COMPUTATIONAL FINANCE

Stéphane CRÉPEY, Évry University, France
stephane.crepey@univ-evry.fr

March 21, 2008

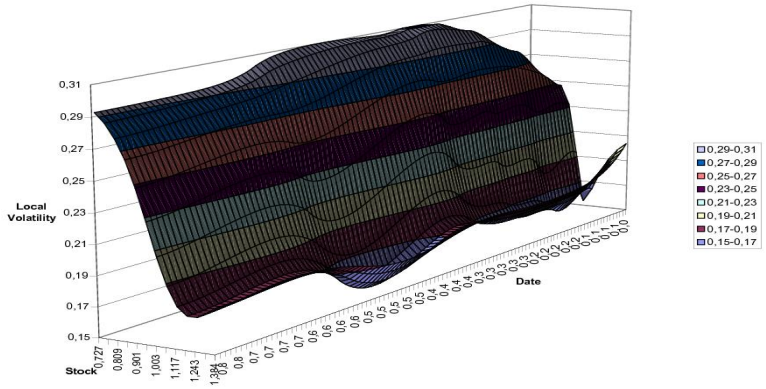

Figure 1: *DAX Index Effective Volatility, June 1 2001*

Contents

Note to the Reader: *Parts IV, V and VI rely to a significant extent on the public releases of the option pricing software and documentation system PREMIA developed since 1999 by the MATHFI project at INRIA and CERMICS, France* (see `www.premia.fr`).

Part I
Introduction and Preliminaries

1 Outline

These notes bear on computational finance (pricing, Greeking and calibration methods), with a focus on *algorithmic* aspects. The related *theoretical results* (*convergence analysis*, etc) are generally stated without proof.

Since the object of these notes is *methods*, not models, we present most methods on simple models, like the *Black–Scholes model* (in general), the *Merton model*, the *Heston model*, etc. Of course the methods themselves are always generic to some degree, hence applicable in a broad range of models to a broad range of financial instruments. In Part II, we recall the basic facts of financial theory necessary to understand how a generic contingent claim pricing equation is derived (equation (57)), in a Markovian risk-neutral primary market model. We then review in Part III the benchmark models on the main derivative markets (equity, interest rate and credit), with the related closed pricing formulae for vanilla derivatives (so no computational methods are required in these models, as far as pricing vanillas is concerned).

In Parts IV and V, we discuss *deterministic* pricing methods (general *finite differences* methods in Part IV, and more specific *tree methods* in Part V). Part VI is about *stochastic simulation* pricing methods (*Monte Carlo methods*).
Note that there is no hermetic frontier between deterministic and stochastic methods. In a sense, Monte Carlo (MC) methods are special cases of tree methods. This is more clearly visible on the problem of pricing by simulation an American option (see section 34.4). Yet it is also true of a standard Monte Carlo algorithm for pricing an European option. Indeed the latter may be interpreted as a one-time-step multi-nomial tree, which provides an exact discretization for the option price in the limit where the number of discretization points in space (tree branches) goes to infinity. In essence, all these numerical schemes are based on the idea of propagating the solution from a surface of the time-space domain on which it is known, along suitable (randomized) 'characteristics' of the problem, in the sense of *Riemann's method of characteristics* for solving hyperbolic first-order equations (see , e.g., [143, Chapter 4]). From the alternative point of view of *control* theory, all these numerical schemes are based on *Bellman's dynamic programming principle* [25].
Of course the difference between tree methods in the usual sense and Monte Carlo methods is that the computation mesh is stochastically generated and unstructured in the case of Monte Carlo methods.

Note that the *prices* of financial instruments are essentially given by the market, and made by offer-and-demand (unless very exotic structures are considered). Market prices are in fact *used* (rather than computed) by models, in the "reverse-engineering" mode that consists in *calibrating* the model to market prices, so that the model be consistent with the market for suitable values of its parameters. This *calibration* process is the object of Part VII. Once calibrated to the market, the model is effectively used for *Greeking* (and/or pricing exotic structures), that is, computing the risk sensitivities of a position, in order to set up a related *hedge*, or complementary position required for off-setting such or such undesired source of

risk.

2 General Set-Up

The evolution of a financial market model is modeled throughout in terms of stochastic processes defined on a continuous time stochastic basis $(\Omega, \mathbb{F}, \widehat{\mathbb{P}})$, where $\widehat{\mathbb{P}}$ denotes the *objective* (or physical, statistical..) probability measure. We may and do assume that the filtration $\mathbb{F}$ satisfies the usual completeness and right-continuity conditions, and that all semimartingales are càdlàg. Finally, since we are always in the context of pricing a contingent claim with maturity T, we further assume that $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ with $\mathcal{F}_0$ trivial and $\mathcal{F}_T = \mathcal{F}$, for simplicity. Moreover, we declare that a *process* on $[0,T]$ (resp. *a random variable*) has to be $\mathbb{F}$-adapted (resp. *$\mathcal{F}$-measurable*), by definition.
We shall typically work under a *risk-neutral* (RN) probability measure $\mathbb{P} \sim \widehat{\mathbb{P}}$, or, more generally, under a martingale probability measure $\mathbb{P}$ relative to a suitable *numeraire*, such that the prices of the primary assets, once properly discounted and adjusted for any dividends, are $\mathbb{P}$ – local martingales. Recall that, under mild technical conditions, existence of such a martingale measure $\mathbb{P}$ is equivalent to a suitable notion of no-arbitrage (NFLVR condition [72], see Part II). In practical applications, it is convenient to think of $\mathbb{P}$ as "the pricing measure chosen by the market" to price a contingent claim. We denote by $\mathcal{T}_t$ (or simply $\mathcal{T}$, in case $t = 0$) the set of $[t,T]$-valued stopping times, and by $\mathbb{E}$ (resp. $\mathbb{E}_t$) the $\mathbb{P}$-expectation (resp. $\mathbb{P}$ – conditional expectation given $\mathcal{F}_t$) operator.

Note that pricing theory can also be developed in discrete time (see, e.g., [87, 124]). The tree (including Monte Carlo) computational methods presented in these notes are directly applicable in this case (in Markovian set-ups), and they are then of course exact in the time direction, since they involve no approximation in time. In particular, these methods can be used in *static* (one-period) set-ups, such as often encountered in *multi-name credit* (like with static *copula models*, see Section 12).

As for *single-name credit* applications, namely the pricing of *defaultable claims* with terminal payoffs of the form $\mathbb{1}_{T<\tau_d}\phi(S_T)$ (or $\mathbb{1}_{\tau<\tau_d}\phi(S_\tau)$ upon exercise at a stopping time τ, in case of American claims), where τ_d represents the *default-time* of a reference entity, it happens that such defaultable claims can be handled in exactly the same way as default-free ones, provided a suitably *credit-risk adjusted* discount factor is used, instead of the usual riskless discount factor (see the references given in the introductory paragraph to Part II). Up to this simple amendment, defaultable contingent claims can be treated in exactly the same way as default-free ones, both from the theoretical and from the practical point of view. This means in particular that all the numerical methods presented in these notes can be used for pricing and Greeking defaultable claims (or calibrating credit-risk models), provided the credit-risk adjusted discount factor is used instead of the original default-free discount factor.
Incidentally, note that the original "default-free" discount factor can itself be interpreted as a default probability (or *killing rate*, in stochastic processes terminology, see, e.g., [161]).

3 Accuracy Requirements and Computational Cost Considerations

A typical benchmark of *accuracy* in computational finance (this varies of course with the application at hand) is a 1bp ($=10^{-4}$) error on *normalized* prices and Greeks, namely prices

and Greeks for a value of the spot scaled to unity.
As for computation times, the benchmark also greatly varies with the application at hand, but as far as 'real-time' option pricing is concerned, 'instantaneous' pricing is the target, and more than half-an-hour computation time (!!) is prohibitive.
Now a resolution within a 1bp normalized error by a finite differences ADI method (the industry standard today as far as deterministic methods are concerned, see Section 18), in space dimension d, typically requires 300 grid points per space dimension (most used numerical schemes are of order 2 of accuracy in space and 1 in time), whence a computation time (and storage cost) as $\mathrm{O}(300^d)$, i.e. a computation time ranging from a few milliseconds for $d = 1$ to twenty minutes or so for $d = 3$. This limits in practice the range of applicability of deterministic pricing methods to problems in space dimension ≤ 3, unless sophisticated sparse grid or grid refinement techniques (mostly available with finite element methods) are used to counter Bellman's 'curse of dimensionality' (referring to the fact that in space dimension d, the computational cost of numerical integration grows exponentially like m_1^d, where m_1 is the number of discretization points in each space direction).

Table 1 provides a crude comparison of the computational costs of typical Monte Carlo and deterministic methods (Monte Carlo algorithm with time discretisation of the underlying factor process, cf. section 33.5, versus ADI PDE method). A rough conclusion (note that we don't detail the *constants* involved in those computational cost estimates) is that deterministic methods are more efficient (but often harder to implement!) in space dimension ≤ 3, otherwise Monte Carlo methods (if applicable) are the best.

	Number of Operations	**Convergence rate**	**Memory Cost**
MC	$\mathrm{O}(nm)$	$\mathrm{O}(n^{-1} + m^{-\frac{1}{2}})$	$\mathrm{O}(1)$
PDE	$\mathrm{O}(nm)$	$\mathrm{O}(n^{-1} + m^{\frac{-2}{d}})$	$\mathrm{O}(m)$

Table 1: *Compared computational costs of Stochastic methods (m simulation runs) versus Deterministic methods (m_1 mesh points per space dimension, i.e. $m = m_1^d$ space mesh points), in space dimension d; n is the number of discretization points in time.*

This leads to the following dictionary (Table 2) of the method to use, depending on the space-dimension and the nature of a pricing problem.
In the upper left corner of this Table, the choice between PDE and Monte Carlo pricing methods should be dictated by the relative interest of performance level with respect to implementation cost and risk, generally higher with deterministic methods.
In the lower right corner of the Table, FBSDE (Forward-Backward Stochastic Differential Equations) methods refers to specific Monte Carlo methods which are available for control problems, cf. the introductory paragraph to Part VI. These methods are not treated in these notes.
Note that our recommendations in the case of American Problems are in fact valid for more general *Control Problems*, e.g., for pricing *game contingent claims*, like *convertible bonds* (see [28, 31]), or for pricing problems in (classes of) model(s) in which the spot volatility is stochastic, but known to remain in a range $(\underline{\sigma}, \overline{\sigma})$ for sure, where $\underline{\sigma}$ and $\overline{\sigma}$ are given constants or functions of (t, S) (Uncertain Volatility model [15]), etc.

Finally an interesting issue hardly mentioned in these notes (see Figure 5(b) p. 65 and Table 3 p. 146, however) is *parallel computing*. Provided dedicated machines or/and networks are used, parallel computing allows one to divide computation time by a factor going from 3 or

	European Problem	**American Problem**
$d \leq 3$	PDE or MC	PDE
$d > 3$	MC	FBSDE

Table 2: *Which method to choose?*

4 to 100 or 1000, depending on the application at hand.

4 Bibliographic Guidelines

Bibliographic references are given along the text and gathered at the end of this document. Various aspects of computational finance are already introduced in Hull's textbook [100] (see in particular Chapter 16 "Numerical procedures"). Here are some more comprehensive references, among others:

- On financial modeling (Part II): [72, 144];
- On market models (Part III): [92, 47, 159, 144, 34, 167];
- On deterministic pricing methods (Parts IV and V): [171, 180, 143, 1, 14, 122, 124, 181, 26];
- On stochastic simulation methods (Part VI): [94, 126, 37, 118];
- On calibration of financial models (Part VII): [57, 82, 149].

Finally, let us mention some related *web resources*:

`www-rocq.inria.fr/mathfi/Premia/index.html`
`screpey.free.fr`
`www.mathfinance.de/frontoffice.html`
`quantlib.org`
`www.nr.com`
`www.gro.creditlyonnais.fr`
`www.iro.umontreal.ca/~lecuyer`
`www.optioncity.net`
`www.defaultrisk.com`

Part II
Martingale Modeling

This Part may be skipped at first reading.

The material in this part is contained in Bielecki et al. [28, 30], where a more general notion of *defaultable option* is considered. Here we only consider the case of default-free European and American options. Note that defaultable claims with terminal payoffs like $\mathbb{1}_{T<\tau_d}\phi(S_T)$ (or $\mathbb{1}_{\tau<\tau_d}\phi(S_\tau)$ upon exercise at a stopping time τ, in case of American claims), where τ_d represents the *default-time* of a reference entity, can be handled in exactly the same way as below, provided the default-free discount factor process β is replaced by a *credit-risk adjusted* discount factor $\alpha = \beta G$, where G_t represents a suitable *survival probability* beyond t (see [28, 29, 30, 31]).

5 Arbitrage Theory

To model a derivative with maturity T, we consider, in the abstract set-up of Section 2, a primary market composed of the savings account B and of d primary risky assets such that on $[0,T]$:
• the *discount factor* process β, that is, the inverse of the savings account B, is a finite variation, continuous, positive and bounded process, with $\beta_0 = 1$;
• the risky assets are locally bounded semimartingales.

The discount factor β is supposed to be absolutely continuous with respect to the Lebesgue measure, namely $\beta_t = \exp(-\int_0^t r_u\,du)$ for a bounded from below *short-term interest rate* process r.

The primary risky assets, with $\mathbb{R}^d$-valued price process X, may pay dividends, whose cumulative value process, denoted by $\mathcal{D}$, is assumed to be an $\mathbb{R}^d$-valued process of finite variation. Given the price process X, we define the *cumulative price* $\widehat{X}$ of the asset as

$$\widehat{X}_t = X_t + \beta_t^{-1}\int_{[0,t]}\beta_u\,d\mathcal{D}_u. \tag{1}$$

In the financial interpretation, the last term in (1) represents the current value at time t of all dividend payments of the asset over the period $[0,t]$, under the assumption that all dividends are immediately reinvested in the savings account B.

Definition 5.1 A *primary trading strategy* (ζ^0,ζ) built on the primary market is an $\mathbb{R}\times\mathbb{R}^{1\otimes d}$-valued process, with ζ predictable and locally bounded, representing the number of units held in the savings account and in the primary risky assets, respectively. The related *wealth process* V is thus given as, for $t\in[0,T]$:

$$V_t = \zeta_t^0 B_t + \zeta_t X_t\ ,\ t\in[0,T] \tag{2}$$

Accounting for dividends, we say that the strategy is *self-financing* if

$$\boxed{dV_t = \zeta_t^0 dB_t + \zeta_t\,(dX_t + d\mathcal{D}_u)}$$

or, equivalently[1]

$$\boxed{d(\beta_t V_t) = \zeta_t \, d(\beta_t \widehat{X}_t)} \tag{3}$$

If, moreover, the discounted wealth process βV is bounded from below, we say that the strategy is *admissible*.

Given the initial wealth V_0 of a *self-financing* primary trading strategy and the strategy ζ in the primary risky assets, the related wealth process is given by, for $t \in [0, T]$:

$$\boxed{\beta_t V_t = \beta_0 V_0 + \int_0^t \zeta_u \, d(\beta_u \widehat{X}_u)} \tag{4}$$

and the process ζ^0 (number of units held in the savings account) is thus uniquely determined as

$$\zeta^0_t = \beta_t \left(V_t - \zeta_t X_t \right) \ .$$

In the sequel we restrict ourselves to self-financing trading strategies. We thus may and do ***redefine*** a (self-financing) primary primary trading strategy as a pair (V_0, ζ), for an initial wealth $V_0 \in \mathbb{R}$ and a $\mathbb{R}^{1 \otimes d}$-valued predictable locally bounded primary strategy in the risky assets ζ, with related wealth process V defined by (4).

We assume that the primary market model is free of arbitrage opportunities (though presumably incomplete), in the sense that the so-called *No Free Lunch with Vanishing Risk* (NFLVR) condition is satisfied. This is a specific no arbitrage condition involving wealth processes of admissible self-financing primary trading strategies (see [72]).
By the now classic arbitrage theory (see, e.g., [72, 53, 28]), the NFLVR condition in a perfect market (without transaction costs, in particular) is equivalent to the existence of a *risk-neutral measure* $\mathbb{P} \in \mathcal{M}$, where $\mathcal{M}$ denotes the set of probability measures $\mathbb{P} \sim \widehat{\mathbb{P}}$ such that $\beta \widehat{X}$ is a $\mathbb{P}$ – local martingale.

Definition 5.2 **(i)** An *European derivative* is a financial product with bounded variation dividend process $D = (D_t)_{t \in [0,T]}$, and with *payment at maturity* T, as seen from the perspective of the option holder, ξ, where ξ denotes a bounded from below real random variable.
(ii) An *American derivative* is a financial product with bounded variation dividend process $D = (D_t)_{t \in [0,T]}$, and with *payment at terminal time* $t \in [0, T]$ *at the holder's convenience* given by, as seen from the perspective of the option holder,

$$\mathbb{1}_{\{t<T\}} L_t + \mathbb{1}_{\{t=T\}} \xi \ , \tag{5}$$

where:
• the *early exercise payment* process $L = (L_t)_{t \in [0,T]}$ is a real-valued, bounded from below, càdlàg process;
• the *payment at maturity* T, ξ, is a bounded from below random variable.

In the simplest case, $D = 0$ and $\xi = (S_T - K)^{\pm}$, for an European vanilla call/put option with maturity T and strike K on $S = X^1$, the first primary risky asset.

[1] This equivalence is very general (cf. Section 8), and it is an easy exercise in the present context where β is a finite variation and continuous process.

Theorem 5.1 **(i)** *For any $\mathbb{P} \in \mathcal{M}$, the process $\Pi = (\Pi_t)_{t\in[0,T]}$ defined by*

$$\boxed{\beta_t \Pi_t = \mathbb{E}_t \int_t^T \beta_u \, dD_u + \beta_T \xi, \quad t \in [0,T]} \tag{6}$$

is an arbitrage price of the related European derivative. Moreover, any arbitrage price is of this form provided

$$\sup_{\mathbb{P}\in\mathcal{M}} \mathbb{E} \int_{[0,T]} \beta_u \, dD_u + \beta_T \xi < \infty \, . \tag{7}$$

(ii) *For any $\mathbb{P} \in \mathcal{M}$, the process $\Pi = (\Pi_t)_{t\in[0,T]}$ defined by*

$$\boxed{\beta_t \Pi_t = \operatorname{esssup}_{\tau\in\mathcal{T}_t} \mathbb{E}_t \int_t^\tau \beta_u \, dD_u + \beta_\tau \big(\mathbb{1}_{\{\tau<T\}} L_\tau + \mathbb{1}_{\{\tau=T\}} \xi \big), \quad t \in [0,T]} \tag{8}$$

is an arbitrage price of the related American derivative as soon as it is a semimartingale. Moreover, any arbitrage price is of this form provided

$$\sup_{\mathbb{P}\in\mathcal{M}} \mathbb{E} \sup_{t\in[0,T]} \int_{[0,t]} \beta_u \, dD_u + \beta_t \big(\mathbb{1}_{\{t<T\}} L_t + \mathbb{1}_{\{t=T\}} \xi \big) < \infty \, . \tag{9}$$

In view of this result, one may interpret an European derivative as a special case of an American derivative with $\beta L = -(c+1)$, where $-c$ is a minorant of $\int_t^T \beta_u \, dD_u + \beta_T \xi$. Henceforth, by default, 'option' means American derivative, including European derivative (with L as above) as a special case. Arbitrage prices like (6)–(8) are called $\mathbb{P}$ – *arbitrage prices*.

6 Connection with Hedging

Given a risk-neutral measure $\mathbb{P} \in \mathcal{M}$, we shall now postulate suitable integrability and regularity conditions *embedded in the standing assumption that a related reflected Backward Stochastic Differential Equation (BSDE, see, e.g., [81]) has a solution*. We shall thus introduce a reflected BSDE $(\mathcal{E})$ under the probability measure $\mathbb{P}$, with data defined in terms of those of a derivative. Assuming that $(\mathcal{E})$ has a solution (for which various sets of sufficient regularity and integrability conditions are known in the literature [98, 68]), we shall deduce explicit hedging strategies with minimal initial wealth for the related derivative.

We assume further that $dD_t = C_t dt$[2], for some progressively measurable time-integrable process C, and we consider the following *reflected BSDE* $(\mathcal{E})$ with data ξ, L :

$$\boxed{\begin{array}{c} \Pi_t = \xi + \int_t^T (C_u - r_u \Pi_u) du + (K_T - K_t) - (M_T - M_t), \quad t \in [0,T] \\ L_t \le \Pi_t, \quad t \in [0,T] \\ \int_0^T (\Pi_u - L_u) \, dK_u = 0 \end{array}} \tag{$\mathcal{E}$}$$

Definition 6.1 By a *($\mathbb{P}$-)solution* to $(\mathcal{E})$, we mean a triplet (Π, M, K) such that all conditions in $(\mathcal{E})$ are satisfied, where:

[2]Note that even if it is not so a priori, it is most of the time possible to transform the pricing problem into an equivalent pricing problem for which this is satisfied, see [30].

• the *state process* Π is a real valued, càdlàg process,
• M is a ($\mathbb{P}$-)martingale vanishing at time 0,
• K is a non-decreasing continuous process (null at time 0).

Note that the first line of $(\mathcal{E})$ is equivalent to

$$\boxed{\beta_t \Pi_t = \beta_T \xi + \int_t^T \beta_u C_u du + \int_t^T \beta_u (dK_u - dM_u), \quad t \in [0,T]} \tag{10}$$

Under mild conditions a solution to $(\mathcal{E})$ exists (and is also unique), with furthermore $K = 0$ in the case of an European derivative. Let henceforth (Π, M, K) denote a solution to $(\mathcal{E})$, with $K = 0$ in the case of an European derivative.

Theorem 6.1 *Π is the $\mathbb{P}$-price process of the related derivative.*

Proof. If (Π, M, K) is a solution to $(\mathcal{E})$, then Π is a semimartingale, and by a standard verification principle, we have that

$$\beta_t \Pi_t = \text{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E}_t \int_t^\tau \beta_u \, C_u du + \beta_\tau \left(\mathbb{1}_{\{\tau < T\}} L_\tau + \mathbb{1}_{\{\tau = T\}} \xi \right) \ , \tag{11}$$

which is also equal to $\mathbb{E}_t \int_t^T \beta_u \, C_u du + \beta_T \xi$, in the case of an European derivative. So the state process Π corresponds to the $\mathbb{P}$-price process of the derivative at hand. □

We shall now see that under mild technical conditions, the $\mathbb{P}$-prices of Theorem 6.1 can be connected with a suitable notion of hedging.

The issuer of a financial derivative immediately sets up a *primary hedging strategy* such that the corresponding wealth process reduces to a *residual cost* (or hedging error) Q, after accounting for the 'dividend cost' $-D$ and for the 'terminal loss' given by $-L$ or $-\xi$. The initial wealth V_0 may then be used as a safe issuer price, up to the hedging error Q, for the derivative at hand.

Definition 6.2 An (issuer) *hedge with residual cost* is a triple (V_0, ζ, Q), where:
• (V_0, ζ) is a (self-financing) primary trading strategy (ζ being called in this context the *hedging strategy*),
• the *residual cost* (or *hedging error*) Q is a real-valued semimartingale with $Q_0 = 0$,
such that the wealth process V of the strategy (V_0, ζ) satisfies, for $t \in [0,T]$:

$$\boxed{\beta_t V_t + \int_0^t \beta_u dQ_u \geq \int_0^t \beta_u dD_u \ + \ \beta_t \Big(\mathbb{1}_{\{t<T\}} L_t + \mathbb{1}_{\{t=T\}} \xi \Big)} \tag{12}$$

Hedges *with no residual cost* (that is, with $Q = 0$) are also called *perfect (super-)hedges*.

In the special case of European derivatives, if moreover equality holds in (12) at $t = T$, so

$$\boxed{\beta_T V_T + \int_0^T \beta_u dQ_u = \int_0^T \beta_u dD_u \ + \ \beta_T \xi} \tag{13}$$

we then deal with a *replicating strategy with residual cost* Q, or simply, in case $Q = 0$, a *perfect replicating strategy*.

Remark 6.3 **(i)** The hedging error Q can also be interpreted as a *financing cost*, that is, cash added ($dQ \geq 0$) or withdrawn ($dQ \leq 0$) from the portfolio in order to get a perfect (but no more self-financing) hedge (and which in the special case where ρ is a $\mathbb{F}$-martingale corresponds to a *mean self-financing* hedge in the sense of Schweizer [169]).
(ii) In relation with admissibility issues (see the end of Definition 5.1), note that the l.h.s. of (12) (discounted wealth process with financing costs included) is bounded from below, for any hedge with residual cost (V_0, ζ, Q).

Now, for any hedging strategy ζ, we define the (discounted) *Profit and Loss* (or *Tracking Error*) process $(e_t)_{t\in[0,T]}$ relative to the price process Π of Theorem 6.1 by setting, for $t \in [0,T]$:

$$\boxed{\beta_t e_t = \Pi_0 - \int_0^t \beta_u C_u du + \int_0^t \zeta_u d(\beta_u \widehat{X}_u) - \beta_t \Pi_t = \int_0^t \left(-d(\beta_u \widehat{\Pi}_u) + \zeta_u d(\beta_u \widehat{X}_u) \right)} \tag{14}$$

where we set (cf. (1))

$$\beta_t \widehat{\Pi}_t := \beta_t \Pi_t + \int_0^t \beta_u C_u du \ .$$

Note that βe is a special semimartingale, by (10). Let further (the $\mathbb{P}$-local martingale) ρ be such that $\rho_0 = 0$ and $\int_0^{\cdot} \beta_t d\rho_t$ is the local martingale component of the special semimartingale βe, so (cf. (14)–(10))

$$\boxed{\beta_t \, d\rho_t = \beta_t \, dM_t - \zeta_t d(\beta_t \widehat{X}_t)} \tag{15}$$

$$\boxed{\beta_t e_t = \int_0^t \beta_u dK_u - \int_0^t \beta_u d\rho_u} \tag{16}$$

and thus in particular

$$\boxed{\beta_t e_t = -\int_0^t \beta_u d\rho_u} \tag{17}$$

in the case of an European derivative with $K = 0$.

Theorem 6.2 **(i)** *For any hedging strategy ζ, (Π_0, ζ, ρ), is an hedge with $\mathbb{P}$ – local martingale residual cost;*
(ii) *Π_0 is the minimal initial wealth of an hedge with $\mathbb{P}$ – local martingale residual cost;*
(iii) *In the special case of an European derivative with $K = 0$, then (Π_0, ζ, ρ) is a replicating strategy with $\mathbb{P}$ – local martingale residual cost. Π_0 is thus also the minimal initial wealth of a replicating strategy with $\mathbb{P}$ – local martingale residual cost.*

Remark 6.4 **(i)** Theorem 6.2 thus *characterizes* the $\mathbb{P}$-price (*arbitrage price* relative to the risk-neutral measure $\mathbb{P}$) of a derivative as the *least initial wealth of a hedge* with $\mathbb{P}$ –local martingale residual cost, under the assumption that the related reflected BSDE $(\mathcal{E})$ has a solution (for related results, see also Föllmer and Sondermann [110] or Schweizer [169]);
(ii) The special case $\rho = 0$ in the previous results corresponds to a suitable form of model completeness (replicability of European options, cf. point (iii) of the theorem), in which the issuer of the option *may and wishes to* hedge all the risks embedded in the option.
The case where $\rho \neq 0$ corresponds to either model incompleteness, or a situation of model completeness in which the issuer may but *wishes not to* hedge all the risks embedded in the product at hand, for instance because she wants to limit transaction costs, or because she

wishes to take some bets in specific risk directions.
(iii) In case where ρ may be taken equal to 0 in Theorem 6.2, the minimality statements in this Theorem may be used to prove uniqueness of the related arbitrage prices (see [29]).

Proof of Theorem 6.2. **(i)** One must show that, for any $t \in [0,T]$:

$$\Pi_0 + \int_0^t \zeta_u d(\beta_u \widehat{X}_u) + \int_0^t \beta_u d\rho_u \geq \int_0^t \beta_u C_u du + \beta_t \Big(\mathbb{1}_{\{t<T\}} L_t + \mathbb{1}_{\{t=T\}} \xi \Big) \tag{18}$$

or, using (15):

$$\Pi_0 + \int_0^t \beta_u dM_u \geq \int_0^t \beta_u C_u du + \beta_t \Big(\mathbb{1}_{\{t<T\}} L_t + \mathbb{1}_{\{t=T\}} \xi \Big) \tag{19}$$

where by the first line in $(\mathcal{E})$:

$$\Pi_0 + \int_0^t \beta_u dM_u = \beta_t \Pi_t + \int_0^t \beta_u C_u du \int_0^t + \beta_u dK_u$$

Using also the facts that $\Pi_T = \xi$ and $\Pi_t \geq L_t$ (terminal condition and obstacle condition in $(\mathcal{E})$), (19) follows by non-negativity of K.
(ii) There exists an (actually, infinitely many) hedge(s) with initial wealth Π_0 and $\mathbb{P}$ – local martingale residual cost Π_0, by (i). Moreover, for any hedge (V_0, ζ, Q) with $\mathbb{P}$ – local martingale residual cost, one has for every $t \in [0,T]$:

$$V_0 + \int_0^t \zeta_u d(\beta_u \widehat{X}_u) + \int_0^t \beta_u dQ_u \geq \int_0^t \beta_u C_u du + \beta_t \Big(\mathbb{1}_{\{t<T\}} L_t + \mathbb{1}_{\{t=T\}} \xi \Big) \tag{20}$$

The l.h.s. is thus a bounded from below local martingale, hence it is a supermartingale. So, by taking expectations in (20):

$$V_0 \geq \mathbb{E} \int_0^t \beta_u C_u du + \beta_t \Big(\mathbb{1}_{\{t<T\}} L_t + \mathbb{1}_{\{t=T\}} \xi \Big) \ .$$

Since this holds for every $t \in [0,T]$, it also holds for every stopping time. Hence $V_0 \geq \Pi_0$ follows, by (8).
(iii) In the special case of an European derivative, the stated results follow by setting $K = 0$ in the previous points of the proof. □

7 Markovian Set-Up

7.1 Markovian FBSDE Approach

For being usable in practice, a (dynamic) pricing model needs to be constructive, or *Markovian* in some sense, relatively to a given derivative. This will be achieved by assuming that the related BSDE $(\mathcal{E})$ is *Markovian* (see, e.g., [66] or Section 4 of [81]).

Remark 7.1 An intuitive criterion to check that a given vector-process has the Markov property consists in verifying that the increment of the process on the time interval $(t, t+h)$ may be simulated by using the value of the process at time t only, without knowledge of any values of the process before t.

Definition 7.2 We say that the BSDE $(\mathcal{E})$ is a *decoupled Markovian Forward-Backward SDE* (Markovian FBSDE, for short), if the input data r, C, ξ and L of $(\mathcal{E})$ are given by Borel-measurable functions of some $\mathbb{P}$-Markov *factor* process $\mathcal{Z}$ with values in a suitable state space (finite-dimensional state space with first component given by time t), so

$$\boxed{r_t = r(\mathcal{Z}_t)\,,\ C_t = C(\mathcal{Z}_t)\,,\ \xi = \xi(\mathcal{Z}_T)\,,\ L_t = L(\mathcal{Z}_t)} \tag{21}$$

(where the related functions are denoted by the same symbols as the corresponding processes or r.v.).

In particular, the system made of the specification of a forward dynamics for $\mathcal{Z}$, together with the BSDE $(\mathcal{E})$, constitutes a decoupled *Markovian forward-backward system of equations* in $(\mathcal{Z}, \Pi, M, K)$. The system is decoupled in the sense that the forward component of the system serves as an input for the backward component ($\mathcal{Z}$ is an input to $(\mathcal{E})$, cf. (21)), but not the other way round.

From the point of view of interpretation, the components of $\mathcal{Z}$ are observable *factors*. The first component of $\mathcal{Z}$ (indexed by 0) is $\mathcal{Z}^0_t = t$. As for the other components of $\mathcal{Z}$, they are typically intimately, though non-trivially, connected with the primary risky asset price process X, as follows:

• Most factors are typically given as primary price processes. The components of $\mathcal{Z}$ that are not included in X (if any) are to be understood as simple factors that may be required to 'Markovianize' the payoffs of the derivative (factors accounting for path dependence in the derivative's payoff and/or non-traded factors such as stochastic volatility in the dynamics of the assets underlying the derivative);

• Some of the primary price processes may not be needed as factors, but are used for hedging purposes.

Note that, due to the nature of our model, observability of the factor process $\mathcal{Z}$ in the mathematical sense of $\mathbb{F}$-adaptedness is not sufficient in practice. In order for the model to be usable in practice, a constructive *mapping* from a collection of meaningful and directly observable economic variables to $\mathcal{Z}$ is really needed. Otherwise, the model will be useless.

Under a rather generic specification for the Markov factor process $\mathcal{Z}$, we shall now derive a related *variational inequality approach* for pricing and hedging the derivative.

7.2 Jump–Diffusion Setting with Regimes

7.2.1 Generator

In this view, given an integer q and a finite set $I = \{y^1, \dots, y^k\}$, we define the following linear operator $\mathcal{A}$ acting on regular real-valued functions $\Pi = \Pi(z)$, for $z = (t, x, y) \in E = [0, T] \times \mathbb{R}^q \times I$:

$$\begin{aligned}
&\mathcal{A}\Pi(z) = \sum_{i=1}^{q} b_i(z)\partial_{x_i}\Pi(z) + \frac{1}{2}\sum_{i,j=1}^{q} a_{ij}(z)\partial^2_{x_i x_j}\Pi(z) \\
&\quad + \int_{\mathbb{R}^q}\Big(\delta\Pi(z,x') - \partial\Pi(z)\delta(z,x')\Big)\gamma(z)h(z,dx') + \sum_{y'\in I}\Delta\Pi(z,y')\lambda(z)\ell(z,y') \\
&= \tfrac{1}{2}\mathrm{Tr}[a(z)\mathcal{H}\Pi(z)] + \partial\Pi(z)\Big(b(z) - \gamma(z)\overline{\delta}(z)\Big) + \gamma(z)\overline{\delta\Pi}(z) + \lambda(z)\overline{\Delta\Pi}(z)
\end{aligned} \tag{22}$$

where:
• the $a(z)$ are q-dimensional *covariance* matrices, with $a(z) = \sigma(z)\sigma(z)^\mathsf{T}$, for some q-dimensional *dispersion* matrices $\sigma(z)$;
• the $b(z)$ are q-dimensional *drift* vector coefficients;
• the *jump intensity function* $\gamma(z)$ is non-negative, the $h(z,\cdot)$ are *conditional jump probability measures* on $\mathbb{R}^q$, and the $\delta(z,x')$ are *jump size functions*, which are supposed to be bounded w.r.t x', locally uniformly in z;
• the *regime switching intensity function* $\lambda(z)$ is non-negative, and the $\ell(z,y')$ for $y' \neq y$ define *regime switching conditional probabilities*;
• $\partial\Pi$ (resp. $\mathcal{H}\Pi$) denotes the *row-gradient* (resp. the *Hessian*) of $\Pi(z)$ with respect to x;
• for any (real-valued, vector-valued or matrix-valued) function f (like Π above) on E, $\delta f(z,\cdot)$ and $\Delta f(z,\cdot)$ (or $\delta f(z)$ and $\Delta f(z)$, for short) denote the functions

$$\mathbb{R}^q \ni x' \overset{\delta f(z)}{\longrightarrow} f(t, x+\delta(z,x'), y) - f(z)\ ,\ I \ni y' \overset{\Delta f(z)}{\longrightarrow} f(t,x,y') - f(z)\ , \tag{23}$$

and we set

$$\overline{\delta f}(z) = \textstyle\int_{\mathbb{R}^q} \delta f(z,x') h(z,dx')\ ,\ \overline{\Delta f}(z) = \sum_{y'\in I} \Delta f(z,y')\ell(z,y')\ ; \tag{24}$$

• $\delta = \delta f$ and $\Delta = \Delta g$ for f and g given as the projections $z = (t,x,y) \mapsto x$ and $z = (t,x,y) \mapsto y$, respectively, so

$$\mathbb{R}^q \ni x' \overset{\delta(z)}{\longrightarrow} \delta(z,x')\ ,\ I \ni y' \overset{\Delta(z)}{\longrightarrow} y' - y\ , \tag{25}$$

and we set, accordingly, $\overline{\delta}(z) = \int_{\mathbb{R}^q} \delta(z,x')h(z,dx')$, $\overline{\Delta}(z) = \sum_{y'\in I} \big(y'-y\big)\ell(z,y')$.

We define further, for any (real-valued, vector-valued or) matrix-valued functions f and g on E such that the matrix-product fg makes sense

$$\begin{array}{c} \overline{\delta f \delta g}(z) = \int_{\mathbb{R}^q} \big(f(t,x+\delta(z,x'),y) - f(z)\big)\big(g(t,x+\delta(z,x'),y) - g(z)\big) h(z,dx') \\ \overline{\Delta f \Delta g}(z) = \sum_{y'\in I} \big(f(t,x,y') - f(z)\big)\big(g(t,x,y') - g(z)\big)\ell(z,y') \end{array} \tag{26}$$

7.2.2 Dynamics

Under appropriate technical conditions (first of which, standard Lipschitz conditions on the model coefficients, see [66], or see Theorems 4.1 and 5.4 in Chapter 4 of Ethier and Kurtz [85] for abstract conditions regarding the existence and uniqueness of a solution to the related martingale problem with generator $\mathcal{A}$), there exists a stochastic basis $(\Omega, \mathbb{F}, \mathbb{P})$ on $[0,T]$, endowed with:
• a q-dimensional Brownian motion W,
• an *integer-valued random measure* μ (see Jacod and Shiryaev [104, Definition II.1.13 p.68]), and
• an $(\Omega,\mathbb{F},\mathbb{P})$-Markov càdlàg process $\mathcal{Z} = (t, \mathcal{X}, \mathcal{Y})$,
such that:
• $\mathcal{X}$ and $\mathcal{Y}$ cannot jump together;
• The $\mathbb{P}$-compensated martingale measure $\widetilde{\nu}$ of the integer-valued random measure ν on I which counts the transitions $\nu_t(y)$ of $\mathcal{Y}$ to state y between time 0 and time t, is given by

$$\boxed{d\widetilde{\nu}_t(y) = d\nu_t(y) - \mathbb{1}_{\{\mathcal{Y}_t \neq y\}} \lambda(\mathcal{Z}_t)\ell(\mathcal{Z}_t, y)\, dt} \tag{27}$$

whence the following special semimartingale canonical representation for $\mathcal{Y}$:

$$\boxed{d\mathcal{Y}_t = \lambda(\mathcal{Z}_t)\overline{\Delta}_t\, dt + \sum_{y\in I}(y - \mathcal{Y}_{t-})\, d\widetilde{\nu}_t(y)\ ,\ t\in[0,T]} \tag{28}$$

• The $\mathbb{P}$-compensated martingale (random) measure $\widetilde{\mu}$ of μ is given by

$$\boxed{\widetilde{\mu}(dx,dt) = \mu(dt,dx) - \gamma(\mathcal{Z}_t)h(\mathcal{Z}_t,dx)dt} \tag{29}$$

and the $\mathbb{R}^q$-valued process $\mathcal{X}$ satisfies, for $t\in[0,T]$,

$$\boxed{d\mathcal{X}_t = b(\mathcal{Z}_t)\, dt + \sigma(\mathcal{Z}_t)\, dW_t + \int_{\mathbb{R}^q}\delta(\mathcal{Z}_{t-},x)\,\widetilde{\mu}(dt,dx)} \tag{30}$$

Moreover, one has the following estimates, for any $p\in[2,+\infty)$:

$$\sup_{[0,T]}|\mathcal{X}|^p \le C_p\big(1+|x|^p\big)\ . \tag{31}$$

Given a further Borel-measurable function $r = r(z)$, such a factor process $\mathcal{Z}$ and the short-term interest rate process $r_t = r(\mathcal{Z}_t)$ can then be used as starting point in the construction of a risk-neutral primary market model relative to $(\Omega,\mathbb{F},\mathbb{P})$. The primary risky price process X and the related primary dividends $\mathcal{D}$ in (1) may thus be defined in terms of $\mathcal{Z}$ (with for instance $\mathcal{D} = \int_0^{\cdot} d(\mathcal{Z}_t)dt$, for some Borel-measurable Markovian dividend rate function d), under the additional constraint that, consistently with arbitrage requirements (see Section 5), $\beta\,\widehat{X}$ be a locally bounded $\mathbb{P}$ – local martingale, and without forgetting to take care about the availability of a well-defined and constructive mapping between $\mathcal{Z}$ and X (cf. section 7.1).

Remark 7.3 **(i)** If we suppose that the *intensity matrix* $\lambda\ell$ of $\mathcal{Y}$ does not depend on t,x, then $\mathcal{Y}$ is an homogenous Markov chain with finite state space I. Alternatively, if we take $\delta(z,x') = x'$, and we suppose that the coefficients σ,b,γ and h do not depend on z, then $\mathcal{X}$ is a Lévy-Poisson process. This model thus defines a rather generic class of Markovian factor processes $\mathcal{Z} = (t,\mathcal{X},\mathcal{Y})$ for a derivative, in the form of a $\mathcal{Y}$-modulated Lévy-like component $\mathcal{X}$ and an $\mathcal{X}$-modulated Markov chain-like component $\mathcal{Y}$.
(ii) From the point of view of interpretation, $\mathcal{Y}$ represents *regimes* that modulate the dynamics of the risk-neutral pricing process. In order to make the calibration of the model possible, various regimes $y\in I$ should correspond to non-overlapping (vector-valued) sets of model parameters.
For $k=1$, that is, in the case when the regime indicator process is constant, the one-dimensional process $\widetilde{\nu}$ in (27) is trivially null and plays no role whatsoever, so that we may and do redefine k as zero.

For simplicity we do not consider the "infinite activity" case, that is, the case of possibly unbounded measures $\gamma h(z,\cdot)$. Note however that reinforcing our local boundness assumption on the jump size function $\delta(z,x')$ into

$$|\delta(z,x')| \le C(1\wedge|x'|) \tag{32}$$

for some constant C locally uniform in z, then most of the results presented here can be extended to more general *Lévy jump measures* $\gamma h(z,\cdot)$ such that

$$\int_{\mathbb{R}^q} (1 \wedge |x|^2)\gamma(z)h(z,dx) < +\infty \,, \tag{33}$$

provided one works with the form defined by the first equality in (22) for the generator $\mathcal{A}$ of $\mathcal{Z}$ (see, e.g., Barles et al. [18]). Regarding this last reservation note that, under assumptions (32)–(33), $\delta\Pi(z,\cdot)$ and $\delta(z,\cdot)$ in (22) may not be integrable separately with respect to $\gamma h(z,\cdot)$, whereas the difference $\delta\Pi(z,\cdot) - \partial\Pi(z)\delta(z,\cdot)$ is always integrable with respect to $\gamma h(z,\cdot)$.

7.2.3 Elementary Reformulation of the Model

It is possible to derive a maybe more intuitive (yet strictly limited to finite jump measure $\gamma h(z,\cdot)$, cf. Remark 7.3(i)) reformulation of the model dynamics (27) to (30) by introducing the following notation, for $t \in [0,T]$:
• $H_t = \nu(I \times [0,t])$, and I_t, a r.v. on $I \setminus \{\mathcal{Y}_{t-}\}$ with conditional law $\ell(\mathcal{Z}_{t-}, y)$ given $\mathcal{Z}_{t-}$;
• $N_t = \mu(\mathbb{R}^q \times [0,t])$ and J_t, a r.v. on $\mathbb{R}^q$ with conditional law $h(\mathcal{Z}_{t-}, dx)$ given $\mathcal{Z}_{t-}$;
• for any (real-valued, vector-valued or matrix-valued) function f on E,

$$\begin{gathered}\delta f_t = \delta f(\mathcal{Z}_{t-}, J_t) \;,\; \overline{\delta f}_t = \mathbb{E}(\delta f_t \,|\, \mathcal{Z}_{t-}) = \overline{\delta f}(\mathcal{Z}_{t-}) \\ \Delta f_t = \Delta f(\mathcal{Z}_{t-}, I_t) \;,\; \overline{\Delta f}_t = \mathbb{E}(\Delta f_t \,|\, \mathcal{Z}_{t-}) = \overline{\Delta f}(\mathcal{Z}_{t-})\end{gathered}$$

and in particular

$$\begin{gathered}\delta_t = \delta(\mathcal{Z}_{t-}, J_t) \;,\; \overline{\delta}_t = \overline{\delta}(\mathcal{Z}_{t-}) \\ \Delta_t = \Delta(\mathcal{Z}_{t-}, I_t) = I_t - \mathcal{Y}_{t-} \;,\; \overline{\Delta}_t = \overline{\Delta}(\mathcal{Z}_{t-}) \,.\end{gathered}$$

Denoting further by (s_l) and (t_l) the ordered sequence of the (random) times of jumps of ν and μ, respectively (note that we deal with *finite* jump and regime switching measures γh and $\lambda\ell$ without common jumps, by assumption), then we have (cf. (27), (29)):

$$\boxed{\begin{array}{l}\sum_{y\in I} d\widetilde{\nu}_t(y) = dH_t - \lambda(\mathcal{Z}_t)dt \\ \int_{\mathbb{R}^q} \widetilde{\mu}(dx,dt) = dN_t - \gamma(\mathcal{Z}_t)dt\end{array}} \tag{34}$$

and equations (30) and (28) may be rewritten as, respectively:

$$\boxed{\begin{array}{l}d\mathcal{X}_t = \big(b(\mathcal{Z}_t) - \gamma(\mathcal{Z}_t)\overline{\delta}_t\big)\, dt + \sigma(\mathcal{Z}_t)\, dW_t + d\left(\sum_{l=1}^{N_t} \delta_{t_l}\right) \\ d\mathcal{Y}_t = d\left(\sum_{l=1}^{H_t} \Delta_{s_l}\right)\end{array}} \tag{35}$$

7.2.4 Itô formula

The following variant of the Itô formula holds (see Bielecki et al. [29] or Jacod [103, Theorem

3.89 p.109]):

$$\boxed{\begin{aligned} d\Pi(\mathcal{Z}_t) &= (\partial_t + \mathcal{A})\Pi(\mathcal{Z}_t)\,dt + \partial\Pi(\mathcal{Z}_t)\,\sigma(\mathcal{Z}_t)\,dW_t \\ &\quad + \int_{\mathbb{R}^q} \delta\Pi(\mathcal{Z}_{t-}, x)\widetilde{\mu}(dt, dx) + \sum_{y\in I} \Delta\Pi(\mathcal{Z}_{t-}, y)d\widetilde{\nu}_t(y) \\ &= (\partial_t + \mathcal{A})\Pi(\mathcal{Z}_t)\,dt + \partial\Pi(\mathcal{Z}_t)\,\sigma(\mathcal{Z}_t)\,dW_t \\ &\quad + \delta\Pi(\mathcal{Z}_{t-})\widetilde{\mu}(dt, d\cdot) + \Delta\Pi(\mathcal{Z}_{t-})d\widetilde{\nu}_t \end{aligned}} \tag{36}$$

for any sufficiently regular function Π on E. Or, equivalently to (36) (cf. (35)):

$$\begin{aligned} d\Pi(\mathcal{Z}_t) &= (\partial_t + \widetilde{\mathcal{A}})\Pi(\mathcal{Z}_t)\,dt + \partial\Pi(\mathcal{Z}_t)\,\sigma(\mathcal{Z}_t)\,dW_t + d\left(\sum_{l=1}^{N_t} \delta\Pi_{t_l}\right) + d\left(\sum_{l=1}^{H_t} \Delta\Pi_{s_l}\right) \\ &= (\partial_t + \mathcal{A})\Pi(\mathcal{Z}_t)\,dt + \partial\Pi(\mathcal{Z}_t)\,\sigma(\mathcal{Z}_t)\,dW_t \\ &\quad + \left(d\sum_{l=1}^{N_t} \delta\Pi_{t_l} - \gamma(\mathcal{Z}_t)\overline{\delta\Pi}_t dt\right) + \left(d\sum_{l=1}^{H_t} \Delta\Pi_{s_l} - \lambda(\mathcal{Z}_t)\overline{\Delta\Pi}_t dt\right) \end{aligned} \tag{37}$$

with

$$\begin{aligned} \widetilde{\mathcal{A}}\Pi(z) &= \frac{1}{2}\mathrm{Tr}[a(z)\mathcal{H}\Pi(z)] + \partial\Pi(z)\Big(b(z) - \gamma(z)\overline{\delta}(z)\Big) \\ &= \mathcal{A}\Pi(z) - \gamma(z)\overline{\delta\Pi}(z) - \lambda(z)\overline{\Delta\Pi}(z) \end{aligned} \tag{38}$$

7.2.5 Brackets

Let X^c and Y^c, resp. ΔX and ΔY, denote the continuous local martingale components, resp. the jump processes, of two given (real-valued) semimartingales X and Y. Recall that the quadratic covariation or *bracket* $[X, Y]$ is given by

$$d[X, Y]_t = d(X_t Y_t) - X_{t-}dY_t - Y_{t-}dX_t \tag{39}$$

$$= d\langle X^c, Y^c\rangle_t + d\left(\sum_{s\le t} \Delta X_s \Delta Y_s\right) \tag{40}$$

with the initial condition $[X, Y]_0 = 0$. The *sharp bracket* $\langle X, Y\rangle$ corresponds to the *compensator* of $[X, Y]$, which is well defined provided $[X, Y]$ is of locally integrable variation (see, e.g., Protter [157]).

For processes X and Y given as $X_t = X(\mathcal{Z}_t)$ and $Y_t = Y(\mathcal{Z}_t)$ in our jump-diffusion setting with regimes $\mathcal{Z}$, we get by application of (40):

$$d[X, Y]_t = \partial X a(\partial Y)^{\mathsf{T}}(\mathcal{Z}_t)dt + d\left(\sum_{l=1}^{N_t} \delta X_{t_l}\delta Y_{t_l}\right) + d\left(\sum_{l=1}^{H_t} \Delta X_{s_l}\Delta Y_{s_l}\right)$$

which obviously admits the compensator $< X, Y >$ with Lebesgue-density given as (cf. (26)):

$$\boxed{\frac{d<X,Y>}{dt} = \partial X a(\partial Y)^{\mathsf{T}}(\mathcal{Z}_t) + \gamma(\mathcal{Z}_t)\overline{\delta X \delta Y}(\mathcal{Z}_t) + \lambda(\mathcal{Z}_t)\overline{\Delta X \Delta Y}(\mathcal{Z}_t)} \tag{41}$$

Moreover, it comes by application of the Itô formula (36) to the functions X, Y and XY, with "$\triangleq$" standing for "equality up to a local martingale term":

$$\begin{aligned} d[X,Y]_t &= d(X_tY_t) - X_{t-}dY_t - Y_{t-}dX_t \\ &\triangleq (\partial_t + \mathcal{A})(XY)(\mathcal{Z}_t)dt - X(\mathcal{Z})(\partial_t + \mathcal{A})Y(\mathcal{Z}_t)dt - Y(\mathcal{Z})(\partial_t + \mathcal{A})X(\mathcal{Z}_t)dt \ . \end{aligned}$$

This comes out onto the following alternative characterization of the compensator $\langle X,Y\rangle$ of $[X,Y]$ (to be compared with (39)):

$$\boxed{d < X,Y >_t = (\partial_t + \mathcal{A})(XY)(\mathcal{Z}_t)dt - X_t(\partial_t + \mathcal{A})Y(\mathcal{Z}_t)dt - Y_t(\partial_t + \mathcal{A})X(\mathcal{Z}_t)dt} \tag{42}$$

We are now ready to prove the following

Proposition 7.1 *For processes X and Y given as $X_t = X(\mathcal{Z}_t)$ and $Y_t = Y(\mathcal{Z}_t)$ in our jump-diffusion setting with regimes $\mathcal{Z}$, the sharp bracket $\langle X,Y\rangle$ is absolutely continuous w.r.t. the Lebesgue measure, with related density*

$$\boxed{\frac{d\langle X,Y\rangle}{dt} = \lim_{h\to 0} h^{-1}\mathbb{C}\mathrm{ov}_t(X_{t+h} - X_t, Y_{t+h} - Y_t)} \tag{43}$$

Proof. For any fixed $h > 0$, we have:

$$\begin{aligned} &\mathbb{C}\mathrm{ov}_t(X_{t+h} - X_t, Y_{t+h} - Y_t) + \mathbb{E}_t(X_{t+h} - X_t)\mathbb{E}_t(Y_{t+h} - Y_t) = \\ &\quad \mathbb{E}_t\left(X_{t+h}Y_{t+h} - X_tY_t\right) - X_t\mathbb{E}_t\left(Y_{t+h} - Y_t\right) - Y_t\mathbb{E}_t\left(X_{t+h} - X_t\right) \ . \end{aligned} \tag{44}$$

Now, we have by the Itô formula (36) applied with $\Pi = X$, $\Pi = Y$ and $\Pi = XY$, respectively:

$$\begin{aligned} \lim_{h\to 0} h^{-1}\mathbb{E}_t(X_{t+h} - X_t) &= (\partial_t + \mathcal{A})X(\mathcal{Z}_t) \\ \lim_{h\to 0} h^{-1}\mathbb{E}_t(Y_{t+h} - Y_t) &= (\partial_t + \mathcal{A})Y(\mathcal{Z}_t) \\ \lim_{h\to 0} h^{-1}\mathbb{E}_t\left(X_{t+h}Y_{t+h} - X_tY_t\right) &= (\partial_t + \mathcal{A})(XY)(\mathcal{Z}_t) \end{aligned}$$

Hence, by (44):

$$\begin{aligned} \lim_{h\to 0} h^{-1}\mathbb{C}\mathrm{ov}_t(X_{t+h} - X_t, Y_{t+h} - Y_t) &= (\partial_t + \mathcal{A})(XY)(\mathcal{Z}_t) \\ -X_t(\partial_t + \mathcal{A})Y(\mathcal{Z}_t) - Y_t(\partial_t + \mathcal{A})X(\mathcal{Z}_t) &= \frac{d\langle X,Y\rangle}{dt} \ , \end{aligned}$$

by (42). □

7.2.6 Special Cases

Jump–diffusions and Lévy-like processes We stated the generic model $\mathcal{Z}$ above for the sake of generality of a factor process underlying a financial derivative. Such a level of generality is actually useful for applications in credit risk modeling (see, e.g., [32]). However in most applications the process $\mathcal{Y}$ is trivial ($k = 0$), in which case $\mathcal{Z}$ reduces to $(t, \mathcal{X})$, and the general Itô formula (36) reduces to the following *Itô-Lévy formula*:

$$\boxed{d\Pi(\mathcal{Z}_t) = (\partial_t + \mathcal{A})\Pi(\mathcal{Z}_t)\,dt + \partial\Pi(\mathcal{Z}_t)\,\sigma(\mathcal{Z}_t)\,dW_t + \delta\Pi(\mathcal{Z}_{t-})\widetilde{\mu}(dt, d\cdot)} \tag{45}$$

with

$$\boxed{\begin{array}{l}\mathcal{A}\Pi(z) = \frac{1}{2}\sum_{i,j=1}^{q} a_{ij}(z)\partial^2_{x_i x_j}\Pi(z) + \sum_{i=1}^{q} b_i(z)\partial_{x_i}\Pi(z) \\ \quad + \int_{\mathbb{R}^q}\Big(\delta\Pi(z,x') - \partial\Pi(z)\delta(z,x')\Big)\gamma(z)h(z,dx') \\ = \frac{1}{2}\mathrm{Tr}[a(z)\mathcal{H}\Pi(z)] + \partial\Pi(z)\Big(b(z) - \gamma(z)\overline{\delta}(z)\Big) + \gamma(z)\overline{\delta\Pi}(z)\end{array}} \tag{46}$$

Or, equivalently to (45) (cf. (37)–(38)):

$$\boxed{d\Pi(\mathcal{Z}_t) = (\partial_t + \widetilde{\mathcal{A}})\Pi(\mathcal{Z}_t)\,dt + \partial\Pi(\mathcal{Z}_t)\,\sigma(\mathcal{Z}_t)\,dW_t + d\left(\sum_{l=1}^{N_t}\delta\Pi_{t_l}\right)} \tag{47}$$

with $\widetilde{\mathcal{A}}$ as in (38).

In the context of more general Lévy jump measures $\gamma h(z,\cdot)$ under assumptions (32)–(33), the Itô-Lévy formula (45) still holds, with $\mathcal{A}$ therein to be understood as defined by the first identity in (46) (since $\overline{\delta\Pi}(z)$ and $\overline{\delta}(z)$ may not exist in the second one, cf. Remark 7.3(i)). Note that in the context of more general Lévy–like processes , the process $\mathcal{X}$ is typically given in the following form (see, e.g., Cont and Tankov [57]):

$$\boxed{d\mathcal{X}_t = b^c(\mathcal{Z}_t)\,dt + \sigma(\mathcal{Z}_t)\,dW_t + d\sum_{l=1}^{N^c_t}\delta(\mathcal{Z}_{t^c_l-}, J_{t^c_l}) + \int_{|x|<1}\delta(\mathcal{Z}_{t-},x)\,\widetilde{\mu}(dt,dx)} \tag{48}$$

for some coefficient (function) b^c and with $N^c_t = \mu(B^c\times[0,t])$, where B^c denotes the complement of the unit ball in $\mathbb{R}^q$, and where the t^c_ls denote the successive times of jumps of $\mu(B^c\times[0,t])$ (which are well defined, in the case of Lévy jump measures $\gamma h(z,\cdot)$). By comparison with (30), we thus have:

$$b(z) = b^c(z) + \int_{|x'|\geq 1}\delta(z,x')\gamma(z)h(z,dx')\ .$$

The following equivalent form of the generator $\mathcal{A}$ in terms of b^c follows (cf. the first identity in (46)):

$$\boxed{\begin{array}{l}\mathcal{A}\Pi(z) = \frac{1}{2}\mathrm{Tr}[a(z)\mathcal{H}\Pi(z)] + \partial\Pi(z)b^c(z) \\ \quad + \int_{\mathbb{R}^q}\Big(\delta\Pi(z,x') - \partial\Pi(z)\delta(z,x')\mathbb{1}_{|x'|<1}\Big)\gamma(z)h(z,dx')\end{array}} \tag{49}$$

Pure Diffusions When there are no jumps in $\mathcal{X}$ either, so

$$\boxed{d\mathcal{X}_t = b(t,\mathcal{X}_t)\,dt + \sigma(t,\mathcal{X}_t)\,dW_t} \tag{50}$$

then (45) reduces further to the *standard Itô formula*:

$$\boxed{d\Pi(t,\mathcal{X}_t) = (\partial_t + \mathcal{A})\Pi(t,\mathcal{X}_t)\,dt + \partial\Pi(t,\mathcal{X}_t)\,\sigma(t,\mathcal{X}_t)\,dW_t} \tag{51}$$

with

$$\boxed{\mathcal{A}\Pi(z) = \tfrac{1}{2}\sum_{i,j=1}^{q} a_{ij}(z)\partial^2_{x_i x_j}\Pi(z) + \sum_{i=1}^{q} b_i(z)\partial_{x_i}\Pi(z)} \tag{52}$$

Continuous Times Markov Chains In other applications the process $\mathcal{X}$ is trivial ($q = 0$), in which case $\mathcal{Z}$ reduces to a Continuous-Time Markov Chain $(t, \mathcal{Y})$, and the general Itô formula (36) reduces to the following elementary *Markov Chains Itô formula*:

$$\boxed{d\Pi(\mathcal{Z}_t) = (\partial_t + \mathcal{A})\Pi(\mathcal{Z}_t)\, dt + \Delta\Pi(\mathcal{Z}_{t-})d\widetilde{\nu}_t} \tag{53}$$

with

$$\boxed{\mathcal{A}\Pi(z) = \lambda(z)\overline{\Delta\Pi}(z)} \tag{54}$$

Or, equivalently to (53) (cf. (37)–(38)):

$$\boxed{d\Pi(\mathcal{Z}_t) = \partial_t\Pi(\mathcal{Z}_t)\, dt + d\left(\sum_{l=1}^{H_t} \Delta\Pi_{s_l}\right)\ .} \tag{55}$$

7.3 Variational Inequality Approach

7.3.1 Reflected BSDEs and PIDEs with obstacles

We are now back to the general factor process $\mathcal{Z} = (t, \mathcal{X}, \mathcal{Y})$ above with related Markovian FBSDE $(\mathcal{E})$ for a derivative. Denote by $\mathcal{P}$ the $\mathbb{F}$-predictable σ-algebra on $\Omega \times [0, T]$ and by $\mathcal{B}(\mathbb{R}^q)$ the Borel σ-algebra on $\mathbb{R}^q$. A solution (Π, M, K) to $(\mathcal{E})$ is then typically sought for with M in the form

$$\boxed{M_t = \int_0^t \widehat{Z}_u\, dW_u + \int_0^t \widetilde{Z}_u\, d\widetilde{\nu}_u + \int_0^t \int_{\mathbb{R}^q} V_u(x)\, \widetilde{\mu}(du, dx)} \tag{56}$$

for $\mathbb{F}$-predictable processes $\widehat{Z}, \widetilde{Z}$, and a $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^q)$-measurable random function $V : \Omega \times [0, T] \times \mathbb{R}^q \to \mathbb{R}$ such that (with the convention that $k = 0$ when there are no regimes):

$$\begin{gathered} \textstyle\sum_{j=1}^q \mathbb{E} \int_0^T (\widehat{Z}_t^j)^2\, dt \\ \textstyle\sum_{j=1}^k \mathbb{E} \int_0^T (\widetilde{Z}_t^j)^2 \lambda(\mathcal{Z}_t)\ell(\mathcal{Z}_t, y^j)\, dt < \infty \\ \mathbb{E} \int_0^T \int_{\mathbb{R}^q} V_t^2(x)\gamma(\mathcal{Z}_t)h(\mathcal{Z}_t, dx)\, dt < \infty \end{gathered}$$

It is shown in [68] that, under mild regularity conditions, $(\mathcal{E})$ has a unique solution (Π, M, K) with M of the form (56), in suitable Hilbert spaces. In the Markovian case, [66, 65] establishes the relation between this solution and the unique solution in some sense (viscosity solution with polynomial growth in the x variable), $\Pi(z)$, to the following *PIDE obstacle problem* (system of k coupled PIDEs with obstacle in space-dimension q):

$$\boxed{\min\left(-\partial_t\Pi - \mathcal{A}\Pi - C + r\Pi, \Pi - L\right) = 0 \text{ on } [0, T) \times \mathbb{R}^q \times I} \tag{57}$$

with terminal condition $\Pi(T, x, y) = \xi(x, y)$. So,

Theorem 7.2 *Under mild conditions, we have*

$$\boxed{\Pi_t = \Pi(\mathcal{Z}_t), \quad t \in [0, T]}$$

and $\Pi_0 = \Pi(\mathcal{Z}_0)$ is the minimal initial wealth of a hedge with $\mathbb{P}$ – local martingale residual cost process for the Claim.

Moreover, in regular cases, we also have in some sense (see [67]), for $t \in [0,T]$:

$$\boxed{\widehat{Z}_t = \partial\Pi\sigma(\mathcal{Z}_t)\ ,\ \widetilde{Z}_t = \Delta\Pi(\mathcal{Z}_{t-})\ ,\ V_t = \delta\Pi(\mathcal{Z}_{t-})}$$

7.3.2 Discussion of Various Hedging Schemes

Let us further assume that the primary risky price process X satisfies likewise $X_t = X(\mathcal{Z}_t)$ for a function X with the same regularity as Π, and that, consistently with the requirement that $\beta\widehat{X}$ is a $\mathbb{P}$ – local martingale:

$$d(\beta_t\widehat{X}_t) = \beta_t\Big(\partial X\sigma(\mathcal{Z}_t)dW_t\ +\ \Delta X(\mathcal{Z}_{t-})d\widetilde{\nu}_t\ +\ \delta X(\mathcal{Z}_{t-})\widetilde{\mu}(dt,d\cdot)\Big) \tag{58}$$

Note that X is an $\mathbb{R}^d$-valued function, so in particular ∂X lives in $\mathbb{R}^{d\otimes q}$, and identity (58) holds in $\mathbb{R}^d$.

