.

Theorem 12.6. *Let X be a càd-làg submartingale of class D. Then:*
(i) *$X = M + A$ for some martingale càdlàg martingale M, and integrable càd-làg nondecreasing adapted process A with $A_0 = 0$;*
(ii) *moreover, A can be chosen to be predictable and, under this additional restriction, the decomposition is unique.*

Proof. We only prove (i). We first observe that we may assume without generality that

$$X_T = 0 \quad \text{and} \quad X_t \leq 0 \ \text{ for all } \ t \in [0, T]. \tag{12.1}$$

Indeed, these restrictions are satisfied by the process $\tilde{X}_t := X_t - \mathbb{E}[X_T|\mathcal{F}_t]$, $t \in [0, T]$. Moreover $\tilde{X}$ is clearly of class D, and inherits the submartingale property of X. Finally, if we prove that the process $\tilde{X}$ has the decomposition $\tilde{X} = \tilde{M} + \tilde{A}$ as in the statement of the theorem, then $X = M + A$ with $A := \tilde{A}$ and $M_t := M_t + \mathbb{E}[X_T|\mathcal{F}_t]$, $t \in [0, T]$, so that M is a càd-làg martingale.

In view of this, we now proceed to the proof of the theorem under the additional restriction (12.1).

1. For all $n \geq 1$, denote $t_j^n := j2^{-n}T$, $j = 0, \ldots, 2^n$, and define the discrete-time submartingale $X^n := \{X_{t_j^n}, 1 \leq j \leq 2^n\}$. By Proposition 12.3, we have the decomposition $X^n = M^n + A^n$ where M^n is a martingale and A^n is a predictable nondecreasing process. Moreover, it follows from (12.1) that

$$M_T^n = -A_T^n \quad \text{and} \quad X_{t_j^n} = A_{t_j^n}^n - \mathbb{E}[A_T^n|\mathcal{F}_{t_j^n}], \;\; j = 0, \ldots, 2^n. \tag{12.2}$$

Also, in preparation for the next step, we introduce for all $c > 0$ the sequence

$$\tau_n^c := 1 \wedge \inf\left\{t_{j-1}^n : A_{t_j^n}^n > c\right\} \quad \text{so that} \quad \{\tau_n^c < 1\} = \{A_T^n > c\}, \;\; n \geq 1.$$

Since A^n is predictable, it follows that τ_n^c is a stopping time, and it follows from the Chebychev inequality and (12.2) that

$$\mathbb{P}\left[\tau_n^c < 1\right] = \mathbb{P}\left[A_T^n > c\right] \leq \frac{\mathbb{E}\left[A_T^n\right]}{c} = \frac{-\mathbb{E}\left[M_T^n\right]}{c} = \frac{\mathbb{E}\left[X_0\right]}{c} \longrightarrow 0, \text{ uniformly in } n. \tag{12.3}$$

2. We now prove that the sequence $(A_T^n)_{n\geq 1}$ and $(M_T^n)_{n\geq 1}$ are uniformly integrable. Since $X_T = M_T^n + A_T^n$ is integrable, it suffices to prove that $(A_T^n)_{n\geq 1}$ is uniformly integrable.

First, by(12.2) and the definition of τ_n^c, we have

$$X_{\tau_n^c} = A^n_{\tau_n^c} - \mathbb{E}\big[A_T^n|\mathcal{F}_{\tau_n^c}\big] \le c - \mathbb{E}\big[A_T^n|\mathcal{F}_{\tau_n^c}\big].$$

We then estimate

$$\begin{aligned}\mathbb{E}\big[A_T^n\mathbf{1}_{\{A_T^n>c\}}\big] &= \mathbb{E}\big\{\mathbf{1}_{\{\tau_n^c<1\}}\mathbb{E}\big[A_T^n|\mathcal{F}_{\tau_n^c}\big]\big\}\\ &\le \mathbb{E}\big[\mathbf{1}_{\{\tau_n^c<1\}}\big(c-X_{\tau_n^c}\big)\big]\\ &= c\mathbb{P}\big[\tau_n^c<1\big]-\mathbb{E}\big[X_{\tau_n^c}\mathbf{1}_{\{\tau_n^c<1\}}\big]. \end{aligned} \tag{12.4}$$

Moreover, since $\{\tau_n^c < 1\} \subset \{\tau_n^{c/2} < 1\}$, it follows from (12.2) and the non decrease of A^n that

$$\begin{aligned}-\mathbb{E}\big[X_{\tau_n^{c/2}}\mathbf{1}_{\{\tau_n^{c/2}<1\}}\big] &= \mathbb{E}\big[\big(A_T^n - A^n_{\tau_n^{c/2}}\big)\mathbf{1}_{\{\tau_n^{c/2}<1\}}\big]\\ &\ge \mathbb{E}\big[\big(A_T^n - A^n_{\tau_n^{c/2}}\big)\mathbf{1}_{\{\tau_n^c<1\}}\big] \;\ge\; \frac{c}{2}\mathbb{P}\big[\tau_n^c<1\big].\end{aligned}$$

Plugging this in (12.4) provides the inequality

$$\mathbb{E}\big[A_T^n\mathbf{1}_{\{A_T^n>c\}}\big] \;\le\; -\mathbb{E}\big[X_{\tau_n^c}\mathbf{1}_{\{\tau_n^c<1\}}\big] - 2\mathbb{E}\big[X_{\tau_n^{c/2}}\mathbf{1}_{\{\tau_n^{c/2}<1\}}\big].$$

In view of (12.3), this implies that $(A_T^n)_{n\ge1}$ is uniformly integrable.
3. By the uniform integrability of $(M_T^n)_{n\ge1}$ established in the previous step, we deduce from Lemma 12.5 that there is a sequence $\{\lambda_k^n, 1\le k\le N_n\}_{n\ge1}$ with $\lambda_k^n \ge 0$ and $\sum_{k=n}^{N_n}\lambda_k^n = 1$, such that $\sum_{k=n}^{N_n}\lambda_k^n M_T^n$ converges in $\mathbb{L}^1$ to some r.v. $M_T \in \mathbb{L}^1$. Then, it follows from the Jensen inequality that

$$\hat M_t^n := \mathbb{E}\big[M_T^n|\mathcal{F}_t\big] \longrightarrow M_t := \mathbb{E}\big[M_T|\mathcal{F}_t\big] \quad \text{in} \quad \mathbb{L}^1 \quad \text{for all} \quad t\in[0,T].$$

Clearly, M is a càd-làg martingale. We also define

$$A_t^n := \sum_{j=1}^n A^n_{t_j^n}\mathbf{1}_{(t_{j-1}^n,t_j^n]}(t), \quad t\in[0,T], \quad \text{and} \quad \hat A^n := \sum_{k=n}^{N_n}\lambda_k^n A^n.$$

Then $\hat A^n_{t_j^k} = X_{t_j^k} - \hat M^n_{t_j^k}$ converges in $\mathbb{L}^1$ to some r.v. $A_{t_j^k} \in \mathbb{L}^1\big(\mathcal{F}_{t_j^k}\big)$, for all k, j. Therefore, after possibly passing to a subsequence, $\hat A^n_{t_j^k} \longrightarrow A_{t_j^k}$, $\mathbb{P}-$a.s. for all k, j and we obtain that the process $A_{t_j^k}$ inherits a.s. the nondecrease of $\hat A^n$. By right-continuity, the convergence and the nondecrease extend to the interval $[0,T]$. ♢

12.2 Optimal stopping

Throughout this section, we consider an $\mathbb{F}-$adapted pathwise continuous process $X : [0,T]\times\Omega \longrightarrow \mathbb{R}$ satisfying the condition.

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|X_t|\Big] \;<\; \infty. \tag{12.5}$$

Our main interest is on the optimal stopping problem

$$V_0 \quad := \quad \sup_{\tau \in \mathcal{T}^T} \mathbb{E}[X_\tau]. \tag{12.6}$$

In order to solve this problem, we shall use the *dynamic programming* approach, which requires to introduce the dynamic version of the optimal stopping problem, i.e. move the time origin from zero to an arbitrary point in the time interval $[0,T]$. The natural definition of such a dynamic version would be $V_t := \sup\{\mathbb{E}[X_\tau|\mathcal{F}_t] : \tau \in \mathcal{T}^T, \tau \geq t\}$. However, we are then faced to the problem of measurability of V_t as the supremum of an infinite uncountable family of measurable functions.

Therefore, we need to introduce a measurable substitute. Our starting point is the *Snell envelop* process:

$$Y_t \quad := \quad \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_t^T} \mathbb{E}[X_\tau|\mathcal{F}_t], \quad t \in [0,T],$$

where $\mathcal{T}_t^T$ is the collection of all stopping times τ with $t \leq \tau \leq T$. We refer to the Appendix section 12.4 for the definition of the essential supremum.

Under this definition, we clearly have that Y is an $\mathbb{F}-$adapted process. In this section, we prove the following characterization of the optimal stopping problem (12.6).

Theorem 12.7. *Let X be an $\mathbb{F}-$adapted pathwise continuous process satisfying the integrability condition* (12.5)*. Then*
(i) *the Snell envelop has a càd-làg version, still denoted by Y,*
(ii) *Y is a supermartingale*
(iii) *$\tau^* := \inf\{t \geq 0 : Y_t = X_t\}$ is an optimal stopping rule, and $Y_{.\wedge\tau^*}$ is a martingale.*

Claims (i) is proved in Lemma 12.11 below, while (ii) is proved in Lemma 12.9 below. Finally Claim (iii) is proved in Corollary 12.13 below.

12.2.1 The dynamic programming principle

We start by a property of the objective function in (12.6).

Lemma 12.8. *For any $t \in [0,T)$, the family $\{\mathbb{E}[X_\tau|\mathcal{F}_t]; \tau \in \mathcal{T}_t^T\}$ satisfies the lattice property, i.e. for all $\tau_1, \tau_2 \in \mathcal{T}_t^T$, there exists $\bar\tau \in \mathcal{T}_t^T$ such that $\mathbb{E}[X_{\bar\tau}|\mathcal{F}_t] \geq \mathbb{E}[X_{\tau_1}|\mathcal{F}_t] \vee \mathbb{E}[X_{\tau_2}|\mathcal{F}_t]$, $\mathbb{P}-a.s.$*

Proof. Let $A := \{\mathbb{E}[X_{\tau_1}|\mathcal{F}_t] \geq \mathbb{E}[X_{\tau_2}|\mathcal{F}_t]\}$, and define $\bar\tau := \tau_1 \mathbf{1}_A + \tau_2 \mathbf{1}_{A^c}$. Clearly, $A \in \mathcal{F}_t$, and $\bar\tau \in \mathcal{T}_t^T$, and we immediately verify that it satisfies the claim in the statement of the lemma. $\diamondsuit$

As a consequence of this property, we deduce from Theorem 12.20 (ii) that for all $t \in [0,T)$, we may find a sequence $(\tau_n)_{n\geq 1} \subset \mathcal{T}_t^T$ such that

$$Y_t \quad = \quad \lim \uparrow_{n\to\infty} \mathbb{E}\big[X_{\tau_n}|\mathcal{F}_t\big]. \tag{12.7}$$

A first crucial property of the Snell envelop is the following.

Lemma 12.9. *Y is a supermartingale with* $\sup_{t\in[0,T]}\mathbb{E}|Y_t|<\infty$ *and* $\mathbb{E}[Y_t]=\sup_{\tau\in\mathcal{T}_t^T}\mathbb{E}[X_\tau]$ *for all* $t\in[0,T]$.

Proof. Denote $|X|_T^* := \sup_{t\in[0,T]}|X_t|$. By the definition of Y, we have

$$\sup_{t\in[0,T]}\mathbb{E}\big[|Y_t|\big] \quad\le\quad \mathbb{E}\big[|X|_T^*\big] \;<\; \infty.$$

For $s\le t$, it follows from Lemma 12.8 and its consequence in (12.7), together with the monotone convergence theorem and the tower property, that:

$$\mathbb{E}[Y_t|\mathcal{F}_s]=\mathbb{E}\big\{\lim\uparrow_{n\to\infty}\mathbb{E}\big[X_{\tau_n}|\mathcal{F}_t\big]\big|\mathcal{F}_s\big\}=\lim\uparrow_{n\to\infty}\mathbb{E}\big[X_{\tau_n}|\mathcal{F}_s\big]\le Y_s,$$

which proves that Y is a supermartingale.

Next, for $\tau\in\mathcal{T}_t^T$, we have $Y_t\ge\mathbb{E}[X_\tau|\mathcal{F}_t]$, a.s. By the tower property, this implies that $\mathbb{E}[Y_t]\ge\mathbb{E}[X_\tau]$. The reverse inequality follows from the lattice property of Lemma 12.8 and its consequence in (12.7). ♢

We now prove the main property for the dynamic programming approach.

Proposition 12.10. (Dynamic programming principle) *For all* $t\in[0,T)$ *and* $\theta\in\mathcal{T}_t^T$*:*

$$Y_t \quad=\quad \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^T}\mathbb{E}\big[X_\tau\mathbf{1}_{\{\tau<\theta\}}+Y_\theta\mathbf{1}_{\{\tau\ge\theta\}}\big|\mathcal{F}_t\big],\quad \mathbb{P}-a.s.$$

Proof. Since $X\le Y$, we have for all $\theta\in\mathcal{T}_t^T$

$$Y_t\le\operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^T}\mathbb{E}^{\mathbb{P}}[X_\tau\mathbf{1}_{\{\tau<\theta\}}+Y_\tau\mathbf{1}_{\{\tau\ge\theta\}}|\mathcal{F}_t]\le\operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^T}\mathbb{E}[X_\tau\mathbf{1}_{\{\tau<\theta\}}+Y_\theta\mathbf{1}_{\{\tau\ge\theta\}}|\mathcal{F}_t],$$

$\mathbb{P}-$a.s., where the last inequality is due to the supermartingale property of Y established in Lemma 12.9. On the other hand, since Y is a supermartingale, we have for all $\tau\in\mathcal{T}_t^T$:

$$Y_t\ge\mathbb{E}[Y_{\theta\wedge\tau}|\mathcal{F}_t]=\mathbb{E}[Y_\theta\mathbf{1}_{\{\tau\ge\theta\}}+Y_\tau\mathbf{1}_{\{\tau<\theta\}}|\mathcal{F}_t]\ge\mathbb{E}[Y_\theta\mathbf{1}_{\{\tau\ge\theta\}}+X_\tau\mathbf{1}_{\{\tau<\theta\}}|\mathcal{F}_t],$$

$\mathbb{P}-$a.s. The proof is completed by taking ess sup over $\tau\in\mathcal{T}_t^T$. ♢

Lemma 12.11. *Y has a càd-làg supermartingale version.*

Proof. Since Y is a martingale, it follows from Chapter 3, Theorem 3.17, that the statement of the lemma is equivalent to the right-continuity of the map $t\longmapsto\mathbb{E}[Y_t]$.

Let $\{t_n\}\subset[0,T]$ be such that $t_n\searrow t$. By Lemma 12.9, we know that $\mathbb{E}[Y_{t_n}]=\sup_{\tau\in\mathcal{T}_{t_n}^T}\mathbb{E}[X_\tau]\le\sup_{\tau\in\mathcal{T}_t^T}\mathbb{E}[X_\tau]\le\mathbb{E}[Y_t]$. On the other hand, for any $\tau\in\mathcal{T}_t^T$, denoting $\tau_n:=\tau\vee t_n$, it follows from the continuity of X and the uniform integrability of X that $\mathbb{E}[X_\tau]=\lim_{n\to\infty}\mathbb{E}[X_{\tau_n}]\le\liminf_{n\to\infty}\mathbb{E}[Y_{t_n}]$. Using again Lemma 12.9, we obtain that $\mathbb{E}[Y_t]\le\liminf_{n\to\infty}\mathbb{E}[Y_{t_n}]$. Hence,

$$\mathbb{E}[Y_t]\quad=\quad\lim_{s\downarrow t}\mathbb{E}[Y_s].$$

♢

12.2.2 Optimal stopping rule

We now introduce for all $\varepsilon \geq 0$ the stopping times:

$$D_\varepsilon := \inf\big\{t \geq 0 : Y_t \leq X_t + \varepsilon\big\}.$$

Lemma 12.12. *For all $\varepsilon > 0$, the stopped process $Y_{.\wedge D_\varepsilon}$ is a martingale.*

Proof. By the dynamic programming principle of Proposition 12.10, we know that $Y_{.\wedge D_\varepsilon}$ is a supermartingale. So, in order to prove the required result it remains to show that $Y_0 \leq \mathbb{E}[Y_{D_\varepsilon}]$.

By the dynamic programming principle of Proposition 12.10, we may find a sequence of stopping times $(\tau_n)_{n\geq 1} \subset \mathcal{T}^T$ such that

$$\begin{aligned} Y_0 &\leq \frac{1}{n} + \mathbb{E}\big[X_{\tau_n}\mathbf{1}_{\{\tau_n<D_\varepsilon\}} + Y_{D_\varepsilon}\mathbf{1}_{\{\tau_n<D_\varepsilon\}}\big] \qquad (12.8)\\ &\leq \frac{1}{n} + \mathbb{E}\big[(Y_{\tau_n}-\varepsilon)\mathbf{1}_{\{\tau_n<D_\varepsilon\}} + Y_{D_\varepsilon}\mathbf{1}_{\{\tau_n<D_\varepsilon\}}\big]\\ &\leq \frac{1}{n} + Y_0 - \varepsilon\mathbb{P}\big[\tau_n < D_\varepsilon\big], \end{aligned}$$

where the last inequality follows from the supermartingale property of Y. This implies that

$$n\varepsilon\,\mathbb{P}\big[\tau_n < D_\varepsilon\big] \quad\leq\quad 1. \qquad (12.9)$$

Returning to (12.8), we then see that

$$\begin{aligned} Y_0 &\leq \frac{1}{n} + \mathbb{E}\big[(X_{\tau_n}-Y_{D_\varepsilon})\mathbf{1}_{\{\tau_n<D_\varepsilon\}}\big] + \mathbb{E}\big[Y_{D_\varepsilon}\big]\\ &\leq \frac{1}{n} + \mathbb{E}\big[2|X|^*_T\mathbf{1}_{\{\tau_n<D_\varepsilon\}}\big] + \mathbb{E}\big[Y_{D_\varepsilon}\big]. \end{aligned}$$

Sending $n \to \infty$, and using (12.9) together with our integrability condition on $|X|^*_T$, this provides the required inequality $Y_0 \leq \mathbb{E}\big[Y_{D_\varepsilon}\big]$. ♢

We are now ready for the characterization of the optimal stopping rule.

Corollary 12.13. *Let $\tau^* := D_0$. Then the process $Y_{.\wedge\tau^*}$ is a martingale, and τ^* is an optimal stopping rule, i.e. $Y_0 = \mathbb{E}\big[X_{\tau^*}\big]$.*

Proof. By Lemma 12.12 together with the definition of D_ε, we have for all $\varepsilon > 0$:

$$Y_0 = \mathbb{E}\big[Y_{D_\varepsilon}\big] \leq \varepsilon + \mathbb{E}\big[X_{D_\varepsilon}\big].$$

From our condition $|X|^*_T \in \mathbb{L}^1$, it follows that the family $\{X_\tau, \tau \in \mathcal{T}^T\}$ is uniformly integrable. We then obtain by sending $\varepsilon \to 0$, and using the inequality $X \leq Y$ and the supermartingale property of Y:

$$Y_0 \leq \mathbb{E}\big[X_{\tau^*}\big] \leq \mathbb{E}\big[Y_{\tau^*}\big] \leq Y_0.$$

♢

12.3 Pricing and hedging American derivatives

12.3.1 Definition and first properties

Definition 12.14. *An American derivative security is defined by an adapted process $\{G_t, t \in [0,T]\}$, where for all $t \in [0,T]$, G_t is the (random) payoff received by the holder upon choosing to exercise the contract at time t.*

G_t is called the intrinsic value of the American option.

The most popular examples are those of American calls and puts on an underlying asset with price process $\{S_t, t \geq 0\}$. The intrinsic value of an American call with strike K is $(S_t - K)^+$, while the intrinsic value of an American put is $(K - S_t)^+$.

We recall from Chapter 1, Section 1.4, that the early exercise of American calls written on non-dividend paying assets is not optimal and, therefore, such a contract is equivalent to the European counterpart.

12.3.2 No-arbitrage valuation and hedging

Similar to European derivatives, we define the price of the American security by the costs of perfect hedging. Let T be the maturity of the American option, and let $G = \{G_t, t \in [0,T]\}$ be the corresponding process of intrinsic values. We would like to define the superhedging cost of the American derivative security as the minimal initial capital which allows the seller of the option to face the payoff defined by the intrinsic value at any possible exercise time. Since such possible exercise times are defined by all stopping times, we are reduced to the following definition of the super-hedging cost:

$$V_0(G) \quad := \quad \inf\left\{X_0 : X^\theta_\tau \geq G_\tau,\ \mathbb{P}-\text{a.s. for all } \tau \in \mathcal{T}^T,\ \text{for some } \theta \in \mathcal{A}\right\},$$

Here, we have used the notations introduced in Chapter 7, Section 7.4. In particular, we recall that the value of the portfolio θ is the solution of the budget constraint equation implied by the self-financing condition:

$$dX^\theta_t \quad = \quad \theta_u \cdot \text{diag}[S_t]^{-1} dS_t + (X^\theta_t - \theta_t) r_t dt,$$

and $\{r_t, t \in [0,T]\}$ is the adapted process of interest rates.

By adapting the approach used for European securities, and using some advanced result from the theory of optimal stopping and the Doob-Meyer decomposition of supermartingales, we can prove the following characterization of the superhedging cost in terms of risk-neutral valuation under the risk-neutral measure $\mathbb{Q}$.

Theorem 12.15. *Assume that the reward process $\{G_t, t \in [0,T]\}$ is pathwise continuous, and $\sup_{t\in[0,1]} \tilde{G}_t \in \mathbb{L}^1(\mathbb{Q})$. Then*

$$V_0(G) \quad = \quad \sup_{\tau \in \mathcal{T}^T} \mathbb{E}^{\mathbb{Q}}[\tilde{G}_\tau],$$

where $\mathcal{T}^T$ is the collection of all stopping times with values in $[0,T)$.

Proof. First, let $X_0 \in \mathbb{R}$ and $\theta \in \mathcal{A}$ be such that $X_\tau^\theta \geq G_\tau$, $\mathbb{P}-$a.s. for all $\tau \in \mathcal{T}^T$. Then, since $\tilde{X}^\theta$ is a $\mathbb{Q}-$supermartingale, it follows from the optional sampling theorem that $X_0 \geq \mathbb{E}^{\mathbb{Q}}[\tilde{G}_\tau]$. By the arbitrariness of $\tau \in \mathcal{T}^T$, this shows that $V_0(G) \geq \sup_{\tau\in\mathcal{T}^T} \mathbb{E}^{\mathbb{Q}}[\tilde{G}_\tau]$.

To prove the converse inequality, we introduce the Snell envelop

$$Y_t := \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^T} \mathbb{E}^{\mathbb{Q}}[\tilde{G}_\tau|\mathcal{F}_t],$$

and we notice that the process $\{\tilde{G}_t, t \in [0,T]\}$ inherits the pathwise continuity of the process $\{G_t, t \in [0,T]\}$. Then, it follows from Theorem 12.7 that $\{Y_t, t \geq 0\}$ may be considered in its càd-làg version, and is a $\mathbb{Q}-$supermatringale. We now apply the Doob-Meyer decomposition of Theorem 12.1 to obtain $Y = M - A$, where M is a $\mathbb{Q}-$local martingale, A is a nondecreasing process, and $A_0 = 0$.

Following the argument in the proof of Theorem 7.17, we next use the martingale representation theorem to identify the local martingale $M = X^\theta$ for some $\theta \in \mathcal{A}$. Then, for all stopping time τ, we see that $G_\tau \leq V_\tau = M_\tau - A_\tau = X_\tau^\theta - A_\tau \leq X_\tau^\theta$, $\mathbb{P}-$a.s. and therefore $X_0^\theta = M_0 = Y_0 \geq V_0$ by the definition of V_0. ♢

Notice that, in general:

$$V^0(G) \quad > \quad \sup_{t\leq T} \mathbb{E}^{\mathbb{Q}}[\tilde{G}_t]. \tag{12.10}$$

Here $\mathbb{E}^{\mathbb{Q}}[\tilde{G}_t]$ is the no-arbitrage price of the $t-$maturity European derivative with payoff G_t at maturity. So, the value $\sup_{t\leq T} \mathbb{E}^{\mathbb{Q}}[\tilde{G}_t]$ is the maximum value over all possible exercise times. The reason for the inequality (12.10) is that the amount $\sup_{t\leq T} \mathbb{E}^{\mathbb{Q}}[\tilde{G}_t]$ only account for the deterministic exercise strategies of the holder of the option. In particular, it does not account for all exercise strategies of the contract holder which would depend on the information. For instance, the rational holder may decide the exercise an American put option whenever the price of the underlying assets hits some level below K. Such a barrier exercise strategy is the typical example of a stopping time, i.e. a random time which can be perceived by the investor.

12.3.3 The valuation equation

In this section, we specialize the discussion to the one-dimensional ($d = 1$) Markovian case. We assume that the underlying asset price process pays no dividends and is generated by the stochastic differential equation:

$$\frac{dS_t}{S_t} \quad = \quad rdt + \sigma(t, S_t)dB_t.$$

Here, B is a Brownian motion under the risk-neutral measure $\mathbb{Q}$, $r \geq 0$ is the (constant) instantaneous interest rate, and σ is a volatility function satisfying the required condition for existence of a unique solution to the above stochastic differential equation.

The American derivative intrinsic value process is $G_t := g(t, S_t)$, for some continuous function $g : [0,T] \times [0,\infty) \longrightarrow \mathbb{R}$. Then, the American option price is:

$$V(t,s) \quad = \quad \sup_{\tau \in \mathcal{T}_t^T} \mathbb{E}_{t,s}\big[e^{-r\tau} g(\tau, S_\tau)\big], \quad t \in [0,T], \ s \in [0,\infty).$$

The following result provides the valuation partial differential equation of the American derivative.

Theorem 12.16. *Assume that the function $V \in C^{1,2}([0,T), \mathbb{R}_+)$. Then V is a solution of:*

$$\min\{-\partial_t V - rsV_s - \frac{1}{2}\sigma^2 s^2 V_{ss} + rV, V - g\} \quad = \quad 0, \quad t < T, s > 0.$$

Proof. In the first part of this proof, we show that V is a supersolution of the valuation equation, i.e.

$$\min\{-\partial_t V - rsV_s - \frac{1}{2}\sigma^2 s^2 V_{ss} + rV, V - g\} \geq 0, \quad t < T, \quad s > 0. \quad (12.11)$$

Then, in the second part, we show that V is a subsolution of the valuation equation:

$$\min\{-\partial_t V - rsV_s - \frac{1}{2}\sigma^2 s^2 V_{ss} + rV, V - g\} \leq 0, \quad t < T, \quad s > 0. \quad (12.12)$$

(i) Since immediate exercise is always possible (although not always optimal), we have $V \geq g$. To prove (12.11), it remains to prove that $-\partial_t V - rsV_s - \frac{1}{2}\sigma^2 s^2 V_{ss} + rV \geq 0$, i.e. V is a supersolution of the Black-Scholes valuation equation. To see this, we observe from the suermartingale property of the process $\{V(t, S_t)\}_{t\in[0,T]}$ that for all sufficiently small $h > 0$, we have:

$$V(t,s) \geq \mathbb{E}_{t,s}\Big[e^{-r(\tau_h - t)} V(\tau_h, S_{\tau_h})\Big],$$

where

$$\tau_h := (t+h) \wedge \inf\{u > t : |\ln(S_u/s)| \geq 1\}.$$

A simple alternative justification of the last inequality is that the right hand-side expresses the price of the American contract with no possible exercise on $[t, \tau_h]$. Since V is assumed to be $C^{1,2}$, it follows from Itô's formula that:

$$\begin{aligned} 0 \quad &\leq \quad -\mathbb{E}_{t,s} \int_t^{\tau_h} d\{e^{-r(t-u)} V(u, S_u)\} \\ &= \quad -\mathbb{E}_{t,s}\Big[\int_t^{\tau_h} e^{-r(t-u)}(-rV + \partial_t V + rsV_s + \frac{1}{2}\sigma^2 s^2 V_{ss})(u, S_u) du\Big]. \end{aligned}$$

By dividing by h, and sending h to zero, we obtain the required inequality.
(ii) We next prove (12.12) by showing that the following equivaent claim holds:

$$(V-g)(t,s)>0 \quad \text{implies that} \quad \{-\partial_t V - rsV_s - \frac{1}{2}\sigma^2 s^2 V_{ss} + rV\}(t,s) = 0.$$

Indeed, if $(V-g)(t,s)>0$, then, for some sufficiently small $h>0$, $V-g>0$ on $B_r(t,s)$, the ball centered at (t,s) with radius r. We then introduce

$$\tau_h := (t+h) \wedge \inf\{u>t : (u,S_u) \notin B_r(t,s)\}$$

By the definition of $r>0$, it is not optimal to exercise the American contrat before the stopping time τ_h for all $h>0$. Then

$$V(t,s) \;=\; \mathbb{E}_{t,s}\Big[e^{-r(\tau_h-t)}V(\tau_h, S_{\tau_h})\Big].$$

We next apply Itô's formula, as in the part (i), divide by h, and send h to zero to conclude that:

$$\{-\partial_t V - rsV_s - \frac{1}{2}\sigma^2 s^2 V_{ss} + rV\}(t,s) \;=\; 0.$$

$\diamondsuit$

Our last result provides conditions for the solution of the valuation equation to coincide with the price of the American option, and is the American counterpart of the corresponding European option result of Chapter 6 Proposition 6.13. Recall the collection of functions with generalized derivatives W^2 introduced in Chapter 6, Remark 6.3, and define similarly the set $W^{1,2}$ of function of the variables (t,s) with generalized partial derivatives in t and s.

Proposition 12.17. *Let $v \in W^{1,2}([0,T),\mathbb{R}_+) \cap C^0([0,T]\times\mathbb{R}_+)$ be a solution of the valuation equation*

$$\min\Big\{-\partial_t v - rsv_s - \frac{1}{2}\sigma^2 s^2 v_{ss} + rv, v-g\Big\} = 0, \quad (t,s)\in[0,T)\times(0,\infty),$$

with $v(T,s)=g(T,s)$ for all $s\ge 0$. Assume further that v has polynomial growth. Then $v(t,s)=V(t,s)=\mathbb{E}^{\mathbb{Q}}_{t,s}\big[e^{-r\tau^*}g\big(\tau^*,S_{\tau^*}\big)\big]$, where

$$\tau^* \;:=\; \inf\{u\ge t : v(u,S_u)=g(u,S_u)\},$$

i.e. τ^* is an optimal exercise rule.

Proof. By Remark 6.3, Itô's formula still holds with the function v under the present regularity. Then, for all stopping time $\tau\in\mathcal{T}_t^T$, we set $\tau_n := \tau \wedge (T-n^{-1}) \wedge \inf\{u\ge t : |\ln(S_u/s)| \ge n$ for $\varepsilon>0$ sufficiently small, and we compute by Itô's formula that:

$$v(t,s) \;=\; v\big(\tau_n, S_{\tau_n}\big) - \int_t^{\tau_n} e^{-r(u-t)}\big[\mathcal{L}v(u,S_u)du + v_s(u,S_u)\sigma(u,S_u)S_u dB_u\big],$$

where $\mathcal{L}v := \partial_t v + rsv_s + \frac{1}{2}\sigma^2 s^2 v_{ss} - rv$. Since v satisfies the valuation equation, we have $v \geq g$ and $\mathcal{L}v \leq 0$. Then, taking conditional expectations under $\mathbb{Q}$,

$$v(t,s) \;=\; \mathbb{E}^{\mathbb{Q}}_{t,s}\Big[v\big(\tau_n, S_{\tau_n}\big) - \int_t^{\tau_n} e^{-r(u-t)}\mathcal{L}v(u,S_u)du\Big] \;\geq\; \mathbb{E}^{\mathbb{Q}}_{t,s}\big[g\big(\tau_n, S_{\tau_n}\big)\big]$$

Since g has polynomial growth, we may send $n \to \infty$ and get $v(t,s) \geq \mathbb{E}^{\mathbb{Q}}_{t,s}[g(\tau, S_\tau)]$ by the dominated convergence theorem. By the arbitrariness of $\tau \in \mathcal{T}_t^T$, this shows that $v \geq V$.

We next review the previous calculation with $\tau = \tau^*$ as defined in the statement of the proposition. Since $v > g$ on $[t, \tau^*)$, it follows from the valuation equation that $\mathcal{L}v(u, S_u) = 0$ on $[0, \tau^*)$. This restores the first equality $v(t,s) = \mathbb{E}^{\mathbb{Q}}_{t,s}\big[v\big(\tau_n^*, S_{\tau_n^*}\big)\big]$. By the polynomial growth condition, we may use the dominated convergence theorem to pass to the limit $n \to \infty$. This provides

$$v(t,s) \;=\; \mathbb{E}^{\mathbb{Q}}_{t,s}\big[v\big(\tau^*, S_{\tau^*}\big)\big] \;=\; \mathbb{E}^{\mathbb{Q}}_{t,s}\big[g\big(\tau^*, S_{\tau^*}\big)\big],$$

where the last equality follows from the definition of τ^*. ♢

12.3.4 The exercise boundary

In this subsection, we specialize further the model by considering a constant volatility model

$$\frac{dS_t}{S_t} \;=\; r\,dt + \sigma\,dB_t.$$

Given an initial condition $S_0 = s$, we denote the corresponding solution by

$$S_t^s = se^{(r-\frac{\sigma^2}{2})t + \sigma B_t}, \quad t \geq 0, \quad s \geq 0.$$

Moreover, we focus on the American put option price

$$P(t,s) \;:=\; \sup_{\tau \in \mathcal{T}_t^T} \mathbb{E}^{\mathbb{Q}}\big[e^{-r(\tau-t)}(K - S_\tau)^+ \big| S_t = s\big].$$

By the time homogeneity of the problem, we have

$$P(t,s) \;=\; \sup_{\tau \in \mathcal{T}^{T-t}} \mathbb{E}^{\mathbb{Q}}\big[e^{-r\tau}(K - S_\tau^s)^+\big].$$

The exercise boundary is defined by

$$\mathcal{E} \;:=\; \big\{(t,s) \in [0,T] \times \mathbb{R}_+ : P(t,s) = (K-s)^+\big\},$$

and represents the location of the time-space domain where it is optimal to exercise the American put option. Similarly, the remaining region

$$\mathcal{S} \;:=\; \big\{(t,s) \in [0,T] \times \mathbb{R}_+ : P(t,s) > (K-s)^+\big\}$$

indicates the region where the contract is worth strictly more than its exercise value, and therefore exercising the option at such a point is strictly sub-optimal.

Proposition 12.18. *There exists a function $t \longmapsto s^*(t)$ such that $\mathcal{E} = \{(t,s) \in [0,T] \times \mathbb{R}_+ : s \le s^*(t)\}$.*

Proof. First, one easily proves that P is uniformly Lipschitz in s, see Exercise 12.22. We next show that

$$P(t,s_0) = (K-s_0)^+ \quad \Longrightarrow \quad P(t,s_1) = (K-s_1)^+ \text{ for all } s_1 \le s_0, \tag{12.13}$$

which implies the required result, given the continuity of the function P in s.

We first observe that $\mathcal{E} \cap [0,T) \times \mathbb{R}_+ \subset [0,T) \cap [0,K)$. Indeed, the inequality $\mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}(K - S^s_{T-t})^+] > 0 = (K-s)^+$, for all $(t,s) \in \mathcal{E} \cap [0,T) \times \mathbb{R}_+$, shows that immediate exercise at any point $s \ge K$ is not optimal.

Next, by the convexity of the function $s \longmapsto (K-s)^+$, we see that for all $s_1 < s_0$, we have $(K - S^{s_1}_\tau)^+ \le K\left(1 - \frac{s_1}{s_0}\right) + \frac{s_1}{s_0}(K - S^{s_0}_\tau)^+$. This implies that

$$P(t,s_1) \quad \le \quad K\Big(1 - \frac{s_1}{s_0}\Big) + \frac{s_1}{s_0}\, P(t,s_0),$$

which provides (12.13). ♢

12.3.5 Perpetual American derivatives

In this section, we consider the one dimensional Black-Scholes model:

$$\frac{dS_t}{S_t} \quad = \quad r dt + \sigma dB_t,$$

where the interest rate $r \ge 0$ and the volatility $\sigma > 0$ are constant parameters.

A perpetual American put option has infinite maturity, i.e. the holder of the option has an exercise right which he may defer forever. Due to the time homogeneity of the problem, the value of the infinite maturity American put does not depend on the time variable:

$$P(s) \quad := \quad \sup_{\tau \in \mathcal{T}_\infty} \mathbb{E}\Big[e^{-r\tau}(K - S_\tau)^+\Big],$$

where $\mathcal{T}_\infty$ is the collection of all finite stopping times.

The PDE approach

The valuation equation reduces to the following ordinary differential equation (ODE):

$$\min\{-rsv_s - \frac{1}{2}\sigma^2 s^2 v_{ss} + rv, v - g\} \quad = \quad 0, \quad s > 0.$$

We now provide a closed-form solution for this nonlinear ordinary differential equation (ODE). We start by guessing that there exists an exercise boundary $s^* < K$ such that

$$\begin{array}{l} v(s) = K - s, \quad s \le s^* \\ -rsv_s - \frac{1}{2}\sigma^2 s^2 v_{ss} + rv = 0, \quad s > s^*. \end{array}$$

We know that the general solution of an ODE is defined by a two-dimensional space determined by two independent solutions. In the present case, it is immediately seen that:

$$v(s) \quad = \quad \begin{cases} K-s, & s \leq s^* \\ As+Bs^{-2r/\sigma^2}, & s > s^*, \end{cases}$$

for some constants A and B. Cearly, the American put option price decreases to 0 as $s \nearrow \infty$. Then $A=0$. We next impose that v is continuous and continuously differentiable at the point s^*:

$$K-s^* = B(s^*)^{-2r/\sigma^2} \quad \text{and} \quad -1 = -\frac{2r}{\sigma^2}B(s^*)^{-1-(2r/\sigma^2)}.$$

This is a linear system of two equations for the unknowns s^* and B, which immediately implies the unique parameters:

$$s^* = \frac{K}{1+\frac{\sigma^2}{2r}} \quad \text{and} \quad B = \frac{\sigma^2}{2r}(s^*)^{1+(2r/\sigma^2)}.$$

It is now immediately checked that the function v induced by these parameters satisfy the conditions of Proposition 12.17. Hence $v=V$, and the first hitting time of the barrier s^* is an optimal exercise rule.

The probabilistic approach

By Proposition 12.18, the optimal exercise rule of the American put option is defined by some barrier $s^*(t)$. In the present infinite horizon context, it follows from the homogeneity of the problem that:

$$P(s) \quad := \quad \sup_{\tau\in\mathcal{T}} \mathbb{E}^{\mathbb{Q}}\big[e^{-r\tau}(K-S^s_\tau)\big] \;=\; \sup_{b<K} \mathbb{E}^{\mathbb{Q}}\big[e^{-rT^s_b}(K-S^s_{T^s_b})^+\big],$$

where

$$S^s_t = se^{(r-\frac{\sigma^2}{2})t+\sigma B_t}, \quad \text{and} \quad T^s_b := \inf\{t>0 : S^s_t \leq b\}.$$

Then, in preparation for the derivation of $P(s)$ in closed form, we start by studying the Laplace transform of hitting times of barriers.

