IMA Journal of Numerical Analysis (2003) **23**, 241–267

Numerical convergence properties of option pricing PDEs with uncertain volatility

D. M. POOLEY† AND P. A. FORSYTH‡
School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

AND

K. R. VETZAL§
Centre for Advanced Studies in Finance, University of Waterloo, Ontario, Canada N2L 3G1

[Received on 5 November 2001; revised on 5 August 2002]

The pricing equations derived from uncertain volatility models in finance are often cast in the form of nonlinear partial differential equations. Implicit timestepping leads to a set of nonlinear algebraic equations which must be solved at each timestep. To solve these equations, an iterative approach is employed. In this paper, we prove the convergence of a particular iterative scheme for one factor uncertain volatility models. We also demonstrate how non-monotone discretization schemes (such as standard Crank–Nicolson timestepping) can converge to incorrect solutions, or lead to instability. Numerical examples are provided.

Keywords: nonlinear PDE; option pricing; convergence; viscosity solution; uncertain volatility.

1. Introduction

An option is a financial contract entered into by two parties, a buyer and a seller. The buyer of the contract obtains the right to trade an underlying asset, such as a stock, for a specified price, called the strike price, on or before a maturity date. Options which provide the right to buy the underlying asset are known as calls, whereas options conferring the right to sell the underlying asset are referred to as puts. When the option contract is entered into, the option buyer pays a price to the seller. In return for this price, the seller agrees to meet any obligations arising from the contract. For example, the seller of a call option agrees to sell the underlying asset to the buyer of that option for the strike price should the buyer exercise their right to purchase. Option buyers are said to have long positions, while option sellers have short positions.

There are many varieties of options. European options may only be exercised on the maturity date. American options may be exercised any time up to and including the maturity date. Path-dependent options have payoffs which depend on the history of the underlying asset, such as the average price (an Asian option) or the maximum price (a

†Email: dmpooley@elora.math.uwaterloo.ca
‡Email: paforsyt@elora.math.uwaterloo.ca
§Email: kvetzal@watarts.uwaterloo.ca

lookback option) over some period of time. Barrier options have payoffs that depend on whether or not the price of the underlying asset reaches specified levels during the contract.

Regardless of the type of option, the seller of the contract is faced with two decisions: what price to charge for the contract, and how to hedge the resulting risk exposure. It turns out that in many situations it is possible (at least in theory) to find hedging strategies which completely eliminate this risk. By the no-arbitrage principle, the fair price to charge for the contract is then simply the cost of this hedge. The principal source of risk is the price of the underlying asset. The seller of a call option is faced with the possibility of having to sell the underlying asset for a much lower price than its prevailing market price should the price of the underlying asset rise dramatically before the option matures. The probability of large movements in the underlying asset price depends on its volatility.

Several models for volatility have been proposed in the option pricing literature. The simplest model assumes constant volatility. This was the approach taken by Black & Scholes (1973) and Merton (1973) in the work which laid the foundations for the modern analysis of options and is still the industry standard. However, it is generally agreed that constant volatility cannot explain observed market prices for options. More complicated models assume volatility surfaces across underlying asset prices and time (see, e.g., Andersen & Brotherton-Ratcliffe, 1998; Coleman *et al.*, 1999, and references therein). These surfaces are often constructed by the implied volatilities under the Black & Scholes model for a variety of currently traded contracts. A third modelling approach uses stochastic volatility, in which the volatility is assumed to follow some random process (Heston, 1993). A downside of stochastic volatility for numerical pricing methods is an increase in the number of state variables that need to be considered.

Another approach, and the one that will be studied in this paper, is uncertain volatility. The uncertain volatility model was independently developed by Lyons (1995) and Avellaneda *et al.* (1995). In this case, volatility is assumed to lie within a range of values. As such, prices obtained under a no-arbitrage analysis are no longer unique. All that can be computed are the best-case and worst-case prices, for a specified long or short position. By assuming the worst case, an investor can hedge his/her position and obtain a non-negative balance in the hedging portfolio, regardless of the actual volatility movement, provided that the volatility remains within the specified range.

Several studies have already considered uncertain volatility for one factor problems (see, e.g., Lyons, 1995; Avellaneda *et al.*, 1995; Dokuchaev & Savkin, 1998; Lyons & Smith, 1999; Forsyth & Vetzal, 2001). These studies show that pricing in uncertain volatility models involves nonlinear partial differential equations (PDEs). For simple options with convex payoffs, the solution reduces to that of a constant volatility problem with one of the extreme volatility values. Most authors therefore choose to study more exotic options with non-convex payoff functions. Barrier options seem to be the most popular. The nonlinearity of the problem also means that portfolio evaluation is more difficult (Avellaneda & Buff, 1999).

When solving a nonlinear PDE, there is always the question of the uniqueness of the solution. For example, it is well known that nonlinear conservation law hyperbolic PDEs do not have unique solutions once shocks form. In this case, the physically correct solution satisfies the E-condition (LeVeque, 1990). In financial applications, the relevant solution is the viscosity solution (Fleming & Sonar, 1993). Provided that the nonlinear PDE satisfies certain technical conditions (the strong comparison principle (Barles, 1997)), it is

known that a stable, consistent and monotone discretization of financial (non-conservative) PDEs converges to the viscosity solution (Barles, 1997). We provide a brief introduction to viscosity solutions in the Appendix of this paper.

The use of an implicit discretization method results in a set of nonlinear algebraic equations which must be solved at each timestep. We show that the iterative scheme developed in this paper is globally convergent. We also derive conditions which ensure that the discrete scheme is monotone and hence converges to the financially relevant solution. We then show, by means of numerical examples, that non-monotone (but implicit) schemes can lead to incorrect solutions, or to instability.

The organization of this paper is as follows. In Section 2, we review the basics of the uncertain volatility model. In Section 3, we discuss some theoretical convergence issues for one-factor problems. In particular, we show how the nonlinear iteration at each timestep is globally convergent, and we discuss monotonicity properties of the numerical scheme (which are important for convergence to the desired viscosity solution). We then proceed to give some numerical examples in Section 4. Finally, we provide some conclusions in Section 5.

2. Basic background

Following standard arguments, the PDE for the fair price of a contingent claim on one asset in the Black–Scholes model with uncertain volatility is given by

$$U_\tau = \frac{\sigma(\Gamma)^2}{2} S^2 U_{SS} + rSU_S - rU, \tag{2.1}$$

where S represents the underlying asset price, T is the maturity time of the option, $\tau = T - t$ is time in the backwards direction, $\Gamma = U_{SS}$, $\sigma(\Gamma)$ is the uncertain volatility (more details below), and r is the risk-free interest rate. At $S = 0$, we have the boundary condition

$$U_\tau = -rU, \tag{2.2}$$

while at $S \to \infty$, we have a Dirichlet condition

$$U \simeq A(\tau)S + B(\tau), \tag{2.3}$$

where A and B can be determined by financial reasoning. In practice, we use a finite computational domain so that condition (2.3) is applied at a finite value $S_{\max}$.

The volatility is assumed to lie within the range

$$\sigma_{\min} \leqslant \sigma(\Gamma) \leqslant \sigma_{\max}.$$

In the following, it is assumed that $\sigma_{\min} > 0$, $r > 0$. With a range of possible volatility values, (2.1) is nonlinear and does not possess a unique solution. Nevertheless, the best/worst case values are expected to be unique. These values are found by either maximizing or minimizing the diffusion term by selecting σ according to the value of $\Gamma = U_{SS}$. Hence, we have written $\sigma(\Gamma)$ in (2.1) to denote the explicit dependence of

volatility on the value of gamma. Specifically, if we consider the worst case for an investor with a long position in the option, then

$$\sigma(\Gamma)^2 = \begin{cases} \sigma_{\max}^2 & \text{if } \Gamma \leqslant 0 \\ \sigma_{\min}^2 & \text{if } \Gamma > 0. \end{cases} \tag{2.4}$$

On the other hand, the best case for an investor with a long position is determined by

$$\sigma(\Gamma)^2 = \begin{cases} \sigma_{\max}^2 & \text{if } \Gamma > 0 \\ \sigma_{\min}^2 & \text{if } \Gamma \leqslant 0. \end{cases} \tag{2.5}$$

Prices for investors with short positions are given by the negative of the solutions when applying (2.4) and (2.5). Note that the Leland (1985) model of transaction costs can be formulated as a nonlinear PDE which is mathematically identical to (2.1) with nonlinearities of the form (2.4)–(2.5) (see Wilmott, 1998, for details).

To solve the nonlinear PDE (2.1) numerically, we must select an appropriate discretization scheme. If an implicit method is used, then an iterative approach (such as Newton's method) must be used at each timestep to solve the discrete equations. The convergence of a particular iterative scheme will be addressed in the next section.