The cost ρ relative to the strategy ζ (cf. (15)) can in turn be expressed in terms of the pricing functions Π and X and the related delta functions.

Theorem 7.3 *Under the previous conditions in the Markovian set-up, the dynamics (15) for the cost process ρ relative to the strategy ζ (and thus the related P&L, cf. (16)) may be rewritten as (cf. (23)):*

$$\boxed{\begin{aligned} d\rho_t = &\Big(\partial\Pi\sigma(\mathcal{Z}_t) - \zeta_t\partial X\sigma(\mathcal{Z}_t)\Big)dW_t \\ &+ \Big(\Delta\Pi(\mathcal{Z}_{t-}) - \zeta_t\Delta X(\mathcal{Z}_{t-})\Big)d\widetilde{\nu}_t \\ &+ \Big(\delta\Pi(\mathcal{Z}_{t-}) - \zeta_t\delta X(\mathcal{Z}_{t-})\Big)\widetilde{\mu}(dt,d\cdot) \end{aligned}} \tag{59}$$

It is thus possible to hedge completely the source risk W (which amounts to *hedging market risk*, or *spread risk* in a context of credit risk modeling, see section 13.2.2) by setting, provided $\partial X\sigma$ is left-invertible,

$$\boxed{\zeta_t = \partial\Pi\sigma(\partial X\sigma)^{-1}(\mathcal{Z}_t)} \tag{60}$$

In the simplest case where $q = d$ and ∂X and σ are invertible this formula further reduces to

$$\boxed{\zeta_t = \partial\Pi\partial X^{-1}(\mathcal{Z}_t)} \tag{61}$$

Note that this strategy actually *creates some jump risk* via the dependence in ζ of the remaining terms in (59).

At the other extreme, it is alternatively possible to hedge completely the source risk ν (which typically amounts to *hedging jump risk*, or *default risk*, in a context of credit risk modeling, see section 13.2.2) by setting, provided $\Delta X(\mathcal{Z}_{t-})$ is left-invertible,

$$\boxed{\zeta_t = \Delta\Pi(\mathcal{Z}_{t-})(\Delta X(\mathcal{Z}_{t-}))^{-1}} \tag{62}$$

This strategy creates additional market risk via the dependence in ζ of the remaining terms in (59).

Of course a perfect hedge ($\rho = 0$) is hopeless unless there are no jumps (or only a finite number of jump sizes) in $\mathcal{X}$. In the context of incomplete markets the choice of a hedging strategy is up to one's *optimality criterion*, relative to the hedging cost (15)–(59). For instance, a trader may wish to minimize the (objective, $\widehat{\mathbb{P}}$ –) variance of $\int_0^T \beta_t d\rho_t$. Yet the related strategy $\widehat{\zeta}^{va}$ is hardly accessible in practice (in particular it typically depends on the objective model drift, a quantity notoriously difficult to estimate on financial data). As a proxy to this strategy, traders commonly use the strategy ζ^{va} which minimizes the *risk-neutral* variance of the error. Note that under mild conditions $\int_0^{\cdot} \beta dM$ and $\beta\widehat{X}$ are square integrable martingales, by estimate (31) on $\mathcal{X}$ combined with the polynomial growth of the functions Π and X in x. The risk-neutral minimal variance strategy ζ^{va} is then given by the following *Galtchouk-Kunita-Watanabe decomposition* of $\int_0^{\cdot} \beta dM$ with respect to $\beta\widehat{X}$ (see, e.g., Protter [157, IV.3, Corollary 1]):

$$\beta_t dM_t = \zeta_t^{va} d(\beta_t \widehat{X}_t) \; + \; \beta_t d\rho_t^{va} \tag{63}$$

for some $\mathbb{R}^d$-valued $\beta\widehat{X}$-integrable process ζ^{va} and a real-valued square integrable martingale $\beta_t d\rho_t^{va}$ strongly orthogonal to $\beta\widehat{X}$. Denoting in vector-matrix form $< X, Y >= (< X^i, Y^j >)_i^j$, $< X >=< X, X >$, we thus have by (63):

$$\boxed{\zeta_t^{va} = \tfrac{d<\Pi,X>}{dt}(\tfrac{d<X>}{dt})^{-1}} \tag{64}$$

where, by (41):

$$\boxed{\begin{array}{l} \frac{d<\Pi,X>}{dt} = \partial\Pi a \partial X^{\mathsf{T}}(\mathcal{Z}_t) + \gamma(\mathcal{Z}_t)\overline{\delta\Pi(\delta X)^{\mathsf{T}}}(\mathcal{Z}_t) + \lambda(\mathcal{Z}_t)\overline{\Delta\Pi(\Delta X)^{\mathsf{T}}}(\mathcal{Z}_t) \\ \frac{d<X>}{dt} = \partial X a \partial X^{\mathsf{T}}(\mathcal{Z}_t) + \gamma(\mathcal{Z}_t)\overline{\delta X(\delta X)^{\mathsf{T}}}(\mathcal{Z}_t) + \lambda(\mathcal{Z}_t)\overline{\Delta X(\Delta X)^{\mathsf{T}}}(\mathcal{Z}_t) \end{array}} \tag{65}$$

Remark 7.4 **(i)** For every fixed $t \in [0, T]$ and $h > 0$ it follows from (63) that $(\zeta_u^{va})_{u\in[t,t+h]}$ minimizes

$$\mathbb{V}\mathrm{ar}_t\Big(\int_t^{t+h} \beta_u dM_u - \int_t^{t+h} \zeta_u d(\beta_u d\widehat{X}_u)\Big)$$

over the set of all (self-financing) trading strategies ζ_t on the time interval $[t, t + h]$. Let likewise $\zeta_t^h =: \zeta_t^{va,h}$ minimize

$$\mathbb{V}\mathrm{ar}_t\Big(\int_t^{t+h} \beta_u dM_u - \zeta_t^h \int_t^{t+h} d(\beta_u d\widehat{X}_u)\Big)$$

over the set of all *buy-and-hold* (self-financing) strategies ζ_t^h on the time interval $[t, t + h]$. The strategy $\zeta_t^{va,h}$ is given as the solution of the linear regression problem of $\int_t^{t+h} \beta_u dM_u$ against $\int_t^{t+h} d(\beta_u d\widehat{X}_u)$, so:

$$\boxed{\zeta_t^{va,h} = \mathbb{C}\mathrm{ov}_t\big(\textstyle\int_t^{t+h} \beta_u dM_u, \int_t^{t+h} d(\beta_u d\widehat{X}_u)\big)\mathbb{V}\mathrm{ar}_t\big(\int_t^{t+h} d(\beta_u d\widehat{X}_u)\big)^{-1}}$$

In view of (44) we deduce that $\zeta_t^{va} = \lim_{h\to 0} \zeta_t^{va,h}$, as one could expect.
(ii) In case of the pure diffusion model $\mathcal{X}$ of section 7.2.6 (cf. formulas (50) to (52)), then sharp brackets coincide with (square) brackets and are independent of the equivalent probability measure under consideration. It follows that the risk-neutral minimal variance strategy ζ^{va} defined by (64) satisfies $\zeta_t^{va} = \lim_{h\to 0} \widehat{\zeta}_t^{va,h}$ where the strategies $\widehat{\zeta}_t^{va,h}$ are the

counterpart under the objective probability measure $\widehat{\mathbb{P}}$ of the strategies $\zeta_t^{va,h}$ introduced in part (i). In the case where there are no jumps in the model the risk-neutral minimal variance strategy ζ^{va} is thus also an objective locally (but possibly not globally) minimal variance strategy.

8 More General Numeraires

Up to this point, as it is always the case by default henceforth, we implicitly chose the savings account B, assumed to be a positive finite variation process, as a *numeraire*, namely a primary asset with positive price process, devoted to be used for discounting other price processes. However for certain applications, like dealing with stochastic interest rates in the field of interest rate derivatives, this choice may not be available (inasmuch as there may not be a riskless asset in the primary market), or it may not be the most appropriate (even if there is one, the choice of another asset as a numeraire may be more convenient). This motivates the extension of the previous developments to the case where B is a general locally bounded positive semimartingale, not necessarily of finite variation. The interpretation of B as savings account and of $\beta = B^{-1}$ as a riskless discount factor is now replaced by the interpretation of B as a simple numeraire, referring to the fact that other price processes will be typically expressed as relative (rather than discounted) prices βX.

Understanding *discounted price* as *relative price*, *risk-neutral model* as *martingale model relative to the numeraire* B, etc., the risk-neutral modeling approach developed in the previous sections holds mutatis mutandis under this relaxed assumption on B. Note in particular that the self-financing condition still writes (3) (see, e.g., [158]), though this is not as obvious as in the special case where B was a finite variation and continuous process. Also note that the notion of arbitrage is now to be understood as the one relative to the numeraire B, where the set of admissible strategies is a numeraire dependent notion.

In this more general situation, let us define a formal correspondence between processes (Π, M, K) and (π, m, k) by setting

$$\pi_t = \beta_t \Pi_t \ , \ dm_t = \beta_t \, dM_t \ , \ dk_t = \beta_t \, dK_t \text{ with } m_0 = 0 \text{ and } k_0 = 0 \tag{66}$$

where β now refers to the discount factor relative to an arbitrarily fixed numeraire. Equation $(\mathcal{E})$ to be solved in (Π, M, K) (with β as just mentioned above) is then equivalent to the following reflected BSDE with data $(c, \eta, \ell) := (\beta C, \beta_T \xi, \ \beta L)$, to be solved in (π, m, k) (cf. (10)):

$$\boxed{\begin{array}{c} \pi_t = \eta + c_T - c_t + k_T - k_t - (m_T - m_t), \quad t \in [0,T] \\ \ell_t \leq \pi_t, \quad t \in [0,T] \\ \int_0^T (\pi_u - \ell_u) \, dk_u = 0 \end{array}} \tag{67}$$

Note that equation (67) has the same structure as $(\mathcal{E})$ (it is in fact equation $(\mathcal{E})$ with input data r, C, ξ, L defined as $0, c, \eta, \ell$).

The conclusions of Theorems 6.1, 6.2 are still valid in this context, provided that "a solution (Π, M, K) to $(\mathcal{E})$" therein is understood as the process (Π, M, K) defined via (66) in terms of a solution (π, m, k) to (67).

The Markovian case now corresponds to the case where (to be compared with (21)):

$$\boxed{c_t = c(\mathcal{Z}_t)\,,\ \eta = \eta(\mathcal{Z}_T)\,,\ \ell_t = \ell(\mathcal{Z}_t)} \tag{68}$$

for a suitable Markov factor process $\mathcal{Z}$.

Remark 8.1 Of course, in order to ensure (68), one needs as a rule to include the numeraire asset B as a component of the factor process $\mathcal{Z}$, which increases by one the dimension of the model, and by a factor 100 or more (number of mesh points in the direction B) the numerical cost of solving the related systems of PIDEs. An important exception to this rule is the case when $\beta_T = 1$ and there are no dividends D nor barrier L involved (case of European derivatives without dividends, see Section 11).

In the Jump–Diffusion Setting with Regimes for $\mathcal{Z}$ with generator $\mathcal{A}$ given as (22) and drivers W, ν and μ (under a valuation measure corresponding to the numeraire under consideration), a solution (Π, m, k) to (67) is typically sought for with m in the form (cf. (56))

$$\boxed{m_t = \int_0^t \widehat{z}_u\, dW_u + \int_0^t \widetilde{z}_u\, d\widetilde{\nu}_u + \int_0^t \int_{\mathbb{R}^q} v_u(x)\, \widetilde{\mu}(du, dx)} \tag{69}$$

The system of PIDEs formally related to the BSDE (67) writes:

$$\boxed{\min\left(-\partial_t \pi - \mathcal{A}\pi - c, \pi - \ell\right) = 0 \text{ on } [0,T) \times \mathbb{R}^q \times I} \tag{70}$$

with terminal condition $\pi(T, x, y) = \eta(x, y)$. Under suitable conditions, the BSDE (67) admits a unique solution (Π, m, k) with m of the form (69), and the PIDE (70) admits a unique solution (in some sense) $\pi = \pi(z)$. The connection between them writes, for $t \in [0, T]$:

$$\boxed{\pi_t = \pi(\mathcal{Z}_t)}$$

and in regular cases:

$$\boxed{\widehat{z}_t = \partial\pi\sigma(\mathcal{Z}_t)\ ,\ \widetilde{z}_t = \Delta\pi(\mathcal{Z}_{t-})\ ,\ v_t(x) = \delta\pi(\mathcal{Z}_{t-})}$$

Let us further assume that the primary risky price process X satisfies likewise $\beta X = \chi(\mathcal{Z}_t)$ for a function χ such that (cf. (58))

$$d(\beta_t \widehat{X}_t) = \partial\chi\sigma(\mathcal{Z}_t) dW_t \ + \ \Delta\chi(\mathcal{Z}_{t-}) d\widetilde{\nu}_t \ + \ \delta\chi(\mathcal{Z}_{t-})\widetilde{\mu}(dt, d\cdot) \tag{71}$$

We then have the following analogs to Theorems 7.2, 7.3.

Theorem 8.1 *Under the previous conditions in the Markovian set-up, $\Pi_0 = B_0\pi(\mathcal{Z}_0)$ is the minimal initial wealth of a hedge with $\mathbb{P}$ – local martingale residual cost process for the related derivative. Moreover the cost process ρ and the related P&L process e (cf. (14), (15), (16)) may be rewritten as, respectively (with $\rho_0 = 0$):*

$$\boxed{\begin{aligned} d\rho_t = &\Big(\partial\pi\sigma(\mathcal{Z}_t) - \zeta_t \partial\chi\sigma(\mathcal{Z}_t)\Big) dW_t \\ &+ \Big(\Delta\pi(\mathcal{Z}_{t-}) - \zeta_t \Delta\chi(\mathcal{Z}_{t-})\Big) d\widetilde{\nu}_t \\ &+ \Big(\delta\pi(\mathcal{Z}_{t-}) - \zeta_t \delta\chi(\mathcal{Z}_{t-})\Big) \widetilde{\mu}(dt, d\cdot) \end{aligned}} \tag{72}$$

$$\boxed{\beta_t e_t = \pi_0 - \int_0^t c_u du \ + \int_0^t \zeta_u d(\beta_u \widehat{X}_u) - \pi_t = \int_0^t dk_u - \int_0^t \beta_u d\rho_u} \tag{73}$$

It is thus possible to hedge completely the market risk represented by W by setting, provided $\partial x \sigma$ is left-invertible,

$$\boxed{\zeta_t = \partial\pi\sigma(\partial\chi\sigma)^{-1}(\mathcal{Z}_t)} \tag{74}$$

In the simplest case where $q = d$ and $\partial\chi$ and σ are invertible this formula further reduces to

$$\boxed{\zeta_t = \partial\pi\partial\chi^{-1}(\mathcal{Z}_t)} \tag{75}$$

Alternatively, it is possible to hedge completely the jump risk ν by setting, provided $\chi(t, \mathcal{X}_{t-}) - \chi(\mathcal{Z}_{t-})$ is left-invertible,

$$\boxed{\zeta_t = \Delta\pi(\mathcal{Z}_{t-})(\Delta\chi(\mathcal{Z}_{t-}))^{-1}} \tag{76}$$

Still another possibility is to use the strategy ζ^{va} which minimizes the risk-neutral variance of the error, and which is given by

$$\boxed{\zeta^{va}_t = \frac{d<\pi,\chi>}{dt}\left(\frac{d<\chi>}{dt}\right)^{-1}} \tag{77}$$

where

$$\boxed{\begin{array}{l}\frac{d<\pi,\chi>}{dt} = \partial\pi a\partial\chi^{\mathsf{T}}(\mathcal{Z}_t) + \gamma(\mathcal{Z}_t)\overline{\delta\Pi(\delta\chi)^{\mathsf{T}}}(\mathcal{Z}_t) + \lambda(\mathcal{Z}_t)\overline{\Delta\Pi(\Delta\chi)^{\mathsf{T}}}(\mathcal{Z}_t) \\ \frac{d<\chi>}{dt} = \partial\chi a\partial\chi^{\mathsf{T}}(\mathcal{Z}_t) + \gamma(\mathcal{Z}_t)\overline{\delta\chi(\delta\chi)^{\mathsf{T}}}(\mathcal{Z}_t) + \lambda(\mathcal{Z}_t)\overline{\Delta\chi(\Delta\chi)^{\mathsf{T}}}(\mathcal{Z}_t)\end{array}} \tag{78}$$

8.1 Changes of Numeraire

In order to assess the effect of a change of numeraire, let us be given another numeraire $\widetilde{B}$ (primary asset with price process defined as a locally bounded positive semimartingale). Let us set $\nu_t := \frac{B_0\widetilde{B}_t}{\widetilde{B}_0 B_t}$. Note that ν is a $\mathbb{P}$ – local martingale, by arbitrage. Assuming that ν is in fact a $\mathbb{P}$ – martingale (for instance because ν is bounded, or in virtue of suitable moment conditions on ν), let us define the measure $\widetilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ (with $\mathcal{F} = \mathcal{F}_T$ as usual) by

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \nu_T \ . \tag{79}$$

Recall that a càdlàg process π is a $\widetilde{\mathbb{P}}$ – local martingale if and only if $\nu\pi$ is a $\mathbb{P}$ – local martingale (see e.g. Jacod–Shiryaev [104, Proposition III.3.8] or Jeanblanc et al. [111]). So

$$\frac{\widehat{X}}{\widetilde{B}} = \nu^{-1}\frac{\nu\widehat{X}}{\widetilde{B}} = \nu^{-1}\frac{B_0\widehat{X}}{\widetilde{B}_0 B}$$

is a $\widetilde{\mathbb{P}}$ – local martingale, and $\widetilde{\mathbb{P}}$ is a martingale measure relative to $\widetilde{B}$.

Consistently with this, denoting by Π the $\mathbb{P}$-price of an European derivative with integrable payoff ξ and no dividends (so $D = 0$ in (6)), the abstract Bayes rule (see, e.g., [111]) and the fact that ν is a $\mathbb{P}$-martingale yield:

$$\widetilde{B}_t\,\mathbb{E}_{\widetilde{\mathbb{P}}}(\widetilde{B}_T^{-1}\xi\,|\,\mathcal{F}_t) = \widetilde{B}_t\frac{\mathbb{E}_{\mathbb{P}}(\widetilde{B}_T^{-1}\xi\nu_T\,|\,\mathcal{F}_t)}{\mathbb{E}_{\mathbb{P}}(\nu_T\,|\,\mathcal{F}_t)} = B_t\mathbb{E}_{\mathbb{P}}(B_T^{-1}\xi\,|\,\mathcal{F}_t) = \Pi_t \ , \ t \in [0,T] \ .$$

So

$$\boxed{\Pi_t = \widetilde{B}_t\, \mathbb{E}_{\widetilde{\mathbb{P}}}(\widetilde{B}_T^{-1}\xi \,|\, \mathcal{F}_t)\ ,\ t \in [0,T]} \tag{80}$$

and $\frac{\Pi}{\widetilde{B}}$ is a (Doob) martingale under $\widetilde{\mathbb{P}}$.

Also note that ν_t is the $\mathcal{F}_t$-measurable Radon-Nikodym density $\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}\,|_{\mathcal{F}_t}$ of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$ on $\mathcal{F}_t$, for any $t \in [0,T]$. Indeed, for any $\mathcal{F}_t$-measurable and bounded random variable X, we have:

$$\mathbb{E}_{\widetilde{\mathbb{P}}}(X) = \mathbb{E}_{\mathbb{P}}(X\nu_T) = \mathbb{E}_{\mathbb{P}}\left[X\mathbb{E}_{\mathbb{P}}(\nu_T \,|\, \mathcal{F}_t)\right] = \mathbb{E}_{\mathbb{P}}(X\nu_t)\ . \tag{81}$$

Remark 8.2 These results will be intensively used in Section 11 with $\widetilde{\mathbb{P}} = \mathbb{P}^T$, the so-called *$T$ – forward-neutral* measure, with related numeraire $\widetilde{B} = B_t^T$ corresponding to a discount bond maturing at time T. Since $B_T(T) = 1$, (80) rewrites in this case:

$$\boxed{\Pi_t = B_t^T\, \mathbb{E}_{\mathbb{P}^T}(\xi \,|\, \mathcal{F}_t)\ ,\ t \in [0,T]} \tag{82}$$

For further study of changes of numeraires from the point of view of the connection between arbitrage prices and hedging through the related BSDEs and/or PIDEs, we refer the reader to Bielecki et al. [32] and Crépey [66].

9 Towards Real-Life Models

9.1 Model Calibration

The knowledge of the martingale valuation measure $\mathbb{P} \in \mathcal{M}$ 'chosen by the market' (assuming no arbitrage) is the key to fair valuation and hedging of financial derivatives. For hedging purposes or/and in order to be able to implement definite bets on specific risk factors, and also, to be able to sell exotic or structured products at fair price, traders need to know the market martingale valuation measure $\mathbb{P}$.

There are two sets of constraints that $\mathbb{P}$ should satisfy, and that can help in the quest for the market martingale valuation measure.

Firstly, $\mathbb{P}$ should satisfy structural requirements coming from the equivalence between $\mathbb{P}$ and the objective probability measure $\widehat{\mathbb{P}}$. For instance, if there are jumps in the time series of the factor process $\mathcal{Z}$, resp. of the market price process X, then $\mathcal{Z}$, resp. X, should also be a jump process under the valuation measure. More generally, the factor process, resp. the primary price process, should have the same *trajectorial properties* under the objective and under the market valuation measure.

Secondly, the discounted cumulative value process $\beta\widehat{X}$ must be a $\mathbb{P}$ – local martingale, and the cross-section $\Pi_t^{\pm}$ of the market prices of European vanillas quoted at time t on S must satisfy (assuming S_T integrable, for call options; cf. Theorem 5.1(i)):

$$\Pi_t^{\pm}(T,K) = \beta_t^{-1}\mathbb{E}_t\beta_T(S_T - K)^{\pm}\ ,\ (T,K) \in \mathrm{obs}_t^{\pm} \tag{83}$$

where $\mathrm{obs}_t^{\pm}$ is the set of quoted European vanilla call/put options on S at time t. Constraints of type (83) are called *calibration constraints*. A model is said to *fit the smile* at a given time

t, if it satisfies the calibration constraints (83). We shall see in the sequel that quite a few classes of models can fit the smile within the bid-ask spread, provided their parameters are suitably calibrated. A further requirement is that a model fits the smile *dynamics* under the market valuation measure. This corresponds to additional calibration constraints associated with *exotic* options prices (see, e.g., [42, 43]).

Finally, for being usable in practice, a pricing model needs to be constructive and implementable in real time. Concretely this leads to seek for (preferably low-dimensional) Markovian approximations of the factor process and of the primary price process under the market valuation measure, with related hedging strategies ζ computed in view of the related cost process given as (59) (see also the discussion of section 7.3.2), rather than for the martingale valuation measure itself (assuming no arbitrage) with abstract hedging strategies and related cost process given as (15). A consequence of this shift of interest is that we shall relax the calibration equality constraints (83) into inequality constraints. So we shall consider in practice approximate Markovian pricing models $\mathcal{Z}$ with related generator $\mathcal{A}_{\mathcal{Z}}$ belonging to the class defined by (22)[3], calibrated to the market within the bid-ask spread.

9.2 Hedging in Practice

Given a model calibrated to the market, it may may then be used for consistent hedging (and also pricing, in the case of exotic options or structured products) purposes. Indeed, when banks sell derivatives, they are left with the responsibility to find ways to be able to provide the derivative's payoff at termination. To this end, they immediately set up a *hedging portfolio* by combination of liquidly traded instruments, such as the asset(s) underlying the derivative, and/or further vanilla derivatives. Of course, for feasibility as well as transaction cost reasons (note that transaction costs are not explicitly considered in our formalism), they restrict themselves to *piecewise constant* self-financing strategies in the primary market (cf. (3)) ζ^h such that

$$\zeta^h_t = \zeta^h_{t_i} \text{ for } t_i < t \leq t_{i+1} \tag{84}$$

where $(t_i)_{0\leq i\leq n}$ is a (deterministic here, for simplicity) partition of $[0,T]$ (assuming that the product was sold at time 0). In practice n may vary from 1 *(static hedging)* to the number of (working) days or weeks between 0 and T (the most common forms of *dynamic hedging*, in discrete time). So, at the two extremes of the spectrum:

• in the static hedging approach, one uses a buy-and-hold hedging portfolio in order to synthetically replicate the derivative's *payoff*;

• in the dynamic hedging approach, one aims at replicating the derivative by a *dynamic self-financing* portfolio, with the *same initial value* and the same *risk sensitivities* as the derivative. Since derivative's (at least, options') payoffs are typically nonlinear with respect to the value of the underlying assets at termination, these sensitivities change all the time, so that, in order to get a perfect hedge, the composition of the hedging portfolio would have to be updated continuously (see (15), (59) and the discussion of section 7.3.2).

Let us thus consider an *European option* with maturity T (without dividends). The picture is the following: A trader sells the option at time 0 at price Π_0, that is she receives at time 0 the amount of money Π_0, but in turn it is mandatory for her to pay at time T the payoff ξ, which may be very high depending on the movements of the underlying(s) between 0 and

[3] Of course in most applications the pricing model corresponds to a rather specific subcase of (22): diffusion model without jumps nor regimes, jump-diffusion model, pure jump Markov chain model..

T. Her strategy consists in rebalancing at every time step t_i a self-financing hedge in the primary market. Accounting also for the facts that $V_0 = \Pi_0$ and for the final payment by the trader of ξ at time T, the discounted P&L of the trader at T is therefore (cf. (14)):

$$\boxed{\beta_T e_T = \Pi_0 + \int_0^T \zeta_t^h d\left(\beta_t \widehat{X}_t\right) - \beta_T \xi = \Pi_0 + \sum_{i=0}^{n-1} \zeta_{t_i}^h \left(\beta_{t_{i+1}} \widehat{X}_{t_{i+1}} - \beta_{t_i} \widehat{X}_{t_i}\right) - \beta_T \xi} \quad (85)$$

or, equivalently:

$$\boxed{\beta_T e_T = \sum_{i=0}^{n-1} \beta_{t_i} \delta_i e} \quad (86)$$

with

$$\boxed{\beta_{t_i} \delta_i e = -(\beta_{t_{i+1}} \Pi_{t_{i+1}} - \beta_{t_i} \Pi_{t_i}) + \zeta_{t_i}^h \left(\beta_{t_{i+1}} \widehat{X}_{t_{i+1}} - \beta_{t_i} \widehat{X}_{t_i}\right)} \quad (87)$$

In the case of an *American option* exercised at the stopping time τ, assumed for simplicity to be of the form $\tau = \nu h$, for a random integer $\nu \in \{0, \dots, n\}$, formulas (85) and (86) become:

$$\boxed{\beta_\tau e_\tau = \Pi_0 + \sum_{i=0}^{\nu-1} \zeta_{t_i}^h \left(\beta_{t_{i+1}} \widehat{X}_{t_{i+1}} - \beta_{t_i} \widehat{X}_{t_i}\right) - \beta_\tau \Pi_\tau = \sum_{i=0}^{\nu-1} \beta_{t_i} \delta_i e} \quad (88)$$

In the case of a complete primary market and for ζ^* making ρ^* equal to 0 in (15), then (cf. (16))

$$e_\tau = \Pi_0 + \int_0^\tau \zeta_t^* d\left(\beta_t \widehat{X}_t\right) - \beta_\tau \Pi_\tau = \int_0^\tau \beta_u dK_u - \int_0^\tau \beta_t d\rho_u^* = \int_0^\tau \beta_u dK_u \geq 0$$

for any stopping time τ in the case of an American option, and $e_T = 0$ in the case of an European option with $K = 0$. Now there are (at least) two sets of reasons why a practical strategy ζ^h typically leads to negative P&Ls in some scenarii of the economy.

First, ζ^h necessarily differs from ζ^*, due to:
• *Model misspecification*: ζ^h is typically computed in a Markovian approximation $\mathcal{Z}$ of the real (obviously non-Markovian) market, so even if ζ^h makes the related cost ρ given by (59) equal to 0 in the model $\mathcal{Z}$, however the cost (15) of the strategy ζ^h in the real market is only approximately equal to zero. Note that hedging ratios are typically *model-dependent*, and even *calibrated model dependent*. So the composition of the hedging portfolio varies between models, and even between models calibrated to the same set of hedging instruments.
• *Hedge slippage*: the strategy ζ^h is applied at discrete times only, whereas it should be applied in continuous time to have a chance to make the (approximate Markovian) cost process ρ given by (59) identically equal to 0.

The second set of reasons includes:
• *Market incompleteness*;
• *Arbitrages* which may occur in the market, and are excluded by standing assumption in the standard theory of mathematical finance;
• *Transaction costs*, and more general forms of market imperfections, which are not explicitly accounted for in our formalism.

The conclusion of this part is that in our generic Markovian market model $\mathcal{Z}$, derivative prices and Greeks are given as solutions to the related risk-neutral *pricing BSDEs/PIDEs*.

We shall see in the next part that in the simplest models (and in fact also in all affine jump-diffusion AJD models, see [75], or in the SABR model [96]), semi-analytic formulae are available for the prices and Greeks of European vanillas. This is very useful for model calibration, which typically goes through intensive pricing of European vanillas (cf. Part (VII)). However, as far as pricing exotic products or dealing with less standard models is concerned, the pricing BSDEs/PIDEs will have to be solved numerically by stochastic simulation methods, or, provided the dimension of the model is not too large, by deterministic PIDE or tree numerical schemes (cf. Parts (IV) to (VI)).

Part III
Benchmark Models

In this part we gather basic facts regarding the benchmark models on the main derivatives markets: equity-derivatives markets (Black–Scholes model and simple extensions), interest-rate derivatives markets (BGM model) and credit derivatives markets (one factor Gaussian copula model).

10 Black–Scholes and Beyond

10.1 Black–Scholes Basics

Let us start by the undisputed star of mathematical finance: the Black–Scholes formula. We consider a primary market composed of the savings account $B = \beta^{-1}$ and of an underlying S, which may represent a stock, an index, a future price, an exchange rate, a forward interest or swap rate, etc. The riskless interest rate r in the economy and the dividend yield q on S are assumed to be constant[4]. So in particular $B_t = e^{rt}$. Recall that by arbitrage (cf. Section 5 and [72]), the discounted wealth process of any admissible self-financing trading strategy in S and B must be a local martingale under a risk-neutral probability measure $\mathbb{P}$. In particular, $\beta_t S_t e^{qt} = S_t e^{-\kappa t}$, where we set $\kappa = r - q$, must be a $\mathbb{P}$ – local martingale. Consistently with this requirement, (the risk-neutral form of) the *Black–Scholes(–Merton)* model [36, 1973] [139, 1973], or *BS model* for short in the sequel, postulates the following spot diffusion, driven by a standard $\mathbb{P}$ – Brownian motion W on $[0, T]$:

$$\boxed{dS_t = S_t(\kappa dt + \sigma dW_t)} \tag{89}$$

or, equivalently:

$$d\left(\beta_t S_t e^{qt}\right) = \sigma\left(\beta_t e^{qt} S_t\right) dW_t\,, \tag{90}$$

for a constant *volatility parameter* σ. as The T – forward price $F_t = S_t e^{\kappa(T-t)}$ of S is thus a $\mathbb{P}$ – Brownian martingale with constant volatility σ. Note for further use that (89) may also be rewritten in terms of the cumulative price $\widehat{S}$ (cf. (1)) as:

$$d(\beta_t \widehat{S}_t) = d(\beta_t S_t) + \beta_t q S_t dt = e^{-qt} d(\beta_t S_t e^{qt}) = \beta_t \sigma S_t dW_t\,. \tag{91}$$

The arbitrage price process of an European vanilla option with (integrable, let us say measurable and bounded for simplicity) payoff $\phi(S_T)$ at T is in turn given by (cf. Section 5):

$$\Pi_t = e^{-r(T-t)} \mathbb{E}_t \phi(S_T)\ ,\ t \in [0, T]\,. \tag{92}$$

In particular, the discounted price $e^{-rt}\Pi_t$ is a (Doob) martingale under the risk-neutral measure $\mathbb{P}$. Moreover, by the Markov property of the risk-neutral BS stock S, we have that

$$\mathbb{E}_t \phi(S_T) = \mathbb{E}\left(\phi(S_T) \,|\, S_t\right)\ ,$$

[4] In fact the interpretation of r and q depends on the nature of the underlying S : stock, interest rate, exchange rate..

and it is possible to show that $\Pi_t = v(t, S_t)$, for a suitable Borel-measurable function v. Assuming v of class $C^{1,2}([0,T] \times (0,\infty))$, an application of the Itô formula (51) yields:

$$e^{rt} d(e^{-rt} v(t, S_t)) = (\partial_t v + \mathcal{L} v)(t, S_t) dt + \sigma S_t \partial_S v(t, S_t) dW_t$$

where $\mathcal{L} v = \kappa S \partial_S v + \frac{\sigma^2 S^2}{2} \partial^2_{S^2} v - rv$. As $e^{-rt} v(t, S_t) = e^{-rt} \Pi_t$ is a martingale, thus

$$\partial_t v + \mathcal{L} v = 0 \, .$$

Accounting for the fact that $\Pi_T = \phi$, this leads to the following *Black–Scholes valuation PDE*:

$$\boxed{\begin{cases} v(T, S) = \phi(S), \ S \in (0, +\infty) \\ \partial_t v + \kappa S \partial_S v + \frac{1}{2} \sigma^2 S^2 \partial^2_{S^2} v - rv = 0 \text{ in } [0,T) \times (0, +\infty) \end{cases}} \tag{93}$$

Conversely, for regular enough and bounded ϕ, this PDE is known to have a unique classic solution v bounded on $[0,T] \times (0, +\infty)$ (see, e.g., Friedman [90]). We then have by application of the Itô formula:

$$\beta_T \phi(S_T) = \beta_t v(t, S_t) + \int_t^T \beta_u \partial_S v(u, S_u) \sigma S_u dW_u$$

where

$$\beta_u \partial_S v(u, S_u) \sigma S_u dW_u = \partial_S v(u, S_u) d(\beta_u \widehat{S}_u) \, ,$$

by (91). So

$$\boxed{\beta_T \phi(S_T) = \beta_t v(t, S_t) + \int_t^T \beta_u \partial_S v(u, S_u) d(\beta_u \widehat{S}_u) \ , \ \mathbb{P}\text{-a.s.}} \tag{94}$$

This demonstrates, in the simple set-up of the Black–Scholes model, the following outputs of Theorem 7.2–7.3:

• Firstly, in view of (92), taking expectations in (94) yields $v(t, S_t) = \Pi_t$ (note that the stochastic integral in the r.h.s. of (94) is a bounded $\mathbb{P}$ – local martingale and therefore a $\mathbb{P}$-martingale);

• Secondly, the self-financing hedging strategy defined by, for $t \in [0,T]$

$$\boxed{\zeta^{bs}_u = \partial_S v(u, S_u)} \tag{95}$$

units of stock S (cf. formula (61) with $X = X(t,S) = S$, here, since the hedging instrument under consideration corresponds to the model factor S) and $\beta_u \left(v(u, S_u) - S_u \partial_S v(u, S_u) \right)$ units of the riskless asset $B_u = e^{ru}$ at time $u \in [t,T]$, replicates the option's payoff $\phi(S_T)$, starting from the wealth $v(t, S_t) = \Pi_t$ at time t.

Remark 10.1 Since $\mathbb{P} \sim \widehat{\mathbb{P}}$, (94) also holds $\widehat{\mathbb{P}}$-a.s.. This replicability of Contingent Claims in the Black–Scholes model explains why the Black–Scholes price does not depend on the physical drift μ, even though μ may be responsible for fat tails or skewness in the *physical* spot returns (under $\widehat{\mathbb{P}}$).

In order to illustrate the flexibility of Theorem 7.3 and of the related formulas, let us now consider the case of a primary market defined by the savings account B as before, and a T – forward contract on S (instead of S above) with strike F_0. The arbitrage price process V of the T – forward contract on S is thus given as, for $t \in [0,T]$:

$$V_t = S_t e^{-q(T-t)} - F_0 e^{-r(T-t)} = V(t, S_t)$$

so $\partial_S V(t,S_t) = e^{-q(T-t)}$. By application of Theorem 7.3 (starting from time t), a perfect replication strategy for the option with initial wealth $v(t,S_t)$ at time t is defined by, at any time $u \in [t,T]$,

$$\boxed{\widetilde{\zeta}_u^{bs} = \partial_S v(u,S_u) = e^{q(T-u)}\partial_S v(u,S_u)} \tag{96}$$

units of T – forward contracts on S, and, therefore, $\beta_u\left(v(u,S_u) - \widetilde{\zeta}_u^{bs} V_u\right)$ units of B.

In the special case of an European call option with payoff function $\phi(S) = (S-K)^+$, a direct computation based on (92) (or verification on (93)) shows that the Black–Scholes call pricing function Π^{bs} and delta function $\Delta^{bs} = \partial_S \Pi^{bs}$ are given by the so-called *Black–Scholes formulae*:

$$\boxed{\begin{array}{c}\Pi^{bs}(T,K;t,S,r,q,\sigma) = Se^{-q\tau}\mathcal{N}(d_+) - Ke^{-r\tau}\mathcal{N}(d_-) \\ \Delta^{bs}(T,K;t,S,r,q,\sigma) = e^{-q\tau}\mathcal{N}(d_+)\end{array}} \tag{97}$$

where $\tau = T-t$, $\mathcal{N}$ is the Gaussian cumulative distribution function and

$$\boxed{d_\pm = \frac{\ln(\frac{S}{K}) + \kappa\tau}{\sigma\sqrt{\tau}} \pm \frac{1}{2}\sigma\sqrt{\tau}} \tag{98}$$

These results admit straightforward extensions to the case where r, q and σ are Borel-measurable bounded functions of time: simply replace $r\tau$, $q\tau$ and $\sigma\sqrt{\tau}$ by $\int_t^T r(u)du$, $\int_t^T q(u)du$ and Σ, with $\Sigma^2 = \int_t^T \sigma^2(u)du$, in (97)–(98).

As is well known, the Black–Scholes model is strongly misspecified, which leads to consider natural extensions of this model, adding stochastic volatility (Heston model) or/and jumps (Merton and Bates model) in the picture.

10.2 Heston Model

The best known Stochastic Volatility model is the Heston model (1993, see [99]), which postulates, for the instantaneous variance process, an affine process $V \geq 0$, namely, under $\mathbb{P}$ (i.e., in risk-neutral form):

$$\left\{\begin{array}{l} dV_t = -\lambda(V_t - \theta)dt + \eta\sqrt{V_t}dZ_t \\ dS_t = S_t(\kappa dt + \sqrt{V_t}dW_t)\ , \end{array}\right. \tag{99}$$

where:
- $d\langle W,Z\rangle_t = \rho dt$,
- λ is the *speed of mean-reversion* of the instantaneous variance,
- θ is the *long-term variance mean*,
- $\frac{\eta}{2\sqrt{V}}$ is the (instantaneous) *volatility of the volatility*.

10.3 Merton Model

The Merton model (1976, see [138]) is obtained by adding an independent compound Poisson jump process to the Black–Scholes model as follows, under $\mathbb{P}$:

$$\boxed{\frac{dS_t}{S_{t-}} = \kappa dt + \sigma dW_t + d(\textstyle\sum_{l=1}^{N_t} J_l - \gamma\bar{J}t)} \tag{100}$$

with related generator $\mathcal{A}_S$ such that, for $v = v(S)$ (cf. (46)):

$$\mathcal{A}_S v = \left(\kappa - \gamma\bar{J}\right) S\partial_S v + \frac{1}{2}\sigma^2 S^2 \partial^2_{S^2} v + \gamma\left[\mathbb{E}v((1+J_1)S) - v(S)\right] \ , \tag{101}$$

where:
• $(N_t)_{t\geq 0}$ is a Poisson process with deterministic *jump intensity* γ;
• the J_l are the *jump size variables*, such that $j_l := \ln(1+J_l) \hookrightarrow \mathcal{N}(\alpha, \beta)$, so

$$\bar{j} := \mathbb{E}j_1 = \alpha, \ \mathbb{E}j_1^2 = \alpha^2 + \beta, \ \bar{J} := \mathbb{E}J_1 = e^{\alpha + \frac{\beta}{2}} - 1 \ ;$$

• W, N and the J_l are independent.

Denoting further $a = b - \gamma\bar{J}$, we have in particular:

$$\mathcal{A}_S \ln(S) = \kappa - \gamma\bar{J} - \frac{1}{2}\sigma^2 + \gamma\bar{j} = a + \gamma\bar{j} \ . \tag{102}$$

By application of the Itô-Lévy formula (47), the model can be rewritten in terms of the log-spot $X_t = \ln(S_t)$ as follows:

$$dX_t = adt + \sigma dW_t + d(\sum_{l=1}^{N_t} j_l) \ , \tag{103}$$

whence $X_T = x + at + \sigma W_t + \sum_{l=1}^{N_t} j_l$, and

$$\boxed{S_T = S_0 e^{at+\sigma W_t} \prod_{l=1}^{N_t}(1+J_l)} \tag{104}$$

10.4 Bates Model

The Bates model [24, 1996] is a mixture of the Heston and Merton models (Heston model with additional independent Merton-style jumps in S), so, under $\mathbb{P}$:

$$\begin{cases} dV_t = -\lambda(V_t - \theta)dt + \eta\sqrt{V_t}dZ_t \\ \frac{dS_t}{S_{t-}} = (\kappa - \gamma\bar{J})dt + \sqrt{V_t}dW_t + d(\sum_{l=1}^{N_t} J_l) \ . \end{cases} \tag{105}$$

Note that from the practical point of view, the Heston and the Merton model are still notoriously misspecified, like the Black–Scholes model. The Bates model may often be considered as a flexible enough and robust alternative.

10.5 Log-Spot Characteristic Functions

The (risk-neutral) characteristic function

$$\Phi_T(u) = \mathbb{E}[\exp(iuX_T)] = \mathbb{E}[X_T^{iu}]$$

of the log-spot $X_t = \ln(S_t)$ is known in closed-form in the previous models (which are all *AJD models* [75]), so that the related vanilla option prices and Greeks can be computed

efficiently by Fourier transform (see Section 37).

Proposition 10.1 *We have in the Black–Scholes, Merton, Heston and Bates models, respectively:*

$$\begin{gathered}\Phi_T^{bs}(u) = \Phi_T^{\kappa}(u)\exp\left[-\tfrac{1}{2}u(i+u)\sigma^2 T\right] \\ \Phi_T^{me}(u) = \Phi_T^{bs}(u)\Phi_T^{j}(u) = \exp\left[\gamma T(e^{iu\alpha-u^2\frac{\beta}{2}}-1) - \tfrac{u^2\sigma^2 T}{2} + iu\Big(x_0 + [b-\gamma e^{\alpha+\frac{\beta}{2}}+\gamma]T\Big)\right] \\ \Phi_T^{he}(u) = \Phi_T^{\kappa}(u)\exp\left[C(u,T)\theta + D(u,T)v\right] \\ \Phi_T^{ba}(u) = \Phi_T^{he}(u)\Phi_T^{j}(u)\ ,\end{gathered}$$

with $x = X_0 = \ln(S_0)$, $v = V_0$, *and*

$$\begin{gathered}\Phi_T^{\kappa}(u) = \exp\left[iu\big(x+\kappa T\big)\right]\ ,\ \Phi_T^{j}(u) = \exp\left[-\gamma T\left(iu\big(e^{\alpha+\frac{\beta}{2}}-1\big) - \big(e^{iu\alpha-u^2\frac{\beta}{2}}-1\big)\right)\right] \\ C(u,T) = \lambda\left[Ty_- - \tfrac{2}{\eta^2}\ln(\tfrac{1-ge^{-pT}}{1-g})\right]\ ,\ D(u,T) = \tfrac{1-e^{-pT}}{1-ge^{-pT}}y_- \qquad (106) \\ p = \sqrt{y^2-4wz}\ ,\ y_\pm = \tfrac{y\pm p}{\eta^2}\ ,\ g = \tfrac{y_-}{y_+} \\ w = -\tfrac{1}{2}u(i+u) = \tfrac{1}{2}ui(ui-1) = -\tfrac{u^2}{2} - \tfrac{ui}{2}\ ,\ y = \lambda - \rho\eta iu\ ,\ z = \tfrac{\eta^2}{2}\end{gathered}$$

Remark 10.2 Setting $\lambda = \rho = 0$ and sending $\eta \to 0+$ yields

$$y = 0,\ y_\pm = \frac{\pm p}{\eta^2},\ p = \sqrt{-2w}\eta,\ 1-g = 2,\ D(u,T) \sim \frac{pT}{1-g}y_- = \frac{-p^2T}{2\eta^2} = wT\ ,$$

and $C(u,T)\theta + D(u,T)v$ reduces to $-\frac{1}{2}u(i+u)vT$, so $\Phi_T^{he}(u)$ reduces to $\Phi_T^{bs}(u)$ for $\sigma = \sqrt{v}$, as expected.
Likewise, it is immediate to check that $\Phi_T^{me}(u)$ reduces to $\Phi_T^{bs}(u)$ for $\alpha = \beta = 0$.

Proof of Proposition 10.1. In the case of the Heston model, we introduce

$$F_t := S_t e^{-\kappa t}$$

so $F_0 = S_0$, and we compute $\Phi_T^{he}(u)$ as

$$\mathbb{E}_0 S_T^{ui} = e^{\kappa uiT}\mathbb{E}_0 F_T^{ui} = e^{\kappa uiT}\Phi(0, F_0, V_0)\ ,$$

with $\Phi(t,F,V) := \mathbb{E}_t F_T^{ui}$ on $\{F_t = F,\, V_t = V\}$. In particular

$$\Phi(T,F,V) = F^{ui}\ . \qquad (107)$$

Consistently with (107), let us seek for an explicit solution for Φ of the form

$$\Phi(t,F,V) = F^{ui}\exp\left[C(u,T-t)\theta + D(u,T-t)V\right]\ ,$$

for suitable coefficients $C = C(u,\tau)$ and $D = D(u,\tau)$ such that

$$C(u,0)\theta + D(u,0)V = 0\ . \qquad (108)$$

Note that $\Phi(t,F_t,V_t)$ is a (Doob) martingale. We thus get by application of the Itô formula (51):

$$\partial_t\Phi + \mathcal{A}_{F,V}\Phi = 0 \text{ on } \{t<T\}\ ,$$

with

$$\mathcal{A}_{F,V} = \frac{1}{2}VF^2\partial^2_{F^2} + \frac{1}{2}\eta^2 V\partial^2_{V^2} + \rho\eta VF\partial^2_{FV} - \lambda(V-\theta)\partial_V \ .$$

This yields the following equation in (C, D) :

$$-\theta\,\partial_\tau C - V\,\partial_\tau D + \frac{1}{2}V(ui)(ui-1) + \frac{1}{2}\eta^2 VD^2 + \eta\rho V(ui)D - \lambda(V-\theta)D = 0$$

on $\{t < T\}$. More specifically, let us look for (C, D) such that for $t < T$:

$$\left\{ \begin{array}{l} \partial_\tau C = \lambda D \\ \partial_\tau D = w + \frac{1}{2}\eta^2 D^2 + \rho\eta uiD - \lambda D = w + zD^2 - yD = z(D-y_+)(D-y_-) \ . \end{array} \right. \tag{109}$$

along with the initial condition $C(u, 0) = D(u, 0) = 0$.

We recognize in the second line of (109) a Riccati equation. The solution $C(u,\tau)$, $D(u,\tau)$ of (109) is given by (106) (with $T = \tau$ therein), as one can check by inspection. This proves the result in the case of the Heston model.

In the case of the Bates model, let us introduce the process

$$L_t = \prod_{l=1}^{N_t}(1+J_l)e^{-\gamma\bar{J}t} \ , \ dL_t = L_{t-}\,d(\sum_{l=1}^{N_t} J_l - \gamma\bar{J}t) \ . \tag{110}$$

Note that, using an obvious notation:

$$d(S^{he}_t L_t) = S^{he}_t dL_t + L_{t-}dS^{he}_t = S^{he}_t L_{t-}\left(\kappa dt + \sqrt{V_t}dW_t + d(\sum_{l=1}^{N_t} J_l - \gamma\bar{J}t)\right) \ ,$$

by (110). Hence $S^{ba} = S^{he}L$, by uniqueness of a solution to (105). Thus $\Phi^{ba}_T(u) = \Phi^{he}_T(u)\Phi^L_T(u)$, by independence. It remains to prove that $\Phi^L_T(u) := \mathbb{E}L^{ui}_T = \Phi^j_T(u)$. In this view, note that

$$\mathcal{A}_L L^{ui} = -\gamma L\bar{J}\partial_L L^{ui} + \gamma L^{ui}\big(\mathbb{E}\big[(1+J_1)^{ui}\big] - 1\big).$$

We thus have by the Itô–Poisson formula (45), with "$\triangleq$" standing for "equality up to a local martingale term":

$$dL^{ui}_t \triangleq L^{ui}_t\big(-\bar{J}ui + \big(\mathbb{E}\big[(1+J_1)^{ui}\big] - 1\big)\big)\gamma dt \ ,$$

whence

$$\mathbb{E}L^{ui}_t = L^{ui}_0 e^{\delta\gamma t} \ , \text{ with } \delta = -\bar{J}ui + \big(\mathbb{E}\big[(1+J_1)^{ui}\big] - 1\big) = -ui\big(e^{\alpha+\frac{\beta}{2}} - 1\big) + \ e^{ui\alpha - \frac{u^2\beta}{2}} - 1 \ ,$$

that is $\mathbb{E}L^{ui}_t = \Phi^j_t(u)$.

Finally, the results in the case of the Black–Scholes and of the Merton models follow by passage to the limit in the previous results, by Remark 10.2. □

11 BGM Model

11.1 Black Formulae

Black formulae extend the Black–Scholes formulae (97)–(98) (general case with time-dependent volatility $\sigma(t)$ and dividend yield $q(t)$) to the case of *stochastic* interest rates r. These formulae are derived within a martingale market model $(\Omega, \mathbb{F}, \mathbb{P}^T)$ relative to the numeraire B^T defined by the price process of a T – discount bond (cf. Section 8, Remark 8.2). We denote $\beta_t^T = \frac{1}{B_t^T}$. Recall that by arbitrage, the T – forward value

$$F_t := \beta_t^T S_t e^{-\int_t^T q(u)du} \tag{111}$$

of S (accounting for the dividends at yield $q(t)$ on S) must be a $\mathbb{P}^T$ – local martingale (see Section 8). Consistently with this requirement, Black's T – forward neutral model postulates that the process $(F_t)_{t\in[0,T]}$ is a $\mathbb{P}^T$ – Brownian local martingale with deterministic volatility process $\sigma(t)$.

By (82) applied with $\xi = (S_T - K)^+ = (F_T - K)^+$, and using the Black–Scholes formulae (97)–(98) with $r = q = 0$ to compute the expectation in the r.h.s. of (82), we thus get (note that π^{bl} and δ^{bl} thus defined are *forward pricing and delta functions*, cf. Theorem 8.1 and the related formulas):

$$\Pi_t^{bl} = B_t^T \pi^{bl}(T, K; t, F_t, \sigma)\ ,\ \delta_t^{bl} := \partial_F \pi^{bl}(T, K; t, F_t, \sigma) = \delta^{bl}(T, K; t, F_t, \sigma)$$

with

$$\boxed{\begin{array}{c}\pi^{bl}(T, K; t, F, \sigma) = F\mathcal{N}(d_+) - K\mathcal{N}(d_-) \\ \delta^{bl}(T, K; t, F, \sigma) = \mathcal{N}(d_+)\end{array}} \tag{112}$$

where

$$\boxed{d_\pm = \frac{\ln(F/K)}{\Sigma} \pm \frac{1}{2}\Sigma \text{ where } \Sigma^2 = \int_t^T \sigma^2(u)du} \tag{113}$$

Recall that δ_t^{bl} is a key ingredient of any hedging scheme for the option, cf. Theorem 8.1. Let us first consider a primary market defined by the numeraire (T – discount bond) B_t^T and a T – forward contract on S with strike price F_0. The relative price process $\beta^T V$ of the T – forward contract on S is thus given by, for $t \in [0, T]$ (cf. (111)):

$$\beta_t^T V_t = \beta_t^T \left(S_t e^{-\int_t^T q(u)du} - F_0 B_t^T \right) = F_t - F_0 = V(t, F_t)$$

So $\partial_F(\beta_t^T V_t)(t, F_t) = 1$. By application of formula (75) (see also Section 10.1 or [41] for a more direct proof in this particular situation), a perfect replication strategy with initial wealth Π_0^{bl} at time 0 is defined by, at any time $t \in [0, T]$,

$$\boxed{\widetilde{\zeta}_t^{bl} = \delta^{bl}(T, K; t, F_t, \sigma)(\partial_F \beta_t^T V_t)^{-1} = \delta^{bl}(T, K; t, F_t, \sigma)}$$

units of T – forward contracts on S, and, therefore,

$$\beta_t^T \left(B_t^T \pi^{bl}(T, K; t, F_t, \sigma) - \delta^{bl}(T, K; t, F_t, \sigma) B_t^T (F_t - F_0) \right) = -K\mathcal{N}(d_-)_t + \delta^{bl}(T, K; t, F_t, \sigma) F_0$$

units of the T – discount bond.

Let us now consider a primary market defined by the T – discount bond B_t^T and the stock $S_t = B_t^T F_t e^{\int_t^T q(u)du}$. By application of formula (75) with

$$\beta_t^T S_t = F_t e^{\int_t^T q(u)du} \ , \ \partial_F(\beta_t^T S_t)(t, F_t) = e^{\int_t^T q(u)du}$$

(see also [41]), a perfect replication strategy with initial wealth Π_0^{bl} for the European call option is thus defined by, at any time $t \in [0, T]$,

$$\boxed{\zeta_t^{bl} = e^{-\int_t^T q(u)du} \delta^{bl}(T, K; t, F_t, \sigma)}$$

units of S, and, therefore, $-K\mathcal{N}(d_-)_t$ units of T – discount bond.

As a reality check of these results, note that in the special case where B_t^T is given as $e^{-\int_t^T r(u)du}$ for a deterministic function $r(t)$, so $B^T = e^{-\int_0^T r(u)du} B$ and $\mathbb{P}^T = \mathbb{P}$, then

$$F_t = S_t e^{\int_t^T (r(u)-q(u))du} \ , \ d_\pm^{bl}(T, K; t, F_t, \sigma) = d_\pm^{bs}(T, K; t, S_t, r, q, \sigma) \ .$$

So (cf. (96))

$$\widetilde{\zeta}_t^{bl} = \mathcal{N}(d_+) = \widetilde{\zeta}_t^{bs} \ .$$

The (F-component of the) replication strategy in the T – forward neutral Black model with factor F and primary assets F and B^T is thus the same as the one in the risk-neutral Black–Scholes model with factor S and primary assets F and B. Likewise (cf. (95)),

$$\zeta_t^{bl} = e^{-\int_t^T q(u)du} \mathcal{N}(d_+) = \zeta_t^{bs} \ ,$$

so the (S-component of the) replication strategy in the Black T – forward neutral model with factor F and primary assets S and B^T is the same as the one in the Black–Scholes risk-neutral model with factor S and primary assets S and B.

Black formulae are the standard way to quote *options on forward contracts*, and *bond options*.

11.2 LIBOR Rates

The (forward) LIBOR rate $L_t(T, U)$ is the simple interest rate locked at time t for an investment on the future time period $[T, U]$, so for $t \leq T$:

$$B_t^T = [1 + h(T, U) L_t(T, U)] \, B_t^U \ , \text{ or } L_t(T, U) = \frac{1}{h(T, U)} \frac{B_t^T - B_t^U}{B_t^U} \ , \tag{114}$$

where $h(T, U) = U - T$ and B_t^T is the price at time t of a T – discount bond.

Given a *tenor* $t_1, ..., t_{n+1}$, let us set, for $i = 1, \ldots, n$:

$$h_i = t_{i+1} - t_i \ , \ L^i = L(t_i, t_{i+1}) \ , \ B_t^i = B_t^{t_{i+1}} \ , \ \mathbb{P}^i = \mathbb{P}^{t_{i+1}} \ , \ E^i = \mathbb{E}_{\mathbb{P}^i} \ ,$$

So for $i = 1, \ldots, n$ and $t \leq t_i$:

$$h_i L_t^i = \frac{B_t^{i-1} - B_t^i}{B_t^i} \ , \ B_t^i = \frac{1}{1 + h_i L_t^i} B_t^{i-1} \ . \tag{115}$$

Setting $B_t^0 = B_t^{t_1}$, we thus get by induction, for any $1 \leq i \leq l + 1 \leq n + 1$ (recall that $B_{t_{i+1}}^i = 1$):

$$\boxed{B_{t_i}^l = \prod_{k=i}^l \frac{1}{1 + h_k L_{t_i}^k}} \tag{116}$$

Note that $B_t^{i-1} - B_t^i$ is the price of a (synthetic) traded asset, hence the relative price $h_i L_t^i = \frac{B_t^{i-1} - B_t^i}{B_t^i}$ must be a local martingale under the pricing measure $\mathbb{P}^i$, by arbitrage (cf. Section 8). Consistently with this requirement, the BGM model postulates the following $\mathbb{P}^i$-dynamics for the factor process L^i :

$$\boxed{dL_t^i = \sigma_i(t) L_t^i dW_t^i, \text{ or } L_t^i = L_0^i \exp\left(\int_0^t \sigma_i(s) dW_s^i - \tfrac{1}{2}\int_0^t \sigma_i^2(s) ds\right)} \tag{117}$$

for some $\mathbb{P}^i$ – Brownian motion W^i and a deterministic volatility function $\sigma_i(t)$ (null after t_i). The process L^i is thus log-normal under $\mathbb{P}^i$, and the variance of $\ln L_t^i$ is equal to $\int_0^t \sigma_i^2(s) ds$.

For the specification of the factor process L to be complete, we still need to specify the correlation structure of L. This is deferred to section 11.4. Indeed, the correlation does not affect the pricing of caps and floors to be considered in the next subsection.

11.3 Caps and Floors

A caplet is a call option with arrear settlement on a LIBOR Rate. Considering the caplet with maturity t_i and strike K, the payoff to the caplet holder at t_{i+1} is:

$$C_{t_{i+1}}^i = h_i \max\{L_{t_i}^i - K, 0\}.$$

Working in the B^i numeraire, we thus have by (82), for $t \leq t_i$:

$$\frac{C_t^i}{B_t^i} = h_i \mathbb{E}^i \left[\max\{L_{t_i}^i - K, 0\} | \mathcal{F}_t\right] .$$

Hence by log-normality postulated in (117) (again δ_t^i is here a forward delta):

$$\boxed{\begin{array}{c} C_t^i = B_t^i \pi^{bl}(t_{i+1}, h_i K; t, h_i L_t^i, \sigma_i) = h_i B_t^i \left[L_t^i \mathcal{N}(d_1^i) - K\mathcal{N}(d_2^i)\right] \\ \delta_t^i = \partial_{h_i L} \pi^{bl}(t_{i+1}, h_i K; t, h_i L_t^i, \sigma_i) = \delta^{bl}(t_{i+1}, K; t, L_t^i, \sigma_i) = \mathcal{N}(d_1^i) \end{array}} \tag{118}$$

with (since in particular σ_i vanishes after t_i, for the last identity)

$$\boxed{d_1^i = \frac{\ln(\frac{L_t^i}{K}) + \frac{\Sigma_i^2}{2}}{\Sigma_i}, \quad d_2^i = d_1^i - \Sigma_i, \text{ where } \Sigma_i^2 = \int_t^{t_i} \sigma_i^2(s) ds}$$

Here δ_t^i is a key ingredient of the composition at time t of any hedging scheme for the caplet, cf. Theorem 8.1 and the subsequent formulas. For instance, in a primary market defined by the numeraire (t_{i+1} – discount bond) and a forward swap over the period (t_i, t_{i+1}), one can show by application of formula (75) (see also [41]) that a perfect replication strategy for the caplet with initial wealth C_0^i is defined by, at any time $t \in [0, T]$,

$$\zeta_t = \delta_t^i$$

units of forward swaps over the period (t_i, t_{i+1}), and the number of t_{i+1} – discount bonds following from the self-financing condition on the replicating portfolio.

The analog formulae for a floorlet are:

$$F_t^i = h_i B_t^i \left[K\mathcal{N}(-d_2^i)_t - L_0^i \mathcal{N}(-d_1^i)_t\right] \ , \ \delta_t^i = -\mathcal{N}(-d_1^i)_t \ . \tag{119}$$

We further get the price and delta of the cap (portfolio of caplets) with maturity t_1 and tenor $t_1, ..., t_{n+1}$ as

$$\boxed{C_t = \sum_{i=1}^n C_t^i \ , \ \delta_t = \sum_{i=1}^n \delta_t^i}$$

where C_t^i and δ_t^i were defined in (118), and the analog formula for floors, using (119). Note in particular that due to the additively separable structure of the related payoffs, the prices of caps and floors only depend on the marginal laws of the Libor rates L^i at times t_i.

11.4 Adding Correlation

More complex derivatives, like swaptions, are also affected by the correlation structure of L. We are thus going to define a correlation structure between the L^is by expressing their dynamics (until the t_is) under the common terminal measure $\mathbb{P}^n$.

Note that by definition of the $\mathbb{P}^i$s, we have for $t \leq t_i$ (cf. (81)):

$$\nu_t^i := \frac{d\mathbb{P}^{i-1}}{d\mathbb{P}^i} \mid_{\mathcal{F}_t} = \mathbb{E}^i \left[\frac{d\mathbb{P}^{i-1}}{d\mathbb{P}^i} \mid \mathcal{F}_t \right] = \frac{B_0^i B_t^{i-1}}{B_0^{i-1} B_t^i} = \frac{B_0^i}{B_0^{i-1}}(1 + h_i L_t^i). \tag{120}$$

We find it convenient to denote

$$\boxed{dL_t^i = s_i(t) L_t^i d\mathbb{W}_t^i = \sigma_i(t) L_t^i dW_t^{i,i}} \tag{121}$$

where $s_i(t) = (0, ..., 0, \sigma_i(t), 0, ..., 0)$ is a row volatility vector and $\mathbb{W}^i = (W^{j,i})_{1 \leq j \leq n}$ is a (correlated) n-dimensional $\mathbb{P}^i$ – Brownian motion. Let $\rho = (\rho_{i,l})_{1 \leq i,l \leq n}$ denote the *correlation matrix* of $\mathbb{W}^n$, so $d\langle W^{i,n}, W^{l,n} \rangle_t = \rho_{i,l} dt$.

