Partial differential equations in finance

Y. Achdou[1], O. Bokanowski[1] & T. Lelièvre[2]

[1] *UFR Mathématiques, Université Paris 7, 75251 Paris Cedex 05, France, et Laboratoire Jacques-Louis Lions, Université Paris 6, 75252 Paris Cedex 05*

[2] *CERMICS, Université Paris-Est, ENPC, 6-8 avenue Blaise Pascal, Champs sur Marne, 77455 Marne La Vallée Cedex 2, France*

CERMICS — Université Paris-Est
ENPC, 6 et 8 avenue Blaise Pascal
Cité Descartes - Champs sur Marne
77455 Marne la Vallée Cedex 2

http://cermics.enpc.fr

Partial differential equations in finance.

Yves Achdou[1], Olivier Bokanowski[1] and Tony Lelièvre[2]

[1] UFR Mathématiques, Université Paris 7, Case 7012, 75251 Paris Cedex 05, France

and Laboratoire Jacques-Louis Lions, Université Paris 6, 75252 Paris Cedex 05.

[2] CERMICS, Ecole Nationale des Ponts et Chaussées, 6 & 8 Av. Pascal, 77455 Champs-sur-Marne, France.

E-mail: {achdou, boka}@math.jussieu.fr, lelievre@cermics.enpc.fr

12th September 2007

Abstract

Various contexts where partial differential equations are useful in finance are presented, in particular for the pricing of European and American options, and for calibration. The emphasis is laid on numerical aspects: discretization methods, algorithms and analysis of the numerical schemes.

1 Introduction

Numerical methods based on partial differential equations (PDEs) in finance are not very popular. Indeed, the models are usually derived from probabilistic arguments and Monte Carlo methods are therefore much more natural. Stochastic methods are also often simpler to implement than the algorithms used for solving the related PDEs. However, when it is possible to efficiently discretize the PDE (which is not always the case, the typical counter example being high-dimensional problems, see Remark 9 below), the algorithms are usually much more efficient. Moreover, the solution to the partial differential equation gives more information. In the context of option pricing with constant parameters, one obtains for example the price of the option for all values of the maturity and for all spot prices, while the probabilistic formulation typically gives the value of the option for a fixed maturity and a fixed spot price. In particular, this is useful for computing derivatives of the option's price (the so-called "greeks").

The PDEs obtained in finance have several characteristics. First, they are posed on a bounded domain in time $(0, T)$, with typically a singular final condition at the maturity $t = T$, and very often in an unbounded domain in the spot variable, which leads to impose suitable "boundary conditions" at infinity to get well-posed problems and to use appropriate numerical approximations (truncation to a bounded domain and artificial boundary conditions). These PDEs are usually of parabolic type, but often with degenerate diffusions. Because of operational constraints, the numerical methods used for the discretization of the PDE must be sufficiently fast and accurate to be useful in practice. These peculiarities of PDEs in finance explain the need for up-to-date and sometimes involved numerical methods.

In this paper, we definitely concentrate on numerical issues and try to review the main numerical methods used for solving PDEs in finance. This presentation heavily relies on the monograph [3]. Other useful monographs are [39, 36, 55].

The paper is organized as follows. In Section 2, we present the main arguments to derive a PDE for the price of various European and American options. In Section 3 the finite difference method and the finite element method for the discretization of such PDEs are introduced, for European options. This is generalized in Section 4 to American options. Finally, Section 5 is devoted to calibration methods.

2 Partial differential equations for option pricing

2.1 A primer: the Black and Scholes model for European options

The aim of this section is to recall the basic tools needed to derive a PDE in the context of option pricing, without providing all the detailed assumptions required on the data to perform this derivation. We refer for example to [36, 39] for a rigorous mathematical presentation.

We adopt the standard Black and Scholes model (see [9, 42]) with a risky asset whose price at time t is S_t and a risk free asset whose price at time t is S_t^0, such that:

$$dS_t = S_t(\mu\, dt + \sigma dB_t), \quad dS_t^0 = rS_t^0\, dt.$$

The process B_t is a standard Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q})$, and μ (the mean rate of return), r (the interest rate) and $\sigma > 0$ (the volatility) are three constants. However, the following can be generalized to the case where μ, r and $\sigma > 0$ are functions of t and S (under suitable smoothness assumptions). We introduce the stochastic process $W_t = B_t + \frac{\mu - r}{\sigma} t$. Under the risk-neutral probability $\mathbb{P}$ defined by its Radon-Nikodym derivative with respect to $\mathbb{Q}$ by

$$\left.\frac{d\mathbb{P}}{d\mathbb{Q}}\right|_{\mathcal{F}_t} = \exp\left(\int_0^t \frac{r-\mu}{\sigma} dB_s - \frac{1}{2}\int_0^t \left(\frac{r-\mu}{\sigma}\right)^2 ds\right),$$

W_t is a Brownian motion and S_t/S_t^0 is a martingale. This is one of the fundamental property of the stochastic process needed in the following. The process S_t satisfies the following stochastic differential equation (SDE) under $\mathbb{P}$:

$$dS_t = S_t(r\, dt + \sigma dW_t). \tag{1}$$

Let us now consider a portfolio with H_t risky assets and H_t^0 no-risk assets. Its value at time t is:

$$P_t = H_t S_t + H_t^0 S_t^0. \tag{2}$$

We suppose that this portfolio is self-financing (any manipulation on this portfolio, *i.e.* any change of the values of H_t or H_t^0, is done without any inflows or outflows of money) which translates into

$$dP_t = H_t dS_t + H_t^0 dS_t^0. \tag{3}$$

The value of a self-financing portfolio changes if and only if the price of the risky asset changes. Using (3), it is possible to show that P_t/S_t^0 is also a martingale.

We consider the following problem: for a given function ϕ (the payoff function) and a given time $T > 0$ (the maturity), is it possible to build a self-financing portfolio such that $P_T = \phi(S_T)$? Classical examples of function ϕ are $\phi(S) = (S-K)_+$ (vanilla call) or $\phi(S) = (S-K)_-$ (vanilla put), where, for any real x, $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$. The answer is positive (this is typically based on a martingale representation theorem, the fact that P_t/S_t^0 is a martingale and the fact that the payoff $\phi(S_T)$ is $\mathcal{F}_T$-measurable), and it is then possible to show that such a portfolio has the following value at time t:

$$P_t = \mathbb{E}\left(\exp\left(-\int_t^T r\,ds\right)\phi(S_T)\Big|\mathcal{F}_t\right). \tag{4}$$

By the so-called arbitrage-free principle, P_t is actually the "fair price", at time t of the option which enables its owner to get the payoff $\phi(S_T)$ at time T. In the particular context of vanilla options, the solution is analytically known, at least if r and σ are constant: these are the celebrated Black and Scholes formula. However, in the case when r and σ are functions of t and S, (4) cannot be estimated without a numerical method. We are interested in deterministic numerical methods, based on a PDE related to (4).

The second fundamental property of the stochastic process S_t required to obtain a PDE formulation of this problem is a Markov property. Roughly speaking, it states that the expectation of any function of $(S_t)_{0\le t\le T}$ conditionally to $\mathcal{F}_t$ is actually a function of the price S_t of the risky asset at time t. In our context, this property shows that P_t writes

$$P_t = p(t, S_t) \tag{5}$$

where p is a function of $t \in [0, T]$ and $S \in [0, \infty)$, called *the pricing function* of the option. Notice that even if (5) only involves the value of p at point (t, S_t), the pricing function p is a deterministic function defined for all values of $t \ge 0$ and $S \ge 0$. By the Markov property of S_t, we also have the following representation formula for p:

$$p(t, x) = \mathbb{E}\left(\exp\left(-\int_t^T r\,ds\right)\phi(S_T^{t,x})\right) \tag{6}$$

where $(S_\theta^{t,x})_{t\le\theta\le T}$ denotes the process solution to (1) starting from x at time t

$$\begin{cases} dS_\theta^{t,x} = S_\theta^{t,x}(r\,d\theta + \sigma dW_\theta), \quad \theta \ge t, \\ S_t^{t,x} = x. \end{cases} \tag{7}$$

By using Itô's calculus and the fact that P_t/S_t^0 is a martingale, we then obtain that p should satisfy the following PDE:

$$\begin{cases} \dfrac{\partial p}{\partial t} + rS\dfrac{\partial p}{\partial S} + \dfrac{\sigma^2 S^2}{2}\dfrac{\partial^2 p}{\partial S^2} - rp = 0, \\ p(T, S) = \phi(S). \end{cases} \tag{8}$$

Conversely, it is possible (using again a martingale representation theorem) to show that if p satisfies (8), then $p(t, S_t)$ is the value of a self-financing port-folio with

value $\phi(S_T)$ at time T. Moreover, one can check that $\frac{\partial p}{\partial S}(t, S_t) = H_t$, which shows that obtaining an accurate approximation of $\frac{\partial p}{\partial S}$ is important in order to estimate the quantity of risky asset H_t needed at time t to build the portfolio with value P_t (this is the hedging strategy). This is an example of so-called Feynman-Kac formulae, which are used in many other contexts (quantum chemistry [13], transport equations [40]) either to give a probabilistic interpretation to a PDE, or to recast the computation of an expectation into a PDE problem.

For the problem (8) to be well posed (*i.e.* for one and only one solution to exist), one needs to supply the system with boundary conditions when $S = 0$ or $S \to \infty$. More precisely, one needs to make precise in which functional space the function p is looked for. This will be explained in Section 3.2.

Remark 1 (Maximum principle) *From the PDE (8) and the so-called maximum principle, it is possible to derive many qualitative properties and* a priori *bounds on the price p (like the call-put parity for example, see [3]). Roughly speaking, the maximum principle states that if the data (initial condition, boundary conditions, right-hand side) for the PDE (8) are positive, then the solution is positive. This property is definitely necessary to hold for a price. It is also an important property to check on the numerical schemes (which is then called a discrete maximum principle, see Remark 6 below).*

Remark 2 (Another derivation) *It is also possible to derive the PDE without introducing the risk-neutral probability (see [55]), by considering a portfolio containing some options and some risky assets and by using an arbitrage-free argument.*

Remark 3 (On the Black and Scholes model) *It is well-known that the Black and Scholes model for the evolution of the risky asset (1) badly compares with experimental data. We will discuss in Section 5.1 some possible refinements which have been introduced in order to better fit the observations.*

2.2 Other options

The derivation presented in Section 2.1 is prototypical. In particular, the derivation of a PDE satisfied by the pricing function of an option always relies on the two fundamental properties stressed above: the martingale and the Markov properties of a suitable stochastic process. In this section, we present PDEs for the prices of various options without providing all the details of the derivation.

2.2.1 Basket options

In many cases, the payoff of the option depends on the values of more than one asset, which typically do not evolve independently. Let us for example consider the case of two assets, which evolve following the following SDE under the neutral risk probability

$$\begin{cases} dS_t^1 = S_t^1 \left(r\, dt + \sigma_1 dW_t^1 \right), \\ dS_t^2 = S_t^2 \left(r\, dt + \sigma_2 dW_t^2 \right). \end{cases}$$

where W_t^1 and W_t^2 are possibly correlated standard Brownian motions. We call ρ the correlation of W_t^1 and W_t^2 : $d\langle W_1, W_2 \rangle_t = \rho dt$. We suppose that the maturity is

$T > 0$ and the payoff is $\phi(S_T^1, S_T^2)$, where ϕ is a given function. It is then possible to show that the price of the option at time t is $p(t, S_t^1, S_t^2)$ where p satisfies

$$\begin{cases} \dfrac{\partial p}{\partial t} + rS_1\dfrac{\partial p}{\partial S_1} + rS_2\dfrac{\partial p}{\partial S_2} + \dfrac{\sigma_1^2 S_1^2}{2}\dfrac{\partial^2 p}{\partial S_1^2} + \dfrac{\sigma_2^2 S_2^2}{2}\dfrac{\partial^2 p}{\partial S_2^2} + \rho\sigma_1\sigma_2 S_1 S_2 \dfrac{\partial^2 p}{\partial S_1 \partial S_2} - rp = 0, \\ p(T, S_1, S_2) = \phi(S_1, S_2). \end{cases} \tag{9}$$

Here again, r, σ_1 and σ_2 may be functions of t and (S_1, S_2). It is possible to solve such PDEs by standard numerical methods up to dimension 3 or 4. To derive appropriate discretization for higher dimension is not an easy task, and is still the subject of current researches (see Remark 9 below).

2.2.2 Barrier options

Again, let us consider an option on a single asset. For some options, the payoff becomes 0 if for a time $t \in [0, T]$ S_t goes below a or above b, where a and b are two given values, $0 < a < b$ (the case $a = 0$ or $b = \infty$ can be treated similarly). Mathematically, the payoff is $1_{\forall t \in [0,T],\, S_t \in [a,b]}\phi(S_T)$ where, for any event $A \subset \Omega$, 1_A denotes the characteristic function of A, and S_t satisfies (1). In this case, the relevant stochastic process for deriving the PDE is $S_{t\wedge\tau}$, where $\tau = \inf\{t \in [0, T], S_t \geq b \text{ or } S_t \leq a\}$ is a stopping time, and, for any real x and y, $x \wedge y = \inf(x, y)$. It can be checked that $S_{t\wedge\tau}$ is a Markov process, and that $S_{t\wedge\tau}/S^0_{t\wedge\tau}$ is a martingale. It is then possible to show that the price of the option at time t is $p(t \wedge \tau, S_{t\wedge\tau})$ where p is defined for $t \in [0, T]$ and $S \in [a, b]$ and satisfies:

$$\begin{cases} \dfrac{\partial p}{\partial t} + rS\dfrac{\partial p}{\partial S} + \dfrac{\sigma^2 S^2}{2}\dfrac{\partial^2 p}{\partial S^2} - rp = 0, \\ p(T, S) = \phi(S), \\ p(t, a) = p(t, b) = 0. \end{cases} \tag{10}$$

Here again, r and σ may be functions of t and S. Moreover, the generalization to basket options is straightforward (see Section 2.2.1). In this case, it is possible to consider more general barriers, namely a payoff of the form $1_{\forall t \in [0,T],\, (S_t^1, S_t^2, \ldots, S_t^d) \in \mathcal{D}}\phi(S_T)$, where d denotes the number of underlying assets and $\mathcal{D}$ is any simple connected domain of $\mathbb{R}^d$. The appropriate discretization for general domains $\mathcal{D}$ is the finite element method (see Section 3.2).