3. Numerical convergence issues

3.1 *A finite difference discretization*

Assuming a European-style option, (2.1) can be discretized by a standard finite difference method with variable timeweighting to give

$$\begin{aligned} U_i^{n+1} - U_i^n = (1-\theta)\left[(-\alpha_i^{n+1} - \beta_i^{n+1} - r\Delta\tau)U_i^{n+1} + \alpha_i^{n+1}U_{i-1}^{n+1} + \beta_i^{n+1}U_{i+1}^{n+1}\right] \\ + \theta\left[(-\alpha_i^n - \beta_i^n - r\Delta\tau)U_i^n + \alpha_i^n U_{i-1}^n + \beta_i^n U_{i+1}^n\right]. \end{aligned} \tag{3.1}$$

Fully implicit and Crank–Nicolson discretizations correspond to cases of $\theta = 0$ and $\theta = 1/2$ respectively. The form of α_i and β_i depends on the choice of finite-difference stencil. Discretizing the first derivative term of (2.1) with central differences leads to

$$\begin{aligned} \alpha_{i,\text{central}}^n &= \left[\frac{\sigma(\Gamma_i^n)^2 S_i^2}{(S_i - S_{i-1})(S_{i+1} - S_{i-1})} - \frac{rS_i}{S_{i+1} - S_{i-1}}\right]\Delta\tau \\ \beta_{i,\text{central}}^n &= \left[\frac{\sigma(\Gamma_i^n)^2 S_i^2}{(S_{i+1} - S_i)(S_{i+1} - S_{i-1})} + \frac{rS_i}{S_{i+1} - S_{i-1}}\right]\Delta\tau. \end{aligned} \tag{3.2}$$

If $\alpha_{i,\text{central}}$ is negative, oscillations may appear in the solution ($\beta_{i,\text{central}}$ is always positive). The oscillations can be avoided by using forward differences at the problem nodes, leading to

$$\begin{aligned} \alpha_{i,\text{forward}}^n &= \frac{\sigma(\Gamma_i^n)^2 S_i^2}{(S_i - S_{i-1})(S_{i+1} - S_{i-1})}\Delta\tau \\ \beta_{i,\text{forward}}^n &= \left[\frac{\sigma(\Gamma_i^n)^2 S_i^2}{(S_{i+1} - S_i)(S_{i+1} - S_{i-1})} + \frac{rS_i}{S_{i+1} - S_i}\right]\Delta\tau. \end{aligned} \tag{3.3}$$

Algorithmically, we decide between a central or forward discretization at each node for (3.1) as follows:

$$
\begin{aligned}
&\text{If } \left[\frac{\sigma_{\min}^2 S_i^2}{(S_i - S_{i-1})(S_{i+1} - S_{i-1})} - \frac{r S_i}{S_{i+1} - S_{i-1}} \right] \geqslant 0 \text{ then} \\
&\qquad \alpha_i = \alpha_{i,\text{central}} \\
&\qquad \beta_i = \beta_{i,\text{central}} \\
&\text{Else} \\
&\qquad \alpha_i = \alpha_{i,\text{forward}} \\
&\qquad \beta_i = \beta_{i,\text{forward}} \\
&\text{EndIf}
\end{aligned}
\tag{3.4}
$$

Note that the use of $\sigma_{\min}$ in the test condition guarantees that α_i and β_i are always positive, regardless of the choice of $\sigma(\Gamma)$. For the second derivative term, and also for the approximation of $\Gamma_i^n = (U_{SS})_i^n$, we use

$$
(U_{ss})_i^n = \Gamma_i^n = \sum_{j \in \eta_i} \frac{2(U_j^n - U_i^n)}{(S_{i+1} - S_{i-1})|S_j - S_i|}, \tag{3.5}
$$

where $\eta_i = \{i+1, i-1\}$.

Note that a finite volume approach as in Zvan *et al.* (2001) would lead to the same form of discretization as (3.1). Wherever forward differencing is used, a finite volume approach would use upstream weighting.† One advantage of finite volume methods is the potential for higher order flux limiters. For typical values of σ, r and grid spacing, forward differencing is rarely required for single factor options. However, for multi-factor options, a flux limiter can be highly beneficial (Zvan *et al.*, 2001). Consequently, for ease of notation in our single factor analysis, we have used a finite difference discretization. Note again that forward differencing is only used at nodes where $\alpha_{i,\text{central}} < 0$. In practice, since this occurs at only a small number of nodes remote from the region of interest, the limited use of a low-order scheme does not result in poor convergence as the mesh is refined. As we shall see, requiring that all α_i and β_i are non-negative has important theoretical ramifications.

The set of algebraic equations (3.1) is non-smooth due to the form of (2.4)–(2.5). The non-smoothness can be made clear by re-writing the discrete equations at each node as

$$
\begin{aligned}
g_i = &-U_i^{n+1} + U_i^n + (1-\theta)\left[\sum_{j \in \eta_i} \Delta\tau \gamma_{ij} (U_j^{n+1} - U_i^{n+1}) - r\Delta\tau U_i^{n+1} \right] \\
&+ \theta \left[\sum_{j \in \eta_i} \Delta\tau \gamma_{ij} (U_j^n - U_i^n) - r\Delta\tau U_i^n \right] + (1-\theta)\Delta\tau \frac{\sigma(\Gamma_i^{n+1})^2 S_i^2}{2} \Gamma_i^{n+1} \\
&+ \theta \Delta\tau \frac{\sigma(\Gamma_i^n)^2 S_i^2}{2} \Gamma_i^n = 0,
\end{aligned}
\tag{3.6}
$$

†On a uniform grid in one space dimension, the finite volume method and finite difference method give identical discretizations. On non-uniform grids, upstream weighting and forward differencing give slight differences, although the form of the discretizations remains identical.

where

$$\gamma_{ij} = \begin{cases} rS_i/(S_{i+1} - S_{i-1}) & \text{if } j = i+1 \text{ and central differences} \\ rS_i/(S_{i+1} - S_i) & \text{if } j = i+1 \text{ and forward differences} \\ -rS_i/(S_{i+1} - S_{i-1}) & \text{if } j = i-1 \text{ and central differences} \\ 0 & \text{if } j = i-1 \text{ and forward differences.} \end{cases} \tag{3.7}$$

For future reference, note that if we define α_i and β_i as in algorithm (3.4), then

$$\gamma_{ij} + \frac{\sigma_{\min}^2 S_i^2}{(S_{i+1} - S_{i-1})|S_j - S_i|} \geqslant 0. \tag{3.8}$$

Consider now a long investment. The nonlinear component of (3.6) can be written as

$$\sigma(\Gamma)^2\Gamma = \begin{cases} \max(\sigma_{\min}^2\Gamma, 0) + \min(\sigma_{\max}^2\Gamma, 0)\ ; & \text{worst-case long} \\ \max(\sigma_{\max}^2\Gamma, 0) + \min(\sigma_{\min}^2\Gamma, 0)\ ; & \text{best-case long.} \end{cases} \tag{3.9}$$

In order to apply Newton iteration to the non-smooth equations (3.6), we must specify the element of the generalized Jacobian that will be used in the Newton iteration (Qi & Sun, 1993; Pang & Qi, 1993; Sun & Han, 1997). We will define the derivatives for best-case long as

$$\frac{\partial\sigma(\Gamma)^2\Gamma}{\partial\Gamma} = \begin{cases} \sigma_{\max}^2 & \text{if } \Gamma > 0 \\ \sigma_{\min}^2 & \text{if } \Gamma \leqslant 0, \end{cases} \tag{3.10}$$

and for worst-case long as

$$\frac{\partial\sigma(\Gamma)^2\Gamma}{\partial\Gamma} = \begin{cases} \sigma_{\min}^2 & \text{if } \Gamma > 0 \\ \sigma_{\max}^2 & \text{if } \Gamma \leqslant 0. \end{cases} \tag{3.11}$$

For further ease of analysis, we can also write the discrete equations (3.1) in matrix form. Let $U^{n+1} = [U_1^{n+1}, U_2^{n+1}, \ldots, U_{\text{imax}}^{n+1}]'$, $U^n = [U_1^n, U_2^n, \ldots, U_{\text{imax}}^n]'$ and

$$\left[\hat{M}^n U^n\right]_i = -\left[(-\alpha_i^n - \beta_i^n - r\Delta\tau)U_i^n + \alpha_i^n U_{i-1}^n + \beta_i^n U_{i+1}^n\right]. \tag{3.12}$$

The first and last rows of $\hat{M}$ are modified as needed to handle the boundary conditions. In our case, the boundary conditions (2.2) and (2.3) are of Dirichlet type. The boundary condition at $S = 0$ is enforced by setting $\alpha_i = \beta_i = 0$ at $i = 1$. We approximate the infinite computational domain $S \in [0, \infty]$ by the finite domain $S \in [0, S_{\max}]$. Denote the node corresponding to $S_i = S_{\max}$ as $S_i = S_{\text{imax}}$. To avoid algebraic complication in the following, we will assume that the Dirichlet condition at $S = S_{\text{imax}}$ is time independent, so that

$$U(\tau, S_{\max}) = U_{\text{imax}} = \text{ Const.}$$

The boundary condition at $i = \text{imax}$ is enforced by setting $U_{\text{imax}}^0 = U_{\text{imax}}$, and setting the last row of $\hat{M}$ to be identically zero. With a slight abuse of notation, we denote this last

row as $(\hat{M})_{\text{imax}} \equiv 0$. In the following, it will be understood that equations of type (3.12) hold only for $i < \text{imax}$, with $(\hat{M})_{\text{imax}} \equiv 0$.

The discrete equations (3.1) can then be written in a compact matrix form:

$$\left[I + (1-\theta)\hat{M}^{n+1}\right] U^{n+1} = \left[I - \theta \hat{M}^{n}\right] U^{n}. \tag{3.13}$$

By using forward differencing to ensure that α_i^n and β_i^n are positive, matrix $\left[I + (1-\theta)\hat{M}^{n+1}\right]$ is an M-matrix—a diagonally dominant matrix with positive diagonals and non-positive off-diagonals. Note that all of the elements of the inverse of an M-matrix are non-negative.

In the analysis of the following sections, we alternate between the discrete equation representation (3.6) and the matrix representation (3.13) as appropriate. Generally speaking, the matrix form is useful when discussing properties of the algorithm (e.g. convergence of the Newton iteration), while the discrete equation form is suitable for determining properties of the equation (e.g. showing monotonicity).