We proceed by backward induction. We first want to specify $\mathbb{W}_t^{n-1}$ as $\mathbb{W}_t^n - \rho \int_0^t \mu_u^{\mathsf{T}} du$, for a properly chosen row vector process μ. In view of Girsanov theorem, it is enough, for this to be a $\mathbb{P}^{n-1}$ – Brownian motion on $[0, t_n]$, that μ satisfies $d\nu_t^n = \mu_t \nu_t^n d\mathbb{W}_t^n$, where ν_t^n is the $\mathcal{F}_t$-measurable Radon-Nikodym density of $\mathbb{P}^{n-1}$ with respect to $\mathbb{P}^n$ on $\mathcal{F}_t$ that was introduced in (120). Now, by (120)–(121), we have that:

$$d\nu_t^n = \nu_t^n \frac{h_n L_t^n s_n(t)}{1 + h_n L_t^n} d\mathbb{W}_t^n \ .$$

We thus choose $\mu_t = \frac{h_n L_t^n s_n(t)}{1 + h_n L_t^n}$. Hence

$$\begin{aligned} dL_t^{n-1} &= s_{n-1}(t) L_t^{n-1} \left(d\mathbb{W}_t^n - \frac{h_n L_t^n \rho s_n^{\mathsf{T}}(t)}{1 + h_n L_t^n} dt \right) \\ &= \sigma_{n-1}(t) L_t^{n-1} dW_t^{n-1,n} - \frac{h_n L_t^n \sigma_n(t) \rho_{n,n-1}}{1 + h_n L_t^n} \sigma_{n-1}(t) L_t^{n-1} dt \ , \end{aligned}$$

and then likewise, for any $i = 1, \ldots, n$:

$$\boxed{dL_t^i = \sigma_i(t) L_t^i dW_t^{i,n} - \sum_{l=i+1}^n \frac{h_l L_t^l \sigma_l(t) \rho_{l,i}}{1 + h_l L_t^l} \sigma_i(t) L_t^i dt} \tag{122}$$

Note that L^n only is a $\mathbb{P}^n$-martingale. For $i \leq n-1$, L^i has a non vanishing $\mathbb{P}^n$-drift depending on the L^ls for $l > i$.

11.4.1 Correlation Structures

The simplest way to fix ρ is to set it as an historical estimate of the correlation matrix of the Libor rates *(historical correlation structure)*. But such estimates are typically noisy and hardly usable in practice.

Various *parametric forms* for ρ may be used instead, and *calibrated* to market prices of correlation-sensitive instruments, like swaptions (cf. next Subsection). Commonly used parameterizations for ρ are $\rho_{i,l} = \exp(-\gamma|i-l|)$, or (Rebonato):

$$\rho_{i,l} = \rho_\infty + (1-\rho_\infty)\exp\left[-|t_i - t_l|\gamma(t_i,t_l)\right]$$

with $\gamma(t_i,t_l) = d_1 - d_2\max(t_i,t_l)$ (note that all such matrices are not correlation matrices, however), or (Coeffey et Schoenmakers):

$$\rho_{i,l} = \exp\left[-\frac{|i-l|}{n-1}\left(\ln\nu_\infty + \eta_1\varphi(n,i,l) + \eta_2\psi(i,l,n)\right)\right],$$

with

$$\varphi(i,l,n) = \frac{i^2+l^2+il-3ni-3nl+3i+3l+2n^2-n-4}{(n-2)(n-3)}$$
$$\psi(i,l,n) = \frac{i^2+l^2+il-ni-nl-3i-3l+3n+2}{(n-2)(n-3)}$$

The latter parameterization always produces a correlation matrix provided $0 \le \eta_2 \le 3\eta_1$ and $0 \le \eta_1+\eta_2 \le -\ln\nu_\infty$, and is reported to be qualitatively acceptable and robust.

11.5 Swaptions

Recall that an *interest-rate swap* with tenor $t_1,\dots,t_{n+1}$ is a contract with cash flows $h_i(L^i_{t_i} - K)$ at the t_{i+1}s, for $i = 1,\dots,n$ (from the point of view of the party receiving the floating payments). By (82), the value of the swap at time $t \le t_1$ is thus given by

$$\sum_{i=1}^n h_i B^i_t \mathbb{E}^i\left[(L^i_{t_i} - K)\,|\,\mathcal{F}_t\right] = \sum_{i=1}^n h_i B^i_t (L^i_t - K)\,,$$

by the $\mathbb{P}^i$-martingale property of L^i, $i = 1,\dots,n$. A *swaption* (or option to enter the swap at the maturity date t_1) with tenor $t_1,\dots,t_{n+1}$ and strike K is thus tantamount to a product paying at t_1 *the positive part* of

$$\sum_{i=1}^n B^i_{t_1} h_i (L^i_{t_1} - K)\,,$$

that is

$$\boxed{(\textstyle\sum_{i=1}^n h_i B^i_{t_1})(S_{t_1} - K)^+}$$

where the *(forward) swap rate* factor process S is defined by, for $t \le t_1$:

$$\boxed{S_t = \frac{B^0_t - B^n_t}{\sum_{i=1}^n h_i B^i_t}}$$

(hence $\sum_{i=1}^n B^i_{t_1} h_i L^i_{t_1} = B^0_{t_1} - B^n_{t_1} = \sum_{i=1}^n B^i_{t_1} h_i S_{t_1}$, by (115)). Note that S is a martingale under the pricing measure $\mathbb{P}^*$ corresponding to the the numeraire $N_t = \sum_{i=1}^n h_i B^i_t$.

The Libor Swap Model (LSM) mentioned in the introductory paragraph to this section, consists precisely in modeling the swap rate S as a $\mathbb{P}^*$ – Brownian martingale (log-normal process) with time-deterministic volatility, hence exact Black formulae for swaptions prices and deltas follow in the LSM.
The swap rate S_{t_1} is *not* $\mathbb{P}^*$-log-normal in the BGM model, yet it is numerically close to log-normal, with integrated squared variance Σ^2 given by the so-called *Rebonato's formula* (see, e.g., [47, p. 248]):

$$\Sigma^2 = \frac{1}{S_0^2}\sum_{i,l=1}^{n} w^i w^l L_0^i L_0^l \rho_{i,l}\int_0^{t_1}\sigma_i(t)\sigma_l(t)dt =: \int_0^{t_1}\sigma_0^2(t)dt\ , \tag{123}$$

where the weights w^l are proportional to the $h_l B_0^l$ (recall that the B_0^l are assumed to be known in this model, in which the underlying primary risky market typically includes the $n+1$ discount bonds with maturities t_i, $i=1,\dots,n+1$, which are used as numeraires for pricing the caplets). We thus have the following *approximate Black formulae* for swaptions in the BGM model (at time 0):

$$P_0 = N_0\pi^{bl}(t_1,K;0,S_0,\sigma_0) = \left(\textstyle\sum_{i=1}^n h_i B_0^i\right)\left[S_0\mathcal{N}(d_1) - K\mathcal{N}(d_2)\right]$$
$$\Delta_0 = \partial_S\pi^{bl}(t_1,K;0,S_0,\sigma_0) = \mathcal{N}(d_1)$$

with

$$d_1 = \frac{\ln(\frac{S_0}{K}) + \frac{\Sigma^2}{2}}{\Sigma},\quad d_2 = d_1 - \Sigma$$

where Σ is given by (123).

11.6 Model Simulation

Interest rate derivatives discounted payoff processes (working under the terminal measure $\mathbb{P}^n$ as usual) are typically defined as suitable functions ϕ of the factors $(L_{t_j}^i)_{1\le i\le j\le n}$. We thus have for a cap, by (80):

$$\begin{aligned} C_0 &= B_0^n\sum_{i=1}^n \mathbb{E}^n\left[\frac{h_i(L_{t_i}^i - K)^+}{B_{t_{i+1}}^n}\right] \\ &= B_0^n\mathbb{E}^n\left[\sum_{i=1}^n \frac{h_i(L_{t_i}^i - K)^+}{B_{t_{i+1}}^n}\right] \\ &= B_0^n\mathbb{E}^n\left[\sum_{i=1}^n h_i(L_{t_i}^i - K)^+\prod_{l=i+1}^n(1+h_l L_{t_{i+1}}^l)\right]\ , \end{aligned}$$

by (116). To properly discount each cash-flow $h_i(L_{t_i}^i - K)^+$ under $\mathbb{P}^n$, we thus need to know the values of the $L_{t_{i+1}}^l$ for $l > i$, beyond that of $L_{t_i}^i$.

For pricing and Greeking by Monte Carlo in the BGM model, it is thus sufficient to $\mathbb{P}^n$-simulate the $L_{t_j}^i$, $1\le j\le i\le n$. To fix ideas, assume the pricing date $0=:t_0\le t_1$. For many purposes (such as pricing swaptions by Monte Carlo in the LMM, see [101]), Discretizing (the logarithmic form of) (122) by a standard Euler scheme (see section 33.3) at tenor dates is accurate enough (provided of course t_0 is close enough to the left of t_1, otherwise the

discretization must be refined between t_0 and t_1). Starting from L_0 given or extracted from $(B_0^i)_{0\leq i\leq n}$ using (114), one thus gets for $l=0,\ldots,n-1$, $i=l+1,\ldots,n$:

$$\boxed{L^i_{t_{l+1}} = L^i_{t_l}\exp\left[\sigma_i(t_l)\sqrt{h_l}(\Lambda\varepsilon)_i - \left(\sum_{k=i+1}^{n}\frac{h_k L^k_{t_l}\sigma_k(t_l)\rho_{k,i}}{1+h_k L^k_{t_l}}\sigma_i(t_l)+\frac{1}{2}\sigma_i^2(t_l)\right)h_l\right]}$$

where $\varepsilon=(\varepsilon_1,...,\varepsilon_n)^\mathsf{T}$ Gaussian standard and $\Lambda\Lambda^\mathsf{T}=\rho$. For example, in case $n=2$:

$$\mathbb{W}^2=\begin{pmatrix}W^{1,2}\\W^{2,2}\end{pmatrix},\quad d\langle W^{1,2},W^{2,2}\rangle_t=\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}dt=\Lambda\Lambda^\mathsf{T}dt$$

with $\Lambda=\begin{pmatrix}1&0\\\rho&\sqrt{1-\rho^2}\end{pmatrix}$.

In general, the computation goes like this:

t	0	t_1	t_2	$\ldots$	t_n
L^1	L^1_0	$L^1_{t_1}$			
L^2	L^2_0	$L^2_{t_1}$	$L^2_{t_2}$		
L^3	L^3_0	$L^3_{t_1}$	$L^3_{t_2}$		
$\ldots$		$\ldots$		$\ldots$	
L^n	L^n_0	$L^n_{t_1}$	$L^n_{t_2}$	$\ldots$	$L^n_{t_n}$

For instance, assuming for notational simplicity a correlation structure of rank one ($\rho_{i,l}=1$ and $\Lambda_{i,l}=\mathbb{1}_{l=1}$ for any i,l, so ρ completely disappears from the picture and it is enough to make one standard univariate Gaussian draw ε per time step), and with $n=4$, $\sigma_i=15\%$, $h_i=0.5$, and $L^i_0=5\%$ (initial term structure flat at level 5%):

t	0	t_1	t_2	t_3	t_4
$\sqrt{h_l}\varepsilon_l$	-0.371379	1.81768	-0.204069	0.512108	
L^1	5%	4.698%			
L^2	5%	4.699%	6.135%		
L^3	5%	4.701%	6.138%	5.918%	
L^4	5%	4.702%	6.142%	5.923%	6.36%

12 One Factor Gaussian Copula Model

We now move to the context of Multi-Name Credit. Let us thus be given d reference entities (firms), with respective default times denoted by τ_l, $l=1,\ldots,d$. Note that in this context, no standard dynamic model of a relevant primary market model (which might typically consist of CDS index contracts on the pool of reference entities, or of individual CDS contracts on the reference entities, see section 13.2.2) nor of a Markovian factor process (possibly given as the portfolio loss process, or, alternatively, as the family of the individual loss processes) have emerged yet. The benchmark model is the one factor Gaussian copula model, or Li's model [133] (see also Laurent–Gregory [128]). In this approach, one postulates a particular form for the joint cumulative distribution function F of $(\tau_1,\ldots,\tau_d)$, which is enough for deriving

semi-closed pricing formulas *at time 0* for standard multi-name credit derivatives like First-to-Default Swaps or CDOs – but not enough for specifying explicit hedging strategies, nor for pricing more complex *correlation derivatives*.

We thus consider d firms with respective nominal N_l, recovery rate R_l, loss given default $M_l = (1 - R_l)N_l$, and default time $\tau_l > 0$ with distribution function $F_l(t) = \mathbb{P}(\tau_l \leq t)$ and indicator process $H_t^l = \mathbb{1}_{\{\tau_l \leq t\}}$, $l = 1, \ldots, d$. The cumulative portfolio loss at time t is thus given by

$$\boxed{\mathcal{L}_t = \sum_{l=1}^d M_l H_t^l}$$

Finally we denote by

$$F(t_1, \ldots, t_d) = \mathbb{P}(\tau_1 \leq t_1, \ldots, \tau_d \leq t_d)$$

the joint cumulative distribution function of $(\tau_1, \ldots, \tau_d)$.

12.1 Single Tranche CDOs

A single tranche CDO (Collateralized Debt Obligation) with lower attachment point K, upper attachment point $\mathcal{K}$ and maturity T is a contract with the following cumulative discounted cash flow, from the point of view of the investor (which is typically the credit protection seller who bets on perceiving the fees and no default occurs, here):

$$\boxed{\int_0^T \beta_t \left[\Sigma(\mathcal{K} - K - L_t)dt - dL_t\right]}$$

where:
• $L_t = \min\left((\mathcal{L}_t - K)^+, \mathcal{K} - K\right)$ is the cumulative *tranche loss*,
• Σ is the *tranche spread* at time 0, and
• $\beta_t = \exp(-\int_0^t r_u\, du)$ is the risk-neutral discount factor (inverse of the savings account, see Part II).
Note that $\min\left((\mathcal{L}_t - K)^+, \mathcal{K} - K\right) = (\mathcal{L}_t - K)^+ - (\mathcal{L}_t - \mathcal{K})^+$, so a CDO tranche can be interpreted as a call-spread with strikes $K, \mathcal{K}$ on the porfolio loss $\mathcal{L}_t$.

Equity (resp. *senior*) tranches refer to tranches with lower attachment point $K = 0$ (resp. $K > 0$). For instance, on the DJ iTraxx market (a family of CDS indices for Europe and Asia), CDO tranches are quoted for $(K, \mathcal{K})$ equal to $(0\%, 3\%)$, $(3\%, 6\%)$, $(6\%, 9\%)$, $(9\%, 12\%)$ and $(12\%, 22\%)$.

By arbitrage, the price process of a tranche is thus given by

$$\mathbb{E}_t \int_0^T \beta_t \left[\Sigma(\mathcal{K} - K - L_t)dt - dL_t\right] \ , \ t \in [0, T]$$

under a risk-neutral probability $\mathbb{P}$ on the primary market (see Part II). The tranche spread Σ at time 0 is typically set such that the tranche is entered at no cost at inception, so

$$\boxed{\Sigma = \frac{\mathbb{E}\int_0^T \beta_t dL_t}{\mathbb{E}\int_0^T \beta_t(\mathcal{K} - K - L_t)dt}}$$

Note that, assuming deterministic interest rates r, the values of either leg of the CDO only depends on the expected tranche losses $\mathbb{E}L_t$, $t \in [0, T]$. Indeed, we have for the fees leg:

$$\boxed{\mathbb{E}\int_0^T \beta_t(\mathcal{K} - K - L_t)dt = \int_0^T \beta_t(\mathcal{K} - K - \mathbb{E}L_t)dt} \tag{124}$$

For the protection leg, observe that $d(\beta_t L_t) = \beta_t(dL_t - r_t L_t dt)$, and $L_0 = 0$. Hence Fubini's theorem yields:

$$\boxed{\mathbb{E}\int_0^T \beta_t dL_t = \beta_T \mathbb{E}L_T + \int_0^T r_t \beta_t \mathbb{E}L_t dt} \tag{125}$$

12.2 Li Model

By definition, a *copula function* C is the joint cumulative distribution function of an $\mathbb{R}^d$-valued random vector with uniform marginals on $[0,1]$. Thus in particular $C(1,\ldots,1,u_l,1,\ldots,1) = u_l$, for any $u_l \in [0,1]$. *Sklar's theorem* states that for any joint multivariate cumulative distribution function $F(t_1,\ldots,t_d)$ with marginal cumulative distribution functions $F_1(t_1),\ldots,F_d(t_d)$, there exists a copula function C such that

$$F(t_1,\ldots,t_d) = C\left[F_1(t_1),\ldots,F_d(t_d)\right] \ . \tag{126}$$

In the one-factor Gaussian copula model (market model for CDOs), it is postulated that the dependence structure between the τ_l is defined by the Gaussian copula

$$C_\rho(u_1,..,u_d) = \mathcal{N}_\rho\left[\mathcal{N}^{-1}(u_1),..,\mathcal{N}^{-1}(u_d)\right] \ , \tag{127}$$

where $\mathcal{N}$ and $\mathcal{N}_\rho$ respectively stand for the standard (univariate) Gaussian cumulative distribution function and the d-variate Gaussian cumulative distribution function with covariance (correlation, in fact) matrix

$$\begin{pmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & \rho \\ \rho & \ldots & \rho & 1 \end{pmatrix} \ . \tag{128}$$

Let us define, for every $l = 1,\ldots,d$ and $t \geq 0$,

$$\boxed{X_l = \mathcal{N}^{-1}(F_l(\tau_l)) \ , \ x_t^l = \mathcal{N}^{-1}(F_l(t))}$$

So $\tau_l = F_l^{-1}(\mathcal{N}(X_l))$, and $\tau_l \leq t$ iff $X_l \leq x_t^l$. Using also (126),(127), it follows:

$$\mathbb{P}(X_1 \leq x_{t_1}^1,\ldots,X_d \leq x_{t_d}^d) = \mathbb{P}(\tau_1 \leq t_1,\ldots,\tau_d \leq t_d) = F(t_1,\ldots,t_d) = \mathcal{N}_\rho\left[x_{t_1}^1,\ldots,x_{t_d}^d\right] \ .$$

Therefore $X = (X_1,\ldots,X_d)$ is a Gaussian vector with covariance (correlation) matrix given by (128). Equivalently,

$$\boxed{X_l = \sqrt{\rho}V + \sqrt{1-\rho}\,\varepsilon_l \ , \ l = 1,\ldots,d} \tag{129}$$

for independent standard Gaussian random variables V (called the *common factor*) and $\varepsilon_1,\ldots,\varepsilon_d$. The X_l (and then the τ_l) are thus *conditionally independent*, given V. Moreover

$$p_t^{l|v} := \mathbb{P}(\tau_l \leq t \,|\, V = v) = \mathbb{P}(X_l \leq x_t^l \,|\, V = v) = \mathcal{N}\left(\frac{x_t^l - \sqrt{\rho}v}{\sqrt{1-\rho}}\right), \tag{130}$$

and for any $(t_1, \ldots, t_d)$

$$F(t_1, \ldots, t_d) = \mathbb{E}\left[\mathbb{P}\left[X_1 \le x^1_{t_1}, \ldots, X_d \le x^d_{t_d} \,|\, V\right]\right]$$
$$= \int_{-\infty}^{\infty} \prod_{l=1}^{d} p_{t_l}^{l|v} g(v) dv \ ,$$

by conditional independence, where g denotes the standard Gaussian density. Likewise, the *moment generating function* of the portfolio loss $\mathcal{L}_t$ satisfies

$$\Psi_{\mathcal{L}_t}(u) = \mathbb{E}[e^{u\mathcal{L}_t}] = \mathbb{E}[\mathbb{E}[e^{u\mathcal{L}_t}|V]] \tag{131}$$
$$= \mathbb{E}[\prod_{l=1}^{d} \mathbb{E}[e^{uM_l H_t^l}|V]] = \int_{-\infty}^{\infty} \prod_{l=1}^{d} \left(1 - p_t^{l|v} + p_t^{l|v} e^{uM_l}\right) g(v) dv \ .$$

In particular, in the *homogenous case* $p_t^{l|v} = p_t^v$, $M_l = M$:

$$\Psi_{\mathcal{L}_t}(u) \quad = \quad \sum_{l=0}^{d} \int_{-\infty}^{\infty} C_d^l (1 - p_t^v)^l (p_t^v)^{d-l} e^{u(d-l)M} g(v) dv \ . \tag{132}$$

12.3 Exact Methods

Assume for a start that the losses M_l, $l = 1, \ldots, d$, are commensurate, or, more precisely without loss of generality, M_l. We thus have, with $\mathcal{M} = \sum_{l=1}^{d} M_l$ and $q_t^k = \mathbb{P}(\mathcal{L}_t = k)$, $k = 0, \ldots, \mathcal{M}$:

$$\Psi_{\mathcal{L}_t}(u) = \sum_{k=0}^{\mathcal{M}} q_t^k e^{uk} \ , \ \mathbb{E}L_t = \sum_{k=0}^{\mathcal{M}} \left((k-K)^+ \wedge (\mathcal{K} - K)\right) q_t^k \ . \tag{133}$$

So the expected tranche loss $\mathbb{E}L_t$ is a simple function of the portfolio loss probabilities q_t^k, $k = 0, \ldots, \mathcal{M}$. Moreover the latter can be computed by Fourier inversion from the values of $\Psi_{\mathcal{L}_t}(u)$ obtained by (131) for $\mathcal{M}+1$ well chosen values of the argument u, in time $\mathrm{O}(\mathcal{M}^2)$. Choosing $\mathcal{M}+1$ as a power of 2 (completing the involved vectors by zero padding if necessary, see [71, 156]), the Fourier inversion may be realized in time $\mathrm{O}(\mathcal{M} \ln \mathcal{M})$, by FFT.

Alternatively, the portfolio loss probability distribution q_t may be computed through the following *recursive relation between the conditional loss probabilities*, denoted by $q_t^{k|v}(i)$, taking into consideration the i first reference entities only (so $\mathbb{P}(\mathcal{L}_t = k|v) =: q_t^{k|v} = q_t^{k|v}(d)$ and $q_t^{\cdot|v}(0) = \delta_0$):

$$q_t^{k|v}(i) = p_t^{i|v} q_t^{k-M_i|v}(i-1) + (1 - p_t^{i|v}) q_t^{k|v}(i-1) \ , \tag{134}$$
$$i = d, \ldots, 1 \ , \ k = 0, \ldots, \mathcal{M} \ .$$

Indeed we have for any $i = d, \ldots, 1$ and $k = 0, \ldots, \mathcal{M}$:

$$q_t^{k|v}(i) = q_t^{k, H_t^i = 1|v}(i) + q_t^{k, H_t^i = 0|v}(i)$$
$$= q_t^{k-M_i, H_t^i = 1|v}(i-1) + q_t^{k, H_t^i = 0|v}(i-1)$$
$$= q_t^{k-M_i|v}(i-1) p_t^{i|v} + q_t^{k|v}(i-1)(1 - p_t^{i|v})$$

where the two first identities are elementary, and the last one follows by conditional independence w.r.t. V. This proves (134). Once the conditional loss probabilities $q_t^{k|v} = q_t^{k|v}(d)$ have been recursively computed using (134), we recover the portfolio loss probabilities by expectation on V, as

$$q_t^k = \int_{-\infty}^{\infty} q_t^{k|v} g(v) dv.$$

Note that recursive relations analogous to (134) can also be used for *Greeking* (computing sensitivities with respect to the input data $F_l(t)$ for $l = 1, \ldots, d$) in the Li model as follows (see Andersen and Sidenius [7]). One first computes $\partial_{F_l(t)} \mathbb{E}(L_t|V)$ as

$$\big(\partial_{p_t^{l|v}} \mathbb{E}(L_t|V)\big)\big(\partial_{F_l(t)} p_t^{l|v}\big) = \big(\partial_{p_t^{l|v}} \mathbb{E}(L_t|V)\big) \frac{\partial_{x_t^l} p_t^{l|v}}{\partial_{x_t^l} F_l(t)} \ . \tag{135}$$

Now, by (130)

$$\partial_{x_t^l} p_t^{l|v} = \frac{1}{\sqrt{2\pi(1-\rho)}} \exp\Big(\frac{-(x_t^l - \sqrt{\rho} V)^2}{2(1-\rho)}\Big) \ , \ \partial_{x_t^l} F_l(t) = \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-(x_t^l)^2}{2}\Big) \ .$$

Moreover (cf. (133))

$$\partial_{p_t^{l|v}} \mathbb{E}(L_t|V) = \sum_{k=0}^{\mathcal{M}} \big((k-K)^+ \wedge (\mathcal{K} - K)\big) \partial_{p_t^{l|v}} q_t^{k|v} \tag{136}$$

in which (cf. (134))

$$\partial_{p_t^{l|v}} q_t^{k|v} = q_t^{k-M_l|v}(\{l\}) - q_t^{k|v}(\{l\})$$

where $q_t^{\cdot|v}(\{l\})$ denotes the conditional distribution of the loss of the portfolio of all names but l, which can be computed recursively along the same lines as (134). One finally recovers the unconditional sensitivity $\partial_{F_l(t)} \mathbb{E} L_t$ by taking expectation with respect to V in (135).

12.4 Approximate Procedures

As opposed to the previous *exact procedures*, much faster *fast approximate methods* may be used to compute the $q_t^{k|v}$. Moreover these approximate methods don't require the assumption of commensurate losses.

Regarding an efficient, easy-to-implement and mathematically justified approach based on *Gauss–Poisson approximations* for the portfolio conditional loss distributions, we refer the reader to El Karoui and Jiao [79, 80, 112].

Related (yet heuristic and harder to implement) *saddle-point methods* (see [8, 184]) are based on the following inverse Laplace transform representation for the $q_t^{k|v}$, *in a weak sense* to be detailed below and formally obtained by integration parallel to the imaginary axis in the complex plane, for any $\eta > 0$:

$$\boxed{q_t^{k|v} = \tfrac{1}{2\pi i} \textstyle\int_{\eta - i\infty}^{\eta + i\infty} \Psi_{\mathcal{L}_t}(u|V) \exp(-uk)\, du} \tag{137}$$

where (cf. (131)):

$$\Psi_{\mathcal{L}_t}(u|V) := \mathbb{E}[e^{u\mathcal{L}_t}|V] = \prod_{l=1}^{d} \left(1 - p_t^{l|v} + p_t^{l|v} e^{uM_l}\right) .$$

Identity (137) in the weak sense means (cf. (253)–(254)):

$$\boxed{\mathbb{E}\left[\varphi(\mathcal{L}_t)|V\right] = \frac{1}{2\pi i} \int_{u=\eta-i\infty}^{\eta+i\infty} \Psi_{\mathcal{L}_t}(u|V) \left(\int_{L=0}^{\infty} \exp(-uL)\varphi(L)dL\right) du} \tag{138}$$

(identity in the strong sense, now), for any regular enough function φ such that the conditional expectation exists in (138). One thus gets by application of (138) to $\varphi(\mathcal{L}_t) = (\mathcal{L}_t - K)^+$ (note that $\partial_x \left[(ux+1)e^{-ux}\right] = -u^2 x e^{-ux}$, so $\int_0^\infty x e^{-ux}\, dx = \frac{1}{u^2}$):

$$\boxed{\mathbb{E}\left[(\mathcal{L}_t - K)^+|V\right] = \frac{1}{2\pi i} \int_{u=\eta-i\infty}^{\eta+i\infty} \frac{\Psi_{\mathcal{L}_t}(u|V)\exp(-uK)}{u^2}\, du} \tag{139}$$

Saddle-point methods are then based on the approximation of $\Psi_{\mathcal{L}_t}(u|V)\exp(-uk)$ in (139) by a suitable Taylor expansion around a well chosen point $u^\star$, so that the resulting approximating integral may be computed in closed form, as an integral w.r.t. a Gaussian kernel. Depending on the expansion point $u^\star$ and the order of the Taylor expansion, one thus gets a whole family of approximate algorithms for pricing the tranche. In the simplest case, we recover the so-called *large portfolio approximation* to the tranche price.

A last possibility to compute the portfolio loss probabilities q_t^k, or, directly, the values of the fees leg (124) and of the protection leg (125) of the tranche, is of course to proceed by simulation (cf. Part VI). But simulation methods are much slower on these problems than any of the previous procedures (Gauss–Poisson or saddle-point approximations to estimate the tranche legs, or even, assuming commensurate M_l, exact convolution FFT or recursive methods to compute the portfolio loss distribution).

Note finally that the numerical integrations involved in these algorithms, which all involve the Gaussian kernel $g(v)dv$, may be done very quickly by Gaussian quadrature (like the Gauss–Hermite quadrature, using for instance the function `gauher` of [156]; see also [71]).

13 Benchmark Models in Practice

13.1 Implied parameters

The reader should not misunderstand the meaning of the previous "benchmark models". In fact, the Black(–Scholes) closed formulae for single-name derivatives, or the Li semi-closed formulas for multi-name credit, are essentially used by traders for *conveying information* on the relative value of the various derivatives available in the market, in a unit of measurement (Black–Scholes implied volatility or Li implied correlation) less sensitive to the characteristics (maturity, strike(s), etc.) of the product at hand than its money-value. So these formulas are but "wrong formulas into which to put a wrong number [the implied volatility of an option or the implied correlation of a tranche] to get the right result [an option market price or a tranche market spread]".

13.1.1 Black(–Scholes) Implied Volatility

Black(–Scholes) formulae are applicable to any European vanilla option on virtually any market (except for multi-name credit): stock, index, future or exchange rate, forward interest or swap rate, etc. But these formulae are in fact used in the 'reverse-engineering' mode that consists in determining, given an European vanilla price observed on the market, the corresponding value of the volatility consistent with that option price through the Black(–Scholes) formula.

More precisely, given the values of r and q inferred from the related riskless bond market for r and by call-put parity for q, the *Black–Scholes implied volatility* of an (European vanilla) option is the value σ_t such that:

$$\boxed{\Pi^{bs}(T,K;t,S_t,r,q,\sigma_t)=\Pi_t^{ma}(T,K)} \tag{140}$$

where $\Pi_t^{ma}(T,K)$ denotes the market price of the option at time t. Since $\Pi^{bs}(T,K;t,S_t,r,q,\sigma)$ is monotone (increasing) w.r.t. σ (in the case of an European vanilla), equation (140) is well-posed in σ_t (provided the market price lies within the related arbitrage bounds) and easy to solve numerically (using for instance a simple dichotomy method or the Newton–Raphson algorithm).

Proceeding in this way for a range of strikes K, one gets the so-called *Black(–Scholes) implied volatility* smile (typical on foreign exchange derivatives markets), *smirk* (typical on interest rates derivatives markets) or *negative skew* (typical on equity derivatives markets) (or *positive skew* on 'negative beta' assets derivatives markets, like options on gold futures, see Figure 2).

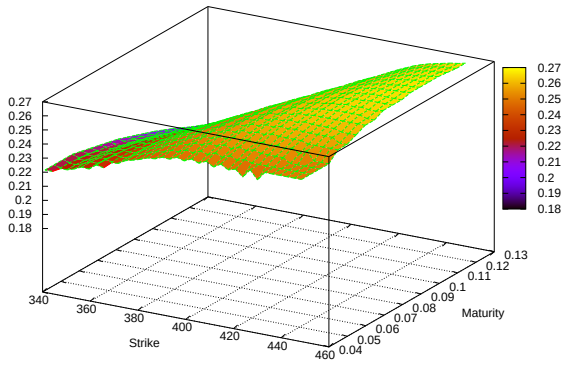

Figure 2: *Golden smile, October 6 2003*

13.1.2 Li implied correlation

The Li model (one factor Gaussian copula model) is the benchmark model for multi-name credit. Yet, as the Black formula for single-name derivatives, it is also used in the reverse-engineering mode for quoting CDO tranches in terms of their *Li implied correlations*, given

the cumulative distribution functions F_l inferred from the respective marginal CDS markets. As for the last point, we have the following arbitrage relations (see [100, 33]):

$$\Sigma^l(T)\int_0^T (1-F_l(t))\,dt - \int_0^T M_l\,dF_l(t) = 0 \text{ for any } l = 1\dots d \text{ and } T \geq 0\,, \qquad (141)$$

where $\Sigma^l(T)$ denotes the fair spread at time 0 of the CDS with maturity T on the l^{th} name of the reference pool. Added to the fact that $F_l(0) = 0$, this allows one to bootstrap the marginal default time cumulative distribution functions F_l from the individual CDS curves at time 0, $\Sigma^l(T)$ (assumed to be observable for every $T \geq 0$). Thus, given the F_l:

• the *compound implied correlation* of a tranche is the value of the correlation $\widetilde{\rho}_t$ in a Li model such that

$$\boxed{\Sigma^{li}(T,K,\mathcal{K};t,(F_l)_{1\leq l\leq d},\widetilde{\rho}_t) = \Sigma_t^{ma}(T,K,\mathcal{K})}$$

where $\Sigma_t^{ma}(T,K,\mathcal{K})$ denotes the market tranche spread at time t;

• the *base implied correlation* of a tranche is the value of the correlation ρ_t in a Li model such that

$$\boxed{\Sigma^{li}(T,0,\mathcal{K};t,(F_l)_{1\leq l\leq d},\rho_t) = \Sigma_t^{ma}(T,0,\mathcal{K})}$$

where $\Sigma_t^{ma}(T,0,\mathcal{K})$ denotes a synthetic market spread computed from the observed market spreads for the tranches with upper attachment point $\mathcal{K}$ and below (see, e.g., [150]).

Base implied correlation is more stable numerically than compound implied correlation, because $\Sigma_t^{li}(T,K,\mathcal{K};t,(F_l)_{1\leq l\leq d},\rho)$ is monotone (decreasing) w.r.t. ρ for $K = 0$, but not for $K > 0$ [150].

Other models are assessed on their ability to reproduce the so-called *market correlation skew*, for suitably calibrated values of their parameters (see, e.g., Figure 3, in which the model correlation skew was derived in the credit migrations set-up of [32]).

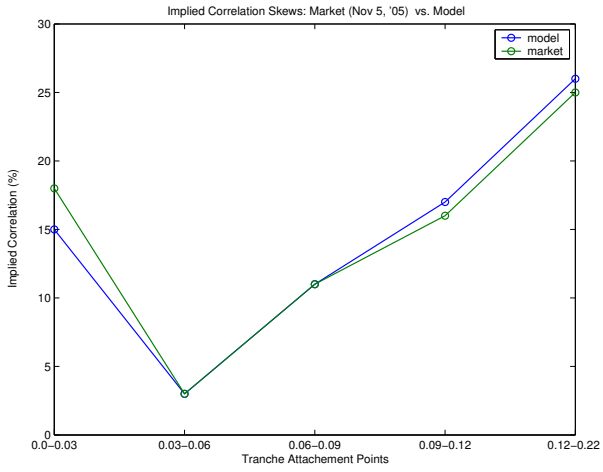

Figure 3: *Market and Model implied correlation skews for CDO tranches.*

13.2 Implied Delta Hedging

The related deltas (Black–Scholes implied volatility delta or Li implied correlation delta) and may also be used for hedging, yet there is no guarantee that the related hedges fairly account for the *volatility or correlation risk*.

13.2.1 Black(–Scholes) Implied Delta Hedging

Let us thus consider the problem of discretely delta-hedging (see section 9.2), at times $t_i = ih, i = 0 \dots n-1$, an *European vanilla call option* with maturity T on an underlying S. We thus sell the option at time 0 at price Π_0, that is we receive at time 0 the amount of money Π_0, but in turn it is mandatory for us to pay at time T the payoff $\xi = (S_T - K)^+$. Our strategy consists in rebalancing at every time step h a self-financing hedge in the underlying and in the savings account. We assume that dividends are paid and kept as new stock shares falling at yield q in the hedging portfolio. We are thus in fact considering a hedging strategy of the following form (to be compared with (84))

$$\zeta_t = \zeta_{t_i} e^{q(t-t_i)} \text{ on } (t_i, t_{i+1}] \ , \ i = 0 \dots n-1$$

where (cf. (91))

$$d(\beta_t \widehat{S}_t) = d(\beta_t S_t) + \beta_t q S_t dt = e^{-qt} d(\beta_t S_t e^{qt}) \tag{142}$$

resulting in

$$\zeta_t d(\beta_t \widehat{S}_t) = \zeta_{t_i} e^{-qt_i} d(\beta_t S_t e^{qt}) \ . \tag{143}$$

Accounting also for the facts that $V_0 = \Pi_0$ and for the final payment by the trader of ξ at time T, the discounted P&L of the trader at T is therefore (cf. (14)), setting $\widetilde{S}_t = S_t e^{qt}$:

$$\boxed{\beta_T e_T = \Pi_0 + \sum_{i=0}^{n-1} e^{-qih} \zeta_{ih} \left(\beta_{(i+1)h} \widetilde{S}_{(i+1)h} - \beta_{ih} \widetilde{S}_{ih} \right) - \beta_T \xi} \tag{144}$$

or, equivalently:

$$\boxed{\beta_T e_T = \sum_{i=0}^{n-1} \beta_{ih} \delta_i e} \tag{145}$$

with

$$\boxed{\beta_{ih} \delta_i e = -(\beta_{(i+1)h} \Pi_{(i+1)h} - \beta_{ih} \Pi_{ih}) + e^{-qih} \zeta_{ih} \left(e^{-r(i+1)h} \widetilde{S}_{(i+1)h} - e^{-rih} \widetilde{S}_{ih} \right)} \tag{146}$$

The previous identities are valid for any discrete hedging scheme $(\zeta_{ih})_{0 \le i \le n-1}$. The *Black–Scholes implied delta hedging scheme* corresponds to the following choice:

$$\zeta_{t_i} = \Delta^{bs}(T, K; t_i, S_{t_i}, \sigma_{t_i}) \ , \ i = 0 \dots n-1$$

where σ_{t_i} denotes the Black–Scholes implied volatility of the option at time t_i (given extracted values of r and q extracted from the market in a suitable way). The rationale behind this strategy is that, were the spot to obey the following *objective* Black–Scholes dynamics (under the *statistical* measure $\widehat{\mathbb{P}}$):

$$dS_t = S_t \left(\mu dt + \sigma d\widehat{W}_t \right)$$

with constant interest rate r and dividend yield q on S, then we would have $\sigma_t \equiv \sigma$, and the previous P&L would go to 0 identically as $h \to 0$. But of course the Black–Scholes

model is strongly misspecified, so that this is far from being true in practice (see the general discussion at the end of section 9.2, and the related simulation issues in Section 35).

To get a better hedge, one can alternatively use a *finite difference Black–Scholes implied delta* in which the Black–Scholes volatility used for valuing the option at the perturbed initial value of the stock is suitably modified with respect to the reference Black–Scholes implied volatility, in order to account for the correlation between stock returns and implied volatility changes. For instance, the *sticky delta rule* stipulates that the implied volatility surface does not move with calendar time t (or moves deterministically with t) provided it is parameterized as a (random) function of time to maturity $T-t$ and moneyness $\ln(\frac{K}{S_t})$ (rather than T and K).

13.2.2 Li Implied Delta Hedging

Our next aim is to hedge the *risk of spread* on a CDO tranche between two successive default times of the reference entities. Note that we do not aim at hedging defaults in this approach (cf. the discussion following Theorem 7.3; for an alternative approach consisting in hedging the defaults only, see Laurent et al. [127]).
The most common hedge is delta-neutral with respect to homogenous bumps on homogenous time buckets of the underlying CDS curves, using *CDS index contracts* as hedging instruments. We recall that a CDS index contract is an insurance contract covering default risk on the pool of names in the index. Index contracts differ slightly from single-name securities. In the case of a credit event, the related entity is removed from the index and the contract continues (with a reduced notional amount) until maturity. CDS index contracts may be considered as kinds of averages of individual CDSs, and they can be priced essentially like the latter, using a relation of the form (141).
Our present goal is to hedge deformations scenarii of the underlying CDS curves given as homogenous bumps on homogenous time buckets of all curves. By using, alternatively, individual CDSs as hedging instruments, one could hedge if wished pertaining deformations scenarii of the related *individual* CDS curves.
Note that the involved CDSs (whether they are CDS index contracts or individual CDSs) are in fact new CDSs freshly emitted at each rebalancing time t.

Given a CDO tranche with maturity T (and strikes $K, \mathcal{K}$), we thus rebalance, every time step h(=1 market day, typically) between two successive default events of the reference entities, a hedging position in a primary market consisting of the savings account $B_t = \beta_t^{-1}$, and of d *CDS index contracts* with increasing maturities T_j, $j = 1, \ldots, d$ (where $T_{d-1} \leq T$). Considering a trader who is short one tranche (sold, i.e. tranche protection was bought) and long ζ_t^j units in the CDS index contract with maturity T_j, the discounted P&L increment of the hedged position on a time interval $(t, t+h)$ *without defaults* writes, cf. (14) (neglecting transaction costs as usual, and with all nominals conventionally taken equal to one):

$$\begin{aligned}\beta_t \delta e &= -\beta_t \delta e^\star + \sum_j \zeta_t^j \beta_t \delta e^j \\ &= (\beta_{t+h} P_{t+h} - \beta_t P_t) - \Sigma_0^*(\beta_{t+h} Q_{t+h} - \beta_t Q_t) - \beta_t \Sigma_0^* h \\ &\quad - \sum_j \zeta_t^j \left((1-R)\beta_{t+h} P_{t+h}^j - \Sigma_t^j \beta_{t+h} Q_{t+h}^j - \beta_t \Sigma_t^j h \right) ,\end{aligned} \tag{147}$$

where:
• $\delta e^\star$ (resp. δe^j) is the increment of the P&L on a unit position on the tranche (resp. on a

unit position on the CDS index contract with maturity T_j),
• Σ_0^* is the spread of the tranche at *inception date* 0, and Σ_t^j is the spread of the CDS index contract with maturity T_j at the *current time* t,
• P and Q, resp. P^j and Q^j, denote the value processes of the Protection and Fees Legs of the tranche, resp. of the CDS index contract with maturity T_j;
• R is a (common and constant) recovery rate on the credit index contracts.

In (147), the terms $\beta_t \Sigma_0^* h$ and $\beta_t \Sigma_t^j h$ account for the *carry* of the various instruments involved, while the remaining terms account for the *slide* of the portfolio between t and $t+h$. Note that the δe^j only depend on the value of the CDS index contracts at time $t+h$. This is due to the fact that the CDS index contracts used for hedging at time t are new CDSs freshly emitted at t. So their value at time t is equal to zero, by definition of the spread Σ_t^j.

Now, in order to hedge the tranche, a strategy based on the Li implied deltas of the tranche consists in setting the (row-)vector of CDS index hedging positions $\zeta_t = (\zeta_t^j)_j$ in (147) as

$$\boxed{\Delta_t^\star = \zeta_t \Delta_t} \tag{148}$$

where $\Delta_t^\star$ and Δ_t respectively stand for the vector of the deltas (sensitivities) of the tranche, and the matrix of the deltas (sensitivities) of the CDS index contracts, with respect to homogenous bp-bumps on the time-buckets $[T_{i-1}, T_i]$ of the underlying CDS curves, for $i = 1, \ldots, d$ (with $T_0 = 0$). Concretely, in (148):
• The Δ_i^js are computed by assessing using (141) the impact of the j^{th} bump on the underlying CDS curves on the Protection and Fees Legs of the i^{th} CDS index contract. Note that the Δ_i^js are independent of the dependence structure (copula function) of $(\tau_1, \ldots, \tau_d)$. Also note that individual CDS deltas (in case where an individual CDSs hedge would be wished) could be computed in the same way;
• The $\Delta_j^\star$s are computed by assessing the impact on the Protection and Fees Legs of the tranche of the j^{th} bump on the underlying CDS curves. More precisely:
(i) one calibrates the marginal cumulative distribution functions F_l to the underlying CDS spread curves using (141), and one computes the related (base) implied correlation ρ_t of the tranche (see section 13.1.2);
(ii) one computes the associated values P and Q of the Protection and Fees Legs of the tranche;
And then, for $j = 1, \ldots, d$:
(iii) one recalibrates as in (i) the marginal cumulative distribution functions $\widetilde{F}_l$ to the underlying CDS curves bumped by +1bp on the time interval $[T_{j-1}, T_j]$, $T_{j-1} \leq T$;
(iv) one computes the values $\widetilde{P}$ and $\widetilde{Q}$ of the Protection and Fees Legs of the tranche in the Li model with marginal cumulative c.d.f. $\widetilde{F}_l$ and correlation parameter obtained by an ad hoc adjustment of ρ_t, meant to account for the correlation between spread moves and implied correlation changes. For instance, the *sticky delta rule* stipulates that the implied correlation smile does not move with calendar time t (or moves deterministically with t) provided it is parameterized as a (random) function of time to maturity $T - t$ and of a suitable notion of *moneyness* of the tranche (which is defined in analogy with the notion of moneyness for vanilla stock options);
(v) one sets

$$\Delta_j^\star = \Sigma_0^*(\widetilde{Q} - Q) - (\widetilde{P} - P) \, .$$

The matrix Δ is obviously lower triangular, so that the linear system (148) is elementary.

To provide a rationale for the previous choice of ζ (and also, to be able to quantify the hedging error (147) by closed-form computations or numerical simulation), one needs to specify a *model dynamics*, beyond the distribution of the portfolio losses at time 0 as defined by the Li Model. Let us thus assume that the deformations of the individual CDS curves are driven by homogenous deformations $d\mathcal{X}_t^i$ on the time buckets $[T_i, T_{i+1}]$ of the curves, for $i = 1, \ldots, d$, where $\mathcal{X}$ may be part of a more general *factor process* $\mathcal{Z} = (t, \mathcal{X}, \mathcal{Y})$ with drivers W, ν, μ as of section 7.2. So the price of the tranche Π_t, resp. of the credit index contracts X_t, is given by $\Pi(\mathcal{Z}_t)$, resp. $X(\mathcal{Z}_t)$. By application of formula (61), it is thus possible to hedge completely the spread risk in this model (represented by the Brownian noise W among the model drivers) by setting, for $t \in [0, T]$ (assuming ∂X and σ invertible in model $\mathcal{Z}$, cf. Subsection 7.3)

$$\zeta_t = \partial\Pi \partial X^{-1}(\mathcal{Z}_t)$$

We thus recover (148), inasmuch as $\Delta_t^\star$ and Δ_t may be considered as reasonable proxies for $\partial\Pi(\mathcal{Z}_t)$ and $\partial X(\mathcal{Z}_t)$. This of course depends a great deal on the way ρ_t is readjusted at step (iv) of the above procedure for computing $\Delta_t^\star$.

As a conclusion to this Part, we wish to emphasize that for a consistent risk-management of financial derivatives, the benchmark models are not good enough. More realistic models are needed (like the Bates model of section 10.4 or beyond, in the case of equity derivatives; see also the generic Markovian model (28)–(30) in Part II). In such models, and/or for pricing and Greeking *exotic products* in *any* model, *numerical procedures* are the only way. *Numerics* will thus be the subject of the remaining Parts of these notes. For simplicity of the presentation we shall present most methods on simple models, like the *Black–Scholes model* (in general), the *Merton model* or the *Heston model*. But of course, the methods themselves are always generic to some degree.

Part IV
Finite Differences Pricing Methods

Note that the *deterministic* pricing methods that we present here: Partial (Integro-)Differential Equation (P(I)DE) methods in this part or tree methods in the next one, can be used in any Markovian model, irrespective of any early exercise features, but with the practical limitation that the space dimension of the model be not too large (no more than three, say; cf. Section 3).

14 Generic Pricing PIDE

First of all, how comes that deterministic methods come into play to compute derivative prices that are in essence *stochastic processes*? In order to answer this question, let us summarize some of the findings of the previous parts.

In a generic Markovian risk-neutral market model $\mathcal{Z} = (t, \mathcal{X}, \mathcal{Y})$, where $\mathcal{X}$ is an $\mathbb{R}^q$-valued jump-diffusion – like component and $\mathcal{Y}$ is a Continuous-Time Markov chain – like component with state space $I = \{y^1, \ldots, y^k\}$, financial derivatives' prices and Greeks can be expressed in terms of the so-called *pricing function*, characterized as the unique solution in some sense to a related risk-neutral *system of pricing P(I)DEs*. In the simplest case this pricing P(I)DE reduces to the well-known Black–Scholes equation, with solution explicitly given by, in the case of European vanilla options, the usual Black–Scholes formulae. Moreover by considering an arbitrary *numeraire* (which is given by the savings account, in the case of risk-neutral models), it is possible to extend the previous statements to more general *relative* (instead of risk-neutral) market models, where "relative" here is a short-hand for *relative to a certain numeraire*. This more general approach comes out, in the simplest case, into the so-called Black formulae, that allow one to cope with stochastic interest rates and discount factors.

The previous approaches thus cover the basic needs for equity and fixed-income derivatives modeling, for which a reasonable model reduces to a (scaled) Brownian motion for the returns $\mathcal{X}$ of the underlying asset (so q is equal to one and there are no jumps in the model, in particular no Markov Chain $\mathcal{Y}$ is needed, meaning that $k = 1$ in the above formalism). Of course, in the context of *credit risk*, financial derivatives are typically given as derivatives on a loss process $\mathcal{L}$, which in the simplest case reduces to a standard Poisson process. So the main component in the model here consists of the jump part of $\mathcal{X}$, or a Markov Chain–like component $\mathcal{Y}$ in case of more general credit derivatives on the *ratings* on a pool of reference names (or for a model of portfolio loss derivatives susceptible to account for the impact of implied ratings on the product at hand, see Bielecki et al. [32]). In the simplest case of an underlying *Poisson process or Markov chain* the related pricing equation reduces to a related *system of Ordinary Differential Equations*.

Let us thus be given, in the jump-diffusion setting with regimes $\mathcal{Z} = (t, \mathcal{X}, \mathcal{Y})$ of section 7.2, a financial derivative with payoffs given as suitable *Borel-measurable functions* of $\mathcal{Z}$. Then under mild technical conditions (see Part II), the $\mathbb{P}$ – *arbitrage price* Π_t of a financial derivative may be represented as $\Pi_t = u(\mathcal{Z}_t)$, for a Borel-measurable (deterministic) function u of t and $\mathcal{X}_t$. So *all the randomness in* Π *is embedded in that of* $\mathcal{Z}$. The related pricing problem is thus reduced to the computation of the (deterministic) function $u^i(t, x) := u(t, x, y^i)$. This is the basic device through which deterministic methods come into play to compute

derivative prices that are in essence stochastic processes.

Now, a generic Markov process $\mathcal{Z}$ admits an *infinitesimal generator* $\mathcal{A} = \mathcal{A}_{\mathcal{Z}}$ with a *canonical structure* in three parts: a *drift*, a *diffusion* and a *jump* component (where the jump component may be given as a Poisson – like or/and a Markov chain – like component). This canonical structure of the generator of a Markov process $\mathcal{Z}$ implies that the $\mathbb{P}$ – *pricing function* u solves a system (because of the Markov chain – like component, if any) of parabolic PIDEs of the following type on the time-space domain $E = [0,T] \times \mathbb{R}^q \times I$, where T is the maturity of the product (cf. the generic pricing equations (57) or (70) and further comments below):

$$\boxed{\begin{cases} F_i(t,x,u(t,x),\partial_t u^i(t,x),\mathcal{I}u^i(t,x),\partial u^i(t,x),\mathcal{H}u^i(t,x)) = 0 \text{ on } [0,T) \times \mathcal{D} \\ u^i(t,x) = \phi^i(t,x) \text{ on } E \setminus \Big([0,T) \times \mathcal{D}\Big) \end{cases}} \tag{149}$$

where:
• $\mathcal{D}$ is an open subset of $\mathbb{R}^q \times I$. In particular the terminal condition at T, which is embedded in the boundary condition ϕ in (149), is given by the derivative payoff at maturity;
• $\mathcal{I}u$ is a *non-local* term defined as

$$\mathcal{I}u^i(t,x) = \gamma(t,x) \int_{\mathbb{R}^q} (u^i(t,x+f(t,x,y)) - u^i(t,x))h(t,x,dy) \ ,$$

for a non-negative *jump intensity function* $\gamma^i(t,x)$ and a *jump size probability measure* $h^i(t,x,dy)$,
• ∂u and $\mathcal{H}u$ denote the (row-)gradient and Hessian matrix of u with respect to x, and $u(t,x) := (u^i(t,x))_{1 \le i \le I}$.

The precise definition of the space domain $\mathcal{D}$ and of the operator F depend on the specifications of the derivative at hand. So, in the case of a *European vanilla* option, $\mathcal{D} = \mathbb{R} \times \mathcal{I}$ (or $(0,\infty) \times I$), and F is a linear operator. In the case of *American* (or *Game*) options, there are further *obstacles* in F. In case there are *barriers* involved, $\mathcal{D}$ is a part of $\mathbb{R}^q \times I$ delimited by the barriers, etc. But the structure of the generator $\mathcal{A}$ of $\mathcal{Z}$ implies that the operator F is always *monotone* in the sense that, for any $(t,x,i) \in [0,T] \times \mathcal{D}$, $u, v \in \mathbb{R}^q$, $I, J, a \in \mathbb{R}$, $p \in \mathbb{R}^{1 \otimes q}$ and $M, N \in \mathbb{R}^{q \otimes q}$:

$$F_i(t,x,u,I,a,p,M) \le F_i(t,x,v,J,a,p,N) \text{ whenever } u \le v, \ I \ge J, \ N \ge M \tag{150}$$

where the inequality $u \le v$ is to be understood componentwise and the inequality $N \ge M$ in the sense of the usual order on the space of real-valued symmetric non-negative matrices (non-negative eigenvalues of $N - M$).

14.1 Maximum Principle

A *classic solution* to the pricing equation (149) is a system $u = (u^j)_{1 \le j \le I}$ of functions $u^j \in \mathcal{C}([0,T] \times \mathbb{R}^q) \cap \mathcal{C}^{1,2}(\{i = j\} \cap ([0,T) \times \mathcal{D}))$ that satisfies (149) everywhere. Likewise, a *classic subsolution* (resp. supersolution) satisfies (149) everywhere, with $=$ replaced by $\le$ (resp. $\ge$) therein.

Let us consider for simplicity the case where $\mathcal{D}$ is bounded and where the monotonicity of F is *strict* in its third argument u, in the sense that the left inequality is strict in (150) as soon as the inequality is strict for at least some component j in $u \le v$ (to be understood componentwise) in (150). We then have the following *comparison principle*.

Proposition 14.1 *$u \le v$, for any classic subsolution u (resp. supersolution v) of (149).*

Proof. The proof essentially relies on the *classic maximum principle*, according to which

$$\partial v(z) = 0,\ \mathcal{H}v(z) \leq 0\ ,$$

for any real-valued function $v \in \mathcal{C}^2(\mathbb{R}^d)$ locally maximum at z.

We proceed by contradiction. Assume that $u \leq v$ does not hold. Then $w = u - v$ admits a positive global maximum at a point $(t, x, i) \in [0, T) \times \mathcal{D}$, so by the classic maximum principle and by definition of $\mathcal{I}u$,

$$\begin{aligned}\partial_t u^i(t,x) = \partial_t v^i(t,x)\ ,\ \mathcal{I}u^i(t,x) \leq \mathcal{I}v^i(t,x)\\ \partial u^i(t,x) = \partial v^i(t,x)\ ,\ \mathcal{H}u^i(t,x) \leq \mathcal{H}v^i(t,x)\ .\end{aligned}$$

Hence the fact that

$$\begin{aligned}F_i(t,x,u(t,x),\partial_t u^i(t,x),\mathcal{I}u^i(t,x),\partial u^i(t,x),\mathcal{H}u^i(t,x)) \leq 0\\ \leq F_i(t,x,v(t,x),\partial_t v^i(t,x),\mathcal{I}v^i(t,x),\partial v^i(t,x),\mathcal{H}v^i(t,x))\end{aligned}$$

implies by monotonicity of F :

$$\begin{aligned}F_i(t,x,u(t,x),\partial_t u^i(t,x),\mathcal{I}u^i(t,x),\partial u^i(t,x),\mathcal{H}u^i(t,x)) \leq\\ F_i(t,x,v(t,x),\partial_t u^i(t,x),\mathcal{I}u^i(t,x),\partial u^i(t,x),\mathcal{H}u^i(t,x))\ ,\end{aligned}$$

which is in contradiction with $u^i(t,x) > v^i(t,x)$, given our strict monotonicity assumption on F. □

Interpreting the problem at hand as comparing the distribution of the temperature in two identical systems of glasses filled with hot liquid, our comparison principle says that the liquid is hotter at a (distributed) inner point of the first system of glasses than at the corresponding point in the second system of glasses, should it be hotter *at any corresponding points outside* the two systems (or at any corresponding points *on the boundary* of the two systems, in the case where there are no jumps in $\mathcal{X}$ (case where $\gamma = 0$).

Proposition 14.1 immediately implies that there may be *at most one* classic solution to the pricing equation (149). In very specific cases, assuming in particular that F is linear, there may be a chance that the pricing equation has a classic solution. However, in general, there is no hope that a classic solution to the pricing equation does exist, and one must resort to suitable notions of *weak* solutions of (149).

14.2 Weak Solutions

We thus need to extend the definition of a solution to (149) so that a solution does exist under rather general circumstances, while preserving uniqueness.

14.2.1 Viscosity Solutions

The theory of *viscosity solutions* [60, 86, 2, 18, 4, 3, 58, 106, 107, 65] defines suitable notions of weak (continuous, or even semi-continuous) solutions, subsolutions and supersolutions of non-linear monotone systems of PIDEs such as (149), such that related maximum and comparison principles still hold true. Thus viscosity solutions of (149) (bounded, or satisfying suitable growth conditions) are unique.

Existence of a viscosity solution of (149) can be established by various means, such as the *Perron's method*, itself based on the related comparison principles.

The precise definition of a viscosity solution to (149) is outside the scope of these notes and, in fact, irrelevant here. It will be enough for us to keep in mind that *under mild technical assumptions, the pricing equation (149) is well-posed (admits a unique solution continuously depending on its input data) in a suitable space of viscosity solutions with growth conditions*. This will enable us to analyze convergence properties of naturally related *finite differences approximation schemes* (including *approximation trees*) for (149).

14.2.2 Weak Solutions in Weighted Sobolev Spaces

Alternatively to working with viscosity solutions, it is possible to derive *weak (variational) formulations* of the pricing equation (149) in suitable *weighted Sobolev functional spaces* $\mathcal{H}$. In this approach, the boundary condition ϕ is typically accounted for by a judicious choice of $\mathcal{H}$.
Existence and uniqueness of a solution to the variational formulation of the problem typically results by application of the *Lax–Milgram Theorem*.
Various choices for $\mathcal{H}$ are possible, all giving rise to well-posed variational problems (see, e.g., Bally et al. [17, 16], Barles–Lesigne [20], Achdou–Pironneau [1], Matache et al. [137], Jaillet et al. [105], Ern et al. [84]). The choice made in [20, 17, 16] has the advantage to give rise to a clear connection (Feynman-Kac formula) between the solution of the PDE problem and the option price process as probabilistically defined by arbitrage (see Part II).
Again, the precise variational formulation of the pricing problem, including the definition of the spaces, is outside the scope of these notes and irrelevant here. We shall simply keep in mind that *under mild technical assumptions, a weak (re)formulation of the pricing PIDE problem (149) is well-posed in a suitable weighted Sobolev space* $\mathcal{H}$. This underlies the theory of naturally related *finite elements and finite volumes approximation schemes*.

15 Numerical Approximation

In the case of European vanilla options in simple models (models of the AJD class [75], SABR model [96],etc.), the pricing equation (149) can be solved analytically (or semi-analytically by Fourier analysis, see Section 37). But in general, (149) must be solved numerically. In order to approximate (149), one can either use *finite differences* methods [171], or resort to more general *finite elements* (or even *finite volumes*) methods [1]. Note that there is in fact no hermetic frontier between these methods. Indeed, schematically:

$$\boxed{\text{Tree Methods} \subset \text{FD Methods} \subset \text{FE Methods} \subset \text{FV Methods}}$$

(and also in a sense, for a complete picture: MC Methods $\subset$ Tree Methods, cf. Section 1).

15.1 Finite differences methods

Finite difference methods are naturally connected with viscosity solutions of monotone equations expressing related *maximum principles* (cf. section 14.1). A practical reason to use finite differences is that finite element methods are potentially more powerful, but they are also heavier, than finite differences methods. The related cost is justified in cases where the geometry of the domain makes it necessary to use a sophisticated unstructured and adaptative discretization mesh. But pricing problems in finance are typically posed on rectangular

domains, for which a simple finite difference grid is good enough (with the limitation that a finite differences grid does not offer as many possibilities as a finite elements mesh for refining the mesh in multi-dimensional problems, however).

15.1.1 Localization, Discretization

The numerical resolution of the pricing equation (149) by finite differences is a four steps process:
(1) *Transformation(s)* of the problem, whenever judged useful, such as:
• changes of variables (e.g. as $x = \ln S$, in the Black–Scholes case),
• changes of unknowns (e.g. solving the equation for $e^{-rt}u$ rather than u),
• changes of probabilities (see [89, Part II A] or [65, III,3.1.3] for a PDE interpretation of the Girsanov transform for diffusion processes);
(2) *Localisation* of the problem, that is, truncating the *integration domain* in the non-local term $\mathcal{I}u$, and the *problem domain*:
• replacing the integration domain $\mathbb{R}^q$ by *bounded domains* $\mathcal{D}_\varepsilon(t,x) \subset \mathbb{R}^q$ such that

$$\int_{\mathbb{R}^q \setminus \mathcal{D}_\varepsilon(t,x)} h^i(t,x,dy) \leq \varepsilon \ ,$$

in the non-local term $\mathcal{I}u$;
• replacing $\mathcal{D}$ by an open *bounded* sub-domain in (149) (in case $\mathcal{D}$ was not bounded from the beginning), still denoted by $\mathcal{D}$ henceforth, and introducing a suitable Dirichlet boundary condition φ on a boundary layer $\partial\mathcal{D}$ ("thick" boundary to account for jumps, see for instance section 16.4.1) around the new domain $\mathcal{D}$ (we also define $\varphi = \phi$ at T);
(3) Discretizing the localized time–space domain and choosing a suitable finite differences numerical scheme for the localized problem;
(4) We thus get a problem of linear algebra (typically: a sparse linear system, or a "system of systems" in the most general case where $q \geq 1$ and $k \geq 2$) in the *values* of the approximate solution at mesh nodes, to be programmed and solved numerically on a computer.

Note that to exploit the parabolic structure of pricing equations in finance, the time dimension is typically treated separately in (3) (as in the so-called 'theta-schemes'), in order to 'save' one dimension in terms of storage cost. The problem is then solved 'linearly in time' (one time step after the other) at step (4).

Convergence, *convergence rate* (and, of course, *computational cost*!) of the resulting approximation scheme, are then the main issues.

15.1.2 Convergence and Convergence rates

Given a suitable mesh with time step h and space step vector $k = (k_1, \ldots, k_m)$ on $[0,T] \times \mathbb{R}^q$, let

$$\boxed{\begin{cases} F_h^k(u_h^k) = 0 \text{ on } [0,T) \times \mathcal{D} \\ u_h^k = \varphi \text{ on } E \setminus \Big([0,T) \times \mathcal{D} \Big) \end{cases}} \tag{151}$$

stand for a fully-discrete finite differences approximation scheme for (149) (or a localized version of (149), in case where the original domain $\mathcal{D}$ or the support of h are unbounded, cf. section 15.1.1). By (151) we mean that the related equalities are satisfied at every mesh point in $[0,T) \times \mathcal{D}$ or $E \setminus \Big([0,T) \times \mathcal{D} \Big)$. We assume that the discrete problem (151) admits a unique solution $u_h^k = (u_h^{j,k})_{1 \leq j \leq I}$ defined at grid points.

Recall that under mild technical assumptions, the pricing equation (149) has a unique solution $u = (u^j)_{1 \le j \le I}$ in a suitable space of viscosity solutions with growth conditions. Moreover we assume that the boundary data ϕ and the solution u are *bounded*, for simplicity.

In the scalar and purely differential case (no jumps, γ in $\mathcal{X}$ and $k = 1$) and in the case of a *linear* operator F, u is typically a *classic* solution of (149). Then the well-known *Lax Equivalence Theorem* states that *any stable and consistent numerical scheme* $F_h^k(u_h^k) = 0$ *is convergent*, which loosely means that u_h^k converges to u at mesh points as $h, k \to 0+$, provided:
• u_h^k is bounded, uniformly over h, k (stability),
• $F_h^k(u) \to 0$ at mesh points as $h, k \to 0$ (consistency).
We refer the reader to [143] for the detail of this theorem. In particular, the previous statements are relative to the specification of a given *norm* in which u_h^k is bounded and $F_h^k(u) \to 0$, $u_h^k \to u$, hence the notions of *stability/convergence in norm* l_∞, l_2, etc.
There is a further notion of *order of consistency*, which measures the speed at which $F_h^k(u) \to 0$ at mesh points as $h, k \to 0$. Note however that the order of consistency of a numerical scheme has no immediate implication in terms of *convergence rates* of u_h^k to u. Convergence rates only follow under additional assumptions, regarding in particular the regularity of the boundary data ϕ, etc.

