

LAGRANGE MULTIPLIER APPROACH WITH OPTIMIZED FINITE DIFFERENCE STENCILS FOR PRICING AMERICAN OPTIONS UNDER STOCHASTIC VOLATILITY

KAZUFUMI ITO* AND JARI TOIVANEN†

Abstract. The deterministic numerical valuation of American options under Heston’s stochastic volatility model is considered. The prices are given by a linear complementarity problem with a two-dimensional parabolic partial differential operator. A new truncation of the domain is described for small asset values while for large asset values and variance a standard truncation is used. The finite difference discretization is constructed by numerically solving quadratic optimization problem aiming to minimize the truncation error at each grid point. A Lagrange approach is used to treat the linear complementarity problems. Numerical examples demonstrate the accuracy and effectiveness of the proposed approach.

Key words. American option pricing, stochastic volatility model, linear complementarity problem, finite difference method, quadratic programming, multigrid method, Lagrange method, penalty method

AMS subject classifications. 35K85, 65M06, 65M55, 65Y20, 91B28

1. Introduction. The seminal papers [2] and [28] by Black & Scholes, and Merton, respectively, laid the foundations of the modern theory of pricing financial options. Since then vast body of scientific work has been devoted to the development of methods for pricing options. Particularly American options are challenging to evaluate due to their early exercise possibility and various approaches to approximate their price have been proposed. The paper [4] by Brennan and Schwartz was one of first ones to formulate a linear complementarity problem (LCP) and then to solve it using a finite difference discretization. Empirical evidence has shown that the assumption on log-normal behavior of the value of the underlying asset is not realistic for many asset categories including shares of companies. To alleviate this, many generalizations have been introduced for the log-normal model. We consider stochastic volatility models which assume that the volatility of the value of the asset follows a stochastic process; see [11] and references therein. Particularly, our model problem for American options is based on Heston’s model [14], but the techniques considered in this paper can be generalized also for other stochastic volatility models like the Hull-White model [17] and the Stein-Stein model [33].

Based on Heston’s stochastic volatility model, a LCP with a parabolic partial differential operator can be derived for the price of American options with the value of underlying asset and its variance being the spatial variables. Near the axes, the first-order derivatives dominate the second-order derivatives in the operator. It also includes the second-order cross derivative. Due to these properties, it is not easy to construct an accurate and stable discretization. In financial literature, most often finite differences are used for the discretization while sometimes also finite elements are used; see [1], [34], [38], for example. It is desirable to use discretizations leading to matrices with the M-matrix property [39]; see also [41]. This property guarantees the stability of the spatial discretization and the feasibility and monotone property

*Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA. (kito@ncsu.edu).

†Institute for Computational and Mathematical Engineering, Building 500, Stanford University, Stanford, CA 94305, USA, (toivanen@stanford.edu).

of the numerical solutions. Many different solution methods have been proposed for the resulting discrete LCPs; see [8], [9], [21], [29], [40]. While these methods have pronounced impact on the efficiency of the valuation of the option the discretization often has more influence on it. In the following, we describe briefly methods considered in the scientific literature.

Clarke and Parrott apply coordinate transform and then perform the discretization using finite differences on uniform grids in [8], [9]. Their mainly central finite difference discretization switches the discretization of the first-order derivative to an upwind difference when the central difference could lead to oscillations. They employ constant and adaptive time steps with a θ -scheme using the value of θ making it close to the Crank-Nicolson method. The resulting LCPs are solved by a projected full approximation scheme (PFAS) multigrid method [3].

In [40], Zvan, Forsyth, and Vetzal use linear finite elements for discretizing the second-order terms and a finite volume discretization for the first-order terms. Furthermore, they use a nonlinear van Leer flux limiter to avoid oscillations. Their grids are nonuniform and adaptive time steps are employed with the Crank-Nicolson method. They consider two penalty methods for LCPs. The arising nonlinear and linear problems are solved using a semismooth Newton method and the BICGSTAB method with an incomplete LU preconditioner, respectively.

Oosterlee uses second-order finite difference discretizations on uniform grids and time integration is performed using the BDF2 method with constant time steps in [29]. The resulting LCPs are solved using the PFAS multigrid method.

In [21], seven point finite difference discretizations are used on nonuniform grids. The discretization is constructed in such a way that an M-matrix is obtained. Due to this, in a part of the domain the discretization is only first-order accurate. The time discretizations are based on the Rannacher scheme [31] with constant time steps. Moreover, in [21], the projected SOR method, a projected multigrid method [32], an operator splitting method [18], [19] with a multigrid, a penalty method [40] with a multigrid, and an componentwise splitting method [20] are compared.

In this paper, we consider techniques to improve the efficiency of the computation of the price of options. Under Heston's model the first-order derivative terms dominates the second-order derivative terms for small asset values. In this part of the domain, a fine grid is required in order to construct a stable discretization leading an M-matrix. Here we prove that the asset values can be truncated to (X_{min}, ∞) without effecting the prices of American options. This way one can remove fine grid for small asset values and significantly improve the computational efficiency. We propose effective choices for the lower bound $X_{min} > 0$ in Section 3. For large asset values and variances a nonuniform grid can be made coarse without increasing error. With such a grid the location of the truncation boundary can be chosen to be fairly far without essentially increasing computational work.

It is not easy to construct accurate discretization leading to an M-matrix due to varying coefficients, vanishing second-order derivative terms, and the presence of a second-order cross derivative term. Reducing the truncation error at grid points improves the accuracy. The M-matrix property can be guaranteed by restricting the offdiagonal weights of the finite element stencil to be non positive. Instead of constructing stencils by hand, we formulate an optimization problem for the stencil aiming to minimize the truncation error under the non positivity constraint for offdiagonal weights. This leads to a small, standard quadratic programming (QP) problem at each grid point. The numerical experiments demonstrate that these QPs can be

solved fast and the resulting discretization leads to very good accuracy. Earlier such an optimization approach has been used to construct stencils for constant coefficient operators near an interface in [26] and its follow-up papers. It is a novel idea to use this approach for varying coefficient operators in the whole computational domain and not just near boundary/interface.

We solve the resulting LCPs using a Lagrange multiplier method [15], [22]; see also [1]. With a suitable choice of parameters (see $\bar{\lambda}$ in (5.2)), this method leads to option prices which always satisfy the early exercise constraint, that is, the prices are arbitrage free in this sense. For the option pricing a penalty method considered in [10] and [40] is also obtained from the Lagrange multiplier method with a trivial choice of the parameters. The penalty method leads to small violations of the early exercise constraint. The numerical experiments show that the Lagrange multiplier method and the penalty method are very similar with respect to the accuracy and computational effort. In [22] an error estimate of the proposed method is obtained. The method uses the semi-smooth Newton [23] and it becomes standard for solving nonsmooth equations. In addition, we refer to [6], [25], [36] and references therein for the theory and applications of the semi-smooth Newton method.

The outline of the paper is the following. Section 2 describes a model LCP obtained from Heston's stochastic volatility model for American put options. The truncation of the computational domain is considered in Section 3. The optimization approach to construct finite difference stencils and time discretization are described in Section 4. Section 5 considers a Lagrange multiplier approach for solving the discrete LCPs. Numerical experiments are presented in Section 6 and the paper ends with conclusions.

2. Model problem. In this section, we describe a linear complementarity problem (LCP) for pricing American put options. Under Heston's stochastic volatility model and with suitable assumptions on the markets a LCP with a two-dimensional parabolic partial differential operator can be derived for the price of American options; see [14], [38], [40]. In general, these LCPs need to be solved numerically; see, for example, [9], [16], [29], [30], [40].

We denote the price of the option by v , the time by t , the price of the underlying asset by x , and its variance by y . Heston's model leads to a spatial operator

$$Av = -\frac{1}{2}y x^2 v_{xx} - \rho\gamma y x v_{xy} - \frac{1}{2}\gamma^2 y v_{yy} - r x v_x - (\alpha(\beta - y) - \vartheta\gamma\sqrt{y}) v_y + r v, \quad (2.1)$$

where the r is the risk free interest rate, β is the mean level of the variance, α is the rate of reversion on the mean level, and γ is the volatility of the variance. The market price of the risk ϑ is assumed to be zero in the rest of the paper similarly to [9], [29], [40]. The correlation between the price of the underlying asset and its variance is ρ .