3.2 *Convergence of the uncertain volatility iteration*

Because of the simple (although non-smooth) form of the Jacobian, we can analyse the Newton iteration in detail. Let $(U^{n+1})^k$ be the kth estimate for U^{n+1}. The Newton iteration at each timestep is then determined by the following scheme:

$$\begin{aligned}
&\text{Let } (U^{n+1})^0 = U^n \\
&\text{For } k = 0, 1, 2, \ldots \text{ until convergence} \\
&\qquad \text{Solve } \left[I + (1-\theta)\hat{M}((U^{n+1})^k)\right](U^{n+1})^{k+1} = \left[I - \theta \hat{M}(U^n)\right] U^n \\
&\qquad \text{If } \max_i \frac{|(U_i^{n+1})^{k+1} - (U_i^{n+1})^k|}{\max(1, |(U_i^{n+1})^{k+1}|)} < \text{ tolerance then quit} \\
&\text{EndFor}
\end{aligned} \tag{3.14}$$

For notational convenience, define

$$\begin{aligned}
\hat{M}^k &\equiv \hat{M}((U^{n+1})^k) \\
\bar{U}^k &\equiv (U^{n+1})^k,
\end{aligned}$$

so that the basic algorithm (3.14) can be written as

$$\left[I + (1-\theta)\hat{M}^k\right] \bar{U}^{k+1} = \left[I - \theta \hat{M}^n\right] U^n. \tag{3.15}$$

Our main result in this section can be summarized in the following theorem.

THEOREM 3.1 (Convergence of the nonlinear iterations) If the matrix $\left[I + (1-\theta)\hat{M}^k\right]$ in (3.15) is an M-matrix, then the nonlinear iteration (3.14) converges to the unique solution to (3.13), given any initial iterate $\bar{U}^0$. Moreover, the iterates converge monotonically. For $\bar{U}^k$ sufficiently close to the solution, convergence is quadratic.

Proof. We will first prove that this algorithm is globally convergent by showing that the iterates form non-increasing (non-decreasing) sequences which are bounded from below (above). Writing (3.15) for iteration k gives

$$\left[I + (1-\theta)\hat{M}^{k-1}\right]\bar{U}^k = \left[I - \theta\hat{M}^n\right]U^n,$$

which can also be expressed as

$$\left[I + (1-\theta)\hat{M}^{k}\right]\bar{U}^k + (1-\theta)\left[\hat{M}^{k-1} - \hat{M}^k\right]\bar{U}^k = \left[I - \theta\hat{M}^n\right]U^n. \tag{3.16}$$

Subtracting (3.16) from (3.15) gives

$$\left[I + (1-\theta)\hat{M}^{k}\right](\bar{U}^{k+1} - \bar{U}^k) = (1-\theta)\left[\hat{M}^{k-1} - \hat{M}^k\right]\bar{U}^k. \tag{3.17}$$

We wish to show that the iterates form a bounded non-increasing or non-decreasing sequence. Expanding the right-hand side of (3.17) using definition (3.12) for node i leads to

$$\begin{aligned}&(1-\theta)\left(\left[\hat{M}^{k-1} - \hat{M}^k\right]\bar{U}^k\right)_i \\ &\quad = (1-\theta)\left[(-\alpha_i^k + \alpha_i^{k-1} - \beta_i^k + \beta_i^{k-1})U_i^k + (\alpha_i^k - \alpha_i^{k-1})U_{i-1}^k + (\beta_i^k - \beta_i^{k-1})U_{i+1}^k\right].\end{aligned} \tag{3.18}$$

This expression can be simplified using definitions (3.2)–(3.5) to obtain

$$(1-\theta)\left(\left[\hat{M}^{k-1} - \hat{M}^k\right]\bar{U}^k\right)_i = (1-\theta)\Delta\tau\frac{S_i^2\left[(\sigma_i^k)^2 - (\sigma_i^{k-1})^2\right]}{2}\Gamma_i^k\,; \quad k \geqslant 1. \tag{3.19}$$

For clarity, we will examine the sign of the right-hand side of (3.19) in different cases. In particular, we must determine the sign of $\left[(\sigma_i^k)^2 - (\sigma_i^{k-1})^2\right]\Gamma_i^k$, as all other factors are clearly positive. Consider first a worst-case long position in the option, so that definition (2.4) applies for the values of $\sigma(\Gamma)$. Then

Case 1. $\Gamma_i^k \leqslant 0$ implies that $(\sigma_i^k)^2 = \sigma_{\max}^2$. Then we have $\left[(\sigma_{\max})^2 - (\sigma_i^{k-1})^2\right]\Gamma_i^k \leqslant 0$.

Case 2. $\Gamma_i^k > 0$ implies that $(\sigma_i^k)^2 = \sigma_{\min}^2$. Then we have $\left[(\sigma_{\min})^2 - (\sigma_i^{k-1})^2\right]\Gamma_i^k \leqslant 0$.

Thus, in either case, and for any i, expression (3.19) is less than or equal to zero. By expression (3.17), and the fact that $\left[I + (1-\theta)\hat{M}^k\right]$ is an M-matrix, it follows that

$$\bar{U}^{k+1} - \bar{U}^k \leqslant 0\,; \quad k \geqslant 1, \tag{3.20}$$

and hence the iterates form a non-increasing sequence.

Consider now a best-case long position so that definition (2.5) applies for the values of $\sigma(\Gamma)$. Then

Case 3. $\Gamma_i^k \leqslant 0$ implies that $(\sigma_i^k)^2 = \sigma_{\min}^2$. Then we have $\left[(\sigma_{\min})^2 - (\sigma_i^{k-1})^2\right]\Gamma_i^k \geqslant 0$.

Case 4. $\Gamma_i^k > 0$ implies that $(\sigma_i^k)^2 = \sigma_{\max}^2$. Then we have $\left[(\sigma_{\max})^2 - (\sigma_i^{k-1})^2\right]\Gamma_i^k \geqslant 0$.

In these cases, for any i, expression (3.19) is greater than or equal to zero. By expression (3.17), and the fact that $\left[I + (1-\theta)\hat{M}^k\right]$ is an M-matrix, it follows that

$$\bar{U}^{k+1} - \bar{U}^k \geqslant 0\,; \quad k \geqslant 1, \tag{3.21}$$

and hence the iterates form a non-decreasing sequence, for all iterations *after* the first iteration ($k \geqslant 1$) at each timestep. A similar analysis can be done for short positions.

Now that we have shown the iterates to be non-increasing or non-decreasing, we need to show that they are bounded. To do this, let $b = [I - \theta\hat{M}^n]U^n$. For a fixed mesh, $\|b\|_\infty$ is bounded. Now, by (3.12) and (3.15), we have

$$\begin{aligned}\left[I + (1-\theta)\hat{M}^k\right]\bar{U}^{k+1} &= b\\ \left[1 + (1-\theta)(\alpha_i^k + \beta_i^k + r\Delta\tau)\right]\bar{U}_i^{k+1} &= (1-\theta)\alpha_i^k\bar{U}_{i-1}^{k+1} + (1-\theta)\beta_i^k\bar{U}_{i+1}^{k+1} + b_i.\end{aligned} \tag{3.22}$$

Now let $U_{\max} = \max_i(\bar{U}_i^{k+1})$, $U_{\min} = \min_i(\bar{U}_i^{k+1})$, $b_{\max} = \max_i(b_i)$ and $b_{\min} = \min_i(b_i)$. Note that $(\hat{M}^k)_{\text{imax}} \equiv 0$ and that $b_{\text{imax}} = U_{\text{imax}}$. Then since all coefficients of the U terms are positive, we have

$$\left[1 + (1-\theta)(\alpha_i^k + \beta_i^k + r\Delta\tau)\right]\bar{U}_i^{k+1} \leqslant (1-\theta)\alpha_i^k U_{\max} + (1-\theta)\beta_i^k U_{\max} + b_{\max}; \quad i < \text{imax}$$

$$U_{\max} \leqslant \max\left(\frac{b_{\max}}{1 + (1-\theta)r\Delta\tau}, b_{\max}\right). \tag{3.23}$$

Similarly,

$$U_{\min} \geqslant \min\left(\frac{b_{\min}}{1 + (1-\theta)r\Delta\tau}, b_{\min}\right). \tag{3.24}$$

Thus $\|U_i^{k+1}\|_\infty \leqslant \|b\|_\infty$, independent of k. Consequently, since the iterates are either non-increasing or non-decreasing, and $\|U_i^{k+1}\|_\infty$ is bounded independent of k, the iteration (3.15) converges. (Note that if $\bar{U}^{k+1} = \bar{U}^k$, then the residual of the nonlinear equations is identically zero, and $\bar{U}^{k+1}$ is an exact solution of (3.13)).

As for uniqueness, suppose we have two solutions to (3.15), U_1 and U_2, such that

$$\left[I + (1-\theta)\hat{M}_1\right]U_1 = \left[I - \theta\hat{M}^n\right]U^n \tag{3.25}$$

$$\left[I + (1-\theta)\hat{M}_2\right]U_2 = \left[I - \theta\hat{M}^n\right]U^n, \tag{3.26}$$

where $\hat{M}_1 = \hat{M}(U_1)$, $\hat{M}_2 = \hat{M}(U_2)$. We can rewrite (3.25) as

$$\left[I + (1-\theta)\hat{M}_2\right] U_1 + (1-\theta)\left[\hat{M}_1 - \hat{M}_2\right] U_1 = \left[I - \theta \hat{M}^n\right] U^n. \tag{3.27}$$

Subtracting (3.26) from (3.27) gives

$$\left[I + (1-\theta)\hat{M}_2\right](U_1 - U_2) = (1-\theta)\left[\hat{M}_2 - \hat{M}_1\right] U_1. \tag{3.28}$$

In component form (note the analogy to (3.19)), this becomes

$$\left(\left[I + (1-\theta)\hat{M}_2\right](U_1 - U_2)\right)_i = (1-\theta)\Delta\tau \frac{S_i^2((\sigma(\Gamma_1)^2)_i - (\sigma(\Gamma_2)^2)_i)}{2}(\Gamma_1)_i. \tag{3.29}$$

Consider a best-case long position. In this situation, the right-hand side of (3.29) is always non-negative, so that we have $U_1 \geqslant U_2$. Interchanging subscripts gives $U_2 \geqslant U_1$, and hence $U_1 = U_2$. Similar arguments can be used for best-case short positions, and worst-case positions.