2.2.3 Options on the maximum

For some options (the so-called lookback options), the payoff involves the maximum of the risky asset. For example, it writes $\phi(S_T, M_T)$ where $M_t = \max_{0 \leq r \leq t} S_r$ and S_t satisfies (1). One can check that (S_t, M_t) is a Markov process. It is then possible to show that the price of the option at time t is $p(t, S_t, M_t)$ where p is defined for $t \in [0, T]$ and $(S, M) \in \{(S, M) \in \mathbb{R}^2,\, 0 \leq S \leq M\}$ and satisfies:

$$\begin{cases} \dfrac{\partial p}{\partial t} + \dfrac{\sigma^2}{2}S^2\dfrac{\partial^2 p}{\partial S^2} + rS\dfrac{\partial p}{\partial S} - rp = 0, \\ p(T, S, M) = \phi(S, M), \\ \dfrac{\partial p}{\partial M}(t, S, S) = 0. \end{cases} \tag{11}$$

If the payoff is of the form $\phi(S,M) = M\tilde{\phi}(S/M)$, it is possible to reduce the problem to a two dimensional one, (including the time variable). Indeed, one can check by straightforward computations that $p(t,S,M) = Mw(t,S/M)$ where w is a function of $t \in [0,T]$ and $\xi \in [0,1]$ which satisfies:

$$\begin{cases} \dfrac{\partial w}{\partial t} + \dfrac{\sigma^2}{2}\xi^2 \dfrac{\partial^2 w}{\partial \xi^2} + r\xi \dfrac{\partial w}{\partial \xi} - rw = 0, \\ w(T,\xi) = \tilde{\phi}(\xi), \\ \dfrac{\partial w}{\partial \xi}(t,1) = w(t,1). \end{cases} \tag{12}$$

Notice that this reduction is not generally possible for (t,S,M)-dependent interest rate and volatility (except for very peculiar dependencies).

2.2.4 Options on the average

Some options (the so-called Asian options) involve the average of the risky asset. More precisely, the payoff writes $\phi(S_T, A_T)$ where $A_t = \frac{1}{t}\int_0^t S_r\, dr$ and S_t satisfies (1). One can check that (S_t, A_t) is a Markov process. Using this property, it is possible to show that the price of the option at time t is $p(t,S_t,A_t)$ where p is defined for $t \in [0,T]$ and $(S,A) \in [0,\infty)^2$ and satisfies:

$$\begin{cases} \dfrac{\partial p}{\partial t} + \dfrac{\sigma^2 S^2}{2}\dfrac{\partial^2 p}{\partial S^2} + rS\dfrac{\partial p}{\partial S} + \dfrac{1}{t}(S-A)\dfrac{\partial p}{\partial A} - rV = 0, \\ p(T,S,A) = \phi(S,A). \end{cases} \tag{13}$$

In some cases (see [48]), it is possible to reduce this problem to a one-dimensional PDE. More precisely, for fixed strike call ($\phi(S,A) = (A-K)_+$) or fixed strike put ($\phi(S,A) = (K-A)_+$), we have $p(t,S,A) = Sf\left(t, \frac{K-tA/T}{S}\right)$ where f satisfies

$$\begin{cases} \dfrac{\partial f}{\partial t} + \dfrac{\sigma^2\xi^2}{2}\dfrac{\partial^2 f}{\partial \xi^2} - \left(\dfrac{1}{T} + r\xi\right)\dfrac{\partial f}{\partial \xi} = 0, \\ f(T,\xi) = \tilde{\phi}(\xi), \end{cases} \tag{14}$$

and $\tilde{\phi}(\xi) = \xi_-$ (resp. $\tilde{\phi}(\xi) = \xi_+$). This reduction of (13) to (14) is also possible for floating strike call ($\phi(S,A) = (S-A)_+$) (resp. for floating strike put ($\phi(S,A) = (A-S)_+$)) by setting $p(t,S,A) = Sf\left(t, -\frac{tA}{TS}\right)$ and $\tilde{\phi}(\xi) = (1+\xi)_+$ (resp. $\tilde{\phi}(\xi) = (1+\xi)_-$). However, this reduction is generally not possible for (t,S,A)-dependent interest rate and volatility (except for very peculiar dependencies).

Remark 4 (Bermudean options) *As a transition between European and American options, we would like to mention that it is very easy to price Bermudean options with the PDE approach. For such options, the contract can be exercised only at certain days between the present time and the maturity. Mathematically, for an option on a single asset (the spot price is called S) and if ϕ denotes the payoff, the pricing function satisfies $p(t_i^+, S) = \max(p(t_i^-, S), \phi(S))$, at each exercising time t_i, and (8) between the exercising times, (see [24] p.211).*

2.3 The case of American options

We have so far presented so-called European options, *i.e.* some options which enables its owner to get $\phi(S_T)$ at a fixed time T. On the other hand, American options can be exercised at any time up to the maturity. Hence the price of an American option of payoff ϕ and maturity T will be the maximum of all possible expectations such as (6) for stopping times τ between t and T, that is, for $t \in [0,T]$ and $x \geq 0$,

$$p(t,x) = \sup_{\tau \in \mathcal{T}_{[t,T]}} \mathbb{E}\left(e^{-\int_t^\tau r\,ds}\phi(S_\tau^{t,x})\right), \tag{15}$$

where $\mathcal{T}_{[t,T]}$ denotes the set of stopping times τ of the filtration $\mathcal{F}_t$, with values in $[t,T]$.

2.3.1 The PDE for American options

We now present the main arguments to derive a PDE on p (or more precisely a set of partial differential inequalities).

Notice first that taking $\tau = t$ in (15) yields the inequality

$$p(t,x) \geq \phi(x). \tag{16}$$

Moreover, we clearly have from (15) $p(T,x) = \phi(x)$.

Let t and δt be such that $0 \leq t \leq t+\delta t \leq T$. From (15) we have:

$$\begin{aligned} e^{-\int_0^{t+\delta t} r\,ds} p\left(t+\delta t, S_{t+\delta t}^{t,x}\right) &= \sup_{\tau \in \mathcal{T}_{[t+\delta t,T]}} \mathbb{E}\left(e^{-\int_0^\tau r\,ds}\phi\left(S_\tau^{t+\delta t, S_{t+\delta t}^{t,x}}\right)\right), \\ &\leq \sup_{\tau \in \mathcal{T}_{[t,T]}} \mathbb{E}\left(e^{-\int_0^\tau r\,ds}\phi\left(S_\tau^{t,x}\right)\right), \\ &\leq e^{-\int_0^t r\,ds} p\left(t,x\right), \end{aligned}$$

where we have used the fact that: $S_\tau^{t+\delta t, S_{t+\delta t}^{t,x}} = S_\tau^{t,x}$. By Itô's calculus (taking the limit $\delta t \to 0$), we thus obtain

$$-\frac{\partial p}{\partial t} + \mathcal{A}p \geq 0 \tag{17}$$

where we have introduced the linear PDE operator

$$\mathcal{A}p = -rS\frac{\partial p}{\partial S} - \frac{\sigma^2 S^2}{2}\frac{\partial^2 p}{\partial S^2} + rp. \tag{18}$$

Combined with (16), we then obtain

$$\min\left(-\frac{\partial p}{\partial t} + \mathcal{A}p, p - \phi\right) \geq 0. \tag{19}$$

Our aim is now to show that the inequality in (19) is actually an equality. This is done in several steps, and requires to identify an optimal stopping time τ^* for which the supremum in (15) is obtained. For a fixed (t,x), let us introduce the stopping time $\tau^* \in \mathcal{T}_{[t,T]}$ defined by

$$\tau^* = \inf\left\{\theta \geq t,\, p\left(\theta, S_\theta^{t,x}\right) = \phi\left(S_\theta^{t,x}\right)\right\},\ a.s. \tag{20}$$

(notice that $\tau^* \leq T$ since $p(T,x) = \phi(x)$). It can be shown (see Section 2.3.2) that

$$p(t,x) = \mathbb{E}\left(e^{-\int_t^{\tau^*} r\,ds}\phi\left(S_{\tau^*}^{t,x}\right)\right) = \mathbb{E}\left(e^{-\int_t^{\tau^*} r\,ds}p(\tau^*, S_{\tau^*}^{t,x})\right). \tag{21}$$

Using a decreasing property (26) proved below, one then obtains that [1] for any $\delta t > 0$,

$$p(t,x) = \mathbb{E}\left(e^{-\int_t^{\tau^*_{\delta t}} r\,ds}p\left(\tau^*_{\delta t}, S_{\tau^*_{\delta t}}^{t,x}\right)\right), \quad \text{where } \tau^*_{\delta t} = (t+\delta t)\wedge\tau^*. \tag{22}$$

This can be seen as a Dynamic Programming Principle (or Bellman's principle).

Now if we suppose that $p(t,x) > \phi(x)$, then for any $\delta t > 0$ we have $\mathbb{P}(\tau^*_{\delta t} > t) = 1$. Considering Itô's formula in (22), and by (17), we obtain $\left(-\frac{\partial p}{\partial t} + \mathcal{A}p\right)(\theta, S_\theta^{t,x}) = 0$ for $t \leq \theta \leq \tau^*_{\delta t}$, thus leading to $\left(-\frac{\partial p}{\partial t} + \mathcal{A}p\right)(t,x) = 0$. This shows that the inequality in (19) is actually an equality.

Hence the PDE for the American option is

$$\begin{cases} \min\left(-\dfrac{\partial p}{\partial t} + \mathcal{A}p, p - \phi\right) = 0, \quad t \in [0,T],\ x \geq 0, \\ p(T,x) = \phi(x), \quad x \geq 0, \end{cases} \tag{23}$$

where $\mathcal{A}$ is defined by (18). The major difference between PDE (23) for American options and PDE (8) for European options is that (23) is a *nonlinear* equation. This makes the theory of existence and uniqueness as well as the numerical approximation more difficult than for European options.

Remark 5 (Viscosity solutions) *In the presentation above, we have used Itô's formula, which requires that p is C^1 in time and C^2 in the spot variable. This is not true in general. It is however possible, following the same lines, to prove that p is a weak solution to* (23) *in the viscosity sense. For an historical derivation of this PDE, see [7] or [26], where a variational formulation of* (23) *is derived (see* (57) *below). We also refer to [44] for an infinite horizon related problem, [21] for general results, [46] for an approach of optimal stopping including jump diffusion processes, and to [4] for the case of a discontinuous payoff ϕ.*

2.3.2 Proof of (21)

First, from the definition (15) of p we have, for any stopping time $\rho \in \mathcal{T}_{[t,T]}$,

$$e^{-\int_t^\rho r\,ds}p\left(\rho, S_\rho^{t,x}\right) = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{[\rho,T]}} \mathbb{E}\left(e^{-\int_t^\tau r\,ds}\phi(S_\tau^{t,x})\middle|\ \mathcal{F}_\rho\right), a.s. \tag{24}$$

where $\mathcal{T}_{[\rho,T]}$ denotes the set of stopping times τ such that $\rho \leq \tau \leq T$. Then it is possible to show that (see for instance [37, Eq. (D.7)]), for any stopping time $\rho \in \mathcal{T}_{[t,T]}$,

$$\mathbb{E}\left(e^{-\int_t^\rho r\,ds}p(\rho, S_\rho^{t,x})\right) = \sup_{\tau\in\mathcal{T}_{[\rho,T]}} \mathbb{E}\left(e^{-\int_t^\tau r\,ds}\phi(S_\tau^{t,x})\right). \tag{25}$$

[1] For a European option we would have more simply $p(t,x) = \mathbb{E}\left(e^{-\int_t^{t+\delta t} r\,ds}p\left(t+\delta t, S_{t+\delta t}^{t,x}\right)\right)$.

We obtain from (25) the decreasing property: for all stopping times $\rho_1, \rho_2 \in \mathcal{T}_{[t,T]}$, such that $\rho_1 \geq \rho_2$,

$$\mathbb{E}\left(e^{-\int_t^{\rho_1} r\,ds} p(\rho_1, S_{\rho_1}^{t,x})\right) \leq \mathbb{E}\left(e^{-\int_t^{\rho_2} r\,ds} p(\rho_2, S_{\rho_2}^{t,x})\right). \tag{26}$$

We deduce from (24) that, for any $\tau \in \mathcal{T}_{[\tau^*,T]}$,

$$\mathbb{E}\left(e^{-\int_t^{\tau} r\,ds}\phi\left(S_\tau^{t,x}\right) \mid \mathcal{F}_{\tau^*}\right) \leq e^{-\int_t^{\tau^*} r\,ds} p(\tau^*, S_{\tau^*}^{t,x}) = e^{-\int_t^{\tau^*} r\,ds}\phi\left(S_{\tau^*}^{t,x}\right), \tag{27}$$

where the last identity comes from the definition (20) of τ^*. Then, for any stopping time $\tau \in \mathcal{T}_{[t,T]}$, we have (by decomposing on the events $\{\tau < \tau^*\}$ and $\{\tau \geq \tau^*\}$), and using (27) for $\tau \geq \tau^*$):

$$\mathbb{E}\left(e^{-\int_t^{\tau} r\,ds}\phi\left(S_\tau^{t,x}\right)\right) \leq \mathbb{E}\left(e^{-\int_t^{\tau\wedge\tau^*} r\,ds}\phi\left(S_{\tau\wedge\tau^*}^{t,x}\right)\right).$$

Hence, by taking the supremum over all the stopping times $\tau \in \mathcal{T}_{[t,T]}$,

$$p(t,x) \leq \sup_{\tau\in\mathcal{T}_{[t,T]}} \mathbb{E}\left(e^{-\int_t^{\tau\wedge\tau^*} r\,ds}\phi\left(S_{\tau\wedge\tau^*}^{t,x}\right)\right) = \sup_{\tau\leq\tau^*,\,\tau\in\mathcal{T}_{[t,T]}} \mathbb{E}\left(e^{-\int_t^{\tau} r\,ds}\phi\left(S_\tau^{t,x}\right)\right). \tag{28}$$

By (15), the right-hand side of (28) is bounded from above by $p(t,x)$, and thus we obtain the equality

$$p(t,x) = \sup_{\tau\leq\tau^*,\,\tau\in\mathcal{T}_{[t,T]}} \mathbb{E}\left(e^{-\int_t^{\tau} r\,ds}\phi\left(S_\tau^{t,x}\right)\right). \tag{29}$$

In fact the supremum in (29) is reached only for $\tau = \tau^*$ a.s.. Indeed, for $\tau \in \mathcal{T}_{[t,T]}$, if $\tau \leq \tau^*$ and $\mathbb{P}(\tau < \tau^*) > 0$, we have, by the definition of τ^*, $\mathbb{E}\left(e^{-\int_t^{\tau} r\,ds}\phi(S_\tau^{t,x})\right) < \mathbb{E}\left(e^{-\int_t^{\tau} r\,ds} p(\tau, S_\tau^{t,x})\right) \leq p(t,x)$. This concludes the proof of (21).