Lemma 12.19. *Let B be a scalar Brownian motion on $(\Omega,\mathcal{F},\mathbb{Q})$, and*

$$X_t := \theta t + B_t, \quad t \geq 0, \quad H_a := \inf\{t>0 : X_t = a\}.$$

Then, the distribution of the hitting time H_a is characterized by its Laplace transform:

$$\mathbb{E}^{\mathbb{Q}}\big[e^{-\lambda H_a}\big] \;=\; e^{-(\theta+\sqrt{\theta^2+2\lambda})|a|} \quad \textit{for all} \quad \lambda > 0. \tag{12.14}$$

Proof. By the symmetry of the Brownian motion, it is sufficient to consider the case $a \geq 0$. By the Girsanov theorem, the process X is a Brownian motion under the probability measure $\hat{\mathbb{Q}}$ defined by the density:

$$\left.\frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}}\right|_{\mathcal{F}_t} = e^{-\theta B_t - \frac{1}{2}\theta^2 t} = e^{-\theta X_t + \frac{1}{2}\theta^2 t}, \quad t \geq 0.$$

Noting that $X_{H_a} = a$, by the continuity of X, we rewrite the Laplace transform as

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\big[e^{-\lambda H_a}\big] &= e^{-\theta^- a}\mathbb{E}^{\mathbb{Q}}\big[e^{\theta^- X_{H_a} - \lambda H_a}\big] \\ &= e^{-\theta^- a} \lim_{n\to\infty} \mathbb{E}^{\mathbb{Q}}\big[e^{\theta^- X_{n\wedge H_a} - \lambda(n\wedge H_a)}\big]\end{aligned}$$

by the dominated convergence theorem, due to the fact that $\lambda \geq 0$ and $0 \leq e^{\theta^- X_{n\wedge H_a}} \leq e^{\theta^- a}$. By changing to the measure $\hat{\mathbb{Q}}$, it follows from another application of the dominated convergence theorem that

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\big[e^{-\lambda H_a}\big] &= e^{-\theta^- a} \lim_{n\to\infty} \mathbb{E}^{\hat{\mathbb{Q}}}\big[e^{(\theta^- + \theta) X_{n\wedge H_a} - \frac{1}{2}\theta^2(n\wedge H_a)} e^{-\lambda(n\wedge H_a)}\big] \\ &= e^{-\theta^- a} \lim_{n\to\infty} \mathbb{E}^{\hat{\mathbb{Q}}}\big[e^{\theta^+ X_{n\wedge H_a} - \frac{1}{2}\gamma^2(n\wedge H_a)}\big] \\ &= e^{-\theta a}\mathbb{E}^{\hat{\mathbb{Q}}}\big[e^{-\frac{1}{2}\gamma^2 H_a}\big],\end{aligned}$$

where $\gamma := \sqrt{2\lambda + \theta^2}$. We are then reduced to the calculation of the Laplace transform of H_a under the probability measure $\hat{\mathbb{Q}}$. Consider the $\hat{\mathbb{Q}}-$martingale $M_t := e^{\gamma X_t - \frac{1}{2}\gamma^2 t}$, $t \geq 0$. Then, for all $n \in \mathbb{N}$, it follows from the optional sampling theorem that

$$\mathbb{E}^{\hat{\mathbb{Q}}}\big[M_{n\wedge H_a}\big] = \mathbb{E}^{\hat{\mathbb{Q}}}\Big[e^{\gamma X_{H_a\wedge n} - \frac{1}{2}\gamma^2(H_a\wedge n)}\Big] = 1.$$

Since $0 \leq M_{H_a\wedge n} \leq e^{\gamma a}$ is uniformly bounded, it follows from the dominated convergence theorem that

$$1 = \mathbb{E}^{\hat{\mathbb{Q}}}\Big[e^{\gamma X_{H_a} - \frac{1}{2}\gamma^2 H_a}\Big] = e^{\gamma a}\mathbb{E}^{\hat{\mathbb{Q}}}\Big[e^{-\frac{1}{2}\gamma^2 H_a}\Big].$$

We then deduce that

$$\mathbb{E}^{\mathbb{Q}}\big[e^{-\lambda H_a}\big] = e^{-(\theta+\gamma)a} = e^{-(\theta+\sqrt{2\lambda+\theta^2})a}.$$

♢

We now proceed to the explicit calculation of the perpetual American put price. For $b < K$, let

$$a := \frac{1}{\sigma}\ln(b/s), \quad \theta := \frac{r}{\sigma} - \frac{\sigma}{2}, \quad \text{and} \quad H_a := \inf\big\{t \geq 0 : B_t + \theta t = a\big\}.$$

We first observe that

$$T_b^s = \mathbf{1}_{\{s\geq b\}}H_a \quad \text{and} \quad S^s_{T_b^s} = s\wedge b.$$

By Proposition 12.18, we have

$$P(s) = \sup_{b<K}(K-s\wedge b)^+\mathbb{E}^{\mathbb{Q}}\Big[e^{-rH_b\mathbf{1}_{\{s\geq b\}}}\Big] = \sup_{b<K}(K-s\wedge b)\mathbb{E}^{\mathbb{Q}}\Big[e^{-rH_b\mathbf{1}_{\{s\geq b\}}}\Big].$$

By considering the two alternative cases $s<K$ and $s\geq K$, we see that

$$P(s) = \sup_{b<s\wedge K}(K-b)\mathbb{E}^{\mathbb{Q}}\Big[e^{-rH_b}\Big] = \sup_{b<s\wedge K}(K-b)e^{\gamma\ln\frac{b}{s}} = \sup_{b<s\wedge K}(K-b)\Big(\frac{b}{s}\Big)^{\gamma}.$$

where $\gamma := \frac{\theta+\sqrt{\theta^2+2r}}{\sigma} = \frac{2r}{\sigma^2}$. The last optimization problem is easily solved, and we obtain the following explicit formula for the perpetual American put price $P(s)$ and the corresponding optimal stopping barrier s^*:

$$P(s) = (K-s)\mathbf{1}_{\{s\leq s^*\}} + \Big(\frac{K}{1+\gamma}\Big)^{1+\gamma}\Big(\frac{\gamma}{s}\Big)^{\gamma}\mathbf{1}_{\{s\geq s^*\}}, \quad \text{with} \quad s^* := \frac{\gamma}{1+\gamma}K.$$

12.4 Appendix: essential supremum

The notion of essential supremum has been introduced in probability in order to face the problem of maximizing random variables over an infinite family $\mathcal{Z}$. The problem arises when $\mathcal{Z}$ is not countable because then the supremum is not measurable, in general.

Theorem 12.20. *Let $\mathcal{Z}$ be a family of r.v. $Z:\Omega\longrightarrow\mathbb{R}\cup\{\infty\}$ on a probability space $(\Omega,\mathcal{F},\mathbb{P})$.*
(i) *Then there exists a unique (a.s.) r.v. $\bar{Z}:\Omega\longrightarrow\mathbb{R}\cup\{\infty\}$ such that:*
 (a) *$\bar{Z}\geq Z$, a.s. for all $Z\in\mathcal{Z}$,*
 (b) *For all r.v. Z' satisfying* (a)*, we have $\bar{Z}\leq Z'$, a.s.*
The r.v. $\bar{Z}$ is called the essential supremum of the family $\mathcal{Z}$, and denoted by $\operatorname{ess\,sup}\mathcal{Z}$.
(ii) *Moreover, there exists a sequence $(Z_n)_{n\in\mathbb{N}}\subset\mathcal{Z}$ such that* $\operatorname{ess\,sup}\mathcal{Z} = \sup_{n\in\mathbb{N}}Z_n$. *If the family $\mathcal{Z}$ has the lattice property, then such a sequence $(Z_n)_{n\in\mathbb{N}}\subset\mathcal{Z}$ may be chosen to be a.s. nondecreasing so that* $\operatorname{ess\,sup}\mathcal{Z} = \lim\uparrow_{n\to\infty}Z_n$.

Proof. By possibly considering a one-to-one transformation of the elements of $\mathcal{Z}$ to the interval $[0,1]$, we first observe that we may revert to the case where $\mathcal{Z}$ only consists take values in the interval $[0,1]$.

The uniqueness of $\bar{Z}$ is an immediate consequence of (b). To prove existence, we consider the set $\mathcal{D}$ of all countable subsets of $\mathcal{Z}$. For all $D\in\mathcal{D}$, we define $Z_D := \sup\{Z : Z\in D\}$, and we introduce the r.v. $\zeta := \sup\{\mathbb{E}[Z_D] : D\in\mathcal{D}\}$.
1. We first prove that there exists $D^*\in\mathcal{D}$ such that $\zeta = \mathbb{E}[Z_{D^*}]$. To see this, let $(D_n)_n\subset\mathcal{D}$ be a maximizing sequence, i.e. $\mathbb{E}[Z_{D_n}]\longrightarrow\zeta$, then $D^* := \cup_n D_n\in\mathcal{D}$

satisfies $\mathbb{E}[Z_{D^*}] = \zeta$. We denote $\bar{Z} := Z_{D^*}$.
2. It is clear that the r.v. $\bar{Z}$ satisfies (b). To prove that property (a) holds true, we consider an arbitrary $Z \in \mathcal{Z}$ together with the countable family $D := D^* \cup \{Z\} \subset \mathcal{D}$. Then $Z_D = Z \vee \bar{Z}$, and $\zeta = \mathbb{E}[\bar{Z}] \le \mathbb{E}[Z \vee \bar{Z}] \le \zeta$. Consequently, $Z \vee \bar{Z} = \bar{Z}$, and $Z \le \bar{Z}$, a.s.

Finally, the first part of the claim (ii) follows from the previous proof of existence. The second part is trivial. ♢

12.5 Problems

Exercise 12.21. In the context of Theorem 12.1, develop a localization argument to prove Claim (i) as a consequence of Claim (ii).

Exercise 12.22. In the context of the Black-Scholes model with constant coefficients, prove that the American put option price $P(t,s)$ is uniformly Lipschitz in s, and $1/2-$H\'older-continuous in t.

Exercise 12.23. Consider a financial market with a risky security and a non-risky asset with constant coefficients, i.e. Black and Scholes model for the security, and constant interest rate. The risky security pays dividends at the continuous rate proportional to the spot price so that the dividend is $\delta S_t dt$ during the time interval $[t, t+dt]$.

1. Provide the partial differential equation satisfied by the price of a perpetual American call option on the above security with maturity T and strike price K.

2. Recall the definition of the stopping region and the continuation region. Guess a reasonable form of the stopping region with convincing justification.

3. By guessing a solution to the valuation equation of the perpetual American call price, prove that:

$$C(S_0) = \frac{1}{\gamma_1}\left(\frac{\gamma_1}{\gamma_1 - 1}\,K\right)^{1-\gamma_1}\;S_0^{\gamma_1}$$

   where$\gamma_1 = \frac{-\left(\frac{2(r-\delta)}{\sigma^2}-1\right)+\sqrt{\left(\frac{2(r-\delta)}{\sigma^2}-1\right)^2+\frac{8r}{\sigma^2}}}{2}$, and provide the optimal exercise strategy.

Chapter 13

Gaussian interest rates models

In this chapter, we provide an introduction to the modelling of the term structure of interest rates. We will develop a pricing theory for securities that depend on default-free interest rates or bond prices. The general approach will exploit the fact that bonds of many different maturities are driven by a few common factors. Therefore, in contrast with the previous theory developed for a finite securities markets, we will be in the context where the number of traded assets is larger (in fact infinite) than the number of sources of randomness.

The first models introduced in the literature stipulate some given dynamics of the instantaneous interest rate process *under the risk neutral measure* $\mathbb{Q}$, which is assumed to exist. The prices of bonds of all maturities are then deduced by computing the expected values of the corresponding discounted payoff under $\mathbb{Q}$. We shall provide a detailed analysis of the most representative of this class, namely the Vasicek model. An important limitation of this class of models is that the yield curve predicted by the model does not match the observed yield curve, i.e. the calibration to the spot yield curve is not possible.

The Heath-Jarrow-Morton approach (1992) solves this calibration problem by taking the spot yield curve as the initial condition for the dynamics of the entire yield curve. The dynamics of the yield curve is driven by a finite-dimensional Broanian motion. In order to exclude all possible arbitrage opportunities, we will assume the existence of a risk neutral probability measure, for all bonds with all maturities. In the present context of a large financial market, This condition leads to the so-called Heath-Jarrow-Morton restriction which states that the dynamics of the yield curve is defined by the volatility process of zero-coupon bonds together with a risk premia process which is common to all bonds with all maturities.

Finally, a complete specification of an interest rates model requires the specification of the volatility of bonds. As in the context of finite securities markets, this is achieved by a calibration technique to the options markets. We therefore provide an introduction to the main tools for the analysis of fixed income derivatives. An important concept is the notion of *forward neutral measure*, which turns the forward price processes with pre-specified maturity into martingales. In the simplest models defined by deterministic volatilities of zero-coupon bonds, this allows to express the prices of European options on zero-coupon bonds in closed form by means of a Black-Scholes type of formula. The structure of implied volatilities extracted from these prices provides a powerfull tool for the calibration of the yield curve to spot interest rates and options.

13.1 Fixed income terminology

13.1.1 Zero-coupon bonds

Throughout this chapter, we will denote by $P_t(T)$ the price at time t of a pure discount bond paying 1 at date $T \geq t$. By definition, we have $P_T(T) = 1$.

In real financial markets, the prices $P_t(T_i)$ are available, at each time t, for various maturities T_i, $i = 1, \ldots, k$. We shall see later that these data are directly available for maturities shorter than one year, and can be extracted from bond prices for larger maturities by the so-called boostrapping technique.

Since the integer k recording the number of maturities is typically large, the models developed below allow for trading the zero-coupon bonds $P_t(T)$ for any $T > t$. We are then in the context of infinitely many risky assets. This is a first major difference with the theory of derivative securities on stocks developed in previous chapters.

The second specificity is that the zero-coupon bond with price $P_t(T)$ today will be a different asset at a later time $u \in [t, T]$ as its time to maturity is shortened to $T - u$. This leads to important arbitrage restrictions.

Given the prices of all zero-coupon bonds, one can derive an un-ambiguous price of any deterministic income stream: consider an asset which pays F_i at each time t_i, $i = 1, \ldots, n$. Then the no-arbitrage price at time 0 of this asset is

$$\sum_{i=1}^{n} F_i \; P_0(t_i).$$

If F_i is random, the above formula does not hold true, as the correlation between F_i and interest rates enters into the picture.

Coupon-bearing bonds are quoted on financial markets. Their prices are obviously related to the prices of zero-coupon bonds by

$$P_0 \;=\; \sum_{i=1}^{n} c \, P_0(T_i) \;+\; K \; P_0(T) \;=\; \sum_{i=1}^{n} \rho K \; P_0(T_i) \;+\; K \; P_0(T)$$

where $c = \rho K$ is the coupon corresponding to the pre-assigned interest $\rho > 0$, $T_1 \leq \ldots \leq T_n \leq T$ are the dates where the coupons are paid, and K is the Principal (or face value) to be paid at the maturity T.

The yield to maturity is defined as the (unique !) scalar Y_0 such that

$$P_0 \quad = \quad \sum_{i=1}^{n} ce^{-Y_0 T_i} \; + \; Ke^{-Y_0 T}$$

The bond is said to be priced

- at par if $\rho = Y_0$,
- below par if $\rho < Y_0$,
- above par if $\rho > Y_0$.

Only short term zero-coupon bonds are quoted on the market (less than one-year maturity). Zero-coupon bonds prices are inferred from coupon-bearing bonds (or interest rates swaps introduced below).

On the US market, Government debt securities are called:

- Treasury bills (T-bills): zero-coupon bonds with maturity ≤ 1 year,
- Treasury notes (T-notes): coupon-bearing with maturity ≤ 10 years,
- Treasury bonds (T-bonds): coupon-bearing with maturity > 10 years.

A government bond is traded in terms of its price which is quoted in terms of its face value.

13.1.2 Interest rates swaps

Let $T_0 > 0$, $\delta > 0$, $T_i = T_0 + i\delta$, $i = 1 \ldots, n$, and denote by $\mathbf{T} := \{T_0 < T_1 < \ldots < T_n\}$ the set of such defined maturities. We denote by $L\,(T_{j-1})$ the *LIBOR rate* at time T_{j-1}, i.e. the floating rate received at time T_j and set at time T_{j-1} by reference to the price of the zero-coupon bond over that period:

$$P_{T_{j-1}}\,(T_j) \quad = \quad \frac{1}{1 + \delta L\,(T_{j-1})}\,.$$

The interest rate swap is defined by the comparison of the two following streams of payments:

- *the floating leg*: consists of the payments $\delta L\,(T_{j-1})$ at each maturity T_j for $j = 1, \ldots, n$, and the unit payment (1) at the final maturity T_n,
- *the fixed leg*: consists of the payments $\delta\kappa$ at each maturity T_j for $j = 1, \ldots, n$, and the unit payment (1) at the final maturity T_n, for some given constant rate κ.

The interest rate swap rate correponding to the set of maturities $\mathbf{T}$ is defined as the constant rate κ which equates the value at time T_0 of the above floating and fixed leg. Direct calculation leads to the following expression of the swap rate:

$$\kappa_{T_0}\,(\delta, n) \quad = \quad \frac{1 - P_{T_0}(T_n)}{\delta\, \sum_{j=1}^{n} P_{T_0}(T_j)}\,.$$

We leave the verification of this formula as an exercise.

13.1.3 Yields from zero-coupon bonds

We define the yields corresponding to zero-coupon bonds by

$$P_t(T) \;=\; e^{-(T-t)R_t(T)}\,, \quad \text{i.e.} \quad R_t(T) \;:=\; \frac{-\ln P_t(T)}{T-t}\,.$$

The term structure of interest rates represents at each time t the curve $T \longmapsto R_t(T)$ of yields on zero-coupon bonds for all maturities $T > 0$. It is also commonly called *the yields curve.*

In practice, there are many different term structures of interest rates, depending on whether zero-coupon bonds are deduced

- from observed bonds prices
- from observed swaps prices
- from bonds issued by the government of some country, or by a firm with some confidence on its liability (rating).

We conclude this section by the some stylized facts observed on real financial markets data:
a- The term structure of interest rates exhibits different shapes: (almost) flat, increasing (most frequently observed), decreasing, decreasing for short term maturities then increasing, increasing for short term maturities then decreasing. Examples of observed yield curves are displayed in Figure 13.1 below.
b- Interest rates are positive: we shall however make use of gaussian models which offer more analytic solutions, although negative values are allowed by such models (but with very small probability).
c- Interest rates exhibit *mean reversion*, i.e. oscillate around some average level and tend to be attracted to it. See Figure 13.2 below.
d- Interest rates for various maturities are not perfectly correlated.
e- Short term interest rates are more volatile than long term interest rates. See Figure 13.3

13.1.4 Forward Interest Rates

The forward rate $F_t(T)$ is the rate at which agents are willing, at date t, to borrow or lend money over the period $[T, T+h]$ for $h \searrow 0$. It can be deduced directly from the zero-coupon bonds $\{P_t(T),\ T \geq t\}$ by the following no-arbitrage argument:

- start from the initial capital $P_t(T)$, and lend it over the short periods $[t, t+h]$, $[t+h, t+2h]$, ..., at rates agreed now; for $h \searrow 0$, this strategy yields the payoff $P_t(T)e^{\int_t^T F_t(u)du}$.
- Since the alternative strategy of buying one discount bond costs $P_t(T)$ yields a certain unit payoff (1), it must be the case that

$$P_t(T) \quad = \quad e^{-\int_t^T F_t(u)du}\,.$$

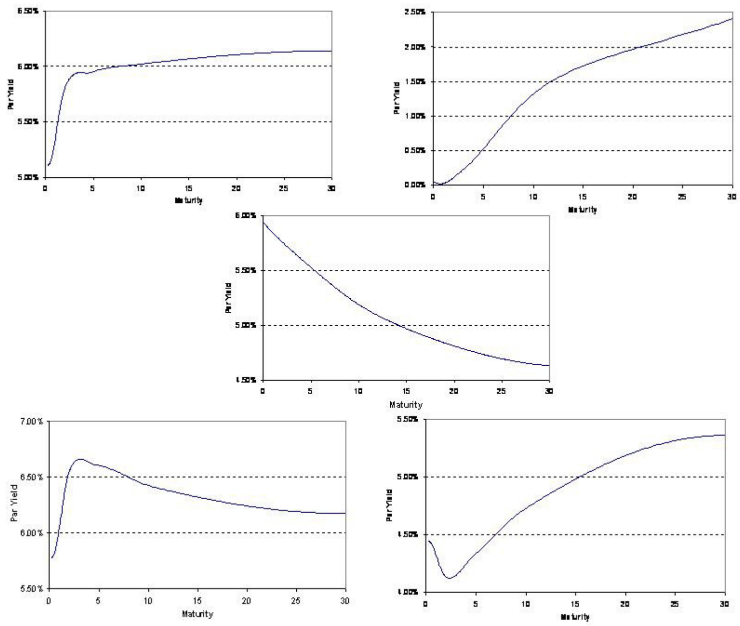

Figure 13.1: Various shapes of yield curves from real data

In terms of yields to maturity, the above relation can be rewritten in:

$$R_t(T) \quad = \quad \frac{1}{T-t}\int_t^T F_t(u)du\,.$$

So zero-coupon bonds and the corresponding yields can be defined from forward rates. Conversely, forward rates can be obtained by

$$F_t(T) \;=\; \frac{\partial}{\partial T}\left\{(T-t)R_t(T)\right\} \quad \text{and} \quad F_t(T) \;=\; \frac{-\partial}{\partial T}\left\{\ln P_t(T)\right\}$$

13.1.5 Instantaneous interest rates

The instantaneous interest rate is given by

$$r_t \quad = \quad R_t(t) \;=\; F_t(t)\,.$$

It does not correspond to any tradable asset, and is not directly observable. However, assuming that the market admits a risk neutral measure $\mathbb{Q}$, it follows from the valuation theory developed in the previous chapter that:

$$P_t(T) \quad = \quad \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_u du}\middle|\,\mathcal{F}_t\right]\,.$$

Hence, given the dynamics of the instantaneous interest rates under $\mathbb{Q}$, we may deduce the prices of zero-coupon bonds.

Figure 13.2: Mean reversion of interest rates

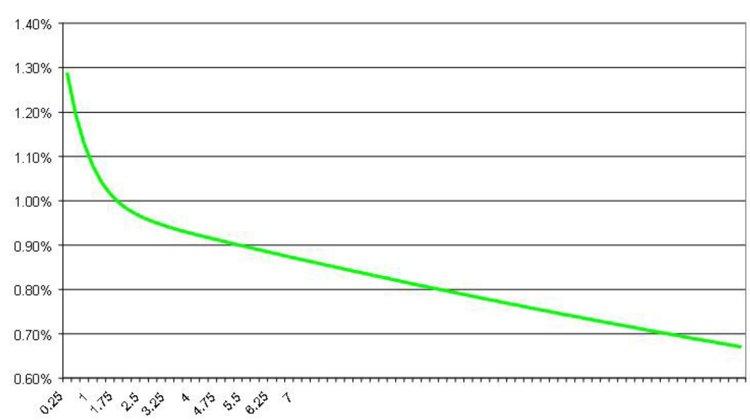

Figure 13.3: Volatility of interest rates is decreasing in terms of maturity

13.2 The Vasicek model

Consider the process ρ defined by:

$$\begin{aligned}\rho_t &= \rho_0 + m\lambda \int_0^t e^{\lambda t} dt + \sigma \int_0^t e^{\lambda t} dB_t \\ &= \rho_0 + m\left(e^{\lambda t} - 1\right) + \sigma \int_0^t e^{\lambda t} dB_t\,,\end{aligned}$$

where B is a Brownian motion under the risk neutral measure $\mathbb{Q}$. Observe that the above stochastic integral is well-defined by Theorem 6.1. The Vasicek model (1977) assumes that the the instantaneous interest rate is given by:

$$r_t \ := \ \rho_t e^{-\lambda t} \quad \text{for} \quad t \geq 0\,.$$

An immediate application of Itô's formula provides the following dynamics of the interest rates process:

$$dr_t \ = \ \lambda(m - r_t)dt + \sigma dB_t\,.$$

this is the so-called Ornstein-Uhlenbeck process in the theory of stochastic processes. Observe that this process satisfies the mean reversion property around the level m with intensity λ:
• if $r_t < m$, then the drift is positive, and the interest rate is pushed upward with the intensity λ,
• if $r_t > m$, then the drift is negative, and the interest rate is pushed downward with the intensity λ.

The process $\{r_t, t \geq 0\}$ is explicitly given by

$$r_t \quad = \quad m + (r_0 - m)\, e^{-\lambda t} + \sigma \int_0^t e^{-\lambda(t-u)} dB_u\,. \tag{13.1}$$

Using the Itô isometry, this shows that $\{r_t, t \geq 0\}$ is a gaussian process with mean

$$\mathbb{E}^{\mathbb{Q}}[r_t] \ = \ m + (r_0 - m)\, e^{-\lambda t}\,, \qquad t \geq 0\,,$$

and covariance function

$$\begin{aligned} \mathbb{C}ov^{\mathbb{Q}}[r_t, r_s] \quad &= \quad \mathbb{E}^{\mathbb{Q}}\left[\int_0^t e^{-\lambda(t-u)} dB_u \int_0^s e^{-\lambda(s-u)} dB_u\right] \\ &= \quad \frac{\sigma^2}{2\lambda}\left(e^{-\lambda|t-s|} - e^{-\lambda(t+s)}\right)\,, \quad \text{for } s, t \geq 0\,. \end{aligned}$$

In particular, this model allows for negative interest rates with positive (but small) probability !

For fixed $t > 0$, the distribution under $\mathbb{Q}$ of r_t is $\mathbf{N}\left(\mathbb{E}^{\mathbb{Q}}[r_t], \mathbb{V}^{\mathbb{Q}}[r_t]\right)$, which converges to the stationary distribution

$$\mathbf{N}\left(m, \frac{\sigma^2}{2\lambda}\right) \quad \text{as} \quad t \longrightarrow \infty\,.$$

13.3 Zero-coupon bonds prices

Recall that the price at time 0 of a zero-coupon bond with maturiyt T is given by

$$P_0(T) \quad = \quad \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r_t dt}\right]\,.$$

In order to develop the calculation of this expectation, we now provide the distribution of the random variable $\int_0^T r_t dt$ by integrating (13.1):

$$\int_0^T r_t dt = mT + (r_0 - m)\int_0^T e^{-\lambda t}dt + \sigma \int_0^T \int_0^t e^{-\lambda(t-u)}dB_u \quad (13.2)$$

$$= mT + (r_0 - m)\frac{1-e^{-\lambda T}}{\lambda} + \sigma \int_0^T \int_0^t e^{-\lambda(t-u)}dB_u \,. \quad (13.3)$$

In order to derive the distribution of the last double integral, we need to reverse the order of integration. To do this, we introduce the process $Y_t := \int_0^t e^{\lambda u} dB_u$, $t \geq 0$, and we compute by Itô's formula that

$$d\left(e^{-\lambda t}Y_t\right) = e^{-\lambda t}dY_t - \lambda e^{-\lambda t}Y_t dt = dB_t - \lambda e^{-\lambda t}Y_t dt \,.$$

Integrating between 0 and T and recalling the expression of Y_t, this provides:

$$\int_0^T \lambda e^{-\lambda t}\int_0^t e^{\lambda u}dB_u dt = B_T - e^{-\lambda T}Y_T = \int_0^T \left(1 - e^{-\lambda(T-t)}\right) dB_t \,.$$

Plugging this expression into (13.2), we obtain:

$$\int_0^T r_t dt = mT + (r_0 - m)\,\lambda(T) + \sigma \int_0^T \Lambda(T-t)dB_t$$

where

$$\Lambda(u) := \frac{1-e^{-\lambda u}}{\lambda} \,.$$

This shows that

$$\int_0^T r_t dt \quad \text{is distributed as} \quad \mathcal{N}\left(\mathbb{E}^{\mathbb{Q}}\left[\int_0^T r_t dt\right], \mathbb{V}^{\mathbb{Q}}\left[\int_0^T r_t dt\right]\right), \quad (13.4)$$

where

$$\mathbb{E}^{\mathbb{Q}}\left[\int_0^T r_t dt\right] = mT + (r_0 - m)\,\Lambda(T)$$

$$\mathbb{V}^{\mathbb{Q}}\left[\int_0^T r_t dt\right] = \sigma^2 \int_0^T \Lambda(u)^2 du \,.$$

Given this explicit distribution, we can now compute the prices at time zero of the zero-coupon bonds:

$$P_0(T) = \exp\left[-mT - (r_0 - m)\,\Lambda(T) + \frac{\sigma^2}{2}\int_0^T \Lambda(u)^2 du\right] \quad \text{for} \quad T \geq 0 \,.$$

Since the above Vasicek model is time-homogeneous, we can deduce the price at any time $t \geq 0$ of the zero-coupon bonds with all maturities:

$$P_t(T) \;=\; \exp\left[-m(T-t) - (r_t - m)\,\Lambda(T-t) + \frac{\sigma^2}{2}\int_0^{T-t}\Lambda(u)^2 du\right] \tag{13.5}$$

for $T \geq t \geq 0$. The term structure of interest rates is also immediately obtained:

$$R_t(T) \;=\; \frac{-\ln P_t(T)}{T-t} \;=\; m + (r_t - m)\,\frac{\Lambda(T-t)}{T-t} - \frac{\sigma^2}{2(T-t)}\int_t^T \Lambda(u)^2 du \tag{13.6}$$

Exercise 13.1. Show that the joint distribution of the pair $\left(r_T\,,\,\int_0^T r_t dt\right)$ is gaussian, and provide its characteristics in explicit form. *Hint*: compute its Laplace transform.

We conclude this section by deducing from (13.5) the dynamics of the price process of the zero-coupon bonds, by a direct application of Itô's formula. An important observation for this calculation is that the drift term in this differential representaion is already known to be $dP_t(T) = P_t(T)r_t dt + \cdots dB_t$, since $\{P_t(T), 0 \leq t \leq T\}$ is the price of a security traded on the financial market. Therefore, we only need to compute the volatility coefficient of the zero-coupon price process. This is immediately obtained from (13.5):

$$\frac{dP_t(T)}{P_t(T)} \;=\; r_t dt - \sigma\Lambda(T-t)dB_t\,, \qquad t < T\,.$$

13.4 Calibration to the spot yield curve and the generalized Vasicek model

An important requirement that the interest rate must satisfy is to reproduce the observed market data for the zero-coupon bond prices

$$B_0^*(T)\,, \qquad T \geq 0\,,$$

or equivalently, the spot yield curve at time zero

$$R_0^*(T)\,, \qquad T \geq 0\,,$$

or equivalently the spot forward rates curve

$$F_0^*(T)\,, \qquad T \geq 0\,,$$

In practice, the prices of the zero-coupon bonds for some given maturities are either observed (for maturities shorter than one year), or extracted from coupon-bearing bonds or interest rates swaps; the yield curve is then constructed by an interpollation method.

Since the Vasicek model is completely determined by the choice of the four parameters r_0, λ, m, σ, there is no hope for the yield curve (13.6) predicted by this model to match some given observe spot yield curve $R_0^*(T)$ for every $T \geq 0$. In other words, the Vasicek model can not be calibrated to the spot yield curve.

Hull and White (1992) suggested a slight extension of the Vasicek model which solves this calibration problem. In order to meet the infinite number of constraints imposed by the calibration problem, they suggest to model the instantaneous interest rates by

$$r_t \;:=\; \rho_t e^{-\lambda t} \quad \text{where} \quad \rho_t \;=\; \rho_0 + \lambda \int_0^t m(t) e^{\lambda t} dt + \sigma \int_0^t e^{\lambda t} dB_t\,,$$

which provides the dynamics of the instantaneous interest rates:

$$dr_t \;=\; \lambda\,(m(t) - r_t)\,dt + \sigma dB_t\,, \tag{13.7}$$

Here, the keypoint is that $m(.)$ is a deterministic function to be determined by the calibration procedure. This extension increases the number of parameters of the model, while keeping the main features of the model: mean reversion, gaussian distribution, etc...

All the calculations of the previous section can be reproduced in this context.

1. We first explicitly integrate the stochastic differential equation (13.7):

$$r_t \;=\; r_0 e^{-\lambda t} + \lambda \int_0^t m(s) e^{-\lambda(t-s)} ds + \sigma \int_0^t e^{-\lambda(t-s)} dB_s.$$

Integrating this expression between t and T, and exchanging the order of the integration as in the context of the Vasicek model, this provides:

$$\begin{aligned}\int_0^T r_t dt \;&=\; r_0 \int_0^T e^{-\lambda t} dt + \lambda \int_0^T\!\!\int_0^t m(s) e^{-\lambda(t-s)} ds dt + \sigma \int_0^T\!\!\int_0^t e^{-\lambda(t-s)} dB_s dt \\ &=\; r_0 \Gamma(T) + \lambda \int_0^T m(s)\Gamma(T-s) ds + \sigma \int_0^T \Gamma(T-s) dB_s,\end{aligned}$$

where $\Gamma(u) := \left(1 - e^{-\lambda u}\right)/\lambda$.

2. By the previous calculation, we deduce that:

$$P_0(T) \;=\; \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r_t dt}\right] \;=\; e^{-r_0\Gamma(T) + \lambda \int_0^T m(t)\Gamma(T-t)dt + \frac{1}{2}\sigma^2 \int_0^T \Gamma(T-t)^2 dt}.$$

Since $\Gamma(0) = 0$, this shows that:

$$\begin{aligned}F_0(T) \;&=\; r_0 \Gamma(T) + \lambda \int_0^T m(t)\Gamma'(T-t) dt - \frac{1}{2}\sigma^2 \int_0^T (2\Gamma\Gamma')(T-t) dt \\ &=\; r_0 e^{-\lambda T} + \lambda \int_0^T m(t) e^{-\lambda(T-t)} dt - \frac{1}{2}\sigma^2 \Gamma(T)^2.\end{aligned}$$

3. We now impose that $F_0(T) = F_0^*(T)$ for all $T \geq 0$, where F_0^* is the observed market forward rate curve. Then:

$$\int_0^T \lambda e^{\lambda t} m(t)dt \quad = \quad r_0 + e^{\lambda T}\Big[F_0(T) + \frac{1}{2}\sigma^2\Gamma(T)^2\Big], \quad T \geq 0.$$

Then, the calibration to the spot forward curve is possible by chosing:

$$m(T) \; = \; \frac{e^{-\lambda T}}{\lambda}\frac{\partial}{\partial T}\left\{e^{\lambda T}\Big[F_0^*(T) + \frac{1}{2}\sigma^2\Gamma(T)^2\Big]\right\}.$$

Of course, such a calibration must be updated at every time instant. This means that the coefficient $m(.)$ which is supposed to be deterministic, will typically be fixed at every time instant by the calibration procedure. Hence, similarly to the implied volatility parameter in the case of European options on stocks, the Hull-White model is based on a gaussian model, but its practical implementation violates its founding assumptions by allowing for a stochastic evolution of the mean reversion level $m(.)$.

13.5 Multiple Gaussian factors models

In the one factor Vasicek model and its Hull-White extension, all the yields-to-maturity $R_t(T)$ are linear in the spot interest rate r_t, see (13.6). An immediate consequence of this model is that yields corresponding to different maturities are perfectly correlated:

$$\mathbb{C}or^{\mathbb{Q}}\left[R_t(T), R_t(T')\right|\mathcal{F}_t] \quad = \quad 1$$

which is not consistent with empirical observation. The purpose of this section is to introduce a simple model which avoids this perfect correlation, while keeping the analytical tractability: the two-factor Hull-White model.

Let X and Y be two factors driven by the centred Vasicek model:

$$\begin{aligned} dX_t &= -\lambda X_t dt + \sigma dB_t \\ dY_t &= -\theta Y_t dt + \xi dB'_t \end{aligned}$$

Here λ, σ, μ, and ξ are given parameters, and B, B' are two independent Brownian motions under the risk neutral measure $\mathbb{Q}$. The instantaneous interest rate is modelled as an affine function of the factors X, Y:

$$r_t \quad = \quad a(t) \; + \; X_t \; + \; Y_t$$

This two-factor model is then defined by four parameters and one deterministic function $a(t)$ to be determined by calibration on the market data.

Exploiting the independence of the factors X and Y, we immediately compute that

$$\begin{aligned} P_t(T) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_u du}\middle|\mathcal{F}_t\right] \\ &= e^{-\int_0^T a(t)dt}\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T X_u du}\middle|\mathcal{F}_t\right]\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T Y_u du}\middle|\mathcal{F}_t\right] \\ &= \exp\left[-A(t,T) - \Lambda(T-t)X_t - \Theta(T-t)Y_t\right] \end{aligned}$$

where

$$A(t,T) \quad := \quad \int_t^T a(u)du - \frac{\sigma^2}{2}\int_t^T \Lambda(u)^2 du - \frac{\xi^2}{2}\int_t^T \Theta(u)^2 du$$

and

$$\Lambda(u) := \frac{1-e^{-\lambda u}}{\lambda}, \qquad \Theta(u) := \frac{1-e^{-\theta u}}{\theta}.$$

The latter explicit expression is derived by analogy with the previous computation in (13.5). The term structure of interest rates is now given by

$$R_t(T) \quad = \quad \frac{1}{T-t}\left[A(t,T) + \Lambda(T-t)X_t + \Theta(T-t)Y_t\right]$$

and the yields with different maturities are not perfectly correlated. Further explicit calculation can be performed in order to calibrate this model to the spot yield curve by fixing the function $a(.)$. We leave this calculation as an exercise for the reader.

We finally comment on the interpretation of the factors. It is usually desirable to write the above model in terms of factors which can be identified on the financial market. A possible parameterization is obtained by projecting the model on the *short and long rates*. This is achieved as follows:

- By direct calculation, we find the expression of the short rate in terms of the factors:

$$r_t \quad = \quad \lambda X_t + \theta Y_t + A_T(t,t),$$

  where

$$A_T(t,t) \quad = \quad \left.\frac{\partial A}{\partial T}\right|_{T=t} \quad = \; a(t) - \frac{\sigma^2}{2}\Lambda(t)^2 - \frac{\xi^2}{2}\Theta(t)^2 .$$

- Fix some positive time-to-maturity $\bar{\tau}$ (say, 30 years), and let

$$\ell_t \quad = \quad R_t(t+\bar{\tau}),$$

  represent the long rate at time t. In the above two-factors model, we have

$$\ell_t \quad = \quad \frac{1}{\bar{\tau}}\left[A(t,t+\bar{\tau}) + \Lambda(\bar{\tau})X_t + \Theta(\bar{\tau})Y_t\right]$$

- In order to express the factors in terms of the short and the long rates, we now solve the linear system

$$\begin{pmatrix} r_t \\ \ell_t \end{pmatrix} \quad = \quad D\begin{pmatrix} X_t \\ Y_t \end{pmatrix} + \begin{pmatrix} A_T(t,t) \\ \frac{A(t,t+\bar{\tau})}{\bar{\tau}} \end{pmatrix} \tag{13.8}$$

where

$$D \quad := \quad \begin{pmatrix} \lambda & \theta \\ \frac{\Lambda(\bar{\tau})}{\bar{\tau}} & \frac{\Theta(\bar{\tau})}{\bar{\tau}} \end{pmatrix}.$$

Assuming that $\lambda \neq \theta$ and $\lambda\theta \neq 0$, it follows that the matrix D is invertible, and the above system allows to express the factors in terms of the short and long rates:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \quad = \quad D^{-1}\left[\begin{pmatrix} r_t \\ \ell_t \end{pmatrix} - \begin{pmatrix} A_T(t,t) \\ A(t,t+\bar{\tau})/\tau \end{pmatrix}\right]. \tag{13.9}$$

Finally, using the dynamics of the factor (X,Y), it follows from (13.8) and (13.9) that the dynamics of the short and long rates are given by:

$$\begin{pmatrix} dr_t \\ d\ell_t \end{pmatrix} \quad = \quad K\left[b(t,\bar{\tau}) - \begin{pmatrix} r_t \\ \ell_t \end{pmatrix}\right]dt + D\begin{pmatrix} \sigma dB_t \\ \xi dB'_t \end{pmatrix}$$

where

$$K \quad := \quad D\begin{pmatrix} \lambda & 0 \\ 0 & \theta \end{pmatrix}D^{-1},$$

and

$$b(t,\bar{\tau}) \quad := \quad K^{-1}\begin{pmatrix} (A_t + A_T)(t,t) \\ \frac{(A_{tT}+A_{TT})(t,t+\bar{\tau})}{\bar{\tau}} \end{pmatrix} + \begin{pmatrix} A_T(t,t) \\ \frac{A(t,t+\bar{\tau})}{\bar{\tau}} \end{pmatrix}.$$

Notice from the above dynamics that the pair (r_t, ℓ_t) is a two-dimensional Hull-While model with mean reversion towrd $b(t, t+\bar{\tau})$.

Exercise 13.2. *Repeat the calculations of this section in the case where* $\mathbb{E}[B_t B'_t] = \rho t$, $t \geq 0$, *for some* $\rho \in (-1,1)$.

13.6 Introduction to the Heath-Jarrow-Morton model

13.6.1 Dynamics of the forward rates curve

These models were introduced in 1992 in order to overcome the two following shortcomings of factor models:

• Factor models ignore completely the distribution under the statistical measure, and take the existence of the risk neutral measure as granted. In particular, this implies an inconsistency between these models and those built by economists for a predictability purpose.