Note that since Γ_i^n as defined by (3.5) is a simple linear function of U_i^n and U_j^n, the non-smooth equations (3.6) are strongly semi-smooth (Qi & Zhou, 2000). This means that convergence will be quadratic in a sufficiently small neighbourhood of the solution (Qi & Sun, 1993). □

Thus, we have shown that iteration (3.14) will converge to a unique solution at each timestep. However, this says nothing about convergence to the viscosity solution of the PDE as the timestep and mesh size are reduced. This topic will be discussed in the next section.

3.3 *Convergence to the viscosity solution*

In the previous section, global convergence of the Newton iteration for the nonlinear algebraic equations at each timestep was proven. However, since the PDE is nonlinear, questions remain about convergence to the financially correct solution as ΔS, $\Delta\tau \to 0$. In a financial context, we would like to ensure convergence to the *viscosity solution* (Crandall *et al.*, 1992). A brief discussion of viscosity solutions is provided in the Appendix. From the work of Barles (1997), we know that a stable, consistent, and monotone discretization will converge to the viscosity solution. To this end, it is helpful to review the concept of a monotone discretization.

The set of discrete equations (3.6) can be written as

$$g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) = 0 \quad \forall i\ ; \quad j \in \eta_i. \tag{3.30}$$

With a monotone discretization, a positive perturbation to any of $\{U_j^{n+1}, U_i^n, U_j^n\}$ produces

a positive perturbation of U_i^{n+1}. If g_i is differentiable, then this is equivalent to stating that

$$\begin{aligned}
\frac{\partial U_i^{n+1}}{\partial U_i^n} &= -\frac{\partial g_i/\partial U_i^n}{\partial g_i/\partial U_i^{n+1}} \geqslant 0 \\
\frac{\partial U_i^{n+1}}{\partial U_j^{n+1}} &= -\frac{\partial g_i/\partial U_j^{n+1}}{\partial g_i/\partial U_i^{n+1}} \geqslant 0 \\
\frac{\partial U_i^{n+1}}{\partial U_j^n} &= -\frac{\partial g_i/\partial U_j^n}{\partial g_i/\partial U_i^{n+1}} \geqslant 0.
\end{aligned} \tag{3.31}$$

In the case of nondifferentiable g_i, which is the case for the uncertain volatility discretization, we will use the following definition of monotonicity.

DEFINITION 1 (Monotone discretizations) A discretization of the form (3.30) is monotone if either

$$\begin{aligned}
g_i(U_i^{n+1}, U_j^{n+1} + \rho_j^{n+1}, U_i^n + \rho_i^n, U_j^n + \rho_j^n) \geqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \quad \forall i\ ; \quad j \in \eta_i \\
\forall \rho_i^n \geqslant 0,\ \forall \rho_j^{n+1} \geqslant 0,\ \forall \rho_j^n \geqslant 0 \\
g_i(U_i^{n+1} + \rho_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \leqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \quad \forall i\ ; \quad j \in \eta_i \\
\forall \rho_i^{n+1} \geqslant 0
\end{aligned} \tag{3.32}$$

or

$$\begin{aligned}
g_i(U_i^{n+1}, U_j^{n+1} + \rho_j^{n+1}, U_i^n + \rho_i^n, U_j^n + \rho_j^n) \leqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \quad \forall i\ ; \quad j \in \eta_i \\
\forall \rho_i^n \geqslant 0,\ \forall \rho_j^{n+1} \geqslant 0,\ \forall \rho_j^n \geqslant 0 \\
g_i(U_i^{n+1} + \rho_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \geqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \quad \forall i\ ; \quad j \in \eta_i \\
\forall \rho_i^{n+1} \geqslant 0
\end{aligned} \tag{3.33}$$

This somewhat longwinded definition, with either condition (3.32) or (3.33), avoids having to define $g_i(\dots)$ with a standard sign convention.

The difficulty in verifying these relations for the discrete equations (3.6) comes from the nonlinear term. However, note that

$$\begin{aligned}
\Gamma_i^n(U_i^n, U_j^n + \rho_j^n) &= \Gamma_i^n(U_i^n, U_j^n) + \rho_j^n \frac{2}{(S_{i+1} - S_{i-1})|S_j - S_i|}\ ; \quad \forall \rho_j^n \geqslant 0 \\
\Gamma_i^n(U_i^n + \rho_i^n, U_j^n) &= \Gamma_i^n(U_i^n, U_j^n) - \rho_i^n \sum_{j \in \eta_i} \frac{2}{(S_{i+1} - S_{i-1})|S_j - S_i|}\ ; \quad \forall \rho_i^n \geqslant 0.
\end{aligned} \tag{3.34}$$

Further, regardless of whether $\sigma(\Gamma)$ is defined for best or worst case, it follows from (2.4)–(2.5) that

$$\begin{aligned}\sigma(\Gamma)^2\Gamma + \sigma_{\max}^2\Delta\Gamma \geqslant \sigma(\Gamma+\Delta\Gamma)^2(\Gamma+\Delta\Gamma) \geqslant \sigma(\Gamma)^2\Gamma + \sigma_{\min}^2\Delta\Gamma\ ; \quad \forall\Delta\Gamma \geqslant 0\\ \sigma(\Gamma)^2\Gamma - \sigma_{\max}^2\Delta\Gamma \leqslant \sigma(\Gamma-\Delta\Gamma)^2(\Gamma-\Delta\Gamma) \leqslant \sigma(\Gamma)^2\Gamma - \sigma_{\min}^2\Delta\Gamma\ ; \quad \forall\Delta\Gamma \geqslant 0.\end{aligned} \tag{3.35}$$

We are now in a position to test (3.6) for monotonicity, and hence determine the conditions under which convergence to the viscosity solution is guaranteed. This will be done separately for the fully implicit and Crank–Nicolson cases.

3.3.1 *Fully implicit discretization.* For a fully implicit discretization, we set $\theta = 0$ in (3.6) to obtain (for $i < \text{imax}$)

$$\begin{aligned} g_i = -U_i^{n+1} + U_i^n + \left[\sum_{j\in\eta_i} \Delta\tau\gamma_{ij}(U_j^{n+1} - U_i^{n+1}) - r\,\Delta\tau U_i^{n+1}\right]\\ +\Delta\tau\frac{\sigma(\Gamma_i^{n+1})^2 S_i^2}{2}\Gamma_i^{n+1} = 0,\end{aligned} \tag{3.36}$$

and for $i = \text{imax}$

$$g_{\text{imax}} = -U_{\text{imax}}^{n+1} + U_{\text{imax}}^n = 0.$$

In order to show convergence to the viscosity solution of this discretization, we will use the following lemmas.

LEMMA 3.1 (Monotonicity of the fully implicit discretization) The fully implicit discretization (3.36) is monotonic, independent of any choice of $\Delta\tau$ and grid spacing.

Proof. Note that all U_j^n terms have disappeared in (3.36). Consider perturbing U_j^{n+1} by an amount $\epsilon > 0$. In this case, using relations (3.34), (3.35), and (3.36) gives

$$\begin{aligned} g_i(U_i^{n+1}, U_j^{n+1}+\epsilon, U_i^n) &\geqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n) + \Delta\tau\epsilon\left(\frac{\sigma_{\min}^2 S_i^2}{(S_{i+1}-S_{i-1})|S_j - S_i|} + \gamma_{ij}\right)\\ &\geqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n),\end{aligned} \tag{3.37}$$

where the last line follows from relation (3.8). Continuing in the same manner, we perturb U_i^{n+1} by $\epsilon > 0$ to get

$$\begin{aligned} g_i(U_i^{n+1}+\epsilon, U_j^{n+1}, U_i^n) &\leqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n) - \epsilon\\ &\quad - \Delta\tau\epsilon\sum_{j\in\eta_i}\left(\frac{\sigma_{\min}^2 S_i^2}{(S_{i+1}-S_{i-1})|S_j - S_i|} + \gamma_{ij}\right) - \Delta\tau\epsilon r\\ &\leqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n).\end{aligned} \tag{3.38}$$

Again, the factor in the summation is guaranteed to be positive by relation (3.8). It is obvious from (3.36) that

$$g_i(U_i^{n+1}, U_j^{n+1}, U_i^n + \epsilon) \geqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n). \tag{3.39}$$

The monotonicity of (3.36) now follows directly from definition (3.32). □

LEMMA 3.2 (Stability of the fully implicit discretization) The fully implicit discretization (3.36) is unconditionally stable, in the sense that

$$\|U^{n+1}\|_\infty \leqslant \|U^n\|_\infty$$

independent of the timestep and mesh spacing.

Proof. Define

$$\begin{aligned} U_{\max}^n &= \max_i U_i^n \\ U_{\min}^n &= \min_i U_i^n. \end{aligned} \tag{3.40}$$

From Lemma 3.1 we have that a fully implicit discretization is unconditionally monotone. It follows that

$$\begin{aligned} U_{\max}^{n+1} &\leqslant \max_i (U_{\max}^*)_i \\ U_{\min}^{n+1} &\geqslant \min_i (U_{\min}^*)_i, \end{aligned} \tag{3.41}$$

where $(U_{\max}^*)_i$, $(U_{\min}^*)_i$ are given from the solutions to

$$\begin{aligned} g_i((U_{\max}^*)_i, (U_{\max}^*)_i, U_{\max}^n) &= 0 \\ g_i((U_{\min}^*)_i, (U_{\min}^*)_i, U_{\min}^n) &= 0. \end{aligned} \tag{3.42}$$

Hence, from (3.36) and (3.42), we see that

$$\max(U_{\max}^n, \frac{U_{\max}^n}{1 + r\Delta\tau}) \geqslant U_i^{n+1} \geqslant \min(U_{\min}^n, \frac{U_{\min}^n}{1 + r\Delta\tau}); \quad \forall i. \tag{3.43}$$

Consequently, we have that $\|U^{n+1}\|_\infty \leqslant \|U^n\|_\infty$. □

Let $\Delta S = \max_i (S_{i+1} - S_i)$. Our main result concerning convergence of the fully implicit discretization is the following.