3 Pricing European options with PDEs

The aim of this section is to present two classes of methods for solving partial differential equations with some applications to the PDEs derived in Section 2. We first introduce the finite difference method which is based on approximation of the differential operators by Taylor expansions, and then the finite element methods which belong to the wider class of Galerkin methods and are based on a variational formulation of the PDE. We try to stress the most important aspects of the numerical methods, and refer for example to [3] for a more comprehensive presentation.

3.1 The finite difference method

3.1.1 Basic schemes

Let us introduce the finite difference method on the simple PDE (8). Let us first concentrate on the discretization of (8) with respect to the variable S. The principle is to divide the interval $[0, S_{\max}]$ into I intervals of length $\delta S = S_{\max}/I$ (where $S_{\max}$

has to be chosen large enough, see below), and to approximate the derivatives by finite differences. A possible semi-discretization of (8) is: for $i \in \{0, 1, \ldots, I\}$,

$$\begin{cases} \dfrac{\partial P_i}{\partial t} + rS_i \dfrac{P_{i+1} - P_{i-1}}{2\delta S} + \dfrac{\sigma^2 S_i^2}{2} \dfrac{P_{i+1} - 2P_i + P_{i-1}}{\delta S^2} - rP_i = 0, \\ P_i(T) = \phi(S_i), \end{cases} \tag{30}$$

where $S_i = i\delta S$ denotes the i-th discretization point, and $P_i(t)$ is intended to be an approximation of $p(t, S_i)$. Now, (30) is a system of coupled ordinary differential equations (ODEs). The generalization to the case of a time and spot dependent r or σ is straightforward.

Notice that for $S = 0$, P_0 can be solved independently (since $S_0 = 0$): $P_0(t) = \phi(0) \exp(-\int_t^T r\, ds)$. In order to obtain a solution of the whole system of ODEs, one needs to define an appropriate boundary condition at $S = S_{\max}$. Indeed, (30) taken at $i = I$ involves P_{I+1} which is *a priori* not defined. There are basically two methods to deal with this issue. The first one consists of using some *a priori* knowledge on the values of $p(t, S)$ when S is large and making some approximations of $p(t, S_{\max})$. In this case, the value of P_I is given as a data (this is a so-called Dirichlet boundary condition), and the unknowns are $(P_i)_{0 \le i \le I-1}$. For example, in the case of a put ($\phi(S) = (S - K)_-$) (resp. a call ($\phi(S) = (S - K)_+$)), it is known that $\lim_{S \to \infty} p(t, S) = 0$ (resp., in the limit $S \to \infty$, $p(t, S) \sim S - K \exp(-\int_t^T r\, ds)$), so that one can set $P_I(t) = 0$ (resp. $P_I(t) = S_{\max} - K \exp(-\int_t^T r\, ds)$). The error introduced by these artificial boundary conditions can be estimated. Another method is based on some knowledge on the asymptotic behavior of the derivatives of p. For example, in the case of the put, one can use the so-called homogeneous Neumann boundary condition which writes $\partial p / \partial S(t, S_{\max}) = 0$ at the continuous level, and $\frac{P_{I+1}(t) - P_I(t)}{\delta S} = 0$ at the discrete level. In this case, the unknowns are $(P_i)_{0 \le i \le I}$. For both methods, $S_{\max}$ should be chosen sufficiently large. In practice, the quality of the method may be assessed by measuring how sensitive the result is to the value of $S_{\max}$.

Let us now consider the time discretization. Here again, the idea is to divide the time interval $[0, T]$ into N intervals of length $\delta t = T/N$ and to replace the time derivative by a finite difference. Three numerical methods are classically used:

$$\begin{cases} \dfrac{P_i^{n+1} - P_i^n}{\delta t} + rS_i \dfrac{P_{i+1}^{n+1} - P_{i-1}^{n+1}}{2\delta S} + \dfrac{\sigma^2 S_i^2}{2} \dfrac{P_{i+1}^{n+1} - 2P_i^{n+1} + P_{i-1}^{n+1}}{\delta S^2} - rP_i^{n+1} = 0, \\ P_i^N = \phi(S_i), \end{cases} \tag{31}$$

$$\begin{cases} \dfrac{P_i^{n+1} - P_i^n}{\delta t} + rS_i \dfrac{P_{i+1}^n - P_{i-1}^n}{2\delta S} + \dfrac{\sigma^2 S_i^2}{2} \dfrac{P_{i+1}^n - 2P_i^n + P_{i-1}^n}{\delta S^2} - rP_i^n = 0, \\ P_i^N = \phi(S_i), \end{cases} \tag{32}$$

$$\begin{cases} \dfrac{P_i^{n+1} - P_i^n}{\delta t} + \dfrac{1}{2}\Big(rS_i \dfrac{P_{i+1}^{n+1} - P_{i-1}^{n+1}}{2\delta S} + \dfrac{\sigma^2 S_i^2}{2} \dfrac{P_{i+1}^{n+1} - 2P_i^{n+1} + P_{i-1}^{n+1}}{\delta S^2} - rP_i^{n+1} \\ \qquad\qquad + rS_i \dfrac{P_{i+1}^n - P_{i-1}^n}{2\delta S} + \dfrac{\sigma^2 S_i^2}{2} \dfrac{P_{i+1}^n - 2P_i^n + P_{i-1}^n}{\delta S^2} - rP_i^n \Big) = 0, \\ P_i^N = \phi(S_i), \end{cases} \tag{33}$$

where P_i^n is intended to be an approximation of $p(t_n, S_i)$, with $t_n = n\delta t$. Notice that using the discretization scheme (31) (the so-called explicit Euler scheme), the values of $(P_i^n)_{0\leq i\leq I}$ are explicitly obtained from the values of $(P_i^{n+1})_{0\leq i\leq I}$. On the contrary, in the two other schemes (32) (implicit Euler scheme) or (33) (Crank-Nicolson scheme), the values of $(P_i^n)_{0\leq i\leq I}$ are obtained from the values of $(P_i^{n+1})_{0\leq i\leq I}$ through the resolution of a linear system, which is more demanding from the computational viewpoint. Various numerical methods can be used for solving this linear system; here, we cannot describe them in details. Let us simply mention that basically, there exists two classes of methods: the direct methods which are based on Gaussian elimination, and the iterative methods which consist of computing the solution as the limit of a sequence of approximations and which only requires matrix-vector multiplications. The method of choice depends on the characteristics of the problem.

3.1.2 Notions of stability and consistency

In order to analyze the convergence of the three discretization schemes (31), (32) and (33), and to understand the differences between these schemes, we need to introduce two important notions. The first notion is *the consistency*. A numerical method is said to be consistent if, when the exact solution is plugged into the numerical scheme, the error tends to zero when the discretization parameters tend to zero. In our context, it consists of replacing P_i^n in (31), (32) or (33) by $p(t_n, S_i)$, where p satisfies (8), and to check that the remaining terms tend to zero when δt and δS tend to zero. By using Taylor expansions, one can check that for (31) and (32) (resp. for (33)), the remaining terms are bounded from above by $C\left(\delta t + \delta S^2\right)$ (resp. by $C\left(\delta t^2 + \delta S^2\right)$), where C denotes a constant which depends on some norms of the derivatives of p. Therefore (31) and (32) (resp. (33)) are consistent discretization schemes of order 2 in the spot variable, and of order 1 (resp. 2) in time. The second important notion is *the stability*. A numerical method is said to be stable if the norm of the solution to the numerical scheme is bounded from above by a constant (independent of the discretization parameters) times the norm of the data (initial condition, boundary conditions, right-hand side). This property is clearly satisfied if the numerical method is convergent, *i.e.* if the numerical approximation converges to the solution of the PDE when the discretization parameters tend to zero. A general result states that, conversely, a consistent and stable discretization scheme is indeed convergent. The estimate of convergence is given by the estimate of consistency error. For example, the error for the EI scheme is bounded from above by $C\left(\delta t + \delta S^2\right)$. Notice that the constant C in these estimates depends on the solution p: for high-order schemes, one needs more regularity on p. For example, for some parameters, it may happen that the results obtained with the CN scheme around $t = T$ are not better than those obtained with an order one scheme (IE or EE) since the solution is not sufficiently regular in time around $t = T$.

To give a precise meaning to all these results would require to specify the norms used to measure the errors. Let us simply mention that two norms are used in practice: the stability in L^∞-norm (the supremum of the absolute values of the components) is related to a discrete maximum principle ; and the stability in L^2-norm (the Euclidean norm of the vector) is related to an energy estimate on the variational formulation. We refer for example to [3] for more details.

$N = 1000$	$I = 150$	$I = 300$	$I = 600$	$I = 1200$
IE	0.165	0.0356	0.00103	0.000452

Table 1: Error on the value of a call in function of the number of intervals I in the variable S, for the implicit Euler (IE) scheme.

$I = 500$	$N = 5$	$N = 10$	$N = 20$	$N = 40$	$N = 80$	$N = 160$
EE	28.53	0.386	0.398	0.0739	0.0162	0.00714
IE	0.0892	0.0449	0.0225	0.0113	0.00554	0.00226
CN	0.0299	0.00758	0.00103	0.00169	0.00169	0.00168

Table 2: Error on the value of a call in function of the number of time-steps N. We observe that the Euler explicit (EE) scheme is unstable for $N = 5$. The convergence in time of the Crank- Nicolson (CN) scheme is much more faster than for the implicit Euler (IE) scheme. The remaining error when N is large is due to the discretization with respect to the variable S.

Remark 6 (Discrete maximum principle) *The discrete maximum principle is the counterpart at the discrete level of the maximum principle at the continuous level (see Remark 1). It states that if the data for the numerical schemes are positive, then the solution is positive. Such schemes are by construction stable in L^∞-norm. The numerical methods based on binomial or trinomial trees can be interpreted as explicit finite difference methods to solve the PDE (8), which naturally satisfy a discrete maximum principle.*

Let us now discuss the properties of the three discretization schemes. We already mentioned that they are all consistent. On the other hand, it can be shown that the explicit scheme (31) is stable under an additional assumption (a so-called CFL condition, see [19]) of the form $\delta t \leq C\delta S^2$, where C denotes a positive constant. The other two schemes (32) and (33) are unconditionnally stable. In conclusion, with the explicit scheme, the values of $(P_i^n)_{0\leq i\leq I}$ can be very rapidly obtained from the values of $(P_i^{n+1})_{0\leq i\leq I}$, but the time step must be sufficiently small with respect to the spot step to guarantee stability and hence convergence. On the other hand, the implicit schemes (32) and (33) require the resolution of a linear system at each time-step, but converge without any restriction on the time-step. This situation is very general for the parabolic PDEs obtained in finance. In terms of computational costs, the balance is generally in favor of the implicit schemes, since the CFL condition appears to be very stringent in practice.

In Tables 1 and 2, we illustrate this analysis by computing the error on the price of a call with $r = 0.1$, $\sigma = 0.01$, $K = 100$, $T = 1$, $S_0 = 100$ and $S_{\max} = 300$ for the three discretization schemes (31), (32) and (33), and various values of the numerical parameters I and N. The reference value ($P = 9.51625$) is obtained by the analytic Black and Scholes formula. In particular, one can check that the rates of convergence with respect to δt and δS are indeed those predicted by the analysis.

Remark 7 (Change of variable) *It is well-known that by a change of variable $x = \ln S$, it is possible to get rid of the dependency in S of the advective and diffusion terms in (8). It is not better to discretize the PDE after this change of variable, since it corresponds to take a grid refined near $S = 0$, which is useless in this case. As we will see below (see Section 3.2), what actually matters is to refine the grid around the singularity of p (*i.e. *around $S = K$). A finite element approach is better suited to implement these refinements.*

3.1.3 Application to Asian options

We now present a less easy implementation of a finite difference method for pricing Asian options (see Section 2.2.4 and [23]). More precisely, we focus on computing numerical solutions to (14) for a fixed strike call:

$$\tilde{\phi}(\xi) = \xi_-. \tag{34}$$

We have seen in the previous section that a simple finite difference scheme leads to very satisfactory result when computing the solution of the classical Black-Scholes equation (8). On the other hand, when one uses a simple finite difference scheme on (14), very bad results are obtained, especially when the volatility σ is small (see Table 1 in [23]). These bad results are due to the fact that when ξ is close to zero, the advective term $\left(\frac{1}{T} + r\xi\right)$ is much larger than the diffusion term $\sigma^2\xi^2/2$ in (14). This is known to deteriorate the stability of the numerical scheme, particularly with respect to the L^∞-norm. In practice, the numerical solution exhibits some oscillations and does not satisfy the discrete maximum principle. Moreover, the finite difference method introduces numerical diffusion which leads to unsatisfactory results for purely advective equations.