Now, a nice feature of the theory of viscosity solutions is that the Lax Equivalence Theorem can be generalized to the case of a general *non linear monotone* operator F and to systems and jumps in (149).
So, for a general monotone, yet still *scalar and purely differential*, operator F in (149), Barles and Souganidis [21] proved the convergence of any *monotone, stable and consistent* approximation scheme for (149), provided (149) satisfies a suitable viscosity solutions comparison principle. We already observed that such comparison principles can indeed be obtained under mild technical conditions on the data of problem (149). The *monotonicity* condition on the scheme is a discrete version of that on F (cf. (150)), which is typically satisfied by natural finite differences approximation schemes for (149). So the convergence conditions in the Barles–Souganidis theorem are essentially reduced to those in the Lax Equivalence Theorem, namely *stability* and *consistency* of the scheme.
Finally, the latter results were extended to (scalar) non-linear equations with jumps (including the case of *unbounded* jump measures, hence beyond the case of *probability measures* h postulated in (149)) by Briani, La Chioma and Natalini [46], and to *systems* of non-linear equations with jumps in Crépey [65]. Note however that in presence of jumps, the monotonicity of the related numerical schemes does not follow as universally as in the purely differential case (see, e.g., [178]).

15.2 Finite Element Methods and Beyond

Finite element and finite volume methods are based on energetic, variational (re-)formulations of the pricing PIDE problem (149).These methods give approximate solutions defined on the whole state space of the continuous (localized) problem (see Figure 4), as opposed to approximate solutions *at grid points* by finite differences methods. They are most naturally connected with equations expressing *(energy) conservation principles*.

As already mentioned in section 15.1, they are heavier and harder to implement (finite volume methods in particular) than finite differences methods. By *heavier*, we mean that they are computationally intensive, particularly in terms of *storage cost*. Indeed, a prerequisite in

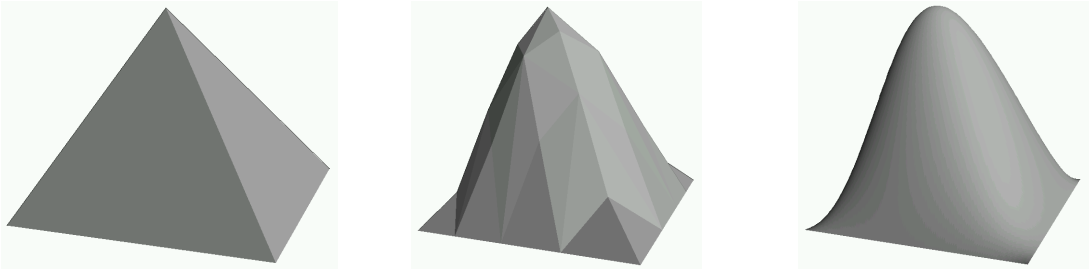

Figure 4: *Convergence of finite elements approximations.*

a finite element method is the construction of a suitable *discretization mesh*, which is typically unstructured and adaptative, and which has to be handled by the computer program all along the numerical resolution. This also means that finite element methods are *harder to implement*. One typically then has to use *finite element toolboxes* (many of them are available for free, see, e.g., http://www.inria.fr/valorisation/logiciels/calcul.en.html), which may induce less flexibility in programming.

However, the related cost is justified in cases where the geometry of the domain makes it necessary to use a sophisticated unstructured and adaptative discretization mesh, such as it is typically the case in *fluid dynamics* (see Figure 5), but also occasionally so in finance. This is for instance the case for pricing barrier options with curved boundaries (see, e.g.,

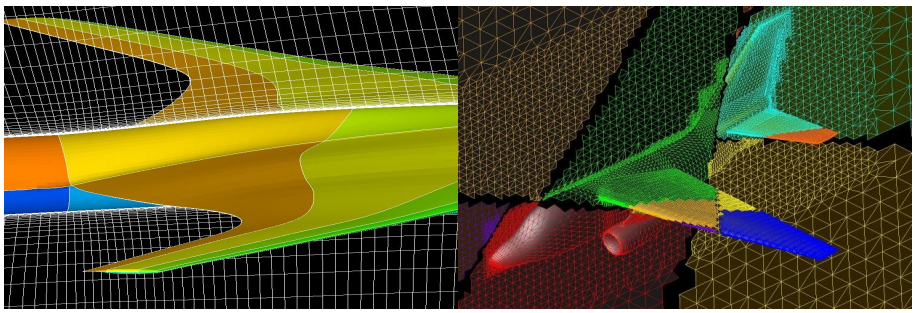

Figure 5: *(a) Detail of viscous mesh for wind tunnel model in aircraft design, (b) Parallel computation on an unstructured mesh showing the domain decomposition of 16 processors of a distributed memory computer.*

Crépey [64]; barrier options with curved boundaries are common on interest-rates), or for the precise approximation of exercise boundaries in the case of American problems (see, e.g., [51], also reported in Crépey [64]).
Another motivation for using finite elements is to counter the curse of dimensionality (see Section 3), by refining the approximation mesh in a more clever way than simply taking the product of uni-dimensional adaptive meshes, as is typically done with finite differences (see Figure 6 and section 15.2.2).

Practically speaking, the numerical resolution of problem (149) by finite elements is a five steps process (cf. the analogous description of finite differences methods in section 15.1.1):
(1) *Transformation(s)* of the problem, whenever judged useful;
(2) *Localisation* of the problem, that is, truncating the integration domain and the problem

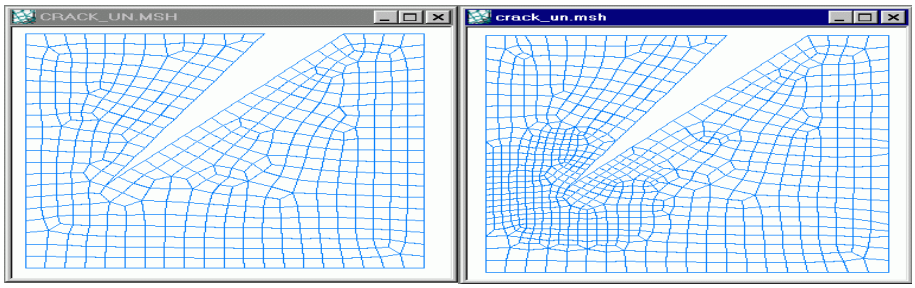

Figure 6: *Finite elements mesh refinement.*

domain, and introducing a suitable Dirichlet boundary condition φ which prolongates ϕ on a "thick" (because of jumps) boundary layer $\partial\mathcal{D}$ around the localized domain;
(3) Derivation of a *weak formulation of the localized problem* in a suitable *weighted Sobolev functional space* $\mathcal{H}$ (the boundary condition φ is typically accounted for by a judicious choice of $\mathcal{H}$);
(4) Projection of the resulting problem onto a finite-dimensional sub-space of *finite elements* $\mathcal{H}^h \subset \mathcal{H}$;
(5) We thus get a high-dimensional, sparse linear system, in the *coefficients* of the approximate solution on a finite element basis, to be programmed and solved numerically on a computer.

Existence and uniqueness of a solution to the weak form of the localized problem at step (3) typically follows by application of the *Lax–Milgram Theorem* (see section 14.2.2).

With the same motivation as for finite differences methods, the time dimension is typically (yet not always, see, e.g., [51, 84]) treated separately at steps (3)–(4). So the problem may be solved 'linearly in time' at constant storage cost. Like with finite differences methods, *convergence*, *convergence rate*, and *computational cost* of the resulting approximation scheme, are the main issues.
An interesting feature of finite elements methods is that a powerful *error theory* (*a priori* and *a posteriori error estimates*) is available.

To solve the high-dimensional sparse linear system arising at step (5), an iterative solver is required. The *Generalized Minimal Residual (GMRES) algorithm* [165], a special form of conjugate gradient method which is both efficient and relatively easy to code, is the industry standard in this regard.

15.2.1 Finite Volumes

Finite volumes methods can be seen in an informal way as counterparts of finite element methods in which the test functions used in the variational formulation of the problem are *indicator (discontinuous) functions*, instead of regular test functions as usual. Finite volumes methods are especially dedicated to PDE convection (or convection dominant) problems with discontinuous data. For instance, the finite volume discretization of the pricing function of

a digital option is exact in a Black–Scholes model with no volatility ($\sigma = 0$).

15.2.2 Sparse Grids

Sparse grids denote numerical techniques to represent, integrate or interpolate high dimensional functions, relying on the seminal works of the Russian mathematician Smolyak, who found a clever quadrature rule to (partially) escape the curse of dimensionality.
This direction of research underlies active developements in finite elements methods (or finite differences methods; see Figure 7). But the related algorithms are difficult to implement, and we shall not go further in this direction at the level of these notes, referring the interested reader to [160, 91].

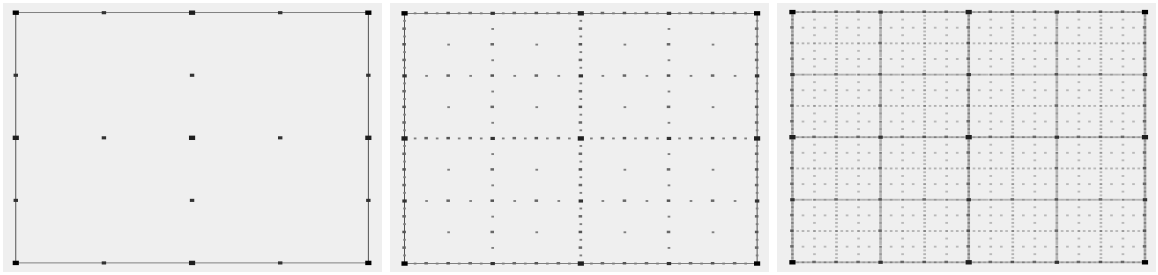

Figure 7: *Sparse Grids Refinement.*

For simplicity, we shall focus on *finite differences methods* in the sequel. Regarding finite element methods in finance, we refer the reader to Achdou–Pironneau [1] or Matache et al. [137].

16 Finite Differences for European Vanilla Options

16.1 Black–Scholes Equation

Recall from section 10.1 that in the risk-neutral Black–Scholes model

$$\boxed{\frac{dS_t}{S_t} = \kappa dt + \sigma dW_t}$$

for a standard $\mathbb{P}$ – Brownian motion W (and with $\kappa = r - q$), the price process of an European vanilla option with (integrable) payoff $\phi(S_T)$ at T is given by $\Pi_t = v(t, S_t)$, where the pricing function v solves the following Black–Scholes valuation PDE:

$$\boxed{\begin{cases} v(T,S) = \phi(S),\ S \in (0,\infty) \\ \partial_t v + \kappa S \partial_S v + \frac{1}{2}\sigma^2 S^2 \partial^2_{S^2} v - rv = 0 \text{ in } [0,T) \times (0,\infty) \end{cases}} \tag{152}$$

After logarithmic transformation $X_t = \ln(S_t)$, the option price process is given as $\Pi_t = u(t, X_t)$, where u solves the parabolic equation

$$\boxed{\begin{cases} u(T,x) = \psi(x),\ x \in \mathbb{R} \\ \partial_t u + b \partial_x u + \frac{1}{2}\sigma^2 \partial^2_{x^2} u - ru = 0 \text{ in } [0,T) \times \mathbb{R} \end{cases}} \tag{153}$$

where $b = \kappa - \frac{\sigma^2}{2}$ and $\psi(x) = \phi(e^x)$, the payoff in the x variable.

For nice enough terminal conditions, these equations are well-posed (i.e. have one and only one solution, which depends continuously on the data of the problem), in a suitable space of classic solutions [90] (and in suitable spaces of viscosity or weak solutions in weighted Sobolev spaces as well, cf. section 14.2).

16.2 Localization and Discretization in space

Let $x = \ln(S_0)$. To solve numerically the Black–Scholes equation in log-returns variable (153), we start by truncating the integration domain in space: the problem will be solved on the bounded domain $\bar{\mathcal{D}} = [x - \ell, x + \ell]$. One chooses ℓ so that

$$\mathbb{P}\left(\exists s \in [0,T],\ |X_s^x - x| \geq \ell\right) \leq \varepsilon\ , \tag{154}$$

which can be achieved by setting

$$\boxed{\ell = |b|T + f\sigma\sqrt{T}} \tag{155}$$

for a related quantile f of the Gaussian law (like $f = 4$ or 5, see, e.g., Crépey [64]). Once $\mathcal{D}$ is chosen, one discretizes it in space, constructing a uniform grid $\{x_j\}$ with

$$x_j = x - \ell + \frac{2j\ell}{m+1}, \quad \text{for } 0 \leq j \leq m+1$$

(with m odd, so that x lies in the space grid; otherwise some kind of interpolation has to be used, cf. Remark 16.2).

One then approximates the differential spatial operator

$$\boxed{\mathcal{A}u = \tfrac{1}{2}\sigma^2\partial^2_{x^2}u + b\partial_x u - ru}$$

by a discrete operator $\mathcal{A}^k$ acting on $\mathbb{R}^m$-valued vectors $u^k = (u^k(t,x_1), \ldots, u^k(t,x_m))$. The easiest and most natural is to take:

$$\boxed{\mathcal{A}^k u^k(x_j) = \frac{1}{2}\sigma^2\delta^2_{x^2}u^k(x_j) + b\delta_x u^k(x_j) - ru^k(x_j)} \tag{156}$$

with

$$\delta_x u^k(x_j) = \tfrac{1}{2k}(u^k(x_{j+1}) - u^k(x_{j-1}))$$
$$\delta^2_{x^2}u^k(x_j) = \tfrac{1}{k^2}(u^k(x_{j+1}) - 2u^k(x_j) + u^k(x_{j-1}))$$

where $u^k(x_0) = u^k(x - \ell)$ and $u^k(x_{m+1}) = u^k(x + \ell)$ are notations for quantities to be defined below using the $u^k(x_j)$, $j = 1 \ldots m$. By suitable Taylor expansions, it is possible to show that δ_x and $\delta^2_{x^2}$ are *consistent approximations of order two* of the spatial differential operators ∂_x and $\partial^2_{x^2}$, respectively, meaning that for regular test functions φ :

$$|\delta_x\varphi(t,x_j) - \partial_x\varphi(t,x_j)| = \mathrm{O}(k^2) \quad , \quad |\delta^2_{x^2}\varphi(t,x_j) - \partial^2_{x^2}\varphi(t,x_j)| = \mathrm{O}(k^2)\ .$$

Remark 16.1 If $|\kappa|/\sigma^2$ is not small, a less precise but more stable finite difference approximation for ∂_x is

$$\delta_x u^k(x_j) = \begin{cases} \frac{1}{k}(u^k(x_j) - u^k(x_{j-1})) & \text{if} \quad b < 0 \\ \frac{1}{k}(u^k(x_{j+1}) - u^k(x_j)) & \text{if} \quad b > 0\ . \end{cases}$$

This alternative discretization for ∂_x, called *upwind discretization*, is based on the consideration of the characteristics of the limiting hyperbolic transport equation with $\sigma = 0$.

One then seeks an $\mathbb{R}^m$-valued time functional $u^k(t)$ such that:
• In the case of *Dirichlet boundary conditions*,

$$\left\{\begin{array}{l} u^k(T,x_j)=\psi(x_j) \\ u^k(t,x\pm\ell)=\psi(x\pm\ell) \\ \frac{d}{dt}u^k(t,x_j)+\mathcal{A}^k u^k(t,x_j)=0 \text{ for } 0\le t<T,\ 1\le j\le m\ ; \end{array}\right. \tag{157}$$

• In the case of *Neumann boundary conditions*,

$$\left\{\begin{array}{l} u^k(T,x_j)=\psi(x_j) \\ u^k(t,x_1)=u^k(t,x-\ell)+k\partial_x\psi(x-\ell) \\ u^k(t,x_m)=u^k(t,x+\ell)-k\partial_x\psi(x+\ell) \\ \frac{d}{dt}u^k(t,x_j)+\mathcal{A}^k u^k(t,x_j)=0 \text{ for } 0\le t<T,\ 1\le j\le m\ . \end{array}\right. \tag{158}$$

Set

$$\boxed{\alpha=\frac{\sigma^2}{2k^2}-\frac{b}{2k}\ ,\ \beta=-\frac{\sigma^2}{k^2}-r\ ,\ \gamma=\frac{\sigma^2}{2k^2}+\frac{b}{2k}} \tag{159}$$

In this notation, the operator $\mathcal{A}^k$ applied to $u^k(t)$ writes:

$$\mathcal{A}^k u^k(t)=A^k u^k(t)+v^k,$$

with in the case of Dirichlet boundary conditions:

$$A^k=\begin{bmatrix} \beta & \gamma & 0 & ,\dots, & 0 & 0 \\ \alpha & \beta & \gamma & 0 & ,\dots, & 0 \\ 0 & \alpha & \beta & \gamma & ,\dots, & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & ,\dots, & \alpha & \beta & \gamma \\ 0 & 0 & 0 & ,\dots, & \alpha & \beta \end{bmatrix},\ v^k=\begin{bmatrix} \psi(x-\ell)\alpha \\ 0 \\ \vdots \\ 0 \\ \psi(x+\ell)\gamma \end{bmatrix} \tag{160}$$

and in the case of Neumann boundary conditions:

$$A^k=\begin{bmatrix} \beta+\alpha & \gamma & 0 & ,\dots, & 0 & 0 \\ \alpha & \beta & \gamma & 0 & ,\dots, & 0 \\ 0 & \alpha & \beta & \gamma & ,\dots, & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & ,\dots, & \alpha & \beta & \gamma \\ 0 & 0 & 0 & ,\dots, & \alpha & \beta+\gamma \end{bmatrix},\ v^k=\begin{bmatrix} -\alpha k\frac{\partial\psi}{\partial x}(x-\ell) \\ 0 \\ \vdots \\ 0 \\ \gamma k\frac{\partial\psi}{\partial x}(x+\ell) \end{bmatrix}. \tag{161}$$

Remark 16.2 **(i)** If m was even, x would not belong to $\{x_j\,;\,0\le j\le m+1\}$. In this case one could use linear interpolation to compute the option value corresponding to the initial stock price at time 0. This price would thus be approximated by $\frac{1}{2}\left(u^k(0,x_{m/2})+u^k(0,x_{m/2+1})\right)$.

(ii) In the case of Dirichlet boundary conditions of variant of (160) (*propagation of the Dirichlet condition at terminal time*), simpler to implement, consists in

$$A^k=\begin{bmatrix} 1 & 0 & 0 & ,\dots, & 0 & 0 \\ \alpha & \beta & \gamma & 0 & ,\dots, & 0 \\ 0 & \alpha & \beta & \gamma & ,\dots, & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & ,\dots, & \alpha & \beta & \gamma \\ 0 & 0 & 0 & ,\dots, & 0 & 1 \end{bmatrix} \tag{162}$$

and $v^k = 0$.

16.3 Theta-schemes

We now discuss discretization in time. The standard *theta-scheme* ($\theta \in [0,1]$) for the parabolic equation (153) may be summarized as follows (see, e.g., [124, 143]): Fix a time discretization step h such that $T = nh$, and construct a fully discrete approximation $u_h^k(t_i, x_j) = u_i(x_j)$ where the u_i, $i = 0 \dots n$ are $\mathbb{R}^m$-valued vectors such that:

$$\begin{cases} u_n = \psi^k \\ \frac{u_{i+1}-u_i}{h} + \mathcal{A}^k(u_{i+1} + \theta(u_i - u_{i+1})) = 0 \text{ for } 0 \le i \le n-1 \end{cases}$$

or

$$\boxed{\begin{cases} u_n = \psi^k \\ \left[\mathrm{Id} - h\theta\mathcal{A}^k\right] u_i = \left[\mathrm{Id} + h(1-\theta)\mathcal{A}^k\right] u_{i+1} \text{ for } 0 \le i \le n-1 \end{cases}} \tag{163}$$

or, equivalently

$$\boxed{\begin{cases} u_n = \psi^k \\ \left[\mathrm{Id} - h\theta A^k\right] u_i = \left[\mathrm{Id} + h(1-\theta)A^k\right] u_{i+1} + hv^k \text{ for } 0 \le i \le n-1 \end{cases}} \tag{164}$$

For $\theta = 0$, we get the so-called *Euler explicit scheme*. For $\theta = 1$ the scheme is the *fully implicit Euler scheme*, and for $\theta = \frac{1}{2}$ it is the *Crank-Nicholson scheme*.

Once we have computed u_h^k, we recover the delta $\Delta = e^{-x}\partial_x u$ by its approximation given by

$$\boxed{\Delta_h^k(t_i, x_j) = e^{-x_j} \tfrac{u_i(x_j+k) - u_i(x_j-k)}{2k}}$$

16.3.1 Explicit Method

First, let us discuss the case $\theta = 0$ (*explicit scheme*). Using the definition of A^k (considering say Dirichlet conditions), the approximating scheme (163) is reduced to (cf. (159))

$$\begin{cases} u_n = \psi^k \text{ , and for } 0 \le i \le n-1 \text{ , } 1 \le j \le m : \\ u_i(x_j) = h\alpha\, u_{i+1}(x_{j-1}) + (1 + h\beta)\, u_{i+1}(x_j) + h\gamma\, u_{i+1}(x_{j+1}) \end{cases}$$

with $u_{i+1}(x \pm \ell) = \psi(x \pm \ell)$.

Considering (159), one can then show that the explicit approximation scheme for the Black–Scholes pricing equation (153) is (cf. section 15.1.2):

- *stable, provided* $h \le \frac{k^2}{\sigma^2 + rk^2}$ (and $\sigma^2 > |b|k$, but this is always satisfied for k small enough);
- *consistent* of order *one in time and two in space*.

16.3.2 Implicit Methods

When we choose $1 \ge \theta > 0$, we have to solve at each time step, a linear system

$$\boxed{Au_i = Bu_{i+1} + hv^k} \tag{165}$$

where $A = \mathrm{Id} - h\theta A^k$ and $B = \mathrm{Id} + h(1-\theta)A^k$ cf. (164)) are tridiagonal matrices of the type

$$\begin{pmatrix} b_1 & c_1 & 0 & ,\dots, & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & ,\dots, & 0 \\ 0 & a_3 & b_3 & c_3 & ,\dots, & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & ,\dots, & a_{m-1} & b_{m-1} & c_{m-1} \\ 0 & 0 & 0 & ,\dots, & a_m & b_m \end{pmatrix}.$$

For example, in the case of natural Dirichlet boundary condition, A is given by, for every j:

$$\boxed{a_j = \theta h(\tfrac{b}{2k} - \tfrac{\sigma^2}{2k^2}), \quad b_j = 1 + \theta h(r + \tfrac{\sigma^2}{k^2}), \quad c_j = -\theta h(\tfrac{b}{2k} + \tfrac{\sigma^2}{2k^2})}$$

and B is given by, for every j:

$$\boxed{a_j = (1-\theta) h(\tfrac{\sigma^2}{2k^2} - \tfrac{b}{2k}), \quad b_j = 1 - (1-\theta)h(r + \tfrac{\sigma^2}{k^2}), \quad c_j = (1-\theta)h(\tfrac{b}{2k} + \tfrac{\sigma^2}{2k^2})}$$

The *fully implicit* and the *Crank–Nicholson* schemes correspond to $\theta = 1$ and $\theta = \frac{1}{2}$, respectively. One can show that the implicit ($\theta \neq 0$) approximation theta-schemes for the Black–Scholes pricing equation (153) are:

- *stable inconditionally* for $\theta \geq \frac{1}{2}$;
- *consistent* of order *one in time and two in space*, with the notable exception of the *Crank–Nicholson scheme* which is consistent of order *two in time and space*.

On Figure 8, we plotted the relative error at time 0 as a function of the spot price S_0, obtained when pricing an European vanilla call option with the explicit, fully implicit and Crank–Nicholson theta-schemes, respectively (for $r = 10\%, \sigma = 20\%, T = 1y, K = 100$). The Crank–Nicholson is more accurate (at least around the money), as expected.

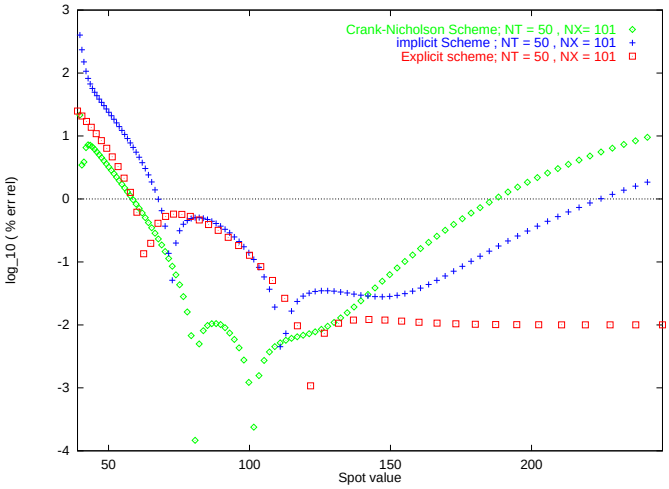

Figure 8: *Pricing of an European call option by theta-schemes for* $\theta = 0, \frac{1}{2}, 1$.

All implicit theta-schemes require at each time step the resolution of a linear system $Au = v$, where u and v are m-dimensional vectors. Let us describe two algorithms of resolution of such linear systems.

Gauss Factorization This algorithm is based on the fact that a regular matrix can be factorized as $A = LU$, where L is a lower triangular matrix, and U is an upper triangular matrix with all ones on its diagonal. The linear system $LUz = v$ is decomposed into $Ly = v, Uz = y$. It is easy to see that A tridiagonal implies that L, U are also tridiagonal and so only the upper diagonal of U and the two diagonals of L need to be found. This results in the following procedure, known as *Thomas' algorithm* [143, 124]:

$$\begin{cases} b'_m = b_m, \quad y_m = v_m \\ \text{For } 1 \le j \le m-1, \ j \text{ decreasing:} \\ b'_j = b_j - c_j a_{j+1}/b'_{j+1}, \\ y_j = v_j - c_j y_{j+1}/b'_{j+1}. \end{cases} \text{, followed by } \begin{cases} z_1 = y_1/b'_1 \\ \text{For } 2 \le j \le m, \ j \text{ increasing:} \\ z_j = (y_j - a_j z_{j-1})/b'_j. \end{cases}$$

Remark 16.3 Note that this method presupposes that all the b'_j (called the pivots) are non zero.

SOR Iterative Methods An alternative, which in the case of tridiagonal systems is justified only by its programming simplicity, is to use the Successive Over-Relaxation *iterative scheme*. The idea is to decompose A as $A = D + R$ where D is the diagonal part of A. The linear system $Au = v$ is then rewritten $Du = v - Ru$. The solution is computed as the limit of a converging sequence

$$u^{p+1} = D^{-1}(v - Ru^p) \ , \tag{166}$$

or better, given a *relaxation paramater* $1 < \omega < 2$:

$$\boxed{u^{p+1} = u^p + \omega(\widetilde{u}^{p+1} - u^p)}$$

where $\widetilde{u}^{p+1}$ stands for the r.h.s. of (166). More precisely, the algorithm writes as follows:

- ***Step 0*** Choose $u^0 \ge 0$, $\varepsilon > 0$, $1 < \omega < 2$. Set $p = 0$.
- ***Step 1 (Jacobi iteration)*** Form an intermediate vector $\widetilde{u}^{p+1} = (h_j^{p+1})_{1 \le j \le m}$ *by*

$$h_j^{p+1} = \frac{1}{A_{jj}}(v_j - \sum_{l<j} A_{jl} u_l^p - \sum_{l>j} A_{jl} u_l^p) \ , \ 1 \le j \le m \tag{167}$$

Here a possible refinement (*Gauss–Seidel iteration*) is to use h_l^{p+1} instead of u_l^p in the first sum in the r.h.s. of (167).

- ***Step 2 (Over-relaxation)*** Define u^{p+1} by

$$u^{p+1} = u^p + \omega(\widetilde{u}^{p+1} - u^p) \ , \tag{168}$$

and set $p = p + 1$.

- ***Step 3*** Repeat steps 2 and 3 until $|u^{p+1} - u^p| < \varepsilon$.

16.4 Adding Jumps

Let us now add jumps in S, assuming that the underlying asset price evolves according to the following risk-neutral jump-diffusion :

$$\frac{dS_u}{S_{u-}} = (\kappa - \gamma\bar{J})du + \sigma dW_u + d(\sum_{l=1}^{N_u} J_l),\ S_{T-t} = s \tag{169}$$

where:
• N is a Poisson process with deterministic jump intensity γ;
• the J_l are iid r.v. > -1 with law ν of $j_1 := \ln(1 + J_1)$, and we set $\bar{J} = \mathbb{E}J_1$;
• the other data in (169) are defined as usual.
Moreover, N, W, and the J_l are independent. So in case $\nu = \mathcal{N}(\alpha, \beta)$ we recover the risk-neutral Merton model of section 10.3.

One can show as in section 16.1 (see also Part II) that the price of an European option in the risk-neutral jump-diffusion (169) can be formulated in terms of the solution to a related Partial Integro-Differential Equation (PIDE, cf. (57)). So, u denoting the pricing function in the returns variable $x = \ln S$, we have formally as in section 16.1 (resorting to the Itô–Poisson formula (45)), with "$\triangleq$" standing for "equality up to a local martingale term":

$$e^{rt}d(e^{-rt}u(t, X_t)) \triangleq (\partial_t u + \mathcal{A}_X u - ru)(t, X_t)dt\ ,$$

where given (103), by (38):

$$\mathcal{A}_X u = a\partial_x u + \frac{1}{2}\sigma^2\partial^2_{x^2}u + \gamma\left[\mathbb{E}u(x + j_1) - u(x)\right]\ .$$

By the usual arbitrage argument, the price at time t of the option is thus given by $\Pi_t = u(t, X_t)$ where u solves the integro-parabolic equation

$$\boxed{\begin{cases} u(T, x) = \psi(x),\ x \in \mathbb{R} \\ \partial_t u + \mathcal{A}u + \mathcal{B}u = 0 \text{ in } [0, T) \times \mathbb{R} \end{cases}} \tag{170}$$

with

$$\boxed{\mathcal{A}u = \frac{1}{2}\sigma^2\partial^2_{x^2}u + a\partial_x u - ru\ ,\ \mathcal{B}u = \gamma\int_{\mathbb{R}}(u(t, x + z) - u(t, x))\nu(dz)}$$

(note that the operator $\mathcal{A}$ of this Section reduces to the operator $\mathcal{A}$ of the previous Section when $\gamma = 0$, since in this case $a = b - \gamma\bar{J} = b$).

16.4.1 Localization

Localization works essentially as in Subsections 16.2, except for the fact that:
• b is replaced by a in (155),
• the boundary $\partial\mathcal{D} = \{x - \ell\} \cup \{x + \ell\}$ is replaced by the "thick" boundary layer $\partial\mathcal{D} = [x - \ell - z^-_{\min}, x - \ell] \cup [x + \ell, x + \ell + z^+_{\max}]$ (cf. step (2) in section 15.1.1), where $z_{\min}$ and $z_{\max}$ are such that

$$\int_{z_{\min}}^{z_{\max}} \nu(dz) \approx \int_{-\infty}^{\infty} \nu(dz) - \varepsilon = 1 - \varepsilon, \quad \varepsilon \ll 1\ ;$$

• we solve the following localized problem on $\bar{\mathcal{D}} := \mathcal{D} \cup \partial\mathcal{D}$:

$$\boxed{\begin{cases} u(T,x) = \varphi(T,x) = \psi(x)\ ,\ x \in \mathbb{R} \\ u(t,x) = \varphi(t,x)\ ,\ (t,x) \in [0,T) \times \partial\mathcal{D} \\ \partial_t u + \mathcal{A}u + \mathcal{B}_\ell u = 0\ ,\ (t,x) \in [0,T) \times \mathcal{D} \end{cases}} \tag{171}$$

where φ is a suitable Dirichlet condition such that $\varphi(T,x) = \psi(x)$ (e.g., $\varphi(t,x) \equiv \psi(x)$), and $\mathcal{B}_\ell$ is such that, for $x \in \mathcal{D}$:

$$\boxed{\begin{array}{c} \mathcal{B}_\ell u(x) = \gamma \int_{\mathcal{D}-x} u(t,x+z)\nu(dz) - \gamma u(t,x) + \\ \\ \gamma \int_{\partial\mathcal{D}-x} \varphi(t,x+z)\nu(dz) \end{array}} \tag{172}$$

16.4.2 Discretization

We define:

$$\begin{gathered} x_{\min} = \min \bar{\mathcal{D}},\ x_{\max} = \max \bar{\mathcal{D}} \\ k = \tfrac{x_{\max} - x_{\min}}{m},\ x_j = x_{\min} + jk\ (j = 0, \dots, m) \\ j_l = \left[\tfrac{z^-_{\min}}{k}\right]\ ,\ j_u = n - \left[\tfrac{z^+_{\max}}{k}\right]. \end{gathered}$$

$\mathcal{B}_\ell$ approximation In order to approximate $\mathcal{B}_\ell$, we set $\nu(dz) = \zeta(z)dz$, where ζ is the density of ν, assumed to exist. Then the $\mathcal{B}_\ell$ operator is decomposed as

$$\boxed{\begin{array}{c} \mathcal{B}_\ell u(t,x) = \left[\gamma \int_{\mathcal{D}-x} u(t,x+z)\zeta(z)dz - \gamma u(t,x)\right] + \\ \left[\gamma \int_{\partial\mathcal{D}-x} \varphi(t,x+z)\zeta(z)dz\right] = \bar{\mathcal{B}}_\ell u(t,x) + \Phi(t,x) \end{array}}$$

Standard approximation For $j = j_l, \dots, j_u$, we approximate

$$\boxed{\begin{array}{c} \bar{\mathcal{B}}_\ell u(\cdot, x_j) \approx \mathcal{B}^k u(\cdot, x_j) = \gamma k \sum_{i=j_l-j}^{j_u-j} u(\cdot, x_{j+i})\zeta(\cdot, x_i) - \gamma u(\cdot, x_j) \\ \Phi(\cdot, x_j) \approx \Phi^k(\cdot, x_j) = \gamma k \sum_{\{i+j<j_l\} \cup \{i+j>j_u\}} \varphi(\cdot, x_{j+i})\zeta(\cdot, x_i) \end{array}}$$

FFT approximation Recall that the *discrete correlation* f of two real-valued functions g_j, h_j, each periodic with period m, is defined by

$$f_j = \sum_{l=0}^{m-1} g_{j+l} h_l.$$

The *discrete correlation theorem* says that the *discrete Fourier transform* F of f is such that

$$Ff = (Fg)(\overline{Fg}) \tag{173}$$

where $\overline{Fh}$ denotes the complex conjugate of Fh. We can thus compute f by FFT as follows: FFT the two data sets g and h, multiply one resulting transform by the complex conjugate of the other and inverse transform the product. The result will formally be a complex vector of length m. However, it will turn out to have all its imaginary parts equal to zero since the original data sets were both real.
We can apply this procedure to the $\mathcal{B}^k$ and Φ^k operator (with the related vectors suitably prolongated by *zero padding* outside $j_l, \dots, j_u$, see,e.g., [97, 156]).

Finite differences in space For the differential operator $\mathcal{A}$, we write:

$$\boxed{\mathcal{A}^k u(\cdot,x_j) = [a\delta_x + \tfrac{\sigma^2}{2}\delta^2_{xx} - r]u(\cdot,x_j)}$$

where

$$\delta^2_{x^2}u(\cdot,x_j) = \tfrac{u(\cdot,x_{j+1})-2u(\cdot,x_j)+u(\cdot,x_{j-1})}{k^2},$$

$$\delta_x u(\cdot,x_j) = \tfrac{u(\cdot,x_j)-u(\cdot,x_{j-1})}{k} + \alpha\tfrac{u(\cdot,x_{j+1})-2u(\cdot,x_j)+u(\cdot,x_{j-1})}{k},$$

where α is chosen such that (for stability reasons):

$$\begin{cases} ka \le \frac{\sigma^2}{2} & \alpha = \frac{1}{2} \\ ka > \frac{\sigma^2}{2} & \begin{cases} \alpha = 0 & a > 0 \\ \alpha = 1 & a < 0 \end{cases} \end{cases}$$

Theta-schemes We define the time grid $\{t_i \mid t_i = hi, \quad i = 0,\dots,n\}$. Denoting $u_i^j = u(t_i,x_j)$, we consider the following discrete operator:

$$\boxed{\begin{array}{l} \frac{u^j_{i+1}-u^j_i}{h} + \mathcal{A}^k[\theta_A u^j_i + (1-\theta_A)u^j_{i+1}] + \\ \\ +\theta_B[\mathcal{B}^k u^j_i + (\Phi^k)^j_i] + (1-\theta_B)[\mathcal{B}^k u^j_{i+1} + (\Phi^k)^j_i] \end{array}} \tag{174}$$

where $\theta_A,\ \theta_B \in [0,1]$. We consider in particular two cases.
Explicit scheme $\theta_A = \theta_B = 0$: Computationally feasible but potentially unstable and suffers from the drawback that convergence in time is only of the first order (in $\mathrm{O}(h)$).
For every $i = n-1,\dots,0$ we solve

$$\begin{cases} u^j_i = \varphi^j_i, \quad j = 0,\dots,j_l-1 \quad \text{and} \quad j = j_u+1,\dots,n \\ \\ u^j_i = p_1 u^{j-1}_{i+1} + p_2 u^j_{i+1} + p_3 u^{j+1}_{i+1} + h\gamma(k\sum_{i=i_l}^{i_u}\zeta_i u^{j+i}_{i+1} - u^j_{i+1} \\ \quad +k \sum_{\{i+j<j_l\}\cup\{i+j>j_u\}} \zeta_i \varphi^{j+i}_{i+1}), \quad j = j_l,\dots,j_u, \end{cases} \tag{175}$$

where

$$p_1 = h\Big(\frac{\sigma^2}{2k^2} + \frac{a}{k}(1-\alpha)\Big),\ \ p_2 = 1 - h\Big(\frac{\sigma^2}{k^2} + \frac{a}{k}(1-2\alpha) + r\Big),\ \ p_3 = h\Big(\frac{\sigma^2}{2k^2} - \frac{a}{k}\alpha\Big).$$

'Asymmetric' scheme: $\theta_A = \frac{1}{2}$, $\theta_B = 0$: Stable and efficient but some accuracy is lost due to the asymmetric treatment of the continuous and jump part.
We have to solve, for $i = n-1,\dots,0$, the linear system $Au_i = B_{i+1}$, where $u_i = (u_i^0,\dots,u_i^{j_l},\dots,u_i^{j_u},\dots,u_i^m)^\mathsf{T}$

$$A = \begin{pmatrix} I_l & & 0 \\ & \widetilde{A} & \\ 0 & & I_u \end{pmatrix}, \tag{176}$$

in which I_l and I_u are two identity matrices, $j_l \times j_l$ and $(m - j_u) \times (m - j_u)$ respectively, and $\widetilde{A}$ is the $(j_u - j_l + 1) \times (j_u - j_l + 1)$ tridiagonal matrix such that

$$\widetilde{A} = \begin{pmatrix} a_1 & a_2 & 0 & \cdot & \cdot & 0 \\ a_0 & a_1 & a_2 & 0 & \cdot & 0 \\ 0 & a_0 & a_1 & a_2 & 0 & \cdot \\ \cdot & 0 & & & & 0 \\ \cdot & \cdot & 0 & a_0 & a_1 & a_2 \\ 0 & \cdot & \cdot & 0 & a_0 & a_1 \end{pmatrix} \quad \begin{array}{c} a_0 = -\frac{h}{2}\Big(\frac{\sigma^2}{2k^2} + \frac{a}{k}(1-\alpha)\Big) \\ a_1 = 1 + \frac{h}{2}\Big(\frac{\sigma^2}{k^2} + \frac{a}{k}(1-2\alpha) + r\Big) \\ a_2 = -\frac{h}{2}\Big(\frac{\sigma^2}{2k^2} - \frac{a}{k}\alpha\Big) \end{array}$$

and finally

$$B_{i+1} = (\varphi_{i+1}^0, \ldots, \varphi_{i+1}^{j_l-1}, f_{i+1}^{j_l}, \ldots, f_{i+1}^{j_u}, \varphi_{i+1}^{j_u+1}, \ldots, \varphi_{i+1}^m)^{\mathsf{T}}, \tag{177}$$

where, for $j = j_l, \ldots, j_u$,

$$f_{i+1}^j = -a_0 u_{i+1}^{j-1} + (2 - a_1) u_{i+1}^j - a_2 u_{i+1}^{j+1} + h\gamma\Big(k \sum_{i=i_l}^{i_u} \zeta_i u_{i+1}^{j+i} - u_{i+1}^j + k \sum_{\{i+j<j_l\} \cup \{i+j>j_u\}} \zeta_i \varphi_{i+1}^{j+i}\Big).$$

17 Finite Differences for American Vanilla Options

17.1 Black–Scholes Variational inequalities

Using the formalism of viscosity or variational solutions of PDEs, the arguments of section 16.1 can be extended to American options. It can thus be shown (see Part II) that the price of an American vanilla option in the Black–Scholes model at time 0 is given by

$$\sup_{\tau \in \mathcal{T}} \mathbb{E} e^{-r\tau} \psi(X_\tau) = u(0, x) \ ,$$

with $X_t = \ln(S_t)$ and $X_0 = x$, where u is the unique viscosity solution with growth conditions to the following *obstacle problem* in the log-return space variable $x = \ln S$:

$$\boxed{\begin{cases} u(T, \cdot) = \psi(T, \cdot) \text{ on } \mathbb{R} \\ \max(\partial_t u + \mathcal{A}u, \psi - u) = 0 \text{ on } [0, T) \times \mathbb{R} \end{cases}} \tag{178}$$

Note that at the intuitive level, the second line of (178) expresses the facts that the American option pricing function is $\geq$ to the related payoff, and also to the related European option pricing function.

Equivalently to (178), u can be characterized as the unique solution to the following variational inequality problem:

$$\begin{cases} u(T, \cdot) = \psi(T, \cdot) \\ \langle \partial_t u + \mathcal{A}u, v - u \rangle \leq 0 \text{ for any } v \geq \psi \end{cases} \tag{179}$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in a suitable weighted Sobolev space.

To support this equivalence, note that provided $u \geq \psi$, (178) implies that, for any $v \geq \psi$:

$$\langle \partial_t u + \mathcal{A}u, v - u \rangle = \langle \partial_t u + \mathcal{A}u, v - \psi \rangle + \langle \partial_t u + \mathcal{A}u, \psi - u \rangle = \langle \partial_t u + \mathcal{A}u, v - \psi \rangle \leq 0,$$

hence (179). Conversely, still assuming $u \geq \psi$, (179) implies that for any non-negative test-function φ, we have $\langle \partial_t u + \mathcal{A}u, \varphi \rangle \leq 0$, hence $\partial_t u + \mathcal{A}u \leq 0$ and $\langle \partial_t u + \mathcal{A}u, \psi - u \rangle \geq 0$. But taking $v = \psi$ in (179) gives $\langle \partial_t u + \mathcal{A}u, \psi - u \rangle \leq 0$, so that finally $\langle \partial_t u + \mathcal{A}u, \psi - u \rangle = 0$, hence (178).

17.2 Splitting methods

Let us first describe the computational treatment of the obstacle problem (178), using the so-called *splitting method*. The idea is to construct recursively, by *dynamic programming*, the approximate solution u_i^j, starting from $u_n = \psi^k(T)$, and computing u_i from u_{i+1}, for $0 \leq i \leq n-1$, *in two steps* as follows (cf. (165)):

$$\boxed{\begin{array}{c} A\widetilde{u}_i = Bu_{i+1} + hv^k \\ u_i = \max\left(\psi^k(t_i), A\widetilde{u}_i\right) \end{array}} \tag{180}$$

By Barles et al. [19, 21], this scheme converges to the unique viscosity solution of (178) satisfying suitable growth conditions, namely the pricing function u.

17.3 Linear Complementarity Problem

Let us now describe the computational treatment of the variational inequalities (179). We refer to [105, 95] for a detailed presentation. It is well-known that after localization and discretization the variational inequalityproblem (179) can be expressed as a Linear Complementarity (LC) problem. At each time step i, we thus have to solve:

$$\left\{ \begin{array}{l} AX \geq G \\ X \geq \Phi \\ (AX - G, X - \Phi) = 0 \end{array} \right. \tag{181}$$

with (using Dirichlet boundary conditions)

$$\left\{ \begin{array}{l} A = I - h\theta A^k \\ X = u_i \\ G = (I + h(1-\theta)A^k)u_{i+1} + hv^k \\ \Phi = \psi^k \, . \end{array} \right.$$

So A is the following tridiagonal matrix:

$$A = \begin{pmatrix} b & c & ,\ldots, & ,\ldots, & 0 \\ a & b & c & ,\ldots, & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & a & b & c \\ 0 & ,\ldots, & ,\ldots, & a & b \end{pmatrix} \text{ with } \left\{ \begin{array}{l} a = \theta h \left(-\frac{\sigma^2}{2k^2} + \frac{1}{2k} b \right) \\ b = 1 + \theta h (\frac{\sigma^2}{k^2} + r) \\ c = -\theta h \left(\frac{\sigma^2}{2k^2} + \frac{1}{2k} b \right) \, . \end{array} \right.$$

There exist three classic algorithms to solve the LC problem (181).

A method by Brennan and Schwarz [45] This method consists in introducing an auxiliary LC problem. A rigorous justication of the convergence of this algorithm, in the case of an American put option, was given in Jaillet et al. [105].

PSOR Method The LC problem (181) can be written as follows: find vectors $W = (w_i)_{1\leq j\leq m}$ and $Z = (z_j)_{1\leq j\leq m}$ in $\mathbb{R}^m$ such that

$$\left\{\begin{array}{ll} W = AZ + V & (182.1)\\ W \geq 0, \quad Z \geq 0 & (182.2)\\ (W, Z) = 0 & (182.3)\end{array}\right. \tag{182}$$

where we have set $Z = X - \Phi$ and $V = A\Phi - G$. Such a LC problem can be solved by a Projected SOR scheme (cf. section 16.3.2). Convergence was established in Cryer [69].

An Algorithm of Cryer This algorithm is based on a direct method, which is an adaptation to LC problems of Saigal's linear programming algorithm. The basic idea of this kind of algorithm is: Choose an initial value which satisfies both (182.1) and (182.2), maintain the two conditions during all steps and have gradually satisfied the null condition (182.3). The solution of problem (182) is then obtained.
Note that the matrix A is a Minkowski matrix, namely a matrix with positive principal minors, positive diagonal entries and non-positive off-diagonal entries. This implies in particular that $A^{-1} \geq 0$. The Cryer algorithm is valid for all Minkowski matrix. In the particular case, such as ours, where A is a *tridiagonal* Minkowski matrix, an implementation of this basic method which minimizes the amount of computation can be found in [70].

18 Finite Differences for bi-dimensional Vanilla Options

The purpose of this Section is to describe various algorithms for pricing options in the Black–Scholes bidimensional setting, relying upon the ADI (Alternate Direction Implicit) method. The ADI method can in fact be used in any multi-dimensional Markov model. It is actually the current industry standard to solve multi-dimensional pricing problems by PDE methods.

In the risk-neutral Black–Scholes bidimensional model, the underlying stock-prices satisfy the following stochastic differential equations:

$$\left\{\begin{array}{l} dS_t^1 = S_t^1(\kappa_1 dt + \sigma_{11} dW_t^1 + \sigma_{12} dW_t^2), \ \ \ln(S_0^1) = x_1\\ dS_t^2 = S_t^2(\kappa_2 dt + \sigma_{21} dW_t^1 + \sigma_{22} dW_t^2), \ \ \ln(S_0^2) = x_2\end{array}\right.$$

for independent Brownian motions W^1, W^2, with

$$\left(\begin{array}{c}\kappa_1\\ \kappa_2\end{array}\right) = \left(\begin{array}{c} r - q_1\\ r - q_2\end{array}\right) \quad , \quad \boxed{\Sigma = \left(\begin{array}{cc}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{array}\right) = \left(\begin{array}{cc}\sigma_1 & 0\\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2\end{array}\right)}$$

so that in particular $\boxed{\Sigma\Sigma^{\mathsf{T}} = \Gamma}$, where Γ is the following *covariance matrix*:

$$\boxed{\Gamma = \left(\begin{array}{cc}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{array}\right)}$$

In order to apply the ADI method, we'd rather work with the underlying *decorrelated* bidimensional Brownian motion. In this view we introduce

$$\varphi(t, w_1, w_2) := \phi(t, e^{x_1 + b_1 t + \sigma_1 w_1}, e^{x_2 + b_2 t + \sigma_2(\rho w_1 + \sqrt{1-\rho^2} w_2)}) \ ,$$

with $(b_1, b_2) = (\kappa_1 - \frac{1}{2}\sigma_1^2, \kappa_2 - \frac{1}{2}\sigma_2^2)$, so that the payoff of the option writes $\varphi(t, W_t^1, W_t^2)$. The price at time 0 of an European option on (S_T^1, S_T^1) is then given by:

$$\boxed{\mathbb{E}e^{-rT}\phi(T, S_T^1, S_T^2) = v(0, S_0^1, S_0^2) = u(0, 0, 0) = \mathbb{E}e^{-rT}\varphi(T, W_T^1, W_T^2)}$$

where u is the unique viscosity solution with suitable growth conditions to the following two dimensional PDE:

$$\boxed{\left\{ \begin{array}{l} u(T, w_1, w_2) = \varphi(T, w_1, w_2) \quad \text{on } \mathbb{R}^2 \\ \partial_t u(t, w_1, w_2) + \frac{1}{2}\partial^2_{w_1^2} u(t, w_1, w_2) + \frac{1}{2}\partial^2_{w_1^2} u(t, w_1, w_2) - ru(t, w_1, w_2) = 0 \quad \text{on } [0, T) \times \mathbb{R}^2 \end{array} \right.} \tag{183}$$

and v satisfies the analog equation in the S variables. Observe that

$$\begin{pmatrix} \ln S^1 \\ \ln S^2 \end{pmatrix} = \begin{pmatrix} x_1 + b_1 T \\ x_2 + b_2 T \end{pmatrix} + \Sigma \begin{pmatrix} W^1 \\ W^2 \end{pmatrix} .$$

Hence

$$\begin{pmatrix} S^1\partial_{S^1} \\ S^2\partial_{S^2} \end{pmatrix} = \Sigma^{-1} \begin{pmatrix} \partial_{W^1} \\ \partial_{W^2} \end{pmatrix} = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \begin{pmatrix} \sigma_2\sqrt{1-\rho^2} & 0 \\ -\sigma_2\rho & \sigma_1 \end{pmatrix} \begin{pmatrix} \partial_{W^1} \\ \partial_{W^2} \end{pmatrix} .$$

The deltas $\partial_S v(0, S_0^1, S_0^2)$ are thus given in terms of u as

$$\boxed{\Delta^1 = \frac{e^{-x_1}}{\sigma_1}\partial_{w_1} u \ , \ \Delta^2 = \frac{e^{-x_2}}{\sqrt{1-\rho^2}}\left(\frac{-\rho\partial_{w_1} u}{\sigma_1} + \frac{\partial_{w_2} u}{\sigma_2}\right)} \tag{184}$$

For the numerical resolution of problem (183) by finite differences:
• we localize in space the problem on a domain $\bar{\mathcal{D}} = [-\ell, \ell]^2$, introducing a suitable boundary condition on $(0, T) \times \partial\mathcal{D}$;
• we introduce a grid of mesh points $(t, w_1, w_2) = (ih, j_1 k_1, j_2 k_2)$ on $[0, T] \times \bar{\mathcal{D}}$, for *mesh steps* (which can be thought of as tending to zero) h, k_1, k_2, and we discretize the localized problem on the time-space grid by a suitable finite differences scheme, such as the *ADI scheme* described in the next section.
Let $u_i^{j_1, j_2}$ denote the solution of the discrete problem. The deltas are then retrieved by using suitable finite differences approximations for $\partial_{w_1} u$ and $\partial_{w_2} u$ in (184), like

$$\partial_{w_1} u \approx \frac{u_0^{j_1+1, j_2} - u_0^{j_1-1, j_2}}{2k} \ , \ \partial_{w_2} u \approx \frac{u_0^{j_1, j_2+1} - u_0^{j_1, j_2-1}}{2k} .$$

18.1 Numerical integration by an ADI Method

Alternate Direction Implicit (ADI) methods (see Peachman–Rachford [153], Morton–Mayers [143]) consist in decomposing each discrete time step in two half-steps, the first one in the first space variable and the second one in the other, as

$$\left\{ \begin{array}{c} \frac{2}{h}(u_{i+1} - u_{i+\frac{1}{2}}) + \frac{1}{2}\delta^2_{w_1^2} u_{i+\frac{1}{2}} + \frac{1}{2}\delta^2_{w_2^2} u_{i+1} - \frac{1}{2}ru_{i+\frac{1}{2}} - \frac{1}{2}ru_{i+1} = 0 \\ \frac{2}{h}(u_{i+\frac{1}{2}} - u_i) + \frac{1}{2}\delta^2_{w_1^2} u_{i+\frac{1}{2}} + \frac{1}{2}\delta^2_{w_2^2} u_i - \frac{1}{2}ru_{i+\frac{1}{2}} - \frac{1}{2}ru_i = 0 \ , \end{array} \right.$$

that is:

$$\boxed{\left\{ \begin{array}{c} [(1 + \frac{hr}{4})\mathrm{Id} - \frac{h}{4}\delta^2_{w_1^2}]u_{i+\frac{1}{2}} = [(1 - \frac{hr}{4})\mathrm{Id} + \frac{h}{4}\delta^2_{w_2^2}]u_{i+1} \\ [(1 + \frac{hr}{4})\mathrm{Id} - \frac{h}{4}\delta^2_{w_2^2}]u_i = [(1 - \frac{hr}{4})\mathrm{Id} + \frac{h}{4}\delta^2_{w_1^2}]u_{i+\frac{1}{2}} \end{array} \right.} \tag{185}$$

with

$$(\delta^2_{w_1^2}u)_i^{j_1,j_2} = \frac{u_i^{j_1-1,j_2} - 2u_i^{j_1,j_2} + u_i^{j_1+1,j_2}}{k^2}$$
$$(\delta^2_{w_2^2}u)_i^{j_1,j_2} = \frac{u_i^{j_1,j_2-1} - 2u_i^{j_1,j_2} + u_{j_1,j_2+1}}{k^2} .$$

The system (185) is of the form

$$\boxed{\begin{cases} (Au_{i+\frac{1}{2}}^{\cdot,j_2})_{j_2} = (Bu_{i+1}^{\cdot,j_2})_{j_1} \text{ , for any } j_2 \\ (Cu_i^{j_1,\cdot})_{j_1} = (Du_{i+\frac{1}{2}}^{j_1,\cdot})_{j_2} \text{ , for any } j_1 \end{cases}} \tag{186}$$

for tridiagonal matrices A, B, C, D. So each time step consists in the resolution of $m_1 + m_2$ implicit one-dimensional problems, solvable by the Gauss method. This is in general a far better alternative than having to solve the $m_1 m_2$-dimensional linear system that would arise from a two-dimensional implicit scheme (the sparseness of the related matrix may be fruitfully used in an *iterative approximation scheme*, however).

In less elementary situations where the correlation cannot be eliminated of the problem, the correlation terms are treated in an explicit way (i.e., aggregated to the r.h.s. of (185) or (186)), which induces a related stability condition.

Also note that the ADI method admits suitable extensions in arbitrary space dimension q.

18.2 American Options

Likewise, the price of a $2D$ American vanilla option at time 0 is given by (see Part II)

$$\sup_{\tau \in \mathcal{T}} \mathbb{E} e^{-r\tau} \varphi(\tau, W_\tau^1, W_\tau^2) = u(0,0,0)$$

where u is the solution to the following variational inequality (obstacle problem):

$$\begin{cases} \max\left(\varphi - u, \partial_t u + \frac{1}{2}\partial^2_{x^2} u + \frac{1}{2}\partial^2_{y^2} u - ru\right) = 0 \text{ on } [0,T) \times \mathbb{R}^2 \\ u(T,x,y) = \varphi(T,x,y) \text{ on } \mathbb{R}^2 . \end{cases} \tag{187}$$

Again:
• we localize in space problem (187) on a domain $\bar{\mathcal{D}} = [-\ell, \ell]^2$, introducing a suitable boundary condition $u = \varphi$ on $(0,T) \times \partial\mathcal{D}$;
• we introduce a grid of mesh points $(t,x,y) = (ih, j_1 k_1, j_2 k_2)$ on $\bar{\mathcal{D}}$ and we discretize and solve the localized problem on the grid.
A suitable finite differences approximation scheme on the grid can be obtained by combining the previous ADI finite difference method with the splitting method of section 17.2 (see, e.g., [177]).

19 Finite Differences for Exotic Options

19.1 Lookback Options

Lookback otions are options with payoffs of the form $\phi(S, M)$, where $M_t = \sup_{0 \le s \le t} S_s$. The Black–Scholes pricing function v of a lookback option satisfies the usual Black–Scholes

pricing PDE in the (t, S) variables, but in the subdomain $S \leq M$ of a *three-dimensional* state space (t, S, M). On the boundary $\{S = M\}$, v satisfies an homogenous oblique Neumann condition $\partial_M uv = 0$ (see [180]).

This can be established by a formal application of the Itô semimartingale formula to the price process $\Pi_t = v(t, S_t, M_t)$. Note that M is a non-decreasing process, yet it is not absolutety continuous w.r.t. the Lebesgue measure, thus the relevant Itô formula is not contained in (36). Here the relevant Itô formula is the general Itô formula for semimartingales (see, e.g., [57, 104]), yielding:

$$\boxed{dv_t = \left(\partial_t v + \tfrac{1}{2}\sigma^2 S^2 \partial^2_{S^2} v + \kappa S \partial_S v - rv\right) dt + \partial_M v \, dM_t + \sigma S \partial_S v \, dW_t}$$

We thus deduce from the local martingale property of the discounted price $e^{-rt}\Pi_t$ that on $\{t < T\}$:

$$\partial_t v + \tfrac{1}{2}\sigma^2 S^2 \partial^2_{S^2} v + \kappa S \partial_S v - rv = 0 \text{ on } \{S < M\}$$
$$\partial_M v = 0 \text{ on } \{S = M\} .$$

Remark 19.1 **(i)** This can also be derived by letting n tend to ∞ in the pricing PDE of the option with approximating payoff $\phi(S_T, M_T(n))$, where

$$M_t(n) = \left(\int_0^t S^n ds\right)^{\frac{1}{n}} , \; dM_t(n) = \frac{1}{n} \frac{S_t^n}{M_t(n)^{n-1}} dt .$$

(ii) When S_t gets close to M_t, then it becomes very likely that the related value M_t of M (*running maximum* of S) will have changed by T. It follows that the option value is insensitive to small changes in the value of M in this case, which provides an intuitive insight into the fact that $\partial_M v = 0$ on $\{S = M\}$.

19.2 Barrier Options

A barrier option is a type of financial option where the option to exercise depends on the underlying crossing or reaching a given barrier level. For instance, barrier *up and out* options at level H correspond to the following payoff process $\phi(S_t, M_t)$, involving in particular the *rebate* R :

$$\phi(S_t, M_t) = \mathbb{1}_{\{M_t < H\}} \phi(S_t) + \mathbb{1}_{\{M_t \geq H\}} R \tag{188}$$

which is paid at time

$$\tau = \inf\{t \geq 0; S_t \geq H\} \wedge T .$$

The related price process is thus given as $\Pi_t = \beta_t^{-1} \mathbb{E}_t \beta_\tau \phi(S_\tau, M_\tau)$, $t \in [0, T]$.

Such options were created as a way to provide the insurance value of an option without charging as much premium. For instance, if a trader believes that IBM will go up this year, but she is willing to bet that it won't go above 100, then she can buy the barrier and pay less premium than the vanilla option.

Other common forms of barrier options are up and in, down and out, down and in, double in, and double out options. In some cases, the rebate may be paid at the maturity time T

rather than at the time of the barrier event. Barrier options are then special cases of lookback options, so they may be handled as in the previous section. Their pricing functions then solve related *two-dimensional Cauchy problems* in the set of variables (t, S, M).

However, in any case, it is better to use the fact that the Black–Scholes *pre-barrier event pricing function* $u = u(t, x) = v(t, S)$ of a barrier option, where

$$\mathbb{1}_{t \leq \tau} \Pi_t = \mathbb{1}_{t \leq \tau} v(t, S_t)$$

and $x = \ln(S)$, generically solves a *uni-dimensional Cauchy–Dirichlet problem* of the following kind:

$$\boxed{\begin{cases} \partial_t u + \frac{1}{2}\sigma^2 \partial^2_{x^2} u + b \partial_x u - ru = 0 \text{ on } [0, T) \times \mathcal{D}, \\ u(T, x) = \varphi(x) \text{ on } \mathcal{D}, \\ u(t, x) = R(t, x) \text{ on } [0, T] \times \partial\mathcal{D} \, . \end{cases}} \tag{189}$$

More precisely, in the case of a rebate paid at the time of the barrier event, we have with $l = \ln(L), h = \ln(H)$, and denoting also by $C(t, x)$ is the pricing function of an European vanilla call with maturity T :

- *Up Out Barrier* h : $\mathcal{D} = (x - \ell, h)$, $\varphi(x) = \phi(S), R(t, h) = R$;
- *Up In Barrier* h : $\mathcal{D} = (x - \ell, h)$, $\varphi(x) = R$, $R(t, h) = C(t, h)$;
- *Down Out Barrier* l : $\mathcal{D} = (l, x + \ell)$, $\varphi(x) = \phi(S), R(t, l) = R$;
- *Down In Barrier* l : $\mathcal{D} = (l, x + \ell)$, $\varphi(x) = R$, $R(t, l) = C(t, l)$;
- *Double Out Barriers* l, h : $\mathcal{D} = (l, h)$, $\varphi(x) = \phi(S), R(t, l) = R(t, h) = R$;
- *Double In Barriers* l, h : $\mathcal{D} = (l, h)$, $\varphi(x) = R$, $R(t, l) = C(t, l)$, $R(t, h) = C(t, h)$.

For computing the (pre-barrier event) pricing function of a barrier option, one may thus apply a standard theta-scheme to (189) (or the analog problem in the $x = \ln(S)$ variable). The use of implicit schemes is recommended, for stability issues.
For the sake of accuracy of the resulting scheme, grid points should be put *on* the barrier. In the case of *curved barriers*, *artificial Dirichlet boundary conditions* may be imposed along the barrier (Dirichlet boundary conditions on fictitious grid points along the barrier, see, e.g., [143, 65]).

Linear interpolation may be used to find the price and delta corresponding to the initial stock price if need be. If the initial stock price is close to barrier, a one-sided second-order finite difference approximation should be used for getting the delta.

19.3 Asian options

Asian options are options with factors S and $I = \int_0^{\cdot} S_t dt$, the *running time-average of* S. Let us consider a few specific examples, in the simple set-up of the Black–Scholes model. The pair (S, I) is then a Markov process with generator $\mathcal{A}_{S,I}$ given by

$$\mathcal{A}_{S,I} = \frac{1}{2}\sigma^2 S^2 \partial^2_{S^2} + \kappa S \partial_S + S \partial_I \, . \tag{190}$$

Note that the generator $\mathcal{A}_{S,I}$ is degenerate in the I variable.