• The calibration of factor models to the spot yield curve is artificial. First, the structure of the yield curve implied by the model has to be computed, then the parameters of the model have to be fixed so as to match the observed yield curve. Furthermore, the calibration must be repeated at any instant in time leading to an inconsistency of the model.

Heath-Jarrow-Morton suggest to directly model the dynamics of the observable yield curve. Given the spot forward rate curve, $F_0(T)$ for all maturities $0 \leq T \leq \bar{T}$, the dynamics of the forward rate curve is defined by:

$$\begin{aligned} F_t(T) &= F_0(T) + \int_0^t \alpha_u(T)du + \int_0^t \sigma_u(T) \cdot dW_u \\ &= F_0(T) + \int_0^t \alpha_u(T)du + \sum_{i=1}^n \int_0^t \sigma_u^i(T)dW_u^i \end{aligned}$$

where W is a Brownian motion under the statistical measure $\mathbb{P}$ with values in $\mathbb{R}^n$, and $\{\alpha_t(T), t \leq \bar{T}\}$, $\{\sigma_t^i(T), t \leq \bar{T}\}$, $i = 1, \ldots, n$, are adapted processes for every fixed maturity T. Throughout this section, we will assume that all stochastic integrals are well-defined, and we will ignore all technical conditions needed for the subsequent analysis.

13.6.2 The Heath-Jarrow-Morton drift condition

The first important question is whether such a model allows for arbitrage. Indeed, one may take advantage of the infinite number of assets available for trading in order to build an arbitrage opportunity. To answer the question, we shall derive the dynamics of the price process of zero-coupon bonds, and impose the existence of a risk neutral measure for these tradable securities. This is a sufficient condition for the absence of arbitrage opportunities, as the discounted wealth process corresponding to any portfolio strategy would be turned into a local martingale under the risk neutral measure, hence to a supermartingale thanks to the finite credit line condition. The latter supermartingale property garantees that no admissible portfolio of zero-coupon bonds would lead to an arbitrage opportunity.

We first observe that

$$\begin{aligned} d\int_t^T F_t(u)du &= -F_t(t)dt + \int_t^T dF_t(u)du \\ &= -r_tdt + \int_t^T \alpha_t(u)dudt + \int_t^T \sigma_t(u)du \cdot dW_t \\ &= -r_tdt + \gamma_t(T)dt + \Gamma_t(T)dW_t \end{aligned}$$

where we introduced the adapted processes:

$$\gamma_t(T) = \int_t^T \alpha_t(u)du \quad \text{and} \quad \Gamma_t(T) = \int_t^T \sigma_t(u)du$$

Since $P_t(T) = e^{-\int_t^T F_t(u)du}$, it follows from Itô's formula that

$$dP_t(T) = P_t(T)\left[\left(r_t - \gamma_t(T) + \frac{1}{2}|\Gamma_t(T)|^2\right) dt - \Gamma_t(T) \cdot dW_t\right]$$

We next impose that the zero-coupon bond price satisfies the risk neutral dynamics:

$$\frac{dP_t(T)}{P_t(T)} \;=\; r_t\, dt \;-\; \Gamma_t(T) \cdot dB_t$$

where

$$B_t \;:=\; W_t + \int_0^t \lambda_u du\,, \qquad t \geq 0\,,$$

defines a Brownian motion under some risk neutral measure $\mathbb{Q}$, and the so-called risk premium adapted $\mathbb{R}^n-$valued process $\{\lambda_t, t \geq 0\}$ is independent of the maturity variable T. This leads to the Heath-Jarrow-Morton drift condition:

$$\Gamma_t(T) \cdot \lambda_t \;=\; \gamma_t(T) - \frac{1}{2}|\Gamma_t(T)|^2\,. \tag{13.10}$$

Recall that

$$\frac{\partial \gamma}{\partial T}(t,T) = \alpha(t,T) \quad \text{and} \quad \frac{\partial \Gamma}{\partial T}(t,T) = \sigma(t,T)\,.$$

Then, differentiating with respect to the maturity T, we see that

$$\sigma_t(T) \cdot \lambda_t \;=\; \alpha_t(T) - \sigma_t(T) \cdot \Gamma_t(T)\,,$$

and therefore

$$dF_t(T) \;=\; \alpha_t(T)dt + \sigma_t(T) \cdot dW_t \;=\; \sigma_t(T) \cdot \Gamma_t(T)dt + \sigma_t(T) \cdot dB_t$$

We finally derive the risk neutral dynamics of the instantaneous interest rate under the HJM drift restriction (13.10). Recall that:

$$r_T \;=\; F_T(T) \;=\; F_0(T) + \int_0^T \alpha_u(T)du + \int_0^T \sigma_u(T) \cdot dW_u$$

Then:

$$\begin{aligned} dr_T \;=\;& \frac{\partial}{\partial T}F_0(T)\; dT + \alpha_T(T)\; dT + \int_0^T \frac{\partial}{\partial T}\alpha_u(T)du\; dT \\ &+ \int_0^T \frac{\partial}{\partial T}\sigma_u(T) \cdot dW_u\; dT + \sigma_T(T) \cdot dW_T\,. \end{aligned}$$

Organizing the terms, we get:

$$dr_t \;=\; \beta_t\, dt \;+\; \sigma_t(t)\, dW_t$$

where

$$\beta_t \;=\; \frac{\partial}{\partial T}F_0(t) + \alpha_t(t) + \int_0^t \frac{\partial}{\partial T}\alpha_u(t)du + \int_0^t \frac{\partial}{\partial T}\sigma_u(t) \cdot dW_u$$

or, in terms of the $\mathbb{Q}-$Brownian motion:

$$dr_t = \beta_t^0 \, dt + \sigma_t(t) \, dB_t$$

where

$$\begin{aligned}
\beta_t^0 &= \frac{\partial}{\partial T} F_0(t) + \sigma_t(t)\Gamma_t(t) + \int_0^t \frac{\partial}{\partial T} \left(\sigma_u(t) \cdot \Gamma_u(t)\right) du + \int_0^t \frac{\partial}{\partial T} \sigma_u(t) \cdot dW_u \\
&= \frac{\partial}{\partial T} F_0(t) + \int_0^t \frac{\partial}{\partial T} \left(\sigma_u(t) \cdot \Gamma_u(t)\right) du + \int_0^t \frac{\partial}{\partial T} \sigma_u(t) \cdot dW_u \, .
\end{aligned}$$

13.6.3 The Ho-Lee model

The Ho and Lee model corresponds to the one factor case ($n = 1$) with a constant volatility of the forward rate:

$$dF_t(T) = \alpha_t(T)dt + \sigma dW_t = \sigma^2(T-t)dt + \sigma dB_t$$

The dynamics of the zero-coupon bond price is given by:

$$\frac{dP_t(T)}{P_t(T)} = r_t dt - \sigma(T-t)dB_t \quad \text{with} \quad r_t = F_0(t) + \frac{1}{2}\sigma^2 t + \sigma B_t$$

By the dynamics of the forward rates, we see that the only possible movements in the yield curve are parallel shifts, i.e. all rates along the yield curve fluctuate in the same way.

13.6.4 The Hull-White model

The Hull and White model corresponds to one-factor case ($n = 1$) with the following dynamics of the forward rates

$$dF_t(T) = \sigma^2 e^{-\lambda(T-t)} \frac{1 - e^{-\lambda(T-t)}}{k} dt + \sigma e^{-\lambda(T-t)} dB_t$$

The dynamics of the zero-coupon bond price is given by:

$$\frac{dP_t(T)}{P_t(T)} = r_t dt - \frac{\sigma}{\lambda}\left(1 - e^{-\lambda(T-t)}\right) dB_t$$

with

$$r_t = a(t) + \int_0^t \sigma e^{-\lambda(t-u)} dB_u \, ,$$

and

$$a(t) := F_0(t) + \frac{\sigma^2}{2\lambda^2}\left(e^{-2\lambda t} - 1\right) + \frac{1 - e^{-\lambda t}}{\lambda} \, .$$

This implies that the dynamics of the short rate are:

$$dr_t = \lambda\left(m(t) - r_t\right) dt + \sigma dB_t \quad \text{where} \quad \lambda a(t) = \lambda m(t) + m'(t)$$

13.7 The forward neutral measure

Let $T_0 > 0$ be some fixed maturity. The T_0-forward neutral measure $\mathbb{Q}^{T_0}$ is defined by the density with respect to the risk neutral measure $\mathbb{Q}$

$$\frac{d\mathbb{Q}^{T_0}}{d\mathbb{Q}} = \frac{e^{-\int_0^{T_0} r_t dt}}{P_0(T_0)},$$

and will be shown in the next section to be a powerful tool for the calculation of prices of derivative securities in a stochastic interest rates framework.

Proposition 13.3. *Let $M = \{M_t, 0 \le t \le T_0\}$ be an $\mathbb{F}-$adapted process, and assume that $\tilde{M}$ is a $\mathbb{Q}-$martingale. Then the process*

$$\phi_t := \frac{M_t}{P_t(T_0)}, \qquad 0 \le t \le T_0,$$

is a martingale under the T_0-forward neutral measure $\mathbb{Q}^{T_0}$.

Proof. We first verify that ϕ is $\mathbb{Q}^{T_0}-$integrable. Indeed:

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}^{T_0}}[|\phi_t|] &= P_0(T_0)^{-1}\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^{T_0} r_u du}\frac{|M_t|}{P_t(T_0)}\right] \\
&= P_0(T_0)^{-1}\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^{t} r_u du}|M_t|\right] = P_0(T_0)^{-1}\mathbb{E}^{\mathbb{Q}}\left[|\tilde{M}_t|\right] < \infty,
\end{aligned}$$

where the second equality follows from the tower property of conditional expectations. We next compute for $0 \le s < t$ that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}^{T_0}}[\phi_t|\mathcal{F}_s] &= \frac{\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^{T_0} r_u du}\frac{M_t}{P_t(T_0)}\Big|\mathcal{F}_s\right]}{\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^{T_0} r_u du}\Big|\mathcal{F}_s\right]} \\
&= \frac{\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_s^{T_0} r_u du}\frac{M_t}{P_t(T_0)}\Big|\mathcal{F}_s\right]}{\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_s^{T_0} r_u du}\Big|\mathcal{F}_s\right]} \\
&= \frac{\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_s^{t} r_u du}M_t\Big|\mathcal{F}_s\right]}{\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_s^{T_0} r_u du}\Big|\mathcal{F}_s\right]} \\
&= \frac{\mathbb{E}^{\mathbb{Q}}\left[\tilde{M}_t\Big|\mathcal{F}_s\right]}{P_s(T_0)} = \phi_s,
\end{aligned}$$

where we used the Bayes rule for the first equality and the tower property of conditional expectation in the third equality. ♢

The above result has an important financial interpretation. Let S be the price process of any tradable security. Then, the no-arbitrage condition ensure that $M = \tilde{S}$ is a martingale under some risk neutral measure $\mathbb{Q}$. By definition ϕ is the price process of the T_0-forward contract on the security S. Hence Proposition 13.3 states that

The price process of the T_0-forward contract on any tradable security is a martingale under the T_0-forward measure $\mathbb{Q}^{T_0}$.

We continue our discussion of the T_0-forward measure in the context of the gaussian Heath-Jarrow-Morton model for the zero-coupon bond prices:

$$\frac{dP_t(T)}{P_t(T)} = r_t dt - \Lambda(T-t)\sigma dB_t \quad \text{where} \quad \Lambda(u) = \frac{1-e^{-\lambda(u)}}{u},$$

which corresponds to the solution

$$P_t(T) = P_0(T) e^{\int_0^t \left(r_u - \frac{1}{2}\sigma^2\Lambda(T-u)^2\right)du - \int_0^t \sigma\Lambda(T-u)dB_u}, \qquad 0 \leq t \leq T(13.11)$$

Recall that this model corresponds also to the Hull-White extension of the Vasicek model, up to the calibration to the spot yield curve. Since $P_T(T) = 1$, it follows from (13.11) that

$$\frac{d\mathbb{Q}^{T_0}}{d\mathbb{Q}} = \frac{e^{-\int_0^{T_0} r_u du}}{P_0(T_0)} = \exp\left(-\frac{1}{2}\int_0^{T_0} \sigma^2\Lambda(T_0-u)^2 du - \int_0^{T_0} \sigma\Lambda(T_0-u)dB_u\right),$$

and by the Cameron-Martin formula, we deduce that the process

$$W_t^{T_0} := B_t + \int_0^t \sigma\Lambda(T_0-u)du, \qquad 0 \leq t \leq T_0, \qquad (13.12)$$

is a Brownian motion under the T^0-forward neutral measure $\mathbb{Q}^{T_0}$.

13.8 Derivatives pricing under stochastic interest rates and volatility calibration

13.8.1 European options on zero-coupon bonds

The objective of this section is to derive a closed formula for the price of a European call option on a zero-coupon bond defined by the payoff at time $T_0 > 0$:

$$G := (P_{T_0}(T) - K)^+ \quad \text{for some} \quad T \geq T_0$$

in the context of the above gaussian Heath-Jarrow-Morton model.

We first show how the use of the forward measure leads to a substantial reduction of the problem. By definition of the T_0-forward neutral measure, the no-arbitrage price at time zero of the European call option defined by the above payoff is given by

$$p_0(G) \;=\; \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^{T_0} r_t dt}\,(P_{T_0}(T)-K)^+\right] \;=\; P_0(T_0)\mathbb{E}^{\mathbb{Q}^{T_0}}\left[(P_{T_0}(T)-K)^+\right].$$

Notice that, while firt expectation requires the knowledge of the joint distribution of the pair $\left(e^{-\int_0^{T_0} r_t dt}, P_{T_0}(T)\right)$ under $\mathbb{Q}$, the second expectation only requires the distribution of $P_{T_0}(T)$ under $\mathbb{Q}^{T_0}$. But, in view of Proposition 13.3, an additional simplification can be gained by passing to the price of the T_0-forward contract on the zero-coupon bond with maturity T:

$$\phi_t \;=\; \frac{P_t(T)}{P_t(T_0)}, \qquad 0 \le t \le T_0\,.$$

Since $P_{T_0}(T_0) = 1$, it follows that $\phi_{T_0} = P_{T_0}(T)$, and therefore:

$$p_0(G) \;=\; P_0(T_0)\mathbb{E}^{\mathbb{Q}^{T_0}}\left[(\phi_{T_0}-K)^+\right].$$

Since the process ϕ is a $\mathbb{Q}^{T_0}-$martingale by Proposition 13.3, we only need to compute the volatility of this process. An immediate calculation by means of Itô's formula shows that

$$\frac{d\phi_t}{\phi_t} \;=\; \sigma\left(\Lambda(T-t)-\Lambda(T_0-t)\right)dW_t^{T_0}\,.$$

By analogy with the previously derived Black-Scholes formula with deterministic coefficients (9.8), this provides:

$$\begin{aligned} p_0(G) &= P_0(T_0)\left[\phi_0\mathbf{N}\left(\mathbf{d}_+(\phi_0,K,v(T_0))\right) - K\mathbf{N}\left(\mathbf{d}_-(\phi_0,K,v(T_0))\right)\right] \\ &= P_0(T)\mathbf{N}\left(\mathbf{d}_+(P_0(T),\tilde{K},v(T_0))\right) - \tilde{K}\mathbf{N}\left(\mathbf{d}_-(P_0(T),\tilde{K},v(T_0))\right), \end{aligned} \tag{13.13}$$

where

$$\tilde{K} := KP_0(T_0)\,, \qquad v(T_0) := \sigma^2\int_0^{T_0}\left(\Lambda(T-t)-\Lambda(T_0-t)\right)^2 dt\,. \tag{13.14}$$

Given this simple formula for the prices of options on zero-coupon bonds, one can fix the parameters σ and λ so as to obtain the best fit to the observed options prices or, equivalently, the corresponding implied volatilities. Of course with only two free parameters (σ and λ) there is no hope to perfectly calibrate the model to the whole strucrue of the implied volatility surface. However, this can be done by a further extension of this model.

13.8.2 The Black-Scholes formula under stochastic interest rates

In this section, we provide an extension of the Black-Scholes formula for the price of a European call option defined by the payoff at some maturity $T > 0$:

$$G \;:=\; (S_T-K)^+ \quad \text{for some exercise price} \quad K > 0\,,$$

to the context of a stochastic interest rate. Namely, the underlying asset price process is defined by

$$S_t \;=\; S_0 e^{\int_0^T \left(r_u - \frac{1}{2}|\Sigma(u)|^2\right)du + \int_0^T \Sigma(u)\cdot dB_u}\,, \qquad t \geq 0\,,$$

where B is a Brownian motion in $\mathbb{R}^2$ under the risk neutral measure $\mathbb{Q}$, and $\Sigma = (\Sigma_1, \Sigma_2) : \mathbb{R}_+ \longrightarrow \mathbb{R}$ is a deterministic C^1 function. The interest rates process is defined by the Heath-Jarrow-Morton model for the prices of zero-coupon bonds:

$$P_t(T) \;=\; P_0(T) e^{\int_0^T \left(r_u - \frac{1}{2}\sigma^2 \Lambda(T-u)^2\right)du - \int_0^T \sigma\Lambda(T-u)dW_u^{0^1}}\,,$$

where

$$\Lambda(t) \;:=\; \frac{1 - e^{-\lambda t}}{\lambda}\,,$$

for some parameters $\sigma, \lambda > 0$. This models allows for a possible correlation between the dynamics of the underlying asset and the zero-coupon bonds.

Using the concept of formard measure, we re-write the no-arbitrage price of the European call option in:

$$\begin{aligned} p_0(G) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r_t dt}\,(S_T - K)^+\right] \\ &= P_0(T)\mathbb{E}^{\mathbb{Q}^T}\left[(S_T - K)^+\right] \;=\; P_0(T)\mathbb{E}^{\mathbb{Q}^T}\left[(\phi_T - K)^+\right]\,, \end{aligned}$$

where $\phi_t := P_t(T)^{-1} S_t$ is the price of the $T-$forward contract on the security S. By Proposition 13.3, the process $\{\phi_t, t \geq 0\}$ is a $\mathbb{Q}^T-$martingale measure, so its dynamics has zero drift when expressed in terms of the $\mathbb{Q}^T-$Brownian motion W^T. We then calculate the volatility component in its dynamics by means of Itô's formula, and we obtain:

$$\frac{d\phi_t}{\phi_t} \;=\; \left(\Sigma_1(t) + \sigma\Lambda(T-t)\right) dW^{T^1} + \Sigma_2(t) dW^{T^2}\,.$$

Hence, under the $T-$forward neutral measure $\mathbb{Q}^T$, the process ϕ follows a time-dependent Black-Scholes model with zero interest rate and time dependent squared volatility $\left(\Sigma_1(t) + \sigma\Lambda(T-t)\right)^2 + \Sigma_2(t)^2$. We can now take advantage of the calculation performed previously in (9.8), and conclude that

$$\begin{aligned} p_0(G) &= P_0(T)\left[\phi_0\, \mathbf{N}\left(\mathbf{d}_+(\phi_0, K, v(T))\right) - K\, \mathbf{N}\left(\mathbf{d}_-(\phi_0, K, v(T))\right)\right] \\ &= S_0\, \mathbf{N}\left(\mathbf{d}_+(S_0, \tilde{K}, v(T))\right) - \tilde{K}\, \mathbf{N}\left(\mathbf{d}_-(S_0, \tilde{K}, v(T))\right)\,, \end{aligned}$$

where

$$\tilde{K} := K P_0(T) \quad \text{and} \quad v(T) := \int_0^T \left(\left(\Sigma_1(t) + \sigma\Lambda(T-t)\right)^2 + \Sigma_2(t)^2\right) dt\,.$$

Chapter 14

Introduction to financial risk management

The publication by Harry Markowitz in 1953 of his dissertation on the theory of risk and return, has led to the understanding that financial institutions take risks as part of their everyday activities, and that these risks must therefore be measured and actively managed. The importance of risk managers within banks and investment funds has been growing ever since, and now the chief risk officer (CRO) is often a member of the top management committee. The job of a risk manager now requires sophisticated technical skills, and risk control departments of major banks employ many mathematicians and engineers.

The role of risk management in a financial institution has several important aspects. The first objective is to identify the risk exposures, that is the different types of risk (see below) which affect the company. These exposures should then be quantified, and measured. It is generally not possible to quantify the risk exposure with a single number, or even associate a probability distribution to it, since some uncertain outcomes cannot be assigned a probability in a reliable manner. Modern risk management usually combines a probabilistic approach (e.g. Value at Risk) with worst-case scenario analysis. These quantitative and qualitative assessments of risk exposures form the basis of reports to senior management, which must convey the global picture in a concise and non-technical way.

More importantly, the risk management must then design a *risk mitigation strategy*, that is, decide, taking into account the global constraints imposed by the senior management, which risk exposures are deemed acceptable, and which must be reduced, either by hedges or by limiting the size of the positions. In a trading environment, this strategy will result in precise exposure limits for each trading desk, in terms of the Value at Risk, the sensitivities to different risk factors, and the notional amounts for different products. For acceptable exposures, provisions will be made, in order to ensure the solvency of the bank if the corresponding risky scenarios are realized.

Finally, it is the role of risk management to monitor the implementation and performance of the chosen risk mitigation strategy, by validating the models and algorithms used by the front office for pricing and computing hedge ratios, and by double-checking various parameter estimates using independent data sources. This part in particular requires extensive technical skills.

With the globalization of the financial system, the bank risk management is increasingly becoming an international affair, since, as we have recently witnessed, the bankruptcy of a single bank or even a hedge fund can trigger financial turmoil around the world and bring the entire financial system to the edge of a collapse. This was the main reason for introducing the successive Basel Capital Accords, which define the best practices for risk management of financial institutions, determine the interaction between the risk management and the financial regulatory authorities and formalize the computation of the *regulatory capital*, a liquidity reserve designed to ensure the solvency of a bank under unfavorable risk scenarios. These agreements will be discussed in more detail in the last section of this chapter.

14.1 Classification of risk exposures

The different risk exposures faced by a bank are usually categorized into several major classes. Of course this classification is somewhat arbitrary: some risk types are difficult to assign a category and others may well belong to several categories. Still, some classification is important since it gives a better idea about the scope of possible risk sources, which is important for effective risk managemet.

14.1.1 Market risk

Market risk is probably the best studied risk type, which does not mean that it is always the easiest to quantify. It refers to the risk associated with movements of market prices of securities and rates, such as interest rate or exchange rate.

Different types of market risk are naturally classified by product class: interest rate risk, where one distinguishes the risk of overall movements of interest rates and the risk of the change in the shape of the yield curve; equity price risk, with a distinction between global market risk and idiosyncratic risk of individual stocks; foreign exchange risk etc. Many products will be sensitive to several kinds of market risk at the same time.

One can also distinguish the risk associated to the underlying prices and rates themselves, and the risk associated to other quantities which influence asset prices, such as volatility, implied volatility smile, correlation, etc. If the volatility risk is taken into accout by the modeling framework, such as, in a stochastic volatility model, it is best viewed as market risk, however if it is ignored by the model, such as in the Black-Scholes framework, then it will contribute to model risk. Note that volatility and correlation are initially interpreted as model parameters or non-observable risk factors, however, nowadays these parameters may be directly observable and interpretable via the quoted market prices of volatility / correlation swaps and the implied volatilities of vanilla options.

Tail risk Yet another type of market risk, the tail risk, or gap risk, is associated to large sudden moves (gaps) in asset prices, and is related to the fat tails of return distributions. A distribution is said to have fat tails if it assigns to very large or very small outcomes higher probability than the Gaussian distribution with the same variance. For a Gaussian random variable, the probability of a downside move greater than 3 standard deviations is 3×10^{-7} but such moves do happen in practice, causing painful losses to financial institutions. These deviations from Gaussianity can be quantified using the skewness $s(X)$ and the kurtosis $\kappa(X)$:

$$s(X) = \frac{E(X-EX)^3}{(\operatorname{Var} X)^{3/2}}, \qquad \kappa(X) = \frac{E(X-EX)^4}{(\operatorname{Var} X)^2}.$$

The skewness measures the asymmetry of a distribution and the kurtosis measures the 'fatness of tails'. For a Gaussian random variable X, $s(X) = 0$ and $\kappa(X) = 3$, while for stock returns typically $s(X) < 0$ and $\kappa(X) > 3$. In a recent study, κ was found to be close to 16 for 5-minute returns of S&P index futures.

Assume that conditionnally on the value of a random variable V, X is centered Guassian with variance $f(V)$. Jensen's inequality then yields

$$\kappa(X) = \frac{E[X^4]}{E[X^2]^2} = \frac{3E[f^2(V)]}{E[f(V)]^2} > 3.$$

Therefore, all conditionnally Gaussian models such as GARCH and stochastic volatility, produce fat-tailed distributions.

The tail risk is related to correlation risk, since large downward moves are usually more strongly correlated than small regular movements: during a systemic crisis all stocks fall together.

Managing the tail risk In a static framework (one-period model), the tail risk can be accounted for using fat-tailed distributions (Pareto). In a dynamic approach this is usually accomplished by adding stochastic volatility and/or jumps to the model, e.g. in the Merton model

$$\frac{dS_t}{S_{t-}} = \mu dt + \sigma_t dW_t + dJ_t,$$

where J_t is the process of log-normal jumps.

To take into account even more extreme events, to which it is difficult to assign a probability, one can add stress scenarios to models by using extreme values of volatility / correlations or taking historical data from crisis periods.

Brazil	BBB-	Japan	AA
China	A+	Russia	BBB
France	AAA	Tunisia	BBB
India	BBB-	United States	AAA

Table 14.1: Examples of credit ratings (source: Standard & Poor's, data from November 2009).

14.1.2 Credit risk

Credit risk is the risk that a default, or a change in credit quality of an entity (individual, company, or a sovereign country) will negative affect the value of the bank's portfolio. The portfolio may contain bonds or other products issued by that entity, or credit derivative products linked to that entity. One distinguishes the default risk from the spread[1] risk, i.e., risk of the depreciation of bonds due to a deterioration of the creditworthiness of their issuer, without default.

Credit rating The credit quality is measured by the *credit rating*: an evaluation of the creditworthiness of the borrower computed internally by the bank or externally by a rating agency. The credit ratings for sovereign countries and large corporations are computed by international rating agencies (the best known ones are Standard & Poor's, Moody's and Fitch ratings) and have letter designations. The rating scale of Standard and Poor's is AAA, AA, A, BBB, BB, B, CCC, CC, C, D, where AAA is the best possible rating and D corresponds to default. Ratings from AA to CCC may be further refined by the addition of a plus (+) or minus (-). Bonds with ratings above and including BBB- are considered 'investment grade' or suitable for long-term investment, whereas all others are considered 'speculative'. Table 14.1 reproduces the S&P credit ratings for several countries as of November 2009.

Credit derivatives The growing desire of the banks and other financial institutions to remove credit risk from their books to reduce regulatory capital, and more generally, to transfer credit risk to investors willing to bear it, has led to the appearance, in the late 90s of several classes of credit derivative products. The most widely used ones are Credit Default Swaps (CDS) and Collateralized Debt Obligations (CDO). The structure, the complexity and the role of these two types of instruments is entirely different. The credit default swap is designed to offer protection against the default of a single entity (let us call it Risky Inc.). The buyer of the protection (and the buyer of the CDS) makes regular premium payments to the seller of the CDS until the default of Risky Inc. or the maturity of the CDS. In exchange, when and if Risky Inc. defaults, the seller of the CDS makes a one-time payment to the seller to cover the losses from the default.

[1] The credit spread of an entity is defined as the difference between the yield to maturity of the bonds issued by the entity and the yield to maturity of the risk-free sovereign bonds.

The CDO is designed to transfer credit risk from the books of a bank to the investors looking for extra premium. Suppose that a bank owns a portfolio of defaultable loans (P). To reduce the regulatory capital charge, the bank creates a separate company called Special Purpose Vehicle (SPV), and sells the portfolio P to this company. The company, in turn, issues bonds (CDOs) which are then sold to investors. This process of converting illiquid loans into more liquid securities is known as securitization. The bonds issued by the SPV are divided onto several categories, or tranches, which are reimbursed in different order from the cash flows of the inital portfolio P. The bonds from the Senior tranche are reimbursed first, followed by Mezzanine, Junior and Equity tranches (in this order). The senior tranche thus (in theory) has a much lower default risk than the bonds in the original portfolio P, since it is only affected by defaults in P after all other tranches are destroyed. The tranches of a CDO are evaluated separately by rating agencies, and before the start of the 2008 subprime crisis, senior tranches received the highest ratings, similar to the bonds of the most financially solid corporations and sovereign states. This explained the spectacular growth of the CDO market with the global notional of CDOs issued in 2007 totalling to almost 500 billion US dollars. However, senior CDO tranches are much more sensitive to systemic risk and tend to have lower recovery rates[2] than corporate bonds in the same rating class. As a typical example, consider the situation when the portfolio P mainly consists of residential mortgages. If the defaults are due to individual circumstances of each borrower, the senior tranche will be protected by diversification effects. However, in case of a global downturn of housing prices, such as the one that happened in the US in 2007–2008, a large proportion of borrowers may default, leading to a severe depreciation of the senior tranche.

Counterparty risk The counterparty risk is associated to a default or a downgrade of a counterparty as opposed to the entity underlying a credit derivative product which may not necessarily be the counterparty. Consider a situation where a bank B holds bonds issued by company C, partially protected with a credit default swap issued by another bank A. In this case, the portfolio of B is sensitive not only to the credit quality of C but also to the credit quality of A, since in case of a default of C, A may not be able to meet its obligations on the CDS bought by B.

14.1.3 Liquidity risk

Liquidity risk may refer to asset liquidity risk, that is, the risk of not being able to liquidate the assets at the prevailing market price, because of an insufficient depth of the market, and funding liquidity risk, that is, the risk of not being able to raise capital for current operations. The standard practice of *marking to market* a portfolio of derivatives refers to determining the price of this portfolio

[2] The recovery rate of a bond is the proportion of the notional recovered by the lender after default.

for accounting purposes using the prevailing market price of its components. However, even for relatively liquid markets, where many buyers and sellers are present, and there is a well-defined market price, the number of buy orders close to this price is relatively small. To liquidate a large number of assets the seller will need to dig deep into the order book[3], obtaining therefore a much lower average price than if he only wanted to sell a single share (see fig. 14.1). This type of risk is even more important for illiquid assets, where the ballance-sheet price is computed using an internal model (marking to model). The model price may be very far from the actual price which can be obtained in the market, especially in the periods of high volatility. This is also related to the issue of model risk discussed below. The 2008 financial crisis started essentially as an asset liquidity crisis, when the market for CDOs suddenly shrunk, leading to massive depreciation of these products in the banks' ballance sheets. The fear of imminent default of major banks, created by these massive depreciations, made it difficult for them to raise money and created a funding liquidity crisis.

14.1.4 Operational risk

Operational risk refers to losses resulting from inadequate or failed internal processes, people and systems or from external events. This includes deliberate fraud by employees, several spectacular examples of which we have witnessed in the recent years. The Basel II agreement provides a framework for measuring operational risk and making capital provisions for it, but the implementation of these quantitative approaches faces major problems due to the lack of historical data and extreme heavy-tailedness of some types of operational risk (when a single event may destroy the entire institution). No provisions for operational risk can be substituted for strict and frequent internal controls, up-to-date computer security and expert judgement.

14.1.5 Model risk

Model risk is especially important for determining the prices of complex derivative products which are not readily quoted in the market. Consider a simple European call option. Even if it is not quoted, its price in the balance sheet may be determined using the Black-Scholes formula. This is the basis of a widely used technique known as marking to model, as opposed to marking to market. However, the validity of this method is conditionned by the validity of the Black-Scholes model assumptions, such as, for example, constant volatility. If the volatility changes, the price of the option will also change. More generally, model risk can be caused by an inadequate choice of the pricing model, parameter errors (due to statistical estimation errors or nonstationary parameters) and inadequate implementation (not necessarily programming bugs but perhaps an unstable algorithm which amplifies data errors). While the effect of parameter estimation errors may be quantified by varying the parameters

[3]The order book contains all outstanding limit orders at a given time.

within confidence bounds, and inadequate implementations can be singled out by scrupulous expert analysis, the first type of model risk (inadequate models) is much more difficult to analyze and quantify. The financial environment is an extremely complex system, and no single model can be used to price all products in a bank's portfolio. Therefore, a specific model is usually chosen for each class of products, and it is extremely important to choose the model which takes into account the risk factors, relevant for a given product class. For example: stochastic volatility may not be really necessary for pricing short-dated European options, but it is essential for long-dated forward start options or cliquets. The selected model will then be fitted to a set of *calibration instruments*, and here it is essential on one hand that the model is rich enough to match the prices of all instruments in the calibration set (for instance it is impossible to calibrate the constant volatility Black-Scholes model to the entire smile), and on the other hand that the model parameters are identifiable in a unique and stable way from the prices of calibration instruments (it is impossible to calibrate a complex stochastic volatility model using a single option price).

Finally, the calibrated model is used for pricing the non-quoted exotic products. The final price is always explained to some extent by the model used and to some extent by the calibration instruments. Ideally, it should be completely determined by the calibration instruments: we want every model, fitted to the prices of calibration instruments, to yield more or less the same price of the exotic. If this is not the case, then one is speaking of model risk. To have an idea of this risk, one can therefore price the exotic option with a set of different models calibrated to the same instruments. See [38] for an example of possible bounds obtained that way and [10] for more details on model risk.

The notion of model risk is closely linked to model validation: one of the roles of a risk manager in a bank which consists in scrutinizing and determining validity limits for models used by the front office. The above arguments show that, for sensible results, models must be validated in conjunction with the set of calibration instruments that will be used to fit the model, and with the class of products which the model will be used to price.

Other risk types which we do not discuss here due to the lack of space include legal and regulatory risk, business risk, strategic risk and reputation risk [13].

14.2 Risk exposures and risk limits: sensitivity approach to risk management

A traditional way to control the risk of a trading desk is to identify the risk factors relevant for this desk and impose limits on the sensitivities to these risk factors. The sensitivity approach is an important element of the risk manager's toolbox, and it is therefore important to understand the strengths and weaknesses of this methodology.

The sensitivity is defined as the variation of the price of a financial product which corresponds to a small change in the value of a given risk factor, when

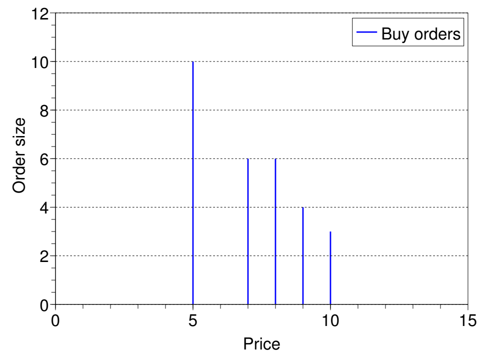

Figure 14.1: Illustration of asset liquidity risk: a sell order for a single share would be executed at a price of 10 euros, while a market sell order for 20 shares would be executed at a weighted average price of 8.05 euros per share.

all other risk factors are kept constant. In mathematical terms, sensitivity is very close to a partial derivative. For example, the standard measure of the interest rate risk of a bond is called DV01, that is, the dollar value of one basis point, or, in other words, the change of the value of a bond corresponding to a 1-bp decrease of its yield to maturity. For an option, whose price will be denoted by C, the basic sensitivities are the delta $\frac{\partial C}{\partial S}$, the vega $\frac{\partial C}{\partial \sigma}$, where σ is the volatility, the theta $\frac{\partial C}{\partial t}$ and the rho $\frac{\partial C}{\partial r}$, where r is the interest rate of the zero-coupon bond with the same maturity as the option. The gamma $\frac{\partial^2 C}{\partial S^2}$ is the price sensitivity of delta, and can be also used to quantify the sensitivity of the option price itself to larger price moves (as the second term in the Taylor expansion).

An option trading desk typically has a limit on its delta and vega exposure (for each underlying), and other sensitivities are also monitored. The vega is typically interpreted as the exposure to Black-Scholes implied volatility, and since the implied volatility depends on the strike and maturity of the option, the sensitivity to the *implied volatility surface* is a natural extension. This is typically taken into account by *bucketing* the smile, that is computing sensitivities to perturbations of specific sections of the smile (for a certain range of strikes and maturities). The same approach is usually applied to compute the sensitivities to the yield curve.

When working with sensitivities it is important to understand the following features and limitations of this approach:

- Sensitivities do not provide information about the actual risk faced by a bank, but only relate changes in the values of derivative products to the changes of basic risk factors.

- Sensitivities cannot be aggregated across different underlyings and across different types of sensitivities (delta plus gamma): they are therefore local measures and do not provide global information about the entire portfolio of a bank.

- Sensitivities are meaningful only for small changes of risk factors: they do not provide accurate information about the reaction of the portfolio to larger moves (jumps).

- The notion of a sensitivity to a given risk factor is associated to a very specific scenario of market evolution, when only this risk factor changes while others are kept constant. The relevance of a particular sensitivity depends on whether this specific scenario is plausible. As an example, consider the delta of an option, which is often defined as the partial derivative of the Black-Scholes price of this option computed using its actual implied volatility: $\Delta_t^{\text{imp}}(T,K) = \frac{\partial \text{BS}}{\partial s}(S_t, \sigma^{\text{imp}}, K, T)$ (see chapter 8). The implicit assumption is that when the underlying changes, the implied volatility of an option with a given strike remains constant. This is the so-called sticky strike behavior, and it is usually not observed in the markets, which tend to have the sticky moneyness behavior, where the implied volatility of an option with a given moneyness level $m = K/S_t$ is close to constant: $\sigma^{\text{imp}} = \sigma^{\text{imp}}(K/S_t)$. Therefore, the true sensitivity of the option price to changes in the underlying is

$$\begin{aligned}\Delta_t &= \frac{d}{ds}\text{BS}(S_t, \sigma^{\text{imp}}(K/S_t), K, T) \\ &= \Delta_t^{\text{imp}}(T,K) - \frac{K}{S_t^2}\frac{\partial \text{BS}(S_t, \sigma^{\text{imp}}, K, T)}{\partial \sigma}\frac{d\sigma^{\text{imp}}}{dm}.\end{aligned}$$

  In equity markets, the implied volatility is usually decreasing as function of the moneyness value (skew effect), and the Black-Scholes delta therefore undervalues the sensitivity of option price to the movements of the underlying.

  When computing the sensitivities to the implied volatility smile or the yield curve, the "bucketing" approach assumes that one small section of the smile/curve moves while the other ones remain constant, which is clearly unrealistic. A more satisfactory approach is to identify possible orthogonal deformation patterns, such as level changes, twists, convexity changes, from historical data (via principal component analysis), and compute sensitivities to such realistic deformations.

14.3 Value at Risk and the global approach

The need to define a global measure of the amount of capital that the bank is actually risking, has led to the appearance of the Value at Risk (VaR). The VaR belongs to the class of *Monetary risk measures* which quantify the risk as a dollar amount and can therefore be interpreted as regulatory capital required to

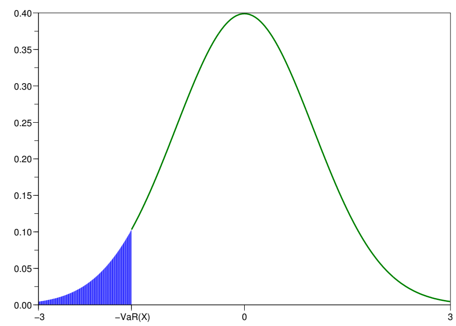

Figure 14.2: Definition of the Value at Risk

cover the risk. The VaR of a position is associated to a time horizon T (usually 1 or 10 days) and confidence level α (usually 99% or 95%), and is defined as the opposite of the $(1-\alpha)$-quantile of the profit and loss of this position in T days (see Fig. 14.2):

$$\mathrm{VaR}_\alpha(X) := -\inf\{x \in \mathbb{R} : P[X \leq x] \geq 1-\alpha\}.$$

The advantages of the VaR include its simplicity (the risk is summarized in a single number, whose meaning is easy to explain to people without technical knowledge) and the fact that it is a portfolio risk measure, that is, it summarizes all risk factors affecting a large portfolio, taking into account correlations and dependencies.

Computing Value at Risk The first step in implementing a VaR computation engine is to identify a reasonable number of risk factors, which span sufficiently well the universe of risks faced by the bank. This is mainly done to reduce the dimension of the problem compared to the total number of products in the portfolio. For example, for a bond portfolio on the same yield curve, three risk factors (short-term, medium-term and long-term yield) may be sufficient. The next step is to identify the dependency of each product in the portfolio on the different risk factors (via a pricing model). Finally, scenarios or probability laws for the evolution of risk factors are identified and a specific method is applied to compute the VaR.