The price at the exercise moment is given by the payoff function ψ . Furthermore, it also defines a final condition

$$v(x, y, T) = \psi(x). \quad (2.2)$$

For a put option with the exercise price K , the payoff function is

$$\psi(x) = \max\{K - x, 0\}. \quad (2.3)$$

Due to the early exercise possibility of the American option the price v has to be at least the same as the payoff function ψ . This leads to the early exercise constraint

$v(x, y, t) \geq \psi(x)$. In the region where the constraint is inactive the price v satisfies a partial differential equation $\frac{d}{dt}v - Av = 0$. By combining these relations, a time dependent linear complementarity problem is obtained for the price of the American option. It is not known a priori where the constraint is active and this makes deriving analytical formulas for the price intractable.

The linear complementarity problem (LCP) for the price of the American put option is

$$\frac{d}{dt}v(x, y, t) - Av \geq 0 \quad \perp \quad v(x, y, t) \geq \psi(x) \quad (2.4)$$

in a domain $\{(x, y, t) \mid x > 0, y > 0, t \in (0, T)\}$ with the final condition (2.2). The symbol \perp indicates that both inequalities are satisfied with at least one of them holding as equality a.e. On the boundaries $x = 0$ and $y = 0$, Dirichlet boundary conditions

$$v(0, y, t) = \psi(0) \quad \text{and} \quad v(x, 0, t) = \psi(x) \quad (2.5)$$

are posed. On far-field the asymptotic behavior of v satisfies the conditions

$$\lim_{x \rightarrow \infty} v_x(x, y, t) = 0 \quad \text{and} \quad \lim_{y \rightarrow \infty} v_y(x, y, t) = 0. \quad (2.6)$$

These conditions are described in [9] and [29], for example.

It is more customary to solve forward problems in time than backward ones. For this reason, we revert the time by defining a price function u to be $u(x, y, t) = v(x, y, T - t)$ and considering the solution of u instead of v . The price u satisfies the LCP

$$\frac{d}{dt}u + Au \geq 0 \quad \perp \quad u \geq \psi \quad (2.7)$$

in the domain $\{(x, y, t) \mid x > 0, y > 0, t \in (0, T)\}$ with an initial condition

$$u(x, y, 0) = \psi(x). \quad (2.8)$$

The boundary conditions are obtained from (2.5) and (2.6) by replacing v by u .

3. Truncation of domain.

3.1. Upper truncation boundary for y . We start by truncating the infinite computational domain in the direction of variance y . For this purpose, we choose sufficiently large value Y . Then the LCP (2.7) is posed in the domain $\{(x, y, t) \mid x > 0, y \in (0, Y), t \in (0, T)\}$ with the boundary condition

$$u_y = 0 \quad \text{on} \quad (0, \infty) \times \{Y\}. \quad (3.1)$$

Based on numerical studies it has been observed that Y should be, say, from four to six times larger than the value of the variance of interest in order to obtain prices with sufficient accuracy for practical purposes [9], [29], [40].

3.2. Black-Scholes model for American options. In order to derive a lower truncation boundary for x , we start by considering the one-dimensional Black-Scholes model for American options given by a LCP of the form

$$\begin{aligned} \frac{d}{dt}\hat{v}(x, t) + \frac{1}{2}\sigma^2x^2\hat{v}_{xx} + rx\hat{v}_x - r\hat{v} &\leq 0 \quad \perp \quad \hat{v}(x, t) \geq \psi(x) \\ \hat{v}(x, T) &= \psi(x) \end{aligned} \quad (3.2)$$

a.e. $(x, t) \in (0, \infty) \times (0, T)$. The volatility of the asset price is $\sigma > 0$. The complementarity system (3.2) has the following interpretation in mathematical finance. The price process X_t is governed by the Ito's stochastic differential equation:

$$dX_t = rX_t dt + \sigma S_t dB_t, \quad (3.3)$$

or

$$X_T = \exp\left(\left(r - \frac{1}{2}\sigma^2\right)(T - t) + \sigma(B_T - B_t)\right) X_t \quad (3.4)$$

where B_t denotes the standard Brownian motion and the value function \hat{v} is represented by

$$\hat{v}(x, t) = \sup_{\tau} E^{x,t} \left[e^{-r(\tau-t)} \psi(X_{\tau}) \right] \quad \text{over all stopping times } t \leq \tau \leq T. \quad (3.5)$$

It can be easily shown from this that $\hat{v}(x, t)$ is monotonically nonincreasing in both $x > 0$ and $t < T$.

To express (3.2) in variational form we define

$$a(\hat{v}, \phi) = \int_{X_{\min}}^{\infty} \left(\left(\frac{1}{2}\sigma^2 x^2 \hat{v}_x + (r - \sigma^2)x \hat{v} \right) \phi_x + (2r - \sigma^2)\hat{v}\phi \right) dx \quad (3.6)$$

for $\hat{v}, \phi \in V$, where V is the completion of the space

$$\left\{ \phi \in H \mid \phi \text{ is absolutely continuous on } (X_{\min}, \infty), \right. \\ \left. \int_{X_{\min}}^{\infty} x^2 |\phi_x|^2 dx < \infty \text{ and } \phi(x) \rightarrow 0 \text{ as } x \rightarrow \infty \text{ and } x \rightarrow X_{\min} \right\} \quad (3.7)$$

under the norm

$$|\phi|_V^2 = \int_{\bar{X}}^{\infty} (x^2 |\phi_x|^2 + |\phi|^2) dx. \quad (3.8)$$

The solution to (3.2) satisfies $\hat{v} - \psi \in V$. Setting $\hat{u}(x, t) = \hat{v}(x, T - t) - \psi$ we arrive at

$$\left\langle \frac{d}{dt} \hat{u}(t), w - \hat{u}(t) \right\rangle + a(\hat{u}(t), w - \hat{u}(t)) - a(\psi, w - \hat{u}(t)) \geq 0 \quad \text{for all } w \in \mathcal{C}, \quad (3.9)$$

$$\hat{u}(x, 0) = 0,$$

where $\mathcal{C} = \{w \in V \mid w \geq 0\}$. Here, we can choose the interval on which the solution \hat{v} is defined, as (X_{\min}, ∞) , $X_{\min} > 0$ and it allows to avoid the singularity at 0. The LCP (3.2) has the equilibrium solution of the form

$$\bar{v}(x) = \begin{cases} K - x, & x \leq X_{\min}, \\ (K - X_{\min}) \left(\frac{x}{X_{\min}} \right)^{-\gamma}, & x \geq X_{\min}, \end{cases} \quad (3.10)$$

where

$$X_{\min} = \frac{K\kappa}{1 + \kappa} \quad \text{and} \quad \kappa = \frac{2r}{\sigma^2}. \quad (3.11)$$

The equilibrium solution satisfies the Cauchy-Euler equation

$$\frac{1}{2}\sigma^2 x^2 \hat{v}_{xx} + r x \hat{v}_x - r \hat{v} = 0 \quad \text{in } (X_{\min}, \infty). \quad (3.12)$$

Thus, it can be expressed as

$$\hat{v}(x) = C_1 x^{s_1} + C_2 x^{s_2}, \quad (3.13)$$

where $s_1 = -\kappa$ and $s_2 = 1$ satisfy

$$\frac{1}{2}\sigma^2 s(s-1) + r s - r = \left(\frac{1}{2}\sigma^2 s + r\right)(s-1) = 0. \quad (3.14)$$

Since $\hat{v} \rightarrow 0$ as $x \rightarrow \infty$, we have $\hat{v} = C_1 x^{-\kappa}$. As $\hat{v} \in H^2(0, \infty)$ [22] we must have

$$\hat{v}(X_{\min}) = K - X_{\min}, \quad \hat{v}_x(X_{\min}) = (K - X_{\min}) \frac{-\kappa}{X_{\min}} = -1, \quad (3.15)$$

which yields (3.10). Since \hat{v} is monotone in t , $\hat{v}(x, t) \leq \bar{v}(x)$ for all $t \leq T$ and $\hat{v}(x, t) \rightarrow \bar{v}(x)$ monotonically as $t \rightarrow -\infty$ for all $x \geq 0$. This validates our claim.