19.3.1 European Fixed Strike Asian Put option

This is the option with the following payoff at T :

$$\xi = \left(K - \frac{I_T}{T}\right)^+ = \phi(I_T) \, , \tag{191}$$

and related price process

$$\Pi_t = \beta_t^{-1}\mathbb{E}_t\beta_T\xi = v(t,S_t,I_t),\ t\in[0,T]. \tag{192}$$

The formally related bi-dimensional pricing problem writes:

$$\boxed{\begin{cases} \partial_t v + \mathcal{A}_{S,I}v = rv, \ \ 0\le t<T \\ v(T,S,I)=\phi(I) \end{cases}} \tag{193}$$

One can thus show that $\Pi_t = v(t,S_t,I_t)$, $t\in[0,T]$, where v is the unique bounded viscosity solution of (193).

Note that the numerical resolution of the PDE (193) requires special care to cope with the degeneracy of the generator $\mathcal{A}_{S,I}$ in the I variable ('PDE in dimension $1\frac{1}{2}$', see Zvan et al. [186]).

Alternatively to the previous approach, it is possible to reduce the pricing problem to a one-dimensional one (see Rogers-Shi [164]), by working in the numeraire S (see section 8.1). To start with, observe that $\frac{\phi(I_T)}{S_T} = \eta_T^+$, where the process η is defined by

$$\eta_t = \frac{1}{S_t}\left(K - \frac{I_t}{T}\right)\ ,\ d\eta_t \ = \ -\frac{dt}{T} - \eta_t(\kappa-\sigma^2)dt - \eta_t\sigma dW_t$$

with related generator

$$\mathcal{A}_\eta = -\left[\frac{1}{T} + (\kappa-\sigma^2)\eta\right]\partial_\eta + \frac{1}{2}\eta^2\sigma^2\partial^2_{\eta^2}\ . \tag{194}$$

Now, in the numeraire S, the price process (192) writes (cf. (79), (80)):

$$\Pi_t = S_t\widetilde{\mathbb{E}}_t S_T^{-1}\phi(I_T) = S_t\widetilde{\mathbb{E}}_t\eta_T^+ = S_t u(t,\eta_t),\ t\in[0,T] \tag{195}$$

where a process X is a $\widetilde{\mathbb{P}}$ – local martingale if and only if $\beta_t S_t X_t$ is a $\mathbb{P}$ – local martingale. This last property allows one to derive the dynamics of η under the valuation measure $\widetilde{\mathbb{P}}$ and to check that η is a $\widetilde{\mathbb{P}}$ – Markov process, which justifies the last equality in (195)[5]. Moreover, since $u(t,\eta_t)$ is a (Doob) $\widetilde{\mathbb{P}}$ – local martingale, thus $\beta_t S_t u(t,\eta_t)$ is a $\mathbb{P}$ – local martingale.

In order to guess the form of the equation solved by u, note further that for u of class $\mathcal{C}^{1,2}$, Itô calculus yields, with "$\triangleq$" standing for "equality up to a local martingale term":

$$e^{rt}d\left[e^{-rt}S_t u(t,\eta_t)\right] \triangleq \Big(-rS_t u(t,\eta_t) + u(t,\eta_t)\kappa S_t dt + S_t\left(\partial_t u + \mathcal{A}_\eta u\right)(t,\eta_t) - \sigma S_t\eta\sigma\partial_\eta u(t,\eta_t)\Big)dt\ .$$

Should $S_t u(t,\eta_t)$ be equal to the price process Π_t, u would then solve the PDE (196) below. Now, this PDE has a unique classic solution u, by Frieman [90]. Conversely, the previous computations thus show that $u(t,\eta_t)$, with u defined as the solution to (196), satisfies (195). This yields the following

Proposition 19.1 $\Pi_t = S_t u(t,\eta_t)$, $t\in(0,T]$, *where the* pricing function in the numeraire S, $u = u(t,\eta)$, *is the unique (classic) solution to the following* one-dimensional PDE*:*

$$\boxed{\begin{cases} \partial_t u - \left(\frac{1}{T} + \kappa\eta\right)\partial_\eta u + \frac{1}{2}\sigma^2\eta^2\partial^2_{\eta^2}u - qu = 0 \\ u(T,\eta) = \eta^+ \end{cases}} \tag{196}$$

[5] On the general issue of the preservation of the Markov property by change of numeraire, see the concluding remarks to section 8.1 and the references given therein.

We thus reduced the original two-dimensional degenerate pricing PDE to a one-dimensional non-degenerate PDE which can be solved numerically by standard finite differences theta-schemes. Note that the advection term is dominant in this equation (due to the first-order coefficient $\frac{1}{T}$), which requires special care from the numerical point of view. This can be handled by a further transformation of the problem, however (see [74]).

19.3.2 American Fixed Strike Asian Put option

This is the American counterpart to the previous option, thus the option with the following payoff process which is paid at a stopping time τ at the holder's convenience between 0 and T :

$$\left(K - \frac{I_\tau}{\tau}\right)^+ = \phi(\tau, I_\tau) \ , \ t \in [0, T] \tag{197}$$

The related price process writes:

$$\Pi_t = \beta_t^{-1} \text{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E}_t \beta_\tau \phi(\tau, I_\tau) = v(t, S_t, I_t), \ t \in [0, T] \tag{198}$$

with associated pricing (obstacle) problem

$$\boxed{\begin{cases} \min(-\partial_t v - \mathcal{A}_{S,I} v + rv \ , \ v - \phi(t, I)) = 0 \ , \quad 0 < t < T \\ v(T, S, I) \ = \ \phi(T, I) \end{cases}} \tag{199}$$

where $\mathcal{A}_{S,I}$ was defined in (190). Note that the obstacle function $\phi(t, I)$ is singular at inception time $t = 0$, so that the pricing problem (199) is only defined on $(0, T] \times (0, +\infty)^2$.

Proposition 19.2 (see Crépey [64]) $\Pi_t = v(t, S_t, I_t)$, $t \in (0, T]$, *where v is the unique bounded viscosity solution to (199). Moreover, the price at inception time 0 is given by the radial limit*

$$\boxed{\Pi_0 = \lim_{t \to 0^+} v(t, S_0, tS_0)}$$

Figure 9 provides a numerical illustration of the last point in Proposition 19.2 (radial convergence at time 0). Numerically it seems that convergence also occurs along radii α other than S_0, yet at a slower rate than for $\alpha = S_0$.

Let us stress again that the two-dimensional PDEs in this Section (PDEs (193) and (199)) are degenerate in the I variable, so that specific treatments are necessary to solve them numerically, particularly in the singular American case (see Zvan et al. [186]).

19.4 Discretely Path-Dependent Options

Discretely path dependent payoffs $\xi = \phi(S_{t_0}, S_{t_1}, \ldots, S_{t_n})$ paid at T, relatively to a set of *monitoring dates* $(t_0 = 0, \ldots, t_n = T)$, can often be priced efficiently by PDE methods after an appropriate extension of the state space. In this Section we shall exemplify this technique on the case of *cliquet options*, *volatility and variance swaps* and *discretely sampled Asian options* (see, e.g., [180, 181, 73, 183]).

Note that the pricing of such products by Monte Carlo methods is obvious in the Black–Scholes model, but it becomes very difficult in the case of American options, or in simple extensions of Black–Scholes like the Uncertain Volatility model in which the volatility process

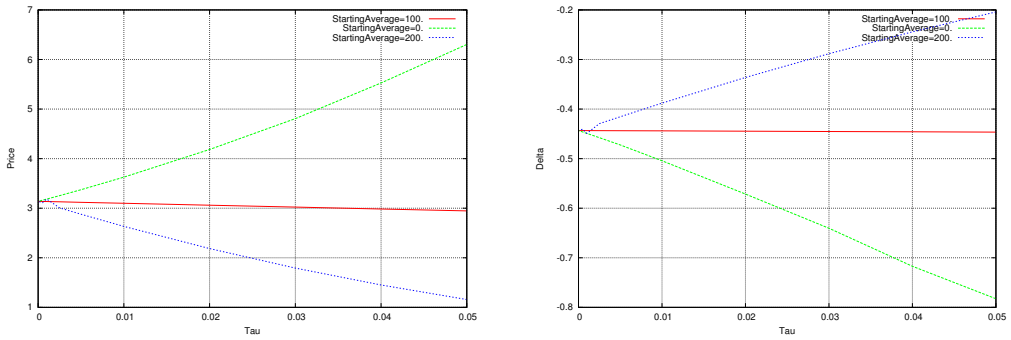

Figure 9: *Convergence of the price and delta of an American Asian put at* $(t, S_0, \alpha t)$ *as* $t \to 0$, *for* $\alpha = 0$, S_0 *or* $2S_0$.

is only known to remain in a constant range $(\underline{\sigma}, \overline{\sigma})$ (see Avellaneda et al. [15]). Yet the PDE method can easily cope with such situations (see, e.g., Windcliff et al. [179, 182]).

19.4.1 Cliquet Options

Let $R_i = \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}$ denote the (simple) spot return on the period $[t_{i-1}, t_i]$, for $i = 1, \ldots, n$. The payoff of a cliquet is defined by (up to a constant *notional* factor):

$$\boxed{\xi = \max(F_g, \min(C_g, \textstyle\sum_{i=1}^{n} \max(F_l, \min(C_l, R_i))))}$$

for given numbers F_l, C_l, F_g and C_g. In order to markovianize this payoff, we introduce two additional state variables: P and Z, such that

$$\boxed{\begin{array}{c} P(t_i \leq t < t_{i+1}) = S(t_i) \\ Z(t_i \leq t < t_{i+1}) = \frac{1}{i} \sum_{k=1}^{i} \max(F_l, \min(C_l, R_k)) \end{array}} \tag{200}$$

Note that the factor process (S, P, Z) is Markovian in the risk-neutral Black–Scholes model for S, with related generator $\mathcal{A}_{S,P,Z} u(S, P, Z)$ given by $\mathcal{A}_S u(S, P, Z)$ on each time interval (t_{i-1}, t_i), where $\mathcal{A}_S$ denotes the usual risk-neutral Black–Scholes generator. Moreover, we have

$$\boxed{\xi = \max(F_g, \min(C_g, nZ_T)) = \phi(Z_T)}$$

In the Black–Scholes model we thus have $\Pi_t = v(t, S_t, P_t, Z_t)$, for a Borel-measurable (deterministic) function v. Given the further jump conditions defined by (200), whereas by arbitrage the pricing process Π must be time-continuous at the t_is, the pricing function v is thus given by $v = \phi =: v_n$ at T and

$$v = v_i \text{ on } [t_i, t_{i+1}) \text{ , for } i = n-1, \ldots, 0$$

where $(v_i)_{0 \leq i \leq n-1}$ is the unique sequence of (viscosity) solutions with suitable growth conditions to the following *PDE cascade*, defined for i decreasing from $n-1$ to 0:

$$\boxed{\left\{ \begin{array}{l} v_i(t_{i+1}, S, P, Z) = v_{i+1}(t_{i+1}, S, P_+, Z_+) \text{ on } (0, \infty)^3 \\ \partial_t v_i + \frac{1}{2}\sigma^2 S^2 \partial^2_{S^2} v_i + \kappa S \partial_S v_i - r v_i = 0 \text{ on } [t_i, t_{i+1}) \times (0, \infty)^3 \end{array} \right.} \tag{201}$$

where P_+ and Z_+ in the third line are obtained via the jump conditions stemming from (200) at the monitoring date t_{i+1}, as (cf. (200))

$$\boxed{P_+ = S \ , \ Z_+ = \tfrac{i}{(i+1)} Z + \tfrac{\rho}{(i+1)}} \tag{202}$$

with $\rho = \max(F_l, \min(C_l, \frac{S-P}{P}))$.

In order to solve (201), we localize in space the problem on a compact set

$$\bar{\mathcal{D}} = [\underline{S}, \overline{S}] \times [\underline{S}, \overline{S}] \times [\underline{Z}, \overline{Z}]$$

with

$$\underline{S} = 0 \ , \ \overline{S} = S_0(1 + f\sigma\sqrt{T}) \ , \ \underline{Z} = \min(0, F_l) \ , \ \overline{Z} = C_l$$

for a suitable factor f, e.g., $f = 5$. Suitable boundary conditions are $\partial^2_{S^2} u = 0$ on $S = \underline{S}$ and $S = \overline{S}$. We then put a finite differences mesh on $\mathcal{D}$. It is recommended to use a non uniform P-mesh $(P^{j_1})_{1 \le j_1 \le m_1}$ concentrating the mesh points near the spot price S_0 (and containing in particular S_0) and to resort to an adaptive grid in S finer near the diagonal of (S, P), defining for instance a (P, S)-grid of the form, for $1 \le j_1 \le m_1$, $1 \le j_2 \le m_2$:

$$(P, S)^{j_1, j_2} = (P^{j_1}, \frac{P^{j_1} P^{j_2}}{S_0})$$

(see Windcliff et al. [182]). An independent uniform grid may be used in Z.
The jump conditions (202) can then be implemented by linear interpolation.
Between monitoring dates, the Black–Scholes equation (201) can be discretized by a standard finite differences scheme (theta-scheme) in the (t, S) variables, defined on the *three-dimensional state space* (S, P, Z).
Finally, interpolation of degree two in space is used for computing the price, the delta and the gamma of the option at $0, S_0, P_0, Z_0$.

19.4.2 Volatility and Variance Swaps

Variance Swaps, resp. Volatility Swaps, correspond to the payoffs at T defined in terms of the *realized variance* $V^2 = \sum_{i=1}^n \ln(\frac{S_{t_i}}{S_{t_{i-1}}})^2$ and the strike K as

$$\boxed{\xi = V^2 - K^2 \ , \text{ resp. } V - K}$$

We use the same notation as in the case of cliquet options, except for the facts that R_i and Z are now to be understood as, respectively, the *continuous spot return* $R_i = \ln(\frac{S_{t_i}}{S_{t_{i-1}}})$, and

$$\boxed{Z(t_i \le t < t_{i+1}) = \tfrac{1}{i} \sum_{k=1}^i R_k^2} \tag{203}$$

In this notation we have in either case:

$$\boxed{\xi = \phi(Z)}$$

for a suitable terminal payoff function ϕ. Like in the case of cliquet option, the Black–Scholes pricing function v of a volatility or variance swap is thus given as $v = v(t, S, P, Z)$

where $v = (v_i)_{0 \le i \le n}$ (with $v_n = \phi$) solves a PDE cascade of the same form as (201) along with the jump conditions (202), where ρ is now to be understood as $\rho = \ln(\frac{S}{P})^2$.

19.4.3 Discretely Monitored Asian Options

In real-life contracts, Asian options are in fact defined in terms of *discretely sampled payoffs*, like

$$\xi = \left(K - \frac{T}{n} \sum_{i=1}^{n} S_{t_k} \right)^+ , \tag{204}$$

Let us introduce the process Y such that

$$\boxed{Y(t_i \le t < t_{i+1}) = \tfrac{1}{i} \textstyle\sum_{k=1}^{i} S_{t_i}} \tag{205}$$

The factor process (S, Y) is Markovian in the risk-neutral Black–Scholes model for S, with related generator $\mathcal{A}_{S,Y} v(S, Y)$ given by the usual Black–Scholes generator $\mathcal{A}_S v(S, Y)$ on each time interval (t_{i-1}, t_i). Moreover, we have

$$\boxed{\xi = (K - TY_T)^+ = \phi(Y_T)}$$

In the Black–Scholes model we then have $\Pi_t = v(t, S_t, Y_t)$, for a Borel-measurable (deterministic) function v. We proceed like in the case of cliquet options, obtaining in this case $v = (v_i)_{0 \le i \le n}$ such that $v_n = \phi$, and for i decreasing from $n - 1$ to 0:

$$\boxed{\begin{cases} v_i(t_{i+1}, S, Y) = v_{i+1}(t_{i+1}, S, Y_+) \text{ on } (0, \infty)^3 \\ \partial_t v_i + \frac{1}{2}\sigma^2 S^2 \partial^2_{S^2} v_i + \kappa S \partial_S v_i - r v_i = 0 \text{ on } [t_i, t_{i+1}) \times (0, \infty)^2 \end{cases}} \tag{206}$$

where Y_+ is obtained via the following jump conditions at the monitoring date t_{i+1} (cf. (205)):

$$\boxed{Y_+ = \tfrac{i}{(i+1)} Y + \tfrac{S}{(i+1)}} \tag{207}$$

Remark 19.2 In the case of discretely monitored Asian options:
(i) The situation is thus simpler (and the computation faster) than for cliquet options or volatility and variance swaps, since the related payoff is a function of the Markovian *pair* (S, Y), instead of the *triple* (S, P, Z);
(ii) This approach is both closer to the clauses of the product, and also easier from the numerical point of view, than the approach by time-continuous integals of S of section 19.3. Indeed, the cost of solving (206) is essentially that of solving m_1 one-dimensional PDE problems, where m_1 is a generic number of mesh points per space dimension. This is a far better alternative than having to solve the two-dimensional degenerate PDEs of section 19.3 (problems in dimension '$1\frac{1}{2}$', unless specific dimension reduction techniques are available, see section 19.3).

Part V
Tree Pricing Methods

Tree methods (historically starting with *binomial tree* methods) are but another point of view on explicit finite differences methods, interpreted probabilistically in terms of a related Markov chain.
Tree methods are natural in finance because the related Markov chains can also be considered as realistic and arbitrage-free models of trading and hedging in discrete time, that can be used per se, without reference to a continuous-time model.
Moreover, tree methods compute prices and Greeks in a *cone* starting from the state $\mathcal{Z}_{t_0}$ of the factor process at the pricing time t_0, rather than on the whole state space by (implicit) finite differences methods. In finance, the main concern is to get prices and Greeks at $\mathcal{Z}_{t_0}$. So the extra-work furnished by finite differences methods for computing the solution everywhere, is in a sense waste of time.

Today it is fair to see that *from a practical point of view*, binomial trees are obsolete as compared with more sophisticated finite differences (or finite elements) technologies. However, in a number of situations, *trinomial* trees remain a simple, flexible and good enough alternative. Moreover, tree methods are easy to customize due to the visualization of the paths of the underlying in the related Markov chain. So it is often straightforward to design a tree algorithm for the pricing of a somewhat involved contingent claim.

Yet from the theoretical point of view (*convergence analysis* of the related numerical schemes), the Markov chain interpretation opens the way to alternative *probabilistic convergence proofs* of the related numerical pricing schemes, as we shall now see.

20 General Markov Chain Approximation Results

Let thus $\mathcal{Z}_t^h = (t, \mathcal{X}_t^h, \mathcal{Y}_t^h)$ stand for a continuous-time Markov chain approximation of our generic Markovian jump-diffusion setting with regimes $\mathcal{Z} = (t, \mathcal{X}, \mathcal{Y})$ on the state space $E = [0,T] \times \mathbb{R}^q \times I$ (cf. section 7.2). Assume $(0, x_0^h, y) =: z_0^h = \mathcal{Z}_0^h \to \mathcal{Z}_0 = z_0 := (0, x_0, y)$ as $h \to 0$, as well as

$$\lim_{h\to 0} h^{-1}\mathbb{E}_t(\Pi(\mathcal{Z}_{t+h}^h) - \Pi(\mathcal{Z}_t^h)) = (\partial_t + \mathcal{A})\Pi(z) \tag{208}$$

on the random set $\{\lim_{h\to 0} \mathcal{Z}_t^h = z\}$, for any $z \in E$ and any sufficiently regular function Π on E. Note that the r.h.s. of (208) can also be interpreted as "$\lim_{h\to 0} h^{-1}\mathbb{E}_t(\Pi(\mathcal{Z}_{t+h}) - \Pi(\mathcal{Z}_t))$ on $\{\mathcal{Z}_t = z\}$", by application of the Itô formula (36) to the process $\Pi(\mathcal{Z}_t)$.

Then, under suitable technical assumptions (see Ethier and Kurtz [85], Kushner and Dupuis [121], Jacod and Shiryaev [104]): the process $\mathcal{Z}^h$ converges in law to $\mathcal{Z}$ as $h \to 0$, thus

$$\boxed{\Pi^h(z_0^h) := \sup_{\tau\in\mathcal{T}} \mathbb{E}e^{-\int_0^\tau r(t,\mathcal{Z}_t^h)dt}\phi(\mathcal{Z}_\tau^h) \to \sup_{\tau\in\mathcal{T}} \mathbb{E}e^{-\int_0^\tau r(t,\mathcal{Z}_t)dt}\phi(\mathcal{Z}_\tau) =: \Pi(z_0)} \tag{209}$$

as $h \to 0$, for any bounded and Borel-measurable function ϕ on E.

20.1 Kushner's theorem

Note that (208) implies in particular, by taking $\Pi(z) = \Pi^i(z) := x^i$ therein, for $i = 1 \dots q$:

$$\boxed{\lim_{h\to 0} h^{-1}\mathbb{E}_t(\mathcal{X}_{t+h}^{h,i} - \mathcal{X}_t^{h,i}) = (\partial_t + \mathcal{A})\Pi^i(z) = b_i(z)} \tag{210}$$

on $\{\lim_{h\to 0} \mathcal{Z}_t^h = z\}$, for any $z \in E$. Moreover, for any fixed $h > 0$, we have for any $i, j = 1 \dots q$: (cf. the proof of Proposition 7.1):

$$\mathbb{C}\mathrm{ov}_t(\mathcal{X}_{t+h}^{h,i} - \mathcal{X}_t^{h,i}, \mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,j}) + \mathbb{E}_t(\mathcal{X}_{t+h}^{h,i} - \mathcal{X}_t^{h,i})\mathbb{E}_t(\mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,j}) =$$
$$\mathbb{E}_t\left(\mathcal{X}_{t+h}^{h,i}\mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,i}\mathcal{X}_t^{h,j}\right) - \mathcal{X}_t^{h,i}\mathbb{E}_t\left(\mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,j}\right) - \mathcal{X}_t^{h,j}\mathbb{E}_t\left(\mathcal{X}_{t+h}^{h,i} - \mathcal{X}_t^{h,i}\right) .$$

Now, on $\{\lim_{h\to 0} \mathcal{Z}_t^h = z\}$, we deduce from (208) applied to $\Pi = \Pi^i = x_i$, $\Pi = \Pi^j = x_j$ and $\Pi = \Pi^i\Pi^j$, respectively:

$$\lim_{h\to 0} h^{-1}\mathbb{E}_t(\mathcal{X}_{t+h}^{h,i} - \mathcal{X}_t^{h,i}) = \mathcal{A}\Pi^i(z) \ , \ \lim_{h\to 0} h^{-1}\mathbb{E}_t(\mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,j}) = \mathcal{A}\Pi^j(z)$$
$$\lim_{h\to 0} h^{-1}\mathbb{E}_t\left(\mathcal{X}_{t+h}^{h,i}\mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,i}\mathcal{X}_t^{h,j}\right) = \mathcal{A}(\Pi^i\Pi^j)(z)$$

As $h \to 0$ we thus deduce from (211), on $\{\lim_{h\to 0} \mathcal{Z}_t^h = z\}$:

$$\lim_{h\to 0} h^{-1}\mathbb{C}\mathrm{ov}_t(\mathcal{X}_{t+h}^{h,i} - \mathcal{X}_t^{h,i}, \mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,j}) = \mathcal{A}(\Pi^i\Pi^j)(z) - x_i\mathcal{A}\Pi^j(z) - x_j\mathcal{A}\Pi^i(z) . \quad (211)$$

Note that the r.h.s. can also be interpreted as "$\frac{d\langle \mathcal{X}_t^i, \mathcal{X}_t^j\rangle}{dt}$ on $\{\mathcal{Z}_t = z\}$", by (42). Therefore, by application of (41) (cf. (26)):

$$\boxed{\lim_{h\to 0} h^{-1}\mathbb{C}\mathrm{ov}_t(\mathcal{X}_{t+h}^{h,i} - \mathcal{X}_t^{h,i}, \mathcal{X}_{t+h}^{h,j} - \mathcal{X}_t^{h,j}) = a_{i,j}(z) + \gamma(z)\overline{\delta_i\delta_j}(z)} \quad (212)$$

In summary, we have by (210) and (212) for any $z \in E$, on $\{\lim_{h\to 0} \mathcal{Z}_t^h = z\}$ (cf. (26)):

$$\boxed{\begin{array}{c}\lim_{h\to 0} h^{-1}\mathbb{E}_t(\mathcal{X}_{t+h}^h - \mathcal{X}_t^h) = b(z) \\ \lim_{h\to 0} h^{-1}\mathbb{V}\mathrm{ar}_t(\mathcal{X}_{t+h}^h - \mathcal{X}_t^h) = a(z) + \gamma(z)\overline{\delta\delta^\mathsf{T}}(z)\end{array}} \quad (213)$$

In the pure diffusion model $\mathcal{X}$ of section 7.2.6 (cf. formulas (50) to (52)), *Kushner's theorem* (see [121, 120]) states that the so-called *local consistency conditions* (213) (with $\gamma = 0$ therein), that is, the matching of the first and second conditional moments of the increments of an approximating Markov chain with those of a continuous-time limiting diffusion process with accuracy $\mathrm{o}(h)$, grants the convergence in law of the process $\mathcal{X}^h$ to $\mathcal{X}$. Hence the convergence of (optimally controlled) expectations of usual functionals follows (see also [121, page 129] for an extension to jump-diffusions).

21 Trees for vanilla options

21.1 Cox-Ross-Rubinstein Binomial Tree

The Cox-Ross-Rubinstein tree [59] may be seen as the approximation to the Black–Scholes model obtained by replacing the risk-neutral Black–Scholes dynamics with the following Markov chain: $S_0^h = S_0$, and for $i = 0, \dots, n-1$:

$$\boxed{S_{(i+1)h}^h = \begin{cases} u\, S_{ih}^h \text{ with probability } p \\ d\, S_{ih}^h \text{with probability } 1-p \end{cases}} \quad (214)$$

with $h = \frac{T}{n}$, where T is the time to maturity of an option (the current date is taken to be zero), n is the number of time steps, and

$$\boxed{u = e^{\sigma\sqrt{h}},\ d = e^{-\sigma\sqrt{h}},\ p = \tfrac{e^{\kappa h} - d}{u - d}} \quad (215)$$

We find it convenient to denote the time in the tree by i rather than ih. In particular, $S_i^h \equiv S_{ih}^h$, $\mathbb{E}_i$ refers to the conditional expectation with respect to the σ-algebra generated by $(S_0^h, \ldots, S_i^h)$, and $\mathcal{T}_i^h$ (with $\mathcal{T}_0^h = \mathcal{T}^h$) stands for the set of stopping times ν taking their values in $\{i, \ldots, n\}$.

The following Proposition shows that the Cox-Ross-Rubinstein tree model shares all the key properties of the Black–Scholes model.

Proposition 21.1 *p is the unique risk-neutral probability in the Cox–Ross–Rubinstein tree, in the sense that the unique replication price process of an European option with payoff $\phi\left(S_n^h\right)$ at time T in the Cox–Ross–Rubinstein tree is given by, for $i = 0, \ldots, n$:*

$$\boxed{\Pi_i^h = e^{-r(T-i)}\mathbb{E}_i\phi(S_n^h) = \Pi_i^h(S_i^h)} \tag{216}$$

with an associated replication strategy given as

$$\boxed{\zeta_i^h = \Delta_i^h(S_i) = \frac{\Pi_{i+1}^h(uS_i^h) - \Pi_{i+1}^h(dS_i^h)}{(u-d)S_i^h}} \tag{217}$$

If the option is American, the (unique) minimal super-hedging price is given by, for $i = 0, \ldots, n$:

$$\boxed{\widetilde{\Pi}_i^h = \max_{\nu \in \mathcal{T}_i^h} \mathbb{E}_i e^{-r(\nu-i)h}\phi(S_\nu^h) = \widetilde{\Pi}_i^h(S_i^h)} \tag{218}$$

with a related (minimal super-)hedging strategy defined as

$$\boxed{\widetilde{\zeta}_i^h = \widetilde{\Delta}_i^h(S_i^h) = \frac{\widetilde{\Pi}_{i+1}^h(uS_i^h) - \widetilde{\Pi}_{i+1}^h(dS_i^h)}{(u-d)S_i^h}} \tag{219}$$

Moreover, denoting here by S^h a generic space level in the tree, the Cox–Ross–Rubinstein pricing function of an European option introduced at the r.h.s. of (216) satisfies $\Pi_n^h(S^h) = \phi(S^h)$, and for $i = n-1, \ldots, 0$:

$$\boxed{\Pi_i^h(S^h) = e^{-rh}\left[p\Pi_{i+1}^h(uS^h) + (1-p)\Pi_{i+1}^h(dS^h)\right]} \tag{220}$$

whereas the pricing function $\widetilde{\Pi}^h$ of an American option introduced at the r.h.s. of (218) satisfies $\widetilde{\Pi}_n^h(S^h) = \phi(S^h)$, and for $i = n-1, \ldots, 0$:

$$\boxed{\widetilde{\Pi}_i^h(S^h) = \max\left(\phi(S^h), e^{-rh}\left[p\widetilde{\Pi}_{i+1}^h(uS^h) + (1-p)\widetilde{\Pi}_{i+1}^h(dS^h)\right]\right)} \tag{221}$$

Note that Proposition 21.1 actually holds irrespective of the precise definition of u and d, provided $d < u$ (with furthermore $1 \in [d, u]$, if one insists on having $p = \frac{1-d}{u-d} \in [0, 1]$; otherwise p only defines a signed probability measure).

Proof. See, for instance, Shreve [170, I] or Lamberton and Lapeyre [124]. □

Moreover, as we shall see now, the Cox-Ross-Rubinstein tree model 'converges' in some sense to the related Black–Scholes model as $h \to 0$. This convergence results by an application of Kushner's theorem where in (209) $\mathcal{X}$ may be taken as the Black–Scholes log-spot $\ln(S)$ and $\mathcal{X}^h$ as the piecewise constant and càdlàg continuous-time Markov chain interpolation of the Cox-Ross-Rubinstein Markov chain $(\ln S_i^h)_{0 \le i \le n}$ on the time intervals $[ih, (i+1)h)_{0 \le i \le n-1}$. However, in the simple situation of the Cox-Ross-Rubinstein model, convergence (at least, for European options) can be established directly as follows.

Convergence of the marginal law of S_T^h Assume $S_0 = 1$, without loss of generality. For $\lambda \in \mathbb{R}$, we have:

$$\begin{aligned}
\mathbb{E}\left[\exp\left(i\lambda \ln S_n^h\right)\right] &= \mathbb{E}\left[\exp\left(i\lambda \ln \prod_{l=0}^{n-1} \frac{S_{l+1}^h}{S_l^h}\right)\right] \\
&= \left(\mathbb{E}\left[\exp\left(i\lambda \ln S_1^h\right)\right]\right)^n \\
&= \left(p\exp\left(i\lambda\sigma\sqrt{h}\right) + (1-p)\exp\left(-i\lambda\sigma\sqrt{h}\right)\right)^n .
\end{aligned}$$

Since $p = \frac{e^{\kappa h}-d}{u-d} \sim \frac{1}{2} + \frac{b}{2\sigma}\sqrt{h} + O(h)$, we get as $h \to 0$:

$$\begin{aligned}
\mathbb{E}\left[\exp\left(i\lambda \ln S_n^h\right)\right] &\sim \left(1 + \left[i\lambda b - \lambda^2\frac{\sigma^2}{2}\right]\frac{T}{n}\right)^n \\
&\to \exp\left(\left[i\lambda b - \lambda^2\frac{\sigma^2}{2}\right]T\right) \\
&= \mathbb{E}\left[\exp\left(i\lambda\left(bT + \sigma W_T\right)\right)\right] \\
&= \mathbb{E}\left[\exp\left(i\lambda \ln(S_T)\right)\right]
\end{aligned}$$

where S denotes the risk-neutral Black–Scholes spot process. We thus have *convergence of the characteristic function* $\Phi_n^h(\lambda) = \mathbb{E}\left[\exp\left(i\lambda \ln S_n^h\right)\right]$ of the risk-neutral Cox–Ross–Rubinstein log-spot $\ln(S_n^h)$ to the characteristic function $\Phi_T(\lambda) = \mathbb{E}\left[\exp\left(i\lambda \ln(S_T)\right)\right]$ of the risk-neutral Black–Scholes log-spot $\ln(S_T)$, hence

$$S_n^h \xrightarrow{\mathcal{L}} S_T \text{ as } n \to \infty . \tag{222}$$

Remark 21.1 **(i)** The limit law depends only on $pe^{i\lambda\ln(u)} + (1-p)\,e^{i\lambda\ln(d)}$ through its Taylor expansion up to $o(h)$. Thus u, d or/and p can be altered as long as the involved terms of the development are not modified.
(ii) The upper and lower value of the spot at maturity are $u^n = e^{\sigma\sqrt{T}\sqrt{n}}$ and $d^n = e^{-\sigma\sqrt{T}\sqrt{n}}$ whereas the ratio of two successive points is $\frac{u}{d} = e^{2\sigma\sqrt{h}} = e^{2\sigma\sqrt{\frac{T}{n}}}$. Thus the scan of the law of S_n^h goes to $(0, +\infty)$ whilst the grid gets more and more dense. It is easy to show that the points visited by the process S^h become eventually dense in $[0,T] \times (0, +\infty)$ as $h \to 0$.

Convergence of Option Prices The convergence in law (222) grants the convergence of the price of European vanilla options with continuous and bounded payoffs, e.g., put options. The convergence of call prices follows by call–put parity, since the Cox–Ross–Rubinstein scheme satisfies the call-put parity relationship.

Convergence of Option Deltas Observe that (cf. (217))

$$\begin{aligned}
S_0\Delta_0^h(S_0) &= \frac{\Pi_1^h(uS_0) - \Pi_1^h(dS_0)}{(u-d)} \\
&= e^{-r(n-1)h}\frac{\mathbb{E}\left[\phi\left(uS_0X^h\right)\right] - \mathbb{E}\left[\phi\left(dS_0X^h\right)\right]}{u-d}
\end{aligned}$$

where X^h is a random variable independent from S_0. Thus

$$S_0\Delta_0^h(S_0) = e^{-r(n-1)h}\mathbb{E}\left[\frac{\phi\left(uS_0X^h\right) - \phi\left(dS_0X^h\right)}{u-d}\right]$$

Assume that ϕ is a function of class C^1. Then

$$\phi\left(uS_0X^h\right) - \phi\left(dS_0X^h\right) = \int_d^u S_0X^h\phi'\left(xS_0X^h\right)dx$$

so

$$\begin{aligned} S_0\Delta_0^h(S_0) &= e^{-r(n-1)h}\frac{1}{(u-d)}\int_d^u \mathbb{E}\left[S_0X^h\phi'\left(xS_0X^h\right)\right]dx \\ &= e^{-r(n-1)h}\mathbb{E}\left[S_0X^h\phi'\left(x^hS_0X^h\right)\right] \end{aligned}$$

for some $x^h \in [d,u]$, by the mean value property. Assuming further $y \mapsto \psi(y) = y\phi'(y)$ to be Lipschitz and bounded, then

$$\lim_{h\to 0}\mathbb{E}\left[S_0X^h\phi'\left(x^hS_0X^h\right)\right] = \lim_{h\to 0}\mathbb{E}\left[\psi\left(S_0X^h\right)\right] \tag{223}$$

where by the standard convergence in law result

$$\begin{aligned} \lim_{h\to 0}\mathbb{E}\left[\psi\left(S_0X^h\right)\right] &= \mathbb{E}\left[\psi\left(S_T\right)\right] \\ &= e^{rT}S_0\Delta^{bs}(0,S_0)\ . \end{aligned}$$

To sum-up, we showed that for bounded and sufficiently regular payoffs, the Cox–Ross–Rubinstein delta of an option $S_0\Delta_0^h(S_0)$ (cf. (217)) converges towards the Black–Scholes delta $\Delta^{bs}(0,S_0) = \partial_S\Pi^{bs}(0,S_0)$ as $h \to 0^+$. This results can be extended to the vanilla put payoff by density, and then to the vanilla call payoff by call–put parity.

Convergence of the prices and deltas still holds true for *American options*, though this cannot proven by elementary computations as in the European case.

21.1.1 Cox–Ross–Rubinstein Algorithm

Input parameters (Black–Scholes parameters,) StepNumber n

Output parameters Price, Delta

The Cox–Ross–Rubinstein algorithm is a backward computation of the option price, based on the dynamic programming equations (220) or (221), after a forward computation of the $n+1$ possible values of the underlying at maturity S_n^h. Since the Cox–Ross–Rubinstein tree is a flat tree, it is easily seen that the value of the underlying at time i and level j (indexing space levels from the bottom of the tree) is the same as that at time $i+2$ and level $j+1$. In particular there are only $2n+1$ possible values of the underlying between time 0 and time n.

For computational purposes, it is clever, in the American case, to compute the corresponding values of the payoff of the option only once at the beginning of the algorithm (storing them in an array for later use).

21.2 Other Binomial Trees

To achieve the convergence in law (222), many other choices of u, d and p may be done, regardless of any arbitrage or financial consideration: the tree algorithm becomes a purely numerical approximation algorithm, the only purpose being to get a good convergence to the limiting price and delta.

Note that a binomial tree is recombining as soon as u and d remain constant within the tree.

21.2.1 The Random Walk scheme

A natural scheme consists in approximating the Brownian motion W by the standard Random Walk in $S_T = S_0 \exp(bT + \sigma W_T)$. This leads to

$$u = e^{bh+\sigma\sqrt{h}},\ d = e^{bh-\sigma\sqrt{h}},\ p = \frac{1}{2}\ .$$

Convergence may be proved in the same way as before. Note that the discretized process is not a martingale.

21.2.2 The matching-three-moments scheme

An alternative route to convergence is Kushner's Theorem (see the introductory paragraph to this part). This leads to the idea of matching the mean and variance of the conditional laws of the approximating chain with those of the continuous risk-neutral Black–Scholes process. The *local consistency conditions* thus give the following equations in u, d, p :

$$\begin{aligned} pu + (1-p)\,d &= e^{\kappa h} \\ pu^2 + (1-p)\,d^2 - e^{2\kappa h} &= e^{2\kappa h}\left(e^{\sigma^2 h} - 1\right)\ . \end{aligned}$$

Since one degree of freedom remains, a natural idea is to match the third moment further, so

$$pu^3 + (1-p)\,d^3 = e^{3\kappa h} e^{3\sigma^2 h}\ .$$

The solution of this system is

$$\begin{aligned} u &= \frac{e^{\kappa h} Q}{2}\left[1 + Q + \sqrt{Q^2 + 2Q - 3}\right] \\ d &= \frac{e^{\kappa h} Q}{2}\left[1 + Q - \sqrt{Q^2 + 2Q - 3}\right] \\ p &= \frac{e^{\kappa h} - d}{u - d} \end{aligned}$$

with $Q = e^{\sigma^2 h}$. Note that $ud = e^{2\kappa h} Q^2 > 1$: this tree is not symmetric any more.

21.3 Trinomial trees

Along this line there is no need to remain stuck with the one node-two sons constraint. One may as well choose a three-points scheme, or a l-points scheme, or even a scheme involving a number of points depending on n (this is useful for other kinds of limiting continuous-time dynamics, like Lévy processes). From the previous computations it is easy to see that the points and probabilities of the chosen scheme should satisfy, for every $\lambda \in \mathbb{R}$:

$$\boxed{\sum p_j \exp(i\lambda \ln u_j) = 1 + \left[i\lambda b - \lambda^2 \tfrac{\sigma^2}{2}\right] h + o(h)}$$

in the sense that this condition ensures convergence of the characteristic functions, hence convergence of the marginal laws of S (cf. section 21.1). Actually we shall see in section 21.4.1 that these conditions are in fact tantamount to the local consistency conditions, which ensure a convergence of a much more general type, by Kushner's theorem.

Note that from a computational point of view, the condition that u_{j+1}/u_j be independent of j must be imposed in order to get a recombining tree.

A feature common to all more-than-binomial trees is that they give an accurate computation of the Greeks (delta, gamma, and theta) at time 0, by finite differences at time h.

21.3.1 The Kamrad–Ritchken tree

The Kamrad–Ritchken tree (see [114]) is the archetype of a trinomial tree. This is a flat tree with $2n+1$ possible values of the underlying S throughout the option's life. Kamrad and Ritchken take a symmetric 3-points approximation with space step k to $\ln\left(\frac{S_h^h}{S_0}\right)$ and match the related first two moments in the risk-neutral Black–Scholes model, so:

$$\begin{aligned} k\left(p_u - p_d\right) &= bh \\ k^2\left(p_u + p_d\right) - k^2\left(p_u - p_d\right)^2 &= \sigma^2 h \quad . \end{aligned}$$

Note that one may simplify the last equality, still maintaining an $o(h)$ matching of the variance (which is enough to ensure convergence of the characteristic functions), as

$$k^2\left(p_u + p_d\right) = \sigma^2 h \ .$$

This yields finally, in terms of a new parameter λ defined by $k = \lambda\sigma\sqrt{h}$:

$$\boxed{\begin{aligned} p_u &= \frac{1}{2\lambda^2} + \frac{b\sqrt{h}}{2\lambda\sigma} \\ p_f &= 1 - \frac{1}{\lambda^2} \\ p_d &= \frac{1}{2\lambda^2} - \frac{b\sqrt{h}}{2\lambda\sigma} \end{aligned}}$$

The parameter λ appears as a free parameter of the geometry of the tree. It is called the *stretch parameter*, and must satisfy $\lambda \geq 1$ to ensure stability of the algorithm. The value $\lambda = 1.22474$ which corresponds to $p_f = \frac{1}{3}$ is reported to be a good choice for an at-the-money call (or put).

Remark 21.2 When they are applied to the *discounted pricing function* $\widetilde{u}(t,x) = e^{-rt}u(t,x)$, the Kamrad–Ritchken tree and the finite differences explicit scheme of section 16.3.1 result in the same approximation scheme, provided one sets the domain half-size ℓ as $n\lambda\sigma\sqrt{h}$ for localizing the Black–Scholes equation (153). Otherwise said, the Kamrad–Ritchken tree and the explicit finite differences scheme for (153) are essentially the same scheme, except for the treatment of the discount factor (ru term in the equation). From the implementation point of view, the Kamrad–Ritchken tree is the better alternative of the two, unless the explicit finite differences scheme is coupled with an implicit scheme in the context of a more general theta-scheme for $\theta \neq 0, 1$ (like a Crank–Nicholson scheme). Indeed in this particular case the explicit scheme needs to be solved on the same (and on the whole) rectangular domain as the coupled implicit scheme.

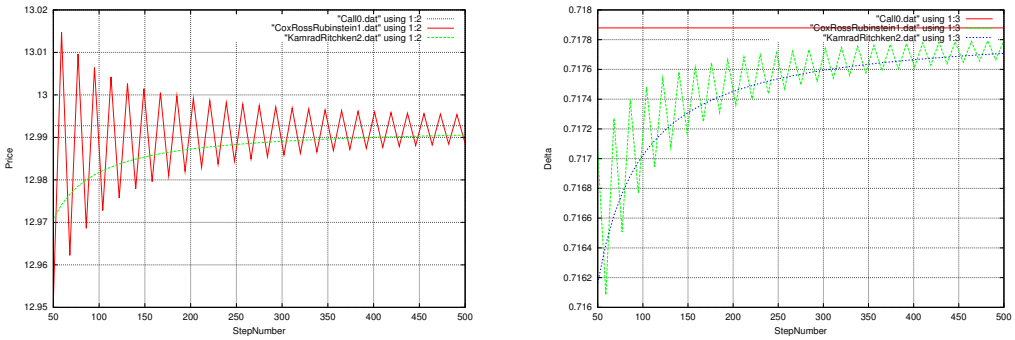

Figure 10: *European vanilla call priced by a Cox–Ross–Rubinstein binomial tree, resp. a Kamrad–Ritchken trinomial tree.*

21.3.2 Trinomial schemes with matching first two moments

Let $u > f > d$ be the possible values of $\frac{S_{i+1}^h}{S_i^h}$, with respective probabilities p_u, p_f, p_d. In order to get a recombining tree we impose $ud = f^2$. The two first moment matching conditions give

$$\begin{aligned} p_u u + p_f f + p_d d &= e^{\kappa h} \\ p_u u^2 + p_f f^2 + p_d d^2 &= e^{2\kappa h} Q \end{aligned}$$

where $Q = e^{\sigma^2 h}$. Since $p_u + p_f + p_d = 1$, thus two unknowns remain. The solution corresponding to the additional constraint

$$p_u = p_f = p_d = \frac{1}{3}$$

is

$$\begin{aligned} u &= V + \sqrt{V^2 - f^2} \\ d &= V - \sqrt{V^2 - f^2} \\ f &= \frac{e^{\kappa h}\,(3 - Q)}{2} \end{aligned}$$

with $V = \frac{e^{\kappa h}(3+Q)}{4}$.

21.4 Miscellaneous Remarks

21.4.1 Local consistency and convergence in law

In a general l-points scheme, let us come back to the equality ensuring *convergence in law*:

$$\sum p_j \exp\left(i\lambda \ln u_j\right) = 1 + \left[i\lambda b - \lambda^2 \frac{\sigma^2}{2}\right] h + \mathrm{o}(h)\,, \tag{224}$$

to be compared with the risk-neutral Black–Scholes *local consistency conditions*:

$$\begin{cases} \sum p_j u_j = \exp(\kappa h) + \mathrm{o}(h) \\ \sum p_j u_j^2 = \exp\left(\left(2\kappa + \sigma^2\right) h\right) + \mathrm{o}(h) \end{cases} \tag{225}$$

Proposition 21.2 *For a multinomial tree of the following form:*

$$\begin{cases} u_j &= 1 + u_{j,1}\sqrt{h} + u_{j,2}h + \mathrm{o}(h) \\ p_j &= p_{j,0} + p_{j,1}\sqrt{h} + p_{j,2}h + \mathrm{o}(h)\ , \end{cases} \tag{226}$$

convergence in law and local consistency are equivalent.

In view of Kushner's Theorem, the fact that local consistency implies convergence in law was expected; what this proposition really says is that for multinomial trees of the form (226), convergence in law implies a convergence of a much more general form.

Proof. Assuming (226), (224) obviously implies

$$\sum p_{j,0} = 1,\ \sum p_{j,1} = \sum p_{j,2} = 0\ .$$

We have

$$\begin{aligned} \exp(i\lambda \ln u_j) &= \exp\left(i\lambda\left(u_{j,1}\sqrt{h} + u_{j,2}h - \frac{u_{j,1}^2}{2}h\right) + \mathrm{o}(h)\right) \\ &= 1 + i\lambda u_{j,1}\sqrt{h} + \left(i\lambda\left(u_{j,2} - \frac{u_{j,1}^2}{2}\right) - \frac{\lambda^2}{2}u_{j,1}^2\right)h + o(h)\ , \end{aligned}$$

so that (224) is in fact equivalent to

$$\begin{aligned} \sum p_{j,0} i\lambda u_{j,1} &= 0 \\ \sum p_{j,0}(i\lambda(u_{j,2} - \frac{u_{j,1}^2}{2}) - \frac{\lambda^2}{2}u_{j,1}^2) + \sum p_{j,1} i\lambda u_{j,1} &= [i\lambda b - \lambda^2\frac{\sigma^2}{2}]\ . \end{aligned}$$

That is, since p and u are real-valued:

$$\begin{aligned} \sum p_{j,0} u_{j,1} &= 0 \\ \sum p_{j,0}\left(u_{j,2} - \frac{u_{j,1}^2}{2}\right) + \sum p_{j,1} u_{j,1} &= b \\ \sum p_{j,0} u_{j,1}^2 &= \sigma^2\ , \end{aligned}$$

or

$$\begin{cases} \sum p_{j,0} u_{j,1} &= 0 \\ \sum p_{j,0} u_{j,2} + \sum p_{j,1} u_{j,1} &= \kappa \\ \sum p_{j,0} u_{j,1}^2 &= \sigma^2\ . \end{cases} \tag{227}$$

Now, this may be rewritten:

$$\begin{aligned} \sum p_{j,0} u_{j,1} &= 0 \\ \sum p_{j,0} u_{j,2} + \sum p_{j,1} u_{j,1} &= \kappa \\ 2\sum p_{j,0} u_{j,1} &= 0 \\ \sum p_{j,0} u_{j,1}^2 + 2\sum p_{j,0} u_{j,2} + 2\sum p_{j,1} u_{j,1} &= \left(2\kappa + \sigma^2\right)\ , \end{aligned}$$

which is another form of the local consistency equations (225). □

21.4.2 Flat trees and American options

The generic multinomial recombining tree algorithm for pricing American options is the natural backward scheme $\Pi_n^h(S^h) = \phi(S^h)$, and for $i = n-1, \ldots, 0$:

$$\Pi_i^h(S^h) = \max\left(\phi\left(S^h\right),\, e^{-rh}\sum p_j \Pi_{i+1}^h(u_j S^h)\right)\ . \tag{228}$$

The algorithm requires the computation of the intrinsic value at each node of the tree. A computational advantage of a flat tree (like the Cox–Ross–Rubinstein or the Kamrad–Ritchken trees) is that one can compute once for all the intrinsic values across all the possible values of the underlying ($n+1$ or $2n+1$ values in the cases of the Cox–Ross–Rubinstein or the Kamrad–Ritchken tree) before performing the backward scheme. This reduces significantly the computational cost of the algorithm.

For pricing *time-dependent American options*, the natural way to modify the basic American tree algorithm is to replace $\phi(S^h)$ by $\phi(ih, S^h)$ in the backward formula (228), where $\phi\,(t, S)$ is the time-dependent payoff of the option. In this case the previous reduction is of course not applicable.

22 Trees for exotic options

22.1 Barrier options

Let us consider the simple case of a Down and Out Call with constant rebate R attached to the barrier L. The first idea to price this option within the Cox–Ross–Rubinstein scheme is to apply the usual backward induction scheme, with price level at or above the barrier constrained to R.
It is indeed possible to show that the resulting price converges to the right Black–Scholes limit. Nevertheless, the convergence is very slow compared with that for vanilla options. The reason is clear: let l denote the node index such that

$$S_0 d^l \geq L > S_0 d^{l+1}\ .$$

Then the algorithm, n being fixed, yields the same result for any value of the barrier between $S_0 d^l$ and $S_0 d^{l+1}$. Therefore the convergence cannot be faster than

$$\partial_L \Pi^{bs}\left(d^l - d^{l+1}\right) = \mathrm{O}(n^{-\frac{1}{2}})$$

(where Π^{bs} denotes the Black–Scholes price of the option), whereas the convergence of the tree for European vanilla options is known to be of the order of $\mathrm{O}(n^{-1})$.

An alternative method due to Ritchken [162] is to feed the algorithm with the right value of the barrier. The idea is to set the stretch parameter λ of a trinomial Kamrad–Ritchken tree (see section 21.3.1) such that the barrier is hit exactly. Recall that λ must be greater than one to ensure stability of the scheme. Intuitively, among the many possibilities for λ, the closer λ to one, the better. The natural way to choose λ is thus the following:

• compute the value of l above, i.e. $l = \left[\frac{\ln\left(\frac{S_0}{L}\right)}{\sigma\sqrt{h}}\right]$, where $[t]$ denotes the integer part of t (greater integer less than or equal to t);

• then set $\lambda = \frac{1}{l}\frac{\ln\left(\frac{S_0}{L}\right)}{\sigma\sqrt{h}}$.

Proceeding in this way, convergence is reported to be as fast as for vanilla options.

22.2 Bermudean Options

In the case of *Bermudean options*, the American right is in force only at a set of prescribed periods, for instance between a fixed date T_1 (included, say) and maturity T. In this case a natural algorithm is to apply the backward formula (228) between step $n-1$ and n_1 and the standard Cox–Ross–Rubinstein scheme before n_1, where $(n_1-1)\,h < T_1 \leq n_1 h$.

This first algorithm is very crude since it gives the same price, n being fixed, for any value of T_1 between $(n_1-1)\,h$ and $n_1 h$. A way to feed the algorithm with the right value is to use a Kamrad–Ritchken tree with a stretch parameter and number of steps λ_1 and n_1 between times 0 and T_1, and another stretch parameter and number of steps λ_2 and n_2 between T_1 and T. In order to get a recombining tree one imposes the following *pasting condition* at time T_1 :

$$\lambda_1\sqrt{\frac{T_1}{n_1}} = \lambda_2\sqrt{\frac{T-T_1}{n_2}}\,.$$

Recall that λ_1 and λ_2 must be greater than one. A possible way to choose the parameters is: first fix $\lambda_1 \geq 1$ (for instance $\lambda_1 = 1.2274$) and n_1, then set

$$n_2 = \left[\frac{n_1\,(T-T_1)}{T_1}\right] + 1\,.$$

So $n_2 T_1 \geq n_1\,(T-T_1)$, thus $\lambda_2^2 = \lambda_1^2 \frac{T_1}{n_1(T-T_1)}\,n_2 \geq \lambda_1^2 \geq 1$.

23 Bidimensional Trees

23.1 Cox-Ross-Rubinstein Tree for Lookback Options

We consider a Lookback Option with payoff function of the form $\phi(S,M)$, where $M_t = \sup_{0\leq s\leq t} S_s$. In a Cox–Ross–Rubinstein bi-dimensional tree model $(S_i^h, M_i^h)_{0\leq i\leq n}$, the related dynamic programming equation writes: $\Pi_n^h(S^h,M^h) = \phi(S^h,M^h)$, and for $i = n-1,\ldots,0$:

$$\boxed{\Pi_i^h(S^h,M^h) = e^{-rh}\left[p\Pi_{i+1}^h(uS^h,\max(uS^h,M^h)) + (1-p)\Pi_{i+1}^h(dS^h,M^h)\right]} \qquad (229)$$

Indeed since $d = e^{-\sigma\sqrt{h}} < 1$ thus $dS_{i+1}^h \leq M_i^h$ holds identically.

23.2 Kamrad–Ritchken Tree for Options on Two Assets

Assume $\left(S^1,S^2\right)$ follows the Black-Scholes dynamic under the risk-neutral probability, so for $l = 1,2$:

$$dS_t^l = \kappa^l S_t^l dt + \sigma_l S_t^l dW_t^l \qquad S_0^l = x_l$$

where (W^1, W^2) is a bidimensional Brownian Motion with correlation ρ. Setting $X = \ln(S)$, then:

$$\left\{ \begin{array}{l} dX_t^1 = b^1 dt + \sigma_1 dW_t^1 \\ dX_t^2 = b^2 dt + \sigma_2 dW_t^2 \end{array} \right. \quad \text{with} \quad \left(\begin{array}{c} b^1 \\ b^2 \end{array} \right) = \left(\begin{array}{c} \kappa^1 - \frac{\sigma_1^2}{2} \\ \kappa^2 - \frac{\sigma_2^2}{2} \end{array} \right) .$$

In order to approximate X we introduce the bidimensional Markov chain ξ_i^h such that

$$\left\{ \begin{array}{l} \left(\Delta \xi_i^h\right)_1 = \left(\kappa^1 - \frac{\sigma_1^2}{2}\right) h + \sigma_1 \sqrt{h} \varepsilon_i^1 \\ \left(\Delta \xi_i^h\right)_2 = \left(\kappa^2 - \frac{\sigma_2^2}{2}\right) h + \sigma_2 \sqrt{h} \varepsilon_i^2 \end{array} \right.$$

where $\left(\varepsilon_i^1, \varepsilon_i^2\right)$ is a sequence of iid r.v. with

$$\begin{array}{l} \mathbb{P}\left(\varepsilon_0^1 = 1, \varepsilon_0^2 = 1\right) = \mathbb{P}\left(\varepsilon_0^1 = -1, \varepsilon_0^2 = -1\right) = \frac{1+\rho}{4} \\ \mathbb{P}\left(\varepsilon_0^1 = 1, \varepsilon_0^2 = -1\right) = \mathbb{P}\left(\varepsilon_0^1 = -1, \varepsilon_0^2 = 1\right) = \frac{1-\rho}{4} \end{array}$$

Then one can show that

$$\lim_{h \to 0} h^{-1} \mathbb{E}_i \Delta \xi_i^h = \left(\begin{array}{c} b^1 \\ b^2 \end{array} \right) \quad , \quad \lim_{h \to 0} h^{-1} \mathbb{C}\text{ov}_i \Delta \xi_i^h = \left(\begin{array}{cc} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array} \right)$$

So Kushner's conditions hold and convergence is granted for the related vanilla options pricing tree algorithms.

Note that the corresponding algorithm is of complexity n^3.

Part VI
Monte Carlo Pricing Methods

This part is a short overview of Monte Carlo Methods in finance. More comprehensive references are given in Section 4. As deterministic methods, *stochastic simulation methods* (the terminology 'Monte Carlo' was introduced in [140]) can also be used in any Markovian model. In the case of European products, they consist of the classic *Monte Carlo (MC) loops*. For products with early exercise features, stochastic simulation methods are available too, as there are numerical methods, more generally so, for *Forward-Backward Stochastic Differential Equations* (methods based on dynamic programming and Monte Carlo computation of conditional expectations, see, e.g., [38] and references therein), but this is outside the scope of these notes.

Monte Carlo methods are attractive by their *genericity* (genericity of the related *theoretical properties*, that are insensitive to the dimension of the problem or to the regularity of the payoff function, and genericity of *implementation* as well), and by the *confidence interval* they provide on the solution (at least for *pseudo Monte Carlo* methods, as opposed to *quasi Monte Carlo* methods, as we shall see below).
But (pseudo) Monte Carlo methods are *slow* (convergence in $\sigma m^{-\frac{1}{2}}$, where m is the sample size and σ is the standard deviation of the sampled payoff), unless specific *variance reduction techniques* are applicable.
An alternative to variance reduction methods is to use *quasi Monte Carlo* methods, which converge faster in practice than (pseudo) Monte Carlo methods. Note however that quasi Monte Carlo methods don't furnish confidence intervals (unless sophisticated *randomized* forms of quasi Monte Carlo methods are used), and that the performances of quasi Monte Carlo algorithms are strongly altered in high dimension. Note that high dimension is to be understood here as (high *effective* dimension, as measured for instance by the number of significant risk dimensions in a PCA of the sampled payoff, as opposed to *nominal dimension*. In financial problems, effective dimension if often much less than nominal dimension, which explains the popularity and success of quasi Monte Carlo methods in finance).
Note finally that Monte Carlo methods are typically very easy to *parallelize*.

24 Random numbers

To sample 'random' vectors in $\mathbb{R}^d$ with specific distributions, one first draws sequences of 'uniform' random vectors u_j over $[0,1]^d$. Then one transforms the u_j into x_j with the required distribution.

'Uniform' random vectors u_j over $[0,1]^d$ may be obtained:
• either by sampling 'iid' random variables over $[0,1]$ with a (pseudo) random generator, and grouping them in buckets of size d,
• or by using quasi random generators (low-discrepancy sequences) in dimension d.

The quotes are here to indicate that simulated pseudo – or quasi – random numbers aim at *emulating* as well as possible the statistical properties of randomness and uniformity (and independence, in the case of pseudo-random numbers). Note however that since the simulated sequences are generated deterministically on a computer, we can always find a statistical test of randomness, uniformity or independence that the sequence will fail.

Of course the quality of a generator puts an upper bound on the quality of any simulation method using it.

25 Pseudo random generators

Pseudo-random generators are used to simulate independent uniform variables over $[0, 1]$ [131, 130, 129, 147, 145, 119].

Definition 25.1 (see L'Ecuyer [131, 130]) A *pseudo-random number generator* is a structure $\mathcal{G} = (s, S, T, U, G)$ where S is a finite set of states, $s \in S$ is the *initial state*, the mapping $T : S \to S$ is the *transition function*, U is a finite set of *outputs symbols*, and $G : S \to U$ is the *output function*.

Since S is finite, the sequence of states is ultimately periodic. The *period* is the smallest positive integer ν such that for any $j \geq$ some j_0, $s_{j+\nu} = s_j$.

25.1 Properties required for a good pseudo-random numbers generator

• *Large period length* At least 2^{60}.
• *Good equidistribution properties and statistical independence of successive pseudo-random draws* The generator should pass statistical tests for uniformity and independence [119, 130]: general tests like chi-square or Kolmogorov-Smirnov tests, and more specific tests like equidistribution test, serial test, gap test, partition test, etc.
• *Little intrinsic structure* Successive values produced by some generators produce a lattice structure in any given dimension.
• *Efficiency, fast generation algorithm, requiring not too much memory space* Especially if we use many generators together or in parallel.
• *Repeatability (fixing a given seed)* Very useful for practical applications. Otherwise one can use the current time (computer clock) to initialize the generators.
• *Portability* The generator should produce exactly the same sequence on different computers or with different compilers.
• *Unpredictability* It means that one cannot predict the next generated value by the algorithm from the previous ones (though this is less important in finance than for other applications like cryptography).

Note that from the point of view of period length the function `rand()` of the C++ standard library may behave very poorly. So `rand()` is a pseudo-random integral number in the range 0 to `RAND_MAX`. `RAND_MAX` is a constant defined in `<cstdlib>`. Its default value may vary between implementations but it is only granted to be at least 32767.

25.2 Constructing pseudo-random number generators

The simplest methods to construct random number generators are *linear methods*. Linear methods use a linear recurrence relation to compute the next value from the previous ones

Linear Congruential Generators (LCG) The j^{th} random number is given by

$$u_j = \frac{U_j}{N} \in [0, 1]$$

where

$$U_j = (aU_{j-1} + b) \bmod N$$

for fixed integers $a > 0$, b and $N > 0$. For instance (this is the choice of the Fortran IMSL Library):

$$a = 397204094\ ,\ b = 0 \text{ and } N = 2^{31} - 1\ .$$

Such generators are prone to produce a lot of regularity in sequences and a lattice structure.

Random numbers generator of L'Ecuyer with Bayes & Durham shuffling procedure This is a combination of two short periods LCG, resulting into a longer period generator with good properties.

25.3 Rejection method

The following lemma is, of course, elementary.

Lemma 25.1 *For every real $x \in (-1, 1)$,*

$$\sum\nolimits_{l\geq 0} x^l = \frac{1}{1-x}, \sum\nolimits_{l\geq 1} l x^{l-1} = \frac{1}{(1-x)^2}\ .$$

The rejection method allows one to draw pseudo-uniform points on "general" subsets of $\mathbb{R}^d$.

Proposition 25.2 *Suppose the U_j are iid uniform random variables over a subset $\mathcal{D}$ of $\mathbb{R}^d$. So U_j has density $\frac{\mathbb{1}_{\mathcal{D}}}{\lambda(\mathcal{D})}$ where λ denotes the Lebesgue measure over $\mathbb{R}^d$. Let $\frac{\lambda(D)}{\lambda(\mathcal{D})} = \alpha \in (0,1)$ for some $D \subseteq \mathcal{D}$. Denoting $\nu_1 = \inf\{l > 0; U_l \in D\}$ and for $j \geq 1$,*

$$\nu_{j+1} = \inf\{l > \nu_j; U_l \in D\}\ ,\ V_j = U_{\nu_j}\ .$$

Then

$$\mathbb{E}(\nu_{j+1} - \nu_j) = \tfrac{1}{\alpha} = \mathbb{E}\nu_1$$

and the V_j define iid uniform random variables over D.