In local valuation methods, the portfolio value and its sensitivities are computed at the current values of risk factors only. In full valuation methods, all products in the portfolio must be repriced with perturbed values of risk factors. A particularly simple method is the Gaussian or normal VaR, where the risk factors are assumed Gaussian, which makes it possible to compute the VaR without simulation. In historical VaR method, the historically observed changes of risk factors are used, and in Monte Carlo VaR method the variations of risk factors are sampled from a fully calibrated model and the VaR value is estimated by Monte Carlo.

We now discuss the three most popular methods for computing VaR using the example of a portfolio of N options written on d different underlyings with values $S_t^1, \dots, S_t^d$. We denote by P_t^i the price of i-th option at time t: $P_t^i = P_i(t, S_t^1, \dots, S_t^d)$. Let w_i be the quantity of i-th option in the portfolio, so that the portfolio value is $V_t = \sum_i w_i P_t^i$.

- The *delta-normal* approach only takes into accout the first-order sensitivities (deltas) of option prices, and assumes that the daily increments of risk factors follow a normal distribution: $\Delta S_t \sim N(0, \Omega_t)$. The means are usually ignored in this approach, and the covariance matrix is estimated from the historical time series using a moving window. The variation of the portfolio is approximated by

  $$\Delta V \approx \sum_i \sum_j w_i \frac{\partial P^i}{\partial S^j} \Delta S^j.$$

  Therefore, the variance of ΔV is given by

  $$\mathbb{V}ar\,[\Delta V] \approx \sum_{ijkl} w_i \frac{\partial P^i}{\partial S^j} w_k \frac{\partial P^k}{\partial S^l} \Omega_{jl},$$

  and the daily Value at Risk for a given confidence level α may be estimated via

  $$\mathrm{VaR}_\alpha = \mathbf{N}(\alpha)\sqrt{\mathbb{V}ar\,[\Delta V]},$$

  where $\mathbf{N}$ is the standard normal distribution function.

  This method is extremely fast, but not so popular in practice due to its imporant drawbacks: it may be very inaccurate for non-linear derivatives and does not allow for fat tails in the distributions of risk factors.

- The *historical approach* is a full valuation method which relies on historical data for obtaining risk factor scenarios. Most often, one uses one year of daily increments of risk factor values: $(\Delta S_i^j)_{i=1\dots 250}^{j=1\dots d}$. These increments are used to obtain 250 possible values for the portfolio price on the next day:

  $$\Delta V_i = \sum_{k=1}^N w_k \{P_k(t+1, S_t^1 + \Delta S_i^1, \dots, S_t^d + \Delta S_i^d) - P_k(t, S_t^1, \dots, S_t^d)\}$$

  The Value at Risk is then estimated as the corresponding empirical quantile of ΔV.

  This method preserves the dependency structure and the distributional properties of the data, and applies to some extent to nonlinear products, however it also has a number of drawbacks. All intertemporal dependencies which may be present in the data, such as stochastic volatility, are destroyed. Some dependency on the current volatility may be preserved by taking a relatively short time window, but in this case important scenarios which have occured in not-so-recent past may be lost. Also, important no-arbitrage relations between risk factors (such as, between stock and option prices) may be violated.

- The *Monte Carlo VaR*, used by most major banks, is the most flexible approach, but it is also the most time consuming. It is a full valuation method where the increments ΔS^i are simulated using a full-fledged statistical model, which may include fat tails, intertemporal dependencies such as stochastic volatility, correlations or copula-based cross-sectional dependencies, and so on. The Monte Carlo computation of a global VaR estimate for the entire bank's portfolio is probably the most time consuming single computation that a bank needs to perform, and may take an entire night of computing on a cluster of processors.

Independently of the chosen method of computation, the VaR engine must be systematically back-tested, that is, the number of daily losses exceeding the previous day's VaR in absolute value must be carefully monitored. For a 95% VaR, these losses should be observed roughly on five days out of 100 (with both a larger and a smaller number being an indication of a poor VaR computation), and they must be uniformly distributed over time, rather than appear in clusters.

Shortcomings of the VaR approach The Value at Risk takes into account the probability of loss but not the actual loss amount above the qualtile level: as long as the probability of loss is smaller than α, the VaR does not distinguish between losing \$1000 and \$1 billion. More generally, VaR is only suitable for everyday activities and loss sizes, it does not accout for extreme losses which happen with small probability. A sound risk management system cannot therefore be based exclusively on the VaR and must include extensive stress testing and scenario analysis.

Because the VaR lacks the crucial subadditivity property (see next section), in some situations, it may penalize diversification: $\mathrm{VaR}\,(A+B) > \mathrm{VaR}\,(A) + \mathrm{VaR}\,(B)$. For example, let A and B be two independent portfolios with distribution

$$
\begin{aligned}
&P[A = -1000\$] = P[B = -1000\$] = 0.04\\
&P[A = 0] = P[B = 0] = 0.96.
\end{aligned}
$$

Then $\mathrm{VaR}_{0.95}(A) = \mathrm{VaR}_{0.95}(B) = 0$ but $P[A + B \leq -1000\$] = 0.0784$ and $\mathrm{VaR}_{0.95}(A+B) = 1000\$ > 0$. Such situations are not uncommon in the domain of credit derivatives where the distributions are strongly non-Gaussian.

More dangerously, this lack of convexity makes it possible for unscrupulous traders to introduce bias into risk estimates, by putting all risk in the tail of the distribution which the VaR does not see, as illustrated in the following example.

Portfolio composition	n stocks	n stocks + n put options
Initial value	nS_0	$n(S_0 + P_0)$
Terminal value	nS_T	$n(S_T + P_0 - (K - S_T)^+)$
Portfolio P&L	$n(S_T - S_0)$	$n(S_T - S_0 - (K - S_T)^+)$
Portfolio VaR	$-nq_{1-\alpha}$	$-nq_{1-\alpha}$.

Table 14.2: The sale of an out of the money put option with exercise probability less than α allows to generate immediate profits but has no effect on VaR_α although the risk of the position is increased (see example 14.1 for details).

Example 14.1. *Consider a portfolio containing n units of stock (S_t), and let q_α denote the $(1-\alpha)$-quantile of the stock return distribution (assumed continuous) for the time horizon T:*

$$q_\alpha := \inf\{x \in \mathbb{R} : P[S_T - S_0 \leq x] \geq \alpha\}.$$

The value at risk of this portfolio for the time horizon T and confidence level α is given by $-nq_{1-\alpha}$ (see table 14.2). Assume now that the trader sells n put options on S_T with strike K satisfying $K \leq S_0 - q_{1-\alpha}$, maturity T and initial price P_0. The initial value of the portfolio becomes equal to $n(S_0 + P_0)$ and the terminal value is $n(S_T + P_0 - (K - S_T)^+)$. Since

$$\begin{aligned} &P[n(S_T - S_0 - (K - S_T)^+) \leq nq_{1-\alpha}] \\ &\qquad = P[n(S_T - S_0) \leq nq_{1-\alpha};\, S \geq K] + P[S_T < K] = 1 - \alpha, \end{aligned}$$

the VaR of the portfolio is unchanged by this transaction, although the risk is increased.

14.4 Convex and coherent risk measures

In the seminal paper [1], Artzner et al. defined a set of properties that a risk measure must possess if it is to be used for computing regulatory capital in a sensible risk management system. Let $\mathcal{X}$ be the linear space of possible pay-offs, containing the constants

Definition 14.2. *A mapping $\rho : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is called a* coherent risk measure *if it possesses the following properties:*

- *Monotonicity: $X \leq Y$ implies $\rho(X) \geq \rho(Y)$.*
- *Cash invariance: for all $m \in \mathbb{R}$, $\rho(X + m) = \rho(X) - m$.*
- *Subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$.*
- *Positive homogeneity: $\rho(\lambda X) = \lambda\rho(X)$ for $\lambda \geq 0$.*

While the first three conditions seem quite natural (in particular, the subadditivity is linked to the ability of the risk measure to encourage diversification), the positive homogeneity property has been questioned by many authors. In particular, a risk measure with this property does not take into account the liquidity risk associated with liquidation costs of large portfolios.

Under the positive homogeneity property, the subadditivity is equivalent to convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$, for $0 \leq \lambda \leq 1$. The Value at Risk possesses the monotonicity, the cash invariance and the positive homogeneity properties, but we have seen that it is not subadditive, and therefore it is not a coherent risk measure.

The smallest coherent risk measure which dominates the VaR is known as the Conditional VaR or expected shortfall, and is defined as the average VaR with confidence levels between α and 1:

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\alpha(X)d\alpha \tag{14.1}$$

The following proposition clarifies the interpretation of ES_α as "the expectation of losses in excess of VaR".

Proposition 14.3. *The Expected Shortfall admits the probabilistic representation*

$$\mathrm{ES}_\alpha(X) = \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}E[(-\mathrm{VaR}_\alpha(X) - X)^+]$$

If the distribution function of X, denoted by $F(x)$, is continuous then in addition

$$\mathrm{ES}_\alpha(X) = -E[X|X < -\mathrm{VaR}_\alpha(X)].$$

Proof. Let $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$ be the left-continuous generalized inverse of F, such that $\mathrm{VaR}_\alpha(X) = -F^{-1}(1-\alpha)$. It is well known that if U is a random variable, uniformly distributed on $[0,1]$ then $F^{-1}(U)$ has the same law as X. Then

$$\begin{aligned}
&\mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}E[(-\mathrm{VaR}_\alpha(X) - X)^+]\\
&= \frac{1}{1-\alpha}E[(F^{-1}(1-\alpha) - F^{-1}(U))^+] - F^{-1}(1-\alpha)\\
&= \frac{1}{1-\alpha}\int_0^{1-\alpha}(F^{-1}(1-\alpha) - F^{-1}(u))du - F^{-1}(1-\alpha) = \frac{1}{1-\alpha}\int_0^{1-\alpha}F^{-1}(u)du,
\end{aligned}$$

which finishes the proof of the first part. The second observation follows from the fact that if F is contunuous, $P[X \leq -\mathrm{VaR}_\alpha(X)] = 1-\alpha$. ♢

The monotonicity, cash invariance and positive homogeneity properties of the expected shortfall are clear from the definition (14.1). The following proposition establishes a dual representation, which shows that unlike the VaR, the expected shortfall is convex (as the upper bound of a family of linear functions). It is therefore, the simplest example of a coherent risk measure.

Proposition 14.4. *The expected shortfall admits the representation*

$$\mathrm{ES}_\alpha(X) = \sup\{E^Q[-X] : Q \in \mathcal{Q}_{1-\alpha}\},$$

where $\mathcal{Q}_{1-\alpha}$ is the set of probability measures on $\mathcal{X}$ satisfying $\frac{dQ}{dP} \leq \frac{1}{1-\alpha}$.

Proof. Let $Q \in \mathcal{Q}_{1-\alpha}$ with $Z = \frac{dQ}{dP}$, and let $x = -\mathrm{VaR}_\alpha(X)$. Then, using the representation of Proposition 14.3,

$$\begin{aligned}\mathrm{ES}_\alpha(X) + E[ZX] &= \frac{1}{1-\alpha}E[(x-X)1_{x\geq X} + (1-\alpha)Z(X-x)] \\ &= \frac{1}{1-\alpha}E[(x-X)1_{x\geq X}(1-(1-\alpha)Z)] \\ &\quad + \frac{1}{1-\alpha}E[(X-x)1_{x<X}(1-\alpha)Z] \geq 0, \qquad (14.2)\end{aligned}$$

which shows that $\mathrm{ES}_\alpha(X) \geq \sup\{E^Q[-X] : Q \in \mathcal{Q}_{1-\alpha}\}$. On the other hand, there exists $c \in [0,1]$ such that $P[X < x] + cP[X = x] = 1 - \alpha$. Taking $Z = \frac{1_{X<x}}{1-\alpha} + \frac{c1_{X=x}}{1-\alpha}$, we get equality in (14.2). ♢

The expected shortfall presents all the advantages of the Value at Risk and avoids many of its drawbacks: it encourages diversification, presents computational advantages compared to VaR and does not allow for regulatory arbitrage. Yet to this day, although certain major banks use the expected shortfall for internal monitoring purposes, the VaR remains by far the most widely used measure for the computation of regulatory capital, and the Expected Shortfall is not even mentioned in the Basel capital accords on banking supervision. In addition to a certain conservatism of the banking environment, a possible reason for that is that the Expected Shortfall is a more conservative risk measure, and would therefore imply higher costs for the bank in terms of regulatory capital.

14.5 Regulatory capital and the Basel framework

Starting from the early 80s, the banking regulators of the developed nations became increasingly aware of the necessity to rethink and standardize the regulatory practices, in order to avoid the dangers of the growing exposure to derivative products and loans to emerging markets on one hand, and ensure fair competition between internationally operating banks on the other hand. The work of the Basel committee for banking supervision, created for the purpose of developping a set of recommendations for regulators, resulted in the publication of the 1988 Basel Accord (Basel I) [3]. The Accord concerned exclusively credit risk, and contained simple rules for computing minimal capital requirements depending on the credit exposures of a bank. More precisely, the Accord defines two minimal capital requirements, to be met by a bank at all times: the assets to capital multiple, and the risk based capital ratio (the Cooke ratio). The assets to capital multiple is the ratio of the total notional amount of the bank's assets to the bank's capital (meaning equity capital, that is, difference between assets and debt, plus some additions). The maximum allowed multiple is 20:

$$\frac{\text{Total assets}}{\text{Capital}} \leq 20.$$

The risk-based capital ratio is the ratio of the capital to the sum of all assets, weighted by their respective risk factors. This ratio must not be less that 8 per cent:

$$\frac{\text{Capital}}{\text{Risk-weighted assets}} \geq 8\%.$$

The risk weights reflect relative riskiness of very broad asset classes: for example, cash, gold and OECD government bonds are considered risk-free and have risk weight zero; claims on OECD banks and public agencies have risk weight 0.20; uninsured residential mortgages have weight 0.50 and all other claims such as corporate bonds have risk weight 1.00. The credit ratings of bond issuers are not explicitly taken into account under the Basel I accord.

The rapid growth of banks' trading activity, especially in the derivative products, has prompted the Basel Committee to develop a set of recommendations for the computation of regulatory capital needed for protection agains market risk, known as the 1996 market risk amendment [4]. This amendment allows the banks to choose between a standard model proposed by the regulator (the standardized approach) and internally developed VaR model (internal models approach). To be eligible for the internal models approach, the banks must have a strong risk management team, reporting only to the senior management, implement a robust back-testing scheme and meet a number of other requirements. This creates an incentive for banks to develop strong risk management, since using internal models allows to reduce the regulatory capital by 20–50 per cent, thanks to the correlations and diversification effects. The capital requirements are computed using the 10-day VaR at the 99% confidence level, multiplied by an adjustment factor imposed by the regulator and reflecting provisions for model risk, quality of ex-post performance etc.

Already in the late 90s, it was clear that the Basel I accord needed replacement, because of such notorious problems as rating-independent risk weights and possibility of regulatory arbitrage via securitization. In 2004, the Basel Committe published a new capital adequacy framework known as Basel II [5]. This framework describes capital provisions for credit, market and newly introduced operational risk. In addition to describing the computation of minimum regulatory requirements (Pillar I of the framework), it also describes different aspects of interaction between banks and their regulators (Pillar II) and the requiremens for disclosure of risk information to encourage market discipline (Pillar III). The basic Cooke Ratio formula remains the same, but the method for computing the risk-weighted assets is considerably modified. For different risk types, the banks have the choice between several approaches, depending on the strengh of their risk management teams. For credit risk, these approaches are the Standardized approach, the Foundation internal ratings based approach

(IRBA) and the Advanced IRBA. Under the standardized approach, the banks use supervisory formulas and the ratings provided by an external rating agency. Under the Foundation IRBA the banks are allowed to estimate their own default probability, and under Advanced IRBA other parameters such as loss given default (LGD) and exposure at default (EAD) are also estimated internally. These inputs are then plugged into the supervisor-provided general formula to compute the capital requirements.

Appendix A

Préliminaires de la théorie des mesures

A.1 Espaces mesurables et mesures

Dans toute cette section, Ω désigne un ensemble quelconque, et $\mathcal{P}(\Omega)$ est l'ensemble des toutes ses parties.

A.1.1 Algèbres, $\sigma-$algèbres

Definition A.1. *Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$. On dit que*
(i) *$\mathcal{A}_0$ est une algèbre sur Ω si $\mathcal{A}_0$ contient Ω et est stable par passage au complémentaire et par réunion.*
(ii) *$\mathcal{A}$ est une $\sigma-$algèbre si c'est une algèbre stable par union dénombrable. On dit alors que $(\Omega, \mathcal{A})$ est un espace mesurable.*

Notons qu'une algèbre doit aussi contenir $\emptyset$, et est stable par intersection et par différence symétrique, i.e.

$$A \cap B \text{ et } A \Delta B := (A \cup B) \setminus (A \cap B) \in \mathcal{A} \quad \text{pour tous} \quad A, B \in \mathcal{A}_0,$$

et qu'une $\sigma-$algèbre est stable par intersection dénombrable. $\mathcal{P}(\Omega)$ est la plus grande $\sigma-$algèbre sur Ω. Il s'avère cependant que cette $\sigma-$algèbre est souvent trop grande pour qu'on puisse y développer les outils mathématiques nécessaires.

En dehors des cas très simples, il est souvent impossible de lister les éléments d'une algèbre ou d'une $\sigma-$algèbre. Il est alors commode de les caractériser par des sous-ensemble "assez riches".

Ainsi, on définit pour tout $\mathcal{C} \subset \mathcal{P}(\Omega)$ la $\sigma-$algèbre $\sigma(\mathcal{C})$ engendrée par $\mathcal{C}$. C'est la plus petite $\sigma-$algèbre sur Ω contenant $\mathcal{C}$, définie comme intersection de toutes les $\sigma-$algèbre sur Ω contenant $\mathcal{C}$.

Example A.2. Si Ω est un espace topologique, la *$\sigma-$algèbre Borelienne*, notée par $\mathcal{B}_\Omega$, est la $\sigma-$algèbre engendrée par les ouverts de Ω. Pour la droite réelle, on peut même simplifier la compréhension de $\mathcal{B}_{\mathbb{R}}$:

$$\mathcal{B}_{\mathbb{R}} = \sigma\left(\pi(\mathbb{R})\right) \quad \text{où} \quad \pi(\mathbb{R}) := \{]-\infty, x] : x \in \mathbb{R}\}$$

(Exercice !)

L'exemple précédent se généralise par la notion suivante:

Definition A.3. *Soit $\mathcal{I} \subset \mathcal{P}(\Omega)$. On dit que $\mathcal{I}$ est un $\underline{\pi-\text{système}}$ s'il est stable par intersection finie.*

Ainsi l'ensemble $\pi(\mathbb{R})$ de l'exemple ci-dessus est un $\pi-$système. L'importance de cette notion apparaîtra dans la proposition A.5 ci-dessous ainsi que dans le théorème des classes monotones A.18 de la section A.2.

A.1.2 Mesures

Definition A.4. *Soit $\mathcal{A}_0$ une algèbre sur Ω, et $\mu_0 : \mathcal{A}_0 \longrightarrow \mathbb{R}_+$ une fonction positive.*
(i) *μ_0 est dite $\underline{\text{additive}}$ si $\mu_0(\emptyset) = 0$ et pour tous $A, B \in \mathcal{A}_0$:*

$$\mu_0(A \cup B) = \mu_0(A) + \mu_0(B) \quad \textit{dès que} \quad A \cap B = \emptyset.$$

(ii) *μ_0 est dite $\underline{\sigma-\text{additive}}$ si $\mu_0(\emptyset) = 0$ et pour toute suite $(A_n)_{n\geq 0} \subset \mathcal{A}_0$:*

$$A = \cup_{n\geq 0} A_n \in \mathcal{A}_0 \textit{ et les } A_n \textit{ disjoints} \quad \Longrightarrow \quad \mu_0(A) = \sum_{n\geq 0} \mu_0(A_n).$$

(iii) *Une fonction $\sigma-$additive $\mu : \mathcal{A} \longrightarrow \mathbb{R}_+$ sur un espace mesurable $(\Omega, \mathcal{A})$ est appelée $\underline{\text{mesure}}$, et on dit que $(\Omega, \mathcal{A}, \mu)$ est un $\underline{\text{espace mesuré}}$.*
(iv) *Un espace mesuré $(\Omega, \mathcal{A}, \mu)$ est dit $\underline{\text{fini}}$ si $\mu(\Omega) < \infty$, et $\underline{\sigma-\text{fini}}$ s'il existe une suite $(\Omega_n)_{n\geq 0} \subset \mathcal{A}$ telle que $\mu(\Omega_n) < \infty$ et $\cup_{n\geq 0}\Omega_n = \Omega$.*

Proposition A.5. *Soient $\mathcal{I}$ un $\pi-$système, et μ, ν deux mesures finies sur l'espace mesurable $(\Omega, \sigma(\mathcal{I}))$. Si $\mu = \nu$ sur $\mathcal{I}$ alors $\mu = \nu$ sur $\sigma(\mathcal{I})$.*

La démonstration est reportée, à titre de complément, dans l'annexe de ce chapitre. Le résultat suivant est essentiel pour construire des mesures "intéressantes".

Theorem A.6. *(extension de Carathéodory) Soient $\mathcal{A}_0$ une algèbre sur Ω, et $\mu_0 : \mathcal{A}_0 \longrightarrow \mathbb{R}_+$ une fonction $\sigma-$additive. Alors il existe une mesure μ sur $\mathcal{A} := \sigma(\mathcal{A}_0)$ telle que $\mu = \mu_0$ sur $\mathcal{A}_0$. Si de plus $\mu_0(\Omega) < \infty$, alors une telle extension μ est unique.*

La démonstration est reportée, à titre de complément, dans l'annexe de ce chapitre. Avec ce résultat, on peut maintenant construire une mesure importante sur l'espace mesurable $(]0,1], \mathcal{B}_{]0,1]})$.

Example A.7. *(Mesure de Lebesgue)* Nous allons définir une mesure sur $\mathcal{B}_{]0,1]}$ qui mesure les longueurs.
1- On remarque tout d'abord que $\mathcal{A}_0$ constitué des parties $A \subset]0,1]$ de la forme

$$A = \cup_{1 \le i \le n}(a_i, b_i] \quad \text{pour} \quad n \in \mathbb{N} \text{ et } 0 \le a_1 \le b_1 \le \ldots \le a_n \le b_n \le 1 \tag{A.1}$$

est une σalgèbre telle que $\mathcal{B}_{]0,1]} = \sigma(\mathcal{A}_0)$. Pour tout $A \in \mathcal{A}_0$ de la forme (A.1), on définit

$$\lambda_0(A) \quad := \quad \sum_{i=1}^{n}(b_i - a_i).$$

2- Alors $\lambda_0 : \mathcal{A}_0 \longrightarrow \mathbb{R}_+$ est une application bien définie et est évidemment additive. On peut montrer qu'elle est $\sigma-$additive (c'est moins évident, voir cours de première année). Comme $\lambda_0(]0,1]) < \infty$, on déduit du théorème de Carathédory l'existence d'une unique extension λ définie sur $\mathcal{B}_{]0,1]}$.
Cette mesure fini λ est appelée <u>*mesure de Lebesgue*</u> sur $]0,1]$. La mesure de Lebesgue sur $[0,1]$ est obtenue par une modification triviale puisque le singleton $\{0\}$ est de mesure de Lebesgue nulle.
3- Par le même raisonnement, on peut construite la mesure de Lebesgue sur $\mathcal{B}_{\mathbb{R}}$ comme extension d'une application d'ensembles sur l'algèbre des unions finies d'intervalles semi-ouverts disjoints. Dans ce cas, la mesure de Lebesgue est seulement $\sigma-$finie.

Definition A.8. (i) *Sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$, un ensemble $N \in \mathcal{A}$ est dit <u>négligeable</u> si $\mu(N) = 0$.*
(ii) *Soit $P(\omega)$ une propriété qui ne dépend que d'un élément $\omega \in \Omega$. On dit que P est vraie <u>$\mu-$presque partout</u>, et on note $\mu-$p.p., si l'ensemble $\{\omega \in \Omega : P(\omega)$ est fausse$\}$ est inclus dans un ensemble négligeable.*

Remark A.9. D'après la propriété de $\sigma-$additivité de la mesure, on voit aisément que *toute union dénombrable de négligeables est négligeable.*

A.1.3 Propriétés élémentaires des mesures

Nous commençons par des propriétés mettant en jeu un nombre fini d'ensembles.

Proposition A.10. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_i)_{i \le n} \subset \mathcal{A}$. Alors:*
(i) $\mu(\cup_{i \le n} A_i) \le \sum_{i \le n} \mu(A_i)$,
(ii) *Si de plus $\mu(\Omega) < \infty$, on a*

$$\mu(\cup_{i \le n} A_i) \quad = \quad \sum_{k \le n}(-1)^{k-1} \sum_{i_1 < \ldots < i_k \le n} \mu(A_{i_1} \cap \ldots A_{i_k}).$$

La preuve de ce résultat est une conséquence immédiate de la définition de mesure. La partie (ii), spécifique aux mesures finies, donne une formule pour la mesure de l'union finie d'ensemble qui alterne entre sur-estimation et sous estimation. Pour $n = 2$ cette formule n'est autre que la propriété bien connue $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ pour $A, B \in \mathcal{A}$.

Le résultat (simple) suivant est fondamental en théorie de la mesure. Pour une suite d'ensembles $(A_n)_n$, nous notons simplement $A_n \uparrow A$ pour indiquer que la suite est croissante ($A_n \subset A_{n+1}$) et $\cup_n A_n = A$. La notation $A_n \downarrow A$ a un sens similaire dans le cas où la suite est décroissante.

Proposition A.11. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)_n$ une suite de $\mathcal{A}$. Alors*
(i) $A_n \uparrow A \Longrightarrow \mu(A_n) \uparrow \mu(A)$,
(ii) $A_n \downarrow A$ *et* $\mu(A_k) < \infty$ *pour un certain entier* $k \Longrightarrow \mu(A_n) \downarrow \mu(A)$,

La démonstration simple de ce résultat est laissée comme exercice. Faisons juste deux remarques:

- Une conséquence de la proposition A.11 est que l'union dénombrable d'ensembles de mesure nulle est de mesure nulle.
- l'exemple $A_n =]n, \infty[$ dans l'espace mesuré $(\mathbb{R}, \mathcal{B}_\mathbb{R}, \lambda)$, λ étant la mesure de Lebesgue sur $\mathbb{R}$, montre que la condition supplémentaire dans (ii) est nécessaire.

Ces résultats permettent de montrer les outils important pour l'analyse de la convergence des mesures des ensembles. On rappelle les notions de liminf et limsup pour une suite d'ensembles $(A_n)_n$:

$$\limsup E_n := \cap_n \cup_{k \geq n} E_k = \{\omega \in \Omega : \ \omega \in E_n \text{ pour une infinité de } n\},$$
$$\liminf E_n := \cup_n \cap_{k \geq n} E_k = \{\omega \in \Omega : \ \omega \in E_n \text{ à partir d'un rang } n_0(\omega)\}.$$

Le résultat suivant est très utile.

Lemma A.12. *(de Fatou pour les ensembles) Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)_n$ une suite dans $\mathcal{A}$. Alors*

$$\mu[\liminf A_n] \quad \leq \quad \liminf \mu[A_n].$$

Proof. Par définition, nous avons $B_n := \cap_{k \geq n} A_k \uparrow B := \liminf A_n$, et on déduit de la proposition A.11 (i) que $\mu[B] = \lim \uparrow \mu[B_n]$. Pour conclure, il suffit de remarquer que $B_n \subset A_n$ et par suite $\mu[B_n] \leq \mu[A_n]$, impliquant que $\lim \uparrow \mu[B_n] \leq \liminf \mu[A_n]$. $\diamondsuit$

Si la mesure est finie, le résultat suivant montre que l'inégalité inverse dans le lemme de Fatou pour les ensembles a lieu en échangeant $\liminf$ et $\limsup$. Nous verrons plus tard que la situation est plus compliquée pour les fonctions...

Lemma A.13. *(inverse Fatou pour les ensembles) Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré fini, et $(A_n)_n$ une suite dans $\mathcal{A}$. Alors*

$$\mu[\limsup A_n] \quad \geq \quad \limsup \mu[A_n].$$

Proof. Par définition, nous avons $C_n := \cup_{k \geq n} A_k \downarrow C := \limsup A_n$. La proposition A.11 (ii), qui requiert que la mesure soit finie, donne $\mu[C] = \lim \downarrow \mu[C_n]$. Pour conclure, il suffit de remarquer que $C_n \supset A_n$ et par suite $\mu[C_n] \geq \mu[A_n]$, impliquant que $\lim \downarrow \mu[C_n] \geq \limsup \mu[A_n]$. ♢

Enfin, nous énonçons le résultat suivant qui sera utilisé à plusieur reprises, et qui sera complété dans la suite quand nous aurons abordé les notions d'indépendance.

Lemma A.14. *(Premier lemme de Borel-Cantelli) Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)_n \subset \mathcal{A}$. Alors*

$$\sum_n \mu[A_n] < \infty \quad \Longrightarrow \quad \mu[\limsup A_n] = 0.$$

Proof. Avec les notations de la démonstration du lemme A.13, on a $\limsup A_n \subset C_n = \cup_{k \geq n} A_k$, et donc $\mu(\limsup A_n) \leq \mu(C_n) \leq \sum_{k \geq n} \mu(A_n)$. Le résultat est obtenu en envoyant n vers l'infini. ♢

A.2 L'intégrale de Lebesgue

Dans cette section, on considère un espace mesuré $(\Omega, \mathcal{A}, \mu)$, et nous développons la théorie d'intégration d'une fonction par rapport à la mesure μ. Si Ω est dénombrable, $\mathcal{A} = \mu(\Omega)$, et $\mu(\{\omega\}) = 1$ pour tout $\omega \in \Omega$, une fonction est identifiée à une suite $(a_n)_n$, et elle est intégrable si et seulement si $\sum_n |a_n| < \infty$, et l'intégrale est donnée par la valeur de la série $\sum_n a_n$. La réelle difficulté est donc pour les espaces non dénombrables.

A.2.1 Fonction mesurable

L'objet central en topologie est la structure des ouverts, et les fonctions continues sont caracrtérisées par la propriété que les images réciproques des ouvert de l'ensemble d'arrivée sont des ouverts de l'ensemble de départ. Dans la théorie de la mesure, les ouverts sont remplacés par les ensembles mesurables, et les fonctions mesurables remplacent les fonctions continues.

Definition A.15. *On dit qu'une fonction $f : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ est* <u>*mesurable*</u> *si l'image réciproque de tout ensemble borélien est dans $\mathcal{A}$. On note par $\mathcal{L}^0(\mathcal{A})$ l'ensemble des fonctions mesurables. Les sous-ensembles des fonctions mesurables positives (resp. bornées) seront notés $\mathcal{L}^0_+(\mathcal{A})$ (resp. $\mathcal{L}^\infty(\mathcal{A})$).*

De manière équivalente $f \in \mathcal{L}^0(\mathcal{A})$ si et seulement l'inverse f^{-1} est bien définie comme une application de $\mathcal{B}_{\mathbb{R}}$ dans $\mathcal{A}$, i.e. $f^{-1} : \mathcal{B}_{\mathbb{R}} \longrightarrow \mathcal{A}$. Si $\mathcal{C} \subset \mathcal{B}_{\mathbb{R}}$ est tel que $\sigma(\mathcal{C}) = \mathcal{B}_{\mathbb{R}}$, alors il suffit de vérifier $f^{-1} : \mathcal{C} \longrightarrow \mathcal{A}$.

Remark A.16. (i) En prenant $\mathcal{C} = \pi(\mathbb{R})$ le $\pi-$système des intervalles de la forme $]-\infty, c]$, $c \in \mathbb{R}$, on voit que

$$f \in \mathcal{L}^0(\mathcal{A}) \quad \text{ssi} \quad \{f \leq c\} \in \mathcal{A} \text{ pour tout } c \in \mathbb{R}.$$

(ii) Sopposons que Ω est un espace topologique, et que $f : \Omega \longrightarrow \mathbb{R}$ est continue. Alors f est $\mathcal{B}_\Omega-$mesurable. En effet, avec $\mathcal{C} = \{\text{ouverts de } \mathbb{R}\}$, la continuité s'écrit $f^{-1} : \mathcal{B}_\mathbb{R} \longrightarrow \mathcal{A}$. On dit que f est une fonction borelienne.
(iii) Soit X une application de Ω dans un ensemble dénombrable $X(\Omega) = \{x_n, n \in \mathbb{N}\}$. On munit $X(\Omega)$ de la plus grande $\sigma-$algèbre $\mathcal{P}(X(\Omega))$ et on remarque que $\mathcal{P}(X(\Omega)) = \sigma(\{\{\omega\} : \omega \in \Omega\})$. Ceci permet de conclure que X est mesurable si et seulement si $\{X = x_n\} \in \mathcal{A}$ pour tout $n \in \mathbb{N}$.

La mesurabilité est conservée par les opérations usuelles pour les fonctions.

Proposition A.17. (i) *Pour $f, g \in \mathcal{L}^0(\mathcal{A})$, $h \in \mathcal{L}^0(\mathcal{B}_\mathbb{R})$, et $\lambda \in \mathbb{R}$, on a $f+g$, λf, fg, $f \circ h$ et $\lambda f \in \mathcal{L}^0(\mathcal{A})$.*
(ii) *Pour une suite $(f_n)_n \subset \mathcal{L}^0(\mathcal{A})$, on a $\inf h_n$, $\liminf h_n$, $\sup h_n$ et $\limsup h_n$ $\in \mathcal{L}^0(\mathcal{A})$.*

La preuve est simple et est laissée en exercice. Avant d'aborder l'objet central de ce chapitre, à savoir la construction de l'intégrale de Lebesgue, nous reportons une version simple du théorème des classes monotones, qui ne sera utilisé que plus tard dans la construction d'espaces mesurés produits.

Theorem A.18. *(classes monotones) Soit $\mathcal{H}$ une classes de fonctions réelles bornées sur Ω vérifiant les conditions suivantes:*
(H1) *$\mathcal{H}$ est un espace vectoriel contenant la fonction constante $\mathbf{1}$,*
(H2) *pour toute suite croissante $(f_n)_n \subset \mathcal{H}$ de fonctions positives telle que $f := \lim \uparrow f_n$ est bornée, on a $f \in \mathcal{H}$.*
Soit $\mathcal{I}$ un $\pi-$système tel que $\{\mathbf{1}_A : A \in \mathcal{I}\} \subset \mathcal{H}$. Alors $\mathcal{L}^\infty(\sigma(\mathcal{I})) \subset \mathcal{H}$.

La démonstration est reportée à titre de complément dans l'annexe de ce chapitre.

A.2.2 Intégration des fonctions positives

Le but de ce paragraphe est de définir pour toute fonction mesurable positive f une notion d'intégrale par rapport à la mesure μ:

$$\int f d\mu \quad \text{que l'on note aussi} \quad \mu(f),$$

qui est un abus de notation comunément accepté ($\mu : \mathcal{A} \longrightarrow \mathbb{R}$!) du fait que notre définition doit vérifier

$$\int \mathbf{1}_A = \mu(A) \quad \text{pour tout} \quad A \in \mathcal{A}.$$

Plus généralement, soit $\mathcal{S}^+$ l'ensemble des fonctions de Ω dans $\mathbb{R}_+$ de la forme

$$g \;=\; \sum_{i=1}^n a_i \mathbf{1}_{A_i}, \tag{A.2}$$

pour un certain entier $n \geq 1$, des ensembles $A_i \in \mathcal{A}$, et des scalaires $a_i \in [0, \infty]$, $1 \leq i \leq n$. Ici, il est commode d'autoriser la valeur $+\infty$, et on utilisera les règles de calcul $0 \times \infty = \infty \times 0 = 0$. l'intégrale sur $\mathcal{S}^+$ est définie par:

$$\mu(g) \quad = \quad \sum_{i=1}^{n} a_i \mu_0(A_i). \tag{A.3}$$

Il est clair que $\mu(g)$ est bien défini, i.e. deux représentations différentes (A.2) d'un élément $f \in \mathcal{S}^+$ donnent la même valeur. Nous étendons à présent la définition de μ à l'ensemble $\mathcal{L}^0_+(\mathcal{A})$ des fonctions $\mathcal{A}-$mesurables positives.

Definition A.19. *Pour $f \in \mathcal{L}^0_+(\mathcal{A})$, l'intégrale de f par rapport à μ est définie par*

$$\mu(f) \quad := \quad \sup\left\{\mu(g) : \; g \in \mathcal{S}^+ \text{ et } g \leq f\right\}.$$

L'ensemble $\{g \in \mathcal{S}^+ : g \leq f\}$, dont la borne supérieure définit l'intégrale, contient la fonction nulle. On peut aussi construire des éléments non triviaux en introduisant la fonction

$$\alpha_n(x) := n\mathbf{1}_{]n,\infty[}(x) + \sum_{i\geq 1}(i-1)2^{-n}\mathbf{1}_{B_i^n}(x), \quad B_i^n := [0,n]\cap](i-1)2^{-n}, i2^{-n}].$$

En effet, pour tout $f \in \mathcal{L}^0(\mathcal{A})$:

$$(\alpha_n \circ f)_n \subset \mathcal{S}^+ \quad \text{est une suite croissante qui converge vers} \quad f. \tag{A.4}$$

La définition de l'intégrale implique immédiatement que

$$\mu(cf) = c\mu(f) \quad \text{pour tous} \quad c \in \mathbb{R}_+ \text{ et } f \in \mathcal{L}^0_+(\mathcal{A}), \tag{A.5}$$

ainsi que la propriété de monotonie suivante.

Lemma A.20. *Pour $f_1, f_2 \in \mathcal{L}^0_+(\mathcal{A})$ avec $f_1 \leq f_2$, on a $0 \leq \mu(f_1) \leq \mu(f_2)$. De plus $\mu(f_1) = 0$ si et seulement si $f_1 = 0$, $\mu-$p.p.*

Proof. Pour la première partie, il suffit de remarquer que $\{g \in \mathcal{S}^+ : \; g \leq f_1\} \subset \{g \in \mathcal{S}^+ : \; g \leq f_2\}$. Pour la deuxième partie de l'énoncé, rappelons que $\mu(\{f > 0\}) = \lim \uparrow \mu(\{f > n^{-1}\})$ d'après la proposition A.11. Si $\mu(\{f > 0\}) > 0$, on a $\mu(\{f > n^{-1}\}) > 0$ pour n assez grand. Alors $f \geq g := n^{-1}\mathbf{1}\{f > n^{-1}\} \in \mathcal{S}^+$, et on déduit de la définition de l'intégrale que $\mu(f) \geq \mu(g) = n^{-1}\mu(\{f > n^{-1}\}) > 0$. ♢

Le résultat à la base de la théorie de l'intégration est l'extension suivante de la propriété de convergence monotone des mesures d'ensembles énoncée dans la proposition A.11 (i).

Theorem A.21. *(convergence monotone) Soit $(f_n)_n \subset \mathcal{L}^0_+(\mathcal{A})$ une suite croissante $\mu-$p.p., i.e. pour tout $n \geq 1$, $f_n \leq f_{n+1}$ $\mu-$p.p. Alors*

$$\mu\left(\lim \uparrow f_n\right) \quad = \quad \lim \uparrow \mu(f_n).$$

Proof. On procède en trois étapes.
<u>*Etape 1*</u> On commence par supposer que $f_n \leq f_{n+1}$ sur Ω. On note $f := \lim \uparrow f_n$. D'après le lemme A.20, la suite des intégrales $(\mu(f_n))_n$ hérite la croissance de la suite $(f_n)_n$ et est majorée par $\mu(f)$. Ceci montre l'inégalité $\lim \uparrow \mu(f_n) \leq \mu(\lim \uparrow f_n)$.