3.3. Lower truncation boundary for x . The corresponding bilinear form a for A defined by (2.1) is given by

$$\begin{aligned} a(u, \phi) = & \int \left(\frac{1}{2} y x^2 u_x \phi_x + (y x - r) u_x \phi + r u \phi \right. \\ & \left. + \frac{1}{2} (\rho \gamma (v_x \phi_y + v_y \phi_x) + \gamma^2 y v_y \phi_y) - \frac{\rho \gamma}{2} (x v_x \phi + y v_y \phi) + (\gamma^2 - \alpha(\beta - y)) v_y \phi \right) dx \end{aligned} \quad (3.16)$$

for $u, \phi \in H^1$. Thus, if $u^+ = \sup\{u, 0\} \in H^1$, then

$$a(u, u^+) \geq \int \left(\frac{1-\rho}{2} (y x^2 |u_x^+|^2 + \gamma^2 y |u_y^+|^2) + \frac{1}{2} (3r - y - \rho \gamma - \gamma^2 - \alpha) |u^+|^2 \right) dx. \quad (3.17)$$

Thus, it can be shown as in [22] that (2.7) has a solution $u \in H_{loc}^1(0, T; L^2) \cap L_{loc}^2(0, T; H^2)$ and that $u(\cdot, t)$ is monotonically nondecreasing in t . Let

$$\tilde{u}(x, y) = \bar{v}(x) \quad (3.18)$$

and u is the steady state solution to (2.7), i.e. satisfies

$$A u \geq 0 \quad \perp \quad u \geq \psi \quad \text{a.e.} \quad (3.19)$$

Then, the following theorem shows that the boundary value problem $v(X_{\min}, y, t) = \psi(X_{\min})$ at the truncated boundary $x = X_{\min}$ provides the exact solution to (2.4).

THEOREM 3.1. *Let \tilde{u} corresponds to the steady state \bar{v} of the one-dimensional model with $\sigma^2 = Y$, where Y is the upper bound of y . Then $\tilde{u} \geq u(x, y)$ and $x = X_{\min}$ provides a lower-bound estimate of the active region defined by $\{u = \psi\}$.*

Proof. On $\{u > \psi\} \cap \{\tilde{u} > \psi\}$

$$A(\tilde{u} - u) = -\frac{1}{2}(y - \sigma^2)\kappa(\kappa + 1) \left(\frac{x}{X_{\min}} \right)^{-\kappa} \geq 0 \quad (3.20)$$

and on $\{u > \psi\} \cap \{\tilde{u} = \psi\}$

$$A(\psi - u) = rK \geq 0. \quad (3.21)$$

The weak maximum principle shows that $\tilde{u} \geq u$ on $\{u > \psi\}$. \square

3.4. Time-dependent lower truncation boundary for \mathbf{x} . Let $\tilde{u}(x, y, t) = \hat{v}(x, T - t)$, where \hat{v} is the solution of (3.2). That is, \tilde{u} is the solution of the one-dimensional LCP

$$\frac{d}{dt}\tilde{u}(x, y, t) - \frac{1}{2}\sigma^2 x^2 \tilde{u}_{xx} - rx\tilde{u}_x + r\tilde{u} \leq 0 \quad \perp \quad \tilde{u}(x, y, t) \geq \psi(x) \quad (3.22)$$

with $\tilde{u}(x, y, 0) = \psi(x)$. The following theorem shows that a lower-bound estimate can be also obtained from the one-dimensional model (3.22) with $\sigma^2 = Y$.

THEOREM 3.2. *It holds that $\tilde{u}(\cdot, t) \geq u(\cdot, t)$ a.e. and, thus, the free surface $S(t) = \partial\{\tilde{u}(x, t) > \psi(x)\} \in (0, K)$ determined by the one-dimensional equation (3.22) with $\sigma^2 = Y$ provides a lower bound estimate for the one for (2.7).*

Proof. Since $\tilde{u}_x \leq 0$ and $\frac{d}{dt}\tilde{u} \geq 0$ a.e., $\tilde{u}_{xx} \geq 0$ holds a.e.. Furthermore, $\tilde{u}_y = 0$. Thus, on $\{u > \psi\} \cap \{\tilde{u} > \psi\}$

$$\frac{d}{dt}(\tilde{u} - u) = -A(\tilde{u} - u) + \frac{1}{2}(\sigma^2 - y)\tilde{u}_{xx} \quad (3.23)$$

and on $\{u > \psi\} \cap \{\tilde{u} = \psi\}$

$$\frac{d}{dt}(\tilde{u} - u) = -A(\tilde{u} - u) + rK. \quad (3.24)$$

Let $\Omega_t = \{(x, y)^T \mid u(x, y, t) > \psi(x)\}$ and $\phi = \inf\{\tilde{u} - u, 0\} \in H^1(0, T; L^2)$. Then for $\omega \geq 0$

$$\frac{d}{dt} \int_{\Omega_t} \frac{1}{2} e^{-\omega t} |\phi|^2 dx = e^{-\omega t} \left(\int_{\Omega_t} \left(\phi \frac{d}{dt}(\tilde{u} - u) - \omega |\phi|^2 \right) dx + \int_{\Gamma_t} \frac{1}{2} |\phi|^2 (h \cdot n) ds \right) \quad (3.25)$$

where $n = \nabla u / |\nabla u|$ and h is the velocity of the interface $\Gamma_t = \partial\Omega_t$. Since $\phi = 0$ at Γ_t and from (3.17) and (3.23)–(3.24)

$$\int_{\Omega_t} \phi \frac{d}{dt}(\tilde{u} - u) - \omega |\phi|^2 dx \leq - \int_{\Omega_t} ((A(\tilde{u} - u), \phi) + \omega |\phi|^2) dx \leq 0.$$

It thus follows from (3.23)–(3.25) that

$$\frac{d}{dt} \int_{\Omega_t} \frac{1}{2} e^{-\omega t} |\inf\{\tilde{u} - u, 0\}|^2 dx \leq 0 \quad (3.26)$$

and thus $\tilde{u} \geq u$ a.e. \square

3.5. Upper truncation boundary for \mathbf{x} . For the one-dimensional Black-Scholes model as well as for its multidimensional generalizations the truncation error has been studied in [24] when a Dirichlet boundary condition is posed on the truncation boundary. They conclude that the usual rule of thumb to truncate at three or four times the exercise price is usually sufficient. Based on numerical studies [9], [29], [40], for Heston's model with the homogenous Neumann boundary condition posed on the truncation boundary two or three times exercise price seems to be sufficient. We denote the location of the truncation boundary by X_{\max} and the Neumann boundary condition posed on it is given by

$$u_y = 0 \quad \text{on} \quad \{X_{\max}\} \times (0, Y). \quad (3.27)$$

4. Discretization. For the computational domain $[X_{\min}, X_{\max}] \times [0, Y] \times [0, T]$, we define a space-time grid

$$(x_i, y_j, t_k) \in \{x_0, \dots, x_m\} \times \{y_0, \dots, y_n\} \times \{t_0, \dots, t_l\}, \quad (4.1)$$

where $x_i < x_{i+1}$, $y_j < y_{j+1}$, $t_k < t_{k+1}$, $x_0 = X_{\min}$, $x_m = X_{\max}$, $y_0 = 0$, $y_n = Y$, $t_0 = 0$, and $t_l = T$.