One way to handle this problem is to use a characteristic method (based on the solution of $d\xi/dt = -1/T$) in order to get rid of the term $1/T$. This means that the following change of variable is introduced:

$$g(t, x) = f(t, x - t/T). \tag{35}$$

One can easily show that g is solution of[2]:

$$\begin{cases} \dfrac{\partial g}{\partial t} + \dfrac{\sigma^2 (x - t/T)^2}{2} \dfrac{\partial^2 g}{\partial x^2} - r\,(x - t/T)\,\dfrac{\partial g}{\partial x} = 0, \\ g(T, x) = \tilde{\phi}(x - 1) = (1 - x)_+. \end{cases} \tag{36}$$

The PDE (36) satisfied by g is such that when the advective term $r\,(x - t/T)$ is small, the diffusion term $\dfrac{\sigma^2 (x - t/T)^2}{2}$ is also small. As shown below, a finite difference scheme applied to (36) will indeed lead to satisfactory results.

An important property of the solution to (36) for $\tilde{\phi}(\xi) = \xi_-$ is that (see [48]) $\forall \xi \leq 0$,

$$f(t, \xi) = \frac{1}{rT}(1 - e^{-r(T-t)}) - \xi e^{-r(T-t)} \tag{37}$$

[2]Notice that the same equation has been considered by Vecer in [53] using some financial arguments.

and therefore, $\forall x \leq t/T$,

$$g(t,x) = \frac{1}{rT}(1 - e^{-r(T-t)}) - (x - t/T)e^{-r(T-t)}. \tag{38}$$

To prove (37), one can notice that f given by (37) is solution to (14) with $\phi(\xi) = -\xi$, and that, due to the fact that the diffusion term is null for $\xi = 0$ and that the advective term is negative, the solution to (14) for $\phi(\xi) = \xi_-$ on $\xi \leq 0$ is the same as the solution to (14) for $\phi(\xi) = -\xi$ on $\xi \leq 0$.

To discretize (36), a Crank-Nicolson time scheme is used, with a uniform time step $\delta t = T/N$. In order to use the fact that g is analytically known on $x \leq t/T$ (see (38)), a mesh that properly discretizes the boundary $x = t/T$ is used. Therefore, the space interval $(0, 1)$ is also discretized with N space steps of length $\delta x = 1/N$ (see Figure 1). The mesh is completed be adding J intervals on the right hand side of $x = 1$, so that $x \in (0, x_{max})$ with $x_{max} = (N + J)\,\delta x$. The value $J = N/2$ has been found to be sufficient to guaranty the independence of the results on the position of x_{max}.

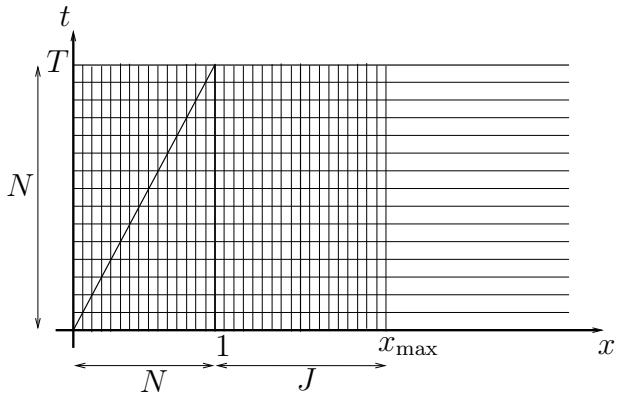

Figure 1: The mesh and the computational domain for the finite difference scheme used to discretize (36).

Notice that at time $t_n = n\delta t$, the number of unknowns is $(N + J - n)$. This means that the dimension of the linear system to solve depends on the time-step.

As far as boundary conditions are concerned, we use a Dirichlet boundary condition on $x = t/T$ (using (38)) and an artificial zero Neumann boundary condition on $x = x_{max}$.

Let us now give some numerical results. In Table 3, a few comparisons of the results obtained with the characteristic method and other methods are given. The characteristic method appears to be accurate for both small and large volatilities. For any values of the parameters, at least 5 digits of precision are obtained in less than one second. Notice that the Thompson bounds and the characteristic method are implemented in Premia [2].

3.2 The finite element method

We would like now to introduce the finite element method. This technique is more flexible than the finite difference method. In particular, it allows for local refinements

σ	K	Characteristic method	Zvan et al. [56]	Večeř [53]
0.05	95	11.09409 ($N = 300$)	11.094	11.094
	100	6.7943 ($N = 1000$)	6.793	6.795
	105	2.7444 ($N = 3000$)	2.748	2.744
0.30	90	16.512 ($N = 300$)	16.514	16.516
	100	10.209 ($N = 300$)	10.210	10.215
	110	5.7304 ($N = 1000$)	5.729	5.736

σ	K	Thompson [52] (low)	Thompson [52] (up)
0.05	95	11.094094	11.094096
	100	6.794354	6.794465
	105	2.744406	2.744581
0.30	90	16.512024	16.523720
	100	10.208724	10.214085
	110	5.728161	5.735488

Table 3: Comparisons of the prices for an Asian fixed call obtained with various methods. Values of parameters: $T = 1$, $r = 0.15$, $S_0 = 100$, $J = N/2$. For the characteristic method, the number of time-steps $N \geq 300$ needed to obtain at least 5 digits of precision is given.

of the spot grid (even in dimension greater than one, and possibly based on local error estimators, see Remark 8 below). This is particularly important for American options, because the pricing function is singular near the exercise boundary and this curve is not known a priori. Let us emphasize that the use of a refined mesh around the singularities of the solution (for example, for vanilla option pricing problems, around $t = T$ and $S = K$) is very important in practice to rapidly obtain accurate results. The finite element method can also be used in a flexible way when the geometry of the computational domain becomes complex, which may be of interest for barrier options in dimension greater than one (see Section 2.2.2). Finally, finite element methods are interesting since they are naturally stable (in L^2-norm) and optimal error bounds (in L^2-norm) can be derived.

In Section 3.2.1, we present the finite element method on a simple example, namely Equation (8). In Section 3.2.2, we show how the finite element method can be used for more complex options. The use of finite element method for American options will be presented in Section 4.2.

3.2.1 The finite element method for (8)

The conforming finite element method is based on two ingredients: a so-called variational formulation of the PDE on a functional space V and the choice of an appropriate sequence of finite dimensional spaces $V_h \subset V$ which tends to V when h (which is the typical diameter of the cells of the space mesh) tends to 0. Let us illustrate this on (8).

To derive a variational formulation of (8), the principle is to multiply the equation by a test function of the spot variable, and to integrate by parts. For these computations to be well defined, the functions need to be sufficiently smooth. We

thus introduce the functional spaces $H = L^2(\mathbb{R}_+) = \{q : [0,\infty) \to \mathbb{R}, \int_0^\infty q^2 < \infty\}$ and $V = \{q \in L^2(\mathbb{R}_+),\ S(\partial q/\partial S) \in L^2(\mathbb{R}_+)\}$. Assuming that ϕ is square integrable, a variational formulation of (8) is then (for a S-dependent volatility σ): find $p \in L^2((0,T),V) \cap \mathcal{C}^0([0,T],H)$ such that for all $q \in V$,

$$\left\{ \begin{array}{l} \dfrac{d}{dt}\displaystyle\int pq - \int \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial S}\frac{\partial q}{\partial S} + \int \left(r - \sigma^2 - S\sigma\frac{\partial \sigma}{\partial S}\right) S\frac{\partial p}{\partial S}q - r\int pq = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad p(T,S) = \phi(S). \end{array} \right. \tag{39}$$

This rewrites: find $p \in L^2((0,T),V) \cap \mathcal{C}^0([0,T],H)$ such that for all $q \in V$,

$$\left\{ \begin{array}{l} \dfrac{d}{dt}\displaystyle\int pq - a(p,q) = 0, \\ \qquad\qquad p(T,S) = \phi(S), \end{array} \right. \tag{40}$$

where a is the bilinear form

$$a(p,q) = \int \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial S}\frac{\partial q}{\partial S} - \int \left(r - \sigma^2 - S\sigma\frac{\partial \sigma}{\partial S}\right) S\frac{\partial p}{\partial S}q + r\int pq. \tag{41}$$

Under suitable assumptions on the data (r, σ and ϕ), it is possible to prove that this variational problem is well posed (see [3]).

The second step is to introduce a sequence of meshes in the spot variable indexed by the maximal step h, and related finite dimensional functional spaces $V_h \subset V$. In the case of (39), the problem is posed on an infinite domain and one needs to first localize the PDE in a finite domain $[0, S_{\max}]$ by using artificial boundary condition at $S = S_{\max}$, as already explained for finite difference discretizations (see Section 3.1.1). We consider for example a zero Neumann boundary condition on $S = S_{\max}$: $\partial p/\partial S(t, S_{\max}) = 0$. Then, a mesh of $[0, S_{\max}]$ consists of a finite number of intervals (S_i, S_{i+1}) with $S_0 = 0$ and $S_I = S_{\max}$. We set $h = \max_{0\le i\le I-1}(S_{i+1} - S_i)$. The intervals (S_i, S_{i+1}) are called elements. We then need to define a functional space V_h associated with the mesh. A classical example is the $P1$ finite element space, which contains continuous and piecewise affine functions (affine in the elements (S_i, S_{i+1})) , namely continuous functions which are affine on each interval (S_i, S_{i+1}), for $0 \le i \le I-1$. In this case, a basis of the vector space V_h is given by the so-called hat functions $q_i \in V_h$ such that for $0 \le i, j \le I$, $q_i(S_j) = \delta_{i,j} = \left\{ \begin{array}{l} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \end{array} \right.$ ($\delta_{i,j}$ is the Kronecker symbol). Notice that higher order finite element methods may be easily obtained by taking continuous and element-wise polynomial functions of degree $k > 1$.

The discretization in the spot price variable now simply consists in replacing the functional space V by the finite dimensional space V_h in (39) or (40) (this is the principle of Galerkin methods): find $p_h \in \mathcal{C}^0([0,T],V_h)$ such that for all $q_h \in V_h$,

$$\left\{ \begin{array}{l} \dfrac{d}{dt}\displaystyle\int p_h q_h - a(p_h, q_h) = 0, \\ \qquad\qquad p_h(T,S) = \phi_h(S), \end{array} \right. \tag{42}$$

where ϕ_h is an approximation of ϕ in the space V_h. One can take for example ϕ_h such that $\int(\phi - \phi_h)q_h = 0$ for all $q_h \in V_h$ (ϕ_h is then the L^2 projection of ϕ onto V_h).

Problem (42) is a finite-dimensional problem in space of the form $M_h dP_h/dt = A_h P_h$, where $P_h(t)$ is a vector of dimension I containing the values of p_h at the nodes of the mesh ($p_h(t,x) = \sum_{j=0}^{I} P_{h,j}(t) q_j(x)$) and M_h, A_h are $I \times I$ matrices. The matrix M_h (resp. A_h), with (i,j)-th component $\int q_i\, q_j$ (resp. $a(q_j, q_i)$) is classically called the mass (resp. stiffness) matrix, because the finite element method was originally popularized by the mechanical engineering community. When using the nodal basis (hat functions), these matrices are very sparse (tridiagonal for one dimensional problems). Problem (42) is somewhat similar to (30) obtained by the finite difference method; the two problems (30) and (42) are actually equivalent if a mesh with uniform space steps is used, and if M_h is replaced by a close diagonal matrix (mass-lumping).
A fundamental result (the Cea's Lemma) states that the norm of $(p - p_h)$ (the discretization error) is bounded from above by a constant times the infimum of the norm of $(p - q_h)$, over all $q_h \in V_h$ (the best fit error). Using this result, if V_h gets closer to V when h tends to 0, *i.e.* if the best fit error tends to 0 when h tends to zero, so does the discretization error. In particular, the finite element discretization is thus naturally stable in this norm. A precise meaning for this statement requires to define the norm and study the best fit error. Let us simply mention that the norms used in this context are related to the L^2-norm introduced for finite difference schemes. We refer to [3, 47] for the details. In our specific example, it is possible to prove that, if the payoff function is regular enough, then

$$\|p - p_h\|_{L^\infty([0,T],H)} + \|p - p_h\|_{L^2([0,T],V)} \leq Ch,$$

and that

$$\|p - p_h\|_{L^2([0,T],H)} \leq Ch^2.$$

For the discretization in time, the situation is exactly the same as for the finite difference method: one can use explicit Euler scheme, implicit Euler scheme or Crank-Nicolson scheme, and the rate of convergence is $O(\delta t)$ for the Euler schemes and $O(\delta t^2)$ for the Crank-Nicolson scheme.

3.2.2 Finite element methods for other options

We have introduced the finite element method in a very simple case. The aim of this section is to explain how it applies for other options.

Let us first consider basket options, or basket options with barriers, in dimension 2 and 3 (see Sections 2.2.1 and 2.2.2). The derivation of a variational formulation for (9) is very similar to the one-dimensional case. However, the construction of the mesh is much more complicated in dimension 2 and 3, than in dimension 1. It consists of partitioning the domain into non-overlapping cells (elements) whose shapes are simple and fixed (for example, triangles or quadrilaterals in dimension 2, or tetrahedra or hexahedra in dimension 3). The functional spaces V_h can then be constructed as in dimension 1, for example by considering continuous piecewise affine functions. One interest of the finite element method in this context is that it is possible to mesh any domain $\mathcal{D}$ for barrier options. In the finite difference method, to mesh non quadrilateral (or non hexahedral) domain is complicated.