THEOREM 3.2 (Convergence of the fully implicit discretization) The fully implicit discretization (3.36) converges unconditionally to the viscosity solution of the nonlinear PDE (2.1), as $\Delta\tau, \Delta S \to 0$.

Proof. In Barles (1997) it is shown that a consistent, stable, monotone discretization converges to the viscosity solution. Since (3.36) is a consistent discretization, Theorem 3.2 follows directly from the results of Barles (1997), and Lemmas 3.1–3.2. □

3.3.2 *Crank–Nicolson discretization.* For a Crank–Nicolson discretization, we set $\theta = 1/2$ in (3.6) to obtain (for $i < \text{imax}$)

$$\begin{aligned} g_i = &-U_i^{n+1} + U_i^n + \frac{1}{2}\left[\sum_{j\in i+1,i-1} \Delta\tau\gamma_{ij}(U_j^{n+1} - U_i^{n+1}) - r\Delta\tau U_i^{n+1}\right] \\ &+ \frac{1}{2}\left[\sum_{j\in\eta_i} \Delta\tau\gamma_{ij}(U_j^n - U_i^n) - r\Delta\tau U_i^n\right] + \frac{1}{2}\Delta\tau\frac{\sigma(\Gamma_i^{n+1})^2 S_i^2}{2}\Gamma_i^{n+1} \\ &+ \frac{1}{2}\Delta\tau\frac{\sigma(\Gamma_i^n)^2 S_i^2}{2}\Gamma_i^n = 0, \end{aligned} \tag{3.44}$$

and for $i = \text{imax}$

$$g_i = -U_i^{n+1} + U_i^n = 0.$$

As with the fully implicit discretization, we will first determine the conditions for monotonicity.

LEMMA 3.3 (Monotonicity of the Crank–Nicolson discretization) The Crank–Nicolson discretization (3.44) is monotone if the timestep is selected such that

$$\Delta\tau < 2\min_i\left(r + \sum_{j\in\eta_i}\left[\frac{\sigma_{\max}^2 S_i^2}{(S_{i+1} - S_{i-1})|S_j - S_i|} + \gamma_{ij}\right]\right)^{-1}. \tag{3.45}$$

Proof. Following the fully implicit analysis above, we immediately obtain in the Crank–Nicolson case (for $\epsilon > 0$)

$$\begin{aligned} g_i(U_i^{n+1}, U_j^{n+1} + \epsilon, U_i^n, U_j^n) &\geqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \\ g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n + \epsilon) &\geqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) \\ g_i(U_i^{n+1} + \epsilon, U_j^{n+1}, U_i^n, U_j^n) &\leqslant g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n). \end{aligned}$$

Using relations (3.34), (3.35), and (3.44) gives ($\epsilon > 0$)

$$\begin{aligned} g_i(U_i^{n+1}, U_j^{n+1}, U_i^n + \epsilon, U_j^n) \geqslant\ & g_i(U_i^{n+1}, U_j^{n+1}, U_i^n, U_j^n) + \epsilon\left(1 - \frac{r\Delta\tau}{2}\right) \\ & - \frac{\Delta\tau\epsilon}{2}\sum_{j\in\eta_i}\left(\frac{\sigma_{\max}^2 S_i^2}{(S_{i+1} - S_{i-1})|S_j - S_i|} + \gamma_{ij}\right). \end{aligned} \tag{3.46}$$

Since $\sigma_{\max} \geqslant \sigma_{\min}$, it follows from (3.8) that

$$\gamma_{ij} + \frac{\sigma_{\max}^2 S_i^2}{(S_{i+1} - S_{i-1})|S_j - S_i|} \geqslant 0. \tag{3.47}$$

Consequently, for the perturbation ($\epsilon > 0$) to produce a positive change, we require

$$0 < \epsilon \left[1 - \frac{r\Delta\tau}{2} - \frac{\Delta\tau}{2} \sum_{j \in \eta_i} \left(\frac{\sigma_{\max}^2 S_i^2}{(S_{i+1} - S_{i-1})|S_j - S_i|} + \gamma_{ij} \right) \right] \forall i. \tag{3.48}$$

To ensure that the timestep condition is satisfied for all i, the worst case is

$$\Delta\tau < 2 \min_i \left(r + \sum_{j \in \eta_i} \left[\frac{\sigma_{\max}^2 S_i^2}{(S_{i+1} - S_{i-1})|S_j - S_i|} + \gamma_{ij} \right] \right)^{-1}, \tag{3.49}$$

which completes the proof. □

Note that condition (3.45) implies that a Crank–Nicolson scheme will be monotone only if the timestep size is less than twice the maximum stable explicit timestep size.

LEMMA 3.4 (Stability of the Crank–Nicolson Discretization) If condition (3.45) is satisfied, then the Crank–Nicolson discretization is stable, in the sense that

$$\|U^{n+1}\|_\infty \leqslant \|U^n\|_\infty \cdot$$

Proof. If condition (3.45) is satisfied, then by Lemma 3.3, the Crank–Nicolson discretization is monotone, and stability of the discrete equations follows by bounding the U_i values as was done in the fully implicit case (see Lemma 3.2). □

Combining these results allows us to state the following theorem.

THEOREM 3.3 (Convergence of the Crank–Nicolson discretization) If condition (3.45) is satisfied, then the Crank–Nicolson discretization (3.44) converges to the viscosity solution of the nonlinear PDE (2.1) as $\Delta\tau, \Delta S \to 0$.

Proof. Again, this follows directly from Lemmas 3.3 and 3.4 and the results of Barles (1997) since the condition of the theorem ensures a monotone, stable, and consistent discretization. □

If condition (3.45) is not satisfied, then for $\theta = 1/2$ (Crank–Nicolson) in (3.13), we have

$$\left[I + \frac{\hat{M}^{n+1}}{2} \right] U^{n+1} = \left[I - \frac{\hat{M}^n}{2} \right] U^n, \tag{3.50}$$

or

$$\begin{aligned} U^{n+1} &= \left[I + \frac{\hat{M}^{n+1}}{2} \right]^{-1} \left[I - \frac{\hat{M}^n}{2} \right] U^n \\ &= \left[I + \frac{\hat{M}^{n+1}}{2} \right]^{-1} \left[I - \frac{\hat{M}^n}{2} \right] \left[I + \frac{\hat{M}^n}{2} \right]^{-1} \left[I - \frac{\hat{M}^{n-1}}{2} \right] \left[I + \frac{\hat{M}^{n-1}}{2} \right]^{-1} \\ &\quad \times \ldots \left[I - \frac{\hat{M}^0}{2} \right] U^0. \end{aligned} \tag{3.51}$$

Assuming that $r > 0$, all rows of $\hat{M}^n$ are diagonally dominant, with positive diagonal entries, except for the last row which is identically zero. It follows that all eigenvalues of $\hat{M}^n$ have non-negative real parts, and hence the eigenvalues of each term

$$\left[I - \frac{\hat{M}^k}{2}\right]\left[I + \frac{\hat{M}^k}{2}\right]^{-1}$$

are less than or equal to one in magnitude, with only one eigenvalue having magnitude one. However, stability does not necessarily follow, since the $\hat{M}^k$ are not normal matrices. In fact, as we shall see, Crank–Nicolson timestepping appears to be unstable with discontinuous payoffs when condition (3.45) is violated.

4. Numerical examples

4.1 *Butterfly spread*

To illustrate the results of Section 3, we begin by examining the uncertain volatility model on a 'butterfly spread'. This is a combination of options with three different strike prices. It can be formed using either call options or put options. Our test problem uses call options, so the payoff can be written as

$$U(S, \tau = 0) = \max(S - K_1, 0) - 2\max(S - (K_1 + K_2)/2, 0) + \max(S - K_2, 0). \tag{4.1}$$

This corresponds to a long position in two calls at strikes K_1, K_2, and a short position in two calls at strike $(K_1 + K_2)/2$. Recall that $\tau = T - t$, so that the payoff is the value of the option at expiry $t = T$, or the initial condition of the PDE at $\tau = 0$. Figure 1 provides a diagram of a sample payoff function. A complete specification of our test problem, including the PDE parameters, is given in Table 1. Note that unless the problem has a non-convex payoff/solution, the sign of gamma (U_{SS}) will not change during the solution process. In these cases the nonlinearity disappears, and σ will always take on one of the extreme volatility values.

Solutions were computed on a sequence of uniformly refined grids, starting with 61 non-uniform points. At each grid refinement, the timestep was halved. The convergence tolerance for nonlinear iteration (3.14) was 10^{-6}. Convergence results using fully implicit and Crank–Nicolson timestepping are given in Table 2. The timestep was selected so that condition (3.45) was violated. We can see that a fully implicit method converges at a linear rate, as we would expect. From Theorem 3.2, we know that this solution is the viscosity solution. However, the Crank–Nicolson method is either converging to a non-viscosity solution, or has a slowly growing instability. Timestepping and nonlinear iteration data are given in Table 3. Note that in the fully implicit case, the average number of nonlinear iterations per step is close to two. For a linear problem, the number of nonlinear iterations at each step would be exactly two (of course, iteration would be unnecessary in this case).

To understand the difficulty with Crank–Nicolson timestepping, consider plots of the solution values, deltas (U_S) and gammas (U_{SS}), as shown in Fig. 2 for both fully implicit and Crank–Nicolson timestepping.† We see that all implicit plots are smooth, as would be

†Note that risk hedging strategies typically involve delta and gamma, so it is important to accurately compute not only the option value but also its first and second derivatives with respect to the price of the underlying asset.