Proof. We have

$$\mathbb{P}(\nu_1 = l) = (1-\alpha)^{l-1}\alpha\ ,$$

so

$$\mathbb{E}\nu_1 = \sum\nolimits_{l\geq 1} l\mathbb{P}(\nu_1 = l) = \sum\nolimits_{l\geq 1} l(1-\alpha)^{l-1}\alpha = \frac{1}{\alpha}\ .$$

Moreover, for any $\Delta \subseteq D$:

$$P(U_{\nu_1} \in \Delta) = \sum\nolimits_{l\geq 1} P(\nu_1 = l, U_l \in \Delta) = \sum\nolimits_{l\geq 1} (1-\alpha)^{l-1}\frac{\lambda(\Delta)}{\lambda(\mathcal{D})} = \frac{1}{\alpha}\frac{\lambda(D)}{\lambda(\mathcal{D})}\frac{\lambda(\Delta)}{\lambda(D)} = \frac{\lambda(\Delta)}{\lambda(D)}$$

and for any $(\Delta_j)_{1\leq j\leq m}$ with $\Delta_j \subseteq D$ for every $j = 1 \ldots m$:

$$\begin{aligned}
&\mathbb{P}\left(U_{\nu_j} \in \Delta_j, j = 1\ldots m\right)\\
&= \textstyle\sum_{j\geq 1, l\geq m-1} \mathbb{P}\left(U_{\nu_1} \in \Delta_1, \ldots, U_{\nu_{m-1}} \in \Delta_{\nu_{m-1}}, \nu_{m-1} = l, \nu_m = l+j, U_{l+j} \in \Delta_m\right)\\
&= \textstyle\sum_{l\geq m-1} \mathbb{P}\left(U_{\nu_1} \in \Delta_1, \ldots, U_{\nu_{m-1}} \in \Delta_{\nu_{m-1}}, \nu_{m-1} = l\right) \sum_{j\geq 1} (1-\alpha)^{j-1}\frac{\lambda(\Delta_m)}{\lambda(\mathcal{D})}\\
&= \left(\textstyle\sum_{l\geq m-1} \mathbb{P}\left(U_{\nu_j} \in \Delta_j, \ldots, U_{\nu_{m-1}} \in \Delta_{\nu_{m-1}}, \nu_{m-1} = l\right)\right)\frac{\lambda(\Delta_m)}{\lambda(D)} = \mathbb{P}\left(U_{\nu_j} \in \Delta_j, \ldots, U_{\nu_{m-1}} \in \Delta_{\nu_{m-1}}\right)\frac{\lambda(\Delta_m)}{\lambda(D)}
\end{aligned}$$

By induction, there comes

$$P\left(U_{\nu_j} \in \Delta_j, j = 1 \dots m\right) = \prod_{j=1}^{m} \frac{\lambda\left(\Delta_j\right)}{\lambda\left(D\right)}$$

so

$$P\left(V_j \in dv_j, j = 1 \dots m\right) = \prod_{j=1}^{m} \frac{1_D\left(v_j\right) dv_j}{\lambda\left(D\right)} \ .$$

□

Note in particular that the *average rejection time* equals $\frac{1}{\alpha}$, which goes to 1 (resp. ∞) as $\alpha \to 1$ (resp. 0).

26 Low-discrepancy sequences

Quasi-random numbers, or successive values of *low-discrepancy sequences* [148, 147, 146, 142], are not independent. But they have good equidistribution properties on $[0,1]^d$, implying good convergence properties of $\frac{1}{m}\sum_{j=1}^{m} \psi(u_j)$ to $\int_{[0,1]^d} \psi(u)du$ as $m \to \infty$.

Let $[|0,x|]$ stand for $\{y \in [0,1]^d \ , \ y \leq x\}$, with $y \leq x$ if and only if $y_j \leq x_j$, $j = 1 \dots d$. Given a $[0,1]^d$-valued sequence $u = (u_j)_{j \geq 1}$, and $x \in [0,1]^d$, we denote

$$\boxed{D_m(u,x) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{[|0,x|]}(u_j) - \mathrm{Vol}([|0,x|])}$$

with $\mathrm{Vol}([|0,x|]) = \prod_1^d x_l$.

Definition 26.1 • A sequence $(u_j)_j \geq 1$ is said to be *equidistributed* on $[0,1]^d$ if

$$\forall x \in [0,1]^d, \lim_{m \to \infty} D_m(u,x) = 0 \ .$$

• The value $D_m(u)$ defined by

$$D_m(u) = \sup_{x \in [0,1]^d} |D_m(u,x)|$$

is called the *(star) discrepancy* for the first m terms of the u sequence.
• A sequence u is said to be a *low-discrepancy sequence* if (the terminology is not completely fixed yet) if its discrepancy satisfies

$$\boxed{D_m = \mathrm{O}(\frac{(\ln m)^d}{m})}$$

or if it is asymptotically better than $\mathrm{O}((\frac{\ln \ln m}{m})^{\frac{1}{2}})$, the discrepancy of a sequence of independent uniform variables (by the law of the iterated logarithm).

26.1 General Remarks on low discrepancy sequences

The discrepancy measures how a given set of points is distributed in $[0,1]^d$. It can be viewed as a quantitative measure for the deviation from the uniform distribution.

Quasi-random numbers combine the advantage of a lattice with the advantage that points can be added incrementally (like points of a random sequence).

But for large dimension d, the theoretical bound $\frac{(\ln m)^d}{m}$ may be meaningful for extremely large values of m only.
Low discrepancy sequences perform very well in low dimension. In high dimension d, a lattice can only be refined by increasing the number of points by a factor 2^d.

Orthogonal projections If a d-dimensional sequence is uniformly distributed in $[0,1]^d$, then two-dimensional sequences formed by pairing coordinates should also be uniformly distributed. The appearance of non-uniformity in such projections (see Figure 11) is an indication of potential problems in using a quasi-random sequence [142].

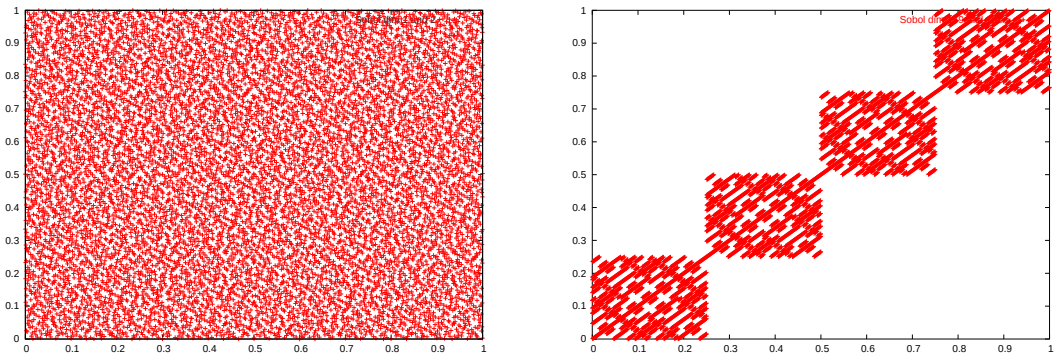

Figure 11: *Orthogonal projection on the first, resp. last two coordinates of the first* 10^4 *points of the Sobol sequence in dimension 160.*

26.2 Sobol sequences

The *Sobol sequence* is one of the most used sequences for Quasi Monte Carlo simulation, and one of the most successful for financial applications. Its construction is based on primitive polynomials in the field $\mathbb{Z}_2$ and its implementation relies on the use of bitwise XOR (Exclusive Or) operations (see [94, 168]).

27 Simulation of non-uniform random variables or vectors

27.1 Inverse method

The inverse method allows one to simulate a random variable with known (*computable*) (generalized) inverse cumulative distribution function F^{-1}, as follows:
• First, simulate a variable u uniformly distributed on $[0,1]$, using a pseudo-random numbers generator or a quasi-random number generator;

• Then set $x = F^{-1}(u)$.

Example 27.1 To simulate an exponential random variable with parameter γ, draw $\mathcal{E}_\gamma = -\frac{1}{\gamma}\ln(1-U)$ (or $-\frac{1}{\gamma}\ln U$), with U uniform.

Example 27.2 A Poisson random variable with parameter μ is such that $\mathbb{P}(\mathcal{P}_\mu = n) = \exp(-\mu)\frac{\mu^n}{n!}$. So, to simulate $\mathcal{P}_\mu$, draw a uniform number U on the unit segment, and set $\mathcal{P}_\mu = \nu$ such that

$$\sum_{n=0}^{\nu} \exp(-\mu)\frac{\mu^n}{n!} \leq U < \sum_{n=0}^{\nu+1} \exp(-\mu)\frac{\mu^n}{n!} .$$

Recall that the number of clients at time t in a queue with exponential inter-arrival times of parameter γ follows a Poisson distribution with parameter $\mu = \gamma t$. So, to simulate $\mathcal{P}_\mu$, another possibility is to draw iid exponential random variables $\mathcal{E}_\mu^j = -\frac{1}{\mu}\ln U^j$ until $\sum_{l=0}^{j}\mathcal{E}_\mu^l > 1$ and to set

$$\mathcal{P}_\mu = \sup\{j \in \mathbb{N}; \sum_{l=0}^{j}\mathcal{E}_\mu^l \leq 1\} = \sup\{j \in \mathbb{N}; \prod_{l=0}^{j} U^l \geq e^{-\mu}\} .$$

In the context of simulation by the inverse method, the following result is useful.

Lemma 27.1 (Change of variables formula for densities) *Let ϕ denote a diffeomorphism between subsets of $\mathbb{R}^d$. The density of $Y \equiv \phi \circ X$ at $y = \phi(x)$ is given in terms of that of X at x (assumed to exist) as*

$$p_Y(y) = \left|\det\left(J\phi^{-1}\right)(y)\right| p_X(x) , \tag{230}$$

where $J\phi^{-1}$ stands for the Jacobian matrix $(\partial_{y_j}\phi_i^{-1})_{1\leq i,j\leq d}$.

Proof (sketched). Perform the change of variables $Y \equiv \phi \circ X$ in the integral expression of $\mathbb{E}f(Y)$ with respect to p_Y, where f denotes a generic bounded Borel-measurable real-valued function on $\mathbb{R}^d$. □

Example 27.3 (Log-Normal density) *For $Y = S_t$ in Black–Scholes, $X = \sigma W_t$ and*

$$\phi^{-1} = x(S) = \ln(\frac{S}{S_0}) - bt,$$

we get

$$p_{S_t}(S) = \frac{1}{\sigma S\sqrt{2\pi t}} e^{-\frac{x(S)^2}{2\sigma^2 t}} = \frac{1}{S} p_{\sigma W_t}[x(S)] .$$

27.2 Simulation of Gaussian variables

One direct method to generate Gaussian variables is the *Box-Müller transformation.*

Box-Müller transformation If (U, V) is uniformly distributed on $[0,1]^2$, then the pair (X, Y) defined by

$$\boxed{\begin{array}{l} X = \sqrt{-2\ln U}\sin(2\pi V) \\ Y = \sqrt{-2\ln U}\cos(2\pi V) \end{array}}$$

is standard Gaussian.

Proof. Let θ be uniform over $[0, 2\pi]$ and ρ^2 be exponential with parameter $\frac{1}{2}$, independent from θ. Then the pair $(X, Y) = (\rho\cos\theta, \rho\sin\theta)$ is standard Gaussian. Indeed, for every measurable and bounded function ϕ, we have:

$$\begin{aligned}\mathbb{E}\phi(X,Y) &= \int_0^\infty \int_0^{2\pi} \phi(\rho\cos\theta, \rho\sin\theta)\frac{1}{2}e^{\frac{-\rho^2}{2}} d(\rho^2)\frac{d\theta}{2\pi} \\ &= \int_0^\infty \int_0^{2\pi} \phi(\rho\cos\theta, \rho\sin\theta)\rho d\rho e^{\frac{-\rho^2}{2}}\frac{d\theta}{2\pi} \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \phi(x,y)e^{-\frac{x^2+y^2}{2}}\frac{dxdy}{2\pi}\ .\end{aligned}$$

□

Note that this method requires two *independent* random values to obtain two Gaussian variables. So *it must not be used when the random numbers u and v are generated from two successive values of a one-dimensional low-discrepancy sequence*, because u and v are not independent in this case. To use the Box-Müller transformation with quasi-random numbers, one should thus take special care to generate u and v independently, e.g., from two different one-dimensional sequences, or from a two-dimensional sequence.

Simulating Gaussian variables by the rejection (or *polar*) method Simulate by the rejection method a uniform point (U, V) on the unit disk. Then the pair (X, Y) defined by

$$\boxed{\begin{aligned} X &= \sqrt{\tfrac{-2\log(\rho^2)}{\rho^2}}U \\ Y &= \sqrt{\tfrac{-2\log(\rho^2)}{\rho^2}}V \end{aligned}}$$

where $\rho^2 = U^2 + V^2$, is a standard Gaussian pair.

Proof. Let (ρ, θ) denote the polar coordinates of (U, V). Then $\left(\rho^2, \frac{\theta}{2\pi}\right)$ is uniformly distributed over the square $(0,1)^2$, by the densities change of variables formula (230) (at the intuitive level, it is clear that ρ is uniformly distributed over the unit disk and that ρ and θ are independent. Moreover, we have $d\mathbb{P}_{U,V}(r,\alpha) = \frac{drrd\alpha}{\pi}$, hence $d\mathbb{P}_\rho(r) = 2rdr = d\left(r^2\right) = d\mathbb{P}_{\rho^2}(r^2)$). As $\left(\rho^2, \frac{\theta}{2\pi}\right)$ is uniformly distributed over the square $(0,1)^2$, we deduce by Box-Müller that

$$\sqrt{-2\log\left(\rho^2\right)}\begin{pmatrix}\cos\left(2\pi\frac{\theta}{2\pi}\right)\\ sin\left(2\pi\frac{\theta}{2\pi}\right)\end{pmatrix} = \sqrt{\frac{-2\log\left(\rho^2\right)}{\rho^2}}\begin{pmatrix}\rho\cos\left(\theta\right)\\ \rho sin\left(\theta\right)\end{pmatrix} = \sqrt{\frac{-2\log\left(\rho^2\right)}{\rho^2}}\begin{pmatrix}U\\ V\end{pmatrix}$$

is Gaussian standard. □

The interest of the polar method is to avoid resorting to trigonometric functions. The disadvantage is that one must proceed by rejection to simulate a uniform point (U, V) on the unit disk, with an average rejection time of $\frac{\pi}{4}$ draw.

Simulating Gaussian variables by the inverse method The inverse Gaussian cumulative distribution function $\mathcal{N}^{-1}$ is not known in closed form. To use the inverse method for simulating Gaussian variables, we thus need an approximation of $\mathcal{N}^{-1}$. *Moro's algorithm* furnishes a very good and quick approximation.

27.3 Simulation of Gaussian vectors

Simulating a d-dimensional Gaussian vector V with zero mean and covariance matrix Γ can be done as follows:
(i) Compute a square root of Γ, namely a matrix Σ such that

$$\boxed{\Gamma = \Sigma\Sigma^{\mathsf{T}}}$$

(ii) Generate d independent standard Gaussian variables ε_i;
(iii) Compute $V = \Sigma G$, with $G = (\varepsilon_1, \ldots, \varepsilon_d)$;
So $V \hookrightarrow \mathcal{N}(0, \Gamma)$.

For Σ in (i), one may take the *lower triangular matrix* obtained by *Cholesky decomposition* of Γ, so for $p = 1 \ldots d$:

$$\begin{array}{l} \Sigma_{pp} = \sqrt{\Gamma_{pp} - \sum_{r=1}^{p-1} \Sigma_{pr}^2} \\ \Sigma_{qp} = \frac{\Gamma_{pq} - \sum_{r=1}^{p-1} \Sigma_{pr}\Sigma_{qr}}{\Sigma_{pp}} \qquad \text{for} \quad q = p+1, \ldots, d \, ; \end{array}$$

An alternative is to proceed to the *Principal Component Analysis* (PCA) of Γ, setting $\Sigma = P\Lambda^{\frac{1}{2}}$, where $P\Lambda P^{\mathsf{T}} = \Gamma$ with P orthonormal and Λ diagonal is the *singular decomposition* of Γ. A variant of this method consists in computing the singular decomposition $Q\Upsilon Q^{\mathsf{T}}$ of the related *correlation matrix* $\Omega = Diag(\sigma_i)^{-1}\Gamma Diag(\sigma_i)^{-1}$, setting $\Sigma = Diag(\sigma_i)Q\Upsilon^{\frac{1}{2}}$.

For instance, denoting in the *two-dimensional case*:

$$\Gamma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \, , \text{ so } \Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \, ,$$

the Cholesky decomposition of Γ yields

$$\boxed{\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix}}$$

whereas the singular decomposition of Ω yields (up to a reordering of the PCA factors)

$$\Sigma = \begin{pmatrix} \sigma_1\sqrt{\frac{1+\rho}{2}} & \sigma_1\sqrt{\frac{1-\rho}{2}} \\ \sigma_2\sqrt{\frac{1+\rho}{2}} & -\sigma_2\sqrt{\frac{1-\rho}{2}} \end{pmatrix}$$

28 Principle of the Monte Carlo Simulation

We want to estimate the following parameter Θ:

$$\Theta = \mathbb{E}[\phi(X)]$$

where ϕ is some function on $\mathcal{D} \subseteq \mathcal{R}^d$ and X is a $\mathcal{D}$-valued random vector. Note that Θ can be expressed as the integral

$$\Theta = \int_{\mathcal{D}} \phi(x) d\mathbb{P}^X(x) \, .$$

(Pseudo) Monte Carlo simulation is a general method for evaluating an integral as an expected value, based on the *Strong Law of Large Numbers* and the *Central Limit Theorem*. It provides an unbiased estimator and the error on the estimate is controlled within a confidence interval.

28.1 Limit theorems

Strong Law of Large Numbers For x_j, X iid with $\mathbb{E}[|\phi(X)|] < \infty$, then

$$\boxed{\frac{1}{m}\sum_{j=1}^{m}\phi(x_j) \xrightarrow{a.s.} \mathbb{E}[\phi(X)] \text{ as } m \to \infty}$$

Central Limit Theorem If, moreover, $\sigma^2 = \mathbb{V}\text{ar}[\phi(X)] < \infty$, the normalized error converges in law to the Gaussian distribution:

$$\boxed{\frac{\sqrt{m}}{\sigma}\left(\frac{1}{m}\sum_{j=1}^{m}\phi(x_j) - \mathbb{E}[\phi(X)]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ as } m \to \infty}$$

28.2 Estimation principle

• An unbiased estimator with m trials for Θ is thus given by

$$\boxed{\Theta_m = \frac{1}{m}\sum_{j=1}^{m}\phi(x_j)}$$

• The variance of this estimator is given by

$$\widetilde{\sigma}_m^2 = \frac{\sigma^2}{m},$$

with unbiased estimator

$$\boxed{\sigma_m^2 = \frac{1}{m-1}\left[\frac{1}{m}\sum_{j=1}^{m}\phi^2(x_j) - \Theta_m^2\right]}$$

This variance of the estimator thus decreases to 0 as $m \to \infty$. It means that the greater m, the more accurate the estimator. The speed of convergence of Θ_m to Θ is $\frac{\sigma}{\sqrt{m}}$, independently of the dimension d.

• A confidence interval $IC = [A, B]$ at the *threshold* (*confidence level*) $1 - 2\alpha$ for Θ is built as

$$\boxed{IC = [\Theta_m - z_\alpha\sigma_m; \Theta_m + z_\alpha\sigma_m]}$$

with $z_\alpha = \mathcal{N}^{-1}(1-\alpha)$. So $\mathbb{P}(A < \Theta < B) = 1 - 2\alpha$. For instance, if the threshold is set at 95%, then $\alpha = 2.5\%$ and $z_\alpha \approx 1.96$.

A natural *stopping criterion* consists in breaking the Monte Carlo loop when $z_\alpha \sigma_m$ becomes less than some proportion (e.g. $10bp = 10^{-3}$) of Θ_m, so that one knows the price with a 10bp relative error, with confidence $1-\alpha$.

28.3 Properties

We briefly summarize some advantages and disadvantages of the Monte Carlo method.
• *Advantages:*
– One can implement this method very easily if one is able to simulate the variable X;
– It does not require regularity or differentiability properties of the function ϕ;
– The estimator is unbiased;
– The error on the estimate can be controlled by the Central Limit Theorem, and one can build a confidence interval.
• *Disadvantages:* One has to realize a lot of simulations to obtain an accurate estimator. Therefore computing time can be very high.

29 Variance Reduction Techniques

The main disadvantage of the standard Monte Carlo Simulation is its =computing time, which is proportional to $\frac{\sigma}{\sqrt{m}}$. So to increase the accuracy by a factor 10, one must increase the number m of simulations by a factor 100.
A better option (this is the topic of *Variance reduction techniques*, or *Accelerated Monte Carlo*) is to rewrite Θ in terms of a new random variable with variance smaller than $\phi(X)$.

29.1 Antithetic Variables

The principle of antithetic variables is to introduce some correlation between the terms of the estimate. When simulation is done by the inverse method, we use uniform numbers u_j on $[0,1]$. In the antithetic variables method, we use each u_j twice, as u_j and $1-u_j$. Note that these two variables have the same law, but they are not independent. Let us denote by x_j and $\bar{x}_j$ the variables respectively generated from u_j and $1-u_j$. An unbiased estimator of Θ with m trials is defined by

$$\bar{\Theta}_m = \frac{1}{2m} \sum_{j=1}^{m} \left(\phi(x_j) + \phi(\bar{x}_j) \right)$$

with x_j, X iid. The variance of the estimator is given by

$$\bar{\sigma}_m^2 = \frac{1}{2m} \left(\mathbb{V}\mathrm{ar}[\phi(X)] + \mathbb{C}\mathrm{ov}(\phi(X), \phi(\bar{X})) \right) ,$$

with $\bar{X} = 1 - X$. The following result gives a simple condition ensuring variance reduction with this method.

Proposition 29.1 *If ϕ is a monotone function, then $\bar{\sigma}_m^2 \leq \frac{1}{2}\widetilde{\sigma}_m^2$.*

Proof. Introducing an independent copy V of U, we have by monotonicity of $\varphi = \phi \circ F_X^{-1}$:

$$\left(\varphi(U) - \varphi(V)\right)\left(\varphi(1-U) - \varphi(1-V)\right) \leq 0 ,$$

identically. Taking expectations, there comes:

$$\mathbb{E}\left[\varphi(U)\varphi(1-U) + \varphi(V)\varphi(1-V)\right] \leq \mathbb{E}\left[\varphi(U)\varphi(1-V) + \varphi(V)\varphi(1-U)\right] .$$

Using the fact that U and V are independent and uniformly distributed over $[0,1]$, it follows that $\mathbb{C}\text{ov}(\phi(X),\phi(\bar{X})) = \mathbb{C}\text{ov}(\varphi(U),\varphi(1-U)) \leq 0$. □

Note that the estimator $\bar{\Theta}_m$ is based on $2m$ terms, instead of m terms for the standard Monte Carlo estimator Θ_m, and that variance reduction by a factor $\frac{1}{2}$ is also ensured by using the standard Monte Carlo estimate based on $2m$ simulations instead of m.

29.2 Control Variables

The principle of this method is to introduce another model for which we have an explicit solution and to estimate the difference between the original parameter Θ and the new one, decomposing

$$\boxed{\Theta = \mathbb{E}[\phi(X)] = \mathbb{E}[\phi(X) - \psi(X)] + \mathbb{E}[\psi(X)]}$$

where ψ is a function such that $\mathbb{E}[\psi(X)] = c$ is known.

An unbiased estimator with m trials of Θ is defined by

$$\boxed{\widehat{\Theta}_m = \frac{1}{m}\sum_{j=1}^{m}(\phi(x_j) - \psi(x_j)) + c}$$

with x_j, X iid. The variance of the estimator is given by

$$\begin{aligned}\widehat{\sigma}_m^2 &= \frac{1}{m}\mathbb{V}\text{ar}[\phi(X) - \psi(X)] \\ &= \frac{1}{m}\left[\mathbb{V}\text{ar}[\phi(X)] + \mathbb{V}\text{ar}[\psi(X)] - 2\mathbb{C}\text{ov}(\phi(X),\psi(X))\right] .\end{aligned}$$

Variance reduction is not guaranteed with regard to standard Monte Carlo simulation, unless the functions ϕ and ψ have a large positive correlation. This supposes an appropriate choice for the control variate ψ.

29.3 Importance Sampling

The basic idea of importance sampling is to concentrate the distribution of the sample points in the most contributive parts of the space. To this end we introduce a new $\mathcal{D}$-valued random vector Y. Assume that the law of X and Y have respective densities μ and ρ, with $\text{Supp}(\phi\mu) \subset \text{Supp}(\rho)$. Denoting $\varphi = \mathbb{1}_{\phi\mu\neq 0}\frac{\phi\mu}{\rho}$, there comes:

$$\int_{\mathcal{D}} \phi(z) d\mathbb{P}^X(z) = \int_{\mathcal{D}} \phi(z)\frac{\mu(z)}{\rho(z)} d\mathbb{P}^Y(z)$$

so

$$\boxed{\Theta = \mathbb{E}[\phi(X)] = \mathbb{E}[\varphi(Y)]}$$

This yields the following estimator for Θ :

$$\boxed{\widehat{\Theta}_m = \frac{1}{m}\sum_{j=1}^{m}\varphi(y_j)}$$

with y_j, Y iid. The variance of the estimator $\widehat{\Theta}_m$ is given by

$$\widehat{\sigma}_m^2 = \frac{1}{m}\mathbb{V}\text{ar}\left[\varphi(Y)\right] .$$

Variance reduction with regard to standard Monte Carlo simulation is not guaranteed. It depends on the choice of the *importance function* ρ. The minimum variance is reached for the following importance function ρ^* :

$$\boxed{\rho^* = \frac{|\phi(x)|\mu(x)}{\int |\phi(y)|\mu(y)dy}} \tag{231}$$

Usually ρ^* is unknown (note that Θ sits in the denominator of the r.h.s. of (231), for $\phi > 0$). In practice, one chooses ρ by using formula (231) applied to an approximation of the payoff ϕ such that the corresponding optimal density is computable.

29.4 Efficiency of the Monte Carlo methods

We now introduce a criterion to compare the efficiency of the various simulation methods: standard simulation or simulation with variance reduction techniques. This criterion takes into account the computing time required by the simulation for each method. Efficiency of the method q with regard to the method p is defined by:

$$\varepsilon(p,q) = \lim_{m_p, m_q \to \infty} \frac{\sigma_{m_p}(p)}{\sigma_{m_q}(q)} \sqrt{\frac{t_{m_p}(p)}{t_{m_q}(q)}} \ .$$

The method q is considered to be more efficient than the method p if $\varepsilon(p,q) \geq 1$. For instance, $\varepsilon(p,q) = 3$ means that method p requires 9 times more time to obtain the same accuracy (as measured by the standard error) than method q, or, equivalently, that for a same computing time the standard error is three times smaller for method q than for method p.

Note that assuming the computing time proportional to the sample size, so $t_{m_p}(p) = k_p m_p$ where k_p is a factor which expresses the complexity of the algorithm for method p, then

$$\varepsilon(p,q) \approx \frac{\sigma_p}{\sigma_q} \sqrt{\frac{k_p}{k_q}} \ .$$

The previous efficiency criterion is thus essentially independent on the sample size.

30 Quasi Monte Carlo Simulation

Defining $\psi = \phi \circ F_X^{-1}$, then $\phi(X) \overset{(law)}{=} \psi(U)$, so $\Theta = E[\phi(X)] = E[\psi(U)]$. Quasi Monte Carlo simulation methods consist in estimating

$$\mathbb{E}[\psi(U)] = \int_{[0,1]^d} \psi(u)du$$

by $\frac{1}{m}\sum_{j=1}^m \psi(\xi_j)$, where ξ is a d-dimensional low-discrepancy sequence.

30.1 Koksma-Hlawka inequality

Theorem 30.1 *For any $\xi_1, \ldots, \xi_m \in [0,1]^d$, we have:*

$$\left| \frac{1}{m}\sum_{j=1}^m \psi(\xi_j) - \int_{[0,1]^d} \psi(x)dx \right| \leq V(\psi) D_m(\xi) \ , \ m \geq 1$$

where $V(\psi)$ denotes the Hardy-Krause variation *of ψ.*

The definition of the Hardy-Krause variation of ψ is rather technical and irrelevant for our purposes. Simply note that in dimension one, the Hardy-Krause variation of a function coincides with the usual notion of variation of a function, e.g., $\int_{[0,1]} \mid \partial_u \psi \mid du$, for ψ of class $\mathcal{C}^1$. Through the Koksma-Hlawka inequality, we understand the interest to have sequences with discrepancy D_m as small as possible.

The Koksma-Hlawka inequality gives an a priori deterministic bound for the error in the approximation of $\int_{[0,1]^d} \psi(x)dx$ by $\frac{1}{m}\sum_{j=1}^m \psi(\xi_j)$. This error is expressed in term of the discrepancy of the sequence and the variation of the function ψ. But it is often difficult to calculate or even to estimate the variation of ψ. Moreover, since for large dimensions d the asymptotic bound $\frac{(\ln m)^d}{m}$ of a low-discrepancy sequence may be meaningful for extremly large values of m only, and because $\frac{(\ln m)^d}{m}$ increases exponentially with d, thus the bound in Koksma-Hlawka inequality typically gives no relevant information for realistic sample sizes m.

Moreover, unlike the (pseudo) Monte Carlo method, the Quasi Monte Carlo approach does not provide a confidence interval for the estimator. The empirical variance of the sample is not meaningful because successive terms of the sequence are not independent. This is due to the construction of low-discrepancy sequences.

Another difference with standard Monte Carlo is that the convergence rate of Quasi Monte Carlo methods depends on the dimension d of the considered model through the discrepancy of the related quasi random sequence.

Let us stress again that low-discrepancy sequences (Sobol sequences in particular) are often more fruitful in finance than in other application areas. This is because the *effective dimension* of financial problems is often much lower than their *nominal dimension*. In such cases, one can benefit from all the power of low-discrepancy sequences by assigning the main risk factors of the problem, ordered by decreasing amount of explained variance, to the successive components of the points of a multi-dimensional low-discrepancy sequence. Thus, though we use a low-discrepancy sequence in the nominal dimension of the problem, which may be high, the fact that the first coordinates of the quasi-random points are assigned to the main risk factors of the problem avoids much of the drawbacks generally associated with high-dimensional low-discrepancy sequences.

31 Greeking by (Quasi) Monte Carlo

Prices *sensitivities*, or *Greeks*, are actually the main issue in financial modeling. Indeed, unless very exotic products are considered, derivatives *prices* are made by offer-and-demand in the market (prices are effectively used in the reverse-engineering mode to *calibrate* the models, see Part VII). Greeks, on the contrary, must be computed within models. They are used to determine the composition of a dynamic hedging portfolio (see Subsections 9.2 and 13.2).

Now it happens that in many cases, Greeks are also given by expressions of the form $\Theta = \mathbb{E}[\phi(X)]$, so that all the Monte Carlo pricing techniques can also be used for Greeking.

In this Section we illustrate this on the problem of (Q)MC computing the delta of an option in a Markovian model. We thus want to compute $\Delta_0 = \partial_s \mathbb{E}[\phi(S_T^s)]$, where S_T^s is the risk-neutral spot with initial condition s at time 0 (we should in fact write $S_T^{s,\bar{s}}$ where $\bar{s}$ is the

initial condition of the remaining model factors; since $\bar{s}$ plays no role in the sequel, we omit it for simplicity).

One obvious method consists in perturbing the initial condition s, computing (Q)MC prices for the original *and perturbed* initial condition, and deducing a finite differences estimator for Δ_0. But this procedure is both costly (since it involves resimulation) and *parameterized and biased* (via the perturbation on s). In many cases, *direct (without resimulation) approaches* are possible: by derivation of the payoff, provided the latter is smooth enough, or by derivation of the transition probability density $p_T(s, S)$ (as a shorthand for $p_T(s, \bar{s}; S)$) of S, provided the latter exists and is smooth enough.

Note that in the case of the Black–Scholes model the related computations are elementary, however in general the latter two approaches ultimately rely on the theory of stochastic flows and on Malliavin calculus, respectively (see, e.g., [94, 50, 88, 89]).

31.1 Finite Differences

For fixed $h > 0$, we may approach Δ_0 by *decentered finite differences*

$$\frac{1}{\varepsilon s}\left(\mathbb{E}[\phi(S_T^{(1+\varepsilon)s})] - \mathbb{E}[\phi(S_T^s)]\right)$$

or, leading to better convergence properties, by *centered finite differences*

$$\boxed{\frac{1}{2\varepsilon s}\left(\mathbb{E}[\phi(S_T^{(1+\varepsilon)s})] - \mathbb{E}[\phi(S_T^{(1-\varepsilon)s})]\right)} \tag{232}$$

The expectations in (232) are estimated by (Q)MC simulation. In order to optimize the algorithms, common random numbers should be used to estimate both expectations in (232). Thus, in the Black–Scholes model, Δ_0 can be best estimated in these approaches by

$$\frac{1}{2\varepsilon s m}\sum_{j=1}^{m}\left(\phi[(1+\varepsilon)se^{bT+\sigma\sqrt{T}\varepsilon_j}] - \phi[(1-\varepsilon)se^{bT+\sigma\sqrt{T}\varepsilon_j}]\right) ,$$

where the ε_j are standard Gaussian (Q)MC draws.

31.2 Derivation of the payoff

In case where ϕ is regular and we know how to derive S_T^s with respect to s, Δ_0 may alternatively be computed as

$$\Delta_0 = \partial_s \mathbb{E}\phi(S_T^s) = \mathbb{E}\partial_s\phi(S_T^s) = \mathbb{E}\phi'(S_T^s)\partial_s S_T^s .$$

In *multiplicative* models (like the Black–Scholes model, or more general *homogenous* models, see section 37.3), we have $\partial_s S_T^s = \frac{S_T^s}{s}$, so

$$\boxed{\Delta_0 = \tfrac{1}{s}\mathbb{E}[\phi'(S_T^s)S_T^s]}$$

provided ϕ is differentiable. Note that ϕ' is allowed to be a weak derivative in the sense of distributions, yet it needs to be a well defined *function* (unlike a Dirac mass for instance) for applicability of the method. So this method is applicable to the computation of the delta of a call option, but not to that of a digital option.

31.3 Derivation of the spot transition probability density

Assuming that S admits a transition probability density $p_T(s,S)$ between 0 and T, differentiable in the first variable s, then under mild conditions:

$$\Delta_0 = \int_{\mathbb{R}} \phi(S)\partial_1 p_T(s,S)\,dS = \int_{\mathbb{R}} \phi(S)\frac{\partial_1 p_T(s,S)}{p_T(s,S)} p_T(s,S)\,dS\ ,$$

so

$$\boxed{\Delta_0 = \mathbb{E}\left[\phi(S_T^s)\partial_1 \ln(p_T(s,S_T^s))\right]}$$

For instance, in the Black–Scholes model:

$$p_T(s,S) = \frac{1}{S\sqrt{2\pi\sigma^2 T}} e^{-\frac{\left(\ln(\frac{S}{s})-bT\right)^2}{2\sigma^2 T}}\ ,\ \partial_1 \ln(p_T(s,S_T^s)) = \frac{W_T}{s\sigma T}\ ,$$

so

$$\boxed{\Delta_0 = \mathbb{E}\left[\phi(S_T^s)\frac{W_T}{s\sigma T}\right]}$$

which can be easily computed by (Q)MC.

32 (Quasi) Monte Carlo Algorithms for Vanilla Options

32.1 (Q)MC BS1D Algorithm

Description Computation, for an European call, put or digital option, of its price and delta by (Pseudo) Monte Carlo or Quasi Monte Carlo simulation. The Pseudo Monte Carlo method provides estimations for price and delta with a confidence interval. The Quasi Monte Carlo method only provides price and delta estimates, without confidence interval. For a call, the implementation is based on the call-put parity relationship.

Input parameters StepNumber m, Generator_Type, Increment ε, Confidence Value.

Output parameters Price Π, Error Price σ_Π, Delta Δ, Error Delta σ_Δ, Price Confidence Interval IC_Π =[Inf Price, Sup Price], Delta Confidence Interval IC_Δ =[Inf Delta, Sup Delta].

The underlying asset price evolves according to the Black–Scholes model, so, under $\mathbb{P}$:

$$S_T = s\exp(bt + \sigma W_t)\ ,$$

where S_T denotes the spot at maturity T, s is the initial spot, and t is time-to-maturity.

The price of an option at $T-t$ is:

$$\Pi = \mathbb{E}\left[e^{-rt}\phi(K,S_T,R)\right]$$

where ϕ denotes the payoff of the option, K is the strike and R the rebate (for the digital option only). The delta is given by:

$$\Delta = \partial_s \mathbb{E}[e^{-rt}\phi(K,S_T,R)]\ .$$

The estimators write:

$$\boxed{\Pi_m = \frac{1}{m}e^{-rt}\sum_{j=1}^{m}\pi(j)}$$

with $\pi(j) = \phi(K, S_T(j), K)$, and

$$\boxed{\Delta_m = \frac{1}{m} e^{-rt} \sum_{j=1}^{m} \partial_s \pi(j) = \frac{1}{m} e^{-rt} \sum_{j=1}^{m} \delta(j)}$$

The values for $\pi(j)$ and $\delta(j)$ are detailed for each option.

• **Put**: The payoff is $(K - S_T)^+$, hence

$$\pi(j) = (K - S_T(j))^+$$

$$\delta(j) = \begin{cases} -\partial_s S_T(j) = -\frac{S_T(j)}{s} & \text{if} \quad \pi(j) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

• **Call**: The payoff is $(S_T - K)^+$. The *call-put parity* relations for price and delta write:

$$C = P + se^{-qt} - Ke^{-rt}$$
$$\Delta_C = \Delta_P + e^{-qt} \ ,$$

where C and P respectively denotes the call and the put prices. These relations may be used for the call simulation (in order to limit variance).

• **Digital option**: The payoff is $R\mathbb{1}_{\{S_T - K \geq 0\}}$, hence

$$\pi(j) = R\mathbb{1}_{\{S_T(j) - K \geq 0\}} \ .$$

To have an estimation of the delta in the case of a digital option, we proceed by finite differences, so

$$\delta(j) = \frac{1}{2\varepsilon s} \left[\phi\left((1+\varepsilon) S_T(j), K, R\right) - \phi\left((1-\varepsilon) S_T(j), K, R\right) \right] \ .$$

32.1.1 Adding Jumps

Let us now add jumps in S, assuming that the underlying asset price evolves according to the risk-neutral Merton model, so, by (104):

$$\boxed{S_T = se^{at + \sigma W_t} \prod_{l=1}^{N_t} (J_l + 1)}$$

with $a = b - \gamma \bar{J}$, where N is a Poisson process with deterministic jump intensity γ and $\ln(1 + J_l) \hookrightarrow \mathcal{N}(\alpha, \beta)$.

From the point of view of (Q)MC pricing European path-independent options with payoff $\phi(S_T)$ (like the European call, put or digital options of section 32.1), essentially nothing changes, except for the fact that $S_T(j)$ in section 32.1 is now given by (cf. (104))

$$\boxed{S_T(j) = se^{at + \sigma W_t(j)} \prod_{l=1}^{N_p} \exp(\alpha + \sqrt{\beta} \varepsilon_j^l)}$$

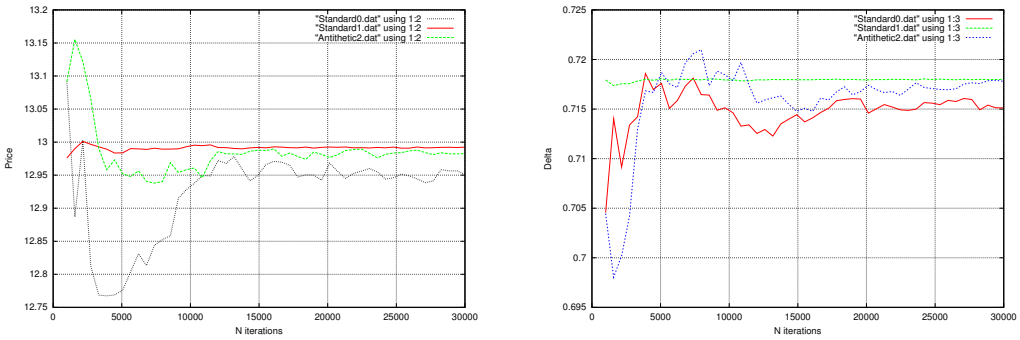

Figure 12: *European vanilla call priced by Pseudo Monte Carlo simulation (L'Ecuyer's generator), Quasi Monte Carlo simulation (Sobol sequence), and Pseudo Monte Carlo simulation with antithetic variables.*

where:
- N_p is a simulated Poisson variable with parameter γT,
- the ε_j^l are independent standard Gaussian draws, and
- $W_t(j) = \sqrt{t}\varepsilon_j$ for a further independent standard Gaussian draw ε_j.

32.2 (Q)MC BS2D Algorithm

Description Computation, for a Call on Maximum, Put on Minimum, Exchange or BestOf European option, of its Price and Deltas by (Pseudo) Monte Carlo or Quasi Monte Carlo simulation. The Pseudo Monte Carlo algorithm also provides confidence intervals.

Input parameters StepNumber m, Generator_Type, Increment ε, Confidence Value.

Output parameters Price Π, Error Price σ_Π, Deltas Δ^1, Δ^2, Errors delta $\sigma_{\Delta^1}, \sigma_{\Delta^2}$, Price and Deltas Confidence Intervals.

The underlying asset prices evolve according to the two-dimensional Black–Scholes dynamics, that is, under $\mathbb{P}$:

$$\begin{cases} dS_u^1 = S_u^1(\kappa_1 du + \sigma_1 dW_u^1), \ S_{T-t}^1 = s_1 \\ dS_u^2 = S_u^2(\kappa_2 du + \sigma_2 dW_u^2), \ S_{T-t}^2 = s_2 \end{cases}$$

where:
- $\kappa_l = r - q_l$,
- s_l is the initial spot value,
- W^1 and W^2 denote two real-valued Brownian motions with instantaneous correlation ρ.

So

$$\begin{cases} S_T^1 = s_1 \exp(b^1 t + \sigma_{11} W_t^1) \\ S_T^2 = s_2 \exp(b^2 t + \sigma_{21} W_t^1 + \sigma_{22} \widetilde{W}_t^2) \end{cases},$$

with

$$\begin{pmatrix} b^1 \\ b^2 \end{pmatrix} = \begin{pmatrix} \kappa_1 - \frac{\sigma_1^2}{2} \\ \kappa_2 - \frac{\sigma_2^2}{2} \end{pmatrix}, \quad \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix}$$

and where $\widetilde{W}$ denotes a real-valued Brownian motion independent of W^1. The price of an option is

$$\Pi = \mathbb{E}\left[e^{-rt}\phi(K, S_T^1, S_T^2)\right] \ ,$$

where ϕ denotes the payoff of the option, K is the strike, and t is time to maturity.
The deltas are given by

$$\Delta^1 = \partial_{s_1}\mathbb{E}[e^{-rt}\phi(K, S_T^1, S_T^2)] \ , \ \Delta^2 = \partial_{s_2}\mathbb{E}[e^{-rt}\phi(K, S_T^1, S_T^2)] \ .$$

The estimators write:

$$\boxed{\begin{array}{c} \Pi_m = \frac{1}{m}e^{-rt}\sum_{j=1}^m \pi(j) \\ \Delta_m^l = \frac{1}{m}e^{-rt}\sum_{j=1}^m \partial_{s_l}\pi(j) = \frac{1}{m}e^{-rt}\sum_{j=1}^m \delta^l(j) \end{array}}$$

The values for $\pi(j)$ and $\delta^l(j)$ are detailed for each option.

• **Put on the Minimum**: The payoff is $(K - \min(S_1, S_2))^+$, so

$$\pi(j) = \left(K - \min(S_T^1(j), S_T^2(j))\right)^+$$

If $\pi(j) \geq 0$, then

$$\delta^1(j) = \begin{cases} -\exp(b^1 t + \sigma_{11}W_t^1) & \text{if} \quad S_T^1(j) \leq S_T^2(j) \\ 0 & \text{otherwise} \end{cases}$$
$$\delta^2(j) =$$
$$\begin{cases} -\exp(b^2 t + \sigma_{21}W_t^1 + \sigma_{22}\widetilde{W}_t^2) & \text{if} \quad S_T^1(j) \leq S_T^2(j) \\ 0 & \text{otherwise} \ . \end{cases}$$

Otherwise $\delta^1(j) = \delta^2(j) = 0$.
• **Call on the Maximum**: The payoff is $(\max(S_1, S_2) - K)^+$, so

$$\pi(j) = \left(\max(S_T^1(j), S_T^2(j)) - K\right)^+$$

If $\pi(j) \geq 0$, then

$$\delta^1(j) = \begin{cases} \exp(b^1 t + \sigma_{11}W_t^1) & \text{if} \quad S_T^1(j) \geq S_T^2(j) \\ 0 & \text{otherwise} \end{cases}$$
$$\delta^2(j) =$$
$$\begin{cases} \exp(b^2 t + \sigma_{21}W_t^1 + \sigma_{22}\widetilde{W}_t^2) & \text{if} \quad S_T^1(j) \geq S_T^2(j) \\ 0 & \text{otherwise} \ . \end{cases}$$

Otherwise $\delta^1(j) = \delta^2(j) = 0$.
• **Exchange Option**: The payoff is $(S_1 - KS_2)^+$, so

$$\pi(j) = \left(S_T^1 - KS_T^2\right)^+$$

$$\delta^1(j) = \begin{cases} \exp(b^1 t + \sigma_{11}W_t^1) & \text{if} \quad \pi(j) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$
$$\delta^2(j) =$$
$$\begin{cases} -K\exp(b^2 t + \sigma_{21}W_t^1 + \sigma_{22}\widetilde{W}_t^2) & \text{if} \quad \pi(j) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

• **BestOf Option**: The payoff is $[\max(S_1 - K_1, S_2 - K_2)]^+$, so

$$\pi(j) = \left[\max(S_T^1 - K_1, S_T^2 - K_2)\right]^+$$

If $\pi(j) \geq 0$, then

$$\delta^1(j) = \begin{cases} \exp(b^1 t + \sigma_{11} W_t^1) & \text{if} \quad S_T^1(j) - K_1 \geq S_T^2(j) - K_2 \\ 0 & \text{otherwise} \end{cases}$$
$$\delta^2(j) =$$
$$\begin{cases} \exp(b^2 t + \sigma_{21} W_t^1 + \sigma_{22} \widetilde{W}_t^2) & \text{if} \quad S_T^1(j) - K_1 \geq S_T^2(j) - K_2 \\ 0 & \text{otherwise .} \end{cases}$$

Otherwise $\delta^1(j) = \delta^2(j) = 0$.

33 Simulation of Processes

Simulating random variables is enough to (Q)MC-compute prices of vanilla options, in cases where the value of the underlying at the maturity time T of the option can be simulated directly, without discretizing the associated SDE. However, to be able to deal with more complex models, or, even in simple models, with path dependent options, one needs to *simulate the whole trajectory* of the underlying between the pricing time t and the maturity T. Even for simple options in simple models, simulating spot *trajectories* is necessary for testing the *hedging performances* of a model (see Section 35). In this case, the 'payoff' we are interested in typically consists of *the Profit and Loss at maturity* of an option's seller, who rebalances a delta hedge in the underlying at regular time intervals. The Profit and Loss at maturity is *a path dependent quantity*, so that the simulation problem in this case is similar to that of Monte Carlo pricing a path dependent option.

33.1 Brownian Motion

We recall that a Brownian motion on the probability space $(\Omega, \mathbb{F}, \mathbb{P})$ is a continuous process W with the properties that $W_0 = 0$ and for $0 \leq s < t$, the increment $W_t - W_s$ is independent of $\mathcal{F}_s$ and is normally distributed with mean zero and variance $t - s$.

Simulation of W_T, $(W_t|W_T)$ and $(W_s, W_t|W_T)$ for $0 \leq s \leq t \leq T$ By Cholesky, we have for a standard Gaussian triple $(\varepsilon_s, \varepsilon_t, \varepsilon_T)$:

$$\begin{cases} W_T = \sqrt{T}\varepsilon_T \\ W_t = \sqrt{t}\Big(\sqrt{\frac{t}{T}}\varepsilon_T + \sqrt{1 - \frac{t}{T}}\varepsilon_t\Big) \\ W_s = \sqrt{s}\Big(\sqrt{\frac{s}{T}}\varepsilon_T + \sqrt{\rho^2}\varepsilon_t + \sqrt{1 - \frac{s}{T} - \rho^2}\Big) \end{cases}$$

where ρ is such that $s = \sqrt{st}\big(\sqrt{\frac{ts}{T^2}} + \sqrt{1 - \frac{t}{T}}\rho\big)$. So, in particular:

$$W_T \overset{(law)}{=} \mathcal{N}(0, T) \ , \ (W_t|W_T) \overset{(law)}{=} \mathcal{N}\left(\frac{t}{T}W_T, \frac{t\,(T-t)}{T}\right) \tag{233}$$

and

$$\mathbb{C}\mathrm{ov}(W_s, W_t|W_T) = \sqrt{ts}\sqrt{1 - \frac{t}{T}}\rho = \sqrt{ts}\big(\sqrt{\frac{s}{t}} - \sqrt{\frac{ts}{T^2}}\big) = s(1 - \frac{t}{T}) \tag{234}$$

Simulation (discretization) of a Brownian trajectory on $[0, T]$ We present two approaches for simulating a Brownian path. Typically, for path dependent options, we have to simulate W over $\{t_0 = 0, t_1, \ldots, t_n = T\}$. Let $h_i = t_{i+1} - t_i$ for $i = 0 \ldots n-1$ (so $h_i = h = \frac{T}{n}$, in the case of a uniform discretisation).

Forward simulation of W is given by

$$\boxed{\begin{array}{l} W_0 = 0 \\ W_{t_{i+1}} = W_{t_i} + \sqrt{h_i}\varepsilon_i \end{array}}$$

where $(\varepsilon_1, ..., \varepsilon_n)$ are independent standard Gaussian variables.

Attention: Simulating independently each W_{t_i} as $\sqrt{t_i}\varepsilon_i$ would be incorrect (e.g., incorrect variance for $W_{t_{i+1}} - W_{t_i}$).

Backward simulation with Brownian Bridge An alternative method is based on the following *Brownian Bridge* property (cf. (233)):

$$\mathcal{L}(W_u, s < u < t | W_s = x, W_t = y) = \mathcal{N}(\tfrac{t-u}{t-s}x + \tfrac{u-s}{t-s}y, \tfrac{(t-u)(u-s)}{t-s}) \ ,$$

hence in particular

$$\mathcal{L}(W_{\frac{t+s}{2}} | W_s = x, W_t = y) = \mathcal{N}(\tfrac{x+y}{2}, \tfrac{t-s}{4}) \quad .$$

The related scheme consists in simulating W as

$$\boxed{\begin{array}{l} W_0 = 0 \\ W_T = \sqrt{T}\varepsilon_1 \\ W_{\frac{T}{2}} = \frac{W_0 + W_T}{2} + \sqrt{\frac{T}{4}}\varepsilon_2 \\ W_{\frac{T}{4}} = \frac{W_0 + W_{\frac{T}{2}}}{2} + \sqrt{\frac{T}{8}}\varepsilon_3 \\ W_{\frac{3T}{4}} = \frac{W_{\frac{T}{2}} + W_T}{2} + \sqrt{\frac{T}{8}}\varepsilon_4 \\ \dots \end{array}}$$

where $(\varepsilon_1, ..., \varepsilon_n)$ are independent standard Gaussian variables.

For this algorithm, one must choose n as a power of 2. The first step is directly from 0 to T. Intermediates steps are filled by taking successive subdivisions of the time intervals into halves. The algorithm can be adapted to non-uniform time subdivisions by considering the conditional law of the Brownian Bridge between s and t.

Remarks on these two schemes for Monte Carlo and Quasi Monte Carlo simulations Both schemes require a vector of n independent Gaussian variables. In the case of a Monte Carlo method, these n variables can be simulated using successive draws from a pseudo random numbers generator. However for a Quasi Monte Carlo simulation we need to take care about the independence property. In this regard a good point with the backward scheme is that the values of W successively determined on each trajectory, are drawn by order of decreasing variance. Therefore the first components (which are also 'the best ones', i.e. 'the more uniform ones') of the simulated Quasi Monte Carlo points are naturally affected to the main directions of risk in the problem.

33.2 Black–Scholes Model

In the risk-neutral (might be *objective with constant drift* μ) Black–Scholes model: $S_t = S_0 \exp(bt + \sigma W_t)$, S_t is a function of W_t. The simulation of price paths is then directly based on the simulation of Brownian motion described in section 33.2.

Forward simulation We have

$$S_{t_{i+1}} = S_{t_i} \exp(bh_i + \sigma\sqrt{h_i}\varepsilon_i) \ ,$$

Backward simulation Denoting $x_t = \ln(S_t) = x_0 + bt + \sigma W_t$, we get

$$\begin{aligned} x_T &= x_0 + bt + \sigma\sqrt{T}\varepsilon_1 \\ x_{\frac{T}{2}} &= \tfrac{x_0+x_T}{2} + \sigma\sqrt{\tfrac{T}{4}}\varepsilon_2 \\ x_{\frac{T}{4}} &= \tfrac{x_0+x_{\frac{T}{2}}}{2} + \sigma\sqrt{\tfrac{T}{8}}\varepsilon_3 \\ x_{\frac{3T}{4}} &= \tfrac{x_{\frac{T}{2}}+x_T}{2} + \sigma\sqrt{\tfrac{T}{8}}\varepsilon_4 \\ &\dots \end{aligned}$$

We finally set $S_{t_i} = \exp(x_{t_i})$.

33.3 General diffusions: Euler and Milshtein schemes

We now consider the d-dimensional diffusion process

$$\boxed{dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t}$$

(under suitable Lipschitz continuous conditions on the coefficients). Unless an explicit solution is known for X (like in the Black–Scholes model, cf. section 33.3), we have to approximate the process by time-discretization. The two best known schemes are the Euler and the Milshtein schemes.

33.3.1 Euler Scheme

It is defined by

$$\widehat{X}_{t_{i+1}} = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i})h_i + \sigma(t_i, \widehat{X}_{t_i})(W_{t_{i+1}} - W_{t_i}) \ .$$

Simulation is obtained with a forward algorithm by

$$\boxed{\widehat{X}_{t_{i+1}} = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i})h_i + \sigma(t_i, \widehat{X}_{t_i})\sqrt{h_i}\varepsilon_i}$$

with ε_i Gaussian iid, for $i = 0, \dots, n-1$.

In the following Theorem we assume $h_i = h$, for notational simplicity.

Theorem 33.1 • L^q-Convergence and Trajectorial Convergence *For b, σ regular enough, such that in particular*

$$|b(t,x) - b(s,x)| + |\sigma(t,x) - \sigma(s,x)| \leq C(1 + |x|)(t-s)^\alpha$$

for positive constants C and α, we have:

$$\mathbb{E}\left(\sup_{0\leq i\leq n} |X_{ih} - \widehat{X}_{ih}|^{2q}\right) \leq Ch^{2q\beta} \ , \ q \geq 1 \tag{235}$$

$$\lim_{h\to 0+} h^{-\alpha} \sup_{0\leq i\leq n} |X_{ih} - \widehat{X}_{ih}| = 0 \ a.s. \ , \ \alpha < \beta \tag{236}$$

with $\beta = \min(\alpha, \frac{1}{2})$. *Moreover, in the homogenous case* $b = b(x), \sigma = \sigma(x)$, *one may take* $\beta = \frac{1}{2}$ *in (235)-(236).*

• **Convergence in law** *For* b, σ *and* ϕ *regular enough, there exists a constant* C_T *such that*

$$|\mathbb{E}\phi(X_T) - \mathbb{E}\phi(\widehat{X}_T)| \leq C_T h \ .$$

So L^2-convergence is basically of order $h^{\frac{1}{2}}$, and almost sure (trajectorial) convergence is of order $h^{\frac{1}{2}-\varepsilon}$, for any $\varepsilon > 0$. Convergence in law (the kind of convergence which is relevant for pricing and Greeking applications) is *linear in* h, in regular cases.

Continuous Euler scheme $(d = 1)$ This is the *continuous time approximation scheme* $(\bar{X}_t)_{t\geq 0}$ defined by interpolation of the previous Euler scheme by a *Brownian Bridge* (see [115, 161]) between $(t_i, \widehat{X}_{t_i})$ and $(t_{i+1}, \widehat{X}_{t_{i+1}})$, for $i = 0, \ldots, n-1$. So, on $[t_i, t_{i+1}]$:

$$\boxed{\bar{X}_t = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i})(t - t_i) + \sigma(t_i, \widehat{X}_{t_i})B^i_t}$$

where $\widehat{X}$ is a (discrete) trajectory of the Euler scheme for X, and B^i is a Brownian Bridge on $[t_i, t_{i+1}]$, such that $\bar{X} = \widehat{X}$ at t_i and t_{i+1}.

33.3.2 Milshtein Scheme $(d = 1)$

It is defined by (assuming $d = 1$ and in the homogenous case $b = b(x), \sigma = \sigma(x)$):

$$\begin{aligned}\widehat{X}_{t_{i+1}} = \widehat{X}_{t_i} + (b(\widehat{X}_{t_i}) - \frac{1}{2}\sigma'(\widehat{X}_{t_i})\sigma(\widehat{X}_{t_i}))h_i + \\ +\sigma(\widehat{X}_{t_i})(W_{t_{i+1}} - W_{t_i}) + \frac{1}{2}\sigma'(\widehat{X}_{t_i})\sigma(\widehat{X}_{t_i})(W_{t_{i+1}} - W_{t_i})^2 \ .\end{aligned}$$

Simulation is obtained with a forward algorithm by

$$\boxed{\begin{array}{c}\widehat{X}_{t_{i+1}} = \widehat{X}_{t_i} + (b(\widehat{X}_{t_i}) - \frac{1}{2}\sigma'(\widehat{X}_{t_i})\sigma(\widehat{X}_{t_i}))h_i + \\ +\sigma(\widehat{X}_{t_i})\sqrt{h_i}\varepsilon_i + \frac{1}{2}\sigma'(\widehat{X}_{t_i})\sigma(\widehat{X}_{t_i})h_i\varepsilon_i^2\end{array}}$$

for Gaussian iid r.v. ε_i, $i = 0, \ldots, n-1$.

Theorem 33.2 • **L^q-Convergence and Trajectorial Convergence** *Assume* $h_i = h$. *For* $b = b(x), \sigma = \sigma(x)$ *regular enough, we have*

$$\mathbb{E}\left(\sup_{0\leq i\leq n} |X_{ih} - \widehat{X}_{ih}|^{2q}\right) \leq Ch^{2q} \ , \ q \geq 1$$

$$\lim_{h\to 0+} h^{-\alpha} \sup_{0\leq i\leq n} |X_{ih} - \widehat{X}_{ih}| = 0 \ a.s. \ , \ \alpha < 1 \ .$$

• **Convergence in law** *Same result as for the Euler scheme.*

As for pricing and Greeking, (weak) convergence is thus again *linear in* h, in regular cases. But the rate of *trajectorial* (or strong) convergence is improved with respect to the Euler scheme.

33.3.3 Example: Heston model

We consider the risk-neutral Heston model (cf. (99)):

$$\begin{cases} dV_t = -\lambda(V_t - \theta)dt + \eta\sqrt{V_t}dZ_t \\ dS_t = S_t(\kappa dt + \sqrt{V_t}dW_t) \end{cases} \tag{237}$$

with $d\langle W, Z\rangle = \rho dt$. A classic discretization for (237) consists in discretizing V by the Milshtein scheme and $\ln(S)$ by the Euler scheme, as follows:

$$\begin{cases} \widehat{V}_{t_{i+1}} - \widehat{V}_{t_i} = -(\lambda(\widehat{V}_{t_i} - \theta) + \dfrac{\eta^2}{4})h_i + \eta\sqrt{\widehat{V}_{t_i}h_i}\varepsilon_1 + \dfrac{h_i\eta^2\varepsilon_1^2}{4} \\ \ln(\widehat{S}_{t_{i+1}}) - \ln(\widehat{S}_{t_i}) = (\kappa - \frac{\widehat{V}_{t_i}}{2})h_i + \sqrt{\widehat{V}_{t_i}h_i}(\rho\varepsilon_1 + \sqrt{1-\rho^2}\varepsilon_2) \end{cases}$$

where $(\varepsilon_1, \varepsilon_2)$ is a standard Gaussian pair. The interest of using the Milshtein scheme for the variance process is to benefit from a better trajectorial convergence than with the Euler scheme. So the process $\widehat{V}$ obtained in this way is less prone to take negative values than it would be the case with an Euler scheme (see also Andersen [5]).

33.4 Jump–Diffusions

Finally we consider the following *jump–diffusion*, driven by a multidimensional Brownian motion W and a Poisson random measure J (under suitable Lipschitz continuous conditions on the coefficients):

$$\boxed{\begin{array}{c} dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t + \int_{x\in\mathbb{R}^d}\delta(t, X_{t-}, x)\mu(dx, dt) \\ = b(t, X_t)dt + \sigma(t, X_t)dW_t + d\left(\sum_{l=1}^{N_t}\delta(t, X_{\tau_l-}, J_{\tau_l})\right) \end{array}} \tag{238}$$

where $\mu(dx, dt)$ is a Poisson random measure with compensator measure $g(t, X_t)h(t, X_t, dx)dt$, for some *intensity* g and *jump size probability measure* h (so $N_t = \mu(\mathbb{R}^q \times [0, t])$ and J_{τ_l} is a r.v. with law $h(t, X_{\tau_l-}, dx)$ in the second line). Depending on the application at hand, we may be interested at simulating X at fixed times, or at the jump times of X.

Euler scheme at fixed times To simulate (238) at fixed times $0 < t_1 <, \ldots, < t_n$, set $\widehat{X}_0 = X_0$, and for $i = 0, \ldots, n-1$:

• simulate

$$\widetilde{X}_{t_{i+1}} = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i})h_i + \sigma(t_i, \widehat{X}_{t_i})\sqrt{h_i}\varepsilon_i \ ,$$

• compute $\widehat{X}_{t_{i+1}}$ by adding to $\widetilde{X}_{t_{i+1}}$, with probability $1 - e^{-g(t_i, \widehat{X}_{t_i})h_i}$ (as dictated by the position of an independent uniform draw u_i with respect to $e^{-g(t_i, \widehat{X}_{t_i})h_i}$), a jump term equal to $\delta(t_i, \widehat{X}_{t_i-}, x)$ with probability $h(t_i, \widehat{X}_{t_i-}, dx)$.

Euler scheme at jump times To simulate (238) at the n first jump times $0 < t_1 < \ldots < t_n$ of X (so the t_i and h_i are random, here), set $\widehat{X}_0 = X_0$, and for $i = 0, \ldots, n-1$:

• simulate t_{i+1} as t_i plus an independent draw in an exponential law with parameter $g(t_i, \widehat{X}_{t_i})$,

• simulate

$$\widetilde{X}_{t_{i+1}} = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i})(h_i) + \sigma(t_i, \widehat{X}_{t_i})\sqrt{h_i}\varepsilon_i \ ,$$

• compute $\widehat{X}_{t_{i+1}}$ by adding to $\widetilde{X}_{t_{i+1}}$ a jump term equal to $\delta(t_{i+1}, \widetilde{X}_{t_{i+1}-}, x)$ with probability $h(t_{i+1}, \widetilde{X}_{t_{i+1}-}, dx)$.

Continuous Euler scheme $(d = 1)$ This is the continuous time approximation scheme defined by interpolation of the previous Euler scheme at jump times by a Brownian bridge between $(t_i, \widehat{X}_{t_i})$ and $(t_{i+1}, \widehat{X}_{t_{i+1}-})$, for $i = 0, \ldots, n-1$.

33.5 Monte Carlo Simulation for Processes

In the case of Monte Carlo simulation for processes, we have

$$\boxed{\mathbb{E}[\phi(X)] - \frac{1}{m}\sum_{j=1}^{m}\phi(\widehat{X}_j) = (\mathbb{E}[\phi(X)] - \mathbb{E}[\phi(\widehat{X})]) + (\mathbb{E}[\phi(\widehat{X})] - \frac{1}{m}\sum_{j=1}^{m}\phi(\widehat{X}_j))}$$

So the error is the sum of two terms, a *discretization error* and a *Monte Carlo error (simulation error)*:
• for usual discretization schemes such as the Euler or the Milshtein scheme, the weak convergence rate is linear in h, so the discretization error is of the order $\mathrm{O}(h)$;
• after scaling by $m^{-\frac{1}{2}}$, the Monte Carlo error is asymptotically distributed as $\mathcal{N}(0, \mathbb{V}\mathrm{ar}[\phi(\widehat{X})])$.

