First proof • Typesetter: RH • March 12, 2002

A Newton method for American option pricing

Thomas F. Coleman, Yuying Li and Arun Verma

Computer Science Department and Cornell Theory Center, Cornell University, Ithaca, NY 14850.

The variational inequality formulation provides a mechanism for determining both the option value and the early exercise curve implicitly (Jaillet, Lamberton and Lapeyre, 1990). Standard finite-difference approximation typically leads to linear complementarity problems with tridiagonal coefficient matrices. However, the second-order upwind finite-difference formulation gives rise to finite-dimensional linear complementarity problems with non-tridiagonal matrices, whereas the upstream weighting finite-difference approach with the van Leer flux limiter for the convection term (Roache, 1972; Zvan, Vetzal and Forsyth, 1997) yields non-linear complementarity problems.

We propose a Newton-type interior-point method to solve discretized complementarity/variational inequality problems arising in the American option valuation. We show that, on average, the proposed method solves a discretized problem in around two to five iterations to an appropriate accuracy. More importantly, the average number of iterations required does not seem to depend on the number of discretization points in the spatial dimension; the average number of iterations actually decreases as the time discretization becomes finer.

The arbitrage condition for the fair value of an American option requires that its delta hedge factor be continuous. We investigate continuity of the delta factor approximation using the complementarity approach, the binomial method, and a simple method of taking the maximum of continuation value and the option payoff (the explicit payoff method). We show that, whereas the (implicit finite-difference) complementarity approach yields continuous delta hedge factors, both the binomial method and the explicit payoff method (with the implicit finite difference) yield discontinuous delta approximations; the early exercise curve computed using the binomial method and the explicit payoff method can be inaccurate. In addition, it is demonstrated that the delta factor computed using the Crank–Nicolson method with the complementarity approach can be oscillatory around the early exercise curve.

Presented at the first annual conference, Computational and Quantitative Finance '98, New York. Research partially supported by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the US Department of Energy under grant DE-FG02-90ER25013.A000, NSF, through grant DMS-9805602 and ONR through grant N00014-96-1-0050, and the Financial Industry Solutions Center, an SGI/Cornell joint venture in New York. The authors would like to thank Jacqueline Huang and an anonymous referee for their comments, which have improved the presentation of this paper.

1 Introduction

An American option gives the holder of the option the right, but not the obligation, to buy or sell the underlying for a fixed exercise price any time before expiry. Valuation of the American option is an optimal stopping (or a free boundary) problem; it is significantly more complex than the European option pricing. Computing the fair value of an American option requires determination of the early exercise curve. This early exercise curve divides the domain of the underlying asset price and time into a continuation region, $\mathcal{C}$, and a stopping region, $\mathcal{S}$; it is optimal to exercise in $\mathcal{S}$ but to continually hold the contract in $\mathcal{C}$. This early exercise curve is characterized by two boundary conditions. First, the option value is equal to the payoff at the early exercise curve. Second, the delta hedge factor is continuous at the early exercise curve.

American option valuation has been an active research area; many methods have been proposed to approximate American option values. Reviews of these methods can be found in, for example, Broadie and Detemple (1996, 1997). A method that is frequently used in practice is the binomial method proposed by Cox, Ross and Rubinstein (1979). Convergence of this method for pricing American options is established in Amin and Khanna (1994). Another popular method in the partial differential equation framework is to approximate the American option value by simply taking the maximum of the continuation value and the payoff; this method is referred to as the "explicit payoff" method in this paper. Convergence of this method is, however, unclear.

In addition, it is shown in Jaillet, Lamberton and Lapeyre (1990) that determination of the early exercise curve can be made implicit with a variational inequality formulation in a generalized Black–Scholes framework. The variational inequality formulation for the American option in a jump–diffusion model is analyzed in Zhang (1997). Advantages of the variational inequality approach include established convergence of computational methods in both the option values as well as the hedge factors (Jaillet, Lamberton and Lapeyre, 1990; Zhang, 1997). Brennan and Schwartz (1977) introduce a simple procedure using the standard implicit finite-difference method for the classical Black–Scholes partial differential operator. Convergence of the Brennan and Schwartz method is established by Jaillet, Lamberton and Lapeyre (1990).

Unfortunately, the Brennan and Schwartz method (1977) can be applied only when the standard finite-difference approximation – which uses central difference to approximate the first-order derivative in the spatial dimension – is used. This finite-difference approximation generates discretized linear complementarity problems with tridiagonal coefficient matrices and negative off-diagonals; the method of Brennan and Schwartz (1977) explicitly exploits this special structure.

Linear complementarity problems with pentadiagonal matrices arise when the second-order upwind finite-difference approximation is used (Huang and Pang, 1998). In addition, linear complementarity problems with non-tridiagonal coefficient matrices can arise in different asset pricing models, eg, a jump–diffusion model (Zhang, 1997). Computational investigation of American option pricing using the discretized linear complementarity has been made in Dempster and Hutton (1997, 1999), Dewynne and Wilmot (1995), and Huang and Pang (1998). In Wilmott, Dewynne and Howison (1993) the projected SOR [?] approach has been considered. In Dempster and Hutton (1999) the discretized linear complementarity problems from the standard finite-difference approximation are solved as linear programming problems by the simplex method. More sophisticated methods, such as Lemke's algorithm and interior-point method, have been used for the discretized linear complementarity problems (Huang, 1999; Huang and Pang, 1998).