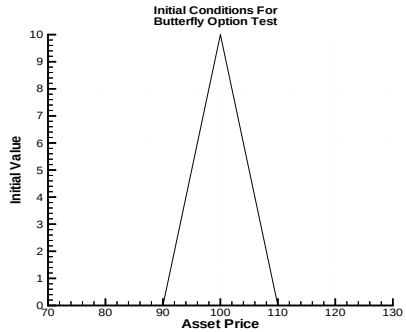


FIG. 1. Sample payoff function for a butterfly spread. Parameters for (4.1) are taken from Table 1.

TABLE 1 *Model parameters for the butterfly spread test problem*

Type:	Butterfly spread (using call options)
Scenario:	Worst case long
Time to expiry:	0.25 years
r:	0.1
K_1:	90
K_2:	110
$\sigma_{\max}$:	0.25
$\sigma_{\min}$:	0.15

TABLE 2 *Convergence results for an at-the-money* ($S = 100$) *butterfly spread with uncertain volatility. Parameters are provided in Table 1. The timestep is halved at each grid refinement. 'Difference' is the absolute value of the change in the solution as the grid is refined. 'Ratio' is the ratio of successive differences. Timestepping data are given in Table 3*

	Fully implicit			Crank–Nicolson		
Nodes	Value	Difference	Ratio	Value	Difference	Ratio
61	2.3501			1.7246		
121	2.3250	0.0251		1.5713	0.1533	
241	2.3116	0.0134	1.87	1.4622	0.1091	1.41
481	2.3047	0.0069	1.95	1.3806	0.0816	1.34
961	2.3012	0.0035	1.97	1.3264	0.0542	1.51

expected. However, a small 'kink' in the Crank–Nicolson solution at $S = 100$ leads to a discontinuity in the solution delta (U_S), and to major oscillations in the solution gamma (U_{SS}) values. Since the uncertain volatility model has a crucial dependence on the sign of gamma, we expect problems.

To further isolate the source of Crank–Nicolson timestepping difficulties, consider the solution after one timestep. In Fig. 3(a), we see that implicit timestepping leads to

TABLE 3 *Timestepping information for an at-the-money* ($S = 100$) *butterfly spread with uncertain volatility. Parameters are provided in Table 1. 'No of iterations' is the total number of nonlinear iterations used during the solution process. 'Average no of iterations' is the number of iterations divided by the number of timesteps. The number of nodes is doubled each time the timestep is halved. The convergence tolerance was* 10^{-6} (3.14). *Convergence data are given in Table 2*

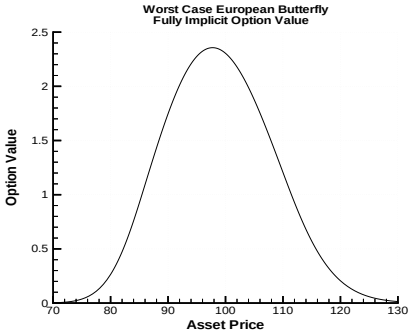
		Fully implicit		Crank–Nicolson	
No of timesteps	$\Delta\tau$	No of iterations	Average no of iterations	No of iterations	Average no of iterations
25	0.01	58	2.32	87	3.48
50	0.005	116	2.32	204	4.08
100	0.002 5	236	2.36	432	4.32
200	0.001 25	461	2.31	886	4.43
400	0.000 625	868	2.17	1858	4.65

a smooth curve. On the other hand, Fig. 3(b) reveals that Crank–Nicolson timestepping has introduced a cusp at the strike price of 100 (other problems at $S = 90$ and $S = 110$ are not shown). For linear problems, such oscillations would eventually be damped out, since Crank–Nicolson is a stable method.

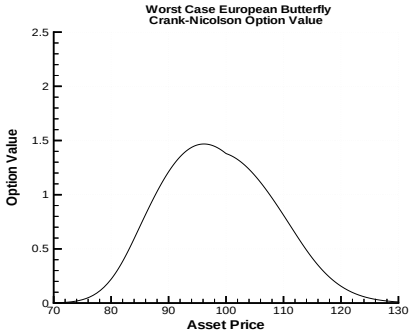
However, for the present problem, the oscillations have caused the computed values of gamma at all nodes where the initial payoff has a discontinuity in delta to have the wrong sign. This is shown in Fig. 4. Consequently, different values of σ will be used at the second timestep at the nodes where oscillations occurred. In this case, since we have the convergence results from a monotone scheme, which is guaranteed to converge to the viscosity solution, we can see that the Crank–Nicolson solution is incorrect.

The oscillations also have an effect on how strongly nonlinear the problem becomes. In Table 3, we see that using implicit timestepping leads to just over two nonlinear Newton iterations per timestep, indicating fairly mild nonlinearity. On the other hand, the average number of nonlinear iterations per timestep for Crank–Nicolson timestepping starts at 3.48 for the the coarsest grid, and increases as the grids are refined. Clearly, the non-monotone discretization amplifies the nonlinear properties of the problem.

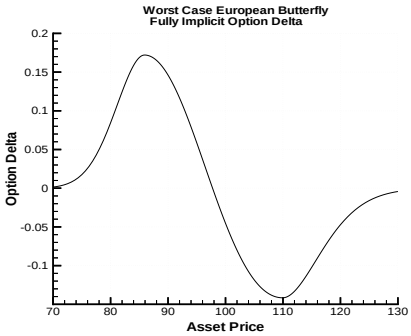
Unfortunately, we would like to use Crank–Nicolson timestepping for the potential of second-order convergence. Since the difficulty appears to be oscillations at the first timestep, it makes sense to start with fully implicit timestepping, and then switch to Crank–Nicolson timestepping. Such a method was discussed in Rannacher (1984), and will be called Rannacher timestepping below. Since only a finite number of fully implicit steps are taken, the overall convergence rate can be shown to be quadratic for linear problems (Rannacher, 1984). Results using this method with two and four initial implicit steps are given in Table 4. Both approaches give (nearly) quadratic convergence, although taking four steps appears to converge at a slightly higher rate. More importantly, even though Rannacher timestepping is not unconditionally monotone (strictly speaking), both methods appear to converge to the correct solution, with no evidence of instability. Assuming a linear rate of convergence, the extrapolated solution using fully implicit timestepping (Table 2) is 2.2977, in excellent agreement with the results in Table 4. Further, as shown



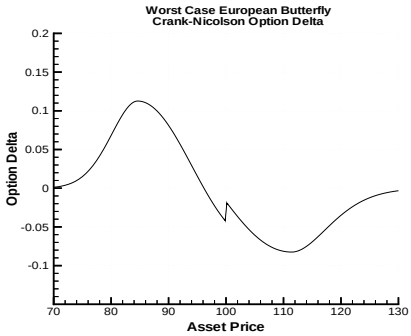
(a) Value, fully implicit timestepping.



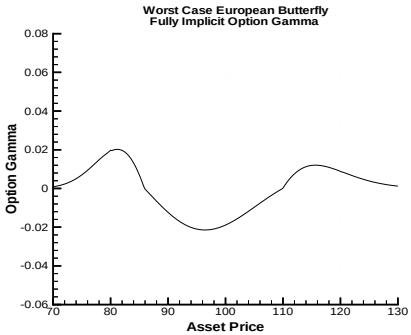
(b) Value, Crank–Nicolson timestepping.



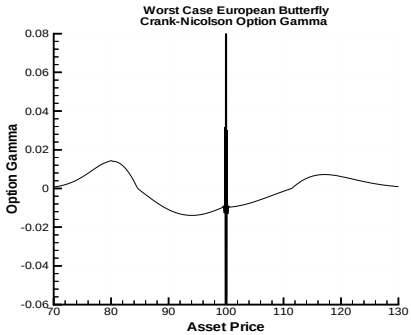
(c) Delta, fully implicit timestepping.



(d) Delta, Crank–Nicolson timestepping.

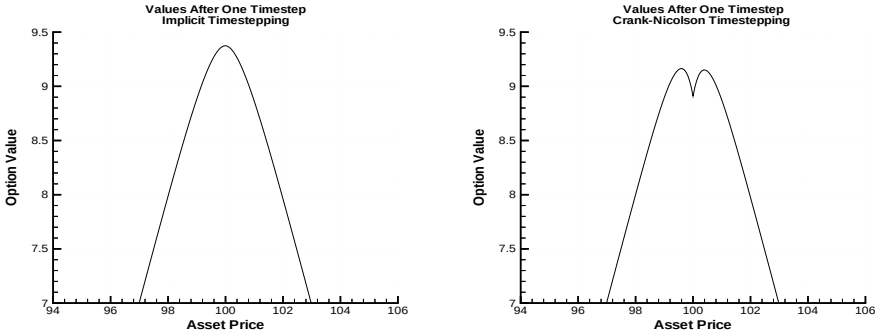


(e) Gamma, fully implicit timestepping.



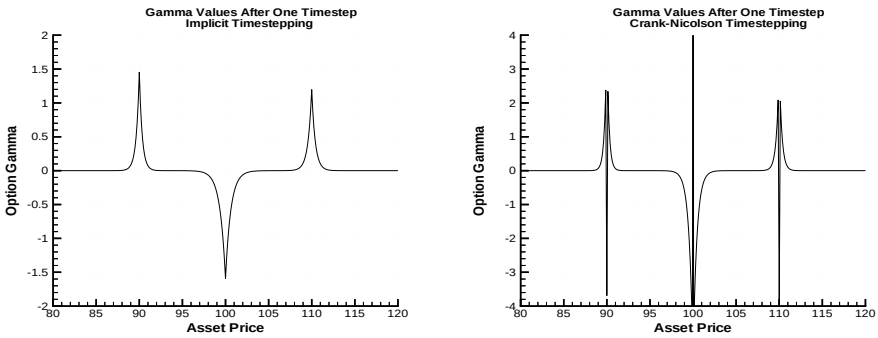
(f) Gamma, Crank–Nicolson timestepping.