Let us now consider lookback options (see Section 2.2.3). It is possible to derive a natural variational formulation for (11). This is the following (written here for a

constant volatility σ): find $p : \mathcal{D} \to \mathbb{R}$ such that, for all $q : \mathcal{D} \to \mathbb{R}$,

$$\begin{aligned} \frac{d}{dt}\int_{\mathcal{D}} pq - \int_{\mathcal{D}} \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial S}\frac{\partial q}{\partial S} - \int_{\mathcal{D}} \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial S}\frac{\partial q}{\partial M} + \int_{\mathcal{D}} \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial M}\frac{\partial q}{\partial S} \\ + \int_{\mathcal{D}} \left(r - \sigma^2\right) S\frac{\partial p}{\partial S}q - r\int_{\mathcal{D}} pq = 0, \\ p(T, S, M) = \phi(S, M), \end{aligned} \tag{43}$$

where $\mathcal{D} = \{(S, M) \in \mathbb{R}^2,\, 0 \le S \le M\}$. The boundary condition $\partial p/\partial M(t, S, S) = 0$ is naturally contained in this variational formulation since, by integration by parts over $\mathcal{D}$:

$$\begin{aligned} &- \int_{\mathcal{D}} \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial S}\frac{\partial q}{\partial S} - \int_{\mathcal{D}} \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial S}\frac{\partial q}{\partial M} + \int_{\mathcal{D}} \frac{\sigma^2 S^2}{2}\frac{\partial p}{\partial M}\frac{\partial q}{\partial S} - \int_{\mathcal{D}} \sigma^2 S\frac{\partial p}{\partial S}q \\ &= \int_{\mathcal{D}} \frac{\sigma^2 S^2}{2}\frac{\partial^2 p}{\partial S^2}q + \frac{1}{\sqrt{2}}\int_{\{S=M\}} \frac{\partial p}{\partial M}q. \end{aligned}$$

The first term corresponds to the diffusion term in (11). The second term is an integral over the boundary $\{S = M\}$ of $\mathcal{D}$ and naturally enforces the boundary condition $\partial p/\partial M(t, S, S) = 0$. On Figure 2, we represent the price of a fixed strike call, obtained using the formulation (11), an implicit Euler scheme and $P1$ finite elements.

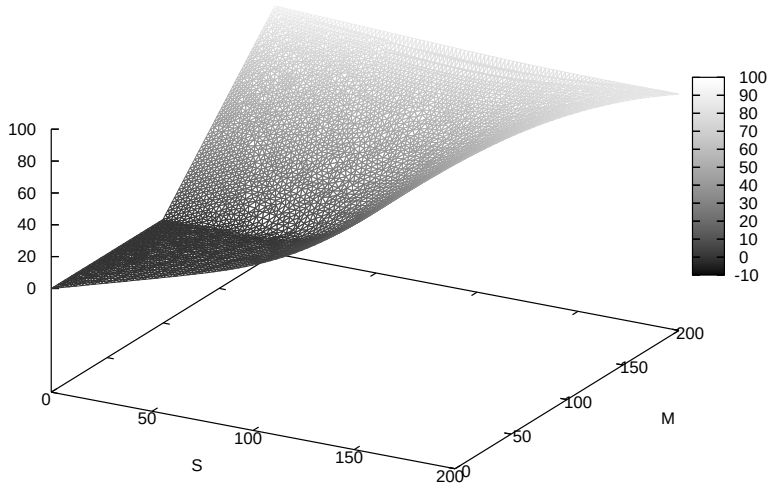

Figure 2: Price of a lookback option for a fixed strike call: $\phi(S, M) = (M - K)_+$. The parameters are: $\sigma = 0.3$, $r = 0.1$, $K = 100$, $T = 1$. Computations made with freeFEM++ [1].

Remark 8 (A posteriori error estimates) *A frequently mentioned advantage of the Monte Carlo methods is that they naturally provide* a posteriori *error bounds through an interval of confidence, typically built upon the central limit theorem. It is*

also possible to obtain such a posteriori *error estimates in the framework of the finite element method (this is one additional advantage of this method compared to finite difference methods). Moreover, these* a posteriori *estimates have two very important features:*

- *they depend on local error indicators*
- *they can be proved to be reliable and efficient,* i.e. *the actual error is bounded above and below by some fixed constants times the* a posteriori *error, and these estimates can be made local.*

Therefore, in the finite element method, the a posteriori *error estimates enable to refine the mesh in space and time adaptively. We refer to Section 4.2 for a numerical illustration for American options and to [27, 3] for more details.*

Remark 9 (High dimensional problems) *In practical problems, options often involve more than* 3 *assets. In this case, the PDE is posed in a space of dimension larger than* 4*, and the finite element or difference methods cannot be used, since the number of unknowns typically grows exponentially with respect to the problem's dimension. People often call that the* curse of dimensionality*. Let us mention that such high-dimensional problems also appear in other scientific fields, like in quantum chemistry for example, and that it is still a subject of current research to build appropriate discretizations for high-dimensional PDEs. Roughly speaking, the problem is to find an appropriate sequence of functional spaces V_h (whose basis is called a Galerkin basis), such that their dimension do not grow too rapidly with the dimension of the problem. One approach is the sparse tensor product [12, 54]. The main difficulty when using this approach is actually to project the initial condition on V_h. Another approach used in other contexts for solving high dimensional problems by deterministic methods is the low separation rank method (see [8]). Let us finally mention that another possible approach for building appropriate Galerkin basis would be the reduced basis method, where some solutions for a given set of parameters are used to approximate the solution for other values of the parameters (see for example [14]).*

4 Pricing American options with PDEs

This section is devoted to the discretization of the system (23) for the price of an American option. Notice that no closed formula such as the Black-Scholes formula are available for American put, or for American call with a dividend rate, so that efficient discretization of this system is needed even for these simple payoffs.

4.1 The finite difference approach

4.1.1 Some finite difference schemes

We consider a regular mesh discretization $S_i = i\delta S$ and a time discretization $t_n = n\delta t$ with $\delta t = \frac{T}{N}$. As in the European case, it is natural to consider the following three

iterative numerical schemes for P_i^n, an approximation of $p(t_n, S_i)$. In all cases, the scheme is initialized by $P_i^N = \phi(S_i)$. Let A be the matrix such that

$$(AP^{n+1})_i = -rS_i \frac{P_{i+1}^{n+1} - P_{i-1}^{n+1}}{2\delta S} - \frac{\sigma^2 S_i^2}{2} \frac{P_{i+1}^{n+1} - 2P_i^{n+1} + P_{i-1}^{n+1}}{\delta S^2} + rP_i^{n+1} \tag{44}$$

The Explicit Euler (EE) scheme for (23) is, for $n = N-1, N-2, \ldots, 0$,

$$\min\left(-\frac{P_i^{n+1} - P_i^n}{\delta t} + (AP^{n+1})_i,\ P_i^n - \phi(S_i) \right) = 0. \tag{45}$$

The scheme computes $P^n = (P_i^n)_{i=0,\ldots,I-1}$ from the knowledge of $P^{n+1} = (P_i^{n+1})_{i=0,\ldots,I-1}$. Similarly, we can propose an implicit Euler (IE) scheme:

$$\min\left(-\frac{P_i^{n+1} - P_i^n}{\delta t} + (AP^n)_i,\ P_i^n - \phi(S_i) \right) = 0, \tag{46}$$

and an (implicit) Crank-Nicolson (CN) scheme

$$\min\left(-\frac{P_i^{n+1} - P_i^n}{\delta t} + \frac{1}{2}\big((AP^n)_i + (AP^{n+1})_i\big),\ P_i^n - \phi(S_i) \right) = 0. \tag{47}$$

In the case of the EE scheme, it is easy to see that we have the equivalent formulation

$$P_i^n = \max\big(((I_d - \delta t A)P^{n+1})_i,\ \phi(S_i)\big), \tag{48}$$

where I_d denotes the identity matrix.

We now have two new difficulties: first, the well-posedness of the schemes (46) or (47) is not immediate (for European options, we obtained a linear system, but this is no longer true for American options), and secondly, studying the convergence is more difficult.

On way to circumvent the first difficulty is to introduce a splitting method (see [6, 5, 41]). For (23), it writes (a similar modification of (47) could also be considered, yielding a *Crank Nicolson-splitting* (CN-S) scheme):

$$\text{compute } P^{n,1} \text{ s.t.} \quad -\frac{P_i^{n+1} - P_i^{n,1}}{\delta t} + (AP^{n,1})_i = 0, \tag{49a}$$

$$\text{and then compute } P_i^n = \max(P_i^{n,1}, \phi(S_i)). \tag{49b}$$

Hereafter, (49) will be refered to as the *implicit Euler-splitting* (IE-S) scheme. The first step (49a) consists of solving a linear system, as in the European case. The second step is a projection on the set $\{v = (v_i),\ v_i \geq \phi(S_i),\ \forall i\}$, as for the EE scheme (48).

Notice that as for European options, we set the equation on a truncated domain $(0, S_{\max})$, and use artificial boundary conditions on $S = S_{\max}$ (see Section 3.1.1). We refer to [5] for error estimates between the truncated problem on $(0, S_{\max})$ and the exact problem.

4.1.2 An abstract convergence result.

Assuming for the moment that the schemes are well posed, it is possible to study the convergence in the general framework of finite different schemes for Hamilton-Jacobi equations. Possibly under some restrictions on the mesh sizes δt and δS, we can obtain convergence to the viscosity solution of the PDE (23). We refer to [4, 5] for a short introduction, and to [21] for a more detailed overview. To give a rough idea of the convergence results involved in such schemes, we consider a general Hamilton-Jacobi equation of the form

$$H\left(t, x, p, \frac{\partial p}{\partial t}, \frac{\partial p}{\partial S}, \frac{\partial^2 p}{\partial S^2}\right) = 0, \tag{50}$$

with a terminal condition on $p(T, \cdot)$, where H is assumed to be continuous and "backward parabolic" in the sense that

$$\text{if } \psi_1 \leq \psi_2 \text{ then } H(t, x, p, u, v, \psi_1) \geq H(t, x, p, u, v, \psi_2), \tag{51a}$$

$$\text{and if } u_1 \leq u_2 \text{ then } H(t, x, p, u_1, v, \psi) \geq H(t, x, p, u_2, v, \psi). \tag{51b}$$

Equation (23) indeed writes (50) with, for $(t, S) \in (0, T) \times (0, S_{\max})$, $H(t, S, p, u, v, \psi) = \min(-u - rSv - \frac{1}{2}\sigma^2 S^2 \psi + rp,\ p - \phi(S))$, which obviously satisfies (51).

First convergence results were given in the fundamental work of Crandall and Lions [22] for Lipschitz continuous final condition ϕ (and without $\frac{\partial^2 p}{\partial S^2}$ dependence in (50)).

An abstract and general convergence result is given by Barles and Souganidis in [6], which we now summarize. We first assume that H satisfies a *comparison principle*, which can be seen as an extension of the maximum principle (see Remark 1) to some nonlinear equations. The comparison principle is roughly the following [21, 4, 46]: assume that u is a subsolution of (50) (resp. supersolution), *i.e.* $H\left(t, x, u, \frac{\partial u}{\partial t}, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}\right) \leq 0$ (resp. $H\left(t, x, v, \frac{\partial v}{\partial t}, \frac{\partial v}{\partial x}, \frac{\partial^2 v}{\partial x^2}\right) \geq 0$) for $(t, S) \in (0, T) \times (0, S_{\max})$, with $u \leq 0$ (resp. $v \geq 0$) on the boundaries $S = S_{\max}$ and $t = T$, then $u \leq v$ everywhere.

Now, suppose that we can write the scheme in the abstract form: $\forall i \in \{0, \ldots, I\}$

$$\mathcal{S}_\rho(t_n, S_i, P_i^n, \tilde{P}) = 0, \tag{52}$$

where $\rho = (\delta t, \delta S)$, and $\tilde{P}$ stands for $(P_j^k)_{0 \leq k \leq N,\ 0 \leq j \leq I}$. We suppose that (52) admits at least one solution denoted P_ρ. Then, in the limit when ρ goes to zero, P_ρ converges to p solution to (50) if the following conditions are satisfied:

> (i) A *stability condition*, which reads $\max_{0 \leq n \leq N,\ 0 \leq i \leq I} |P_i^n| \leq C$, for some constant C independent of N and I (*i.e.*, independent of ρ).
> (ii) A *consistency condition*: for any regular function ψ, and $\xi \in \mathbb{R}$,
>
> $$\lim_{\rho \to 0,\, t_n \to t,\, S_i \to S} \mathcal{S}_\rho(t_n, S_i, \psi(t_n, S_i) + \xi, \tilde{\psi} + \xi) = H\left(t, S, \psi, \frac{\partial \psi}{\partial t}, \frac{\partial \psi}{\partial S}, \frac{\partial^2 \psi}{\partial S^2}\right)(t, S)$$
>
> (where $\tilde{\psi} + \xi = (\psi(t_k, S_j) + \xi)_{0 \leq k \leq N,\ 0 \leq j \leq I}$). For a weaker statement see [6].
> (iii) A *monotonicity condition*, which reads
>
> $$\tilde{P} \leq \tilde{Q} \quad \Rightarrow \quad \mathcal{S}_\rho((t_n, S_i), P_i^n, \tilde{P}) \geq \mathcal{S}_\rho((t_n, S_i), P_i^n, \tilde{Q}).$$

For most standard finantial options, a comparison principle holds. The stability and consistency conditions are close to the stability and consistency conditions already introduced in Section 3.1.2 for European options. Hence the new condition to check is the monotonicity assumption (which is related to the property (51b) satisfied by H). It is actually related to a discrete maximum principle (see Remark 6).