Pour établir l'inégalité inverse, nous devons montrer que $\lim \uparrow \mu(f_n) \geq \mu(g)$ pour tout $g = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i} \in \mathcal{S}^+$ vérifiant $g \leq f$. Pour tout $c \in [0,1[$, on déduit du lemme A.20 et de (A.5) que:

$$\mu(f_n) \geq \mu(f_n \mathbf{1}_{\{f_n \geq cg\}}) \geq c\mu(g\mathbf{1}_{\{f_n \geq cg\}}) = c\sum_{i=1}^{k} a_i \mu(A_i \cap \{f_n \geq ca_i\}).$$

En utilisant la propriété de convergence monotone des mesures d'ensembles énoncée dans la proposition A.11 (i), on obtient alors:

$$\lim \uparrow \mu(f_n) \quad \geq \quad c\sum_{i=1}^{l} a_i \mu(A_i) = c\mu(g) \longrightarrow \mu(g) \text{ quand } c \to 1.$$

<u>*Etape 2*</u> Dans le reste de la preuve, on veut passer de la monotonie de la suite $(f_n)_n$ sur Ω à la monotonie $\mu-$p.p. Pour celà, introduisons $\Omega_0 = \{\omega \in \Omega : (f_n(\omega))_n \text{ croissante}\}$, la suite croissante (sur Ω) $\tilde{f}_n := f_n \mathbf{1}_{\Omega_0}$, et les appapproximations croissantes (sur Ω) par des fonctions simples $\left(\alpha^k \circ f_n\right)_k$, $\left(\alpha^k \circ \tilde{f}_n\right)_k$ de $f_n, \tilde{f}_n$, comme dans (A.4). La définition de l'intégrale pour les fonctions simples donne trivialement $\mu(\alpha^k \circ f_n) = \mu(\alpha^k \circ \tilde{f}_n)$, et par suite $\mu(f_n) = \mu(\tilde{f}_n)$ d'après l'étape 1. Le résultat du théorème est enfin obtenu en appliquant le résultat de l'étape 1 à la suite $(\tilde{f}_n)_n$. ♢

Remark A.22. Par le même argument que l'étape 2 ci-dessus (approximation par les fonctions simples (A.4) et utilisation du théorème de convergence monotone), on montre facilement que:
(i) Pour $f_1, f_2 \in \mathcal{L}^0_+(\mathcal{A})$ telles que $f_1 = f_2$ $\mu-$p.p., on a $\mu(f_1) = \mu(f_2)$.
(ii) Pour $f_1, f_2 \in \mathcal{L}^0_+(\mathcal{A})$, on a $f_1 + f_2 \in \mathcal{L}^0_+(\mathcal{A})$ et $\mu(f_1 + f_2) = \mu(f_1) + \mu(f_2)$.

Voici une conséquence simple et très utile du théorème de convergence monotone.

Lemma A.23. *(Fatou) Pour une suite de fonctions $(f_n)_n$ de $\mathcal{L}^0_+(\mathcal{A})$, on a*

$$\mu(\liminf f_n) \leq \liminf \mu(f_n).$$

Proof. D'après la monotonie de l'intégrale, $\inf_{k \geq n} \mu(f_n) \geq \mu(\inf_{k \geq n} f_k)$ pour tout $n \geq 1$, et on obtient le résultat par application du théorème de convergence monotone. ♢

A.2.3 Intégration des fonctions réelles

Pour une fonction $f \in \mathcal{L}^0(\mathcal{A})$, on note $f^+ := \max\{f, 0\}$ et $f^- := \max\{-f, 0\}$ si bien que $|f| = f^+ - f^-$. Ces deux fonctions héritent la $\mathcal{A}-$mesurabilité de f.

Definition A.24. *Une fonction $f \in \mathcal{L}^0(\mathcal{A})$ est dite $\underline{\mu-\text{intégrable}}$ si $\mu(|f|) = \mu(f^+) + \mu(f^-) < \infty$, et son intégrale est définie par*

$$\mu(f) \quad := \quad \mu(f^+) - \mu(f^-).$$

On note par $\mathcal{L}^1(\mathcal{A}, \mu)$ l'ensemble des fonctions $\mu-$intégrables.

On voit immédiatement que $\mathcal{L}^1(\mathcal{A}, \mu)$ est un espace vectoriel dont on donnera d'autres propriétés topologiques dans la suite.

Avant de continuer, levons tout de suite une source d'ambiguité concernant l'intégration d'une fonction $f \in \mathcal{L}^1(\mathcal{A}, \mu)$ sur une partie $A \in \mathcal{A}$. En effet celle-ci peut se faire soit en intégrant la fonction intégrable $f\mathbf{1}_A$, soit en intégrant la restriction $f|_A$ par rapport à la restriction μ_A de μ à l'espace mesurable $(A, \mathcal{A}_A)$, où $\mathcal{A}_A$ est la $\sigma-$algèbre définie par $\mathcal{A}_A := \mathcal{P}(A) \cap \mathcal{A}$.

Proposition A.25. *Pour tout $f \in \mathcal{L}^1(\mathcal{A}, \mu)$ et $A \in \mathcal{A}$, on a $\mu(f\mathbf{1}_A) = \mu_A(f|_A)$.*

Proof. Tout d'abord, cette propriété est vraie pour les fonctions $f = \mathbf{1}_B$, $B \in \mathcal{A}$, puisque dans ce cas $\mu(\mathbf{1}_B\mathbf{1}_A) = \mu(A \cap B) = \mu_A(\mathbf{1}_B|_A)$. Par linéarité, cette égalité reste vraie pour les fonctions simples, puis par convergence monotone pour les fonctions mesurables positives. Enfin, pour $f \in \mathcal{L}^1(\mathcal{A}, \mu)$, on décompose $f = f^+ - f^-$, et on obtient le résultat voulu en appliquant l'égalité à f^+ et f^-. ♢

Voici un résultat qui rappelle une propriété classique sur les intégrales de Riemann, éventuellement impropres.

Lemma A.26. *Soit $f \in \mathcal{L}^1(\mathcal{A}, \mu)$ et $\varepsilon > 0$. Alors, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{A}$ vérifiant $\mu(A) < \delta$, on a $\mu(|f|\mathbf{1}_A) < \varepsilon$.*

Proof. Supposons, au contraire, qu'il existe ε_0 et une suite $(A_n)_n \subset \mathcal{A}$ tels que $\mu(A_n) < 2^{-n}$ et $\mu(|f|\mathbf{1}_{A_n}) \geq \varepsilon_0$. D'après le premier lemme de Borel-Cantelli, lemme A.14, on déduit que $A := \limsup A_n$ est négligeable. En particulier $\mu(|f|\mathbf{1}_A) = 0$, et on obtient une contradiction en remarquant que $\mu(|f|\mathbf{1}_A) = \mu(|f|) - \mu(|f|\mathbf{1}_{A^c}) \geq \mu(|f|) - \liminf \mu(|f|\mathbf{1}_{A_n^c}) = \limsup \mu(|f|\mathbf{1}_{A_n}) \geq \varepsilon_0$, où on a utilisé le lemme de Fatou. ♢

A.2.4 De la convergence p.p. à la convergence $\mathcal{L}^1$

Theorem A.27. *(convergence dominée) Soient $(f_n)_n \subset \mathcal{L}^0(\mathcal{A})$ une suite telle que $f_n \longrightarrow f$ $\mu-$a.e. pour une certaine fonction $f \in \mathcal{L}^0(\mathcal{A})$. Si $\sup_n |f_n| \in \mathcal{L}^1(\mathcal{A}, \mu)$, alors*

$$f_n \longrightarrow f \ \textit{dans}\ \mathcal{L}^1(\mathcal{A}, \mu) \quad \textit{i.e.} \quad \mu(|f_n - f|) \longrightarrow 0.$$

En particulier, $\mu(f_n) \longrightarrow \mu(f)$.

Proof. On note $g := \sup_n f_n$, $h_n := f_n - f$. Alors, les fonctions $2g+h_n$ et $2g-h_n$ sont positives, on obtient par le lemme de Fatou que $\liminf \mu(g-f_n) \geq \mu(g-f)$ et $\liminf \mu(g+f_n) \geq \mu(g+f)$. Du fait que g est intégrable, on peut utiliser la linéarité de l'intégrale, et on arrive à $\mu(f) \leq \liminf \mu(f_n) \leq \limsup \mu(f_n) \leq \mu(f)$. ♢

Le résultat suivant donne une condition nécessaire et suffisante pour qu'une suite convergente $\mu-$p.p. soit convergente dans $\mathcal{L}^1(\mathcal{A})$.

Lemma A.28. *(Scheffé) Soit $(f_n)_n \subset \mathcal{L}^1(\mathcal{A},\mu)$ telle que $f_n \longrightarrow f$ $\mu-$p.p. pour une certaine fonction $f \in \mathcal{L}^1(\mathcal{A},\mu)$. Alors:*

$$f_n \longrightarrow f \text{ dans } \mathcal{L}^1(\mathcal{A},\mu) \quad \text{ssi} \quad \mu(|f_n|) \longrightarrow \mu(|f|).$$

Proof. L'implication "$\Longrightarrow$" est triviale. Pour l'inégalité inverse, on procède en deux étapes.
Etape 1 Supposons que $f_n, f \geq 0$, $\mu-$p.p. Alors $(f_n - f)^- \leq f \in \mathcal{L}^1(\mathcal{A})$, et on déduit du théorème de convergence dominée que $\mu\left((f_n-f)^-\right) \longrightarrow 0$. Pour conclure, on écrit que $\mu(|f_n - f|) = \mu(f_n) - \mu(f) + 2\mu\left((f_n-f)^-\right) \longrightarrow 0$.
Etape 2 Pour f_n et f de signe quelconque, on utilise le lemme de Fatou pour obtenir $\mu(|f|) = \lim\{\mu(f_n^+) + \mu(f_n^-)\} \geq \mu(f^+) + \mu(f^-) = \mu(|f|)$ et par suite toutes les inégalit{es sont des égalité, i.e. $\lim \mu(f_n^+) = \mu(f^+)$ et $\lim \mu(f_n^-) = \mu(f^-)$. On est alors ramené au contexte de l'étape 1, qui permet d'obtenir $f_n^+ \longrightarrow f^+$ and $f_n^- \longrightarrow f^-$ dans $\mathcal{L}^1(\mathcal{A})$, et on conclut en écrivant $|f_n - f| \leq |f_n^+ - f^+| + |f_n^- - f^-|$ et en utilisant la monotonie de l'intégrale. ♢

Exercise A.29. *Soient $(\Omega,\mathcal{A},\mu)$ un espace mesuré, I un intervalle ouvert de $\mathbb{R}$, et $f : I \times \Omega \longrightarrow \mathbb{R}$ une fonction telle que $f(x,.) \in \mathcal{L}^0(\mathcal{A})$ pour tout $x \in I$.*

1. *On suppose qu'il existe une fonction $g \in \mathcal{L}^1_+(\mathcal{A},\mu)$ telle que $|f(x,.)| \leq g$, $\mu-$p.p. Montrer alors que, si $f(.,\omega)$ est continue en en un point $x_0 \in I$, $\mu-$p.p., la fonction $\phi : I \longrightarrow \mathbb{R}$ définie par*

$$\phi(x) := \int f(x,\omega)d\mu(\omega); \quad x \in I,$$

   est bien définie, et qu'elle est continue au point x_0.

2. *On suppose que la dérivée partielle $f_x := (\partial f/\partial x)$ existe pour tout $x \in I$, $\mu-$p.p. et qu'il existe une fonction $h \in \mathcal{L}^1_+(\mathcal{A},\mu)$ telle que $|f_x(x,.)| \leq h$, $\mu-$p.p. Montrer alors que ϕ est dérivable sur I, et*

$$\phi'(x) = \int \frac{\partial f}{\partial x}(x,\omega)d\mu(\omega); \quad x \in I.$$

3. *Donner des conditions qui assurent que ϕ soit continuement dérivable sur I.*

A.2.5 Intégrale de Lebesgue et intégrale de Riemann

Dans ce paragraphe, nous donnons quelques éléments qui expliquent l'avantage de l'intégrale de Lebesgue par rapport à celle de Riemann. Pour être plus concret, on considère le problème d'intégration sur $\mathbb{R}$.

(a) *L'intégrale de Riemann est construite sur un intervalle* $[a, b]$ *compact de* $\mathbb{R}$. Il y a bien une extension par les intégrales impropres, mais celà conduit à un cadre assez restrictif.

(b) *L'intégrale de Riemann est construite en approximant la fonction par des fonctions en escalier*, i.e. constantes sur des sous-intervalles de $[a, b]$ de longueur petite. Sur un dessin, il s'agit d'une approximation verticale. Par contre, *l'intégrale de Lebesgue est construite en découpant l'intervalle image et en approximant f sur les images réciproques de ces intervalles.* Il s'agit dans ce cas d'une approximation horizontale de la fonction à intégrer.

(c) *Les fonctions Riemann intégrables sont Lebesgue intégrables.* Montrons ceci dans $[0, 1]$. Soit f une fonction Riemann integrable bornée sur $\Omega = [0, 1]$ d'intégrale (au sens de Riemann) $\int_0^1 f(x)dx$. Alors f est Lebesgue intégrable d'intégrale $\lambda(f) = \int_0^1 f(x)dx$. Si f est une fonction en escalier, ce résultat est trivial. Pour une fonction Rieman intégrable f arbitraire, on peut trouver deux suites de fonctions en escalier $(g_n)_n$ et $(h_n)_n$ croissante et décroissante, respectivement, telles que $g_n \le f \le h_n$ et $\inf_n \int_0^1 (g_n - h_n)(x)dx = \lim_{n\to\infty} \int_0^1 (g_n - h_n)(x)dx = 0$. Sans perte de généralité, on peut supposer $h_n \le 2\|f\|_\infty$. Les fonctions $f_* := \sup_n g_n$ et $f^* := \inf_n h_n = 2M - \sup_n(-2M + h_n)$ sont boreliennes, et on a $f_* \le f \le f^*$. D'après la monotonie de l'intégrale:

$$0 \le \mu(f^* - f_*) = \mu\left(\inf(h_n - g_n)\right) \le \inf_n \mu(h_n - g_n) = 0,$$

et par suite $f = f^* = f_*$. Enfin:

$$\mu(f_*) = \lim \uparrow \mu(g_n) = \lim \uparrow \int_0^1 g_n(x)dx = \int_0^1 f(x)dx$$

La réciproque n'est pas vraie. Par exemple, la fonction $f = \mathbf{1}_{\{\mathbb{Q}\cap[0,1]\}}$ est $\lambda-$intégrable, mais n'est pas Riemann-intégrable.

(d) *Le théorème de convergence dominée n'a pas son équivalent dans le cadre de l'intégrale de Riemann*, et permet d'obtenir un espace de fonctions intégrables complet (on verra ce résultat plus tard). Par contre, on peut construire des exemples de suites de Cauchy de fonctions Riemann intégrables dont la limite n'est pas Riemann intégrable.

(e) *Pour les fonctions définies par des intégrales, les résultats de continuité et de dérivabilité sont simplement obtenus* grâce au théorème de convergence dominée. Leur analogue dans le cadre des intégrales de Riemann conduit à des résultats assez restrictifs.

(f) *L'intégrale de Lebesgue se définit naturellement dans* $\mathbb{R}^n$, alors que la situation est un peu plus compliquée pour l'intégrale de Riemann. En particulier, le théorème de Fubini est d'une grande simplicité dans le cadre de l'intégrale de Lebesgue.

A.3 Transformées de mesures

A.3.1 Mesure image

Soit $(\Omega_1, \mathcal{A}_1, \mu_1)$ un espace mesuré, $(\Omega_2, \mathcal{A}_2)$ un espace mesurable et $f : \Omega_1 \longrightarrow \Omega_2$ une fonction mesurable, i.e. $f^{-1} : \mathcal{A}_2 \longrightarrow \mathcal{A}_1$. On vérifie immédiatement que l'application:

$$\mu_2(A_2) := \mu_1\left(f^{-1}(A_2)\right) \quad \text{pour tout} \quad A_2 \in \Omega_2,$$

définit une mesure sur $(\Omega_2, \mathcal{A}_2)$.

Definition A.30. *μ_2 est appelée mesure image de μ_1 par f, et est notée $\mu_1 f^{-1}$.*

Theorem A.31. *(transfert) Soient $\mu_2 := \mu_1 f^{-1}$, la mesure image de μ_1 par f, et $h \in \mathcal{L}^0(\mathcal{A}_2)$. Alors $h \in \mathcal{L}^1(\mathcal{A}_2, \mu_2)$ si et seulement si $h \circ f \in \mathcal{L}^1(\mathcal{A}_1, \mu_1)$. dans ces conditions, on a*

$$\int_{\Omega_2} hd(\mu_1 f^{-1}) \quad = \quad \int_{\Omega_1} (h \circ f) d\mu_1. \tag{A.6}$$

Proof. On commence par vérifier la formule de transfert (A.6) pour les fonctions positives. La formule est vraie pour les fonctions $\mathbf{1}_{A_2}$, $A_2 \in \mathcal{A}_2$, puis oar linéarité pour les fonctions simples positives, et on conclut par le biais du théorème de convergence monotone. Pour h de signe arbitraire intégrable, on applique le résultat précédente à h^+ et h^-. Enfin, la formule de transfert montre que $h \in \mathcal{L}^1(\mathcal{A}_2, \mu_2)$ ssi $h^+ \circ f$ et $h^- \circ f \in \mathcal{L}^1(\mathcal{A}_1, \mu_1)$, et l'équivalence découle du fait que $h^+ \circ f = (h \circ f)^+$ et $h^- \circ f = (h \circ f)^-$. $\diamondsuit$

A.3.2 Mesures définies par des densités

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et soit $f \in \mathcal{L}^0_+(\mathcal{A})$ une fonction mesurable positive finie. On définit

$$\nu(A) := \mu(f\mathbf{1}_A) = \int_A f d\mu \quad \text{pour tout} \quad A \in \mathcal{A}.$$

Exercise A.32. *Vérifier que ν est une mesure sur $(\Omega, \mathcal{A})$.*

Definition A.33. (i) *La mesure ν est appelée mesure de densité f par rapport à μ, et on note $\nu = f \cdot \mu$.*
(ii) *Soient μ_1, μ_2 deux mesures sur un espace mesurable $(\Omega, \mathcal{A})$. On dit que μ_2 est absoluement continue par rapport à μ_1, et on note $\mu_2 \prec \mu_1$, si $\mu_2(A) = 0 \Longrightarrow \mu_1(A)$ pour tout $A \in \mathcal{A}$. Sinon, on dit que μ_2 est étrangère à μ_1. Si $\mu_2 \prec \mu_1$ et $\mu_1 \prec \mu_2$, on dit que μ_1 et μ_2 sont équivalentes, et on note $\mu_1 \sim \mu_2$. Enfin, si $\mu_2 \not\prec \mu_1$ et $\mu_1 \not\prec \mu_2$, on dit que μ_1 et μ_2 sont singulières.*

Ainsi, la mesure $f \cdot \mu$ est absoluement continue par rapport à μ.

Theorem A.34. (i) *Pour $g : \Omega \longrightarrow [0, \infty]$ $\mathcal{A}-$mesurable positive, on a $(f \cdot \mu)(g) = \mu(fg)$.*
(ii) *Pour $g \in \mathcal{L}^0_+(\mathcal{A})$, on a $g \in \mathcal{L}^1(\mathcal{A}, f \cdot \mu)$ ssi $fg \in \mathcal{L}^1(\mathcal{A}, \mu)$, et alors $(f \cdot \mu)(g) = \mu(fg)$.*

Exercise A.35. *Prouver le théorème A.34 (en utilisant le shémas de démonstration habituel).*

A.4 Inégalités remarquables

Dans ce paragraphe, nous énonçons trois inégalités qui sont très utiles. Afin d'habituer le lecteur à la manipulation des mesures et de l'intégration, nous formulons les résultats sous forme d'exercices.

Exercise A.36. *(Inégalité de Markov) Soit f une fonction $\mathcal{A}-$mesurable, et $g : \mathbb{R} \longrightarrow \mathbb{R}_+$ une fonction borelienne croissante positive.*

1. *Justifier que $g \circ f$ est une fonction mesurable, et pour tout $c \in \mathbb{R}$:*
$$\mu(g \circ f) \quad \geq \quad g(c)\mu(\{f \geq c\}). \tag{A.7}$$
2. *Montrer que*
$$\begin{aligned} c\mu(\{f \geq c\}) \leq \mu(f) &\quad \text{pour tout} \quad f \in \mathcal{L}^0_+(\mathcal{A}) \text{ et } c > 0, \\ c\mu[|f| \geq c] \leq \mu(|f|) &\quad \text{pour tout} \quad f \in \mathcal{L}^1(\mathcal{A}, \mu) \text{ et } c > 0. \end{aligned}$$
3. *Montrer l'inégalité de Chebychev:*
$$c^2\mu[|f| \geq c] \leq \mu(f^2) \quad \text{pour tout} \quad f^2 \in \mathcal{L}^1(\mathcal{A}, \mu) \text{ et } c > 0.$$
4. *Montrer que*
$$\mu(\{f \geq c\}) \leq \inf_{\tau > 0} e^{-\tau c}\mathbb{E}[e^{\tau f}] \quad \text{pour tout} \quad f \in \mathcal{L}^0(\mathcal{A}) \text{ et } c \in \mathbb{R}.$$

Exercise A.37. *(Inégalité de Schwarz) Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $f, g : \mathcal{A} \longrightarrow \mathbb{R}_+$ deux fonctions mesurables positives telle que $\mu(f^2) + \mu(g^2) < \infty$.*

1. *Montrer que $\mu(fg) < \infty$.*

2. *Montrer que* $\mu(fg)^2 \le \mu(f^2)\mu(g^2)$ *(Indication: considérer la fonction* $xf + g$, $x \in \mathbb{R}$*).*

3. *Montrer que l'inégalité de Schwarz dans la question 2 est valable sans la condition de positivité de* f *et* g.

Exercise A.38. *(Inégalité de Hölder, inégalité de Minkowski) On admet l'inéglité de Jensen, valable pour une mesure* ν *sur* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *telle que* $\nu(\mathbb{R}) = 1$*:*

$$\nu(c(f)) \ge c(\nu(f)) \quad \textit{pour} \quad f, c(f) \in \mathcal{L}^1(\mathcal{B}_{\mathbb{R}}, \nu) \ \textit{ et } c(.) \textit{ convexe},$$

qui sera démontrée dans le chapitre B, theéorème B.6.

Soient $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré et* $f, g : \Omega \longrightarrow \mathbb{R}$ *deux fonctions mesurables avec*

$$\mu(|f^p|) < \infty \ \mu(|g|^q) < \infty \quad \textit{où} \quad p > 1, \ \frac{1}{p} + \frac{1}{q} = 1. \tag{A.8}$$

1. *On suppose* $f, g \ge 0$ *et* $\mu(f^p) > 0$*. Montrer l'inégalité de Hölder:*

$$\mu(|fg|) \le \mu(|f|^p)^{1/p}\mu(|g|^q)^{1/q} \quad \textit{pour} \quad p > 1, \ \frac{1}{p} + \frac{1}{q} = 1 \ \textit{ et } \mu(|f|^p) + \mu(|g|^q) < \infty,$$

(Indication: introduire la mesure $\nu := \frac{f^p}{\mu(f^p)} \cdot \mu$*.)*

2. *Montrer que l'inégalité de Hölder de la question 1 est valable sous les conditions* (A.8) *sans les conditions supplémentaires de la question précédente.*

3. *En déduire l'inégalité de Minkowski:*

$$\mu(|f + g|^p)^{1/p} \le \mu(|f|^p)^{1/p} + \mu(|g|^p)^{1/p} \quad \textit{pour} \quad p > 1 \ \textit{ et } \mu(|f|^p) + \mu(|g|^p) < \infty.$$

(Indication: décomposer $|f + g|^p = (f + g)|f + g|^{p-1}$*.)*

A.5 Espaces produits

A.5.1 Construction et intégration

Dans ce paragraphe, nous faisons la construction de la mesure produit sur le produit de deux espaces mesurés.

Soient $(\Omega_1, \mathcal{A}_1, \mu_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés. Sur l'espace produit $\Omega_1 \times \Omega_2$, on vérifie immédiatement que $\mathcal{A}_1 \times \mathcal{A}_2$ est un $\pi-$système. On définit alors la $\sigma-$algèbre qu'il engendre

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \quad := \quad \sigma\left(\mathcal{A}_1 \times \mathcal{A}_2\right).$$

Sur cette structure d'espace mesurable $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, on veut définir une mesure μ telle que

$$\mu(A \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \text{pour tous} \quad (A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2, \tag{A.9}$$

puis définir l'intégrale d'une fonction $f : \Omega_1 \times \Omega_2 \longrightarrow \mathbb{R}$ intégrable:

$$\int_{\Omega_1\times\Omega_2} f d\mu.$$

Une question importante est de relier cette quantité aux intégrales doubles

$$\int_{\Omega_2}\left(\int_{\Omega_1} f d\mu_1\right) d\mu_2 \quad \text{et} \quad \int_{\Omega_1}\left(\int_{\Omega_2} f d\mu_2\right) d\mu_1,$$

qui pose tout d'abord les questions de
(1a) la μ_1-mesurabilité de la fonction$f_2^{\omega_2} : \omega_1 \longmapsto f(\omega_1, \omega_2)$,
(2a) la μ_2-mesurabilité de la fonction $f_1^{\omega_1} : \omega_2 \longmapsto f(\omega_1, \omega_2)$,
puis, une fois ces questions règlées,
(1b) la $\mathcal{A}_1-$mesurabilité de la fonction $I_1^f : \omega_1 \longmapsto \int f(\omega_1, \omega_2) d\mu_2(\omega_2)$,
(2b) la $\mathcal{A}_2-$mesurabilité de la fonction $I_2^f : \omega_2 \longmapsto \int f(\omega_1, \omega_2) d\mu_1(\omega_1)$.
Ces deux problèmes sont résolus aisément grâce au théorème des classes monotones:

Lemma A.39. (a) *Soit $f \in \mathcal{L}^\infty(\mathcal{A}_1 \otimes \mathcal{A}_2)$. Alors, pour tous $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$:*

$$f_1^{\omega_1} \in \mathcal{L}^\infty(\mathcal{A}_2) \quad \text{et} \quad f_2^{\omega_2} \in \mathcal{L}^\infty(\mathcal{A}_1).$$

(b) *Supposons de plus que μ_1 et μ_2 soient finies. Alors $I_i^f \in \mathcal{L}^1(\mathcal{A}_i, \mu_i)$ pour $i = 1, 2$ et*

$$\int_{\Omega_1} I_1^f d\mu_1 \quad = \quad \int_{\Omega_2} I_2^f d\mu_2.$$

Proof. (a) Soit $\mathcal{H} := \{f \in \mathcal{L}^\infty(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) : f_i^{\omega_i} \in \mathcal{L}^\infty(\Omega_i, \mathcal{A}_i), i = 1, 2\}$. Les condition H1 et H2 sont trivialement satisfaites par $\mathcal{H}$. De plus, rappelons que $\mathcal{A}_1 \times \mathcal{A}_2$ est un $\pi-$système engendrant $\mathcal{A}_1 \otimes \mathcal{A}_2$, par définition. Il est claire que $\mathcal{H} \supset \{\mathbf{1}_A : A \in \mathcal{A}_1 \times \mathcal{A}_2\}$. Le théorème des classes monotones permet de conclure que $\mathcal{H} = \mathcal{L}^1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.
(b) Il suffit de refaire le même type d'argument que pour (a). $\diamondsuit$

Grâce au dernier résultat, nous pouvons maintenant définir un candidat pour la mesure sur l'espace produit $\Omega_1 \times \Omega_2$ par:

$$\mu(A) := \int\left(\int \mathbf{1}_A d\mu_1\right) d\mu_2 = \int\left(\int \mathbf{1}_A d\mu_2\right) d\mu_1 \quad \text{pour tout} \quad A \in \mathcal{A}_1 \otimes \mathcal{A}_2.$$

Theorem A.40. *(Fubini) L'application μ est une mesure sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, appelée mesure produit de μ_1 et μ_2, et notée $\mu_1 \otimes \mu_2$. C'est l'unique mesure sur $\Omega_1 \times \Omega_2$ vérifiant* (A.9)*. De plus, pour tout $f \in \mathcal{L}^0_+(\mathcal{A}_1 \otimes \mathcal{A}_2)$,*

$$\int f d\mu_1 \otimes \mu_2 \quad = \quad \int\left(\int f d\mu_1\right) d\mu_2 \;=\; \int\left(\int f d\mu_2\right) d\mu_1 \;\in\; [0, \infty] \tag{A.10}$$

Enfin, si $f \in \mathcal{L}^1(\mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$, les égalités (A.10) *sont valides.*

Proof. On vérifie que $\mu_1 \otimes \mu_2$ est une mesure grâce aux propriétés élémentaires de l'intégrale de Lebesgue. L'unicité est une conséquence immédiate de la proposition A.5. Les égalités (A.10) ont déjà été établies dans le lemme A.39 (b) pour f bornée et des mesures finies. Pour généraliser à des fonctions f mesurables positives, on introduit des approximations croissantes, et utilise le théorème de convergence monotone. Enfin, pour des fonctions $f \in \mathcal{L}^1(\mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$, on applique le résultat précédent à f^+ et f^-. $\diamondsuit$

Remark A.41. (i) La construction de ce paragraphe, ainsi que les résultats d'intégration ci-dessous, s'étendent sans difficulté pour la construction du produit de n espaces mesurés au prix de notations plus encombrantes.
(ii) Soit maintenant $(\Omega_i, \mathcal{A}_i)_{i\geq 1}$ une famille dénombrable d'espaces mesurés, et $\Omega := \prod_{i\geq 1} \Omega_i$. Pour tout sous-ensemble fini $I \subset \mathbb{N}$, et pour tous $A_i \in \mathcal{A}_i$, $i \in I$, on définit le *cylindre*

$$\mathcal{C}(A_i, i \in I) \quad := \quad \{\omega \in \Omega : \omega_i \in A_i \text{ pour } i \in I\}.$$

La $\sigma-$algèbre produit est alors définie par

$$\mathcal{A} := \otimes_{n\geq 1} \mathcal{A}_i := \sigma\left(\mathcal{C}(A_i, i \in I) : \ I \subset \mathbb{N}, \ \mathrm{card}(I) < \infty\right\}.$$

A.5.2 Mesure image et changement de variable

Soit $\mathcal{O} = \mathbb{R}^n$, ou d'un espace de dimension n. Les outils développés dans les paragraphes précédents permettent de définir la mesure de Lebesgue sur $\mathbb{R}^n$ à partir de notre construction de la mesure de Lebesgue sur $\mathbb{R}$.

Dans ce paragraphe, on considère une fonction

$$g : \Omega_1 \longrightarrow \Omega_2 \quad \text{où} \quad \Omega_1, \Omega_2 \text{ ouverts de } \mathbb{R}^n.$$

On note $g = (g_1, \ldots, g_n)$. Si g est différentiable en un point $x \in \Omega_1$, on note par

$$Dg(x) := \left(\frac{\partial g_i}{\partial x_j}\right)_{1\leq i,j\leq n} \quad \text{et} \quad \det[Dg(x)]$$

la matrice jacobienne de f en x et son déterminant. Rappelons enfin que g est un C^1-difféomorphisme si g est une bijection telle g et g^{-1} sont de classe C^1, et que dans ce cas

$$\det[Dg^{-1}(y)] \quad = \quad \frac{1}{\det[Dg \circ g^{-1}(y)]}.$$

Theorem A.42. *Soit μ_1 une mesure sur $(\Omega_1, \mathcal{B}_{\Omega_1})$ de densité par rapport à la mesure de Lebesgue $f_1 \in \mathcal{L}^0_+(\mathcal{B}_{\Omega_1})$, i.e. $\mu_1(dx) = \mathbf{1}_{\Omega_1} f_1(x) \cdot dx$. Si g est un C^1-difféomorphisme, la mesure image $\mu_2 := \mu g^{-1}$ est absolument continue par rapport à la mesure de Lebesgue de densité*

$$f_2(y) = \mathbf{1}_{\Omega_2}(y) f\left(g^{-1}\right) |\det[Dg^{-1}(y)]| \ \text{et} \ \int_{\Omega_1} h \circ g(x) f_1(x) dx = \int_{\Omega_2} h(y) f_2(y) dy$$

pour toute fonction $h : \Omega_2 \longrightarrow \mathbb{R}$ positive ou μ_2-intégrable.

Pour la démonstration, on renvoit au cours de première année.

A.6 Annexe du chapitre A

A.6.1 $\pi-$système, $d-$système et unicité des mesures

Commençons par introduire une notion supplémentaire de classes d'ensembles.

Definition A.43. *Une classe $\mathcal{D} \subset \mathcal{P}(\Omega)$ est appelée $d-$système si $\Omega \subset \mathcal{D}$, $B \setminus A \in \mathcal{D}$ pour tous $A, B \in \mathcal{D}$ avec $A \subset B$, et $\cup_n A_n \in \mathcal{D}$ pour toute suite croissante $(A_n)_n \subset \Omega$.*

Lemma A.44. *Une classe $\mathcal{C} \subset \mathcal{P}(\Omega)$ est une $\sigma-$algèbre si et seulement si $\mathcal{C}$ est un $\pi-$système et un $d-$système.*

La preuve facile de ce résultat est laissée en exercice. Pour toute classe $\mathcal{C}$, on définit l'ensemble

$$d(\mathcal{C}) \quad := \quad \cap \{\mathcal{D} \supset \mathcal{C} : \ \mathcal{D} \text{ est un } d-\text{système}\},$$

qui est le plus petit $d-$système contenant $\mathcal{C}$. L'inclusion $d(\mathcal{C}) \subset \sigma(\mathcal{C})$ est évidente.

Lemma A.45. *Pour un $\pi-$système $\mathcal{I}$, on a $d(\mathcal{I}) = \sigma(\mathcal{I})$.*

Proof. D'après le lemme A.44, il suffit de montrer que $d(\mathcal{I})$ est un $\pi-$system, i.e. que $d(\mathcal{I})$ est stable par intersection finie. On définit l'ensemble $\mathcal{D}' := \{A \in d(\mathcal{I}) : A \cap B \in d(\mathcal{I})$ pour tout $B \in d(\mathcal{I})\}$, et on va montrer que $\mathcal{D}' = d(\mathcal{I})$ ce qui termine la démonstration.
1- On commence par montrer que l'ensemble $\mathcal{D}_0 := \{B \in d(\mathcal{I}) : B \cap C \in d(\mathcal{I})$ pour tout $C \in Ic\}$ est un $d-$système. En effet:
- $\Omega \in \mathcal{D}$;
- soient $A, B \in \mathcal{D}_0$ tels que $A \subset B$, et $C \in \mathcal{I}$; comme $A, B \in \mathcal{D}_0$, on a $(A \cap C)$ et $(B \cap C) \in d(\mathcal{I})$, et du fait que $d(\mathcal{I})$ est un $d-$système, on voit que $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C) \in d(\mathcal{I})$;
- enfin, si $\mathcal{D}_0 \ni A_n \uparrow A$ et $C \in \mathcal{I}$, on a $A_n \cap C \in d(\mathcal{I})$ et donc $\lim \uparrow (A_n \cap C) = A \cap C \in d(\mathcal{I})$ du fait que $d(\mathcal{I})$ est un $d-$système;
2- par définition $\mathcal{D}_0 \subset d(\mathcal{I})$, et comme on vient de montrer que c'est un $d-$système contenant $\mathcal{I}$, on voit qu'on a en fait $\mathcal{D}_0 = d(\mathcal{I})$; on vérifie maintenant que ceci implique que $\mathcal{I} \in \mathcal{D}'$;
3- enfin, en procédant comme dans les étapes précédentes, on voit que $\mathcal{D}'$ est un $d-$système. $\diamondsuit$

Preuve de la proposition A.5 On vérifie aisément que l'ensemble $\mathcal{D} := \{A \in \sigma(\mathcal{I}) : \mu(A) = \nu(A)\}$ est un $d-$système (c'est à ce niveau qu'on utilise que les mesures sont finies afin d'éviter des formes indéterminées du type $\infty - \infty$). Or, par hypothèse, $\mathcal{D}$ contient le $\pi-$système $\mathcal{I}$. On déduit alors du lemme A.45 que $\mathcal{D}$ contient $\sigma(\mathcal{I})$ et par suite $\mathcal{D} = \sigma(\mathcal{I})$. $\diamondsuit$

A.6.2 Mesure extérieure et extension des mesures

Le but de ce paragraphe est de démonrer du théorème de Caracthéodory A.6 dont nous rappeleons l'énoncé.

Théorème A.6 *Soient $\mathcal{A}_0$ une algèbre sur Ω, et $\mu_0 : \mathcal{A}_0 \longrightarrow \mathbb{R}_+$ une fonction $\sigma-$additive. Alors il existe une mesure μ sur $\mathcal{A} := \sigma(\mathcal{A}_0)$ telle que $\mu = \mu_0$ sur $\mathcal{A}_0$. Si de plus $\mu_0(\Omega) < \infty$, alors une telle extension μ est unique.*

Pour préparer la démonstration, nous considérons une $\sigma-$algèbre $\mathcal{A}' \subset \mathcal{P}(\Omega)$, et une application $\lambda : \mathcal{A}' \longrightarrow [0, \infty]$ vérifiant $\lambda(\emptyset) = 0$.

Definition A.46. *On dit que λ est une mesure extérieure sur $(\Omega, \mathcal{A}')$ si*
(i) $\lambda(\emptyset) = 0$,
(ii) *λ est croissante: pour $A_1, A_2 \in \mathcal{A}'$, $\lambda(A_1) \le \lambda(A_2)$ dès que $A_1 \subset A_2$,*
(iii) *λ est $\sigma-$sous-additive: pour $(A_n)_n \subset \mathcal{A}'$, on a $\lambda(\cap_n A_n) \le \sum_n \lambda(A_n)$.*

Definition A.47. *On dit qu'un élément $A \in \mathcal{A}'$ est un $\lambda-$ensemble si*

$$\lambda(A \cap B) + \lambda(A^c \cap B) = \lambda(B) \quad \textit{pour tout} \quad B \in \mathcal{A}_0,$$

(en particulier, $\lambda(\emptyset) = 0$). On note par $\mathcal{A}'_\lambda$ l'ensemble de tous les $\lambda-$ensembles de $\mathcal{A}'$.

Le résultat suivant utilise uniquement le fait que $\mathcal{A}'$ est une algèbre.

Lemma A.48. *L'ensemble $\mathcal{A}'_\lambda$ est une algèbre, et la restriction de λ à $\mathcal{A}'_\lambda$ est additive et vérifie pour tout $B \in \mathcal{A}'$:*

$$\lambda\left(\cup_{i=1}^n (A_i \cap B)\right) = \sum_{i=1}^n \lambda(A_i \cap B) \quad \textit{dès} \quad A_1, \ldots, A_n \in \mathcal{A}'_\lambda \textit{ sont disjoints.}$$

Ce lemme, dont la démonstration (facile) est reportée pour la fin du paragraphe, permet de montrer le résultat suivant:

Lemma A.49. *(Carathéodory) Soit λ une mesure extérieure sur $(\Omega, \mathcal{A}')$. Alors $\mathcal{A}'_\lambda$ est une $\sigma-$algèbre, et la restriction de λ à $\mathcal{A}'_\lambda$ est $\sigma-$additive, et par suite λ est une mesure sur $(\Omega, \mathcal{A}'_\lambda)$.*

Proof. En vue du lemme A.48, il reste à montrer que pour une suite d'ensembles disjoints $(A_n)_n \subset \mathcal{A}'_\lambda$, on a

$$L := \cup_n A_n \in \mathcal{A}'_0(\lambda) \quad \text{et} \quad \lambda(\cup_n A_n) = \sum_n \lambda(A_n). \tag{A.11}$$

Notons $\bar{A}_n := \cup_{i \le n} A_i$, $\bar{A} := \cup_n A_n$, et remarquons que $\bar{A}^c \subset \bar{A}_n^c$. D'après le lemme A.48, $\bar{A}_n \in \mathcal{A}'_\lambda$ et pour tout $B \in \mathcal{A}'$:

$$\lambda(B) = \lambda(\bar{A}_n^c \cap B) + \lambda(\bar{A}_n \cap B) \ge \lambda(\bar{A}^c \cap B) + \lambda(\bar{A}_n \cap B) = \lambda(\bar{A}^c \cap B) + \sum_{i \le n} \lambda(A_i \cap B).$$

On continue en faisant tendre n vers l'infini, et en utilisant (deux fois) la sous-additivité de λ:

$$\lambda(B) \geq \lambda(\bar{A}^c \cap B) + \sum_n \lambda(A_i \cap B) \geq \lambda(\bar{A}^c \cap B + \lambda(A \cap B) \geq \lambda(B).$$

On déduit que toutes les inégalités sont des égalités, prouvant que $\bar{A} \in \mathcal{A}'_\lambda$, et pour $B = \bar{A}$ on obtient la propriété de sous-additivité de λ, finissant la preuve de (A.11). $\diamondsuit$

Nous avons maintenant tous les ingrédients pour montrer le théorème d'extension de Carathéodory.