Remark 33.1 The overall error is thus of the order of $\mathrm{O}(h)+\mathrm{O}(m^{-\frac{1}{2}})$, to be compared with $\mathrm{O}(h) + \mathrm{O}(m_1^{-2})$, in the case of a typical finite differences numerical scheme with a generic number m_1 of mesh points per space dimension. Taking m as m_1^d so that both schemes have comparable computing costs, we thus see that the Monte Carlo method is more efficient for $d > 4$, and less efficient for $d < 4$ (cf. Table 1 and the related discussion in Section 3).

Therefore in order to balance the two terms in the error a natural choice is to take m of the order of n^2.

34 (Quasi) Monte Carlo methods for Exotic Options

A nice feature of Monte Carlo methods is that they can easily deal with path dependent payoffs. Note however that specific treatments must be applied in order to preserve convergence rates. The idea to efficiently (Q)MC price path dependent payoffs is to use the continuous Euler scheme for the underlying assets. (see sections 33.3.1 and 33.4).

34.1 Lookback options

We consider a Lookback option with payoff $\phi(X_T, M_T)$ at T, where X is given by the following one-dimensional diffusion:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$$

(under suitable Lipschitz continuous conditions on the coefficients), and $M_t = \sup_{0\le s\le t} X_s$. The following results give a way of simulating the pair $(\bar{X}, \bar{M})$, where $\bar{M}_t = \sup_{[0,t]} \bar{X}$.

Lemma 34.1 *Denoting*

$$W_t^\lambda = W_t + \lambda t\ ,\ M_t^\lambda = \sup_{0\le s\le t} W_s^\lambda\ ,$$

where λ is a real number (drift parameter), the random variable

$$Z_t^\lambda = (2M_t^\lambda - W_t^\lambda)^2 - (W_t^\lambda)^2$$

is, conditionally on W_t^λ, $\frac{1}{2t}$ – exponentially distributed.

Proof By the Cameron–Martin formula, we have for any $x, y \in \mathbb{R}$:

$$\widehat{\mathbb{P}}(W_t^\lambda \le x) = \nu^\lambda \widehat{\mathbb{P}}(W \le x)\ ,\ \widehat{\mathbb{P}}(W_t^\lambda \le x,\, M_t^\lambda \le y) = \nu^\lambda \widehat{\mathbb{P}}(W \le x,\, M_t \le y)$$

for a suitable factor ν^λ. Therefore

$$\widehat{\mathbb{P}}(M_t^\lambda \le y \,|\, W_t^\lambda \le x) \ = \ \widehat{\mathbb{P}}(M_t \le y \,|\, W_t \le x)\ .$$

This shows that the law of M_t^λ conditional on W_t^λ does not depend on λ. So we may and do reduce attention to the case where $\lambda = 0$. Now it is well known that the law of the pair (W_t, M_t) admits the following transition probability density between times 0 and t :

$$p(x,y) = \mathbb{1}_{y \ge x^+} \frac{2(2y-x)}{\sqrt{2\pi t^3}} \exp\left[-\frac{(2y-x)^2}{2t}\right] \tag{239}$$

This can for instance be established by taking $-\partial^2_{xy}$ in the following identity (*mirror's lemma*, see, e.g., [115, 161]), which is valid for any $x \in \mathbb{R}$ and $y \in [x^+, +\infty)$:

$$\widehat{\mathbb{P}}(W_t \ge 2y - x) = \widehat{\mathbb{P}}(W_t \le x,\, M_t \ge y)\ .$$

Denoting $Z_t = (2M_t - W_t)^2 - (W_t)^2$ and introducing the one-to-one mapping

$$[x^+, +\infty) \ni y \mapsto z = (2y-x)^2 - x^2 \in \mathbb{R}_+\ ,$$

we have:

$$\widehat{\mathbb{P}}(Z_t \in dz \,|\, W_t = x) = \widehat{\mathbb{P}}(M_t \in dy \,|\, W_t = x)\ .$$

So, by (239):

$$\widehat{\mathbb{P}}(Z_t \in dz \,|\, W_t = x) = \frac{1}{2t} \exp(\frac{-z}{2t}) dz\ .$$

□

Proposition 34.2 *Set $\widehat{X} = (\widehat{X}_{t_i})_{0 \le i \le n}$. The law of $(\widetilde{M}^i)_{0 \le i \le n-1}$, with $\widetilde{M}^i = (\sup_{ih \le t \le t_{i+1}} \bar{X}_t \,|\, \widehat{X})$, $i = 0, \dots, n-1$, can be simulated by, for $i = 0, \dots, n-1$:*

$$\frac{1}{2}\left(\widehat{X}_{t_i} + \widehat{X}_{t_{i+1}} + \sqrt{(\widehat{X}_{t_i} - \widehat{X}_{t_{i+1}})^2 - 2\sigma(t_i, \widehat{X}_{t_i})^2 h_i \ln(1-U_i)}\right) =: F_i^{-1}(U; \widehat{X}_{t_i}, \widehat{X}_{t_{i+1}}) \tag{240}$$

where $(U_i)_{0 \le i \le n-1}$ is a sequence of independent uniform r.v. on $[0,1]$.

Proof. Setting $\lambda_i = \frac{b(t_i, \widehat{X}_{t_i})}{\sigma(t_i, \widehat{X}_{t_i})}$, we have in the notation of Lemma 34.1, for $t \in [t_i, t_{i+1}]$:

$$\frac{(\bar{X}_t \,|\, \widehat{X}) - \widehat{X}_{t_i}}{\sigma(t_i, \widehat{X}_{t_i})} = \left(W_t^{\lambda_i} - W_{t_i}^{\lambda_i} \,|\, W_{t_{i+1}}^{\lambda_i} - W_{t_i}^{\lambda_i} = \frac{\widehat{X}_{t_{i+1}} - \widehat{X}_{t_i}}{\sigma(t_i, \widehat{X}_{t_i})}\right)$$

and

$$\frac{\widetilde{M}^i - \widehat{X}_{t_i}}{\sigma(t_i, \widehat{X}_{t_i})} = \left(\sup_{t \in [t_i, t_{i+1}]} W_t^{\lambda_i} - W_{t_i}^{\lambda_i} \,|\, W_{t_{i+1}}^{\lambda_i} - W_{t_i}^{\lambda_i} = \frac{\widehat{X}_{t_{i+1}} - \widehat{X}_{t_i}}{\sigma(t_i, \widehat{X}_{t_i})} \right)$$

By application of Lemma 34.1 to the drifted Brownian motion $W_{t_i+h}^{\lambda_i} - W_{t_i}^{\lambda_i}$ on $[0, h_i]$, we thus get, conditionally on $\widehat{X}_{t_i}$ and $\widehat{X}_{t_{i+1}}$:

$$\left(2\frac{\widetilde{M}^i - \bar{X}_{t_i}}{\sigma(t_i, \widehat{X}_{t_i})} - \frac{\widehat{X}_{t_{i+1}} - \widehat{X}_{t_i}}{\sigma(t_i, \widehat{X}_{t_i})}\right)^2 \overset{(law)}{=} \left(\frac{\widehat{X}_{t_{i+1}} - \widehat{X}_{t_i}}{\sigma(t_i, \widehat{X}_{t_i})}\right)^2 + \mathcal{E}_{\frac{1}{2h}} \ ,$$

where $\mathcal{E}_{\frac{1}{2h}} \overset{(law)}{=} -2h \ln(U)$. □

Thus, to generate a pair $(\bar{X}_T, \bar{M}_T)$:
• first simulate a (discrete) trajectory $\widehat{X}$ of the Euler scheme for X, using n independent uniform random draws;
• then simulate $(\widetilde{M}^i)_{0 \le i \le n-1}$ as in Proposition 34.2, using n further independent uniform random draws;
• set $\bar{X}_T = \widehat{X}_T$, $\bar{M}_T = \max_i \widetilde{M}^i$.

Note that if quasi-random numbers are used, one must use a $2n$-dimensional low-discrepancy sequence. Yet the use of high-dimensional low-discrepancy sequences should be considered with caution.

In the special case of the Black–Scholes model, the Euler discretization is exact, provided one works in returns variable $x = \ln(S)$. In this case one can take n equal to one and use a 2-dimensional low-discrepancy sequence.

Remark 34.1 In the case of a Lookback option on the running *minimum* of X, one can proceed likewise, relying on the following transition probability density q (deduced by symmetry from (239)) for the law of the pair (W_t, m_t), where m_t denotes the running minimum of W between 0 and t:

$$q(x, y) = \mathbb{1}_{y \le -x^-} \frac{2(x - 2y)}{\sqrt{2\pi t^3}} \exp\left[-\frac{(x - 2y)^2}{2t}\right] \tag{241}$$

34.1.1 Andersen and Brotherton-Ratcliffe Algorithm

Description Computation, for a Lookback option, of its Price and Delta by Monte Carlo Simulation (see [6]).

Let $S_T = s \exp(bt + \sigma W_t)$ denote the risk-neutral Black–Scholes spot. We note $S_T^* = \max_{[T-t,T]} S$ the maximum reached before maturity. The price and delta of a lookback option with payoff ϕ and strike K write:

$$\Pi = \mathbb{E}\left[e^{-rt}\phi(K, S_T, S_T^*)\right] \ , \ \Delta = \partial_s \mathbb{E}[e^{-rt}\phi(K, S_T, S_T^*)] \ ,$$

with related estimators:

$$\Pi_m = \tfrac{1}{m} e^{-rt} \textstyle\sum_{j=1}^m \pi(j)$$
$$\Delta_m = \tfrac{1}{m} e^{-rt} \textstyle\sum_{j=1}^m \partial_s \pi(j) = \tfrac{1}{m} e^{-rt} \textstyle\sum_{j=1}^m \delta(j)$$

• **Fixed Strike Lookback Call** The payoff is $(\max_{[T-t,T]} S - K)^+$, so

$$\pi(j) = (S_T^*(j) - K)^+$$

$$\delta(j) = \begin{cases} \partial_s S_T^*(j) = \frac{S_T^*(j)}{s} & \text{if} \quad \pi(j) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

• **Floating Strike Lookback Put** The payoff is $(\max_{[T-t,T]} S - S_T)$, so

$$\pi(j) = (S_T^*(j) - S_T(j))$$

$$\delta(j) = \partial_s S_T^*(j) - \partial_s S_T(j) = \frac{S_T^*(j) - S_T(j)}{s} = \frac{\pi(j)}{s}$$

Simulation of the maximum S_T^* The conditional cumulative distribution function of the maximum M of $X = \ln S$ given $X_{T-t} = x_1$ and $X_T = x_2$ writes:

$$F(x; x_1, x_2) = \left(1 - \exp\Big[-\frac{2}{\sigma^2 t}(x - x_1)(x - x_2)\Big]\right) \mathbb{1}_{x > x_1 \vee x_2}$$

so for $y \in [0, 1]$, consistently with (240):

$$F^{-1}(y; x_1, x_2) = \frac{1}{2}\Big(x_1 + x_2 + \sqrt{(x_1 - x_2)^2 - 2\sigma^2 t \ln(1 - y)}\Big) \, .$$

At run j, $(S_T^*)^j$ is simulated as follows:
• S_T^j is generated as $s \exp\left(X_T^j\right)$, with $X_T^j = bt + \sigma\sqrt{t}\varepsilon_j$ for a standard Gaussian variable ε_j;
• U_j is generated as a uniform variable on $[0, 1]$;
• $(S_T^*)j = \exp(F^{-1}(U_j; \ln(s), X_T^j))$;

34.2 Barrier options

We now consider the special case of a Lookback option corresponding to a Barrier Up and Out option with payoff function (considering the case of a rebate R paid at T):

$$\psi(X_T, M_T) = \phi(X_T)\mathbb{1}_{\{M_T < L\}} + R\mathbb{1}_{\{M_T \geq L\}}$$

(similar techniques are applicable to the other common forms of barrier options). An approximation of the option price is thus given by

$$e^{-r(T-t)}\mathbb{E}\psi(\bar{X}_T, \bar{M}_T)$$

with $(\bar{M}_T \,|\, \widehat{X}) = \max_{0 \leq i \leq n-1} \widetilde{M}^i$ as before. Now, we have:

$$\begin{aligned} & \mathbb{E}\left(\phi(\bar{X}_T)\mathbb{1}_{\{\bar{M}_T \leq L\}} \,|\, \widehat{X}\right) \\ &= \phi(\widehat{X}_T) \prod_{i=0}^{n-1} \mathbb{P}(\sup_{ih \leq t \leq t_{i+1}} \bar{X}_t \leq L \,|\, \widehat{X}) \\ &= \phi(\widehat{X}_T) \prod_{i=0}^{n-1} F_i(L; \widehat{X}_{t_i}, \widehat{X}_{t_{i+1}}) \, , \end{aligned}$$

with

$$F_i(L;x,y) = \left(1 - \exp(-\frac{2}{\sigma(t_i,x)^2 h_i}(L-x)(L-y))\right) \mathbb{1}_{L>x\vee y} \,.$$

Likewise,

$$\mathbb{E}\left(R\mathbb{1}_{\{\bar{M}_T>L\}}\right)(\widehat{X}) = R\left(1 - \prod_{i=0}^{n-1} F_i(L;\widehat{X}_{t_i},\widehat{X}_{t_{i+1}})\right) \,.$$

Therefore the desired approximation writes

$$= R + \mathbb{E}\left[\left(\phi(\widehat{X}_T) - R\right)\prod_{i=0}^{n-1} F_i(L;\widehat{X}_{t_i},\widehat{X}_{t_{i+1}})\right].$$

In this case, no random draws are needed other than those used for simulating $\widehat{X}_T$, namely n random draws by simulation run — where n can be taken equal to one, in the special case of the Black–Scholes model.

34.3 Asian options

We next consider (European) Asian options with payoffs of the form $\phi(S_T, I_T)$ with $I_t = \int_0^T S_u\,du$ (e.g., $\phi(x,y) = (\frac{I}{T} - K)^+$ in the case of a *fixed strike Asian call*), on a Black–Scholes underlying S.

Straightforward application of the Euler scheme suggests to approximate I_T by the Riemann sum $\widehat{I}_T^1 = \sum_{i=0}^{n-1} h_i \widehat{S}_{t_i}$. But this discretization works poorly in practice. A better discretization is given by the trapezoid rule $\widehat{I}_T^2 = \sum_{i=0}^{n-1} h_i \frac{\widehat{S}_{t_i}+\widehat{S}_{t_{i+1}}}{2}$. However one can show by appropriate Taylor expansions that this is tantamount to approximating

$$\mathbb{E}\phi\left(\bar{S}_T, \bar{I}_T\right) = \mathbb{E}\left[\mathbb{E}\left[\phi\left(\bar{S}_T, \bar{I}_T\right)|\widehat{S}\right]\right] \tag{242}$$

by $\mathbb{E}\left[\phi\left(\bar{S}_T, \mathbb{E}\left[\bar{I}_T|\widehat{S}\right]\right)\right]$, where we set $\bar{I}_t = \int_0^T \bar{S}_u\,du$. To eliminate the related bias (due to the non linearity of ϕ), a better alternative (see Lapeyre and Temam [125]) is to compute directly the r.h.s. in (242), which can be done by noting that, conditionally on $\widehat{S}$, $\bar{I}_T$ is well approximated by

$$\begin{aligned}
&\sum_{i=0}^{n-1} \widehat{S}_{t_i} \int_{t_i}^{t_{i+1}} \left(1 + \kappa(t-t_i) + \sigma(t_i, \widehat{S}_{t_i}) B_t^i dt\right) \\
&\qquad \approx \sum_{i=0}^{n-1} h_i \widehat{S}_{t_i} \left(1 + \frac{\kappa h_i}{2} + \frac{\sigma(t_i,\widehat{S}_{t_i})}{h_i}\int_{t_i}^{t_{i+1}} B_t^i dt\right) =: \widehat{I}_T^3 \,,
\end{aligned}$$

where B^i is a Brownian bridge between (t_i, W_{t_i}) and $(t_{i+1}, W_{t_{i+1}})$, so that $\int_{t_i}^{t_{i+1}} B_t^i dt =: \varepsilon_i$ is a Gaussian random variable with mean

$$\mathbb{E}\varepsilon_i = \int_{t_i}^{t_{i+1}} \left(W_{t_i} + \frac{(t-t_i)}{h_i}(W_{t_{i+1}} - W_{t_i})\right) dt = \frac{h_i}{2}(W_{t_i} + W_{t_{i+1}})\,,$$

and variance, using the fact that $\mathbb{C}\mathrm{ov}(B_t^i, B_u^i) = (t-t_i)(1 - \frac{u-t_i}{h_i})$ (cf. (234)):

$$\mathbb{V}\mathrm{ar}\varepsilon_i = 2\int_{u=t_i}^{t_{i+1}} \int_{t=t_i}^{u} \mathbb{C}\mathrm{ov}(B_t^i, B_u^i) dt du = 2\int_{v=0}^{h_i} (1 - \frac{v}{h_i})\frac{v^2}{2} dv = \frac{h_i^3}{12}\,.$$

The previous discretizations schemes can be used in conjunction with suitable variance reduction techniques, first introduced on this problem (see Kemna and Vorst [116], Lapeyre and Temam [125]). Thus note that the arithmetic average $A_T = \frac{I_T}{T}$ is close to the geometric average $G_T \exp\left(\frac{1}{T}\int_0^T \ln(S_t)\,dt\right)$, for r and σ small. This suggests to choose the modified payoff $\phi(S_T, TG_T)$ as a control variable to price the payoff $\phi(S_T, I_T)$. In the case of the fixed strike Asian call in the Black–Scholes model, we have

$$\phi(S_T, TG_T) = (TG_T - K)^+$$

with G_T log-normally distributed, so that $\mathbb{E}\phi(S_T, TG_T)$ is known explicitly. More precisely, we have

$$G_T = \exp\left(\frac{1}{T}\int_0^T \ln(S_t)\,dt\right) = S_0 \exp\left(\frac{1}{T}\int_0^T (\sigma W_t + bt)\,dt\right) = S_0 \exp\left(\frac{\sigma}{T}\int_0^T W_t\,dt + \frac{bT}{2}\right)$$

with $\mathbb{V}\mathrm{ar}(\int_0^T W_t\,dt) = 2\int_{u=0}^T \int_{t=0}^u t dt du = 2\int_{u=0}^T du = \frac{T^3}{3}$. Therefore

$$TG_T = \widetilde{S}_0 \exp\left(\widetilde{\sigma}\sqrt{T}\varepsilon - \frac{\widetilde{\sigma}^2 T}{2}\right)$$

with $\widetilde{\sigma} = \frac{\sigma}{\sqrt{3}}$, $\widetilde{S}_0 = TS_0 \exp\left(\frac{bT}{2} + \frac{\widetilde{\sigma}^2 T}{2}\right)$. So

$$\mathbb{E}(TG_T - K)^+ = \pi^{bl}(T, K; 0, \widetilde{S}_0, \widetilde{\sigma})\,.$$

34.4 American Options

Recall the generic multinomial (recombining) tree algorithm for pricing American options (cf. (228)): $\Pi_n(j) = \phi(S^j)$ for $j = 1 \dots m$, and then for $i = n-1, \dots, 0$, for $j = 1 \dots m$ (where i and j index the time and space step in the algorithm, respectively):

$$\Pi_i^j = \max\left(\phi\left(S^j\right), e^{-rh}\sum_{l\in\mathbb{Z}} p_l \Pi_{i+1}^{j+l}\right)\,. \tag{243}$$

For pricing an American option by Monte Carlo, a procedure consists in writing the generic dynamic programming equation (243) on a *stochastically generated (hence, non recombining) mesh* $(S_i^j)_{0\le i\le n, 1\le j\le m}$, using an appropriate discretization (Euler,..) scheme for the underlying diffusion. We thus get the following amendment to (243): $\Pi_n^j = \phi(S_n^j)$ for $j = 1 \dots m$, and then for $i = n-1, \dots, 0$, for $j = 1 \dots m$:

$$\Pi_i^j = \max\left(\phi\left(S_i^j\right), e^{-rh}\mathbb{E}_{i,j}\Pi_{i+1}\right) \tag{244}$$

where $\mathbb{E}_{i,j}\Pi_{i+1}$ stands for the conditional expectation of Π_{i+1} given $S_i = S_i^j$. The problem thus reduces to computing conditional expectations (for $i \ge 1$, since for $i = 0$ the conditional expectation reduces to a simple expectation).

Recall that under mild conditions the conditional expectation $\mathbb{E}(X|Y)$ is equal to the Hilbert space projection $\mathbb{EL}(X|Y^0, Y^1, Y^2, \dots)$, or more generally $\mathbb{EL}(X|\varphi^0(Y), \varphi^1(Y), \varphi^2(Y), \dots)$, for a suitable basis $\varphi = (\varphi^l)_{l\in n}$ of the set of the univariate real functions. At step $i \ge 1$

the involved conditional expectations may thus be computed in an elementary way by linear regression of the $(\Pi_{i+1}^j)_{1\leq j\leq m}$ (which are already known at step i of the algorithm) against the $(\varphi^l(S_i^j))_{1\leq j\leq m, 0\leq l\leq q}$ where the integer q is a parameter in the method.

The computational cost of the regression is of the order of $\mathrm{O}(m^2q^2)$, hence an overall computational cost as $\mathrm{O}(nm^2q^2)$. This is obviously too much for typical values of the parameters (e.g., $n = 100, m = 10^5, q = 7$). Fortunately this can be improved in a number of ways, leading to methods perfectly amenable to a practical resolution, for problems of dimension up to 10 or more (whereas deterministic methods are out of scope for dimensions greater than 4).

The previous approach is the one developed in Longstaff and Schwartz [135]. For many alternative methods for computing the conditional expectations: more general *non-parametric regression* methods, Malliavin Calculus methods, quantization methods, etc., we refer the reader to the literature (see Broadie and Glassermann [49, 48], Lions and Régnier [134], Pages and Bally [152], Tsitsiklis and VanRoy [174, 175], Bouchard et al. [39], among many).

Note that a *confidence interval* is not available in this case. It is possible however to derive an upper bound on the price by resorting to a suitable dual Monte Carlo approach (see Rogers [163]). Since most pricing methods provide lower bounds, we thus end up with an interval.

34.5 Adding Jumps

The previous Monte Carlo schemes for Exotic Options in diffusion models, can be extended to more general jump-diffusion models of the form (238), using if need be the related continuous time Euler approximation scheme (see the last paragraph of section 33.4). In the generic market model $\mathcal{Z}$ of section 7.2, a further Continous-Time Markov Chain – like model component is considered. For simulation in this generic model, see Bielecki et al. [32].

35 Backtesting

Before a model may be used in production, it is *backtested*, so the hedging performances of the model are assessed using both simulated trajectories in relevant market models and real data sets. Let us thus consider the problem of discretely delta-hedging at times $t_i = ih, i = 0 \dots n-1$, an *European vanilla call option* with maturity T on an underlying S (cf. section 9.2 and section 13.2.1). To fix ideas, let us assume that the spot obeys the following *objective* Black–Scholes dynamics (under the *statistical* measure $\widehat{\mathbb{P}}$):

$$dS_t = S_t\left(\mu_t dt + \sigma d\widehat{W}_t\right)$$

with constant interest rate r and dividend yield q on S. If a trader were able to hedge continuously, she could perfectly hedge the option and her P&L would be equal to 0. But in practice the trader hedges at discrete times, for example every day at closing price, so that her P&L deviates from 0. More precisely, her discounted P&L at maturity is given by (144), (145) as

$$e^{-rT}e_T = \sum_{i=0}^{n-1} e^{-rih}\delta_i e \tag{245}$$

with

$$e^{-rih}\delta_i e = -(e^{-r(i+1)h}\Pi_{(i+1)h} - e^{-rih}\Pi_{ih}) + \Delta_{ih}e^{-qih}\left(e^{-\kappa(i+1)h}S_{(i+1)h} - e^{-\kappa ih}S_{ih}\right) \quad (246)$$

The behavior of the P&L process may then be assessed numerically by simulation. Assuming constant μ, the solution of the objective Black–Scholes SDE satisfies:

$$S_{(i+1)h} = S_{ih}\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)h + \sigma\left(\widehat{W}_{(i+1)h} - \widehat{W}_{ih}\right)\right) \quad (247)$$

where $\widehat{W}_{(i+1)h} - \widehat{W}_{ih} = \sqrt{h}\varepsilon$ and $\varepsilon \hookrightarrow \mathcal{N}(0,1)$. Thus we can simulate the value of the spot step by step (see section 33.2), multiplying S_{ih} by $\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)h + \sigma\sqrt{h}\varepsilon\right)$ with $\varepsilon \hookrightarrow \mathcal{N}(0,1)$ to get $S_{(i+1)h}$. We then get the related values of $\Pi_{(i+1)h}$ by application of the Black–Scholes formulae, and deduce the corresponding increment of the P&L process by application of formula (246).

In the course of this process, it is interesting to be able to filter some noteworthy paths of the spot, selecting for example spot trajectories passing by a target level at a target time. We can thus observe the behaviour of a pricing routine or a hedging scheme when the spot's trajectory passes at a critical point. For example, in the case of a (reverse) barrier option, we can impose the spot to reach the barrier level at an interesting date like the maturity (recall that the delta of a reverse barrier option explodes at this point).

This can be done with the *Brownian Bridge*, wich may be defined as a centered Gaussian process defined on $[0,1]$ with covariance $\Gamma(s,t) = s(1-t)$ on $s \leq t$ (see also Subsections 33.1 and 34.1). The easiest way to prove that Γ is a covariance is to observe that the process $B_t = \widehat{W}_t - t\widehat{W}_1$ satisfies $\widehat{\mathbb{E}}[B_s B_t] = s(1-t)$ for $s \leq t$. This also gives us a continuous version of the Brownian Bridge. Observe that $B_1 = 0$, so (almost) all the paths go from 0 at time 0 to 0 at time 1, hence the name of this process. Of course the notion of Brownian Bridge may be extended to higher dimensions and to intervals other than $[0,1]$.

In our case, we want the spot trajectory to pass by S_{T_1} at time T_1. For this we would like to replace the Brownian motion in (247) by a Brownian Bridge so that S passes by the desired target point. We thus set:

$$B_{0,x}^{T_1,y}(t) = x + \frac{t}{T_1}(y-x) + \widehat{W}_t - \frac{t}{T_1}\widehat{W}_{T_1}\ ,$$

where T_1 is the target time, and x and y are the Brownian Bridge's starting and target levels, respectively. This gives us a Brownian Bridge between $(0,x)$ and (T_1,y). Since we want:

$$S_{T_1} = S_0\exp\left[\left(\mu - \frac{\sigma^2}{2}\right)T_1 + \sigma B_{0,x}^{T_1,y}(T_1)\right]\ ,$$

we choose x and y such that

$$B_{0,x}^{T_1,y}(0) = x = 0$$
$$B_{0,x}^{T_1,y}(T_1) = y = \frac{1}{\sigma}\left(\ln\left(\frac{S_{T_1}}{S_0}\right) - \left(\mu - \frac{\sigma^2}{2}\right)T_1\right)\ .$$

We finally obtain

$$S_{t+h} = S_t\exp\left[\left(\mu - \frac{\sigma^2}{2}\right)h + \sigma\left(B_{0,x}^{T_1,y}(t+h) - B_{0,x}^{T_1,y}(t)\right)\right]\ ,$$

with x and y thus determined. All we need now is to simulate the quantity

$$\Gamma = B_{0,x}^{T_1,y}(t+h) - B_{0,x}^{T_1,y}(t) \ .$$

After some computations, we find that

$$\Gamma \hookrightarrow N\left(\left(y - B_{0,x}^{T_1,y}(t)\right)\frac{h}{T_1 - t}; h\left(1 - \frac{h}{T_1 - t}\right)\right) \ .$$

This gives us the formula for the next value of the spot:

$$S_{t+h} = S_t \exp\left[(\mu - \frac{\sigma^2}{2})h + \sigma\Big((y - B_{0,x}^{T_1,y}(t))\frac{h}{T_1 - t} + \sqrt{\frac{h(T_1 - t - h)}{T_1 - t}}\varepsilon\Big)\right] \ , \qquad (248)$$

where $\varepsilon \hookrightarrow N(0,1)$. Note that formula (248) is only applicable for $t \leq T_1$; for $(t > T_1)$ we have to use formula (247).

We may thus simulate a number of spot trajectories with these formulas at discrete times, and calculate for each path the corresponding P&L given by the discrete hedging formulas (245)–(246). We may then compute pertaining statistics, like the mean, the square deviation or further moments of e_T[6]. We may also identify some relevant spot's trajectories, like those generating extremal or median P&L positions (see Figures 13 and 14). Such dynamic tests allow risk managers or traders to assess the performance of a hedging scheme, and they may help developers in detecting some problems of a pricing routine behaviour.

[6] The standard deviation of the P&L at maturity of a daily rebalanced delta-hedged vanilla option position in the Black–Scholes model is commonly found to be of the order of 1% of the initial option premium Π_0 (see, e.g., [151]).

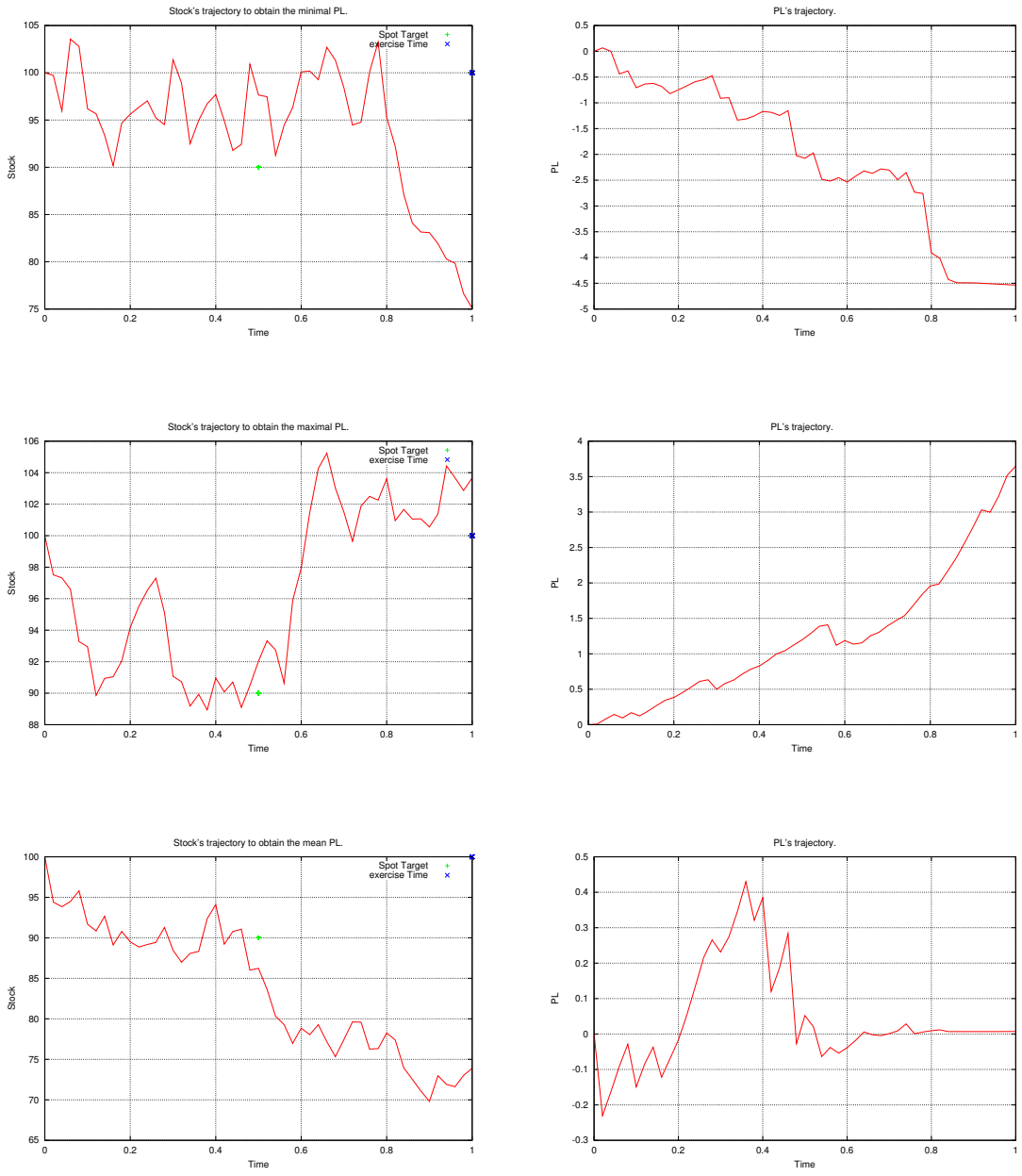

Figure 13: *Weekly hedging P&L simulations with Brownian Bridge at level 90 at* $t = \frac{1}{2}$.

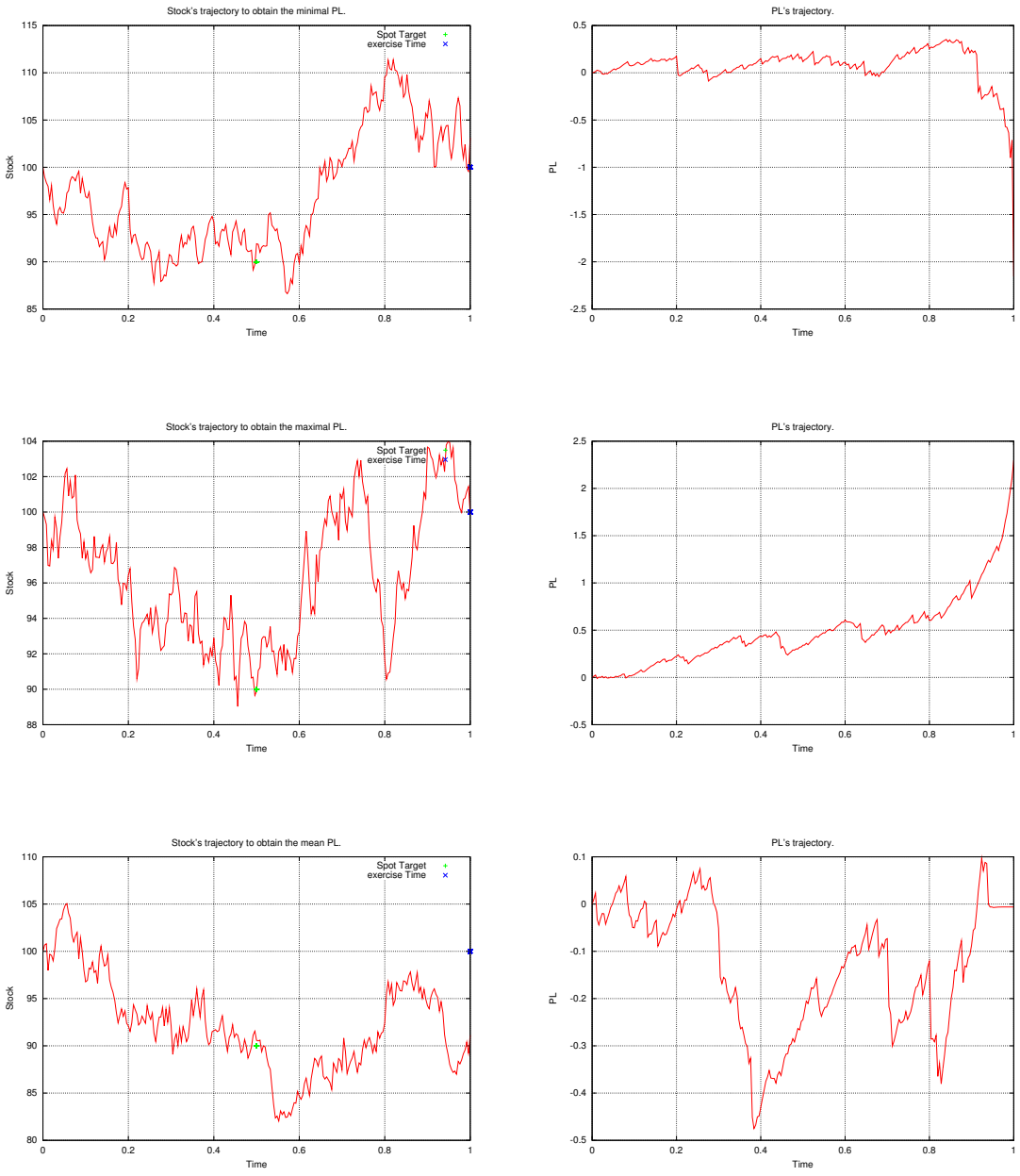

Figure 14: *Daily hedging P&L simulations with Brownian Bridge at level 90 at* $t = \frac{1}{2}$.

Part VII
Calibration Methods

An important issue in quantitative finance is *model calibration*. In financial modelling, calibrating a (class of) model(s) means finding numerical values of its parameters such that the related instance of the model is consistent with the market, in the sense that the prices of some financial instruments *at the current time*, or *calibration input data*, are the same in the market and in the model thus calibrated, or, practically speaking, that the model fits the currently observed market prices within the bid-ask spread.
The calibration problem is thus the *inverse* of the pricing problem. Instead of computing prices in a model with given values for its parameters, one wishes to compute the values of the model parameters that are consistent with observed prices (up to the bid–ask spread).

The simplest example of a calibration problem was encountered in Section 13, when we discussed the notions of implied volatility of an option or implied correlation of a CDO tranche. These problems can indeed be interpreted as the calibration problem of a Black–Scholes model or a Li model, using an observed option price or a CDO tranche market spread as calibration input data. Of course in these cases the calibration problem is very easy, since there is only one parameter to calibrate to exactly one market data, so the task is easily done by dichotomy.

36 The ill-posed Inverse Calibration Problem

Since the calibration input data are typically *derivative prices* given by (risk-neutral) *expectations* of the related payoffs (see Part II), the calibration problem can be seen as a *moment problem*. Physicists would call this problem, the *estimation* of the model. However, in finance, the term 'estimation' specifically refers to *statistical estimation* of the model, namely the estimation of the model parameters using *historical data*, by Maximum likelihood estimation (MLE) or any other statistical procedure. So, in finance, estimation is a *backward-looking* process (using historical data), whereas calibration is said to be a *forward-looking* process, referring to the fact that derivative prices at the current time are based on the *anticipations* of the market regarding the dynamics of the underlying in the future.
It is generally acknowledged that whenever option data are available, it is better to use them to *calibrate* the model, rather than to *estimate* the model on past data. Of course, in the absence of calibration data (no observed prices), one has no other means than to estimate the model statistically, but this is another story.

Now, it is well-known by physicists that such inverse problems are typically *ill-posed*. Recall that a problem is *well-posed* (as defined by Hadamard) if its solution exists, is unique, and depends continuously on its input data. Thus there are three reasons for which a problem might be ill-posed:
• it admits no solution, or/and
• it admits more than one solution, or/and
• the solution(s) of the inverse problem do(es) not depend on the input data in a continuous way.
In the case of calibration problems in finance, except for trivial situations, there exists typically *no instance* of a given class of models which is *exactly* consistent with a full calibration

data set, including a number of option prices, a zero-coupons curve, an expected dividend yield curve on the underlying, etc. But there are often *various* instances of a given class of models that fit the data *within the bid–ask spread*. In this case, if one perturbs the data (e.g., if the observed prices move from some small amount between today and tomorrow), it is quite typical that a numerically determined best fit solution of the calibration problem switches from one 'basin of attraction' to the other, thus the numerically determined solution is *not stable* either.

Recall that option prices (or pricing functions in Markovian models, more precisely) are solutions of PIDEs with model parameters related coefficients (see Part II). Thus model parameters can be expressed in terms of partial derivatives of the option prices with respect to the model factors. Therefore, calibrating the model is tantamount to numerically differentiating derivatives pricing functions. But *numerical differentiation* is the canonical example of an ill-posed problem, see, e.g., [82] (so two functions may be arbitrarily close to one another in sup norm, though their derivatives differ significantly).

In order to get a well-posed problem, we need to introduce some *regularization*. The most widely known and applicable regularization method is *Tikhonov(–Phillips)* regularization method [172, 154, 82].

36.1 Tikhonov regularization of non-linear inverse problems

We consider a Hilbert space $\mathcal{H}$, a closed convex non-void subset $\mathcal{A}$ of $\mathcal{H}$, a direct operator ('pricing functional')

$$\mathcal{H} \supseteq \mathcal{A} \ni a \xrightarrow{\Pi} \Pi(a) \in \mathbb{R}^d ,$$

(so a corresponds to the set of model parameters), noisy data ('observed prices') π^δ, and a *prior* $a_0 \in \mathcal{H}$ (a priori guess for a). The Tikhonov regularization method for *inverting* Π *at* π^δ, or estimating the model parameter a given the observation π^δ, consists in:
• reformulating the inverse problem as the following *nonlinear least squares problem*:

$$\boxed{\min_{a\in\mathcal{A}} \left\|\Pi(a) - \pi^\delta\right\|^2} \tag{249}$$

to ensure *existence* of a solution,
• selecting the solutions of the previous nonlinear least squares problem that minimize $\|a - a_0\|^2$ over the set of all solutions, and
• introducing a trade-off between accuracy and regularity, parameterized by a level of regularization $\alpha > 0$, to ensure *stability*.
More precisely, we introduce the following *cost criterion*:

$$J_\alpha^\delta(a) \equiv \left\|\Pi(a) - \pi^\delta\right\|^2 + \alpha \|a - a_0\|^2 \ . \tag{250}$$

Given α, δ and a further parameter η, where η represents an error tolerance on the minimization, we define a *regularized solution to the inverse problem for* Π *at* π^δ, as any model parameter $a_\alpha^{\delta,\eta} \in \mathcal{A}$ such that

$$J_\alpha^\delta\left(a_\alpha^{\delta,\eta}\right) \leq J_\alpha^\delta(a) + \eta \quad , \quad a \in \mathcal{A} .$$

Under suitable assumptions, one can show that the regularized inverse problem is well-posed, as follows. We first postulate that the direct operator Π satisfies the following regularity

assumption.

Assumption 36.1 (Compactness) $\Pi(a_n)$ converges to $\Pi(a)$ in $\mathbb{R}^d$ if a_n weakly-converges to a in $\mathcal{H}$.

We then have the following *stability* result.

Theorem 36.1 (Stability) *Let $\pi^{\delta_n} \to \pi^{\delta}$, $\eta_n \to 0$ when $n \to \infty$. Then any sequence of regularized solutions $a_\alpha^{\delta_n,\eta_n}$ admits a subsequence which converges towards a regularized solution $a_\alpha^{\delta,\eta=0}$.*

Assuming further that the data lie in the range of the model leads to *convergence* properties of regularized solutions to (unregularized) solutions of the inverse problem as $\alpha \to 0$. Let us then make the following additional assumption on Π.

Assumption 36.2 (Range property) $\pi \in \Pi(\mathcal{A})$.

By an *a_0 – solution* to the inverse problem for Π at π, we mean any $a \in \underset{\{\Pi(a)=\pi\}}{\operatorname{Argmin}} \|a - a_0\|$. Note that the set of a_0-solutions is non-empty, by Assumption 36.2.

Theorem 36.2 (Convergence; see, for instance, Theorem 2.3 of Engl *et al* [83]) *Let the perturbed parameters $\alpha_n, \delta_n, \eta_n$ and the perturbed data $\pi_n \in \mathbb{R}^d$ satisfy*

$$(n \in \mathbb{N}) \quad \|\pi - \pi_n\| \leq \delta_n,$$

$$(n \to \infty) \quad \alpha_n \quad , \quad \delta_n^2/\alpha_n \quad , \quad \eta_n/\alpha_n \quad \longrightarrow \quad 0.$$

Then any sequence of regularized solutions $a_{\alpha_n}^{\delta_n,\eta_n}$ admits a subsequence which converges towards an a_0-solution a of the inverse problem for Π at π. In particular, in case when this problem admits a unique a_0-solution a, then $a_{\alpha_n}^{\delta_n,\eta_n}$ converges to a.

Remark 36.3 In the special case where the direct operator Π is linear, Tikhonov regularization thus appears as an approximating scheme for the pseudo-inverse of Π.

Finally, assuming further regularity of Π, one can get *convergence rates* estimates, uniform over all data $\pi \in \Pi(\mathcal{A})$ sufficiently close and smooth with respect to the prior a_0 (so that the additional *source condition* (251) is satisfied). Let us thus make the following additional assumption on Π.

Assumption 36.4 (Twice Gateaux differentiability) There exists linear and bilinear forms $d\Pi(a)$ on $\mathcal{H}$ and $d^2\Pi(a)$ on $\mathcal{H}^2$ such that

$$\Pi(a + \varepsilon h) = \Pi(a) + \varepsilon d\Pi(a) \cdot h + \tfrac{\varepsilon^2}{2} d^2\Pi(a) \cdot (h,h) + o(\varepsilon^2) \quad ; \quad a, a+h \in \mathcal{A}$$
$$\|d\Pi(a) \cdot h\| \leq C \|h\| \quad , \quad \left\|d^2\Pi(a) \cdot (h,h')\right\| \leq C \|h\| \|h'\| \quad ; \quad a \in \mathcal{A}\,, \quad h, h' \in \mathcal{H}$$

where C is a constant independant of $a \in \mathcal{A}$.

In the following theorem the operator

$$d\Pi(a)^* : \mathbb{R}^d \ni \lambda \mapsto d\Pi(a)^* \lambda \in \mathcal{H}^1$$

denotes the *adjoint* of

$$d\Pi(a) : \mathcal{H}^1 \ni h \mapsto d\Pi(a) h \in \mathbb{R}^d,$$

in the sense that (see [82]):

$$\langle h, d\Pi(a)^\star \lambda \rangle_{\mathcal{H}^1} \;=\; \lambda' d\Pi(a).h \;; \quad (h, \lambda) \in \mathcal{H}^1 \times \mathbb{R}^d \,.$$

Theorem 36.3 (Convergence Rates; see, for instance, Theorem 10.4 of Engl *et al* [82]) *Assume*

$$(n \in \mathbb{N}) \quad \|\pi - \pi_n\| \leq \delta_n,$$

$$(n \to \infty) \quad \alpha_n \longrightarrow 0 \quad , \quad \alpha_n \sim \delta_n \quad , \quad \eta_n = \mathrm{O}\left(\delta_n^2\right) \,.$$

Then $\|a_{\alpha_n}^{\delta_n,\eta_n} - a\| = \mathrm{O}(\sqrt{\delta_n})$, *for any* a_0*-solution* a *of the inverse problem for* Π *at* π *such that*

$$a - a_0 = d\Pi(a)^* \lambda \tag{251}$$

for some λ *sufficiently small in* $\mathbb{R}^d$ *(in particular, there exists at most one such* a_0*-solution* a*).*

Remark 36.5 An interesting feature of Tikhonov regularization is that the data set π does not need to belong to the range of the direct operator for applicability of the method — even if Assumption 36.2 is the simplest assumption for the previous results regarding convergence and convergence rates (in fact a minimal assumption for such results is the existence of a least squares solution to the inverse problem, see Proposition 3.2 of Binder *et al* [35]).

An important issue in practice is the choice of the *regularization parameter* α, that determines the trade-off between accuracy and regularity in the method. To set α, the two main approaches are:
• *a priori* methods, in which the choice of α only depends on δ, the level of noise on the data (such as the size of the bid–ask spread, in the case of market prices data in finance);
• more general *a posteriori* methods, in which α may depend on the data in a less specific way.
In applications to calibration problems in finance, the most commonly used method for choosing α is the a posteriori method based on the so-called *discrepancy principle*, which consists in choosing the greatest level of α for which the 'distance' $\left\|\Pi\left(a_\alpha^{\delta,\eta}\right) - \pi^\delta\right\|$ (for given δ, η) does not exceed the level of noise δ on the observations (as measured by the bid-ask spread).

36.2 Nonlinear Optimization

In the case of *parametric* models in finance, namely models with a small number of *scalar* parameters, such as the Heston model, the Merton model, etc (as opposed to models with *functional*, e.g., time-dependent, parameters), the choice of a suitable regularization term

is generally not obvious. In this case, the calibration industry standard rather consists in solving the unregularized non linear least squares problem (249). So Tikhonov regularization is rather used for calibrating *non parametric* financial models.
Thus, in practice, calibration problems are essentially reduced to nonlinear minimization problems (non linear least squares problems, typically with Tikhonov regularization, at least for calibrating *non parametric models* with functional coefficients), on suitable closed non-void convex subsets $\mathcal{A}$ of related Hilbert spaces $\mathcal{H}$.

Recall that a real function J on $\mathcal{A}$ (the *cost criterion* J, that will typically be given by a regularized non linear least squares criterion $J = J_\alpha^\delta$ in (250)), is said to be *lower semi-continuous* at $a \in \mathcal{A}$ iff 'it cannot exceed its limits at a', i.e. $J(a) \leq \liminf_a J$. The following theorem extends to (infinite dimensional, presumably) Hilbert spaces, under an additional convexity assumption, the well-known fact that a lower semi-continuous function on a compact subset of $\mathbb{R}^k$ admits a global minimum.

Theorem 36.4 *If J is lower semi-continuous,* convex*, and goes to infinity as a goes to infinity in $\mathcal{A}$, then J admits a global minimum on $\mathcal{A}$. Moreover, if J is strictly convex on $\mathcal{A}$, this minimum is unique.*

Of course, when it comes to implementation, the minimization problem (250) is discretized, thus becoming effectively a *nonlinear minimization problem* on (some subset of) $\mathbb{R}^k$ (see, e.g., [149]), where k is the number of model parameters to be estimated.

In the case of a *strictly convex* cost criterion J in (250), and if, additionally, J is differentiable, one can prove the convergence to the (unique) minimum of various *gradient descent algorithms*. These consist in moving at each step from some amount (fixed step descent *vs* optimal step descent) in a direction defined by the gradient ∇J at the current step of the algorithm, in combination with, in some variants of the method *(conjugate gradient method, quasi-Newton algorithms, etc)*, the gradient(s) ∇J at the previous step(s).

In the *non strictly convex* case, (actually, in the context of calibration problems in finance, J is typically not even convex w.r.t. a), or if the cost criterion is only almost everywhere differentiable (as in the *American calibration problem* , see section 38.2), such algorithms can still be used, in which case they typically converge to one among many *local minima* of J.
When there are no constraints (case $\mathcal{A} = \mathcal{H}$), the minimization problem is, in practice, much easier, and many implementations of the related gradient descent algorithms are available (see for instance [156]). As for constrained problems, a state-of-the-art open-source implementation of the quasi-Newton method for minimizing a function on a box, the lbfgs algorithm, is available on `www.ece.northwestern.edu/~nocedal/lbfgsb.html`.

When the gradient ∇J is not computable in closed form, and not computable numerically with the required accuracy either, an alternative to gradient descent methods is to use the *nonlinear simplex method* (not to be confused with the simplex algorithm for solving linear programming problems, see [156]). As opposed to gradient descent methods, the nonlinear simplex algorithm only uses the *values* (and not the *gradient*) of J, but the convergence of the algorithm is not proved in general, and there are known counter-examples in which it does not converge.

37 A method using the Characteristic Function for European Vanillas

Calibrating a model typically involves massive computation of vanilla option prices for various strikes, maturities, and numerical sets of model parameters. Thus for calibration purposes it is essential to have efficient vanilla pricing methods. In many models, such methods are provided by Fourier calculus.

37.1 Fourier Transform Miscellanea

Recall that the *Fourier transform* $\mathcal{F}f$ of an absolutely integrable function f from $\mathbb{R}$ to itself is defined by, for $u \in \mathbb{R}$:

$$\boxed{\mathcal{F}f(u) = \int_{-\infty}^{\infty} e^{iux} f(x)dx}$$

Remark 37.1 The *characteristic function* of a random variable X with law of density p, defined by:

$$\boxed{\Phi(u) = \mathbb{E}[\exp(iuX)] = \int_{-\infty}^{\infty} e^{iux} p(x)dx \ , u \in \mathbb{R}}$$

is thus given by the Fourier transform $\mathcal{F}p$.

The differentiation operator translates into multiplication by iu in the Fourier space:

$$\boxed{\mathcal{F}f^{(m)}(u) = (-iu)^m \mathcal{F}f(u) \ , u \in \mathbb{R}}$$

for differentiable f.

For *regular* f, the *inverse Fourier transform formula* writes, for $x \in \mathbb{R}$:

$$\boxed{f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \mathcal{F}f(u)du} \tag{252}$$

Moreover, the Fourier transform $\mathcal{F}f$ may be extended to complex values of its argument u (resulting on the so-called *complex Fourier transform*, still denoted $\mathcal{F}f$, of f), for u in suitable *strips of analyticity* of $\mathcal{F}f$ parallel to the real axis. Resorting to this notion of complex Fourier transform, one can show that the law of any Fourier-integrable random variable X, with characteristic function denoted by Φ, admits a *weak density* p formally given by, for any $\eta > 0$:

$$\boxed{p(x) = \frac{1}{2\pi} \int_{-\eta i-\infty}^{-\eta i+\infty} e^{-iux} \Phi(u)du} \tag{253}$$

meaning that for any regular enough function φ such that $\mathbb{E}\varphi(X)$ is well defined, we have (in the strong sense, now):

$$\boxed{\mathbb{E}\varphi(X) = \frac{1}{2\pi} \int_{u=-\eta i-\infty}^{-\eta i+\infty} \Phi(u) \left(\int_{y=-\infty}^{\infty} e^{-iuy} \varphi(y)dy \right) du} \tag{254}$$

In particular, in case where the r.v. X has a density in the usual sense, one may apply (254) to $\varphi(y) = \mathbb{1}_{y>x}$, for any $x \in \mathbb{R}$. Therefore

$$\begin{aligned} G(x) &= \mathbb{P}\Big(X > x\Big) \\ &= \frac{1}{2\pi} \int_{u=-\eta i-\infty}^{-\eta i+\infty} \Phi(u) \left(\int_{y=x}^{\infty} e^{-iuy} dy \right) du = \frac{1}{2\pi i} \int_{u=-\eta i-\infty}^{-\eta i+\infty} \frac{e^{-iux}}{u} \Phi(u)\, du \ . \end{aligned}$$

Sending $\eta \to 0+$, one can then show by application of the Cauchy residue formula (see, e.g., Titchmarsh [173] or Lewis [132]; note that the integrals in (255) are only defined as principal values):

$$G(x) = \frac{1}{2} + \frac{1}{2\pi} \lim_{\varepsilon \to 0+} \int_{|u|>\varepsilon} \frac{e^{-iux}\Phi(u)}{iu}\, du\ . \tag{255}$$

Since $\frac{e^{iux}\Phi(-u)}{-iu}$ is the conjugate of $\frac{e^{-iux}\Phi(u)}{iu}$, one has:

$$\lim_{\varepsilon \to 0+} \int_{|u|>\varepsilon} \frac{e^{-iux}\Phi(u)}{iu}\, du = 2 \lim_{\varepsilon \to 0+} \int_{\varepsilon}^{\infty} Re\left[\frac{e^{-iux}\Phi(u)}{iu}\right] du \quad .$$

Hence,

$$\boxed{F(x) = 1 - G(x) = \tfrac{1}{2} + \tfrac{1}{\pi} \int_0^{\infty} Re\left[\frac{ie^{-iux}\Phi(u)}{u}\right]} \tag{256}$$

37.2 Option Pricing by Fourier Transform

Let us now consider the problem of valuing a European vanilla call of maturity T and strike K. For simplicity we assume deterministic (constant, say) riskless interest rate r and dividend yield q on the underlying S. We denote as usual $\kappa = r - q$. Of course in a context of stochastic interest rates the subsequent results admit obvious amendments relative to a T – forward neutral model of T – future price F on S (instead of a risk-neutral model on S, in the text below; cf. section 11.1).

Theorem 37.1 *Whenever a risk-neutral law of $x_T = \ln(S_T)$ has a density, the related call value at time 0, $C_0 = \mathbb{E}e^{-rT}(S_T - K)^+$, is given by:*

$$\boxed{C_0 = S_0 e^{-qT}\Pi_1 - Ke^{-rT}\Pi_2} \tag{257}$$

where the pseudo-probabilities *Π_1 and Π_2 are given in terms of the characteristic function $\Phi_T(u) = \mathbb{E}[\exp(iux_T)]$ as:*

$$\boxed{\begin{cases} \Pi_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} Re\left[\frac{e^{-iuk}\Phi_T(u-i)}{iu\Phi_T(-i)}\right] du \\ \Pi_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} Re\left[\frac{e^{-iuk}\Phi_T(u)}{iu}\right] du \end{cases}} \tag{258}$$

(with in the first line $\Phi_T(-i) = \mathbb{E}S_T = S_0 e^{\kappa T}$, by arbitrage).

Proof. In order to establish (257), we first decompose

$$C_0 = \mathbb{E}\left(e^{-rT}e^{x_T}\mathbb{1}_{\{x_T>k\}}\right) - Ke^{-rT}\mathbb{P}(x_T > k)\ ,$$

where $k = \ln K$. Now, by (256),

$$\Pi_2 = \mathbb{P}\left(x_T > k\right) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} Re\left[\frac{e^{-iuk}\Phi_T(u)}{iu}\right] du.$$

Let us introduce the probability measure $\widetilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ defined by $\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \frac{e^{x_T}}{\mathbb{E}(e^{x_T})} (= \frac{S_T}{S_0 e^{\kappa T}})$. So

$$\begin{aligned}\mathbb{E}(e^{-rT} e^{x_T} \mathbb{1}_{\{x_T > k\}}) &= \int_{x \in \mathbb{R}} e^{-rT} e^{x} \mathbb{1}_{\{x > k\}} d\mathbb{P}\{x_T = x\} \\ &= \mathbb{E}(e^{-rT} e^{x_T}) \widetilde{\mathbb{P}}(x_T > k) = S_0 e^{-qT} \widetilde{\mathbb{P}}(x_T > k) \ .\end{aligned}$$

The characteristic function of x_T under $\widetilde{\mathbb{P}}$ is given by:

$$\widetilde{\mathbb{E}}(e^{iux_T}) = \int_{x \in \mathbb{R}} e^{iux} d\widetilde{\mathbb{P}}\{x_T = x\} = \frac{\mathbb{E}(e^{x_T} e^{iux_T})}{\mathbb{E}(e^{x_T})} = \frac{\Phi_T(u - i)}{\Phi_T(-i)}.$$

Hence, by (256) again,

$$\Pi_1 = \widetilde{\mathbb{P}}\Big(x_T > k\Big) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re\left[\frac{e^{-iuk} \Phi_T(u - i)}{iu \Phi_T(-i)}\right] du \ .$$

□

37.3 Derivation of the delta in the case of homogenous models

Homogenous models mean models in which the European vanilla option prices are degree one-homogenous with respect to the pair (S_0, K) (for given values of the remaining parameters and risk factors in the model), so:

$$\boxed{C_0(T, \alpha K; \alpha S_0) = \alpha C_0(T, K; S_0) \ , \ \alpha > 0}$$

or, equivalently:

$$\boxed{C_0(T, K; S_0) = S_0 \partial_S C_0(T, K; S_0) + K \partial_K C_0(T, K; S_0)} \tag{259}$$

Note that in general

$$\partial_K C_0 = e^{-rT} \partial_K \mathbb{E}(S_T - K)^+ = -e^{-rT} \mathbb{E} \mathbb{1}_{\{S_T > K\}} = -e^{-rT} \mathbb{P}(\{S_T > K\}) = -e^{-rT} \Pi_2$$

(where Π_2 was defined in (258)). So in a homogenous model, using also (257):

$$S_0 \partial_S C_0(T, K; S_0) = C_0(T, K; S_0) - K \partial_K C_0(T, K; S_0) = S_0 e^{-qT} \Pi_1 \ .$$

Thus

$$\boxed{\Delta_0 = \partial_S C(T, K; 0, S_0) = e^{-qT} \Pi_1}$$

37.4 Numerical Algorithm

In many models the characteristic function Φ_T is known and computable. This is for instance the case in all AJD models [75] (see section 10.5). Knowing Φ_T, (257) enables one to compute C_0 numerically by quadrature.

If $\Phi_T(u)$ is computable, both Π_1 and Π_2 can be computed by plugging the expression for $\Phi_T(u)$ in (258), and discretizing the related integrals by the trapezoid method. More

precisely, we choose a step h (e.g. $h = 0.01$) and a time mesh $0 = x_0 < x_1 <, \ldots, < x_{\frac{M}{h}} = M$ (e.g. $M = 100$) such that $x_{j+1} - x_j = h$. Denoting by f either integrand in (258), one approximates $\int_0^M f(u)\,du$ by $\frac{h}{2}\left(f(x_0) + f(x_{\frac{M}{h}}) + 2\sum_{j=1}^{\frac{M}{h}-1} f(x_j)\right)$.

For implementation details, including the use of more efficient *Gaussian quadratures* rather than the simple trapezoid method above, as well as on numerical issues related to the evaluation of the multi-valued complex functions involved in the integrands, we refer the reader to Kahl and Jackel [113].

37.5 An alternative Formula

Unfortunately, FFT cannot be used to evaluate the integrals in (258), due to the singularity of the integrands at $u = 0$. We now present an alternative formula, due to Carr and Madan [52], which is amenable to evaluation by the FFT. A further advantage is that it allows one to compute the prices of *a whole family* of options with various strikes and maturity T at time 0, which is precisely what is required for calibration purposes (for other alternative approaches, see [52, 57]).

The idea is to compute the Fourier transform of the call price viewed as a function $C(k)$ of the log-strike $k = \ln K$. However the function $C(k)$ is not integrable (since $\lim_{k\to-\infty} C(k) = e^{-rT}\mathbb{E}S_T = S_0 e^{-qT} > 0$). We thus define the *modified price* $c(k) = e^{\alpha k}C(k)$, for a fixed $\alpha > 0$. The modified price is integrable, and we have, for any $u \in \mathbb{R}$:

$$\begin{aligned}
\mathcal{F}c(u) &= \int_{-\infty}^{\infty} e^{iku}c(k)dk = \int_{-\infty}^{\infty} e^{(\alpha+iu)k}C(k)dk \\
&= e^{-rT}\int_{-\infty}^{\infty} e^{(\alpha+iu)k}\left(\int_{-\infty}^{\infty}\left(e^x - e^k\right)^+ p_T(x)dx\right)dk \\
&= e^{-rT}\int_{-\infty}^{\infty} p_T(x)\left(\int_{-\infty}^{x} e^{(\alpha+iu)k}\left(e^x - e^k\right)dk\right)dx \\
&= e^{-rT}\int_{-\infty}^{\infty} p_T(x)\left(\frac{e^x}{\alpha+iu}\left[e^{(\alpha+iu)k}\right]_{-\infty}^{x}\right. \\
&\qquad \left. -\frac{1}{\alpha+iu+1}\left[e^{(\alpha+iu+1)k}\right]_{-\infty}^{x}\right)dx \\
&= e^{-rT}\int_{-\infty}^{\infty} p_T(x)\left(\frac{e^x e^{(\alpha+iu)x}}{\alpha+iu} - \frac{e^{(\alpha+iu+1)x}}{\alpha+iu+1}\right)dx\,,
\end{aligned}$$

where the last equality follows from the fact that $\lim_{k\to-\infty} e^{(a+iu)k} = 0$, for any $a > 0$ and $u \in \mathbb{R}$. Hence

$$\mathcal{F}c(u) = e^{-rT}\int_{-\infty}^{\infty} p_T(x)\frac{e^{(\alpha+iu+1)x}}{(\alpha+iu)(\alpha+iu+1)}dx = \frac{e^{-rT}\Phi_T(u-(\alpha+1)i)}{(\alpha+iu)(\alpha+iu+1)} \tag{260}$$

where Φ_T is the characteristic function of $x_T = \ln(S_T)$ (assumed well-defined).