Remark 10 (error estimates) *Error estimates can also be obtained for the finite difference schemes* (45)–(46)–(47). *For example, for the EE scheme, an error estimate of order $\delta S^{1/2}$ in L^∞-norm can be proved under a CFL condition and for Lipschitz initial data [35]. In the context of finite element method (see Section 4.2) an error estimate of order δS^2 can be proved, but in the weaker L^2-norm.*

4.1.3 Implementation and convergence of the finite difference schemes.

It is easy to see, in view of (48), that the EE scheme is stable and monotone if the components of the matrix $(I_d - \delta t A)$ are nonnegative. This is exactly what is needed to prove a discrete maximum principle in the European case (see Remark 6). This property holds under a CFL condition of the form $\delta t \leq C\delta S^2$, C constant, and with an appropriate discretization of the advective term. The CN scheme is also stable and monotone under a CFL-like condition. On the other hand, it can be shown that the IE-S scheme as well as the IE scheme are stable and monotone without any CFL condition.

Now let us explain how to solve the implicit schemes (46) or (47) in practice. Let us consider the IE scheme (46). At each time step, setting $b = P^{n+1}$, $B = I_d + \delta t A$ and $g = (\phi(S_i))_i$, the problem is equivalent to find $x = P^n$ such that

$$\min((Bx - b)_i, (x - g)_i) = 0, \quad \forall i. \tag{53}$$

The *Howard algorithm* [32] (also called the *policy iteration algorithm*) is the method of choice to solve (53). To present this algorithm, we rewrite (53) in the following form: find x such that,

$$\min_{\alpha \in \{0,1\}^I} ((B(\alpha)x - b(\alpha))_i) = 0, \quad \forall i, \tag{54}$$

where $B_{i,j}(\alpha) = \begin{cases} B_{i,j} & \text{if } \alpha_i = 0 \\ \delta_{i,j} & \text{if } \alpha_i = 1 \end{cases}$ (where $\delta_{i,j}$ is again the Kronecker symbol, *i.e.* the (i,j)-th component of I_d) and $b_i(\alpha) = \begin{cases} b_i & \text{if } \alpha_i = 0 \\ g_i & \text{if } \alpha_i = 1 \end{cases}$. The i-th component of $B(\alpha)x - b(\alpha)$ only depends on the i-th component of α, so that the minimum for the i-th component in (54) is indeed taken with respect to the i-th component of α. Thus, for a given x and α realizing the minimum in (54), the component α_i is equal to 0 (resp. to 1) if, at the i-th node, the minimum in (53) is $(Bx - b)_i$ (resp. $(x - g)_i$). For an initial value[3] $\alpha^0 \in \{0,1\}^I$, the algorithm writes as follows: iterate for $k \geq 0$,

(i) Compute x^k such that $B(\alpha^k)x^k = b(\alpha^k)$,
(ii) $\alpha_i^{k+1} = \arg\min_{\alpha_i \in \{0,1\}} (B(\alpha)x - b(\alpha))_i$.

[3] A good initial guess is indeed the vector α obtained at the previous time iteration

We refer to [50] and [10] for some convergence results. Under suitable assumptions on the matrix B (which are satisfied for the schemes considered above which satisfy the monotonicity condition), it can be proved that this method converges in at most I iterations. In practice, only a very few iterations are needed for solving (46).

This algorithm can also be seen as:

- a Newton's method on the function F defined by $F_i(x) = \min((Bx-b)_i, (x-g)_i)$. More precisely, it is a semi-smooth Newton's method applied to the *slantly differentiable* function[4] F.
- a *primal-dual algorithm* to implement the fully Implicit Euler scheme (46) as introduced in [31].

Remark 11 (PSOR) *Another well-known method for solving* (53) *is the Projected Successive Over Relaxation (PSOR) method, which is a modification of the Successive Over Relaxation (SOR) method to solve iteratively systems of linear equations (see [49]). In its simplest form, it consists of decomposing* $B = L + U$ *where* L *is a lower triangular matrix and* U *is an upper triangular matrix with zero coefficients on the diagonal. The algorithm consists of choosing an initial guess* x^0 *and then computing iteratively for* $n \geq 1$, *for* $i = 1, \ldots, I$, $x_i^n = \arg\min \left\{ (Lx^n - (b - Ux^{n-1}))_i,\ (x^n - g)_i \right\}$. *This method converges only if* B *is a diagonal dominant matrix, and the convergence is rather slow in practice.*

Remark 12 (The one-dimensional case) *For the specific case of an American put option on a single asset, a fast algorithm was proposed by Brennan and Schwartz [11] for solving* (46) *and proved to converge in [34]. Also in this case it can be proved that the region of exercise (namely* $\Gamma_t = \{x \in \mathbb{R}_+, p(t,x) > \phi(x)\}$*), is of the form* $\Gamma_t = [\gamma(t), \infty[$ *where* γ *is continuous under some regularity assumption of the data. Then the Howard algorithm takes a simple form which is known as the front-tracking algorithm (see for instance [3]). However these algorithms are very specific to the one-dimensional case and do not apply for general payoff functions.*

4.1.4 Numerical results for the American put option

In Table 4, we give numerical results obtained with the IE-S and IE schemes, for an American put option with $r = 0.1$, $\sigma = 0.1$, $K = 100$, $T = 1$, $S = 100$ and $S_{\max} = 150$. We have computed all error values by taking the reference value $P = 1.63380$ (obtained with a Cox-Ross-Rubinstein algorithm [20, 39] with $N = 10^5$, CPU-time= $1750\,s.$). In this example, the IE scheme is one digit more accurate than the IE-S scheme. With these numerical parameters, the EE scheme would yield bad results since the CFL condition is not respected. The IE scheme has been implemented using the Howard algorithm. The remaining error when I is large is due to the time discretization.

In Table 5, we compare the EE, IE-S and IE schemes. Since the error is of order of $O(\delta t) + O(\delta S^2)$, we have used parameters N and I such that $\delta t \simeq \delta S^2$ (*i.e.*

[4] F is slantly differentiable if there exist $C > 0$ and a matrix $G(x)$ such that $\forall x$, $||G(x)^{-1}||_\infty < C$ and $F(x+h) = F(x) + G(x+h)h + o(h)$ as $h \to 0$. Here $G(x)$ can be defined by $G(x)_{ij} = B_{ij}$ if $(Bx-b)_i \leq (x-g)_i$, and $G(x)_{ij} = \delta_{ij}$ otherwise.

$(N = 1000)$	$I = 100$	$I = 200$	$I = 400$	$I = 800$	$I = 1600$
IE-S	0.00267	0.0361	0.00180	0.00210	0.00210
IE	0.00379	0.0146	0.00011	0.00024	0.00018

Table 4: Error on the value of an American put in function of the number I of intervals in the variable S (and for $N = 1000$).

	$I = 100$ $N = 100$	$I = 200$ $N = 400$	$I = 400$ $N = 1600$	$I = 800$ $N = 6400$	$I = 1600$ $N = 25000$
EE	0.00593	0.00069	0.00045	0.00003	0.00003
CPU-time (sec.)	0.01	0.10	0.5	2.6	10.7
IE-S	0.01177	0.00616	0.00098	0.00029	0.00007
CPU-time (sec.)	0.05	0.22	1.23	7.06	44.31
IE	0.00201	0.00181	0.00016	0.00004	0.00001
CPU-time (sec.)	0.2	0.9	7.3	75.0	1033.0

Table 5: Error and CPU-times for the value of an American put as a function of the number N of time-steps N and the number I of intervals in the variable S.

$N \simeq I^2$), and such that the CFL condition is satisfied. We remark that the EE scheme gives satisfactory results in less than a few seconds. The IE is more accurate but more costly than IE-S. Hence in view of the CPU-time it is more advantageous here to use simply the EE or the IE-S scheme. This conclusion holds for a finite difference discretization, but may be different for a finite element discretization, or for another set of parameters.

Remark 13 (Markov Chains approximations) *Markov Chain schemes [38] are based on the approximation of the Dynamic Programming principle between times t and $t + \delta t$ and on the use of a spatial interpolation over a mesh (S_j). This leads to another class of schemes that are also in finite difference form. Their convergence can be established by showing the convergence to the Dynamic Programming equation, or by using the Barles-Souganidis theorem [6]. Finite difference schemes enter this framework as well as Semi-Lagrangian schemes [15, 28] (an inversed CFL condition, typically of the form $\delta S^2/\delta t \overset{\delta t, \delta S \to 0}{\longrightarrow} 0$ can then be needed). Notice that the Cox-Ross-Rubinstein algorithm [20] can also be seen as a discrete Markov Chain approximation scheme using a very particular spatial mesh such that no interpolation appears at the end.*

Remark 14 (Portfolio optimization) *A portfolio optimization problem (or stochastic optimization problem) is typically of the form*

$$p(t,x) = \max_{\alpha \in L^\infty([t,T],K)} \mathbb{E}\left(\int_t^T f(u, X_u^{t,x,\alpha}, \alpha(u))\, du + e^{-\int_t^T r(s)\, ds} \phi(X_T^{t,x,\alpha}) \right) \qquad (55)$$

where K is compact, and with

$$\begin{cases} dX_u^{t,x,\alpha} = b(u, x, \alpha(u))du + \sigma(u, x, \alpha(u))dW_u, \quad u \geq t, \\ X_t^{t,x,\alpha} = x. \end{cases}$$

The corresponding PDE can be shown to be

$$\min_{\alpha\in K}\left(-\frac{\partial p}{\partial t}-\frac{1}{2}\sigma^2(t,S,\alpha)\frac{\partial^2 p}{\partial S^2}-b(t,S,\alpha)\frac{\partial p}{\partial S}+rp-f(t,S,\alpha)\right)=0$$

in the viscosity sense (see [45]). Finite difference schemes similar to those presented above for American options can be applied. Implicit schemes, if considered, can be solved by the Howard algorithm. This can also be generalized to optimal stopping time problem, adding in (55) *a supremum over stopping times with values in* $[t,T]$ *(as in* (15)*). For such general HJB equations, a discretization by a finite element approach is not always possible because an appropriate variationnal formulation cannot always be obtained (see [7]).*

4.2 The finite element approach

As in the European case, the finite element approach requires a variational formulation of the PDE (23). Let us consider the case of the American put option. Let V be the functional space as in (3.2.1) and

$$K=\{q\in V,\ q\geq\phi\}.$$

We first notice that (23) is equivalent to the set of inequalities[5] (together with $p(T,S)=\phi(S)$)

$$\begin{cases} p-\phi\geq 0, \\ -\dfrac{\partial p}{\partial t}+\mathcal{A}p\geq 0, \\ \left(-\dfrac{\partial p}{\partial t}+\mathcal{A}p\right)(p-\phi)=0. \end{cases} \tag{56}$$

We can check that this is equivalent (for sufficiently smooth function p) to the following variational formulation for (23): find $p\in L^2([0,T],K)\cap C^0([0,T],L^2(\mathbb{R}_+))$ such that for all $t\in[0,T)$,

$$\forall q\in K,\ -\int\frac{\partial p}{\partial t}(q-p)+a(p,q-p)\geq 0 \tag{57}$$

where a is the bilinear form (41) defined above (recall that for compactly supported functions p and q, $a(p,q)=\int\mathcal{A}p\,q$), with the final condition

$$p(T,S)=\phi(S).$$

Indeed, by writing $q-p=(q-\phi)-(p-\phi)$, it is clear that (56) implies (57). Conversely, choosing a sufficiently large $q\in K$ in (57), we obtain that $-\frac{\partial p}{\partial t}+\mathcal{A}p\geq 0$. Taking then $q=\phi$ in (57), we obtain that $\left\langle-\frac{\partial p}{\partial t}+\mathcal{A}p,\ \phi-p\right\rangle\geq 0$, but this inequality is actually an equality since $-\frac{\partial p}{\partial t}+\mathcal{A}p\geq 0$ and $\phi-p\leq 0$.

Notice that if we take $K=V$ in (57), we recover the variational formulation (40) for European option. Precise existence and uniqueness results for such variational inequalities can be found in [7]. For results and applications in the finance context, we refer to [3].

[5] Such a problem is called a linear complementarity problem.

Now, as in Section 3.2.1, we introduce a sequence of finite dimensional functional spaces $V_h \subset V$, such that the functions in V are better and better approximated by functions in V_h as h goes to 0. One can for example consider a P1 finite element space on a mesh $(S_i)_{0\leq i\leq I}$. Remember that a basis of V_h is given by a set of functions $(q_i)_{0\leq i\leq I}$. As in Section 3.2.1, the finite element approximation of (57) is obtained by replacing V by V_h: find $p_h \in \mathcal{C}^0([0,T], K\cap V_h)$ such that for all $t\in[0,T)$,

$$\forall q_h \in K\cap V_h,\ -\int \frac{\partial p_h}{\partial t}(q_h - p_h) + a(p_h, q_h - p_h) \geq 0, \tag{58}$$

with the final condition $p_h(T) = \phi_h$, where $\phi_h \in V_h$ is an approximation of ϕ.

For time discretization, one can again use the schemes we have introduced in Section 3.1.1. For example, the implicit Euler scheme applied to (58) is naturally defined as follows: find $p_h^N, p_h^{N-1}, \ldots, p_h^0$ in $V_h\cap K$ such that $p_h^N = \phi_h$ (initialization) and, for $n = N-1,\ldots,0$:

$$\forall q_h \in V_h\cap K, \quad -\int \frac{p_h^{n+1} - p_h^n}{\delta t}(q_h - p_h^n) + a(p_h^n, q_h - p_h^n) \geq 0. \tag{59}$$

One can easily check that

$$q_h \in V_h\cap K \quad \Leftrightarrow \quad q_h \in V_h \text{ and } q_h(S_i) \geq \phi(S_i),\ \forall i.$$

Denoting A_h and M_h the mass and stiffness matrices as in Section 3.2, and reasoning as for the equivalence between (23), (56) and (57), it can be checked that (59) is equivalent to solve in $\mathbb{R}^I$:

$$\min\left(\big(-M_h\frac{P^{n+1}-P^n}{\delta t} + A_hP^n\big)_i,\ \big(P^n - g\big)_i\right) = 0, \quad \forall i,$$

where $g_i = \phi(S_i)$ and $P_i^n = p_h^n(S_i)$. Equivalently, the problem is to find P^n such that

$$\min\big(\big((M_h + \delta t A_h)P^n - M_hP^{n+1}\big)_i,\ \big(P^n - g\big)_i\big) = 0, \quad \forall i.$$

This is a similar problem as for the IE finite difference scheme (see (53)) where the matrix $(I_d + \delta t A)$ is now replaced by $(M_h + \delta t A_h)$. It can be solved by the Howard algorithm presented in Section 4.1.3. For the particular American put problem under some assumptions on the mesh steps δt and h, it can also be solved by the Brennan and Schwartz algorithm or the front-tracking algorithm mentioned in Remark 12 above.