Preuve du théorème A.6 On considère la $\sigma-$algèbre $\mathcal{A}' := \mathcal{P}(\Omega)$, et on définit l'application sur Ω:

$$\lambda(A) := \inf \left\{ \sum_n \mu_0(B_n) : \ (B_n)_n \subset \mathcal{A}_0, \ B_n \text{ disjoints et } A \subset \cup_n B_n \right\}.$$

Etape 1 Montrons que λ est une mesure extérieure sur $(\Omega, \mathcal{P})$, ce qui implique par le lemme A.49 que

$$\lambda \quad \text{est une mesure sur} \quad (\Omega, \mathcal{A}'_\lambda). \tag{A.12}$$

Il est clair que $\lambda(\emptyset) = 0$, et que λ est croissante, il reste donc à vérifier que λ est $\sigma-$sous-additive. Soit une suite $(A_n)_n \subset \mathcal{P}$ telle que $\lambda(A_n) < \infty$ pour tout n, et soit $A := \cup_n A_n$. Pour tout $\varepsilon > 0$ et $n \geq 1$, on considère une suite $\varepsilon-$optimale $(B_i^{n,\varepsilon})_i \subset \mathcal{A}_0$ du problème de minimisation $\lambda(A_n)$, i.e. $B_i^{n,\varepsilon} \cap B_j^{n,\varepsilon} = \emptyset$,

$$A_n \subset \cup_k B_k^{n,\varepsilon} \quad \text{et} \quad \lambda(A_n) > \sum_k \mu_0(B_k^{n,\varepsilon}) - \varepsilon 2^{-n}.$$

Alors, $\lambda(A) \leq \sum_{n,k} \mu_0(B_k^{n,\varepsilon}) < \varepsilon + \sum_n \lambda(A_n) \longrightarrow \sum_n \lambda(A_n)$ quand $\varepsilon \to 0$.
Etape 2 Rappelons que $\sigma(\mathcal{A}_0) \subset \mathcal{A}'_\lambda$. Alors, pour finir la démonstration de l'existence d'une extension, il nous reste à montrer que

$$\mathcal{A}_0 \subset \mathcal{A}'_\lambda \quad \text{et} \quad \lambda = \mu_0 \text{ sur } \mathcal{A}_0, \tag{A.13}$$

pour ainsi définir μ comme la restriction de λ à $\sigma(\mathcal{A}_0)$.
1- Commençons par montrer que $\lambda = \mu_0$ sur $\mathcal{A}_0$. L'inégalité $\lambda \leq \mu_0$ sur $\mathcal{A}_0$ est triviale. Pour l'inégalité inverse, on considère $A \in \mathcal{A}_0$ et une suite $(B_n)_n \subset \mathcal{A}_0$ d'éléments disjoints telle $A \subset \cup_n B_n$. Alors, en utilisant la $\sigma-$additivité de μ_0 sur $\mathcal{A}_0$:

$$\mu_0(A) = \mu_0\left(\cup_n A \cap B_n\right) = \sum_n \mu_0(A \cap B_n) \leq \sum_n \mu_0(B_n) = \lambda(A).$$

2- Montrons maintenant que $\mathcal{A}_0 \subset \mathcal{A}'_\lambda$. Soient $A \in \mathcal{A}'$, $\varepsilon > 0$ et $(B_n)_n \subset \mathcal{A}_0$ une suite $\varepsilon-$optimale pour le problème de minimsation $\lambda(A)$. Alors, pour tout $A_0 \in \mathcal{A}_0$, on a

$$\begin{aligned}
\lambda(A) + \varepsilon \geq \sum_n \mu_0(B_n) &= \sum_n \mu_0(A_0 \cap B_n) + \sum_n \mu_0(A_0^c \cap B_n) \\
&\geq \lambda((A_0 \cap A) + \lambda((A_0^c \cap A) \\
&\geq \lambda(A),
\end{aligned}$$

où les deux dernières inégalités découlent respectivement de la monotonie et la sous-linéarité de λ. Comme $\varepsilon > 0$ est arbitraire, ceci montre que A_0 est un $\lambda-$ensemble, i.e. $A_0 \in \mathcal{A}'_\lambda$. $\diamondsuit$

Preuve du lemme A.48 1- Commençons par montrer que $\mathcal{A}'_\lambda$ est une algèbre. Il est clair que $\Omega \in \mathcal{A}'_\lambda$ et que $\mathcal{A}'_\lambda$ est stable par passage au complémentaire. Il reste à montrer que $A = A_1 \cap A_2 \in \mathcal{A}_0(\lambda)$ pour tous $A_1, A_2 \in \mathcal{A}_0(\lambda)$. En utilisant successivement le fait que $A_2 \in \mathcal{A}'_\lambda$ et que $A_2 \cap A^c = A_1^c \cap A_2$, $A_2^c \cap A^c = A_2^c$, on calcule directement:

$$\lambda(A^c \cap B) = \lambda(A_2 \cap A^c \cap B) + \lambda(A_2^c \cap A^c \cap B) = \lambda(A_1^c \cap A_2 \cap B) + \lambda(A_2^c \cap B).$$

On continue en utilisant le fait que $A_1, A_2 \in \mathcal{A}'_\lambda$:

$$\lambda(A^c \cap B) = \lambda(A_2 \cap B) - \lambda(A \cap B) + \lambda(A_2^c \cap B) = \lambda(B) - \lambda(A \cap B).$$

2- Pour des ensembles disjoints $A_1, A_2 \in \mathcal{A}'_\lambda$, on a $(A_1 \cup A_2) \cap A_1 = A_1$ et $(A_1 \cup A_2) \cap A_1^c = A_2$, et on utilise le fait que $A_1 \in \mathcal{A}'_\lambda$ pour voir que $\lambda((A_1 \cup A_2) \cap B) = \lambda(A_1 \cap B) + \lambda(A_2 \cap B)$, ce qui est l'égalité annoncée pour $n = 2$. L'extension pour un n plus grand est triviale, et la $\sigma-$addditivité de λ en est une conséquence immédiate. $\diamondsuit$

A.6.3 Démonstration du théorème des classes monotones

Rappelons l'énoncé.

Théorème A.18 *Soit $\mathcal{H}$ une classes de fonctions réelles bornées sur Ω vérifiant les conditions suivantes:*
(H1) *$\mathcal{H}$ est un espace vectoriel contenant la fonction constante $\mathbf{1}$,*
(H2) *pour toute suite croissante $(f_n)_n \subset \mathcal{H}$ de fonctions positives telle que $f := \lim \uparrow f_n$ est bornée, on a $f \in \mathcal{H}$.*
Soit $\mathcal{I}$ un $\pi-$système tel que $\{\mathbf{1}_A : A \in \mathcal{I}\} \subset \mathcal{H}$. Alors $\mathcal{L}^\infty(\sigma(\mathcal{I})) \subset \mathcal{H}$.

Proof. D'après les conditions H1 et H2, on voit immédiatement que l'ensemble $\mathcal{D} := \{F \subset \Omega : 1_F \in \mathcal{H}\}$ est un $d-$système. De plus, comme $\mathcal{D}$ contient le

$\pi-$système $\mathcal{I}$, on déduit que $\sigma(\mathcal{I}) \subset \mathcal{D}$. Soit maintenant $f \in \mathcal{L}^\infty_(\sigma(\mathcal{I}))$ bornée par $M > 0$, et

$$\phi_n(\omega) := \sum_{i=0}^{M2^{-n}} i2^{-n}\mathbf{1}_{A_i^n}(\omega), \quad \text{où} \quad A_i^n := \{\omega \in \Omega : \ i2^{-n} \leq f^+(\omega) < (i+1)2^{-n}\}.$$

Comme $A_i^n \in \sigma(\mathcal{I})$, on déduit de la structure d'espace vectoriel (condition H1) de $\mathcal{H}$ que $\phi_n \in \mathcal{H}$. De plus $(\phi_n)_n$ étant une suite croissante de fonctions positives convergeant vers la fonction bornée f^+, la condition H2 assure que $f^+ \in \mathcal{H}$. On montre de même que $f^- \in \mathcal{H}$ et, par suite, $f = f^+ - f^- \in \mathcal{H}$ d'après H1. $\diamondsuit$

Appendix B

Préliminaires de la théorie des probabilités

Dans ce chapitre, on spécialise l'analyse aux cas d'une mesure de probabilité, i.e. une mesure $\mathbb{P} : \mathcal{A} \longrightarrow \mathbb{R}_+$ telle que $\mathbb{P}[\Omega] = 1$. On dit alors que $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.

Bien évidemment, tous les résultats du chapitre précédent sont valables dans le cas présent. En plus de ces résultats, nous allons exploiter l'intuition probabiliste pour introduire de nouveaux concepts et obtenir de nouveaux résultats.

Ainsi, l'ensemble Ω s'interprète comme l'ensemble des événements élémentaires, et tout point $\omega \in \Omega$ est un événement élémentaire. La $s-$algèbre $\mathcal{A}$ est l'ensemble de tous les événements réalisables.

On remplacera systématiquement la terminologie $\mathbb{P}-$p.p. par $\mathbb{P}-$*presque surement*, notée $\mathbb{P}-$p.s. ou plus simplement p.s. s'il n'y a pas de risque de confusion.

Les fonctions $\mathbb{P}-$mesurables sont appelées variables aléatoires (on écrira v.a.), et on les notera, le plus souvent, par des lettres majuscules, typiquement X. La loi image $\mathbb{P}X^{-1}$ est appelée distribution de la v.a. X, et sera notée $\mathcal{L}_X$ s'il n'y a pas besoin de rappeler la probabilité $\mathbb{P}$.

B.1 Variables aléatoires

B.1.1 $\sigma-$algèbre engendrée par une v.a.

Nous commençons par donner un sens précis à l'information révélée par une famille de variables aléatoires.

Definition B.1. *Soient $\mathbb{T}$ un ensemble, et $\{X_\tau, \tau \in \mathbb{T}\}$ une famille quelconque de v.a. La $\sigma-$algèbre engendrée par cette famille $\mathcal{X} := \sigma(X_\tau : \tau \in \mathbb{T})$ est la plus petite $\sigma-$algèbre sur Ω telle que X_τ est $\mathcal{X}-$mesurable pour tout $\tau \in \mathbb{T}$, i.e.*

$$\sigma(X_\tau : \tau \in \mathbb{T}) \quad = \quad \sigma\left(\{X_\tau^{-1}(A) : \tau \in \mathbb{T} \text{ et } A \in \mathcal{B}_{\mathbb{R}}\}\right). \qquad \text{(B.1)}$$

Il est clair que si les X_τ sont $\mathcal{A}-$mesurables, alors $\sigma(X_\tau : \tau \in \mathbb{T}) \subset \mathcal{A}$.

Lemma B.2. *Soient X et Y deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ prenant valeurs respectivement dans $\mathbb{R}$ et dans $\mathbb{R}^n$. Alors X est $\sigma(Y)-$mesurable si et seulement si il existe une fonction borélienne $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ telle que $X = f(Y)$.*

Proof. Seule la condition nécessaire est non triviale. Par ailleurs quitte à transformer X par une fonction bijective bornée, on peut se limiter au cas où X est bornée. On définit

$$\mathcal{H} := \{f(Y) : f \in \mathcal{L}^\infty(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})\},$$

et on remarque que $\{1_A : A \in \sigma(Y)\} \subset \mathcal{H}$: d'après (B.1), pour tout $A \in \sigma(Y)$, il existe $B \in \mathcal{A}$ tel que $A = Y^{-1}(B)$, et par suite $\mathbf{1}_A = \mathbf{1}_B(Y)$.

Pour conclure, il nous suffit de montrer que $\mathcal{H}$ vérifie les conditions du théorème des classes monotones. Il est cair que $\mathcal{H}$ est un espace vectoriel contenant la v.a. constante 1. Soient $X \in \mathcal{L}^\infty_+(\mathcal{A}, \mathbb{P})$ et $(f_n(Y))_n$ une suite croissante de $\mathcal{H}$ telle que $f_n(Y) \uparrow X$. Alors $X = f(Y)$, où $f = \limsup f_n$ est $\mathcal{B}_{\mathbb{R}^n}-$mesurable bornée (puisque X l'est). ♢

B.1.2 Distribution d'une v.a.

La distribution, ou la loi, d'une v.a. X sur $(\Omega, \mathcal{A}, \mathbb{P})$ est définie par la mesure image $\mathcal{L}_X := \mathbb{P}X^{-1}$. En utilisant le $\pi-$système $\pi(\mathbb{R}) = \{]-\infty, c]) : c \in \mathbb{R}\}$, on déduit de la proposition A.5 que la loi $\mathcal{L}_X$ est caractérisée par la fonction

$$F_X(c) \;\; := \;\; \mathcal{L}_X(]-\infty, c]) \; = \; \mathbb{P}[X \leq c], \; c \in \mathbb{R}. \qquad \text{(B.2)}$$

La fonction F_X est appelée fonction de répartition.

Proposition B.3. (i) *La fonction F_X est croissante continue à droite, et $F_X(-\infty) = 0$, $F_X(\infty) = 1$,*
(ii) *Soit F une fonction croissante continue à droite, et $F(-\infty) = 0$, $F(\infty) = 1$. Alors il existe une variable aléatoire $\tilde{X}$ sur un espace de probabilité $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ telle que $F = F_{\tilde{X}}$.*

Proof. (i) est triviale. Pour (ii), une première approche consiste à construire une loi $\tilde{\mathcal{L}}$ en suivant le schémas de construction de la mesure de Lebesgue dans l'exemple A.7 qui utilise le théorème d'extension de Carathéodory; on prend alors $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \tilde{\mathcal{L}})$ et $X(\omega) = \omega$. La remarque suivante donne une approche alternative. ♢

Remark B.4. Etant donnée une fonction de répartition, ou une loi, voici une construction explicite d'une v.a. lui correspondant. Cette construction est utile, par exemple, pour la simulation de v.a. Sur l'espace de probabilité $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) := ([0, 1], \mathcal{B}_{[0,1]}, \lambda)$, λ étant la mesure de Lebesgue, on définit

$$\overline{X}(\omega) := \inf\{u : F(u) > \omega\} \quad \text{et} \quad \underline{X}(\omega) := \inf\{u : F(u) \geq \omega\}$$

1- $F_{\underline{X}} = F$: nous allons montrer que

$$\omega \leq F(c) \iff \underline{X}(\omega) \leq c, \tag{B.3}$$

et par suite $\mathbb{P}[\underline{X} \leq c] = F(c)$.

L'implication $\Longrightarrow$ découle de la définition. Pour l'inclusion inverse, on observe que $F(\underline{X}(\omega)) \geq \omega$. En effet, si ce n'était pas le cas, on déduirait de la continuité à droite de F que $F(\underline{X}(\omega)+\varepsilon) < \omega$ pour $\varepsilon > 0$ assez petit, impliquant l'absurdité $\underline{X}(\omega) + \varepsilon \leq \underline{X}(\omega)$!

Avec cette observation et la croissance de F, on voit que $\underline{X}(\omega) \leq c$ implique $\omega \leq F(\underline{X}(\omega)) \leq F(c)$ implique $\omega \leq F(c)$.

2- $F_{\underline{X}} = F$: par définition de $\overline{X}$, on a $\omega < F(c)$ implique $\overline{X}(\omega) \leq c$. Mais $\overline{X}(\omega) \leq c$ implique $\underline{X}(\omega) \leq c$ puisque $\underline{X} \leq \overline{X}$. On en déduit que $F(c) \leq \mathbb{P}[\overline{X} \leq c] \leq \mathbb{P}[\underline{X} \leq c] = F(c)$.

B.2 Espérance de variables aléatoires

Pour une v.a. $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, l'espérance dans le vocabulaire probabiliste est l'intégrale de X par rapport à $\mathbb{P}$:

$$\mathbb{E}[X] \; := \; \mathbb{P}(X) \; = \; \int_\Omega X d\mathbb{P}.$$

Pour une v.a. positive, $\mathbb{E}[X] \in [0, \infty]$ est toujours bien définie. Bien sûr, toutes les propriétés du chapitre A sont valides. Nous allons en obtenir d'autres comme conséquence de $\mathbb{P}[\Omega] = 1$.

B.2.1 Variables aléatoires à densité

Revenons à présente à la loi $\mathcal{L}_X$ sur $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ d'une v.a. X sur $(\Omega, \mathcal{A}, \mathbb{P})$. Par définition, on a:

$$\mathcal{L}_X(B) = \mathbb{P}[X \in B] \quad \text{pour tout} \quad B \in \mathcal{B}_\mathbb{R}.$$

Par linéarité de l'intégrale (par rapport à $\mathcal{L}_X$), on obtient $\mathbb{E}[g(X)] = \mathcal{L}_X(g) = \int_\mathbb{R} hd\mathcal{L}_X$ pour toute fonction simple $g \in \mathcal{S}^+$. On étend alors cette relation aux fonction g mesurables positives, par le théorème de convergence monotone, puis à $\mathcal{L}^1$ en décomposant $g = g^+ - g^-$. Ceci montre que $g(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ ssi $g \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}_\mathbb{R}, \mathcal{L}_X)$ et

$$\mathbb{E}[g(X)] \;\; = \;\; \mathcal{L}_X(g) \; = \; \int_\mathbb{R} hd\mathcal{L}_X. \tag{B.4}$$

Definition B.5. *On dit que X a une densité de probabilité f_X si $\mathcal{L}_X$ est absoluement continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et:*

$$\mathbb{P}[X \in B] = \int_B f_X(x)dx \quad \textit{pour tout} \quad B \in \mathcal{B}_\mathbb{R}.$$

Le lien entre la densité de probabilité, si elle existe, et la fonction de répartition (qui existe toujours) est facilement établie en considérant $B =]-\infty, c]$:

$$F_X(c) = \int_{]-\infty,c]} f_X(x)dx \quad \text{pour tout} \quad c \in \mathbb{R}.$$

qui exprime "f_X est la dérivée de F_X" aux points de continuité de f. Enfin, pour une v.a. X à densité f_X, on peut reformuler (B.4) sous la forme:

$$g(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P}) \quad \underline{\text{ssi}} \quad \int_{\mathbb{R}} |g(x)| f_X(x)dx < \infty$$

et

$$\mathbb{E}[g(X)] \quad = \quad \int_{\mathbb{R}} g(x) f_X(x)dx.$$

B.2.2 Inégalités de Jensen

Une fonction convexe $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ est au dessus de son hyperplan tangeant en tout point de l'intérieur du domaine. Si on admet ce résultat, alors, on peut écrire pour une v.a. intégrable X que

$$g(X) \quad \geq \quad g(\mathbb{E}[X]) + \langle p_{\mathbb{E}[X]}, X - \mathbb{E}[X] \rangle,$$

où $p_{\mathbb{E}[X]}$ est le gradient de g au point $\mathbb{E}[X]$, si g est dérivable en ce point. si g n'est pas dérivable ce résultat est encore valable en remplaçant le gradient par la notion de sous-gradient... Dans la démonstration qui va suivre, nous allons éviter de passer par cette notion d'analyse convexe, et utiliser un argument d'approximation. En prenant l'espérance dans la dernière inégalité, on obtient l'inégalité de Jensen:

Theorem B.6. *Soit $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ et $g : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{\infty\}$ une fonction convexe telle que $\mathbb{E}[|g(X)|] < \infty$. Alors $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$.*

Proof. Si g est dérivable sur l'intérieur du domaine, le résultat découle de la discussion qui précéde l'énoncé.

Dans le cas général, on commence par supposer que X est bornée, et on considère une approximation de g par une suite de fonctions $(g_n)_n$ telle que g_n est différentiable, convexe, $-\infty < -\sup_n \|g_n^-\|_\infty \leq g_n \leq g$ pour tout n, et $g_n \longrightarrow g$.[1] On écrit alors que $g_n(X)$ est au dessus de son hyperplan tangeant au point $\mathbb{E}[X]$, et on obtient en prenant l'espérance $\mathbb{E}[g_n(X)] \geq g_n(\mathbb{E}[X])$. Le théorème de convergence dominée permet de conclure.

Pour une variable aléatoire X intégrable, on applique le résultat précédent à $X_n := (-n) \vee X \wedge n$, et on passe à la limite par un argument de convergence dominée. ♢

[1] Un exemple d'une telle fonction est donné par l'inf-convolution $g_n(x) := \inf_{y \in \mathbb{R}^n} \left\{ f(y) + n|y-x|^2 \right\}$, voir Aubin [2].

B.2.3 Fonction caractéristique

Dans tout ce paragraphe X désigne un vecteur aléatoire sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $\mathbb{R}^n$.

Definition B.7. *On appelle fonction caractéristique de X la fonction $\Phi_X : \mathbb{R}^n \longrightarrow \mathbb{C}$ définie par*

$$\Phi_X(u) := \mathbb{E}\left[e^{i\langle u,X\rangle}\right] \quad \textit{pour tout} \quad u \in \mathbb{R}^n.$$

La fonction caractéristique dépend uniquement de la loi de X:

$$\Phi_X(u) \quad = \quad \int_{\mathbb{R}^n} e^{i\langle u,x\rangle} d\mathcal{L}_X(x),$$

est n'est rien d'autre que la transformée de Fourier de $\mathbb{P}_X$ au point $-u/2\pi$. L'intégrale de Lebesgue d'une fonction à valeurs complexes est définie de manière naturelle en séparant partie réelle et partie imaginaire. La fonction caractéristique est bien définie pour tout $u \in \mathbb{R}$ comme intégrale d'une fonction de module 1. Enfin, pour deux v.a. X et Y, on a

$$\Phi_X(u) = \overline{\Phi_{-X}(u)} \text{ et } \Phi_{aX+b}(u) = e^{ib}\Phi_X(au) \quad \text{pour tous} \quad u \in \mathbb{R}^n, a, b \in \mathbb{R}.$$

Les propriés suivantes des fonctions caractéristiques peuvent être démontrés facilement grâce au théorème de convergence dominée.

Lemma B.8. *Soit Φ_X la fonction caractéristique d'une v.a. X. Alors $\Phi_X(0) = 1$, et Φ_X est continue bornée (par 1) sur $\mathbb{R}^n$.*

Proof. $\Phi_X(0) = 1$ et $|\Phi_X| \leq 1$ sont des propriétés évidentes, la continuité est une conséquence immédiate du théorème de convergence dominée. ♢

Exercise B.9. *1. Pour un vecteur gaussien X de moyenne b et de matrice de variance V, montrer que*

$$\Phi_X(u) = e^{\langle u,b\rangle - \frac{1}{2}\langle u,Vu\rangle}.$$

(Il s'agit d'une formule utile à retenir.)

2. Si $\mathcal{L}_X$ est symétrique par rapport à l'origine, i.e. $\mathcal{L}_X = \mathcal{L}_{-X}$, montrer que Φ_X est à valeurs réelles.

3. Pour une v.a. réelle, supposons que $\mathbb{E}[|X|^p] < \infty$ pour un certain entier $p \geq 1$. Montrer que Φ_X est p fois dérivable et

$$\Phi_X^{(k)}(0) = i^k \mathbb{E}[X^k] \quad \textit{pour} \quad k = 1, \ldots, p.$$

Le but de ce paragraphe est de montrer que la fonction caractéristique permet, comme son nom l'indique, de caractériser la loi $\mathcal{L}_X$ de X. Ceci donne un moyen alternatif d'aborder les vecteurs aléatoires pour lesquels la fonction de répartition est difficile à manipuler. Cependant, l'intérêt de cette notion ne se limite pas à la dimension $n = 1$. Par exemple, la manipulation de sommes de v.a. est souvent plus simple par le biais des fonctions caractéristiques.

Dans ces notes, nous nous limitons à montrer ce résultat dans le cas unidimensionnel.

Theorem B.10. *Pour une v.a. rélle, la fonction Φ_X caractérise la loi $\mathcal{L}_X$. Plus précisément*

$$\frac{1}{2}\mathcal{L}_X(\{a\}) + \frac{1}{2}\mathcal{L}_X(\{b\}) + \mathcal{L}_X(]a,b[) = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\Phi_X(u)\frac{e^{-iua}-e^{-iub}}{iu}du$$

pour tous $a < b$. De plus, si Φ_X est inégrable, $\mathcal{L}_X$ est absoluement continue par rapport à la mesure de Lebesgue, de densité

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux}\Phi_X(u)du, \qquad x \in \mathbb{R}.$$

Proof. Nous nous limitons au cas unidimensionnel $n = 1$ pour simplifier la présentation. Pour $a < b$, on vérifie sans peine que la condition d'application du théorème de Fubini est satisfaite, et on calcule que:

$$\begin{aligned}
\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-iua}-e^{-iub}}{iu}\Phi_X(u)du &= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-iua}-e^{-iub}}{iu}\left(\int_{\mathbb{R}} e^{iuv}d\mathcal{L}_X(v)dv\right)du \\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\left(\int_{-T}^{T}\frac{e^{iu(v-a)}-e^{iu(v-b)}}{iu}du\right)d\mathcal{L}_X(v).
\end{aligned}$$

Puis, on calcule directement que

$$\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{iu(v-a)}-e^{iu(v-b)}}{iu}du = \frac{S((v-a)T)-S((v-b)T)}{\pi T}, \qquad \text{(B.5)}$$

où $S(x) := \mathrm{sgn}(x)\int_0^{|x|}\frac{\sin t}{t}dt$, $t > 0$, et $\mathrm{sgn}(x) = \mathbf{1}_{\{x>0\}} - \mathbf{1}_{\{x<0\}}$. On peut vérifier que $\lim_{x\to\infty} S(x) = \frac{\pi}{2}$, que l'expression (B.5) est uniformément bornée en x et T, et qu'elle converge vers

$$0 \text{ si } x \notin [a,b], \quad \tfrac{1}{2} \text{ si } x \in \{a,b\}, \quad \text{et } 1 \text{ si } x \notin]a,b[.$$

On obtient alors le résultat annoncé par le théorème de convergence dominée.

Supposons de plus que $\int_{\mathbb{R}}|\phi_X(u)|du < \infty$. Alors, en prenant la limite $T \to \infty$ dans l'expression du théorème, et en supposant dans un premier temps que $\mathcal{L}_X$ n'a pas d'atomes, on obtient:

$$\mathcal{L}_X(]a,b] = F_X(b) - F_X(a) = \frac{1}{2}\int_{\mathbb{R}}\frac{e^{-iua}-e^{-iub}}{iu}du$$

par le théorème de convergence dominée. On réalise alors que le membre de droite est continu en a et b et, par suite, $\mathcal{L}_X$ n'a pas d'atomes et l'expression ci-dessus est vraie. Pour trouver l'expression de la densité f_X, il suffit de prendre la limite $b \to a$ après normalisation par $b - a$, et d'utiliser le théorème de convergence dominée. $\diamondsuit$

B.3 Espaces $\mathcal{L}^p$ et convergences fonctionnelles des variables aléatoires

B.3.1 Géométrie de l'espace $\mathcal{L}^2$

On désigne par $\mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ l'espace vectoriel des variables aléatoires rélles de carré $\mathbb{P}-$intégrable. Une application simple de l'inégalité de Jensen montre montre que $\mathcal{L}^2 \subset \mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$.

L'application $(X, Y) \longmapsto \mathbb{E}[XY]$ définit un produit scalaire sur $\mathcal{L}^2$ si on identifie les v.a.égales p.s. On note la norme correspondantes par $\|X\|_2 := \mathbb{E}[X^2]^{1/2}$. En particulier, ceci garantit l'inégalité de Schwarz (valable pour les mesures, voir exercice A.37):

$$|\mathbb{E}[XY]| \le \mathbb{E}[|XY|] \le \|X\|_2 \|Y\|_2 \quad \text{pour tous} \quad X, Y \in \mathcal{L}^2,$$

ainsi que l'inégalité triangulaire

$$\|X + Y\|_2 \le \|X\|_2 + \|Y\|_2 \quad \text{pour tous} \quad X, Y \in \mathcal{L}^2.$$

(On peut vérifier que les preuves de ces résultats ne sont pas perturbées par le problème d'identification des v.a. égales p.s.)

En probabilité, l'espérance quantifie la moyenne de la v.a. Il est aussi important, au moins intuitivement, d'avoir une mesure de la dispersion de la loi. ceci est quantifié par la notion de variance et de covariance:

$$\mathbb{V}[X] \;:=\; \mathbb{E}[(X - \mathbb{E}X)^2] \;=\; \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

et

$$\mathbb{C}ov[X, Y] \;:=\; \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \;=\; \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Si X est à valeurs dans $\mathbb{R}^n$, ces notions sont étendus de manière naturelle. Dans ce cadre $\mathbb{V}[X]$ est une matrice symétrique positive de taille n.

Enfin, la corrélation entre les v.a. X et Y est définie par

$$\mathbb{C}or[X, Y] := \frac{\mathbb{C}ov[X, Y]}{\|X\|_2 \|Y\|_2} = \frac{\langle X, Y \rangle_2}{\|X\|_2 \|Y\|_2},$$

i.e. le cosinus de l'angle formé par les vecteurs X et Y. L'inégalité de Schwarz garantit que la corrélation est un réel dans l'intervalle $[-1, 1]$. Le théorème de Pythagore s'écrit

$$\mathbb{E}[(X + Y)^2] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] \quad \text{dès que} \quad \mathbb{E}[XY] = 0,$$

ou, en termes de variances,

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] \quad \text{dès que} \quad \mathbb{C}ov[X, Y] = 0.$$

Attention,la variance n'est pas un opérateur linéaire, la formule ci-dessus est uniquement valable si $\mathbb{C}ov[X, Y] = 0$. Enfin, la loi du parallélogramme s'écrit

$$\|X + Y\|_2^2 + \|X - Y\|_2^2 = 2\|X\|_2^2 + 2\|Y\|_2^2 \quad \text{pour tous} \quad X, Y \in \mathcal{L}^2.$$

B.3.2 Espaces $\mathcal{L}^p$ et $\mathbb{L}^p$

Pour $p \in [1, \infty[$, on note par $\mathcal{L}^p := \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ l'espace vectoriel des variables aléatoires X telles que $\mathbb{E}[|X|^p] < \infty$. On note $\|X\|_p := (\mathbb{E}[|X|^p])^{1/p}$. Remarquons que $\|X\|_p = 0$ implique seulement que $X = 0$ p.s. donc $\|.\|_p$ ne définit pas une norme sur $\mathcal{L}^p$.

Definition B.11. *L'espace $\mathbb{L}^p$ est l'ensemble des classes d'équivalence de $\mathcal{L}^p$ pour la relation définie par l'égalité p.s.*

Ainsi l'espace $\mathbb{L}^p$ identifie les variables aléatoires égales p.s. et $\|.\|$ définit bien une norme sur $\mathbb{L}^p$.

Nous continuerons tout de même à travailler sur l'espace $\mathcal{L}^p$ et nous ne passerons à $\mathbb{L}^p$ que si nécessaire.

Par une application directe de l'inégalité de Jensen, on voit que

$$\|X\|_p \leq \|X\|_r \text{ si } 1 \leq p \leq r < \infty \quad \text{pour tout} \quad X \in \mathcal{L}^r, \tag{B.6}$$

en particulier, $X \in \mathcal{L}^p$. Ceci montre que $\mathcal{L}^p \supset \mathcal{L}^r$ dès que $1 \leq p \leq r < \infty$.

Nous allons montrer que l'espace $\mathcal{L}^p$ peut être transformé (toujours par quotionnement par la classe des v.a. nulles p.s.) en un espace de Banach.

Theorem B.12. *Pour $p \geq 1$, l'espace $\mathbb{L}^p$ est un espace de Banach, et $\mathbb{L}^2$ est espace de Hilbert. Plus précisément, soit $(X_n)_n$ une suite de Cauchy dans $\mathcal{L}^p$, i.e. $\|X_n - X_m\|_p \longrightarrow 0$ pour $n, m \to \infty$. Alors il existe une v.a. $X \in \mathcal{L}^p$ telle que $\|X_n - X\|_p \longrightarrow 0$.*

Proof. Si $(X_n)_n$ est une suite de Cauchy, on peut trouver une suite croissante $(k_n)_n \subset \mathbb{N}$, $k_n \uparrow \infty$, telle que

$$\|X_m - X_n\|_p \leq 2^{-n} \quad \text{pour tous} \quad m, n \geq k_n. \tag{B.7}$$

Alors, on déduit de l'inégalité (B.6) que

$$\mathbb{E}[|X_{k_{n+1}} - X_{k_n}|] \quad \leq \quad \|X_{k_{n+1}} - X_{k_n}\|_p \ \leq \lambda 2^{-n},$$

et que $\mathbb{E}[\sum_n |X_{k_{n+1}} - X_{k_n}|] < \infty$. Alors la série $\sum_n (X_{k_{n+1}} - X_{k_n})$ est absolument convergente p.s. Comme il s'agit d'une série télescopique, ceci montre que

$$\lim_n X_{k_n} = X \text{ p.s.} \quad \text{où} \quad X := \limsup_n X_{k_n}.$$

Revenant à (B.7), on voit que pour $n \geq k_n$ et $m \geq n$, on a $\mathbb{E}|X_n - X_{k_m}|^p] = \|X_n - X_{k_m}\|_p^p \leq 2^{-np}$. Pour $m \to \infty$, on déduit du lemme de Fatou que $\mathbb{E}[|X_n - X|^p] \leq 2^{-np}$. ♢

B.3.3 Espaces $\mathcal{L}^0$ et $\mathbb{L}^0$

On note par $\mathcal{L}^0 := \mathcal{L}^0(\mathcal{A})$ l'espace vectoriel des variables aléatoires $\mathcal{A}-$mesurables sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et on introduit l'espace quotient $\mathbb{L}^0$ constitué des classes d'équivalence de $\mathcal{L}^0$ pour la relation définie par l'égalité p.s.

Definition B.13. *(Convergence en probabilité) Soient $(X_n)_n$ et X des v.a. dans $\mathcal{L}^0$. On dit que $(X_n)_n$ converge en probabilité vers X si*

$$\lim_{n\to\infty} \mathbb{P}\left[|X_n - X| \geq \varepsilon\right] = 0 \quad \textit{pour tout} \quad \varepsilon > 0.$$

Cette notion de convergence est plus faible que la convergence p.s. et que la convergence dans $\mathbb{L}^p$ dans le sens suivant.

Lemma B.14. (i) *La convergence p.s. implique la convergence en probabilité,* (ii) *Soit $p \geq 1$. La convergence en norme dans $\mathbb{L}^p$ implique la convergence en probabilité.*

Proof. (i) découle d'une application immédiate du théorème de la convergence dominée. Pour (ii), il suffit d'utiliser l'inégalité de Markov de l'exercice A.36. ♢

Le but de ce paragraphe est de montrer que la convergence en probabilité est métrisable et qu'elle confère à $\mathbb{L}^0$ une structure d'espace métrique complet. Pour cela, on introduit la fonction $D : \mathcal{L}^0 \times \mathcal{L}^0 \longrightarrow \mathbb{R}_+$ définit par:

$$D(X, Y) = \mathbb{E}[|X - Y| \wedge 1] \quad \text{pour tous} \quad X, Y \in \mathcal{L}^0. \tag{B.8}$$

On vérifie imédiatement que D est une distance sur $\mathbb{L}^0$, mais ne l'est pas sur $\mathcal{L}^0$, pour les mêmes raisons que celles du paragraphe précédent.

Lemma B.15. *La convergence en probabilité est équivalente à la convergence au sens de la distance D.*

Proof. Pour $X \in \mathcal{L}^0$, on obtient par l'inégalité de Markov de l'exercice A.36:

$$\mathbb{P}[|X| \geq \varepsilon] = \mathbb{P}[|X| \wedge 1 \geq \varepsilon] \leq \frac{\mathbb{E}[|X| \wedge 1]}{\varepsilon},$$

qui permet de déduire que au sens de D implique la convergence en probabilité. Pour l'implication inverse, on estime:

$$\mathbb{E}[|X| \wedge 1] = \mathbb{E}[(|X| \wedge 1)\mathbf{1}_{|X| \geq \varepsilon}] + \mathbb{E}[(|X| \wedge 1)\mathbf{1}_{|X| < \varepsilon}] \leq \mathbb{P}[|X| \geq \varepsilon] + \varepsilon,$$

d'où on tire que la convergence en probabilité implique la convergence au sens de D. ♢

Theorem B.16. $(\mathbb{L}^0, D)$ *est un espace métrique complet.*

Proof. Soit $(X_n)_n$ une suite de Cauchy pour D. Alors c'est une suite de Cauchy pour la convergence en probabilité d'après le lemme B.15, et on peut construire une suite $(n_k)_k \uparrow \infty$ telle que

$$\mathbb{P}\left[|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}\right] \leq 2^{-k} \quad \text{pour tout} \quad k \geq 1,$$

et par suite $\sum_k \mathbb{P}\left[|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}\right] < \infty$. Le premier lemme de Borel-Cantelli (lemme A.14) implique alors que $\mathbb{P}\left[\cup_n \cap_{m\geq n} \{|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}\}\right] = 1$ et, par suite, pour presque tout $\omega \in \Omega$, $\left(X_{n_k}(\omega)\right)_n$ est une suite de Cauchy dans $\mathbb{R}$. Ainsi, la v.a. $X := \limsup_n X_{k_n}$ vérifie $X_{n_k} \longrightarrow X$ p.s. donc en probabilité, et on termine comme dans la démonstration du théorème B.12. $\diamondsuit$

B.3.4 Lien entre les convergences $\mathbb{L}^p$, en proba et p.s.

Nous avons vu que la convergence en probabilité est plus faible que la convergence p.s. Le résultat suivant établit un lien précis entre ces deux notions de convergence.

Theorem B.17. *Soient $\{X_n, n \geq 1\}$ et X des v.a. dans $\mathcal{L}^0$.*
(i) $X_n \longrightarrow X$ *p.s.* *ssi* $\sup_{m\geq n} |X_m - X| \longrightarrow 0$ *en probabilité.*
(ii) $X_n \longrightarrow X$ *en probabilité* *ssi* *de toute suite croissante d'entiers $(n_k)_k$, on peut extraire une sous-suite $(n_{k_j})_j$ telle que $X_{n_{k_j}} \longrightarrow X$ p.s.*

La démonstration est reportée à la fin de ce paragraphe. On continue par une conséquence immédiate du téorème B.17 (ii).

Corollary B.18. *(Slutsky) Soient $(X_n)_n$ une suite à valeur dans $\mathbb{R}^d$, et $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}^d$ une fonction continue. Si $X_n \longrightarrow X$ en probabilité, alors $\phi(X_n) \longrightarrow \phi(X)$ en probabilité.*

Ceci est une conséquence immédiate du théorème B.17 (ii). En particulier, il montre que la convergence en probabilité est stable pour les opérations usuelles d'addition, de multiplication, de min, de max, etc...

Avant de démontrer le théorème B.17, énonçons le résultat établissant le lien précis entre la convergence en probabilité et la convergence dans $\mathbb{L}^1$.

Definition B.19. *Une famille $\mathcal{C}$ de v.a. est dite uniformément intégrable, et on note U.I. si*

$$\lim_{c\to\infty} \sup_{X\in\mathcal{C}} \mathbb{E}[|X|\mathbf{1}_{\{|X|\geq c\}} \quad = \quad 0.$$

Theorem B.20. *Soient $\{X_n, n \geq 1\}$ et X des v.a. dans $\mathcal{L}^1$. Alors $X_n \longrightarrow X$ dans $\mathbb{L}^1$ si et seulement si*
(a) $X_n \longrightarrow X$ *en probabilité,*
(b) $(X_n)_n$ *est U.I.*

La démonstration de ce résultat est reportée à la fin de ce paragraphe. L'exercice suivant regroupe les résultats essentiels qui concernent l'uniforme intégrabilité.