FIG. 2. Butterfly spread solution value (U), delta (U_S), and gamma (U_{SS}) for both fully implicit and Crank–Nicolson timestepping. Parameters are provided in Table 1.



(a) Fully implicit timestepping. (b) Crank–Nicolson timestepping.

FIG. 3. Butterfly spread solution value (U) after the first timestep.



(a) Fully implicit timestepping. (b) Crank–Nicolson timestepping.

FIG. 4. Butterfly spread solution gamma (U_{SS}) after the first timestep. Note that the sign of gamma is different at the strike prices of 90, 100, and 110 where the delta is discontinuous.

in Table 5, Rannacher timestepping reduces the average number of nonlinear iterations per timestep to approximately the same levels as for implicit timestepping.

It is interesting to observe that difficulties normally arise in nonlinear financial PDEs when the PDE degenerates to a nonlinear hyperbolic problem. In our case, no degeneracy occurs. However, the payoff has a discontinuous first derivative, which seems to be enough to cause difficulty. The use of a few fully implicit steps at the start smooths the solution, curing the problem.

An obvious approach which avoids having to solve a set of nonlinear algebraic equations at each step is to evaluate the uncertain volatility explicitly. More precisely,

TABLE 4 *Convergence results for an at-the-money* ($S = 100$) *butterfly spread with uncertain volatility and Rannacher timestepping. Parameters are provided in Table 1. The timestep is halved at each grid refinement. 'Difference' is the absolute value of the change in the solution as the grid is refined. 'Ratio' is the ratio of successive differences. Timestepping data are given in Table 5*

	Rannacher (2 steps)			Rannacher (4 steps)		
Nodes	Value	Difference	Ratio	Value	Difference	Ratio
61	2.298 5785			2.304 0657		
121	2.298 0535	0.000 5250		2.299 6153	0.004 4504	
241	2.297 7860	0.000 2675	1.96	2.298 1945	0.001 4208	3.13
481	2.297 7116	0.000 0744	3.60	2.297 8172	0.000 3773	3.77
961	2.297 6910	0.000 0206	3.61	2.297 7178	0.000 0994	3.80

TABLE 5 *Timestepping information for an at-the-money* ($S = 100$) *butterfly spread with uncertain volatility and Rannacher timestepping. Parameters are provided in Table 1. 'No of iterations' is the total number of nonlinear iterations used during the solution process. 'Average no of iterations' is the number of iterations divided by the number of timesteps. The number of nodes is doubled as the timestep sized is halved. The convergence tolerance was* 10^{-6} (3.14). *Convergence data are given in Table 4*

		Rannacher (2 steps)		Rannacher (4 steps)	
No of timesteps	$\Delta\tau$	No of iterations	Average no of iterations	No of iterations	Average no of iterations
25	0.01	59	2.36	59	2.36
50	0.005	116	2.32	118	2.36
100	0.002 5	235	2.35	235	2.35
200	0.001 25	459	2.30	456	2.28
400	0.000 625	847	2.12	847	2.12

(3.13) becomes

$$\left[I + (1-\theta)\hat{M}^n\right] U^{n+1} = \left[I - \theta\hat{M}^n\right] U^n. \tag{4.2}$$

Note that this is equivalent to forcing one nonlinear iteration per timestep. Since $\left[I + (1-\theta)\hat{M}^n\right]$ is a diagonally dominant M-matrix, algorithm (4.2) with $\theta = 0$ is a positive coefficient discretization (Jameson, 1995). Consequently, algorithm (4.2) with $\theta = 0$ is unconditionally stable, but not monotone. As a result, convergence to the viscosity solution is not guaranteed.

Results using this approach are given in Table 6. Algorithm (4.2) with $\theta = 0$ leads to linear convergence, and apparently converges to the viscosity solution. Algorithm (4.2) with $\theta = 1/2$ leads to values away from the correct solution, with perhaps some instability. Setting $\theta = 1/2$ in algorithm (4.2) and using Rannacher timestepping (with four initial steps using $\theta = 0$) leads to linear convergence. The potential for quadratic convergence is lost by only taking one nonlinear iteration per timestep. However, if high accuracy is

TABLE 6 *Convergence results for an at-the-money* ($S = 100$) *butterfly spread with uncertain volatility, solved by forcing one nonlinear iteration per timestep* (*as per* (4.2)). *Parameters are provided in Table 1. 'Difference' is the change in the solution as the grid is refined. 'Ratio' is the ratio of successive differences. The timestep is halved as the grid is refined*

	Algorithm (4.2) ($\theta = 0$)			Algorithm (4.2) ($\theta = 1/2$)			Algorithm (4.2) ($\theta = 1/2$) and Rannacher (4 steps)		
Nodes	Value	Difference	Ratio	Value	Difference	Ratio	Value	Difference	Ratio
61	2.423 4699			3.210 0384			2.376 6897		
121	2.367 6732	0.055 80		3.259 3402	0.049 30		2.341 9918	0.034 70	
241	2.334 7374	0.032 94	1.69	3.312 7764	0.053 44	0.92	2.321 2272	0.020 76	1.67
481	2.316 9350	0.017 80	1.85	3.317 6336	0.004 86	11.0	2.309 9994	0.011 22	1.85
961	2.307 5759	0.009 36	1.90	3.314 1664	-0.003 46	-1.40	2.304 0654	0.005 93	1.89

not required, the simplicity and speed of using fully implicit timestepping with the linear approximation (explicit evaluation of the volatility) may be advantageous.

4.2 *Digital call options*

We have seen how a non-smooth payoff condition can cause problems for non-monotone schemes. This situation will be even more problematic for discontinuous payoffs. A digital call option has the payoff

$$U(S, \tau = 0) = \begin{cases} 1 & \text{if } S \geqslant K \\ 0 & \text{if } S < K. \end{cases} \tag{4.3}$$

(Recall that $\tau = T - t$, so that the payoff is the value of the option at expiry $t = T$.) Previous research has shown that when solving PDEs with discontinuous initial conditions, it is beneficial to smooth the initial conditions (Pooley *et al.*, 2002). The most theoretically sound method for doing this is projecting the initial conditions onto the set of basis functions used to discretize the equations (Wahlbin, 1980). We have used projection onto the space of linear basis functions in the following digital option tests (note that such a projection would have no effect on the butterfly payoff above, and strictly speaking, is not required for implicit timestepping).

The digital option problem is solved on the same set of grids as for the butterfly spread. At each refinement stage, the number of nodes is doubled and the timestep is halved. The convergence tolerance for the nonlinear iteration was 10^{-6} (3.14). The remaining parameters are given in Table 7.

Results for fully implicit, Crank–Nicolson, and Rannacher timestepping (four implicit steps) are given in Table 8. Implicit timestepping leads to consistent linear convergence. The Crank–Nicolson values appear to be unstable, having gone negative and approximately doubling in magnitude at each refinement. This indicates that the timestep restriction (3.45) is of practical importance. Using Rannacher timestepping with four implicit steps restores convergence, but only at a linear rate. Similar observations apply to the average number of nonlinear iterations per timestep, as shown in Table 9. Both fully implicit and Rannacher timestepping take an average of just over two nonlinear iterations, while Crank–Nicolson requires over four iterations on average.

TABLE 7 *Model parameters for the digital call option test problem*

Type:	Digital call
Scenario:	Worst case long
Time to expiry:	0.25 years
r:	0.1
K:	100
$\sigma_{\max}$:	0.25
$\sigma_{\min}$:	0.15

TABLE 8 *Convergence results for an at-the-money* ($S = 100$) *digital call option with uncertain volatility. 'Difference' is the absolute value of the change in the solution as the grid is refined. 'Ratio' is the ratio of successive differences. The timestep is halved at each grid refinement. Timestepping data are given in Table 9*

	Fully Implicit			Crank–Nicolson			Rannacher (4 steps)		
Nodes	Value	Difference	Ratio	Value	Difference	Ratio	Value	Difference	Ratio
61	0.443 1338			-0.255 73486			0.443 2821		
121	0.442 5493	0.000 5845		-1.299 4648	1.044		0.442 6167	0.000 6654	
241	0.442 2251	0.000 3242	1.80	-3.437 9189	2.138	0.488	0.442 2552	0.000 3615	1.84
481	0.442 0542	0.000 1709	1.90	-7.766 4872	4.329	0.494	0.442 0673	0.000 1878	1.92
961	0.441 9641	0.000 0901	1.90	-16.644 796	8.878	0.488	0.441 9698	0.000 0975	1.93

TABLE 9 *Timestepping information for an at-the-money* ($S = 100$) *digital call option with uncertain volatility. 'No of iterations' is the total number of nonlinear iterations used during the solution process. 'Average no of iterations' is the number of iterations divided by the number of timesteps. The number of nodes is doubled as the timestep is halved. The convergence tolerance was* 10^{-6} (3.14). *Convergence data are given in Table 8*

		Fully implicit		Crank–Nicolson		Rannacher (4 steps)	
No of timesteps	$\Delta\tau$	No of iterations	Average no of iterations	No of iterations	Average no of iterations	No of iterations	Average no of iterations
25	0.01	55	2.20	105	4.20	55	2.20
50	0.005	110	2.20	212	4.24	109	2.18
100	0.002 5	212	2.12	442	4.42	208	2.08
200	0.001 25	407	2.04	877	4.39	406	2.03
400	0.000 625	805	2.01	1797	4.49	803	2.01

Given the lack of quadratic convergence (even for Rannacher timestepping), one may think that the simplicity of linearizing the problem as per equation (4.2) is even more advantageous for digital options. However, as shown by the values in Table 10, this is not necessarily true. The results for algorithm (4.2) ($\theta = 0$) and algorithm (4.2) ($\theta = 1/2$) with Rannacher timestepping (first four steps use $\theta = 0$ in (4.2), and $\theta = 1/2$ thereafter) appear to be convergent, but only at a sub-linear rate. Values for algorithm (4.2) using $\theta = 1/2$ at all steps are again unstable. Unless very low accuracy is desired, solving the nonlinear

TABLE 10 *Convergence results for an at-the-money* ($S = 100$) *digital call option with uncertain volatility, solved by forcing one nonlinear iteration per timestep* (*as per* (4.2)). *Parameters are provided in Table 7. 'Difference' is the absolute value of the change in the solution as the grid is refined. 'Ratio' is the ratio of successive differences. The timestep is halved at each grid refinement*

	Algorithm (4.2) ($\theta = 0$)			Algorithm (4.2) ($\theta = 1/2$)			Algorithm (4.2) ($\theta = 1/2$) and Rannacher (4 steps)		
Nodes	Value	Difference	Ratio	Value	Difference	Ratio	Value	Difference	Ratio
61	0.459 603			0.933 258			0.459 550		
121	0.456 200	0.003 40		1.374 314	0.441		0.456 176	0.003 37	
241	0.452 571	0.003 63	0.94	2.172 795	0.798	0.552	0.452 566	0.003 61	0.93
481	0.449 795	0.002 78	1.31	3.746 378	1.57	0.507	0.449 796	0.002 77	1.30
961	0.447 641	0.002 16	1.29	6.941 466	3.20	0.493	0.447 642	0.002 15	1.29

equations with fully implicit timestepping may be the best choice for digital options with uncertain volatility.