Notice that a Crank-Nicolson scheme can be derived in a similar way. The expected error (in L^2-norm) is (as in the European case) $O(h^2) + O(\delta t)$ for the IE scheme and $O(h^2) + O(\delta t^2)$ for the CN scheme.

We conclude this section by a numerical illustration of the mesh refinement procedure mentioned in Remark 8 above applied to the pricing of an American option on two assets. Such an automatic refinement procedure is particularly useful for American options because the pricing function is not smooth at any given time $t\in[0,T]$. Figures 3 and 4 illustrate such a mesh refinement for a typical two assets American option with payoff $\phi(S_1,S_2) = (K - \max(S_1,S_2))_+$. The artificial boundary Γ_0 is $\{\max(S_1,S_2) = \bar{S} = 200\}$. Homogeneous Dirichlet conditions are imposed on Γ_0.

We have chosen two examples. In the first example, the parameters are $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $r = 0.05$, $\rho = -0.6$ and $K = 100$. In the second example, the parameters are $\sigma_1 = \sigma_2 = 0.2$, $r = 0.05$, $\rho = 0$ and $K = 50$. The implicit Euler scheme has been used with a uniform time step of $1/250$ year. For the variables S_1 and S_2, we have used adaptive finite elements. For solving the linear complementarity problems, we have used the Howard algorithm. Mesh adaption in the (S_1, S_2) variable has been performed every $1/10$ year. In Figure 3, we have plotted the adapted mesh (left) and the contours of the pricing function (right) one year to maturity for the first example. Note that the contours exhibit right angles in the exercise region. In Figure 4, we plot the exercise region one year to maturity for the first example (left) and for the second example (right). One sees that the exercise boundary has singularities. It is also visible that the mesh has been adapted near the exercise boundary.

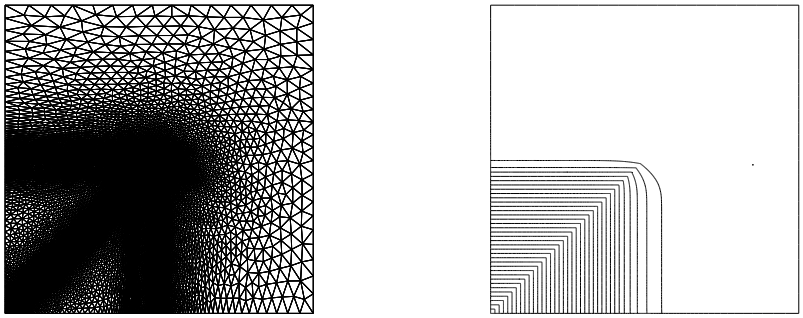

Figure 3: The adapted mesh and the contours of P one year to maturity. $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\rho = -0.6$.

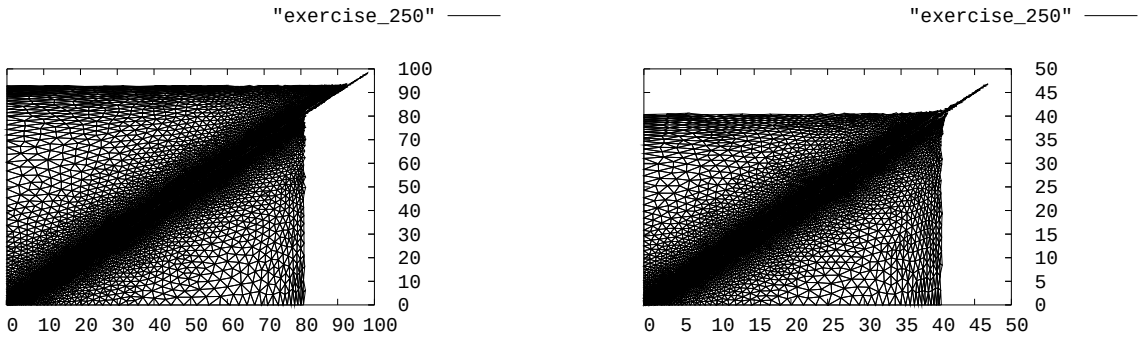

Figure 4: The exercise region one year to maturity. Left: $K = 100$, $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\rho = -0.6$. Right: $K = 50$, $\sigma_1 = \sigma_2 = 0.2$, $\rho = 0$.

5 Calibration

5.1 Limitation of the Black-Scholes model: the need for calibration

Consider a European style option on a given stock with a maturity T and a payoff function ϕ, and assume that this option is on the market. Call p its present price. Also, assume the risk-free interest rate is the constant r. One may associate with p the so-called *implied volatility*, *i.e.* the volatility σ_{imp} such that the price given by formula (4) at time $t = 0$ with $\sigma = \sigma_{imp}$ coincides with p. If the Black-Scholes model was sharp, then the implied volatility would not depend on the payoff function ϕ. Unfortunately, for *e.g.* vanilla European puts or calls, it is often observed that the implied volatility is far from constant. Rather, it is often a convex function of the strike price. This phenomenon is known as the *volatility smile*. A possible explanation for the volatility smile is that the deeply out-of-the-money options are less liquid, thus relatively more expensive than the options in-the-money.
This shows that the critical parameter in the Black-Scholes model is the volatility σ. Assuming σ constant and using (8) often leads to poor predictions of the options'prices. The volatility smile is the price paid for the too great simplicity of Black-Scholes'assumptions.
Let us now discuss some of the possible enrichment of the Black-Scholes model that have been mentioned in Remark 3:

- local volatility models: the volatility is a function of time and of the spot price, *i.e.* $\sigma_t = \sigma(t, S_t)$. With suitable assumptions on the regularity and the behavior at infinity of the function σ, (4) holds, and $P_t = p(t, S_t)$, where p satisfies the final value problem (8), in which σ varies with t and S. Calibrating the model consists of tuning the function σ in such a way that the prices computed *e.g.* with the PDE coincide with the observed prices.

- stochastic volatility models: one assumes that $\sigma_t = f(y_t)$, where y_t is a continuous time stochastic process, correlated or not to the process driving S_t, see [29] for a nice presentation. Several such models have been proposed, among which

  1. Hull-White model [33]: $f(y) = \sqrt{y}$ and y_t is a lognormal process.
  2. Scott model: $f(y) = \sqrt{y}$ and y_t is a mean reverting Ornstein-Uhlenbeck process:
  $$dy_t = \alpha(m - y_t)dt + \beta dZ_t, \tag{60}$$
  where α and β are positive constants, Z_t is a Brownian motion.
  3. Heston model [30]: $f(y) = \sqrt{y}$ and y_t is a Cox-Ingersoll-Ross process,
  $$dy_t = \kappa(m - y_t)dt + \lambda\sqrt{y_t}dZ_t, \tag{61}$$
  where κ, m and λ are positive constants.
  4. Stein-Stein model[51]: $f(y) = \sqrt{y}$ and y_t is a mean reverting Ornstein-Uhlenbeck process.

There are two risk factors, one for the stock price and the other for the volatility. If the two driving processes are not completely correlated, it is not possible to construct a hedged portfolio containing simply one option and shares of the underlying asset. One says that the market is *incomplete*. Nevertheless, if one fixes the contribution of the second source of randomness dZ_t to the risk premium, *i.e.* the *market price of the volatility risk* or the *risk premium factor* as a function of t, S_t and y_t, then it is possible to prove that the option's price is of the form $P_t = p(t, S_t, y_t)$, where the pricing function satisfies a PDE in the variables (t, S, y). The PDE may be degenerate for the values of y corresponding to volatility cancellation. Calibrating the model consists of tuning the parameters of the process y_t and the function f in order to match the observed prices.

- Lévy driven spot price: one may generalize the Black-Scholes model by assuming that the spot price is driven by a more general stochastic process, *e.g.* a Lévy process [17, 43, 16]. Lévy processes are processes with stationary and independent increments which are continuous in probability. For a Lévy process X_τ on a filtered probability space with probability $\mathbb{P}^*$, the Lévy-Khintchine formula says that there exists a function $\chi : \mathbb{R} \to \mathbb{C}$ such that

$$\mathbb{E}^*(e^{iuX_\tau}) = e^{-\tau\chi(u)},$$
$$\chi(u) = \frac{\sigma^2 u^2}{2} - i\beta u + \int_{|z|<1} (e^{iuz} - 1 - iuz)\nu(dz) + \int_{|z|>1} (e^{iuz} - 1)\nu(dz),$$

for $\sigma \geq 0$, $\beta \in \mathbb{R}$ and a positive measure ν on $\mathbb{R}\backslash\{0\}$ such that $\int_{\mathbb{R}} \min(1, z^2)\nu(dz) < +\infty$. The measure ν is called the Lévy measure of X. We focus on Lévy measure with a density, $\nu(dz) = k(z)dz$. It is assumed that the discounted price of the risky asset is a square integrable martingale under $\mathbb{P}^*$, and that it is represented as the exponential of a Lévy process:

$$e^{-r\tau} S_\tau = S_0 e^{X_\tau}.$$

The martingale property is that $\mathbb{E}^*(e^{X_\tau}) = 1$, *i.e.*

$$\int_{|z|>1} e^z \nu(dz) < \infty, \quad \text{and} \quad \beta = -\frac{\sigma^2}{2} - \int_{\mathbb{R}} (e^z - 1 - z1_{|z|\leq 1})k(z)(dz).$$

and the square integrability comes from the condition $\int_{|z|>1} e^{2z} k(z)dz < \infty$. With such models, the pricing function for a European option is obtained by solving a partial integrodifferential equation (PIDE), with a nonlocal term. Calibrating the model consists of tuning σ and the function k in such a way that the prices computed *e.g.* with the PIDE match the observed prices, see [18].

5.2 Local volatility and Dupire's formula

We consider a local volatility model and call $(t, S) \mapsto C(t, S, \tau, x)$ the pricing function for a vanilla European call with maturity τ and strike x. It satisfies the final

value problem:

$$\begin{array}{rcll} \dfrac{\partial C}{\partial t} + \dfrac{\sigma^2(t,S)S^2}{2}\dfrac{\partial^2 C}{\partial S^2} + (r-q)S\dfrac{\partial C}{\partial S} - rC &=& 0 & \text{in } [0,\tau)\times\mathbb{R}_+, \\ C(\tau,S) &=& (S-x)_+ & \text{in } \mathbb{R}_+, \end{array} \tag{62}$$

where we have supposed that the underlying asset yields a distributed dividend, $qS_t dt$. By reasoning directly on (4) or by using PDE arguments, it can be proved that the function $(\tau, x) \mapsto C(t, S, \tau, x)$ (now t and S are fixed) satisfies the forward parabolic PDE:

$$\frac{\partial C}{\partial \tau} - \frac{1}{2}\sigma^2(\tau,x)x^2\frac{\partial^2 C}{\partial x^2} + (r-q)x\frac{\partial C}{\partial x} + qC = 0, \tag{63}$$

for $\tau > t$ and $x \in \mathbb{R}_+$. This observation was first made by B. Dupire [25], and the proof of (63) by PDE arguments can be found in [3].
Using (63) is useful for two reasons:
First, consider a family of calls on the same stock with different maturities and strikes (τ_i, x_i), $I \in I$, where I is a finite set. Assume that the spot price is known, *i.e.*, $S = S_0$. In order to numerically compute the prices of the calls, *i.e.* $C(0, S_0, \tau_i, x_i)$, $i \in I$, one may solve (63) for $\max_{i\in I}\tau_i > \tau > 0$ and initial data $C(\tau = 0, x) = (S_0 - x)_+$ with *e.g.* a finite difference or a finite element method. Only one initial value problem is needed. On the contrary, using (8) would necessitate solving $\#I$ initial value problems. We see that (63) may save a lot of work.
Second, (63) may be used for local volatility calibration. Indeed, if all the possible vanilla options were on the market, the local volatility in (62) could be computed:

$$\sigma^2(\tau,x) = 2\frac{\frac{\partial C}{\partial \tau}(\tau,x) + (r-q)x\frac{\partial C}{\partial x}(\tau,x) + qC(\tau,x)}{x^2\frac{\partial^2 C}{\partial x^2}(\tau,x)}. \tag{64}$$

This is known as *Dupire's* formula for the local volatility. In practice, (64) cannot be used directly because only a finite number of options are on the market.
Assume that the observations are the prices $(\bar{C}_i)_{i\in I}$ of a family of calls with maturity/strike $(\tau_i, x_i)_{i\in I}$. Finding a function $(\tau, x) \mapsto \sigma(\tau, x)$ such that the solution of (63) with $C(0, x) = (S_0 - x)_+$ takes the value $\bar{C}_i$ at (τ_i, x_i), $i \in I$ is called an *inverse problem.*
A natural idea is to somehow interpolate the observed prices by a sufficiently smooth function $\bar{C} : [0, \max_{i\in I}\tau_i] \to \mathbb{R}_+$, then use (64) with $C = \bar{C}$. For examples, bicubic splines may be used. This approach has several serious drawbacks:

- it is difficult to design an interpolation process such that $\frac{\partial^2 \bar{C}}{\partial x^2}$ does not take the value 0, and such that the right hand side of (64) is non negative.
- There is an infinity of possible interpolations of $\bar{C}_i$ at (τ_i, x_i), $i \in I$, and for two possible choices, the volatility obtained by (64) may differ considerably.