4.1. Spatial finite difference discretization. The D th degree Taylor polynomial of the two-dimensional function $u(\mathbf{x} + \Delta \mathbf{x}^i)$ at $\mathbf{x} = (x, y)^T$ and the order of the remainder is given by

$$u(\mathbf{x} + \Delta \mathbf{x}^i) = u(\mathbf{x}) + \sum_{k=1}^M d_{k,i} \partial^{\delta^k} u(\mathbf{x}) + \mathcal{O}(\|\Delta \mathbf{x}^i\|^{D+1}), \quad (4.2)$$

where $M = \frac{1}{2}(D+1)(D+2) - 1$ and multi-indices δ^k s are defined as

$$\delta^k = (j - k, k + i - j), \quad i = \frac{1}{2}(\sqrt{9 + 8k} - 3), \quad j = \frac{1}{2}(i + 1)(i + 2) - 1. \quad (4.3)$$

That is, the nine first multi-indices δ^k are given by

$$\begin{aligned} \delta^1 &= (1, 0), & \delta^2 &= (0, 1), \\ \delta^3 &= (2, 0), & \delta^4 &= (1, 1), & \delta^5 &= (0, 2), \\ \delta^6 &= (3, 0), & \delta^7 &= (2, 1), & \delta^8 &= (1, 2), & \delta^9 &= (0, 3). \end{aligned} \quad (4.4)$$

The coefficients $d_{k,i}$ in (4.2) are defined as

$$d_{k,i} = \frac{1}{|\delta^k|!} \binom{|\delta^k|}{\delta_2^k} (\Delta \mathbf{x}^i)^{\delta^k} = \frac{1}{\delta_1^k! \delta_2^k!} (\Delta x^i)^{\delta_1^k} (\Delta y^i)^{\delta_2^k}. \quad (4.5)$$

Let us have N grid points $\mathbf{x} + \Delta \mathbf{x}^i$, $i = 1, \dots, N$, around \mathbf{x} and let the corresponding finite difference weights at these points be denoted by w_i . Furthermore, let the weight at \mathbf{x} be w_0 and $\Delta \mathbf{x}^0 = \mathbf{0}$. The standard compact nine point stencil and the notations are shown in Figure 4.1. Then an $N + 1$ point finite difference sum has an approximate Taylor polynomial presentation

$$\begin{aligned} \sum_{i=0}^N w_i u(\mathbf{x} + \Delta \mathbf{x}^i) &\approx \sum_{i=0}^N w_i u(\mathbf{x}) + \sum_{i=1}^N w_i \sum_{k=1}^M d_{k,i} \partial^{\delta^k} u(\mathbf{x}) \\ &\approx u(\mathbf{x}) \sum_{i=0}^N w_i + \left(\partial^{\delta^1} u(\mathbf{x}) \quad \dots \quad \partial^{\delta^M} u(\mathbf{x}) \right) \mathbf{D} \mathbf{w}, \end{aligned} \quad (4.6)$$

where we have use the notations

$$\mathbf{D} = \begin{pmatrix} d_{1,1} & \dots & d_{1,N} \\ \vdots & \ddots & \vdots \\ d_{m,1} & \dots & d_{m,N} \end{pmatrix} \quad \text{and} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix}. \quad (4.7)$$

With the finite difference sum (4.6), we approximate the operator

$$b_0 u(\mathbf{x}) + \sum_{k=1}^M b_k \partial^{\delta^k} u(\mathbf{x}) = b_0 u(\mathbf{x}) + \left(\partial^{\delta^1} u(\mathbf{x}) \quad \dots \quad \partial^{\delta^M} u(\mathbf{x}) \right) \mathbf{b}, \quad (4.8)$$

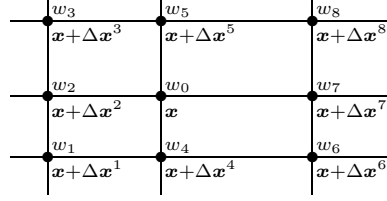


FIG. 4.1. The standard nine point finite difference stencil at \mathbf{x} with the coordinates of the grid points and the associated weights w_i .

where $\mathbf{b} = (b_1 \cdots b_N)^T$. For example, for the operator A in (2.1), the b_i s are given by

$$\begin{aligned} b_0 &= r, & b_1 &= -rx, & b_2 &= -\alpha(\beta - y), \\ b_3 &= -\frac{1}{2}yx^2, & b_4 &= -\rho\gamma yx, & b_5 &= -\frac{1}{2}\gamma^2 y, \end{aligned} \quad (4.9)$$

and $b_k = 0$, $k = 6, \dots, M$. By using (4.6) and (4.8), we obtain for the truncation error an estimate

$$\begin{aligned} & \sum_{i=0}^N w_i u(\mathbf{x} + \Delta \mathbf{x}^i) - b_0 u(\mathbf{x}) - \sum_{k=1}^M b_k \partial^{\delta^k} u(\mathbf{x}) \\ & \approx u(\mathbf{x}) \left(\sum_{i=0}^N w_i - b_0 \right) + \left(\partial^{\delta^1} u(\mathbf{x}) \quad \cdots \quad \partial^{\delta^M} u(\mathbf{x}) \right) (\mathbf{D}\mathbf{w} - \mathbf{b}). \end{aligned} \quad (4.10)$$

We eliminate the first term in the truncation error (4.10) by choosing the diagonal weight to be

$$w_0 = b_0 - \sum_{i=1}^N w_i. \quad (4.11)$$

There are many ways to choose the other N weights. One approach is to eliminate more terms from the truncation error. This was done in [26] and after this the distance from some given stencil added with a weighted sum of squares of terms in the truncation error was minimized. Our aim is to obtain an M-matrix which means that the weights w_i , $i = 1, \dots, N$, have to be non positive. Due to the varying coefficients of the operator A and the use of nonuniform grids, we do not know a priori how many terms we can eliminate and still obtain an M-matrix.

Our approach is to minimize the truncation error under the non positivity constraints for the offdiagonal weights. The truncation error depends on partial derivatives of $u(\mathbf{x})$ which we do not generally know a priori. In such a case, a natural approach is to minimize weighted squares of their coefficients in the truncation error. By choosing the weights of the squares to be ones, we obtain a least squares problem

$$\min_{\mathbf{w} \leq \mathbf{0}} \|\mathbf{D}\mathbf{w} - \mathbf{b}\|_2^2, \quad (4.12)$$

which yields a finite difference stencil leading to an M-matrix. The problem (4.12) is equivalent with a standard quadratic programming problem (QP)

$$\min_{\mathbf{w} \leq \mathbf{0}} \frac{1}{2} \mathbf{w}^T \widehat{\mathbf{D}} \mathbf{w} + \widehat{\mathbf{b}}^T \mathbf{w}, \quad (4.13)$$

where $\widehat{\mathbf{D}} = \mathbf{D}^T \mathbf{D}$ and $\widehat{\mathbf{b}} = -\mathbf{D}^T \mathbf{b}$. This quadratic optimization problem is formed and solved for each grid point. We denote the discretized operator A by a matrix \mathbf{A} . One optimization yields the weights w_i , $i = 0, \dots, N$, and from them the row of \mathbf{A} corresponding to the grid point under consideration is assembled.

The partial derivatives appearing in the truncation error can be estimated using an approximation for u which can be computed on fairly coarse grid. This a posteriori information can be used to choose the weights in the least squares problem for \mathbf{w} . This way more optimal finite difference stencils can be constructed. Any a priori information on the relative amplitude of partial derivatives can be used in the same manner.

The spatial discretization of (2.7) leads to a semi-discrete LCP

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{A} \mathbf{u} \geq \mathbf{0} \quad \perp \quad \mathbf{u} \geq \boldsymbol{\psi}, \quad (4.14)$$

where the vectors \mathbf{u} and $\boldsymbol{\psi}$ contain the grid point values of u and ψ , respectively. The symbol \perp indicates that both inequalities are satisfied componentwise with at least one of them is holding as equality for each component. The initial value of \mathbf{u} is given by $\boldsymbol{\psi}$.

4.2. Temporal discretization. The option pricing model has a nonsmooth initial value. Although the Crank-Nicolson method is popular it does not have good stability properties. In many cases the Crank-Nicolson method can lead to a numerical solution with oscillation because it is not L-stable. In order to obtain numerical solution without undesired oscillations we use the Rannacher time-stepping method [31] in our numerical experiments. This type of method is also applied in the option pricing in [10], [12], [30], for example.

The Rannacher time-stepping scheme performs a few first time steps with the implicit Euler method and after that it uses the Crank-Nicolson method. This way the scheme has good damping properties and second-order accuracy. For the semi-discrete LCP in (4.14) the Rannacher time-stepping scheme reads

$$\begin{aligned} (\mathbf{I} + \frac{1}{2} \Delta t_{[k]} \mathbf{A}) \mathbf{u}^{(k+1/2)} &\geq \mathbf{u}^{(k)} \quad \perp \quad \mathbf{u}^{(k+1/2)} \geq \boldsymbol{\psi}, & k = 0, 1/2, 1, 3/2, \\ (\mathbf{I} + \frac{1}{2} \Delta t_k \mathbf{A}) \mathbf{u}^{(k+1)} &\geq (\mathbf{I} - \frac{1}{2} \Delta t_k \mathbf{A}) \mathbf{u}^{(k)} \quad \perp \quad \mathbf{u}^{(k+1)} \geq \boldsymbol{\psi}, & k = 2, \dots, l-1, \end{aligned} \quad (4.15)$$

where the time step is $\Delta t_k = t_{k+1} - t_k$ and $\mathbf{u}^{(0)} = \boldsymbol{\psi}$. The implicit Euler method is applied with step sizes $\Delta t_0/2$ and $\Delta t_1/2$ in the first four time steps and the time-stepping is continued with the Crank-Nicolson method using step size Δt_k , $k \geq 2$. The purpose of shorter time steps with the Euler method is to improve accuracy.