Exercise B.21. *Soit $(X_n)_n$ une suite de v.a. à valeurs réelles.*

1. *Supposons que $(X_n)_n$ est U.I.*
   (a) *Montrer que $(X_n)_n$ est bornée dans $\mathcal{L}^1$, i.e.* $\sup_n \mathbb{E}[|X_n|] < \infty$.
   (b) *Sur l'espace probabilisé $([0,1], \mathcal{B}_{[0,1]}, \lambda)$, λ étant la mesure de Lebesgue, on considère la suite $Y_n := n\mathbf{1}_{[0,1/n]}$. Montrer que $(Y_n)_n$ est bornée dans $\mathcal{L}^1$, mais n'est pas U.I.*
2. *Supposons que $\mathbb{E}[\sup_n |X_n|] < \infty$. Montrer que (X_n) est U.I. (Indication: utiliser la croissance de la fonction $x \longmapsto x\mathbf{1}_{\{x \geq c\}}$ $\mathbb{R}_+$).*
3. *Supposons qu'il existe $p > 1$ tel que $(X_n)_n$ est bornée dans $\mathcal{L}^p$.*
   (a) *Montrer que $\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n| \geq c\}}] \leq \|X_n\|_p \mathbb{P}[|X_n \geq c]^{1-1/p}$*
   (b) *En déduire que (X_n) est U.I.*

Nous allons maintenons passer aux démonstrations des théorèmes de ce paragraphe.

Preuve du théorème B.17 (i) Remarquons que

$$C := \{X_n \longrightarrow X\} \quad = \quad \cap_k \cup_n \cap_{m \geq n} \{|X_m - X| \leq k^{-1}\} \; = \; \lim \downarrow_k \cup_n A_n$$

où $A_n := \cap_{m \geq n}\{|X_m - X| \leq k^{-1}\}$. La convergence p.s. de X_n vers X s'écrit $\mathbb{P}[C] = 1$, et est équivalente à $\mathbb{P}[\cup_n A_n] = 1$ pour tout $k \geq 1$. Comme la suite $(A_n)_n$ est croissante, ceci est équivalent à $\lim \uparrow_n \mathbb{P}[A_n] = 1$ pour tout $k \geq 1$, ce qui exprime exactement la convergence en probabilité de $\sup_{m \geq n} |X_m - X|$ vers 0.
(ii) Supposons d'abord que $X_n \longrightarrow X$ en probabilité. Soit (n_k) une suite croissante d'indices, et $\bar{X}_k := X_{n_k}$. On définit

$$k_j \quad := \quad \inf\left\{i : \mathbb{P}[|\bar{X}_i - X| \geq 2^{-j}] \leq 2^{-j}\right\}.$$

Alors, $\sum_j \mathbb{P}[|\bar{X}_{k_j} - X| \geq 2^{-j}] < \infty$, et on déduit du premier lemme de Borel Cantelli, lemme A.14 que $|\bar{X}_{k_j} - X| < 2^{-j}$ pour j assez grand, p.s. En particulier, ceci montre que $\bar{X}_{k_j} \longrightarrow X$, p.s.

Pour la condition suffisante, supposons au contraire que $X_n \not\longrightarrow X$ en probabilité. Alors, d'après le lemme B.15, il existe une sous-suite (n_k) croissante et $\varepsilon > 0$ tels que $D(X_{n_k}, X) \geq \varepsilon$. On arrive à une contradiction en extrayant une sous-suite $(X_{n_{k_j}})_j$ qui converge p.s. vers X, et en évoquant le théorème de convergence dominée pour le passage à la limite. $\diamondsuit$

Preuve du théorème B.20 Supposons d'abord que les conditions (a) et (b) sont satisfaites. La fonction $\varphi_c(x) := -c \vee x \wedge c$, $x \in \mathbb{R}$ est lipschitzienne, et vérifie $|\varphi_c(x) - x| \leq |x|\mathbf{1}_{|x| \geq c}$. On déduit alors
- de l'U.I. de $(X_n)_n$ et l'intégrabilité de X que, quand $c \to \infty$:

$$\mathbb{E}[|\varphi_c(X_n) - X_n|] \longrightarrow 0 \quad \text{pour tout } n \text{ et} \quad \mathbb{E}[|\varphi_c(X) - X|] \longrightarrow 0,$$

- et de la convergence en probabilité de X_n vers X, et du corollaire B.18, que

$$\varphi_c(X_n) \longrightarrow \varphi_c(X) \quad \text{en probabilité.}$$

On peut maintenant conclure que $X_n \longrightarrow X$ dans $\mathcal{L}^1$ en décomposant

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n - \varphi_c(X_n)|] + \mathbb{E}[|\varphi_c(X_n) - \varphi_c(X)|] + \mathbb{E}[|\varphi_c(X) - X|.$$

Réciproquement, supposons que $X_n \longrightarrow X$ dans $\mathcal{L}^1$, alors la convergence en probabilité (a) est une conséquence immédiate de l'inégalité de Markov (A.7) (exercice A.36). Pour montrer (b), on se donne $\varepsilon > 0$. La convergence $\mathcal{L}^1$ de $(X_n)_n$ montre l'existence d'un rang N à partir duquel

$$\mathbb{E}|X_n - X| < \varepsilon \quad \text{pour tout} \quad n > N. \tag{B.9}$$

Par ailleurs, d'après le lemme A.26, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{A}$:

$$\sup_{n \leq N} \mathbb{E}[|X_n|\mathbf{1}_A] < \varepsilon \text{ et } \mathbb{E}[|X|\mathbf{1}_A] < \varepsilon \quad \text{dès que} \quad \mathbb{P}[A] < \delta. \tag{B.10}$$

Nous allons utiliser cette inégalité avec les ensembles $A_n := \{|X_n| > c\}$ qui vérifient bien

$$\sup_n \mathbb{P}[A_n] \leq c^{-1} \sup_n \mathbb{E}[|X_n|] < \delta \quad \text{pour} \quad c \text{ assez grand,} \tag{B.11}$$

où nous avons utilisé l'inégalité de Markov (A.7) (exercice A.36), ainsi que la bornitude dans $\mathcal{L}^1$ de la suite $(X_n)_n$ du fait de sa convergence dans $\mathcal{L}^1$. Ainsi, on déduit de (B.10) et (B.11) que

$$\begin{aligned}
\sup_n \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] &= \max\left\{\sup_{n\leq N} \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] \ , \ \sup_{n>N} \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}]\right\} \\
&\leq \max\left\{\varepsilon \ , \ \sup_{n>N} \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}]\right\} \\
&\leq \max\left\{\varepsilon \ , \ \sup_{n>N} \mathbb{E}[|X|\mathbf{1}_{\{|X_n|>c\}} + \mathbb{E}[|X - X_n|\mathbf{1}_{\{|X_n|>c\}}]\right\} \\
&\leq \max\left\{\varepsilon \ , \ \sup_{n>N} \mathbb{E}[|X|\mathbf{1}_{A_n} + \mathbb{E}[|X - X_n|]\right\} \ < \ 2\varepsilon,
\end{aligned}$$

où la dernière inégalité est due à (B.9), (B.10) et (B.11). $\diamondsuit$

B.4 Convergence en loi

Dans ce paragraphe, nous nous intéressons à la convergence des loi. Remarquons immédiatement qu'il ne peut s'agir que d'un sens de convergence plus faible que les convergences fonctionnelles étudiées dans le paragraphe précédent puis qu'on ne pourra en général rien dire sur les variables aléatoires sous-jacentes. A titre d'exemple, si X est une v.a. de loi gaussienne centrée, alors $-X$ a la même loi que X (on écrit $X \stackrel{L}{=} -X$). Pire encore, on peut avoir deux v.a. réelles X et Y sur des espaces probabilisés différents $(\Omega_2 \mathcal{A}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ qui ont la même distribution.

Dans ce paragraphe, on désignera par $C_b(\mathbb{R})$ l'ensemble des fonctions continues bornées sur $\mathbb{R}$ et $\Sigma(\mathbb{R})$ l'ensemble des mesures de probabilité sur $\mathbb{R}$.

B.4.1 Définitions

Soient μ et $\mu_n, n \in \mathbb{N} \in \Sigma(\mathbb{R})$. On dit que $(\mu_n)_n$ converge *faiblement*, ou *étroitement*, vers μ si: $\mu_n(f) \longrightarrow \mu(f)$ pour toute fonction $f \in C_b(\mathbb{R})$.

Soient X et $X_n, n \in \mathbb{N}$ des v.a. dans $\mathcal{L}^0(\mathcal{A}, \mathbb{P})$. On dit que $(X_n)_n$ converge *en loi* vers X si $(\mathcal{L}_{X_n})_n$ converge faiblement vers $\mathcal{L}_X$, i.e.

$$\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)] \quad \text{pour tout} \quad f \in C_b(\mathbb{R}).$$

Dans la dernière définition, il n'est pas nécessaire que les v.a. $X, X_n, n \in \mathbb{N}$ soient définies sur le même espace probabilisé. Montrons maintenant que les convergences intrduites dans les chapitres précédents sont plus fortes que la convergence en loi.

Proposition B.22. *La convergence en probabilité implique la convergence en loi.*

Proof. Supposons que $X_n \longrightarrow X$ en probabilité, et soient $g \in C_b(\mathbb{R})$. La suite réelles $u_n := \mathbb{E}[g(X_n)], n \in \mathbb{N}$, est bornée. Pour montrer la convergence en loi, il suffit de vérifier que toute sous-suite convergente $(u_{n_k})_k$ converge vers $\mathbb{E}[g(X)]$. Pour celà, il suffit d'utiliser le lemme B.17 et le théorème de convergence dominée. $\diamondsuit$

Comme la convergence en probabilité est plus faible que la convergence $\mathbb{L}^1$ et la convergence p.s. on le schémas suivant expliquant les liens entre les différents types de convergence rencontrés:

$$\begin{array}{ccccc} \mathbb{L}^p & \Longrightarrow & \mathbb{L}^1 & & \\ & & \Downarrow & & \\ \text{p.s.} & \Longrightarrow & \mathbb{P} & \Longrightarrow & \text{Loi} \end{array}$$

B.4.2 Caractérisation de la convergence en loi par les fonctions de répartition

Toute loi $\mu \in \Sigma(\mathbb{R})$ est caractérisée par la fonction de répartition correspondante $F(x) := \int x d\mu(x)$. Ainsi, si $F, F_n, n \in \mathbb{N}$ des fonctions de répartition sur $\mathbb{R}$, on dira que $(F_n)_n$ converge en loi vers F si la convergence en loi a lieu pour les mesures correspondantes.

Dans ce paragraphe, nous allons exprimer la définition de la convergence faible de manière équivalente en terme des fonctions de répartition.

Remark B.23. Les points de discontinuité de F, s'il y en a, jouent un role particulier: Sur $([0,1], \mathcal{B}_{[0,1]}, \lambda)$, soit $\mu_n := \delta_{1/n}$ la masse de Dirac au point $1/n$ (c'est la loi de la v.a. déterministe $X_n = 1/n$). Alors (μ_n) converge en loi vers δ_0, la masse de Dirac au point 0. Maisn pour tout $n \geq 1$, $F_n(0) = 0 \not\longrightarrow F_{\delta_0}(0)$.

Theorem B.24. *Soient $F, F_n, n \in \mathbb{N}$ des fonctions de répartition sur $\mathbb{R}$. Alors, (F_n) converge en loi vers F si et seulement si*

$$\textit{Pour tout} \quad x \in \mathbb{R}, \quad F(x-) = F(x) \implies F_n(x) \longrightarrow F(x).$$

Proof. 1- Pour $\eta > 0$ et $x \in \mathbb{R}$, on définit les fonctions

$$g_1(y) := \mathbf{1}]-\infty, x+\eta] - \frac{y-x}{\eta}\mathbf{1}]x, x+\eta] \quad \text{et} \quad g_2(y) := g_1(y+\eta),\ y \in \mathbb{R},$$

et on observe que $\mathbf{1}_{]-\infty,x]} \leq g_1 \leq \mathbf{1}_{]-\infty,x+\eta]}$, $\mathbf{1}_{]-\infty,x-\eta]} \leq g_2 \leq \mathbf{1}_{]-\infty,x]}$ et, par suite

$$F_n(x) \leq \mu_n(g_1),\ \mu(g_1) \leq F(x+\eta), \quad \text{et} \quad F_n(x) \geq \mu_n(g_2),\ \mu(g_2) \leq F(x-\eta)$$

Comme $g_1, g_2 \in C_b(\mathbb{R})$, on déduit de a convergence faible de $(F_n)_n$ vers F que $\mu_n(g_1) \longrightarrow \mu(g_1)$, $\mu_n(g_2) \longrightarrow \mu(g_2)$, et

$$F(x-\eta) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x+\delta) \quad \text{pour tout} \quad \eta > 0,$$

qui implique bien que $F_n(x) \longrightarrow F(x)$ si x est un point de continuité de F.
2- Pour la condition suffisante, on définit comme dans la remarque B.4 les v.a. $\overline{X}, \underline{X}, \overline{X}_n, \underline{X}_n$ qui ont pour fonction de répartition F et F_n. Par définition de $\overline{X}$, pour tout $x > \overline{X}(\omega)$ on a $F(x) > \omega$. Si x est un point de continuité de F, ceci implique que $F_n(x) > \omega$ pour n assez grand et, par suite, $x \geq \overline{X}_n(\omega)$. Comme F est croissante, l'ensemble de ses points de discontinuité est au plus dénombrable. On peut donc faire tendre x vers $\overline{X}(\omega)$ lelong de points de continuité de F, et on tire l'inégalité $\overline{X}(\omega) \geq \overline{X}_n(\omega)$ pour n assez grand. On obtient le résultat symétrique en raisonnant sur $\underline{X}$ et $\underline{X}_n$. D'où:

$$\overline{X}(\omega) \geq \overline{X}_n(\omega) \geq \underline{X}_n(\omega) \geq \underline{X} \quad \text{pour} \quad n \text{ assez grand.}$$

Comme $\mathbb{P}[\overline{X} = \underline{X}] = 1$, ceci montrer que $\overline{X}_n \longrightarrow \overline{X}$ p.s. et donc en loi. ♢

B.4.3 Convergence des fonctions de répartition

L'importance de la convergence en loi provient de la facilité d'obtenir des théorèmes limites. En effet, les suites de mesures convergent en loi "à peu de frais"', lelong d'une sous-suite, vers une limite qui n'est cependant pas nécessairement une loi. Si la limite n'est pas une loi, on dit qu'il y a perte de masse.

Avant d'énoncer un résultat précis, expliquons les idées qu'il y a derrière ces résultats profonds. Les fonctions de répartition ont une structure très spéciale: on regardant le graphe d'une fonction de répartition dans les coordonnée $(x +$

$y, -x + y)$ (obtenu par rotation des coordonnées initiale de 45°), le graphe devient celui d'une fonction dont la valeur absolue de la pente est majorée par 1: les pentes -1 et 1 correspondent respectivement aux "plats" et aux sauts de la fonction de répartition. Ainsi dans ce système de coordonnées le graphe perd la proprié de croissance, mais devient $1-$Lipschitzien. Le théorème d'Ascoli nous garantit alors l'existence d'une sous-suite convergente. La démonstration ci-dessous utilise un argument encore plus élémentaire.

Lemma B.25. *Soit $(F_n)_n$ une suite de fonctions de répartition sur $\mathbb{R}$. Alors, il existe une fonction croissante continue à droite $F : \mathbb{R} \longrightarrow [0, 1]$, et une sous-suite (n_k) telles que $F_{n_k} \longrightarrow F$ simplement en tout point de continuité de F.*

Proof. On dénombre les éléments de l'ensemble des rationnels $\mathbb{Q} = \{q_i, i \in \mathbb{N}\}$. La suite $(F_n(q_1))_n$ est bornée, donc converge le long d'une sous-suite $F_{n_k^1}(q_1) \longrightarrow G(q_1)$ quand $k \to \infty$. De même la suite $\left(F_{n_k^1}(q_2)\right)_n$ est bornée, donc converge le long d'une sous-suite $F_{n_k^2}(q_2) \longrightarrow G(q_2)$ quand $k \to \infty$, etc... Alors, en posant $k_j := n_j^j$, on obtient

$$F_{k_j}(q) \longrightarrow G(q) \quad \text{pour tout} \quad q \in \mathbb{Q}.$$

Il est clair que G est croissante sur $\mathbb{Q}$ et à valeurs dans $[0, 1]$. On définit alors la fonction F par

$$F(x) := \lim_{\mathbb{Q} \ni q \searrow x} G(q) \quad \text{pour tout} \quad x \in \mathbb{R},$$

qui vérifie les propriétés annoncées dans le lemme. ♢

Afin d'éviter la perte de masse à la limite, on introduit une nouvelle notion.

Definition B.26. *Une suite $(F_n)_{n\geq 1}$ de fonctions de répartition sur $\mathbb{R}$ est dite tendue si pour tout $\varepsilon > 0$, il existe $K > 0$ tel que*

$$\mu_n([-K, K]) := F_n(K) - F_n(-K) > 1 - \varepsilon \quad \textit{pour tout} \quad n \geq 1.$$

Le résultat suivant est une conséquence directe du lemme précédent.

Lemma B.27. *Soit $(F_n)_n$ une suite de fonctions de répartition sur $\mathbb{R}$.*
(i) *Si $F_n \longrightarrow F$ en loi, alors $(F_n)_n$ est tendue.*
(ii) *Si $(F_n)_n$ est tendue, alors il existe une fonction de répartition F sur $\mathbb{R}$, et une sous-suite (n_k) telles que $F_{n_k} \longrightarrow F$ en loi.*

B.4.4 Convergence en loi et fonctions caractéristiques

La fonction caractéristique caractérise une loi de distribution tout aussi bien que la fonction de répartition. Le résultat suivant donne la caractérisation de la convergence en loi en termes de fonctions caractéristiques.

Theorem B.28. *(convergence de Lévy) Soit $(F_n)_n$ une suite de fonctions de répartitions sur $\mathbb{R}$, et $(\phi_n)_n$ la suite de fonctions caractéristiques correspondantes. Supposons qu'il existe une fonction ϕ sur $\mathbb{R}$ telle que*

$$\phi_n \longrightarrow \phi \quad \textit{simplement sur } \mathbb{R} \textit{ et } \quad \phi \textit{ continue en } 0.$$

Alors ϕ est une fonction caractéristique correspondant à une fonction de répartition F, et $F_n \longrightarrow F$ en loi.

Proof. 1- Montrons d'abord que

$$(F_n)_n \quad \text{est tendue.} \tag{B.12}$$

Soit $\varepsilon > 0$. D'après la continuité de ϕ en 0, il existe $\alpha > 0$ tel que $|1-\phi| < \varepsilon$ sur $[-\alpha, \alpha]$. Il est clair que $2 - \phi_n(u) - \phi_n(-u) \in \mathbb{R}_+$ et que cette propriété est héritée par ϕ à la limite. Alors $0 \leq \int_0^\alpha [2 - \phi(u) - \phi(-u)]du \leq 2\varepsilon\alpha$, et on déduit de la convergence de ϕ_n vers ϕ et du théorème de convergence dominée qu'à partir d'un certain rang $n \geq N$:

$$\begin{aligned}
4\varepsilon &\geq \frac{1}{\alpha}\int_0^\alpha [2 - \phi_n(u) - \phi_n(-u)]du \\
&= \frac{1}{\alpha}\int_{-\alpha}^\alpha \int_{\mathbb{R}} \left(1 - e^{iu\omega}\right) dF_n(\omega)du \\
&= \frac{1}{\alpha}\int_{\mathbb{R}} \int_{-\alpha}^\alpha \left(1 - e^{iu\omega}\right) du dF_n(\omega) \ = \ 2\int_{\mathbb{R}} \left(1 - \frac{\sin(\alpha\omega)}{\alpha\omega}\right) dF_n(\omega)
\end{aligned}$$

par le théorème de Fubini. Comme $\sin x \leq x$ pour tout $x \in \mathbb{R}$, on déduit alors que pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que:

$$4\varepsilon \geq 2\int_{|\omega| \geq 2\alpha^{-1}} \left(1 - \frac{\sin(\alpha\omega)}{\alpha\omega}\right) dF_n(\omega) \geq \int_{|\omega| \geq 2\alpha^{-1}} dF_n(\omega),$$

prouvant (B.12).
2- Comme $(F_n)_n$ est tendue, on déduit du lemme B.27 que $F_{n_k} \longrightarrow F$ en loi lelong d'une sous-suite $(n_k)_k$, où F est une foncion de répartition. D'après la définition de la convergence en loi, on a aussi convergence des fonctions caractéristiques correspondantes $\phi_{n_k} \longrightarrow \Phi_F$. Alors $\phi = \Phi_F$.
3- Il reste à monter que $F_n \longrightarrow F$ en loi. Supposons au contraire qu'il existe un point de continuité x tel que $F_n(x) \not\longrightarrow F(x)$. Alors, il existe une sous-suite $(n_k)_k$ telle que

$$F(x-) = F(x) \text{ et } |F_{n_k}(x) - F(x)| \geq \varepsilon \quad \text{pour tout} \quad k. \tag{B.13}$$

Comme $(F_{n_k})_k$ est tendue d'après l'étape 1, on a $F_{n_{k_j}} \longrightarrow \tilde{F}$ en loi lelong d'une sous-suite $(n_{k_j})_j$, où $\tilde{F}$ est une foncion de répartition. Raisonnant comme dans l'étape précédente, on voit que $\phi_{n_{k_j}} \longrightarrow \Phi_{\tilde{F}} = \phi = \Phi_F$, et on déduit que $\tilde{F} = F$ par injectivité. Ainsi $F_{n_{k_j}} \longrightarrow F$ en loi, contredisant (B.13). ♢

B.5 Indépendance

B.5.1 $\sigma-$algèbres indépendantes

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $(\mathcal{A}_n)_n \subset \mathcal{A}$ une suite de $\sigma-$algèbres. On dit que les $(\mathcal{A}_n)_n$ sont indépendantes (sous $\mathbb{P}$) si pour tous entiers $n \geq 1$ et $1 \leq i_1 < \ldots < i_n$:

$$\mathbb{P}\left[\cap_{k=1}^n A_{i_k}\right] = \prod_{k=1}^n \mathbb{P}\left[A_{i_k}\right] \quad \text{pour tous} \quad A_{i_k} \in \mathcal{A}_{i_k},\ 1 \leq k \leq n. \tag{B.14}$$

Remarquons que le théorème de convergence monotone permet d'affirmer que (B.14) est aussi valide pour $n = \infty$, i.e.

$$\mathbb{P}\left[\cap_{k\geq 1} A_{i_k}\right] = \prod_{k\geq 1} \mathbb{P}\left[A_{i_k}\right] \quad \text{pour tous} \quad A_{i_k} \in \mathcal{A}_{i_k},\ k \geq 1. \tag{B.15}$$

A partir de cette définition générale pour les $\sigma-$algèbres, on étend l'indépendance à des sous-familles arbitraires de $\mathcal{A}$ et aux v.a.

Definition B.29. *On dit que les événements $(A_n)_n \subset \mathcal{A}$ sont indépendants si $(\sigma(A_n))_n$ sont indépendantes ou, de manière équivalente, si les v.a. $(\mathbf{1}_{X_n})_n$ sont indépendantes.*

Dans la partie (ii) de la définition précédentes, il est inutile de vérifier (B.14) pour tous les choix possibles dans les $\sigma-$algèbres $\sigma(A_n) = \{\Omega, \emptyset, A_n, An^c\}$. En effet, on peut facilement montrer qu'il suffit de vérifier que

$$\mathbb{P}\left[\cap_{k=1}^n A_{i_k}\right] = \prod_{k=1}^n \mathbb{P}\left[A_{i_k}\right] \quad \text{pour} \quad n \geq 1 \text{ et } 1 \leq i_1 < \ldots < i_n.$$

Voici une formulation plus générale de ce résultat.

Lemma B.30. *Soit $(\mathcal{I}_n)_n \subset \mathcal{A}$ une suite de $\pi-$systèmes. Alors les sous-$\sigma-$algèbres $(\sigma(\mathcal{I}_n))_n$ sont indépendantes si et seulement si* (B.14) *est vraie pour les événements des $\mathcal{I}_n$, i.e. si pour tous entiers $n \geq 1$ et $1 \leq i_1 < \ldots < i_n$, on a:*

$$\mathbb{P}\left[\cap_{k=1}^n I_{i_k}\right] = \prod_{k=1}^n \mathbb{P}\left[I_{i_k}\right] \quad \textit{pour tous} \quad I_{i_k} \in \mathcal{I}_{i_k},\ 1 \leq k \leq n.$$

Proof. il suffit de vérifier le résultat pour deux $\pi-$systèmes $\mathcal{I}_1, \mathcal{I}_2$. Fixons un événement $I_1 \in \mathcal{I}_1$, et introduisons les applications de $\sigma(\mathcal{I}_2)$ dans $[0, \mathbb{P}[I_1]]$ définies par $\mu(I_2) := \mu(I_1 \cap I_2)$ et $\nu(I_2) := \nu(I_1)\nu(I_2)$. Il est clair que μ et ν sont des mesures sur $\sigma(\mathcal{I}_2)$ égales sur le $\pi-$système $\mathcal{I}_2$. Alors elles sont égales sur $\sigma(\mathcal{I}_2)$ d'après la proposition A.5. Il suffit maintenant d'évoquer le rôle arbitraire de $I_1 \in \mathcal{I}_1$, et de répéter exactement le même argument en inversant $\mathcal{I}_1$ et $\mathcal{I}_2$.
$\diamondsuit$

B.5.2 variables aléatoires indépendantes

Definition B.31. *On dit que des v.a.* $(X_n)_n$ *sont indépendantes si les sous-*$\sigma-$*algèbres correspondantes* $\left(\sigma(X_n)\right)_n$ *sont indépendantes.*

Une application directe du lemme B.30 et du théorème de Fubini permet d'établir le critère suivant d'indépendance de v.a.

Proposition B.32. *Les v.a.* $(X_n)_n$ *sont indépendantes si et seulement si pour tous* $n \geq 1$ *et* $1 \leq i_1 < \ldots < i_n$*, l'une des assertions suivantes est vérifiée:*
(a) $\mathbb{P}\left[X_{i_k} \leq x_k \text{ pour } 1 \leq k \leq n\right] = \prod_{k=1}^n \mathbb{P}\left[X_{i_k} \leq x_k\right]$ *pour tous* $x_1, \ldots, x_k \in \mathbb{R}$*,*
(b) $\mathbb{E}\left[\prod_{k=1}^n f_{i_k}(X_{i_k})\right] = \prod_{k=1}^n \mathbb{E}\left[f_{i_k}(X_{i_k})\right]$ *pour toutes* $f_{i_k} : \mathbb{R} \longrightarrow \mathbb{R}$*,* $1 \leq k \leq n$*, mesurables bornées.*
(c) $\mathcal{L}_{(X_{i_1},\ldots,X_{i_n})} = \mathcal{L}_{X_{i_1}} \otimes \ldots \otimes \mathcal{L}_{X_{i_n}}$

Exercise B.33. *Montrer la proposition B.32.*

Remark B.34. Si X, Y sont deux v.a. rélles indépendantes, la proposition B.32 implique que la fonction caractéristique du couple se factorise:

$$\Phi_{(X,Y)}(u,v) = \Phi_X(u)\Phi_Y(v) \quad \text{pour tous} \quad u, v \in \mathbb{R}.$$

Remark B.35. Soient X, Y deux v.a. rélles indépendantes intégrables, alors d'après la proposition B.32, on a

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \ \mathbb{C}ov[X,Y] = 0 \quad \text{et} \quad \mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y].$$

Observons que *la nullité de la covariance n'implique pas l'indépendance*, en général. Dans le cas très particulier où le couple (X, Y) est un vecteur gaussien, on a cependant équivalence entre l'indépendance et la nullité de la covariance.

Si les $(X_n)_n$ sont des v.a. indépendantes à densité, alors on déduit de l'assertion (a) ci-dessus que le vecteur aléatoire $(X_{i_1}, \ldots, X_{i_n})$ est absoluement continu par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ de densité

$$f_{(X_{i_1},\ldots,X_{i_n})}(x_1, \ldots, x_n) \quad := \quad f_{X_{i_1}}(x_1) \ldots f_{X_{i_n}}(x_n). \tag{B.16}$$

Réciproquement si le vecteur aléatoire $(X_{i_1}, \ldots, X_{i_n})$ est absoluement continu par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ de densité séparable, comme dans (B.16) $f_{(X_{i_1},\ldots,X_{i_n})}(x_1, \ldots, x_n) = \varphi_1(x_1) \ldots \varphi_n(x_n)$ alors, les v.a. X_{i_k} sont indépendantes à densité $f_{X_{i_k}} = \varphi_k$.

B.5.3 Asymptotique des suites d'événements indépendants

Le résultat suivant joue un rôle central en probabilités. Remarquons tout de suite que la partie (i) reprend le résultat établi plus généralement pour les mesures dans le lemme A.14.

Lemma B.36. *(Borel-Cantelli) Soit $(A_n)_n$ une suite dévénements d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.*
(i) *Si $\sum_n \mathbb{P}[A_n] < \infty$, alors $\mathbb{P}[\limsup_n A_n] = 0$,*
(ii) *Si $\sum_n \mathbb{P}[A_n] < \infty$ et $(A_n)_n$ sont indépendants, alors $\mathbb{P}[\limsup_n A_n] = 1$.*
(iii) *Si $(A_n)_n$ sont indépendants, alors soit $\limsup_n A_n$ est négligeable, soit $(\limsup_n A_n)^c$ est négligeable.*

Proof. Il reste à montrer (ii). Par définition de l'indépendance et (B.15), on a

$$\mathbb{P}\left[\cap_{m\geq n} A_m^c\right] = \prod_{m\geq n}(1-\mathbb{P}[A_m]) \leq \prod_{m\geq n} e^{-\mathbb{P}[A_m]} = e^{-\sum_{m\geq n}\mathbb{P}[A_m]} = 0.$$

Ainsi, pour tout $n \geq 1$, l'événement $\cap_{m\geq n} A_m^c$ est négligeable. L'union dénombrable de négligeables $(\limsup_n A_n)^c = \cup_{n\geq 1}\cap_{m\geq n} A_m^c$ est alors négligeable. ♢

Le résultat suivant est assez frappant, et est une conséquence du Lemma de Borel-Cantelli.

Theorem B.37. *Soient $(X_n)_n$ une suite de v.a. indépendantes, et $\mathcal{T} := \cap_n \sigma(X_m, m > n)$ la $\sigma-$algèbre de queue associée. Alors $\mathcal{T}$ est triviale, c'est à dire:*
(i) *Pour tout événement $A \in \mathcal{T}$, on a $\mathbb{P}[A]\mathbb{P}[A^c] = 0$,*
(ii) *Toute v.a. $\mathcal{T}-$mesurable est déterministe p.s.*

Proof. (i) De l'indépendance des $(X_n)_n$, on déduit que pour tout $n \geq 1$, les $\sigma-$algèbres $\mathcal{A}_n := \sigma(X_1, \ldots, X_n)$ et $\mathcal{T}_n := \sigma(X_m, m > n)$ sont indépendantes. Comme $\mathcal{T} \subset \mathcal{T}_n$, on voit que $\mathcal{A}_n$ et $\mathcal{T}$ sont indépendantes, et par suite $\cup_n \mathcal{A}_n$ et $\mathcal{T}$ sont indépendantes. En observant que $\cup_n \mathcal{A}_n$ est un $\pi-$système, on déduit du lemme B.30 que $\mathcal{A}_\infty := \sigma(\cup_n \mathcal{A}_n)$ et $\mathcal{T}$ sont indépendants.

Or, $\mathcal{T} \subset \mathcal{A}_\infty$, donc l'indépendance entre $\mathcal{T}$ et $\mathcal{A}_\infty$ implique que $\mathcal{T}$ est indépendant de lui même, et pour tout $A \in \mathcal{T}$, $\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2$.
(ii) Pour tout $x \in \mathbb{R}$, l'événement $\mathbb{P}[\xi \leq x] \in \{0, 1\}$ d'après (i). Soit $c := \sup\{x : \mathbb{P}[\xi \leq x] = 0\}$. Si $c = -\infty$, ou $c = +\infty$, on voit immédiatement que $\xi = c$ (déterministe), p.s. Si $|c| < \infty$, la définition de c implique que $\mathbb{P}[\xi \leq c - \varepsilon] = \mathbb{P}[\xi > c+\varepsilon] = 0$ pour tout $\varepsilon > 0$. Alors $1 \geq \mathbb{E}[\mathbf{1}_{]c-\varepsilon,c+\varepsilon]}(\xi)] = \mathbb{P}[< \xi \leq c+\varepsilon] = 1$, i.e. $\mathbf{1}_{]c-\varepsilon,c+\varepsilon]}(\xi) = 1$ p.s. et on termine la preuve en envoyanty ε vers 0. ♢

La $\sigma-$algèbre de queue introduite dans le théorème B.37 contient de nombreux événements intéressants comme par exemple

$$\{\lim_n X_n \text{ existe}\},\ \{\sum_n X_n \text{ converge}\},\ \{\lim_n \frac{1}{n}\sum_{i=1}^n X_i \text{ existe}\}$$

Un exemple de v.a. $\mathcal{T}-$mesurable est $\limsup \sum_n X_n$, $\liminf \frac{1}{n}\sum_{i=1}^n X_i$, ...

B.5.4 Asymptotique des moyennes de v.a. indépendantes

Dans ce paragraphe, nous manipulerons des suites de v.a. indépendantes et identiquement distribuées, on écrira plus simplement iid.

On commencera par énoncer la loi des grands nombres pour les suites de v.a. iid intégrables, la démonstration par l'approche des martingales est disponible dans le polycopié de MAP 432 [41]. Puis, nous montrerons le théorème central limite.

Theorem B.38. *(Loi forte des grands nombres) Soit $(X_n)_n$ une suite de v.a. iid intégrables. Alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \quad \longrightarrow \quad \mathbb{E}[X_1] \ p.s.$$

Si les v.a. iid sont de carré intégrable, le théorème central limite donne une information précise sur le taux de convergence de la moyenne empirique vers l'espérance, ou la moyenne théorique.

Theorem B.39. *Soit $(X_n)_n$ une suite de v.a. iid de carré intégrable. Alors*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1]\right) \quad \longrightarrow \quad \mathcal{N}(0, \mathbb{V}[X_1]) \ \textit{en loi,}$$

où $\mathcal{N}(0, \mathbb{V}[X_1])$ désigne la loi normale centrée de variance $\mathbb{V}[X_1]$.

Proof. On note $\bar{X}_i = X_i - \mathbb{E}[X_1]$ et $G_n := \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} \bar{X}_i$. En utilisant les propriétés de la fonction caractéristique pour les variables iid $(X_i)_i$, on obtient que

$$\Phi_{G_n}(u) = \Phi_{\sum_{i=1}^{n} \frac{\bar{X}_i}{\sqrt{n}}}(u) = \prod_{i=1}^{n} \Phi_{\frac{\bar{X}_i}{\sqrt{n}}}(u) = \left(\Phi_{\bar{X}_i}\left(\frac{u}{\sqrt{n}}\right)\right)^n.$$

D'après 2.7 et le fait que $\mathbb{E}[\bar{X}_1 = 0]$ et $\mathbb{E}[\bar{X}_1^2] = \mathbb{V}[X_1] < \infty$, on peut écrire le développement au second ordre suivant:

$$\Phi_{G_n}(u) \ = \ \left(1 - \frac{1}{2}\frac{u^2}{n}\mathbb{V}[X_1] + \circ\left(\frac{1}{n}\right)\right)^n \quad \longrightarrow \quad \phi(u) := e^{-\frac{u^2}{2}\mathbb{V}[X_1]}.$$

On reconnait alors que $\phi = \Phi_{\mathcal{N}(0,\mathbb{V}[X_1])}$, voir question 1 de l'exercice B.9, et on conclut grâce au théorème B.28 de convergence de Lévy. ♢

Appendix C

Conditional expectation

C.1 Premières intuitions

C.1.1 Espérance conditionnelle en espace d'états fini

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et X, Y deux v.a. réelles. Dans ce paragraphe, on suppose que X et Y prennent un nombre fini de valeurs respectivement dans les ensembles $\{x_i, 1 \le i \le n\}$ et $\{y_j, 1 \le j \le m\}$. Il est alors naturel de définir la distribution conditionnelle de X sachant $Y = y_j$ par

$$\mathbb{P}[X = x_i | Y = y_j] \quad := \quad \frac{\mathbb{P}[(X,Y) = (x_i, y_j)]}{\mathbb{P}[Y = y_j]}, \; 1 \le i \le n, \; 1 \le j \le m, \tag{C.1}$$

i.e. parmi tous les événements où la modalité y_j de Y s'est réalisée, la quantité $\mathbb{P}[X = x_i | Y = y_j]$ exprime la fréquence de réalisation de la modalité x_i de X. En notant $\mathbb{P}^{Y=y_j} := \mathbb{P}[X = x_i | Y = y_j]$, on vérifie immédiatement que pour tout j, $\mathbb{P}^{Y=y_j}$ définit une mesure de probabilité sur Ω. L'espérance conditionnelle est alors naturellement définie par :

$$\mathbb{E}[X | Y = y_j] = \sum_{i=1}^{n} x_i \mathbb{P}[X = x_i | Y = y_j], \; 1 \le j \le m.$$

Nous pouvons ainsi définir une v.a. $\xi := \mathbb{E}[X|Y]$ par $\xi(\omega) = \mathbb{E}[X | Y = Y(\omega)]$, appelée espérance conditionnelle de X sachant Y. Notons que $\mathbb{E}[X|Y]$ est complètement déterminée par la réalisation de Y. On retrouve la notion de mesurabilité puisque ceci peut s'écrire mathématiquement ξ est $\sigma(Y)-$mesurable, ou de manière équivalente $\xi = \varphi(Y)$ pour une certaine fonction déterministe φ, voir le lemme B.2.

Par aileurs, un calcul direct montre que pour toute fonction déterministe φ :

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[X|Y]\varphi(Y)\right] &= \sum_{j=1}^{m} \mathbb{P}[Y=y_j]\varphi(y_j)\mathbb{E}[X|Y=y_j] \\
&= \sum_{j=1}^{m} \mathbb{P}[Y=y_j]\varphi(y_j)\sum_{i=1}^{n} x_i \frac{\mathbb{P}[X=x_i, Y=y_j]}{\mathbb{P}[Y=y_j]} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{n} \mathbb{P}[X=x_i, Y=y_j]\varphi(y_j)x_i \;=\; \mathbb{E}\left[X\varphi(Y)\right].
\end{aligned}$$

Notons $\langle , \rangle_2$ le produit scalaire dans $\mathbb{L}^2$, et écrivons ce dernier résultat sous la forme

$$\langle X - \mathbb{E}[X|Y], \varphi(Y)\rangle_2 = 0 \quad \text{pour tout fonction} \quad \varphi\ : \mathbb{R} \longrightarrow \mathbb{R}.$$

Il s'agit de la condition d'orthogonalité de $X - \mathbb{E}[X|Y]$ à l'espace vectoriel de toutes les fonctions de Y ou, de manière équivalente, l'espace vectoriel de toutes les v.a. $\sigma(Y)-$mesurable. Ainsi, $\mathbb{E}[X|Y]$ s'interprète géométriquement comme la projection orthogonale, au sens de $\mathbb{L}^2$, de X sur l'e.v. des v.a. $\sigma(Y)-$mesurable, et est la solution du problème variationnel :

$$\min_f \|X - f(Y)\|^2.$$

Cette interprétation géométrique montre que $\mathbb{E}[X|Y]$ est la meilleure approximation, au sens de $\mathbb{L}^2$, de X par une fonction de Y.

C.1.2 Cas des variables à densités

Supposons maintenant que le couple (X, Y) est à valeurs dans $\mathbb{R}^2$ et admet une distribution absoluement continue par rapport à la mesure de Lebesgue dans $\mathbb{R}^2$, de densité

$$f_{(X,Y)}(x, y)dxdy$$

La loi marginale de Y est obtenue par intégration par rapport à la variable x :

$$f_Y(y) \quad = \quad \int f_{(X,Y)}(x, y)dx.$$

Dans ce cas, la probabilité conditionnelle est naturellement définie par :

$$f_{X|Y=y}(x) \quad = \quad \frac{f_{(X,Y)}(x, y)}{f_Y(y)} \;=\; \frac{f_{(X,Y)}(x, y)}{\int f_{(X,Y)}(x, y)dx},$$

qui est l'analogue de (C.1) dans le contexte présent. Il est clair que pour tout y fixé, la fonction $f_{X|Y=y}(x)$ définit une densité sur $\mathbb{R}$, et qu'on peut lui associer l'opérateur d'espérance :

$$\mathbb{E}[X|Y=y] \quad = \quad \int x f_{X|Y=y}(x)dx$$

qui définit encore une variable aléatoire $\mathbb{E}[X|Y]$, appelée espérance conditionnelle de X sachant Y. Comme dans le cas fini, $\mathbb{E}[X|Y]$ est une fonction déterministe de Y, elle est donc $\sigma(Y)-$mesurable. La condition d'orthogonalité de la section précédente se vérifie aussi par un calcul direct : pour toute fonction $\phi\ : \mathbb{R} \longrightarrow \mathbb{R}$ bornée

$$\begin{aligned}\langle \mathbb{E}[X|Y], \phi(Y)\rangle_{\mathbb{L}^2} &= \mathbb{E}\left[\mathbb{E}[X|Y]\phi(Y)\right] \\ &= \int \phi(y) f_Y(y) \int x f_{X|Y=y}(x) dx dy \\ &= \iint x\phi(y) f_{(X,Y)}(x,y) dx dy \ = \ \mathbb{E}[X\phi(Y)].\end{aligned}$$

On retrouve alors l'interprétation géométrique de l'espérance conditionnelle comme projection orthogonale, au sens de $\mathbb{L}^2$, de la v.a. X sur l'e.v. des v.a. $\sigma(Y)-$mesurables.