We see that financially relevant additional information have to be added to the interpolation process.

5.3 Least-square methods

Here, we show how (63) can be used for calibration. The first idea is to use least squares, *i.e.* to minimize a functional $J : \sigma \mapsto \sum_{i\in I} \omega_i |\bar{C}_i - C(\tau_i, x_i)|^2$ for σ in a suitable function set Σ, where ω_i are positive weights, and the pricing function C is the solution of (63) with $C(0, x) = (S_0 - x)_+$. The evaluation of J requires the solution of an initial value problem. The set Σ where the volatility is to be found must be chosen in order to ensure that from a minimizing sequence one can extract at least a subsequence that converges in Σ, and that its limit is indeed a solution of the least square problem. For example, Σ may be a compact subset of a Hilbert space W (in principle W could be a more general Banach space but it is easier to work in Hilbert spaces if gradients are needed) such that the mapping J is continuous in W. In practice, W has a finite dimension and is compactly embedded in the space of bounded and continuous functions σ such that $x\partial_x\sigma$ is bounded. Thus, the existence of a solution to the minimization problem is most often guaranteed. What is more difficult to guarantee is uniqueness and stability: is there a unique solution to the least square problem? If yes, is the solution insensitive to small variations of the data? the answer to these questions is no in general, and we say that the problem is *ill-posed*.

As a possible cure to ill-posedness, one usually modifies the problem by minimizing the functional $\sigma \mapsto J(\sigma) + J_R(\sigma)$ instead of J, where J_R is a sufficiently large strongly convex functional defined on W and containing some financially relevant information. For example, one may choose $J_R(\sigma) = \omega\|\sigma - \bar{\sigma}\|^2$, where ω is some positive weight, $\|.\|$ is a norm in W and $\bar{\sigma}$ is a prior local volatility, which may come from a historical knowledge. The difficulty is that ω must not be too large not to perturb the inverse problem too much, but not too small to guarantee some stability. The art of the practitioner lies in the choice of J_R.

Once the least square problem is chosen, we are left with proposing a strategy for the construction of minimizing sequences. If J and J_R are C^1 functional, then gradient methods may be used. The drawbacks and advantages of such methods are well known: on the one hand, they do not guarantee convergence to the global minimum if the functional is not convex, because the iterates can be trapped near a local minimum. On the other hand, they are fast and accurate when the initial guess is close enough to the minimum. For these reasons, gradient methods are often combined with techniques that permit to localize the global minimum but that are slow, like simulated annealing or evolutionary algorithms.

Anyhow, gradient methods require the evaluation of the functional's gradient. Since J_R explicitly depends on σ, its gradient is easily computed. The gradient of J is more difficult to evaluate, because the prices $C(\tau_i, x_i)$ depend of σ in an indirect way: one needs to evaluate the variations of $C(\tau_i, x_i)$ caused by a small variation of σ; calling $\delta\sigma$ the variation of σ and δC the induced variation of C, one sees by differentiating (63) that $\delta C(\tau = 0, \cdot) = 0$ and

$$\partial_\tau \delta C - \frac{\sigma^2(\tau, x)x^2}{2}\partial^2_{xx}\delta C + (r - q)x\partial_x\delta C + q\delta C = \sigma\delta\sigma x^2\partial^2_{xx}C. \tag{65}$$

To express δJ in terms of $\delta\sigma$, an adjoint state function P is introduced, as the solution to the adjoint problem: find the function P such that $P(\bar{\tau}, \cdot) = 0$ and for

$\tau < \bar{\tau}$,

$$\partial_\tau P + \partial^2_{xx}\big(\frac{\sigma^2 x^2}{2} P\big) - \partial_x\big(P(r-q)x\big) - qP = 2\sum_{i\in I} \omega_i (C(\tau_i, x_i) - \bar{C}_i)\delta_{\tau_i, x_i}, \qquad (66)$$

where $\bar{\tau}$ is an arbitrary time greater than $\max_{i\in I} \tau_i$ and in the right hand side, the δ_{τ_i,x_i} denote Dirac functions in time and strike at (τ_i, x_i). The meaning of (66) is the following:

$$-\int_Q \left(\partial_\tau v - \frac{\sigma^2 x^2}{2}\partial^2_{xx} v + (r-q)x\partial_x v + qv\right) P = 2\sum_{i\in I} \omega_i (C(\tau_i, x_i) - \bar{C}_i) v(\tau_i, x_i). \qquad (67)$$

where $Q = (0, \bar{\tau}) \times \mathbb{R}_+$, and v is any function such that $v \in L^2((0,\bar{\tau}), V)$ with $\partial_\tau v \in L^2(Q)$ and $x^2\partial^2_{xx} v \in L^2(Q)$. Taking $v = \delta C$ in (67) and using (65), one finds

$$\begin{aligned} 2\sum_{i\in I} \omega_i (C(\tau_i, x_i) - \bar{C}_i)\delta C(\tau_i, x_i) &= 2\sum_{i\in I} \omega_i (C(\tau_i, x_i) - c_i)\langle \delta_{\tau_i,x_i}, \delta C\rangle \\ &= -\int_Q \left(\partial_\tau \delta C - \frac{\sigma x^2}{2}\partial^2_{xx}\delta C + (r-q)x\partial_x \delta C + q\delta C\right) P \\ &= -\int_Q \sigma\delta\sigma x^2 P \partial^2_{xx} C. \end{aligned}$$

We have worked in a formal way, but all the integrations above can be justified. This leads to the estimate

$$\left|\delta J + \int_Q \sigma\delta\sigma x^2 P\partial^2_{xx} C\right| \le c\|\delta\sigma\|^2_{L^\infty(Q)},$$

which implies that J is differentiable, and that its differential at point σ is given by

$$DJ(\sigma) : \eta \mapsto -\int_Q \sigma\eta x^2 P(\sigma)\partial^2_{xx} C(\sigma),$$

where $P(\sigma)$ satisfies (66), and $C(\sigma)$ satisfies (63). We see that the gradient of J can be evaluated. When (63) is discretized with *e.g.* finite elements, all what has been done can be repeated (with a discrete adjoint problem), and the gradient of the functional can be evaluated in the same way. Let us stress that the gradient $DJ(\sigma)$ is computed exactly, which would not be the case with *e.g.* a finite difference method.

Local volatility can also be calibrated with American options, but it is not possible to find the analogue of Dupire's equation. Thus, in the context of a least square approach, the evaluation of the cost function requires the solution of $\#I$ variational inequalities, which is computationally expensive, see [3]. In this case, it is also possible to find necessary optimality conditions involving an adjoint state.

Acknowledgements. O.B. would like to thank H. Pham and F. Delarue for very helpfull discussions.

References

[1] freefem. http://www.freefem.org/.

[2] Premia: an option pricer. http://cermics.enpc.fr/~premia/index.html.

[3] Y. Achdou and O. Pironneau. *Computational methods for option pricing.* Frontiers in applied mathematics. SIAM, 2005.

[4] G. Barles. *Solutions de viscosité des équations de Hamilton-Jacobi*, volume 17 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Paris, 1994.

[5] G. Barles, Ch. Daher, and M. Romano. Convergence of numerical schemes for parabolic equations arising in finance theory. *Math. Models Methods Appl. Sci.*, 5(1):125–143, 1995.

[6] G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Anal.*, 4(3):271–283, 1991.

[7] A. Bensoussan and J.-L. Lions. *Applications des inéquations variationnelles en contrôle stochastique.* Dunod, Paris, 1978. Méthodes Mathématiques de l'Informatique, No. 6.

[8] G. Beylkin and M.J. Mohlenkamp. Numerical operator calculus in higher dimensions. *Proc Natl Acad Sci*, 99(16):10246–10251, 2002.

[9] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

[10] O. Bokanowski, S. Maroso, and H. Zidani. Super linear convergence results for the Howard algorithm. *Preprint*, 2007.

[11] M. J. Brennan and E. S. Schwartz. The valuation of the american put option. *J. of finance*, 32:449–462, 1977.

[12] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[13] E. Cancès, B. Jourdain, and T. Lelièvre. Quantum Monte Carlo simulations of fermions. A mathematical analysis of the fixed-node approximation. *Mathematical Models and Methods in Applied Sciences*, 16(9):1403–1440, 2006.

[14] E. Cancès, C. Le Bris, Y. Maday, and G. Turinici. Towards reduced basis approaches in ab initio electronic structure computations. *Journal of Scientific Computing*, 17(1-4):461–469, 2002.

[15] I. Capuzzo-Dolcetta and M. Falcone. Discrete dynamic programming and viscosity solutions of the Bellman equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 6:161–183, 1989. Analyse non linéaire (Perpignan, 1987).

[16] P. Carr, H. Geman, D. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation. *Journal of business*, 75(2):305–332, 2002.

[17] R. Cont and P. Tankov. *Financial modelling with jump processes.* Chapman and Hall, 2003.

[18] R. Cont and P. Tankov. Nonparametric calibration of jump-diffusion option pricing models. *Journal of Computational Finance*, 7(3):1–49, 2004.

[19] R. Courant, K. Friedrichs, and H. Lewy. On the partial difference equations of mathematical physics. *IBM J. Res. Develop.*, 11:215–234, 1967.

[20] J. Cox, S. Ross, and M. Rubinstein. Option princing: A simplified approach. *J. Financial Economics*, 7:44–50, 1979.

[21] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[22] M. G. Crandall and P.-L. Lions. Two approximations of solutions of Hamilton-Jacobi equations. *Math. Comp.*, 43(167):1–19, 1984.

[23] F. Dubois and T. Lelièvre. Efficient pricing of Asian options by the PDE approach. *Journal of Computational Finance*, 8(2):55–64, 2005.

[24] D. Duffie. *Dynamic asset pricing theory.* Princeton University Press, 1992.

[25] B. Dupire. Pricing with a smile. *Risk*, pages 18–20, 1994.

[26] N. El Karoui. Les aspects probabilistes du contrôle stochastique. In *Ninth Saint Flour Probability Summer School—1979 (Saint Flour, 1979)*, volume 876 of *Lecture Notes in Math.*, pages 73–238. Springer, Berlin, 1981.

[27] A. Ern, S. Villeneuve, and A. Zanette. Adaptive finite element methods for local volatility european option pricing. *Int. J. Theor. Appl. Finance*, 7(6):659–684, 2004.

[28] M. Falcone and R. Ferretti. Discrete time high-order schemes for viscosity solutions of Hamilton-Jacobi-Bellman equations. *Numer. Math.*, 67(3):315–344, 1994.

[29] J.-P. Fouque, G. Papanicolaou, and R. Sircar. *Derivatives in financial markets with stochastic volatility.* Cambridge University Press, 2000.

[30] S. Heston. A closed form solution for options with stochastic volatility with application to bond and currency options. *Review with Financial Studies*, 6:327–343, 1993.

[31] M. Hintermüller, K. Itô, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Opim.*, 13(3):865–888 (electronic) (2003), 2002.

[32] R. A. Howard. *Dynamic programming and Markov processes.* The Technology Press of M.I.T., Cambridge, Mass., 1960.

[33] J.C. Hull and A. White. The pricing of options on assets with stochastic volatilities. *J. Finance*, 42:281–300, 1987.

[34] P. Jaillet, D. Lamberton, and B. Lapeyre. Variational inequalities and the pricing of American options. *Acta Appl. Math.*, 21(3):263–289, 1990.

[35] E. R. Jakobsen. On the rate of convergence of approximation schemes for Bellman equations associated with optimal stopping time problems. *Math. Models Methods Appl. Sci.*, 13(5):613–644, 2003.

[36] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, 1988.

[37] I. Karatzas and S.E. Shreve. *Methods of mathematical finance*. Springer, 1998.

[38] H. J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*. Applications of Mathematics Vol. 24. Springer-Verlag, 2001.

[39] D. Lamberton and B. Lapeyre. *Introduction au calcul stochastique appliqué à la finance*. Ellipses, 1997.

[40] B. Lapeyre, E. Pardoux, and R. Sentis. *Introduction to Monte-Carlo methods for transport and diffusion equations*. Oxford University Press, 2003.

[41] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.

[42] R.C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4:141–183, 1973.

[43] R.C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

[44] B. Øksendal and K. Reikvam. Viscosity solutions of optimal stopping problems. *Stochastics Stochastics Rep.*, 62(3-4):285–301, 1998.

[45] H. Pham. *Optimisation et Contrôle Stochastique Appliqués à la Finance*. Springer Verlag, 2006.

[46] Huyên Pham. Optimal stopping of controlled jump diffusion processes: a viscosity solution approach. *J. Math. Systems Estim. Control*, 8(1):27 pp. (electronic), 1998.

[47] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer, 1997.

[48] L.C.G. Rogers and Z. Shi. The value of an Asian option. *J. Appl. Probability*, 32:1077–1088, 1995.

[49] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[50] Manuel S. Santos and John Rust. Convergence properties of policy iteration. *SIAM J. Control Optim.*, 42(6):2094–2115 (electronic), 2004.

[51] E. Stein and J. Stein. Stock price distributions with stochastic volatility : an analytic approach. *The review of financial studies*, 4(4):727–752, 1991.

[52] G.W.P. Thompson. Fast narrow bounds on the value of Asian options. *Working paper Judge Institute U. of Cambridge*, 1999.

[53] J. Večeř. A new PDE approach for pricing arithmetic average Asian options. *Journal of Computational Finance*, 4(4):105–113, 2001.

[54] T. von Petersdoff and C. Schwab. Numerical solutions of parabolic equations in high dimensions. *Math. Model. Numer. Anal.*, 38:93–128, 2004.

[55] P. Wilmott, J. Dewynne, and S. Howison. *Option pricing: mathematical models and computation.* Oxford financial press, 1993.

[56] R. Zvan, P.A. Forsyth, and K. Vetzal. Robust numerical methods for PDE models of Asian options. *Journal of Computational Finance*, 1(2):39–78, 1998.