The LCPs appearing in the Rannacher time-stepping (4.15) have a general form

$$\mathbf{B} \mathbf{u} \geq \mathbf{f} \quad \perp \quad \mathbf{u} \geq \boldsymbol{\psi}, \quad (4.16)$$

where we have omitted indices indicating the dependence on the time step in order to simplify the notations. The matrix \mathbf{B} has an M-matrix property due to the same property of \mathbf{A} .

5. Lagrange multiplier approach. We adopt the Lagrange multiplier approach proposed in [15] and further considered in [1] and [22] to treat the complementarity conditions. The LCP (4.16) is approximated by

$$\mathbf{B} \mathbf{u} = \mathbf{f} + \max \{ \bar{\boldsymbol{\lambda}} + c(\boldsymbol{\psi} - \mathbf{u}), \mathbf{0} \}, \quad (5.1)$$

where c is a prescribed positive constant and $\bar{\lambda}$ is a given vector with non negative components. The purpose of the term $c(\psi - \mathbf{u})$ is to force the solution towards feasible one. One of the main motivations to include the vector $\bar{\lambda}$ is that it leads to a feasible solution \mathbf{u} when it is chosen according to

$$\bar{\lambda} = \max \{ \mathbf{B}\psi - \mathbf{f}, \mathbf{0} \}. \quad (5.2)$$

The following theorem states this property. An alternative choice $\bar{\lambda} = \mathbf{0}$ leads to a penalty method. For pricing American options the penalty method has been considered in [10] and [40], for example.

THEOREM 5.1. Let \mathbf{B} be an M-matrix and let $\bar{\lambda}$ be chosen according to (5.2). Then the solution \mathbf{u} of (5.1) with any non negative c satisfies the constraint $\mathbf{u} \geq \psi$.

Proof. By subtracting $\mathbf{B}\psi$ from both side of (5.1), we obtain

$$\mathbf{B}(\mathbf{u} - \psi) = \mathbf{f} - \mathbf{B}\psi + \max \{ \bar{\lambda} + c(\psi - \mathbf{u}), \mathbf{0} \}. \quad (5.3)$$

By using the definition of $\bar{\lambda}$ in (5.2) and rearranging terms, we have

$$\mathbf{B}(\mathbf{u} - \psi) = \max \{ \max \{ \mathbf{f} - \mathbf{B}\psi, \mathbf{0} \} + c(\psi - \mathbf{u}), \mathbf{f} - \mathbf{B}\psi \}. \quad (5.4)$$

If we have $c = 0$ then this simplifies to

$$\mathbf{B}(\mathbf{u} - \psi) = \max \{ \mathbf{f} - \mathbf{B}\psi, \mathbf{0} \}. \quad (5.5)$$

Due to the non negativity of the right-hand side and the M-matrix property of \mathbf{B} , we have $\mathbf{u} - \psi \geq \mathbf{0}$ and, thus, we have proven the result for the case $c = 0$.

In the following, we assume that $c > 0$. We define $\mathbf{v} = \min\{\mathbf{u} - \psi, \mathbf{0}\}$. Then there exists a permutation matrix \mathbf{P} such that

$$\tilde{\mathbf{v}} = \mathbf{P}\mathbf{v} = \begin{pmatrix} \tilde{\mathbf{u}}_1 - \tilde{\psi}_1 \\ \mathbf{0} \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{u}}_1 - \tilde{\psi}_1 \leq \mathbf{0}, \quad (5.6)$$

where $\tilde{\mathbf{u}} = \mathbf{P}\mathbf{u}$ and $\tilde{\psi} = \mathbf{P}\psi$. Then we have

$$\begin{aligned} \mathbf{v}^T \mathbf{B}(\mathbf{u} - \psi) &= \tilde{\mathbf{v}}^T \tilde{\mathbf{B}}(\tilde{\mathbf{u}} - \tilde{\psi}) \\ &= (\tilde{\mathbf{u}}_1 - \tilde{\psi}_1)^T \tilde{\mathbf{B}}_{11}(\tilde{\mathbf{u}}_1 - \tilde{\psi}_1) + (\tilde{\mathbf{u}}_1 - \tilde{\psi}_1)^T \tilde{\mathbf{B}}_{12}(\tilde{\mathbf{u}}_2 - \tilde{\psi}_2), \end{aligned} \quad (5.7)$$

where $\tilde{\mathbf{B}} = \mathbf{P}\mathbf{B}\mathbf{P}^T$. The matrix $\tilde{\mathbf{B}}$ is an M-matrix, since it is obtained from an M-matrix by permuting the rows and columns in the same way. Due to this the quadratic term in (5.7) with $\tilde{\mathbf{B}}_{11}$ is non negative. Furthermore, $(\tilde{\mathbf{u}}_1 - \tilde{\psi}_1)^T$ and $\tilde{\mathbf{B}}_{12}$ are non positive while $\tilde{\mathbf{u}}_2 - \tilde{\psi}_2$ is non negative and, thus, their product which is the term in (5.7) with $\tilde{\mathbf{B}}_{12}$ is non negative. Thus, we have

$$\mathbf{v}^T \mathbf{B}(\mathbf{u} - \psi) \geq 0. \quad (5.8)$$

By premultiplying (5.4) with \mathbf{v}^T and using this inequality, we obtain

$$\mathbf{v}^T \max \{ \max \{ \mathbf{f} - \mathbf{B}\psi, \mathbf{0} \} + c(\psi - \mathbf{u}), \mathbf{f} - \mathbf{B}\psi \} \geq 0. \quad (5.9)$$

Let I be set of indices defined by $I = \{i \mid \mathbf{u}_i < \boldsymbol{\psi}_i\}$. Then we have

$$\begin{aligned} & \sum_i \min\{\mathbf{u}_i - \boldsymbol{\psi}_i, 0\} \max\{\max\{(\mathbf{f} - \mathbf{B}\boldsymbol{\psi})_i, 0\} + c(\boldsymbol{\psi}_i - \mathbf{u}_i), (\mathbf{f} - \mathbf{B}\boldsymbol{\psi})_i\} \\ &= \sum_{i \in I} (\mathbf{u}_i - \boldsymbol{\psi}_i) \max\{\max\{(\mathbf{f} - \mathbf{B}\boldsymbol{\psi})_i, 0\} + c(\boldsymbol{\psi}_i - \mathbf{u}_i), (\mathbf{f} - \mathbf{B}\boldsymbol{\psi})_i\} \geq 0. \end{aligned} \quad (5.10)$$

For any $i \in I$ by using $c > 0$, we obtain

$$\begin{aligned} & (\mathbf{u}_i - \boldsymbol{\psi}_i) \max\{\max\{(\mathbf{f} - \mathbf{B}\boldsymbol{\psi})_i, 0\} + c(\boldsymbol{\psi}_i - \mathbf{u}_i), (\mathbf{f} - \mathbf{B}\boldsymbol{\psi})_i\} \\ & \leq -c(\boldsymbol{\psi}_i - \mathbf{u}_i)^2 < 0. \end{aligned} \quad (5.11)$$

From (5.10) and (5.11), it follows that the set I must be empty and, thus, $\mathbf{u} \geq \boldsymbol{\psi}$. \square

The free boundary between the active set $\{(x, y) \mid u(x, y) = \psi(x)\}$ and the inactive set $\{(x, y) \mid u(x, y) > \psi(x)\}$ can be approximated by

$$\{(x, y) \mid \bar{\lambda}(x, y) + c(\psi(x) - u(x, y)) = 0\}, \quad (5.12)$$

where u and $\bar{\lambda}$ are interpolated functions defined using the grid point values given by \mathbf{u} and $\bar{\boldsymbol{\lambda}}$, respectively. The accuracy of this approximation is at most the order of $1/c$; see [22]. This suggest that with a second-order discretization of A a natural choice for c would be the order of one over grid step size squared.