The call price function $C(k)$ may then be retrieved numerically by discrete Fourier transform.

Indeed, we have by the inverse Fourier transform formula (252):

$$C(k) = e^{-\alpha k} c(k) \;=\; \frac{e^{-\alpha k}}{2\pi} \int_{-\infty}^{\infty} e^{-iku} \mathcal{F}c(u) du$$
$$= \frac{e^{-\alpha k}}{\pi} Re \left[\int_0^{\infty} e^{-iku} \mathcal{F}c(u) du \right] ,$$

where the second equality holds because the function $c(k)$ is real, which implies that $\mathcal{F}c(-u) = \overline{\mathcal{F}c(u)}$, for any $u \in \mathbb{R}$. We thus get by numerical integration using Simpson's rule

$$C(k) \;\approx\; \frac{e^{-\alpha k}}{\pi} \sum_{j=0}^{N-1} e^{-iku_j} \mathcal{F}c(u_j) w_j \tag{261}$$

with N even (for validity of Simpson's integration rule) and for $j = 0 \ldots N-1$:

$$u_j = jh \;,\; w_j = \frac{h}{3} \big(3 + (-1)^{j+1} - \mathbb{1}_{j=0 \text{ or } N-1} \big) \;.$$

Here the motivation for using Simpson's integration rule is to have a good accuracy for a relatively small value of N. In particular, for k of the form $k_n = \underline{k} + \frac{2\pi n}{Nh}$ where $\underline{k}$ will be fixed later, we have:

$$ku_j = k_n u_j = \underline{k} u_j + \frac{2\pi n}{N} j \;.$$

Plugging this into (261), we get

$$C(k_n) \;\approx\; \frac{e^{-\alpha k_n}}{\pi} \sum_{j=0}^{N-1} e^{-2\pi i \frac{jn}{N}} e^{-i\underline{k} u_j} \mathcal{F}c(u_j) w_j \;.$$

In the last sum, we recognize the *discrete Fourier transform* of

$$f = \big(e^{-i\underline{k} u_j} \mathcal{F}c(u_j) w_j \big)_{0 \le j \le N-1}$$

at the evaluation point n. Recall that the discrete Fourier transform $(Ff_n)_{0 \le n \le N-1}$ of a vector $f = (f_j)_{0 \le j \le N-1}$ writes:

$$Ff_n = \sum_{j=0}^{N-1} e^{-2i\pi \frac{jn}{N}} f_j \;,\; 0 \le n \le N-1 \;.$$

Choosing $\underline{k} = \ln(S_0) + \kappa T - \frac{\pi}{h}$, we can thus price a call for N values of the strike K distributed around the T-forward value of the stock $F_0 = S_0 e^{\kappa T}$ by computing the discrete Fourier transform of an explicitly known function, which provided N is a power of 2 can be made at cost $\mathrm{O}(N \ln N)$, by FFT.

This approach is successfully tested by Carr and Madan [52] in a Variance Gamma model using the following method parameters: $N = 4096$, $h = 0.25$, $\alpha = 1.5$. As for α, a general recommendation in view of (260), (261) is to choose α such that

$$\Phi_T(-(\alpha+1)i) = \mathbb{E} S_T^{\alpha+1} < +\infty \;.$$

38 Extracting Effective Volatility

In this Section we consider the problem of inferring an *effective volatility function* $\sigma(t,S)$ (see Dupire [76]) from observed option prices, namely European vanilla calls and/or puts with various strikes and maturities on the underlying S. The effective volatility function thus inferred may then be used to price exotic options and/or Greeking, consistently with the market (see, for instance, Crépey [62]).

38.1 Local versus Effective Volatility

Local Volatility (LV) models [76] are the straightforward generalization of the Black–Scholes model in which the volatility parameter is not a constant anymore, but is given by a (positively bounded Borel-measurable) function $\sigma = \sigma(t,S_t)$ of t and S_t. As the Black–Scholes model, Local Volatility models are complete, and we have under the related risk-neutral probability measure $\mathbb{P}$:

$$\boxed{dS_t = S_t(\kappa dt + \sigma(t,S_t)dW_t)} \tag{262}$$

for a standard $\mathbb{P}$ – Brownian motion W. Thus European vanilla call options on S have unique Local Volatility arbitrage price processes given by, for $t \in [0,T]$:

$$\Pi_t(T,K) = e^{-r\tau}\mathbb{E}\left((S_T-K)^+ \,|\, S_t\right) = \Pi(T,K;t,S_t,\sigma)\ . \tag{263}$$

We don't have closed pricing formulae in a general Local Volatility model. But the pricing function $\Pi(T,K;t,S,\sigma)$ defined by (263) is the unique $W^{1,2}_{p,loc}$-solution (for $p>2$; a.e. solutions with related partial derivatives in $L_{p,loc}$ in time-space) of the *Black–Scholes equation* in the variables (t,S) :

$$\boxed{\left\{\begin{array}{rcl} -\partial_t\Pi - \kappa S\partial_S\Pi - \frac{1}{2}\sigma(t,S)^2S^2\partial^2_{S^2}\Pi + r\Pi &=& 0, \quad t<T \\ \Pi|_T &=& (S-K)^+ \end{array}\right.} \tag{264}$$

and of the dual *Dupire equation* in the variables (T,K) :

$$\boxed{\left\{\begin{array}{rcl} \partial_T\Pi + \kappa K\partial_K\Pi - \frac{1}{2}\sigma(T,K)^2K^2\partial^2_{K^2}\Pi + q\Pi &=& 0, \quad T>t \\ \Pi|_t &=& (S-K)^+ \end{array}\right.} \tag{265}$$

(see [63, theorem 4.3]).

These equations can be used to compute Local Volatility option prices and Greeks numerically. They also imply that whatever the market risk-neutral price process may be, there is always, at any date t_0, a Local Volatility model (dependent on t_0!) with the same spot marginals as the market risk-neutral price process at t_0. Given a suitable interpolation $\Pi_0 \in W^{1,2}_{p,loc}$ of the set obs_0 of observed European vanilla call prices $\pi_0 \equiv \{\pi_0(T,K)\,;\,(T,K)\in \text{obs}_0\}$ at time t_0, this 'tangent diffusion process' of the market risk-neutral price process corresponds to the volatility function given by the following *Dupire's formula* [76], for any $(T,K)\in[t_0,\infty)\times(0,\infty)$:

$$\boxed{\sigma_0(T,K)^2 = 2\,\frac{\partial_T\Pi_0(T,K)+\kappa K\partial_K\Pi_0(T,K)+q\Pi_0(T,K)}{K^2\partial^2_{K^2}\Pi_0(T,K)}} \tag{266}$$

provided Dupire's ratio in the r.h.s. of (266) defines a positively bounded Borel-measurable function. Practically speaking, it is virtually always possible to find an interpolation (and/or approximation within the bid ask spread) $\Pi_0 \in W^{1,2}_{p,loc}$ of π_0 for which this is satisfied, and we shall refer to the related volatility function $\sigma_0(T,K)$, as the *market* (or *market model*, in case where the prices π_0 are in fact model prices) *Effective Volatility (EV) function* σ_0.

The practical problem of extracting effective volatility from market prices is tantamount to calibrating a Local Volatility model (see section 38.2). Once available, effective volatility is a useful tool in various tasks, like hedging, calibrating more general stochastic volatility models, arbitraging basket options, etc.

38.2 The Local Volatility Calibration problem

The Local Volatility calibration problem amounts to inferring a Local Volatility function σ from observed option prices, namely European calls or puts with various strikes and maturities. This is both an under-determined (since the set of observed prices is finite whereas the nonparametric function σ has an infinity of degrees of freedom) and ill-posed problem. So a naïve approach based on numerical differentiation using the so-called *Dupire's formula* [76] gives a local volatility which is highly oscillatory (see Figure 15), and thus unstable, for instance when performing a day-to-day calibration.

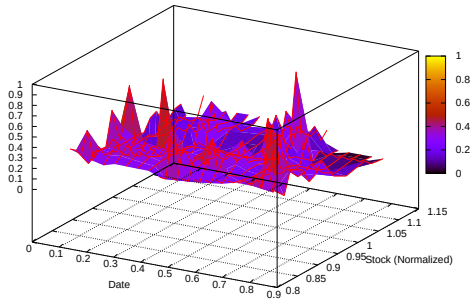

Figure 15: *Local Variance $\sigma(t,S)^2$ obtained by application of Dupire's formula on the DAX index, May 2 2001.*

To meet this issue, the first idea that comes to mind is to look for σ within a parameterized family of functions. However finding classes of functions with all the flexibility required for fitting implied volatility surfaces with several hundred of implied volatility points and a variety of shapes, turns out to be a very challenging task (unless a large family of splines is considered, see Coleman et al. [55], in which case the ill-posedness of the problem shows up again).

The best way to proceed is to stay non-parametric, and to use regularization methods to stabilize the calibration procedure. Since we use a non-parametric local volatility, the model contains a sufficient number of degrees of freedom to provide a perfect fit to virtually any market smile. And the regularization method guarantees that the local volatility thus calibrated is nice and smooth.

38.3 Approach by Tikhonov regularization

Among the various regularization methods at hand, the most popular one is the Tikhonov regularization method of section 36.1. One thus rewrites the local volatility calibration problem as the following nonlinear minimization problem:

$$\min_{\{\sigma\equiv\sigma(t,S);\underline{\sigma}\leq\sigma\leq\overline{\sigma}\}} J(\sigma) = \|\Pi(\sigma) - \pi\|^2 + \alpha\|\sigma - \sigma_0\|^2_{\mathcal{H}^1} \tag{267}$$

where:
• the bounds $\underline{\sigma}$ and $\overline{\sigma}$ are given positive constants specifying the abstract set $\mathcal{A}$ of Section 36.1,
• π is the vector of market prices observed at the calibration time,
• $\Pi(\sigma)$ is the related vector of prices in the Dupire model with volatility function σ,
• σ_0 is a suitable prior (a priori guess on σ), and for $u \equiv u(t,S)$:

$$\|u\|^2_{\mathcal{H}^1} := \int_{t_0}^{\infty}\int_0^{\infty} \left[u(t,S)^2 + (\partial_t u(t,S))^2 + (\partial_S u(t,S))^2\right] dtdS\ .$$

Problem (267) and a related gradient descent approach to solve it numerically (cf. section 36.2) were introduced in Lagnado and Osher [123]. Crépey [63] (see also Egger and Engl [78]) further showed that the general conditions of Section 36.1 are satisfied in this case. Stability and convergence of the method follow.

In Crépey [62] an efficient trinomial tree implementation of this approach was presented, based on an exact computation of the gradient of the (discretized) cost criterion J in (267). Figure 16 displays the local variance surface $\sigma(t,S)^2$ (to be compared with that of Figure 15), the corresponding implied volatility surface and the accuracy of the calibration, obtained by running this algorithm on the DAX index European options data set of May 2, 2001 (consisting of about 300 European vanilla option prices distributed throughout 6 maturities with moneyness $K/S_0 \in [0.8, 1.2]$). At the initiation of the algorithm, the norm of the gradient of the cost criterion J in (267) was equal to `5.73E-02`, and upon convergence after 65 iterations of the gradient descent algorithm, a local minimum of the cost criterion was found, with related value of the norm of the gradient of the cost criterion equal to `6.83E-07`. In the accuracy graph, `implied volatility mismatch` refers to the difference between the Black–Scholes implied volatility corresponding to the market price of an option and its price in the calibrated local volatility model, for each option in the calibration data set.

Such calibration procedures are typically computationally intensive, however it is possible to make them faster by resorting to *parallel computing* (see Table 3 and Crépey [62]).

n × nproc	**1**	**3**	**6**
54	25s	9s	10s
101	4m30s	1m57s	1m36s

Table 3: *Calibration CPU times on a cluster of nproc 1.3 GHz processors connected on a fast Myrinet network, using a calibration tree with n time steps (thus $n^2/2$ nodes in the tree).*

This approach by Tikhonov regularization can be extended to the problem of calibrating a local volatility function using *American* observed option prices as input data (see Crépey

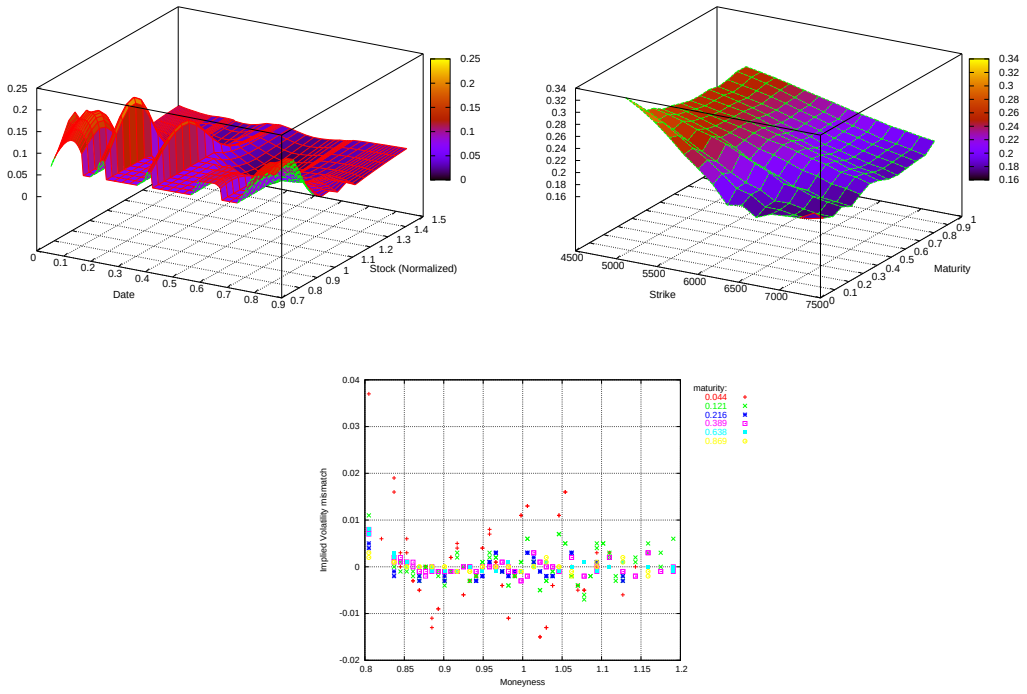

Figure 16: *Local variance, implied volatility and calibration accuracy obtained by application of the Tikhonov regularization method on the DAX index (European options), May 2 2001.*

[62]). Figure 17 thus displays the (squared) local volatility surface $\sigma(t,S)^2$, the corresponding implied volatility surface and the accuracy of the calibration, obtained by running the algorithm of Crépey [62] on the FTSE index American options data set of January 4, 1999 (about 300 option prices distributed throughout 6 maturities with moneyness $K/S_0 \in [0.9, 1.1]$). At the initiation of the algorithm, the norm of the gradient of the cost criterion (267) was equal to `2.84E-02`, and upon convergence after 61 iterations of the gradient descent algorithm, a local minimum of the cost criterion was found, with related value of the norm of the gradient of the cost criterion equal to `6.62E-05`.

Note that this approach by Tikhonov regularization is also applicable to the problem of calibrating a *Lévy model with local jump measure* (see Cont and Rouis [56], Kindermann et al. [117]).

38.4 Approach by entropic regularization

An alternative approach is to use a pseudo-entropic regularization criterion, rewriting the calibration problem as the following nonlinear minimization problem (see Avellaneda et al. [13], Samperi [166]):

$$\min_{\{\sigma\equiv\sigma(t,S);\underline{\sigma}\leq\sigma\leq\overline{\sigma}\}} J(\sigma) = \|\Pi(\sigma) - \pi\|^2 + \alpha\|\sigma - \sigma_0\|^2_{\mathcal{L}^2} \tag{268}$$

where

$$\|\sigma - \sigma_0\|^2_{\mathcal{L}^2} := \mathbb{E}\int_{t_0}^{\infty} (\sigma(t,S_t) - \sigma_0(t,S_t))^2 dt\ .$$

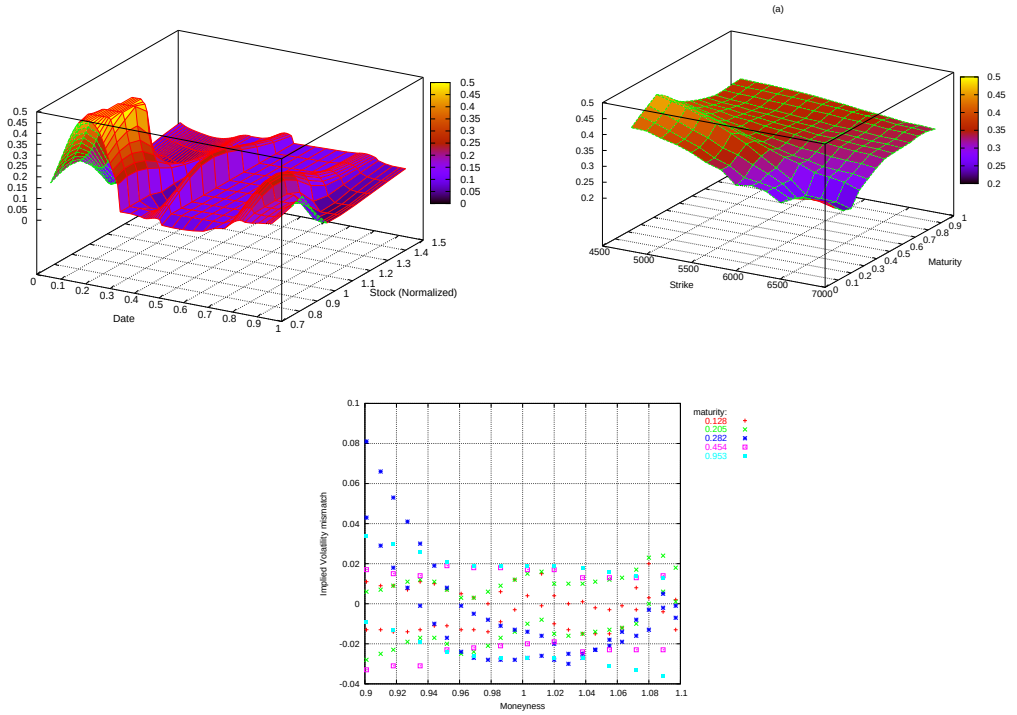

Figure 17: *Local Variance, implied volatility and calibration accuracy obtained by application of the Tikhonov regularization method on the SEI index (American options), January 4, 1999.*

Using a *dual formulation* of a related *stochastic control problem* (cf. Section 39), the optimization problem (268) may be solved at cost $\mathrm{O}(|obs|)$, versus $\mathrm{O}(n^2)$ in the case of Tikhonov regularization. The resolution is thus typically faster, but it also happens to be less robust, and the regularization is less efficient (since it does not involve the gradient of σ, but only the differences $\sigma - \sigma_0$) than with Tikhonov regularization.

Figure 18 (to be compared with Figure 16) displays the (squared) local volatility surface $\sigma(t,S)^2$, the corresponding implied volatility surface and the accuracy of the calibration, obtained by this pseudo-entropic regularization algorithm using a calibration tree with $n = 75$ time steps on the DAX European options data set of May 2, 2001 (same date set as the one of Figure 16). At the initiation of the algorithm, the norm of the gradient of the cost criterion (267) was equal to `4.10E00`, and upon convergence a local minimum of the cost criterion was found, with related value of the norm of the gradient of the cost criterion equal to `5.05E-03`.

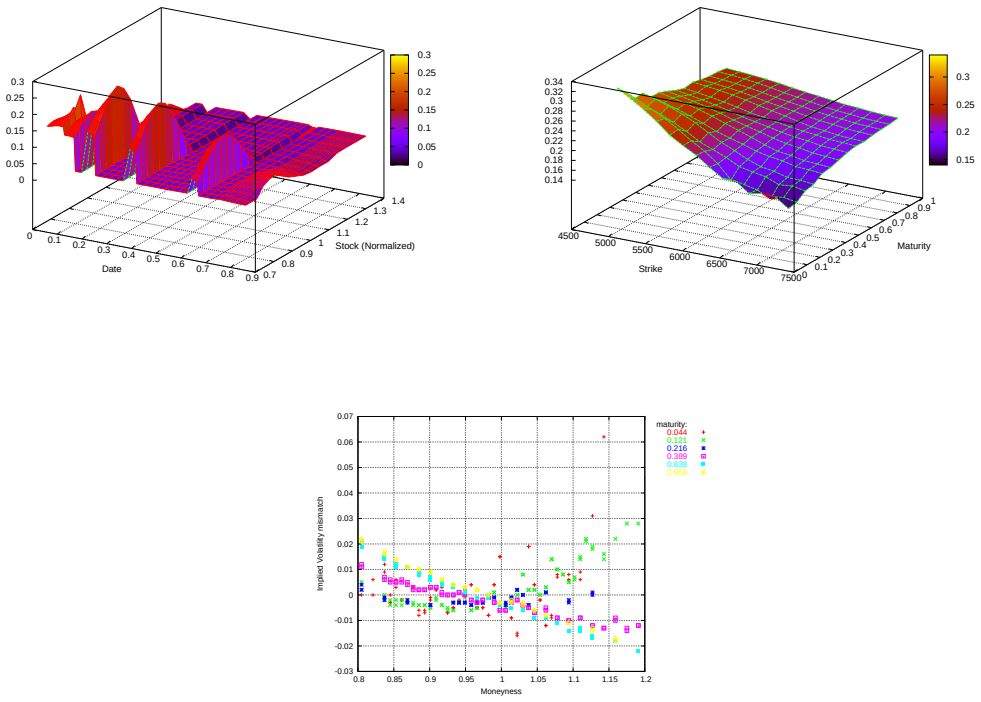

Figure 18: *Local Variance, implied volatility and calibration accuracy obtained by application of the entropic regularization method on the DAX index (European options), May 2 2001.*

39 Weighted Monte Carlo

Let us finally mention a *simulation based* technique for Calibrating Asset Pricing Models, the so-called *Weighted Monte Carlo method* of Avellaneda et al. [10, 11, 12]. In this approach, one first simulates m trajectories following a given (risk-neutral) *prior model*. The simulated trajectories are thus equi-probable in the prior model, by construction. One then

re-ponderates these trajectories, seeking for associated (risk-neutral) probabilities such that option prices are the same in the re-ponderated model and in the market. Calibration here thus takes the form of a re-ponderation of the possible model scenarii, in the spirit of the Girsanov's theorem of stochastic analysis.

Mathematically, we are led to the following program to be solved in the probability measure $p \in \mathcal{P}$, where $\mathcal{P}$ stands for the set of (strictly, for arbitrage issues) positive probability measures over the set of simulated trajectories: $\Pi(p) = \pi$, so

$$\sum_{j=1}^{m} p_j \xi_l^j = \pi_l \ , \ l = 1 \dots d \tag{269}$$

Here π_l denotes the market price of the l^{th} option in the calibration input data set, with payoff ξ_l, and ξ_l^j stands for the payoff of the l^{th} option on the j^{th} trajectory.

Now, since the number of simulated trajectories m is typically much larger than the number d of quoted vanillas (typically $m = 10^4$ at least versus $d =$ a few hundreds at most on major index derivatives markets), one further imposes that the probability measure p be as close as possible to the (prior) uniform distribution on the simulated trajectories in the sense of a suitable *entropy criterion*, in order to stabilize the calibration procedure.

Recall that the *relative entropy* of the probability measure μ relative to the probability measure ν is defined as

$$\mathcal{E}_\nu(\mu) = \begin{cases} \mathbb{E}_\nu \ln(\frac{d\mu}{d\nu}) & \text{if} \quad \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}$$

The mapping $\mu \mapsto \mathcal{E}_\nu(\mu)$ is strictly convex, and strictly minimum and null at $\mu = \nu$.

In the present context, let us denote $\mathcal{E}(p) = \sum_{j=1}^{m} p_j \ln(p_j)$, which is equal to the relative entropy of the probability measure p relative to the uniform distribution on the space of simulated trajectories, up to a constant $\ln(m)$. We then consider the following problem (cf. (269)):

$$\inf_{p \in \mathcal{P} \, ; \, \Pi(p) = \pi} \mathcal{E}(p) \tag{270}$$

By strict convexity of $\mathcal{E}$, this problem admits a unique solution provided the constraints are feasible.

Remark 39.1 Feasibility of the constraints typically holds true in practice since the number of unknowns m is typically much larger than the number of constraints d, however there is no theoretical guarantee that it should hold in general (and it is easy to devise simple examples in which it does not hold). For a robust approach not subject to this constraints feasibility condition, see section 39.2.

39.1 Dual Approach

Introducing Lagrange multipliers, we get the following equivalent form of (270):

$$\sup_{p \in \mathcal{P}} \inf_{\lambda \in \mathbb{R}^d} \mathcal{J}(\lambda, p) := -\mathcal{E}(p) + \sum_{l=1}^{d} \lambda_l (\sum_{j=1}^{m} p_j \xi_l^j - \pi_l) \tag{271}$$

The dual formulation of (271) writes:

$$\inf_{\lambda\in\mathbb{R}^d}\sup_{p\in\mathcal{P}}\mathcal{J}(\lambda,p) \tag{272}$$

Observe that Sion's Minimax Theorem (1958) is applicable here, so that there is no duality gap between (271) and (272). So

$$\sup_{p\in\mathcal{P}}\inf_{\lambda\in\mathbb{R}^d}\mathcal{J}(\lambda,p)=\inf_{\lambda\in\mathbb{R}^d}\sup_{p\in\mathcal{P}}\mathcal{J}(\lambda,p)\ ,$$

and, assuming feasibility of the constraints, the unique solution (p^*,λ^*) to (271) is also the unique solution (p^*,λ^*) to (272), the p^* component of which is in turn the unique solution to (270).

Now, due to the properties of the entropy criterion $\mathcal{E}$, for any given $\lambda\in\mathbb{R}^d$, the probability measure p^λ which maximizes $\mathcal{J}(\lambda,p)$ is explicitly known as, for $j=1\ldots m$:

$$p_j^\lambda=\frac{1}{Z_\lambda}\exp(\sum_{l=1}^d\lambda_l\xi_l^j) \tag{273}$$

where the normalization factor Z_λ is thus given as

$$Z_\lambda=\sum_{j=1}^m\exp(\sum_{l=1}^d\lambda_l\xi_l^j)\ .$$

So, using (273), for every $1\leq k,l\leq d$:

$$\mathcal{E}(p^\lambda)=\textstyle\sum_{j=1}^m p_j^\lambda\ln(p_j^\lambda)=-\ln(Z_\lambda)+\sum_{j=1}^m\sum_{l=1}^d p_j^\lambda\lambda_l\xi_l^j \tag{274}$$

$$\partial_{\lambda_l}\ln(Z_\lambda)=\tfrac{\partial_{\lambda_l}Z_\lambda}{Z_\lambda}=\textstyle\sum_{j=1}^m p_j^\lambda\xi_l^j=\mathbb{E}_{p^\lambda}\xi_l \tag{275}$$

$$\partial^2_{\lambda_k\lambda_l}\ln(Z_\lambda)=\mathbb{C}\mathrm{ov}_{p^\lambda}(\xi_k,\xi_l) \tag{276}$$

In view of (274), problem (272) is thus effectively reduced to

$$\inf_{\lambda\in\mathbb{R}^d}\mathcal{F}(\lambda):=-\mathcal{E}(p^\lambda)+\sum_{l=1}^d\lambda_l(\sum_{j=1}^m p_j^\lambda\xi_l^j-\pi_l)=\ln(Z_\lambda)-\sum_{l=1}^d\lambda_l\pi_l \tag{277}$$

where we have by (275)–(276), for every $1\leq k,l\leq d$:

$$\partial_{\lambda_l}\mathcal{F}(\lambda)=\partial_{\lambda_l}\ln(Z_\lambda)-\pi_l=\mathbb{E}_{p^\lambda}\xi_l-\Pi_l^\lambda$$
$$\partial^2_{\lambda_k\lambda_l}\mathcal{F}(\lambda)=\partial_{\lambda_{kl}}\mathbb{E}_{p^\lambda}\xi_l=\mathbb{C}\mathrm{ov}_{p^\lambda}(\xi_k,\xi_l)$$

So $\mathcal{F}$ in (277) is convex wrt λ, and if λ^* is a critical point of $\mathcal{F}$ with related probability measure $p^*=p^{\lambda^*}$, then λ^* *minimizes* $\mathcal{F}$, and for $l=1\ldots d$:

$$0=\partial_{\lambda_l}\mathcal{F}(\lambda^*)=\partial_{\lambda_l}\ln(Z_\lambda)-\pi_l=\mathbb{E}_{p^*}\xi_l-\pi_l$$

Thus the calibration constraints are satisfied at the probability measure $p^*=p^{\lambda^*}$ relative to a critical (minimum) point λ^* of $\mathcal{F}$ in (277), which is consistent with the fact that the

related probability measure p^* should then solve (270) (whereas the pair (p^*, λ^*) solves (271), (272)).

39.1.1 Algorithm

Finally the related *algorithm* goes as follows:
(i) Simulate m trajectories in the prior model by using a suitable (say Euler) scheme,
(ii) Compute the related payoffs ξ_l^j for $l = 1 \dots d$, $j = 1 \dots m$,
(iii) Use a suitable (descent gradient) optimization routine (like the `lbfgs` algorithm mentioned in section 36.2) in order to minimize $\mathcal{F}(\lambda)$ in (277),
(iv) Compute by (273) the related probabilities p_j^* for $j = 1 \dots m$.

Note that in theory, the values of the probabilities p^λ and of $\mathcal{F}(\lambda)$ are computable in closed-form using (273) and (277), respectively. However the exponentials sitting in p^λ are often found to be infinite numerically. A solution to this problem may be found in re-expressing the related formulas in terms of the differences (typically small, or not too big) $\xi_l^j - \pi_l$ as much as possible.

39.2 Least Squares Approach

In order to get rid of the constraints feasability condition in the previous approach, and, more generally, to improve the stability of the calibration procedure, one may relax the equality constraints into inequality constraints in (269), which results in the following problem

$$\inf_{p \in \mathcal{P}} \mathcal{E}_\omega(p) := \mathcal{E}(p) + \frac{1}{2} \sum_{l=1}^{d} \frac{1}{\omega_l} (\mathbb{E}_p(\xi_l) - \pi_l)^2 \tag{278}$$

for a further vector (parameter) of positive *weights* ω. Note that in the limit $w \to 0$ problem (278) reduces to (269), whereas in the limit $\omega \to +\infty$ the solution of (278) obviously converges to $\mathcal{U}_{\{1 \dots m\}}$.

Now, one can show (see Section 4 of [11]) that problem (278) admits the following equivalent dual formulation:

$$\inf_{\lambda \in \mathbb{R}^d} \sup_{p \in \mathcal{P}} \{ -\mathcal{E}(p) + \sum_{l=1}^{d} \lambda_l (\mathbb{E}_p(\xi_l) - \pi_l) \} + \frac{1}{2} \sum_{l=1}^{d} \omega_l \lambda_l^2 \tag{279}$$

$$= \inf_{\lambda \in \mathbb{R}^d} \mathcal{F}(\lambda) + \frac{1}{2} \sum_{l=1}^{d} \omega_l \lambda_l^2 =: \inf_{\lambda \in \mathbb{R}^d} \mathcal{F}_\omega(\lambda) \tag{280}$$

where $\mathcal{F}(\lambda)$ is the function that was defined in (277).

The related calibration algorithm is then exactly the same as that of section 39.1.1, except for the fact that $\mathcal{F}_\omega(\lambda)$ is minimized instead of $\mathcal{F}(\lambda)$ therein. One thus get a calibrated model achieving a trade-off, ruled by the vector of weights ω, between distance (in the sense of entropy) to the prior and accuracy of the calibration (recall that market prices are in fact defined up to a bid/ask spread, so that a perfect calibration fit has no meaning anyway).

Note that the quadratic term in $\mathcal{F}_\omega(\lambda)$ makes $\mathcal{F}_\omega(\lambda)$ a *proper* function, which means that $\mathcal{F}_\omega(\lambda) \to \infty$ as $|\lambda| \to \infty$. This property may also help in the numerical minimization of $\mathcal{F}_\omega$.

39.3 Applications

The calibrated probability measure p^* can then be used to price OTC (barrier..) options, or to do hedging, consistently with the market smile. This approach is in fact more particularly suited to the application of *statically hedging* a derivative with payoff χ by using the calibration input data options as hedging instruments.

Let us thus formalize this static hedging problem as

$$\min_{\zeta} \mathbb{V}\mathrm{ar}_{p^*}[-\chi + \zeta\xi] \tag{281}$$

where ζ is the (row-)vector of static positions to be held in the calibration input data options, with vector of payoffs $\xi = (\xi_l)_{1\leq l\leq d}$. We thus aim at minimizing the (risk-neutral, here) variance of a static portfolio with a short unit position in the derivative with payoff χ and a long position ζ in the options of the calibration data set. The solution of this problem is then explicitly given by the *multi-linear regression formula*

$$\zeta^* = \mathbb{C}\mathrm{ov}_{p^*}(\chi, \xi)\mathbb{V}\mathrm{ar}_{p^*}(\xi)^{-1} \tag{282}$$

Acknowledgements. These notes grew out of a graduate course of the Master Program in Financial Engineering at Evry University (M2IF program, see www.univ-evry.fr/m2if). I wish to express warm thanks to the students, with a special mention to Arslan Bendimerad and Rémi Boyer, who produced a preliminary version of the material of section 37.5 and Section 11 in the context of their master project.

References

[1] Y. ACHDOU AND O. PIRONNEAU, Computational Methods for Option Pricing. *Vol. 30 of Frontiers in Applied Mathematics*, SIAM, Philadelphia, PA, 2005.

[2] O. ALVAREZ AND A. TOURIN. Viscosity solutions of nonlinear integro-differential equations.*Ann. Inst. H. Poincaré Anal. Non Lineaire*, 13(3):293–317, 1996.

[3] A. L. AMADORI. The obstacle problem for nonlinear integro-differential operators arising in option pricing. *Submitted.*

[4] A. L. AMADORI. Nonlinear integro-differential evolution problems arising in option pricing: a viscosity solutions approach. *Journal of Differential and Integral Equations*, 16(7):787–811, 2003.

[5] L. ANDERSEN. Efficient Simulation of the Heston Stochastic Volatility Model. *Banc of America Securities*, January 23, 2007.

[6] L. ANDERSEN R. BROTHERTON-RATCLIFFE. Exact exotics, *Risk*, 9:85–89, Oct 1996.

[7] L. ANDERSEN AND J. SIDENIUS. Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings, *Journal of Credit Risk*, Vol. 1, No. 1 (Winter 2004), pp. 29–70.

[8] A. ANTONOV, S. MECHKOV, AND T. MISIRPASHAEV. Analytical techniques for synthetic CDOs and credit default risk measures, *NumeriX*, 2005.

[9] M. AVELLANEDA, D. BOYER-OLSON, J. BUSCA AND P. FRIZ. Reconstruction of Volatility: Pricing index options using the steepest-descent approximation, *RISK*, October 2002.

[10] M. AVELLANEDA. Minimum-Entropy Calibration of Asset-Pricing Models, *IJTAF* (1999).

[11] AVELLANEDA M., BU R., FRIEDMAN C., GRANDCHAMP N., KRUK L. AND NEWMAN J. Weighted Monte Carlo: a new technique for calibrating asset-pricing models, *IJTAF*, (2001) 4, 1–29.

[12] M. AVELLANEDA, R. GAMBA. Conquering the Greeks in Monte Carlo, *Proceedings of the First Bachelier Congress* Proc. Courant Seminar Vol. III (2001).

[13] M. AVELLANEDA, C. FRIEDMAN, R. HOLMES AND D. SAMPERI. Calibrating volatility surfaces via relative-entropy minimization, *Applied Math. Finance*, 41 (1997), pp. 37–64.

[14] M. AVELLANEDA, P. LAURENCE. Quantitative Modeling of Derivative Securities from Theory to Practice, *Chapman & Hall*, 2000.

[15] M. AVELLANEDA, A. LEVY, AND A. PARAS. Pricing and hedging derivative securities in markets with uncertain volatilities, *Applied Math. Finance*, 1995.

[16] BALLY V., CABALLERO E., EL-KAROUI N. AND B. FERNANDEZ. Reflected BSDE's, PDE's and Variational Inequalities. *Bernouilli*, To appear.

[17] BALLY, V. AND MATOUSSI, A. Weak solutions for SPDEs and Backward doubly stochastic differential equations, *Journal of Theoretical Probability,* Vol. **14,** No. 1, 125-164 (2001).

[18] BARLES, G., BUCKDAHN, R. AND PARDOUX, E.. Backward Stochastic Differential Equatiions and Integral-Partial Diffrential Equations, *Stochastics and Stochastics Reports*, Vol. **60**, pp. 57-83 (1997).

[19] G.BARLES, C. DAHER AND M. ROMANO. Convergence of numerical schemes for parabolic equations arising in finance theory, *Math. Models Methodes App. Sci.*, 5, n^{o}1, pp. 125–143, 1995.

[20] BARLES, G. AND L. LESIGNE. SDE, BSDE and PDE. El Karoui, N. and Mazliak, L., (Eds.), *Backward Stochastic differential Equatons. Pitman Research Notes in Mathematics Series* 364, pp. 47-80 (1997).

[21] G. G. BARLES AND P.E. SOUGANIDIS. Convergence of approximation schemes for fully nonlinear second order equations, *Asymptotic Anal.*, (4), pp. 271–283, 1991.

[22] J. BARRAQUAND AND T. PUDET. Pricing of American Path-Dependent Contingent Claims, *Mathematical Finance*, 6, n^{o}1, pp. 17–51, 1996.

[23] J. BARRAQUAND AND D.MARTINEAU. Numerical Valuation of High Dimensional Multivariate American Securieties, *J. Of Finance and Quantitative Analysis*, 1995, 30, 383–405.

[24] D. BATES. Jumps and SV: exchange rate processes implicit in deutsche mark options, *Review of Financial Studies*, 9 (1) (1996), 69–107.

[25] R. BELLMAN. Dynamic Programming, *Princeton Univ Press*, 1957.

[26] A. BENSOUSSAN AND J.L. LIONS. Application des Inéquations Variationnelles en Contrôle Stochastique, *Dunod*, 1978.

[27] BERESTYCKI, H., J. BUSCA, AND I. FLORENT. Asymptotics and calibration of local volatility models, *Quant. Finance* 2 (1) (2002).

[28] BIELECKI, T.R., CRÉPEY, S., JEANBLANC, M. AND RUTKOWSKI, M. Arbitrage pricing of defaultable game options with applications to convertible bonds, *Quantitative Finance*, Forthcoming.

[29] BIELECKI, T.R., CRÉPEY, S., JEANBLANC, M. AND RUTKOWSKI, M. Valuation and hedging of defaultable game options in a hazard process model, *Submitted*, 2006.

[30] BIELECKI, T.R., CRÉPEY, S., JEANBLANC, M. AND RUTKOWSKI, M. Defaultable options in a Markovian intensity model of credit risk, *Mathematical Finance*, Forthcoming.

[31] BIELECKI, T.R., CRÉPEY, S., JEANBLANC, M. AND RUTKOWSKI, M. Convertible Bonds in a Defaultable Diffusion Model, *Submitted*.

[32] BIELECKI, T.R., CRÉPEY, S., JEANBLANC, M. AND RUTKOWSKI, M., Valuation of basket credit derivatives in the credit migrations environment. *Handbook of Financial Engineering*, J. Birge and V. Linetsky eds., Elsevier, 2006, forthcoming.

[33] BIELECKI, T.R., JEANBLANC, M. AND RUTKOWSKI, M. Pricing and trading credit default swaps. *Submitted*, 2005.

[34] BIELECKI, T.R. AND RUTKOWSKI, M. Credit Risk: Modeling, Valuation and Hedging. *Springer-Verlag*, Berlin, 2002.

[35] A. BINDER, H. W. ENGL, C. W. GROETSCH, A. NEUBAUER, AND O. SCHERZER. Weakly closed nonlinear operators and parameter identification in parabolic equations by Tikhonov regularization, *Appl. Anal.* 55 (1994), pp. 13–25.

[36] F.BLACK M.SCHOLES. The pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81:635–654, 1973.

[37] N. BOULEAU AND D. LEPINGLE. Numerical methods for stochastic processes, *Wiley*, 1993.

[38] BOUCHARD B. AND ELIE R. Discrete time approximation of decoupled Forward-Backward SDE with jumps, *Submitted*, 2006.

[39] BOUCHARD B., EKELAND I. AND TOUZI N. The Malliavin approach to Monte Carlo approximations to conditional expectations, *Preprint*, 2002.

[40] BRACE, A., DUN, T. AND BARTON, G., Towards a Central Interest Rate Model, *FMMA notes worhing paper*, 1998.

[41] A. BRACE, D. GATAREK AND M. MUSIELA. The market model of interest rate dynamics, *Mathematical Finance*, 7(2) (1997), pp. 127–154.

[42] A. BRACE, B. GOLDYS, J. VAN DER HOEK, and R. WOMERSLEY. Market models in the stochastic implied volatility framework, *Working Paper* S02-11, Department of Statistics, University of New South Wales (2002).

[43] A. BRACE, B. GOLDYS, F. KLEBANER, and R. WOMERSLEY. Market model of stochastic implied volatility with application to the BGM model, *Working Paper* S01-1, Department of Statistics, University of New South Wales (2001).

[44] BREEDEN, D. AND R. LITZENBERGER. Prices of State-contingent Claims Implicit in Options Prices, *Journal of Business*, 51, 621–651 (1978).

[45] M.J. BRENNAN E.S. SCHWARTZ. The valuation of the American put option, *J. of Finance*, 32:449–462, 1977.

[46] BRIANI M, LA CHIOMA C AND NATALINI R. Convergence of numerical schemes for viscosity solutions to integro-differential degenerate parabolic problems arising in financial theory, *Numer. Math.*, 2004, vol. 98, no4, pp. 607-646.

[47] D. BRIGO AND F. MERCURIO. Interest Rate Models - Theory and Practice, *Springer Finance*, 2001.

[48] M.BROADIE P.GLASSERMANN. *A Stochastic Mesh method for Pricing High-Dimensional American Options*, *Journal of Computational Finance*, 7, pp. 35-72, 2004.

[49] M. BROADIE AND P. GLASSERMANN. Pricing American-style securities using simulation, *J.J.of Economic Dynamics and Control*, 21:1323–1352, 1997.

[50] M. BROADIE AND P. GLASSERMANN. Estimating Security Price Derivatives Using Simulation, *Management Science*, 42, 2, 269–285, 1996.

[51] J. BUSCA. A finite elements method for the valuation of American Options, *Unpublished manuscript*, 1998.

[52] P. CARR AND D.MADAN. Option Valuation using the fast Fourier transform,, *Journal of Computational Finance*, (1998), 2, 61-73.

[53] CHERNY, A. AND SHIRYAEV, A. Vector stochastic integrals and the fundamental theorems of asset pricing, *Proceedings of the Steklov Institute of mathematics*, 237 (2002), pp. 6–49.

[54] L. CLEWLOW AND A. CARVEHILL. On the simulation of contingent claims, *Journal of Derivatives*, 66–73, Winter 1994.

[55] T. COLEMAN, Y. LI AND A. VERMA. Reconstructing the unknown volatility function. *Journal of Computational Finance*, 2 (1999), 3, pp. 77–102.

[56] R. CONT AND M. ROUIS. Estimating exponential Lévy models from option prices via Tikhonov regularization, *Working Paper*.

[57] R. CONT AND P. TANKOV. Financial Modelling with Jump Processes, *Chapman & Hall/CRC*, 2003.

[58] R. CONT AND E. VOLTCHKOVA. Finite difference methods for option pricing in jump- diffusion and exponential Lévy models, *SIAM J. Numer. Anal.* 43 (2005), 1596–1626.

[59] J. COX, S. ROSS M. RUBINSTEIN. Option pricing: a simplified approach, *J. of Economics*,January 1978.

[60] M. CRANDALL, H. ISHII AND P.-L. LIONS. User's guide to viscosity solutions of second order partial differential equations, *Bull. Amer. Math. Soc.*, 1992.

[61] S. CRÉPEY. Delta-hedging Vega Risk? *Quantitative Finance*, 4 (October 2004), pp. 559–579.

[62] S. CRÉPEY. Calibration of the Local Volatility in a trinomial tree using Tikhonov regularization. *Inverse Problems*, 19 (2003), pp. 91–127.

[63] S. CRÉPEY. Calibration of the Local Volatility in a generalized Black–Scholes model using Tikhonov regularization. *SIAM Journal on Mathematical Analysis*, Vol. 34 No 5 (2003), pp. 1183–1206.

[64] S. CRÉPEY. Contribution à des méthodes numériques appliquées à la Finance et aux Jeux Différentiels. *PhD Thesis*, Ecole Polytechnique, France, January 2001.

[65] S. CRÉPEY: About the Pricing Equation in Finance. *Submitted.*

[66] S. CRÉPEY: Markovian Reflected and Doubly Reflected BSDEs in a Jump–Diffusion Setting with Regimes. *Submitted.*

[67] CRÉPEY, S., MATOUSSI, A.: About the Greeking Equation in Finance. *Work in preparation.*

[68] CRÉPEY, S., MATOUSSI, A.: Reflected and Doubly Reflected BSDEs with Jumps: A Priori Estimates and Comparison Principle, *Submitted.*

[69] C.W. CRYER. The solution of a quadratic programming problem using systematic overrelaxation, *SIAM J. Control*,(9):385–392, 1971.

[70] C.W. CRYER. The efficient solution of linear complementarity problems for tridiagonal minkowski matrices, *ACM Trans. Math. Softwave*, (9):199–214, 1983.

[71] A. DEBUYSSCHER AND M. SZEGÖ. The Fourier Transform Method – Technical Document, *Moody's Investors Service*, Working paper, Jan 2003.

[72] F. DELBAEN AND W. SCHACHERMAYER. The Mathematics of Arbitrage, *Springer Finance*, 2005.

[73] K. DEMETERFI, E. DERMAN, M. KAMAL AND J. ZOU. More than you ever wanted to know about volatility swaps, *Quantitative Strategies Research Notes*, March 1999. pp. 78–95.

[74] F. DUBOIS, T. LELIEVRE. Efficient Pricing of Asian Options by the PDE Approach, *Journal of Computational Finance*, Vol. 8, No. 2, pp. 55-64, Winter 2004/05.

[75] D. DUFFIE, J. PAN AND K. SINGLETON. Transform Analysis and Asset Pricing for Affine Jump-Diffusions, *Econometrica*, 68 (2000), 6, pp. 1343–1376.

[76] B. DUPIRE. Pricing with a smile, *Risk*, 7 (1994), pp. 18–20.

[77] B. DUMAS, J. FLEMING AND R. WHALEY. Implied volatility functions: empirical tests, *J. of Finance*, 53 (1998), 6, pp. 2059–2106.

[78] H. EGGER AND H. W. ENGL. Tikhonov Regularization Applied to the Inverse Problem of Option Pricing: Convergence Analysis and Rates, *Inverse Problems*, 21 (2005) 1027-1045

[79] EL KAROUI, N., JIAO, Y. Gauss and Poisson Approximation: Applications to CDOs Tranche Pricing. *Working Paper*, 2007.

[80] EL KAROUI, N., JIAO, Y. SteinŠs Method and Zero Bias Transformation for CDOs tranche pricing. *Working Paper*, 2007.

[81] EL KAROUI, N., PENG, S., AND QUENEZ, M.-C. Backward stochastic differential equations in finance. *Mathematical Finance* 7 (1997), 1–71.

[82] H. W. ENGL, M. HANKE, AND A. NEUBAUER. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[83] H. W. ENGL, K. KUNISCH, AND A. NEUBAUER. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems*, 5 (1989), 4, pp. 523–540.

[84] A. ERN S. VILLENEUVE A. ZANETTE. Adaptive Finite element methods for local volatility european option pricing. *International Journal of Theoretical and Applied Finance* 7(6), 2004.

[85] H.J. ETHIER AND T.G. KURTZ. Markov Processes. Characterization and Convergence. *Wiley*, 1986.

[86] W. FLEMING AND H. SONER. *Controlled Markov processes and viscosity solutions, Second edition*, Springer, 2006.

[87] FÖLLMER H. AND SCHIED A.. *Stochastic Finance, An Introduction in Discrete Time*, De Gruyter, 2002.

[88] E. FOURNIÉ, J-M. LASRY, J. LEBUCHOUX, P-L. LIONS AND N. TOUZI. An application of Malliavin calculus to Monte Carlo methods in Finance, *Finance & Stochastics*, vol. 4, no 3, 391–412, 1999.

[89] E. FOURNIÉ, J-M. LASRY, J. LEBUCHOUX AND P-L. LIONS. Applications of Malliavin calculus to Monte Carlo methods in finance. II, *Finance & Stochastics*, 5, 2001, p. 201-236.

[90] A. FRIEDMAN.*Partial Differential Equations of Parabolic Type*, Prentice Hall, 1964.

[91] A. GARCKE. *Sparse grid tutorial*, 2006.

[92] J. GATHERAL. The Volatility Surface: A Practitioner's Guide. *Wiley*, 2006.

[93] M. GIBSON. Understanding the Risk of Synthetic CDOs, *FEDS Working Paper* No. 2004-36, July 2004.

[94] P. GLASSERMAN. Monte Carlo Methods in Financial Engineering, *Springer*, 2004.

[95] R. GLOWINSKI, J-L. LIONS AND R. TREMOLIERES. Analyse Numérique des Inéquations Variationnelles, *Dunod* , 1976.

[96] P. HAGAN, D. KUMAR, A. LESNIEWSKI AND D. WOODWARD. Managing Smile Risk. *Wilmott Magazine*, 2002.

[97] Y. D'HALLUIN, P.A. FORSYTH, K.R. VETZAL. Robust numerical methods for contingent claims under jump diffusion processes. *IMA Journal on Numerical Analysis*, 25, 87–112, 2005.

[98] S. HAMADÈNE, Y. OUKNINE: Reflected backward stochastic differential equation with jumps and random obstacle. *Electronic Journal of Probability*, Vol. 8 (2003) , p. 1–20.

[99] S. HESTON. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, 6(2) 327–43, 1993.

[100] J. HULL. Options, futures, & other derivatives, *Prentice Hall* , 5th edition, 2002.

[101] J. HULL AND A. WHITE. forward rate volatilities, swap rate volatilities, and the implementation of the libor market model, *Journal of Fixed Income*, 10(2) 46–62, 2000.

[102] P. JÄCKEL. Stochastic Volatility Models: Past, Present and Future, *The Best of Wilmott 1*, Chapter 23, 2005.

[103] JACOD, J.: Calcul Stochastique et Problèmes de Martingales. Springer, Berlin Heidelberg New York, 2003.

[104] JACOD, J. AND SHIRYAEV, A.N. Limit Theorems for Stochastic Processes. *Springer*, 2003.

[105] P. JAILLET D. LAMBERTON AND B. LAPEYRE. Variational Inequalities and the Pricing of American Options, *Acta Applicandae Mathematicae 21*, pp. 263–289, 1990.

[106] E. R. JAKOBSEN AND K. H. KARLSEN. A "maximum principle for semicontinuous functions" applicable to integro-partial differential equations, *NoDEA Nonlinear Differential Equations Appl.* 13:137-165, 2006.

[107] E. R. JAKOBSEN AND K. H. KARLSEN. Continuous dependence estimates for viscosity solutions of integro-PDEs, *J. Differential Equations* 212(2): 278-318, 2005.

[108] F. JAMSHIDIAN. Valuation of credit default swap and swaptions. *Finance and Stochastics* 8, 343-371, 2004.

[109] F. JAMSHIDIAN. Libor and swap market models and measures, *Working paper, Sakura Global Capital*, 1996.

[110] Föllmer, H., AND SONDERMANN, D. Hedging of non-redundant contingent claims, *Contributions to Mathematical Economics*, A. Mas-Colell and W. Hildenbrand ed., North-Holland, Amsterdam, 205–223, 1986.

[111] M. JEANBLANC, M. YOR AND M. CHESNEY. Mathematics methods for financial Markets, Springer, 2007.

[112] Y. JIAO. Le risque de crédit: modélisation et simulation numérique. *PhD Thesis* (in English), École polytechnique.

[113] C. KAHL, P. JACKEL. Not-so-complex logarithms in the Heston model. *Working Paper*, 2006.

[114] B. KAMRAD AND P. RITCHKEN. Multinomial approximating models for options with k state variables, *Management Science*, 37:1640–1652, 1991.

[115] I. KARATZAS AND S. SHREVE. Brownian Motion and Stochastic Calculus, *Springer*, 1988.

[116] G.Z. KEMNA AND A.C.F. VORST. A pricing method for options based on average asset values, *J. Banking Finan.*, 113–129, March 1990.

[117] S. KINDERMANN, P. MAYER, H. ALBRECHER, H. W. ENGL. Identification of the local speed function in a Lévy model for option pricing, *Submitted.*

[118] P. E. KLOEDEN AND E. PLATEN. The Numerical Solution of Stochastic Differential Equations, *Springer-Verlag*, 1992

[119] D.E. KNUTH. The Art of Computer programming, Seminumerical Algorithms, volume 2, *Addison-Wesley*, 1981.

[120] H. KUSHNER. Probability Methods for Approximations in Stochastic Control and for Elliptic Equations, *Academic Press*, 1977.

[121] H. KUSHNER AND P.G. DUPUIS. Numerical Methods for Stochastic Control Problems in Continous Time, *Springer-Verlag*, 1992.

[122] Y.W. KWOK. Mathematical models of financial derivatives, *Springer Finance*, 1998.

[123] R. LAGNADO AND S. OSHER. A Technique for Calibrating Derivative Security Pricing Models: Numerical Solution of an Inverse Problem, *Journal of Computational Finance*, 1 (1997), 1, pp. 13–25.

[124] D. LAMBERTON AND B. LAPEYRE. Introduction to Stochastic Calculus Applied to Finance, *Chapman and Hall*, 1996.

[125] B. LAPEYRE AND E. TEMAM.. Competitive Monte Carlo methods for the pricing of Asian options , *Journal of Computational Finance*, Volume 5 / Number 1, Fall 2001

[126] B. LAPEYRE, E. PARDOUX AND R. SENTIS. Méthodes de Monte Carlo pour les équations de transport et de diffusion, *Math. & Applic., SMAI*, No 29, Springer, 1998.

[127] J.-P. LAURENT, A. COUSIN, J-D. FERMANIAN. Hedging default risks of CDOs in Markovian contagion models, *Working Paper.*

[128] J.-P. LAURENT AND J. GREGORY. Basket Default Swaps, CDOs and Factor Copulas, *Journal of Risk*, Forthcoming.

[129] P. L'ECUYER.. Random numbers for simulation, *Communications of the ACM*, 33(10), Octobre 1990.

[130] P. L'ECUYER. Uniform random number generation, *The Annals of Operations Research*, 53:77–120, 1994.

[131] P. L'ECUYER. Random number generation, In *The Hanbook of Simulation*, 1998.

[132] LEWIS, A.. A simple option, formula for general jump- diffusion and other exponential Levy processes, available from http://www.optioncity.net, 2001.

[133] D. LI. On default correlation: A copula function approach, In *Journal of Fixed Income*, 115–118, 2000.

[134] LIONS, P.-L., REGNIER, H.: *Calcul du prix et des sensibilites d'une option americaine par une méthode de Monte Carlo*, *Working Paper*, 2001.

[135] F.A.LONGSTAFF E.S.SCHWARTZ. *Valuing American Options by simulations:A Simple Least-Squares Approach*, *Review of Financial Studies*, Volume 14, Number 1, 2001 , pp. 113-147(35).

[136] J. MA AND J. YONG. Forward-backward stochastic differential equations and their applications. *Springer, Lecture Notes in Math.*, vol. 1702 (1999).

[137] A.-M. MATACHE, T. VON PETERSDORFF AND C. SCHWAB: Fast deterministic pricing of options on Lévy driven assets, *Mathematical Modelling and Numerical Analysis*, 38, 1, 37–72, 2004.

[138] R.C. MERTON. Option pricing when underlying stock returns are discontinuous, *J. Financial Economics*, 3 (1976), pp. 125–144.

[139] R.C. MERTON. The theory of rational option pricing, *Bell J. Econ. Man. Sc.*, 4 (1973), pp. 141–183.

[140] METROPOLIS, N. AND ULAM, S. The Monte Carlo Method, *J. Amer. Stat. Assoc.*, 44, 335-341 (1949).

[141] K. MILTERSEN, K. SANDMANN AND D. SONDERMANN, Closed form solutions for term structure derivatives with log-normal interest rates,

[142] W.J. MOROKOFF AND R.E. CAFLISH. Quasi-random sequences and their discrepancies, *SIAM, Journal of Scientific Computin*, 5(6):1251–1279, nov 1994.

[143] K. MORTON AND D. MAYERS. Numerical Solution of Partial Differential Equations, *Cambridge university press*, 1994.

[144] M. MUSIELA AND M. RUTKOWSKI. Martingale Methods in Financial Modelling, *Springer*, 2004.

[145] H. NIEDERREITER. New developments in uniform pseudorandom number and vector generation, In Springer, editor, In Lecture Notes in Statistics, 106, *Monte Carlo and Quasi Monte Carlo Methods in Scientific Computing*, volume 106, pages 87–120, 1994.

[146] H. NIEDERREITER A.B.OWEN AND J.SHIUE Editors. Randomly permuted (t,m,s)-Nets and (t,s)-sequences, in *Montecarlo and Quasi Montecarlo methods in Scientific Computing*, Springer, New York, 1995.

[147] H. NIEDERREITER. Random Number Generation and Quasi Monte Carlo Methods, *Society for Industrial and Applied mathematics*, 1992.

[148] H. NEIDERREITER. Points sets ans sequences with small discrepancy, *Monatsh.Math*, 104:273–337, 1987.

[149] J. NOCEDAL, S. J. WRIGHT. Numerical Optimization, 2^{nd} edition, *Springer*, 2006.

[150] D. O'KANE AND M. LIVESEY. Base Correlation Explained, *Lehman Brothers Fixed Income Quantitative Credit Research*, Nov 2004.

[151] M. OVERHAUS ET AL. Equity Derivatives: Theory and Applications , *Wiley Finance*, 2002.

[152] G.PAGES V.BALLY. *A quantization method for the discretization of BSDE's and Reflected BSDE's*, *Technical report 628, université Paris 6*, 2000.

[153] D.W. PEACEMAN AND H.H. RACHFORD Jr. The numerical solution of parabolic and elliptic differential equations, *J.of Siam*, 3:28–42, 1955.

[154] D. PHILLIPS. A technique for the numerical solution of certain integral equations of the first kind, *J Assoc Comput Mach*, 9, 84-97 (1962).

[155] D.M.POOLEY P.A.FORSYTH AND K.R.VETZAL, Numerical Convergences properties of option pricing PDEs with uncertain volatility, *IMA Journal of Numerical Analysis*, 23, 241-267, 2003.

[156] W.H. PRESS, B.P. FLANNERY, S.A. TEUKOLSKY AND W.T. VETTERLING, Numerical Recipes in C: The Art of Scientific Computing (Second Edition), *Cambridge University Press*, 1992.

[157] PROTTER, P. E., Stochastic Integration and Differential Equations, Second edition, *Springer*, 2004.

[158] PROTTER, P. E., A partial introduction to financial asset pricing theory, *Stochastic Processes and their Applications*, 91, 2, pp. 169-203(35), 2001.

[159] R. REBONATO. Modern Pricing of Interest-Rate Derivatives: The LIBOR Market Model and Beyond, *Princeton University Press*, 2002.

[160] REISINGER, C. Analysis of linear difference schemes in the sparse grid combination technique, *SIAM Journal on Numerical Analysis*, 2006.

[161] D. REVUZ AND M. YOR. Continuous Martingales and Brownian Motion, 3rd edition. *Springer*, 2001.

[162] P. RITCHKEN. On pricing barrier options, *Journal Of Derivatives*, pages 19–28, Winter 95 1995.

[163] L.C.G. ROGERS. Monte Carlo Valuation of American Options, *Mathematical Finance*, 32(4):1077–1088, 1995. Vol. 12, pp. 271-286, 2002.

[164] L.C.G. ROGERS AND Z. SHI. The value of an asian option, , 32(4):1077–1088, 1995.

[165] SAAD, Y. AND SCHULTZ, M.H. GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear System, *SIAM J. Sci. Statist. Comput.*, 7, 856-869, 1986.

[166] D. SAMPERI. Calibrating a Diffusion Pricing Model with Uncertain Volatility: Regularization and Stability, *Mathematical Finance*, 12 (2002), 1, pp. 71–87.

[167] P.J. SCHÖNBUCHER. Credit Derivatives Pricing Models: Model, Pricing and Implementation, *Wiley Finance*, 2003.

[168] I.M. SOBOL. The distribution of points in a cube and the approximate evaluation of integrals, *U.S.S.R Comput. Maths. Math. Phys*, 16:236–242, 1976.

[169] SCHWEIZER, M.: Option hedging for semimartingales. *Stochastic Processes and their Applications* 37 (1991), 339–363.

[170] S. SHREVE. Stochastic Calculus for Finance I and II, *Springer*, 2005.

[171] D TAVELLA AND C RANDALL. Financial Instruments: The Finite Difference Method, *Wiley Financial Engineering*, 2000.

[172] A. TIKHONOV. Solution of incorrectly formulated problems and the regularization method, *Soviet Math Dokl* 4, 1035-1038, English translation of *Dokl Akad Nauk SSSR* 151, 501-504, 1963.

[173] TITCHMARSH, E.C.. The Theory of Functions, 2nd ed., *Oxford* 1997.

[174] J.N.TSITSIKLIS B.VAN ROY. *Regression methods for Pricing complex American-style Options*, *Working Paper MIT*, 1-22, 2000.

[175] J.N.TSITSIKLIS B.VAN ROY. *Optimal Stopping of Markov Processes: Hilbert Spaces theory, Approximations Algorithms and an application to pricing high-dimensional financial derivatives*, *IEEE Transactions on Automatic Control*, 44 ,10, 1840-1851, 1999

[176] O. VASICEK. Limiting loan loss probability distribution, *Moody's KMV*, 1991.

[177] S. VILLENEUVE AND A. ZANETTE. Parabolic ADI methods for pricing American option on two stocks, *Mathematics of Operations Research*, Vol. 27, Issue 1, 121–149, 2002

[178] E. VOLTCHKOVA. Integro-differential evolution equations: numerical methods and applications in finance, *PhD thesis, Ecole Polytechnique* (in French), Paris, 2005.

[179] P. WILMOTT, Cliquet Options and Volatility Models, *Wilmott Magazine*, 78-83, 2002.

[180] P. WILMOTT. Derivatives: the theory and practice of financial engineering, *Wiley*, 1998.

[181] P. WILMOTT, J. DEWYNNE AND S. HOWISON. Option Pricing, *Oxford Financial Press* (1993).

[182] H.A. WINDCLIFF P.A. FORSYTH AND K.R. VETZAL, Numerical Methods for Valuing Cliquet Options. *Applied Mathematical Finance*, Forthcoming.

[183] H. WINDCLIFF, P.A. FORSYTH AND K.R.VETZAL. Pricing methods and hedging strategies for volatility derivatives, *Journal of Banking and Finance*, 30 (2006) 409–431.

[184] YANG, J., T. HURD AND X. ZHANG. Saddlepoint Approximation Method for Pricing CDOs, *Journal of Computational Finance*, Vol. 10, No. 1, (Fall 2006).

[185] Y. ZHU, X. WU, I-L. CHERN. Derivative Securities and Difference Methods, *Springer Finance*, 2004.

[186] R. ZVAN, P.A. FORSYTH AND K.R. VETZAL. Robust numerical methods for pde models of asian option, *Journal of Computational Finance*, 1:39–78, 1998.