5. Conclusions

If an implicit method is used to discretize the nonlinear PDE for pricing options with uncertain volatility, then we are faced with having to solve a set of nonlinear algebraic equations. Provided that the discretization satisfies certain conditions, we have shown that a non-smooth Newton iteration scheme is globally convergent, with quadratic convergence near the solution. For either the fully implicit method or Crank–Nicolson with Rannacher timestepping, we observed convergence occurring in just over two iterations per timestep, on average.

We have proven that a fully implicit discretization is monotone, and hence converges to the viscosity solution of the PDE. On the other hand, Crank–Nicolson is only conditionally monotone. Numerical examples show that Crank–Nicolson can generate incorrect (i.e. not viscosity) solutions to the PDE, or even unstable results, if a timestep is used which results in a non-monotone discretization.

Numerical experiments further show that, for Crank–Nicolson timestepping, we can converge to the viscosity solution if we take a small number (2–4) of fully implicit steps at the beginning, followed by Crank–Nicolson thereafter. For continuous, but non-smooth payoffs, numerical experiments indicate convergence at a quadratic rate, which is an improvement over the linear convergence of a fully implicit method.

Unfortunately, quadratic convergence (using Rannacher timestepping) could not be achieved for discontinuous payoffs (e.g. digital options). In this case, fully implicit timestepping should be used, since convergence to the viscosity solution is guaranteed.

Evaluating the nonlinear term explicitly avoids the need to solve nonlinear equations at each step, but this approach appears to converge at a slower rate than implicit methods. This method is also not guaranteed to converge to the viscosity solution, although we have not seen this occur in our numerical tests.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and RBC Financial Group.

REFERENCES

ANDERSEN, L. B. G. & BROTHERTON-RATCLIFFE, R. (1998) The equity option volatility smile: An implicit finite-difference approach. *J. Comput. Finance*, **1**, 5–37.

AVELLANEDA, M. & BUFF, R. (1999) Combinatorial implications of nonlinear uncertain volatility models: the case of barrier options. *Appl. Math. Finance*, **6**, 1–18.

AVELLANEDA, M., LEVY, A. & PARÁS, A. (1995) Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl. Math. Finance*, **2**, 73–88.

BARLES, G. (1997) Convergence of numerical schemes for degenerate parabolic equations arising in finance. *Numerical Methods in Finance*. (L. C. G. Rogers & D. Talay, eds). Cambridge: Cambridge University Press, pp. 1–21.

BLACK, F. & SCHOLES, M. (1973) The pricing of options and corporate liabilities. *J. Political Economy*, **81**, 637–659.

COLEMAN, T. F., LI, Y. & VERMA, A. (1999) Reconstructing the unknown local volatility function. *J. Comput. Finance*, **2**, 77–102.

CRANDALL, M. G., ISHII, H. & LIONS, P. L. (1992) User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc.*, **27**, 1–67.

CRANDALL, M. G. & LIONS, P. L. (1983) Viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, **277**, 1–42.

DOKUCHAEV, N. G., SAVKIN, A. V. &, (1998) The pricing of options in a financial market with transaction costs and uncertain volatility. *J. Multinational Financial Management*, **8**, 353–364.

FLEMING, W. H. & SONAR, H. M. (1993) Controlled Markov Processes and Viscosity Solutions. *Applications of Mathematics, 25*. New York: Springer.

FORSYTH, P. A. & VETZAL, K. R. (2001) Implicit solution of uncertain volatility/transaction cost option pricing models with discretely observed barriers. *Appl. Numer. Math.*, **36**, 427–445.

HESTON, S. L. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Studies*, **6**, 327–343.

JAMESON, A. (1995) Positive schemes and shock modelling for compressible flow. *Internat. J. Numer. Meth. Fluids*, **20**, 743–776.

LELAND, H. E. (1985) Option pricing and replication with transaction costs. *J. Finance*, **40**, 1283–1301.

LEVEQUE, R. J. (1990) *Numerical Methods for Conservation Laws*. Basel: Birkhäuser.

LYONS, T. (1995) Uncertain volatility and the risk free synthesis of derivatives. *Appl. Math. Finance*, **2**, 117–133.

LYONS, T. & SMITH, A. T. (1999) Uncertain volatility. *Risk*, **12**, 106–109.

MERTON, R. C. (1973) Theory of rational option pricing. *Bell J. Economics and Management Science*, **4**, 141–183.

PANG, J.-S. & QI, L. (1993) Nonsmooth equations: motivation and algorithms. *SIAM J. Optimiz.*, **3**, 443–465.

POOLEY, D. M., FORSYTH, P. A. & VETZAL, K. R. (2002) Convergence remedies for non-smooth payoffs in option pricing. *J. Comput. Finance*, accepted.

QI, L. & SUN, J. (1993) A nonsmooth version of Newton's method. *Mathematical Programming*, **58**, 353–367.

QI, L. & ZHOU, G. (2000) A smoothing Newton method for minimizing a sum of Euclidean norms. *SIAM J. Optimiz.*, **11**, 389–410.

RANNACHER, R. (1984) Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, **43**, 309–327.

SUN, D. & HAN, J. (1997) Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. *SIAM J. Optimiz.*, **7**, 463–480.

WAHLBIN, L. B. (1980) A remark on parabolic smoothing and the finite element method. *SIAM J. Numer. Anal.*, **17**, 33–38.

WILMOTT, P. (1998) *Derivatives*. Chichester: Wiley.

ZVAN, R., FORSYTH, P. A. & VETZAL, K. R. (2001) A finite volume approach for contingent claims valuation. *IMA J. Numer. Anal.*, **21**, 703–731.

Appendix. Viscosity solutions

The concept of a viscosity solution is closely related to the entropy condition (E-condition) of weak solutions to conservation law problems (LeVeque, 1990). First introduced in Crandall & Lions (1983), a full description of the theory is given in Crandall *et al.* (1992). We give here a brief overview of the concept of a viscosity solution.

To understand the notion of a viscosity solution, consider the general form of a second-order parabolic PDE:

$$\frac{\partial V}{\partial t} + F(x, V, V_x, V_{xx}) = 0. \tag{A.1}$$

We assume that F satisfies the ellipticity condition

$$F(x, V, V_x, V_{xx} + \epsilon) \leqslant F(x, V, V_x, V_{xx}) \text{ if } \epsilon \geqslant 0. \tag{A.2}$$

(Note the sign convention for F as given in (A.1), which is standard in the literature.) This property is crucial for the definition of the viscosity solution. For discrete equations, the concept of ellipticity is replaced by monotonicity. This is why a monotone discretization is required (at least in theory) to guarantee convergence to the viscosity solution.

To motivate the definition, consider functions $v(t, x) \in C^2$ and $V(t, x) \in C^2$. Let (t_0, x_0) be a local maximum of $V - v$. From basic calculus, we know that $\partial V/\partial t = \partial v/\partial t$, $V_x = v_x$, and $v_{xx} \geqslant V_{xx}$ near (t_0, x_0). Using these relations and the ellipticity property (A.2) gives

$$\frac{\partial V}{\partial t}(t_0, x_0) + F(x_0, V(t_0, x_0), v_x(t_0, x_0), v_{xx}(t_0, x_0)) \leqslant 0. \tag{A.3}$$

If (A.3) holds for all $v \in C^2$, then V is said to be a *viscosity subsolution* of (A.1). In a sense, we have used the functions v to provide an upper bound to possible solutions. Similarly, $V(t, x)$ is a *viscosity supersolution* of (A.1) if $\forall v \in C^2$, and if (t_0, x_0) is a local minimum point of $V - v$, then

$$\frac{\partial V}{\partial t}(t_0, x_0) + F(x_0, V(t_0, x_0), v_x(t_0, x_0), v_{xx}(t_0, x_0)) \geqslant 0. \tag{A.4}$$

A viscosity solution of (A.1) is a solution that is both a viscosity subsolution and a viscosity supersolution. Note that a classical solution to equation (A.1) is also a viscosity solution (this can be verified by letting $v = V$).

However, we can still use the definitions (A.3)–(A.4) in the case that V is not smooth, since we do not require the existence of the first and second derivatives of V. In this case, we can define non-smooth solutions to (A.1). This definition essentially sandwiches the desired non-smooth viscosity solution between smooth solutions that are 'above the PDE' or 'below the PDE', in the sense of definitions (A.3)–(A.4).

It is shown in Fleming & Sonar (1993) that the viscosity solution of a nonlinear parabolic option pricing equation is the desirable solution in financial applications. Further, Barles (1997) proves that a stable, consistent and monotone discretization of option pricing problems must converge to the desired viscosity solution.