The problems (5.1) are nonlinear and nonsmooth. Following [40], we employ a semismooth Newton method to solve these problems; see [15] for analysis of this case. The $(p+1)$ th iterant \mathbf{u}^{p+1} is given by

$$\mathbf{u}^{p+1} = \mathbf{u}^p + \mathbf{d}^p, \quad (5.13)$$

where the vector \mathbf{d}^p is the solution of the system of linear equations

$$\mathbf{J}(\mathbf{u}^p)\mathbf{d}^p = \mathbf{f} + \max\{\bar{\boldsymbol{\lambda}} + c(\boldsymbol{\psi} - \mathbf{u}^p), \mathbf{0}\} - \mathbf{B}\mathbf{u}^p = \mathbf{r}^p. \quad (5.14)$$

The matrix $\mathbf{J}(\mathbf{u}^p)$ in (5.14) belongs to the generalized Jacobian [7] at x^p and it is chosen to be

$$[\mathbf{J}(\mathbf{u}^p)]_{i,j} = \mathbf{B}_{i,j} + \begin{cases} c, & \text{if } i = j \text{ and } \bar{\lambda}_i + c(\boldsymbol{\psi}_i - \mathbf{u}_i^p) > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (5.15)$$

6. Numerical results. The American option pricing problem under stochastic volatility (2.7) is solved with the parameter values

$$K = 10, \quad T = 0.25, \quad r = 0.1, \quad \alpha = 5, \quad \beta = 0.16, \quad \gamma = 0.9, \quad \text{and} \quad \rho = 0.1.$$

The same values of the parameters have been used also in [8], [21], [29], and [40]. Similarly to [21] and [29], we truncate domain from right at $X_{\max} = 20$ and from up at $Y = 1$. The left truncation boundary is obtained by approximating the active set of the one-dimensional linear complementarity problem (3.22). The lower bound (3.11) would give us a truncation boundary $x = 5/3$. By solving numerically (3.22), we obtain approximately $X_{\min} = 4.25$, which we will use in the following numerical experiments.

For accuracy, it is beneficial to have a finer grid near the strike price K and for small values of volatility. Such a grid is obtained by choosing it according to

$$x_i = X_{\min} + \left(1 + \frac{\sinh(\eta(i/n - a))}{\sinh(\eta a)}\right) (K - X_{\min}), \quad i = 0, 1, \dots, n, \quad (6.1)$$

and

$$y_j = \left(\frac{b^{j/m} - 1}{b^{-1} - 1}\right) Y, \quad j = 0, 1, \dots, m, \quad (6.2)$$

where the constants a and b control the amount of refinement near the strike price K and for small y , respectively. The constant η is solved numerically from the equation $x_n = X_{\max}$. Similar grid generating functions have been considered in [34], for example. The price of the put option changes more rapidly close to the expiry T and, thus, it is more efficient to take smaller time steps near the expiry. This is accomplished by choosing the approximation times t_k according to

$$t_k = \left(\frac{d^{k/l} - 1}{d^{-1} - 1}\right) T, \quad k = 0, 1, \dots, l, \quad (6.3)$$

where d is a constant greater than one controlling the refinement. Alternatively t_k s could be chosen adaptively; see [1], [9], [10]. Based on experiments with different parameters aiming to improve the efficiency, we chose the values $a = 0.44$, $b = 40$, and $d = 64$. Figure 6.1 shows 33×17 grid constructed using the above a and b .

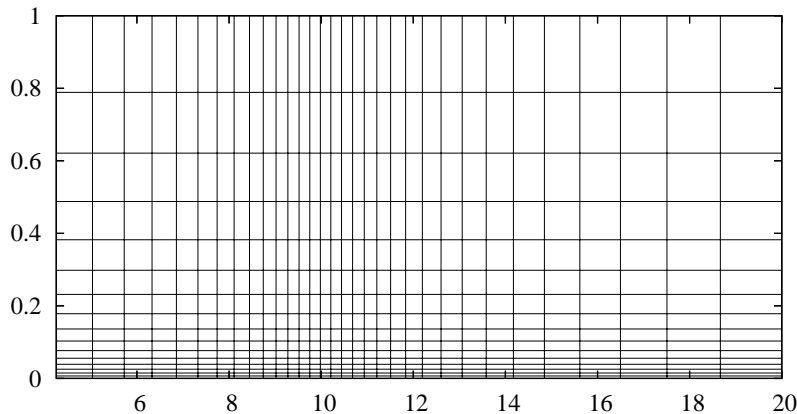


FIG. 6.1. The 33×17 refined space grid.

The following numerical experiments use seven point finite difference stencils which is obtained from the nine point stencils in Figure 4.1 by setting $w_3 = 0$ and $w_6 = 0$. Such stencils were used with good success for the operator A in [21], for example. Computationally it is beneficial to use seven point stencils instead of nine point ones. The quadratic programming problems (4.13) are smaller and, thus, the construction of the discretization is faster. Furthermore, the matrix \hat{D} in (4.13) is better conditioned which makes it easier to solve the QPs. Based on a few experiments it seems that the use of nine point stencils leads to only a minor improvement in the accuracy. In the construction of the QPs, we included first, second, and third order terms. Thus, using the notations in Section 4.1, we have $N = 6$ and $M = 9$.

Figure 6.2 shows the seven point stencils without the diagonal weight for the 257×129 grid. From this figure it can be seen that only in a small part of the computational domain some of the weights are zero which corresponds to the red color in the figure. The norm of the vectors $\mathbf{D}\mathbf{w} - \mathbf{b}$ appearing in (4.12) is shown for the same grid in Figure 6.3. This quantity estimates the truncation error of the finite difference stencil. The prices of options are sought in the middle lower part of the domain and in there the norm is less than 10^{-3} . Moving from there to right and up the norm grows due to coarser grid and larger coefficients in the operator A . In the following, we present results on the grids: 17×9 , 33×17 , 65×33 , 129×65 , 257×129 , and 513×257 . The maximum values of $\|\mathbf{D}\mathbf{w} - \mathbf{b}\|_2$ for these grids are: 24.9, 10.1, 3.00, 0.808, 0.209, and 0.0586. Thus, on finer grids the reduction factor of the norm is close to four. With a solution having bounded partial derivatives up to third-order, the reduction factor being four implies second-order accuracy.

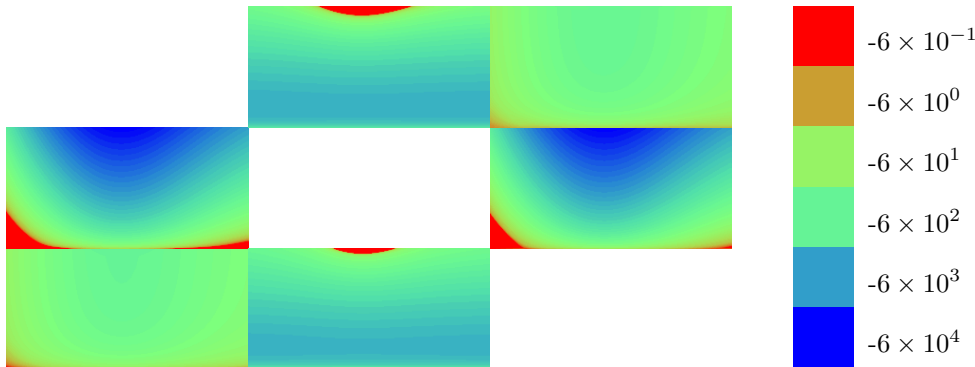


FIG. 6.2. The finite difference stencils for the 257×129 grid without the diagonal weight. Each pixel in one rectangle corresponds to a grid point and the corresponding pixels in the six rectangles gives the six optimized weights.