C.2 Définition et premières propriétés

Considérons maintenant le cadre général d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et soit $\mathcal{F}$ une sous-$\sigma-$algèbre de $\mathcal{A}$.

Les arguments intuitifs du paragraphe précédent suggèrent d'introduire la notion d'espérance conditionnelle par la projection orthogonale au sens du produit scalaire de $\mathbb{L}^2$

$$P_{\mathcal{F}}(X) \quad := \quad \text{Argmin}\left\{\|X-Y\|_2\ : Y \in \mathcal{L}^2(\mathcal{F}, \mathbb{P})\right\},\ X \in \mathcal{L}^2(\mathcal{A}, \mathbb{P}).$$

Ceci est en effet rendu possible grâce à la structure d'espace de Hilbert de l'espace quotient $\mathbb{L}^2$ muni de la norme $\|.\|_2$.

Lemma C.1. *L'opérateur de projection orthogonale $P_{\mathcal{F}}$ est bien défini sur $\mathbb{L}^2(\mathcal{A}, \mathbb{P})$, et vérifie*

$$\mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[P_{\mathcal{F}}(X)\mathbf{1}_F] \quad \textit{pour tout} \quad F \in \mathcal{F} \textit{ et } X \in \mathbb{L}^2(\mathcal{A}, \mathbb{P}).$$

De plus, on a les propriétés suivantes :
(i) $X \geq 0$ *p.s.* $\Longrightarrow P_{\mathcal{F}}(X) \geq 0$ *p.s.*
(ii) $\mathbb{E}\left[P_{\mathcal{F}}(X)\right] = \mathbb{E}[X]$.

Proof. On travaille avec l'espace quotient $\mathbb{L}^2(\mathcal{A}, \mathbb{P})$ identifiant ainsi les v.a. égales p.s. La projection orthogonale $P_{\mathcal{F}}$ est bien définie car l'e.v. $\mathbb{L}^2(\mathcal{A}, \mathbb{P})$ et le s.e.v. $\mathbb{L}^2(\mathcal{F}, \mathbb{P})$ sont complets. Alors, on sait que pour tout X, il existe une (unique) v.a. $Z := P_{\mathcal{F}}(X) \in \mathbb{L}^2(\mathcal{F}, \mathbb{P})$ vérifiant les conditions d'orthogonalité

$$\mathbb{E}[(X-Z)Y] = 0 \quad \text{pour tout} \quad Y \in \mathbb{L}^2(\mathcal{F}, \mathbb{P}).$$

En particulier, pour tout $F \in \mathcal{F}$, la v.a. $Y = \mathbf{1}_F \in \mathbb{L}^2(\mathcal{F}, \mathbb{P})$ induit la condition d'orthogonalité $\mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[Z\mathbf{1}_F]$.

Supposons maintenant que $X \geq 0$ p.s., notons $Z := P_{\mathcal{F}}(X)$, et prenons $F := \{Z \leq 0\} \in \mathcal{F}$. Alors $0 \leq \mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[Z\mathbf{1}_F] = -\mathbb{E}[Z^-] \leq 0$, et $Z^- = 0$ p.s. montrant la propriété (i).
Pour la propriété (ii), il suffit de remarquer que $F = \Omega \in \mathcal{F}$ du fait que $\mathcal{F}$ est une $\sigma-$algèbre. Alors (i) donne le résultat voulu. ♢

Theorem C.2. *Pour toute v.a. $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$, il existe une v.a. Z telles que*
(a) *Z est $\mathcal{F}-$mesurable,*
(b) $\mathbb{E}[|Z|] < \infty$,
(c) *Pour tout événement $F \in \mathcal{F}$, on a $\mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[Z\mathbf{1}_F]$.*
De plus, si $\tilde{Z}$ est une autre v.a. vérifiant (a,b,c), alors $Z = \tilde{Z}$ p.s.

Definition C.3. *Une v.a. vérifiant les propriétés (a)-(b)-(c) est appelée version de l'espérance conditionnelle de X sachant $\mathcal{F}$, notée $\mathbb{E}[X|\mathcal{F}]$, et est unique à l'égalité p.s. près.*

Si $\mathcal{F} = \sigma(Y_1, \ldots, Y_n)$, on écrit simplement $\mathbb{E}[X|Y_1, \ldots, Y_n]$.

Preuve du théorème C.2 Commençons par montrer l'unicité. Si Z et $\tilde{Z}$ vérifient (a,b,c), alors $\mathbb{E}[(Z - \tilde{Z})\mathbf{1}_F] = 0$ pour tout $F \in \mathcal{F}$. Mais Z et $\tilde{Z}$ étant $\mathcal{F}-$mesurable, on peut choisir $F = \{Z - \tilde{Z} \geq 0\} \in \mathcal{F}$, et l'égalité précédente implique que $(Z - \tilde{Z})^+ = 0$ p.s. Le choix $F = \{Z - \tilde{Z} \leq 0\} \in \mathcal{F}$ conduit à $(Z - \tilde{Z})^- = 0$ p.s. et par suite $Z = \tilde{Z}$ p.s.

Pour l'existence, il suffit de traiter le cas $X \geq 0$ et d'utiliser la décomposition $X = X^+ - X^-$ pour conclure le cas où X a un signe arbitraire (ou plutôt n'a pas de signe !). La v.a. $X_n := X \wedge n$ est bornée, donc dans $\mathbb{L}^2(\mathcal{A}, \mathbb{P})$. La v.a. $Z_n := P_{\mathcal{F}}(X_n)$ est alors bien définie d'après le lemme C.1 et vérifie par définition les conditions (a,b,c). Observons que la suite $(Z_n)_n$ est croissante, comme conséquence de la propriété (i) du lemme C.1 et de la linéarité de la projection $P_{\mathcal{F}}$. On introduit alors la v.a.

$$Z \quad := \quad \lim \uparrow_n Z_n.$$

Il est clair que Z hérite la $\mathcal{F}-$mesurabilité des Z_n, et que $\mathbb{E}[Z\mathbf{1}_F] = \mathbb{E}[X\mathbf{1}_F]$ pour tout $F \in \mathcal{F}$ par le théorème de convergence monotone. Pour la condition (b), remarquons par que $\mathbb{E}[Z] = \lim_{n\to\infty} \uparrow \liminf_n \mathbb{E}[Z_n] = \mathbb{E}[X \wedge n] \leq \mathbb{E}[X]$, où on a utilisé les propriétés (i) et (ii) du lemme C.1. ♢

Remark C.4. Regardons deux cas extrêmes pour la $\sigma-$algèbre $\mathcal{F}$.
- Soit $\mathcal{F} = \{\emptyset, \Omega\}$ la plus petite $\sigma-$algèbre correspondant à l'absence totale d'information. Alors la condition (a) dit que Z est déterministe, i.e. $Z = \mathbb{E}[Z]$, et la condition d'orthogonalité (c) permet d'identifier cette constante $\mathbb{E}[X] = \mathbb{E}[Z] = Z$. Ainsi l'espérance conditionnelle dans ce cas se confond avec l'espérance.
- Soit $\mathcal{F} = \sigma(X)$. Alors la condition d'orthogonalité (c) avec $F^+ := \{X - Z \geq 0\} \in \sigma(X)$ donne $\mathbb{E}[(X - Z)\mathbf{1}_{\{X-Z\geq 0\}}] = 0$, soit $(Y - Z)^+ = 0$ p.s., et avec $F^- = \{X - Z \leq 0\} \in \sigma(X)$ donne $(X - Z)^- = 0$ p.s. Ainsi $\mathbb{E}[X|\sigma(X)] = X$ qui vérifie bien les autres conditions (a) et (b).

C.3 Propriétés de l'espérance conditionnelle

Commençons par les propriétés déjà esquissées dans le paragraphe précédent.

Proposition C.5. *L'opérateur d'espérance conditionnelle* $\mathbb{E}[.|\mathcal{F}]$ *est linéaire, et pour tout* $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$*, on a :*
(i) $\mathbb{E}\left[\mathbb{E}[X|\mathcal{F}]\right] = \mathbb{E}[X]$,
(ii) *si* X *est* $\mathcal{F}-$*mesurable,* $\mathbb{E}[X|\mathcal{F}] = X$ *p.s.*
(iii) *si* $X \geq 0$, $\mathbb{E}[X|\mathcal{F}] \geq 0$ *p.s.*

Exercise C.6. *Prouver la proposition C.5.*

Nous montrons maintenant que l'espérance conditionnelle jouit des mêmes propriétés de passage à la limite que l'espérance.

Proposition C.7. *Pour* $X, X_n \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$*,* $n \in \mathbb{N}$*, on a :*
(1- Convergence monotone) *si* $0 \leq X_n \uparrow X$*, alors* $\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}]$,
(2- Fatou) *si* $X_n \geq 0$*, alors* $\mathbb{E}[\liminf_n X_n|\mathcal{F}] \leq \liminf_n \mathbb{E}[X_n|\mathcal{F}]$,
(3- Convergence dominée) *si* $\sup_n |X_n| \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ *et* $X_n \longrightarrow X$*, p.s. alors* $\mathbb{E}[X_n|\mathcal{F}] \longrightarrow \mathbb{E}[X|\mathcal{F}]$ *p.s.*

Proof. 1- la suite $Z_n := \mathbb{E}[X_n|\mathcal{F}]$ est croissante d'après la proposition C.5 (iii). On définit alors la variable $Z = \lim \uparrow Z_n$ qui est par définition $\mathcal{F}-$mesurable positive et, par Fatou, $\mathbb{E}[Z] \leq \liminf_n \mathbb{E}[Z_n] = \liminf_n \mathbb{E}[X_n] \leq \mathbb{E}[X] < \infty$. Enfin, pour tout $F \in \mathcal{F}$, on a $\mathbb{E}[X_n \mathbf{1}_F] = \mathbb{E}[Z_n \mathbf{1}_F]$ et on déduit du théorème de convergence monotone que $\mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[Z\mathbf{1}_F]$. Ainsi Z vérifie les propriétés (a,b,c) du théorème C.2 et $Z = \mathbb{E}[X|\mathcal{F}]$, p.s.
2- D'après la monotonie de l'opérateur d'espérance conditionnelle due à (iii) de la proposition C.5, on a

$$\inf_{k \geq n} \mathbb{E}[X_n|\mathcal{F}] \geq \mathbb{E}\left[\inf_{k \geq n} X_n|\mathcal{F}\right] \quad \text{pour tout} \quad n \geq 1,$$

et on conclut en utilisant le résultat de convergence monotone démontré en première partie de cette preuve.
3- Avec $Y_n := |X_n - X|$ et $\bar{Y} := \sup_n Y_n$, on vérifie que $\bar{Y} \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$, et on applique le lemme de Fatou conditionnel, qu'on vient de démontrer, à la v.a. $\bar{Y} - Y_n$. Le résultat s'en déduit immédiatement. ♢

L'inégalité de Jensen s'étend ausi aux espérances conditionnelles :

Proposition C.8. *Soit* $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$*, et* $g \ : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{\infty\}$ *une fonction convexe telle que* $\mathbb{E}[|g(X)|] < \infty$*. Alors* $\mathbb{E}[g(X)|\mathcal{F}] \geq g\left(\mathbb{E}[X|\mathcal{F}]\right)$.

Proof. Il suffit de répéter les arguments de la preuve de l'inégalité de Jensen dans le cas non conditionnel, théorème B.6... ♢

La propriété suivante est très utile, et est une conséquence de la propriété des projections itérées en algèbre linéaire.

Proposition C.9. *(Projections itérées) Pour $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et $\mathcal{F}, \mathcal{G}$ des sous-$\sigma-$algèbre de $\mathcal{A}$:*

$$\mathcal{F} \subset \mathcal{G} \implies \mathbb{E}\big[\mathbb{E}\{X|\mathcal{G}\}|\mathcal{F}\big] = \mathbb{E}[X|\mathcal{F}].$$

Proof. 1-On observe que $\mathcal{L}^2(\mathcal{F}, \mathbb{P}) \subset \mathcal{L}^2(\mathcal{G}, \mathbb{P})$, et que par suite le résultat dans le cas $X \in \mathcal{L}^2(\mathcal{A}, \mathbb{P})$ est une conséquence immédiate du théorème de projections itérées en algèbre linéaire. Puis, le théorème de convergence monotone permet de l'étendre aux variables $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$. $\diamondsuit$

La propriété suivante généralise celle de la proposition C.5 (i).

Proposition C.10. *Soient $\mathcal{F}$ une sous-$\sigma-$algèbre de $\mathcal{A}$, $X \in \mathcal{L}^0(\mathcal{A})$ et $Y \in \mathcal{L}^0(\mathcal{F})$. On suppose $\mathbb{E}[|X|] < \infty$ et $\mathbb{E}[|XY|] < \infty$. Alors*

$$\mathbb{E}[XY|\mathcal{F}] \quad = \quad Y\mathbb{E}[X|\mathcal{F}].$$

Proof. On commence par le cas $Y = \mathbf{1}_A$, $A \in \mathcal{F}$. Alors pour tout $F \in \mathcal{F}$, on a $\mathbb{E}\left[Y\mathbb{E}[X|\mathcal{F}]\mathbf{1}_F\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_{A\cap F}\right] = \mathbb{E}[X\mathbf{1}_{A\cap F}] = \mathbb{E}[XY\mathbf{1}_F]$ d'après la définition de $\mathbb{E}[X|\mathcal{F}]$ et du fait que $A \cap F \in \mathcal{F}$. Ainsi, la proposition est vraie pour les indicatrices d'événements de $\mathcal{F}$.

Si X est une v.a. positive, la propriété précédente s'étend par linéarité à $\mathcal{S}^+$, l'ensemble des v.a. simples positives, et par le théorème de convergence monotone à l'ensemble des v.a. positives telles que $\mathbb{E}[|XY|] < \infty$ et $\mathbb{E}[|X|] < \infty$ (pour que l'espérance conditionnelle ait un sens).

Enfin, pour des variables X, Y générales, on décompose $X = X^+ - X^-$, $Y = Y^+ - Y^-$, et on applique le résultat établi pour les v.a. positives. $\diamondsuit$

Les deux dernières propriétés donnent des résultats utiles sur l'éspérance conditionnelle en présence d'indépendance.

Proposition C.11. *Soient $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et $\mathcal{F}, \mathcal{G}$ des sous-$\sigma-$algèbres de $\mathcal{A}$ telles que $\mathcal{G}$ est indépendante de $\sigma(\sigma(X), \mathcal{F})$. Alors*

$$\mathbb{E}[X|\sigma(\mathcal{F}, \mathcal{G})] \quad = \quad \mathbb{E}[X|\mathcal{F}].$$

Proof. Il suffit de vérifier pour $X \in \mathcal{L}^1_+(\mathcal{A}, \mathbb{P})$ que

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_A\right] \quad \text{pour tout} \quad A \in \sigma(\mathcal{F}, \mathcal{G}).$$

En remarquant que $A \longmapsto \mathbb{E}[X\mathbf{1}_A]$ et $A \longmapsto \mathbb{E}\left[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_A\right]$ sont des mesures sur Ω, on déduit de la proposition A.5 qu'il suffit de vérifier l'égalité ci-dessus pour les événements A dans le $\pi-$système $\mathcal{F} \cap \mathcal{G}$. Soient alors $F \in \mathcal{F}$ et $G \in \mathcal{G}$. En utilisant l'indépendance entre $\mathcal{G}$ et $\sigma(\sigma(X), \mathcal{F})$, la proposition B.32, et la définition de $\mathbb{E}[X|\mathcal{F}]$, on voit que :

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_{F\cap G}\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_F\right]\mathbb{E}[\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_F]\mathbb{E}[\mathbf{1}_G] = \mathbb{E}\left[X\mathbf{1}_{F\cap G}\right].$$

$\diamondsuit$

Proposition C.12. *Soient* (X,Y) *deux v.a. à valeurs dans* $\mathbb{R}^n$ *et* $\mathbb{R}^m$*, respectivement, et* $g\ :\mathbb{R}^n\times\mathbb{R}^m\longrightarrow\mathbb{R}$ *une fonction telle que* $\mathbb{E}[|g(X,Y)|]<\infty$*. Si* X *et* Y *sont indépendantes, alors*

$$\mathbb{E}[g(X,Y)|X]=G(X)\quad \text{où}\quad G(x):=\mathbb{E}[g(x,Y)]\ \text{pour tout}\ x\in\mathbb{R}^n.$$

Proof. Pour tout $A\in\sigma(X)$, on doit vérifier que $\mathbb{E}[g(X,Y)\mathbf{1}_A]=\mathbb{E}[G(X)\mathbf{1}_A]$. Comme X et Y sont indépendantes, la loi du couple (X,Y) est la loi produit $\mathbb{P}^X\otimes\mathbb{P}^Y$, et on obtient immédiatement par le théorème de Fubini que

$$\begin{aligned}\mathbb{E}[g(X,Y)\mathbf{1}_A] &= \int g(x,y)\mathbf{1}_A(x)\mathbb{P}^X\otimes\mathbb{P}^Y(dx,dy)\\ &= \int\left(\int g(x,y)\mathbb{P}^Y(dy)\right)\mathbf{1}_A(x)\mathbb{P}^X(dx)=\int G(x)\mathbf{1}_A(x)\mathbb{P}^X(dx),\end{aligned}$$

ce qui est exactement le résultat recherché. ♢

C.4 Application au filtre de Kalman-Bucy

On considère une récurrence aléatoire définie par

$$X_k = F_kX_{k-1}+f_k+\varepsilon_k \tag{C.2}$$

$$Y_k = H_kX_k+h_k+\eta_k, \tag{C.3}$$

où $F_k\in\mathcal{M}_{\mathbb{R}}(n,n)$, $f_k\in\mathbb{R}^n$, $H_k\in\mathcal{M}_{\mathbb{R}}(m,n)$, $h_k\in\mathbb{R}^m$ sont les paramètres du système, connus par l'utilisateur, et $(\varepsilon_k)_k$, $(\eta_k)_k$ sont des suites de v.a. que l'on supposera indépendantes et même gaussiennes...

Les v.a. $(X_k)_k$ à valeurs dans $\mathbb{R}^n$ constituent l'objet d'intérêt, mais **ne sont pas observables**. Les v.a. $(Y_k)_k$ à valeurs dans $\mathbb{R}^m$ sont observées par l'utilisateur. On dit que (C.2) est l'équation d'état, et (C.3) est l'équation d'observation. Le problème de filtrage consiste à *chercher, à partir des variables observables* $(Y_k)_k$*, la meilleure approximation des variables inobservables* $(X_k)_k$.

L'indice k joue un rôle important. Il représente la date de l'observation, et induit une structure de l'information naturelle et de sa progression. Ceci représente une composante importante du problème : *à chaque date* k*, on cherche la meilleure approximation de* X_k *sur la base des observations* $\{Y_j, j\le k\}$. D'après la définition de l'espérance conditionnelle dans $\mathcal{L}^2$, comme projection orthogonale sur l'espace vectoriel engendré par les v.a. de $\mathcal{L}^2$ mesurables par rapport au conditionnement, la meilleure approximation est donnée par

$$\mathbb{E}[X_k|Y_0,\ldots,Y_k],$$

qui est ainsi la quantité que l'utilisateur cherche à calculer à chaque date k.

Enfin, en vue d'une application réelle, l'aspect de l'implémentation est une composante essentielle du problème. Quand la taille des observations croit, l'effort numérique pour calculer l'espérance conditionnelle ci-dessus peut devenir rapidement gigantesque. Des méthodes de type **mise à jour** sont très désirables pour l'utilisateur : *utiliser l'observation la plus récente pour mettre à jour l'approximation de la date précédente.*

C.4.1 Lois conditionnelles pour les vecteurs gaussiens

Dans ce paragraphe, nous isolons un résultat sur les vecteurs gaussiens qui sera crucial pour la résolution du problème de filtrage de Kalman-Bucy. Rappelons qu'un vecteur aléatoire Z, à valeurs dans $\mathbb{R}^n$, est gaussien si $a \cdot Z = \sum_{i=1}^n a_i Z_i$ est gaussien sur $\mathbb{R}$ pour tout $a \in \mathbb{R}^n$.

Proposition C.13. *Soit (X,Y) un vecteur gaussien à valeurs dans $\mathbb{R}^n \times \mathbb{R}^m$ de moyenne et de matrice de variances-covariances*

$$\mu = \begin{pmatrix} \mu^X \\ \mu^Y \end{pmatrix} \quad et \quad V = \begin{pmatrix} V^X & (V^{XY})^{\mathrm{T}} \\ V^{XY} & V^Y \end{pmatrix}.$$

Supposons que la matrice $\mathbb{V}[Y] = V^Y$ est inversible. Alors, la loi conditionnelle de X sachant $Y = y$ est gaussienne de moyenne et variance :

$$\mathbb{E}[X|Y=y] = \mu^X + V^{XY}(V^Y)^{-1}(y-\mu^Y), \ et \ \mathbb{V}[X|Y=y] = V^X - V^{XY}(V^Y)^{-1}(V^{XY})^{\mathrm{T}}$$

Proof. 1- On note $Z = (X,Y)$, et on suppose d'abord que V est inversible. Alors, pour $z = (x,y) \in \mathbb{R}^n \times \mathbb{R}^m$, on calcule la densité de la loi de X conditionnellement à $\{Y = y\}$ au point x par :

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f_Z(z)}{f_Y(y)} \\ &= (2\pi)^{-n/2}\sqrt{\frac{\det(V^Y)}{\det(V)}}\, e^{\left[-\frac{1}{2}(z-\mu)^{\mathrm{T}}V^{-1}(z-\mu)+\frac{1}{2}(y-\mu^Y)^{\mathrm{T}}(V^Y)^{-1}(y-\mu^Y)\right]}. \end{aligned} \tag{C.4}$$

On note $W := V^X - V^{XY}(V^Y)^{-1}(V^{XY})^{\mathrm{T}}$, et on vérifie par un calcul direct que

$$\begin{pmatrix} I & -V^{XY}(V^Y)^{-1} \\ 0 & I \end{pmatrix} V \begin{pmatrix} I & 0 \\ -(V^Y)^{-1}(V^{XY})^{\mathrm{T}} & I \end{pmatrix} = \begin{pmatrix} W & 0 \\ 0 & V^Y \end{pmatrix}, \tag{C.5}$$

dont on tire que

$$\frac{\det(V)}{\det(V^Y)} = \det(W). \tag{C.6}$$

On déduit aussi de (C.5) que

$$V^{-1} = \begin{pmatrix} I & 0 \\ -(V^Y)^{-1}(V^{XY})^{\mathrm{T}} & I \end{pmatrix} \begin{pmatrix} W^{-1} & 0 \\ 0 & (V^Y)^{-1} \end{pmatrix} \begin{pmatrix} I & -V^{XY}(V^Y)^{-1} \\ 0 & I \end{pmatrix}$$

qui implique que :

$$\begin{aligned} (z-\mu)^{\mathrm{T}}V^{-1}(z-\mu) &= (x-M(x))^{\mathrm{T}}W^{-1}(x-M(x)) \\ &\quad +(y-\mu^Y)^{\mathrm{T}}(V^Y)^{-1}(y-\mu^Y), \end{aligned} \tag{C.7}$$

où $M(x) := \mu^X + V^{XY}(V^Y)^{-1}(y-\mu^Y)$. en injectant (C.6) et (C.7) dans (C.4), on obtient une expression de la densité conditionnelle $p_{X|Y=y}$ qui correspond à la loi annoncée dans la proposition.

2- Pour une matrice V générale, on considère l'approximation $Z_n := Z + \frac{1}{n}G$, où G est une v.a. indépendante de Z de loi gaussienne centrée réduite dans $\mathbb{R}^{n+m}$. Alors, on peut appliquer le résultat de la première étape et obtenir l'expression de la densité $f_{X_n|Y_n=y}(y)$ qui converge vers la densité annoncée dans la proposition, et par convergence dominéeles il y a convergence des fonctions de répartition correspondantes en tout point. Ainsi, pour finir la démonstration, il ne reste plus qu'à vérifier que $X_n|Y_n = y$ converge en loi vers $X|Y = y$, par exemple en montrant que $\Phi_{X_n|Y_n=y} \longrightarrow \Phi_{X|Y=y}$ simplement. Pour celà, on calcule directement :

$$\begin{aligned}\Phi_{X_n|Y_n=y}(u) &= \mathbb{E}\left[e^{iu\cdot X_n}|Y_n = y\right] \\ &= \mathbb{E}\left[e^{iu\cdot(X+G_1^n)}|Y + G_2^n = y\right],\end{aligned}$$

où $G^n = (G_1^n, G_2^n) := \frac{1}{n}G$. En remarquant que G_1^n est indépendant de (X, Y, G_2^n), puis que G_2^n est indépendant de (X, Y), on obtient alors :

$$\begin{aligned}\Phi_{X_n|Y_n=y}(u) &= \Phi_{G_1^n}(u)\mathbb{E}\left[\mathbb{E}\left[e^{iu\cdot X}|Y = y - g\right]_{g=G_2^n}\right] \\ &= \Phi_{G_1^n}(u)\mathbb{E}\left[\{\Phi_{X|Y=y-g}(u)\}_{g=G_2^n}\right] \longrightarrow \Phi_{X|Y=y}(u)\end{aligned}$$

par utilisation du théorème de convergence dominée. $\diamondsuit$

C.4.2 Filtre de Kalman-Bucy

Précisions à présent les hypothèses sur les composantes aléatoires du système d'état-observation (C.2)-(C.3) :

- l'état initial X_0 est gaussien de moyenne $\mu_0^X := \mathbb{E}[X_0]$ et de variance $V_0^X := \mathbb{V}[X_0]$,
- l'aléa générant l'état $(\varepsilon_k)_k$ est une suite de v.a. indépendantes gaussiennes centrées de variances V_k^ε,
- le bruit d'observation $(\eta_k)_k$ est une suite de v.a. indépendantes gaussiennes centrées de variances V_k^η,
- les bruits $(\varepsilon_k)_k$, $(\eta_k)_k$ et l'état initial X_0 sont mutuellement indépendants.

On modélise l'information en introduisant les $\sigma-$algèbres $\mathcal{F}_k := \sigma(Y_0, \ldots, Y_k)$. Pour tout N, le vecteur aléatoire $(X_k, Y_k)_{0\le k\le N}$ est gaussien. La proposition C.13 garantit que la loi conditionnelle de X_k sachant $\mathcal{F}_k$ est une gaussienne dont il suffit de calculer la moyenne et la variance pour la caractériser :

$$\hat{X}_k := \mathbb{E}[X_k|\mathcal{F}_k], \qquad V_k := \mathbb{V}[X_k|\mathcal{F}_k].$$

Notons que, d'après la proposition C.13, la matrice de covariances conditionnelle V_k est indépendante du conditionnement :

$$V_k := \mathbb{E}\left[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^{\mathrm{T}}\right].$$

Pour un calcul efficace des caractéristiques de la loi du filtre $\hat{X}_k$ et V_k, on procède en deux étapes :

1. Etape de prédiction : étant donnés $(\hat{X}_{k-1}, V_{k-1})$, on voit de l'équation d'état (C.2) que la loi de X_k sachant $\mathcal{F}_{k-1}$ est gaussienne de caractéristiques :

$$\hat{X}_k^{\mathrm{pr}} := \mathbb{E}[X_k|\mathcal{F}_{k-1}],\ V_k^{\mathrm{pr}} := \mathbb{V}[X_k|\mathcal{F}_{k-1}] = \mathbb{E}\left[(X_k - \hat{X}_k^{\mathrm{pr}})(X_k - \hat{X}_k^{\mathrm{pr}})^{\mathrm{T}}\right].$$

   où on a encore utilisé que la matrice de variance est indépendante du conditionnement, comme conséquence de la proposition C.13.

2. Etape de correction : on utilise l'information supplémentaire Y_k de la date k ou, plus précisément, *l'innovation*

$$I_k \quad := \quad Y_k - \mathbb{E}[Y_k|\mathcal{F}_{k-1}] \ = \ H_k(X_k - \hat{X}_k^{\mathrm{pr}}) + \eta_k,$$

   où on a utilisé l'équation d'observation (C.3) et le fait que η_k est indépendant de $\mathcal{F}_{k-1}$. En particulier, on voit que

$$\begin{array}{rcl} (I_k)_k \text{ est gaussien et } \mathbb{E}[I_k] & = & \mathbb{E}[I_k|\mathcal{F}_{k-1}] = 0, \\ \mathbb{V}[I_k] & = & \mathbb{V}[I_k|\mathcal{F}_{k-1}] = H_k V_k^{\mathrm{pr}} H_k^{\mathrm{T}} + V_k^{\eta}. \end{array} \tag{C.8}$$

Theorem C.14. *(Kalman-Bucy) Supposons que la matrice de variance du bruit d'observation V_k^{η} est inversible pour tout $k \geq 0$. Alors, les caractéristiques de l'étape de prédiction sont données par $\hat{X}_0^{\mathrm{pr}} = \mu_0^X = \mathbb{E}[X_0]$, $V_0^{\mathrm{pr}} = V_0^X = \mathbb{V}[X_0]$,*

$$\hat{X}_k^{\mathrm{pr}} = F_k \hat{X}_{k-1} + f_k, \qquad V_k^{\mathrm{pr}} = F_k V_{k-1} F_k^{\mathrm{T}} + V_k^{\eta},\ k \geq 1, \tag{C.9}$$

et celles de l'étape de correction :

$$\hat{X}_k = \hat{X}_k^{\mathrm{pr}} + K_k[Y_k - (H_k \hat{X}_k^{\mathrm{pr}} + h_k)],\ V_k = (I - K_k H_k) V_k^{\mathrm{pr}},\ k \geq 0, \tag{C.10}$$

où $K_k := V_k^{\mathrm{pr}} H_k^{\mathrm{T}} (H_k V_k^{\mathrm{pr}} H_k^{\mathrm{T}} + V_k^{\eta})^{-1}$ est appelée matrice de gain de Kalman.

Proof. On décompose en trois étapes :
1- Initialisation de la prédiction. Le vecteur aléatoire (X_0, Y_0) est gaussien de moyenne et variance

$$\begin{pmatrix} \mu_0^X \\ H_0 \mu_0^X \end{pmatrix}, \qquad \begin{pmatrix} V_0^X & V_0^X H_0^{\mathrm{T}} \\ H_0 V_0^X & H_0 V_0^X H_0^{\mathrm{T}} + V_0^{\eta} \end{pmatrix}.$$

On déduit de la proposition C.13 que la loi de X_0 conditionnellement à Y_0 est gaussienne de caractéristiques

$$\begin{array}{rcl} \hat{X}_0 & = & \mu_0^X + V_0^X H_0^{\mathrm{T}} (H_0 V_0^X H_0^{\mathrm{T}} + V_0^{\eta})^{-1} [Y_0 - (H_0 \mu_0^X + h_0)], \\ V_0 & = & V_0^X - V_0^X H_0^{\mathrm{T}} (H_0 V_0^X H_0^{\mathrm{T}} + V_0^{\eta})^{-1} H_0 V_0^X. \end{array}$$

<u>*2- Prédiction.*</u> D'après la proposition C.13, la loi de X_k sachant $\mathcal{F}_{k-1}$ est gaussienne dont on calcule la moyenne et la variance à partir de l'équation d'état :

$$\begin{aligned}\hat{X}_k^{\mathrm{pr}} &= F_k \hat{X}_{k-1} + f_k \\ V_k^{\mathrm{pr}} &= \mathbb{E}\left[(X_k - \hat{X}_k^{\mathrm{pr}})(X_k - \hat{X}_k^{\mathrm{pr}})^{\mathrm{T}}\right] \\ &= \mathbb{E}\left[(F_k(X_{k-1} - \hat{X}_{k-1}) + \varepsilon_k)(F_k(X_{k-1} - \hat{X}_{k-1}) + \varepsilon_k)^{\mathrm{T}}\right] \\ &= F_k V_{k-1} F_k^{\mathrm{T}} + V_k^{\eta},\end{aligned}$$

d'après nos hypthèses sur le bruit η_k.
<u>*3- Correction.*</u> D'après la proposition C.13, la loi de X_k sachant $\mathcal{F}_k$ est gaussienne. D'après (C.8) et la proposition C.11, on calcule :

$$\hat{X}_k = \hat{X}_k^{\mathrm{pr}} + \mathbb{E}\left[X_k - \hat{X}_k^{\mathrm{pr}} | \mathcal{F}_{k-1}, I_k\right] = \hat{X}_k^{\mathrm{pr}} + \mathbb{E}\left[X_k - \hat{X}_k^{\mathrm{pr}} | I_k\right] \tag{C.11}$$

Par suite, $X_k - \hat{X}_k = (X_k - \hat{X}_k^{\mathrm{pr}}) - \mathbb{E}\left[X_k - \hat{X}_k^{\mathrm{pr}} | I_k\right]$, et

$$V_k = \mathbb{E}\left[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^{\mathrm{T}}\right] = \mathbb{E}\left[\mathbb{V}[X_k - \hat{X}_k^{\mathrm{pr}} | I_k]\right]. \tag{C.12}$$

Pour calculer (C.11) et (C.12), on observe que le vecteur aléatoire $(X_k - \hat{X}_k^{\mathrm{pr}}, I_k)$ est gaussien centré de matrice de variance

$$\begin{pmatrix} V_k^{\mathrm{pr}} & V_k^{\mathrm{pr}} H_k^{\mathrm{T}} \\ H_k V_k^{\mathrm{pr}} & H_k V_k^{\mathrm{pr}} H_k^{\mathrm{T}} + V_k^{\eta} \end{pmatrix}.$$

La matrice $H_k V_k^{\mathrm{pr}} H_k^{\mathrm{T}} + V_k^{\eta}$ est inversible, du fait de l'hypothèse d'inversibilité de V_k^{η}, on déduit alors de la proposition C.11 que

$$\begin{aligned}\hat{X}_k &= \hat{X}_k^{\mathrm{pr}} + V_k^{\mathrm{pr}} H_k^{\mathrm{T}} (H_k V_k^{\mathrm{pr}} H_k^{\mathrm{T}} + V^{\eta})^{-1} I_k, \\ V_k &= V_k^{\mathrm{pr}} - V_k^{\mathrm{pr}} H_k^{\mathrm{T}} (H_k V_k^{\mathrm{pr}} H_k^{\mathrm{T}} + V^{\eta})^{-1} H_k V_k^{\mathrm{pr}}.\end{aligned}$$

♢

Bibliography

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, *Coherent measures of risk*, Mathematical Finance, (1999).

[2] J.-P. Aubin (1975), *L'analysenon linéaire et ses motivations économiques*, Masson.

[3] Basel Committee on Banking Supervision, *International convergence of capital measurement and capital standards*, Bank of International Settlements, 1988.

[4] Basel Committee on Banking Supervision, *Amendment to the capital accord to incorporate market risks*, Bank of International Settlements, 1996.

[5] Basel Committee on Banking Supervision, *Basel II: International convergence of capital measurement and capital standards: A revised framework*, Bank of International Settlements, 2005.

[6] M. Beiglbock, M., Schachermayer, W. and Veliyev,

[7] Black F. and Scholes M. (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-654.

[8] H. Berestycki, J. Busca, and I. Florent, *Asymptotics and calibration of local volatility models*, Quant. Finance, 2 (2002), pp. 61–69.

[9] Carr P., Ellis K. and Gupta V. (1998), Static Hedging of Exotic Options, *Journal of Finance*, June 1998, pp. 1165-90.

[10] R. Cont, *Model uncertainty and its impact on the pricing of derivative instruments*, Mathematical Finance, 16 (2006), pp. 519–542.

[11] Conze A. and Viswanathan R. (1991), Path dependent options : the case of lookback options, *Journal of Finance*, 46, 1893-1907.

[12] J. C. Cox, *The constant elasticity of variance option pricing model*, Journal of Portfolio Management, 22 (1996), pp. 15–17.

[13] M. Crouhy, D. Galai, and R. Mark, *The Essentials of Risk Management*, McGraw-Hill, 2005.

[14] Cox J.C., Ross S.A. and Rubinstein M. (1979), Option pricing : a simplified approach, *Journal of Financial Economics*, 7, 229-263.

[15] S. Crépey, *Calibration of the local volatility in a trinomial tree using Tikhonov regularization*, Inverse Problems, 19 (2003), pp. 91–127.

[16] Dana R.A. and Jeanblanc-Piqué M. (1998), *Marchés financiers en temps continu, valorisation et équilibre*, 2ème édition, Economica.

[17] F. Delbaen and H. Shirakawa, *A note on option pricing for the constant elasticity of variance model*, Asia-Pacific Financial Markets, 9 (2002), pp. 85–99.

[18] Dellacherie C. and Meyer P.A. (1975), *Probabilités et potentiel*, Hermann.

[19] Demange G. and Rochet J.C. (1992), *Méthodes mathématiques de la finance*, Economica.

[20] Duffie D. (1996), *Dynamic Asset Pricing Theory*, 2nd edition, Princeton University Press, Princeton (New Jersey).

[21] B. Dupire, *Pricing with a smile*, RISK, 7 (1994), pp. 18–20.

[22] N. El Karoui, *Couverture des risques dans les marchés financiers.* Lecture notes for master 'Probability and Finance', Paris VI university.

[23] El Karoui N. and Gobet E. (2005), *Modèles Stochastique en Finance*, Polycopié de cours, Ecole Polytechnique, Département de Mathématiques Appliquées.

[24] El Karoui N. and Rochet J.C. (1989), A pricing formula for options on zero-coupon bonds. Preprint.

[25] P. S. Hagan and D. E. Woodward, *Equivalent black volatilities.* Research report, 1998.

[26] Harrison J.M. and Kreps D. (1979), Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory*, 20, 381-408.

[27] Harrison J.M. and Pliska S. (1981), Martingales and stochastic integrals in the theory of continuous trading : complete markets, *Stochastic Processes and their Applications*, 11, 215-260.

[28] J. Jacod et P. Protter (2002), *Probability Essentials*, 2nd Edition. Universitext, Springer-Verlag.

[29] Jacod, J. and Shiryaev, A.N.(1998), Local martingales and the fundamental asset pricing theorems in the discrete-time case. *Finance and Stochastics* 2, 259-273.

[30] Karatzas, I. and Shreve, S. (1991), *Brownian Motion and Stochastic Calculus*, Springer Verlag.

[31] R. Lagnado and S. Osher, *A technique for calibrating derivative security pricing models: numerical solution of the inverse problem*, J. Comput. Finance, 1 (1997).

[32] Lamberton D. and Lapeyre B. (1997), *Introduction au calcul stochastique appliqué à la finance*, 2ème édition, Ellipses.

[33] Musiela M. and Rutkowski M. (1997), *Martingale Methods in Financial Modelling*, Springer.

[34] Pliska S. (1997), *Introduction to Mathematical Finance*, Blackwell, Oxford.

[35] P. Protter, *Stochastic integration and differential equations*, Springer, Berlin, 1990.

[36] Revuz D. and Yor M. (1991), *Continuous Martingales and Brownian Motion*, Springer Verlag.

[37] Schachermayer W. (1992), A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete-time, *Insurance : Mathematics and Economics*, 11, 291-301.

[38] W. Schoutens, E. Simons, and J. Tistaert, *A perfect calibration ! now what ?*, Wilmott Magazine, March (2004).

[39] A. Shiryaev (1989), *Probability*, 2nd Edition. Graduate Texts in Mathematics 95, Springer-Verlag.

[40] Shiryaev A. (1998), *Essentials of Stochastic Finance: Facts, Models and Theory*. World Scientific.

[41] Touzi, N. (2009), *Chaînes de Markov et martingales en temps discret*, Polycopié MAP432, Ecole Polytechnique, Paris.

[42] Williams D. (1991), *Probability with martingales*, Cambridge Mathematical Textbooks.

[43] Yamada T. and Watanabe S. (1971), On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* 11, p155-167.