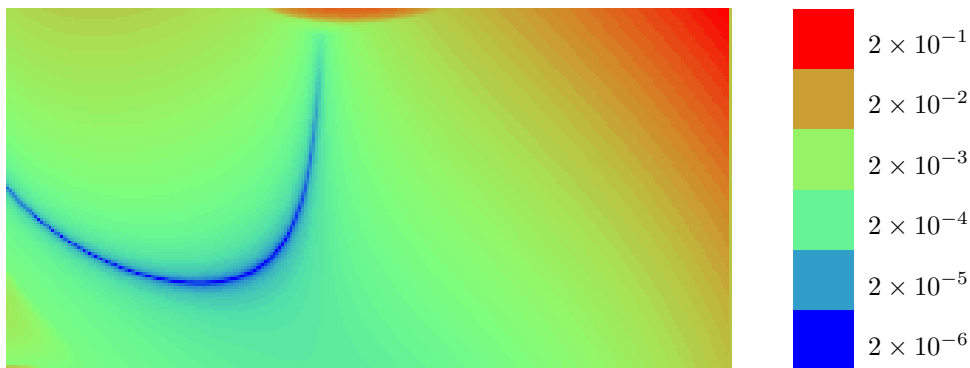


FIG. 6.3. The norm $\|\mathbf{D}\mathbf{w} - \mathbf{b}\|_2$ in (4.12) which estimates the truncation error for the 257×129 grid with each pixel being a grid point.

The parameter c in (5.1) is chosen to be $c = (n/Y)^2$. As we discussed in Section 5, this choice makes the approximation of the LCP (4.16) second-order accurate. We solve the systems of linear equations in (5.14) using a multigrid method with an ADI smoother; for details, see [21], [29]. For general description of multigrid methods

see [5], [13], [35], [37].

We start by computing reference prices using a fine grid defined by $m = 2048$, $n = 1024$, and $l = 1024$. For this computation, we have used $\bar{\lambda}$ given by (5.2) and fairly strict stopping criteria for the iterative methods. In Table 6.1, the reference prices together with three other sets of prices published in [8], [29], and [40] are reported for the variances 0.0625, 0.25, and for the asset prices 8, 9, 10, 11, 12. Based on the convergence study in the following, we conclude that for the pricing problem in the truncated domain all reference prices have at least five correct decimals. Furthermore, based on experiments not reported here with larger computational domains the error due the truncation from up and right seems to appear in the sixth decimal or later. Thus, we expect that the reference prices have five correct decimals. Based on this, we have computed the l_2 errors for the price sets and given them in Table 6.1.

TABLE 6.1

The reference prices, the prices published in scientific literature and their l_2 errors based on the reference prices.

reference & space grid	y	x					error
		8	9	10	11	12	
this paper (2049, 1025)	0.0625	2.000000	1.107621	0.520030	0.213677	0.082044	0
	0.25	2.078364	1.333632	0.795977	0.448273	0.242810	
[8] (513, 193)	0.0625	2.0000	1.1080	0.5316	0.2261	0.0907	0.02244
	0.25	2.0733	1.3290	0.7992	0.4536	0.2502	
[40] (177, 103)	0.0625	2.0000	1.1076	0.5202	0.2138	0.0821	0.00026
	0.25	2.0784	1.3337	0.7961	0.4483	0.2428	
[29] (257, 257)	0.0625	2.00	1.107	0.517	0.212	0.0815	0.00371
	0.25	2.079	1.334	0.796	0.449	0.243	

For our next experiments, we chose the stopping criterion for the iterative methods to be

$$\|\mathbf{r}\| \leq \frac{1}{10 m n} \|\mathbf{f}\|, \quad (6.4)$$

where \mathbf{r} is the residual vector and \mathbf{f} is the right-hand side vector. The termination of iterations using (6.4) leads to an additional error in the prices which is less than 5%. Table 6.2 reports the results for the six different grids and for two choices for $\bar{\lambda}$ in (5.1). The columns of the table gives the average number of Newton and multigrid iterations, the l_2 error of the prices based on the reference prices in Table 6.1, the ratios of the consecutive errors, the maximum violation of the constraint $u(\cdot, T) \geq \psi$, and CPU times in seconds on a PC with 3.8 GHz Xeon processor. The column 'init' gives the time required for forming the matrix \mathbf{A} using the QP optimization for the finite difference stencils and 'solve' gives the time required for the time stepping. For completeness and in order to make future comparisons possible, we have also give the prices for the nonzero $\bar{\lambda}$ in Table 6.3.

Based on Table 6.2, the discretization seems to be second-order accurate as the ratios of consecutive errors are about four. As could be expected the initialization CPU time required to optimize the finite difference stencil grows by the factor of four when the number of grid steps is doubled. The number of iterations increases slightly mainly due to the stopping criteria which becomes more tight on finer grids. Partly due to this the CPU time required for the solution grows about the factor of ten instead of eight when the number of steps in each direction is doubled. Due to the different scalings of the initialization and solution times, the initialization requires a

TABLE 6.2

The average number of iterations, the errors at the ten reference points, the ratios of consecutive errors, the maximum violations of constraint, and the CPU times in seconds for all methods on six grids.

$\bar{\lambda}$	grid (m, n, l)	Newton		error	ratio	violation	CPU time	
		iter.	iter.				init	solve
as in (5.2)	(16, 8, 8)	1.6	1.0	0.019027		0		
	(32, 16, 16)	1.6	1.0	0.004121	4.62	0	0.01	0.01
	(64, 32, 32)	1.5	1.0	0.000881	4.68	0	0.04	0.09
	(128, 64, 64)	1.7	1.2	0.000173	5.09	0	0.13	0.94
	(256, 128, 128)	1.7	1.3	0.000044	3.97	0	0.52	9.46
	(512, 256, 256)	1.8	1.5	0.000010	4.33	0	2.04	92.58
$\mathbf{0}$	(16, 8, 8)	1.6	1.0	0.019622		1.8×10^{-3}		
	(32, 16, 16)	1.7	1.0	0.004144	4.73	2.3×10^{-4}	0.01	0.01
	(64, 32, 32)	1.9	1.0	0.000862	4.81	3.0×10^{-5}	0.04	0.10
	(128, 64, 64)	2.0	1.1	0.000168	5.14	3.9×10^{-6}	0.14	1.04
	(256, 128, 128)	2.1	1.2	0.000042	3.97	5.0×10^{-7}	0.52	10.21
	(512, 256, 256)	2.2	1.4	0.000010	4.35	6.2×10^{-8}	2.06	99.59

considerable part of the total time only on the coarsest grids while on the finest grid it requires only a couple percent of the time.

The behavior of the solution procedure is fairly similar for both choices of $\bar{\lambda}$. The main difference is that the choice (5.2) satisfies the early exercise constraint as Theorem 5.1 states while $\bar{\lambda} = \mathbf{0}$ leads to small violations which tend to zero as grid is refined. The choice (5.2) leads to a slightly faster convergence of the Newton method while the converse is true for the multigrid method. Thus, the CPU times are essentially the same. Also, the accuracy of both choice is essentially the same under the choice $c = (n/Y)^2$ for the parameter c in (5.1).

We can compare the accuracy of the discretization proposed in this paper and the ones employed in [8], [40], and [29] using Table 6.1 and Table 6.2. The discretization in [8] is the most inaccurate. The finite difference discretization in [29] uses uniform grids without any coordinate transform. This could explain why it is less accurate than the discretization in here and in [40]. The discretization in [40] is fairly accurate. On the (177, 103) grid, the l_2 error based on the reference prices in Table 6.1 is 2.6×10^{-4} . With the optimized stencils, we obtained the error 1.7×10^{-4} on the (129, 65) grid. Thus, the optimized discretization is a couple of times more accurate which is a small difference when compared to the other discretizations. The paper [21] does not report the ten prices, but it does give an l_2 error based on different, fairly accurate reference prices. The error in [21] for the componentwise splitting method on the (161, 65) grid is 6.2×10^{-3} . Thus, the prices in [21] are more than one order of magnitude less accurate than the ones in here and [40] while their accuracy seems to be comparable with the prices in [29]. An essential differences between the optimized discretization in here and the ones in [8], [40], and [29] is that here the M-matrix property is guaranteed while the other discretization usually do not satisfy it. The discretizations in [8] and [29] are fairly easy-to-implement while the ones in here and [40] are more complicated. Thus, there is a clear tradeoff between implementation effort and accuracy.

7. Conclusions. For American put options, we proposed a truncation of a region with small asset values away from the computational domain under Heston's stochastic volatility model. This makes accurate discretization easier and reduces the number of unknowns in the resulting linear complementarity problems (LCPs).

We described optimization approach for generating finite difference stencils for the

TABLE 6.3

The prices computed using $\bar{\lambda}$ in (5.2) and their l_2 errors based on the reference prices.

grid (m, n, l)	y	x				error
		8	9	10	11	
(16, 8, 8)		2.002061	1.098723	0.511742	0.206673	0.079616
		2.078986	1.325264	0.788562	0.443066	0.241243
(32, 16, 16)		2.000036	1.105806	0.518176	0.212456	0.081646
		2.078813	1.331793	0.794061	0.447137	0.242572
(64, 32, 32)		2.000003	1.107355	0.519686	0.213503	0.081723
		2.078608	1.333307	0.795588	0.448095	0.242487
(128, 64, 64)		2.000000	1.107624	0.519987	0.213564	0.081988
		2.078387	1.333613	0.795938	0.448189	0.242759
(256, 128, 128)		2.000000	1.107626	0.520021	0.213651	0.082034
		2.078383	1.333627	0.795963	0.448252	0.242804
(512, 256, 256)		2.000000	1.107620	0.520026	0.213672	0.082043
		2.078358	1.333630	0.795974	0.448270	0.242810

fairly complicated operator. The aim of the optimization is to minimize the truncation error when the resulting matrix is required to have the M-matrix property. The numerical results showed that the arising quadratic programming problems can be solved much faster than the LCPs with the exception of very coarse grids. We proposed to use a Lagrange multiplier approach to approximate the solutions of LCPs. It can be considered as a generalization of the penalty method described by Zvan, Forsyth, and Vetzal in [40]. With the Lagrange multiplier approach it is possible to guarantee that the prices satisfy the early exercise constraint while the penalty method leads always to small violations.

The numerical results demonstrated that the proposed approach leads to option prices which are more accurate than the prices in the literature. By using a multigrid method proposed by Oosterlee in [29] to solve the resulting systems of linear equations an effective and scalable method to price American options is obtained. The results showed that accurate discretizations are essential for pricing American options under Heston's model accurately and fast. One possibility to further improve the accuracy of discretizations is to use adaptive grids; see [1], [27], for example.

Acknowledgements. We thank Dr. Samuli Ikonen for many fruitful discussions on numerical methods for option pricing.

REFERENCES

- [1] Y. ACHDOU AND O. PIRONNEAU, *Computational methods for option pricing*, vol. 30 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 2005.
- [2] F. BLACK AND M. SCHOLES, *The pricing of options and corporate liabilities*, J. Political Economy, 81 (1973), pp. 637–654.
- [3] A. BRANDT AND C. W. CRYER, *Multigrid algorithms for the solution of linear complementarity problems arising from free boundary problems*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 655–684.
- [4] M. J. BRENNAN AND E. S. SCHWARTZ, *The valuation of American put options*, J. Finance, 32 (1977), pp. 449–462.
- [5] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A multigrid tutorial*, SIAM, Philadelphia, PA, second ed., 2000.
- [6] X. CHEN, Z. NASHED, AND L. QI, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216.
- [7] F. H. CLARKE, *Optimization and nonsmooth analysis*, John Wiley & Sons Inc., New York, 1983.

- [8] N. CLARKE AND K. PARROTT, *The multigrid solution of two-factor American put options*, Tech. Rep. 96-16, Oxford Computing Laboratory, Oxford, 1996.
- [9] ———, *Multigrid for American option pricing with stochastic volatility*, Appl. Math. Finance, 6 (1999), pp. 177–195.
- [10] P. A. FORSYTH AND K. R. VETZAL, *Quadratic convergence for valuing American options using a penalty method*, SIAM J. Sci. Comput., 23 (2002), pp. 2095–2122.
- [11] J.-P. FOUQUE, G. PAPANICOLAOU, AND K. R. SIRCAR, *Derivatives in financial markets with stochastic volatility*, Cambridge University Press, Cambridge, 2000.
- [12] M. B. GILES AND R. CARTER, *Convergence analysis of Crank-Nicolson and Rannacher time-marching*, J. Comput. Finance, 9 (2006), pp. 89–112.
- [13] W. HACKBUSCH, *Multigrid methods and applications*, vol. 4 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1985.
- [14] S. HESTON, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Rev. Financial Stud., 6 (1993), pp. 327–343.
- [15] M. HINTERMÜLLER, K. ITO, AND K. KUNISCH, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.
- [16] J. HUANG AND J.-S. PANG, *Option pricing and linear complementarity*, J. Comput. Finance, 2 (1998), pp. 31–60.
- [17] J. HULL AND A. WHITE, *The pricing of options on assets with stochastic volatilities*, J. Finance, 42 (1987), pp. 281–300.
- [18] S. IKONEN AND J. TOIVANEN, *Operator splitting methods for American option pricing*, Appl. Math. Lett., 17 (2004), pp. 809–814.
- [19] ———, *Operator splitting methods for pricing American options with stochastic volatility*, Tech. Rep. B11/2004, Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland, 2004. Submitted.
- [20] ———, *Componentwise splitting methods for pricing American options under stochastic volatility*, Int. J. Theor. Appl. Finance, 10 (2007), pp. 331–361.
- [21] ———, *Efficient numerical methods for pricing American options under stochastic volatility*, Numer. Methods Partial Differential Equations, 24 (2008), pp. 104–126.
- [22] K. ITO AND K. KUNISCH, *Parabolic variational inequalities: The Lagrange multiplier approach*, J. Math. Pures Appl., 85 (2006), pp. 415–449.
- [23] ———, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, Philadelphia, 2008.
- [24] R. KANGRO AND R. NICOLAIDES, *Far field boundary conditions for Black-Scholes equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1357–1368.
- [25] B. KUMMER, *Newton's method for nondifferentiable functions*, in Advances in mathematical optimization, vol. 45 of Math. Res., Akademie-Verlag, Berlin, 1988, pp. 114–125.
- [26] Z. LI AND K. ITO, *Maximum principle preserving schemes for interface problems with discontinuous coefficients*, SIAM J. Sci. Comput., 23 (2001), pp. 339–361.
- [27] P. LÖTSTEDT, J. PERSSON, L. VON SYDOW, AND J. TYSK, *Space-time adaptive finite difference method for European multi-asset options*, Comput. Math. Appl., 53 (2006), pp. 1159–1180.
- [28] R. C. MERTON, *Theory of rational option pricing*, Bell J. Econom. and Management Sci., 4 (1973), pp. 141–183.
- [29] C. W. OOSTERLEE, *On multigrid for linear complementarity problems with application to American-style options*, Electron. Trans. Numer. Anal., 15 (2003), pp. 165–185.
- [30] D. M. POOLEY, K. VETZAL, AND P. A. FORSYTH, *Remedies for non-smooth payoffs in option pricing*, J. Comput. Finance, 6 (2003), pp. 25–40.
- [31] R. RANNACHER, *Finite element solution of diffusion problems with irregular data*, Numer. Math., 43 (1984), pp. 309–327.
- [32] C. REISINGER AND G. WITTUM, *On multigrid for anisotropic equations and variational inequalities: Pricing multi-dimensional European and American options*, Comput. Vis. Sci., 7 (2004), pp. 189–197.
- [33] E. STEIN AND J. STEIN, *Stock price distributions with stochastic volatility: An analytical approach*, Rev. Financial Stud., 4 (1991), pp. 727–752.
- [34] D. TAVELLA AND C. RANDALL, *Pricing financial instruments: The finite difference method*, John Wiley & Sons, Chichester, 2000.
- [35] U. TROTTEMBERG, C. W. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press Inc., San Diego, CA, 2001.
- [36] M. ULBRICH, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2002), pp. 805–842.
- [37] P. WESSELING, *An introduction to multigrid methods*, John Wiley & Sons Ltd., Chichester, 1992.

- [38] P. WILMOTT, *Derivatives*, John Wiley & Sons Ltd., Chichester, 1998.
- [39] G. WINDISCH, *M-matrices in numerical analysis*, vol. 115 of Teubner Texts in Mathematics, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1989.
- [40] R. ZVAN, P. A. FORSYTH, AND K. R. VETZAL, *Penalty methods for American options with stochastic volatility*, J. Comput. Appl. Math., 91 (1998), pp. 199–218.
- [41] ———, *Negative coefficients in two-factor option pricing models*, J. Comput. Finance, 7 (2003), pp. 37–73.