Numerical consideration may require use of a more sophisticated finite-difference method – for example, when the Black–Scholes operator has little or no diffusion (Zvan, Vetzal and Forsyth, 1997). This occurs for some path-dependent exotic options. For the Asian option, the diffusion term in one of the spatial dimensions is absent in a Black–Scholes partial differential equation (Zvan, Vetzal and Forsyth, 1997). Standard finite difference using central weighting for the convection term can produce solutions with spurious oscillations (Zvan, Vetzal and Forsyth, 1997). To prevent oscillations, an upstream weighting for the convection term can be used to introduce numerical diffusion. Unfortunately this can lead to excessive numerical diffusion. To overcome this, a non-linear flux limiter with the upstream weighting can be used for the convection term (Zvan, Vetzal and Forsyth, 1997). In Section 2 we demonstrate that the upstream weighting with the van Leer limiter leads to discretized non-linear complementarity problems. The Brennan and Schwartz (1977) method or a linear programming method cannot be applied to these discretized problems.

In the partial differential equation framework, hundreds or even thousands of discretized problems need to be solved sequentially backwards in time. Efficient computational methods for solving discretized problems are crucial in pricing American options, particularly when a multi-factor model is used. In this paper we propose a Newton-type interior-point method to solve the discretized complementarity problem for American option pricing. The proposed method is applicable both to discretized linear complementarity problems and to non-linear complementarity problems; hence, the standard finite-difference approximation as well as the more complex upstream weighting with the van Leer flux limiter can be used. Our proposed method is based on the observation that these discretized problems, linear or non-linear, are closely related: the difference between the solutions of the consecutive problems decreases as the size of the time step decreases. The proposed method uses a local Newton process for fast convergence; this Newton process is then globalized using a quadratic penalty function.

In Section 4 we show that, on average, about two to five iterations are required to solve each discretized problem using the proposed method (motivated and described in Section 3. More importantly, the number of iterations required seems to be insensitive to the number of discretization points in the spatial dimension; it actually decreases as the number of discretization points in time increases. This property becomes particularly attractive for a multi-factor model or for exotic options.

The arbitrage condition for the fair value of the American option requires that the delta hedge factor be continuous. Satisfaction of this condition is important in computing an accurate delta hedge factor, the early exercise curve, and option values. We demonstrate in Section 4 that the delta hedge factor computed using the (implicit) complementarity approach is continuous, thus satisfying the continuity requirement of the early exercise curve for no arbitrage. On the other hand, the delta hedge factors computed using the popular explicit payoff method and the binomial method typically have jumps.

The stability of a numerical method for a partial differential complementarity problem is different from that of a partial differential equation. We show, in Section 4, that the Crank–Nicolson finite-difference method is typically unstable for the partial differential complementarity problem; the computed delta hedging factor can be oscillatory.

2 Discretized problems for American options

To compute an option value numerically, the associated partial differential operator is approximated through discretization and finite-difference approximation. To illustrate, we describe the discretized non-linear complementarity formulation for American option pricing when the upstream weighting finite difference with a non-linear flux limiter is used.

For simplicity, we consider the generalized Black–Scholes one-factor model; assume that the underlying price satisfies

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \sigma dW_t \tag{1}$$

where W_t is a standard Brownian motion, and $\sigma(S, t)$ denotes a deterministic local volatility function satisfying some regularity conditions so that the stochastic differential equation (1) admits a unique solution. Let $\Lambda(S)$ denote the payoff function of an American option when the underlying equals S at time t. The no-arbitrage assumption leads to the requirement that $\partial V/\partial S$ is continuous and $V(S, t) = \Lambda(S)$ at the early exercise curve (eg, Duffie, 1996).

It is established in a generalized Black–Scholes model (Jaillet, Lamberton and Lapeyre, 1990), and in a jump–diffusion model (Zhang, 1997), that necessary and sufficient conditions for the American option value $V(S, t)$ are for $V(S, t)$ to solve a partial differential complementarity/variational inequality problem. Assume that r is the risk free interest and q is the continuous dividend yield. If the volatility σ is a function of time only, and the payoff function is a convex function of S satisfying some technical conditions, then the American option value is a solution to the partial differential complementarity problem

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-q)S\frac{\partial V}{\partial S} - rV \le 0$$

$$V(S,t) - \Lambda(S) \ge 0$$

$$\big(V(S,t) - \Lambda(S)\big)\left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-q)S\frac{\partial V}{\partial S} - rV\right) = 0 \tag{2}$$

with the final condition $V(S,T) = \Lambda(S)$. Subsequently, for simplicity we omit notationally the dependence of σ on S and t. Partial differential complementarity problems for many exotic options, for example, Asian options, can be formulated similarly in this Black–Scholes partial differential complementarity framework (eg, Wilmott, Dewynne and Howison, 1993).

The payoff constraint $V(S,t) \ge \Lambda(S)$ comes directly from the no-arbitrage assumption. In contrast to the Black–Scholes equation for the European option, the inequality

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-q)S\frac{\partial V}{\partial S} - rV \le 0 \tag{3}$$

reflects the asymmetric relationship between the long and short positions of the American contract; only the holder has the early exercise option.

If we let $t^* = T - t$, the partial differential inequality in (3) becomes $\mathcal{L}_{\mathrm{BS}}[V] \ge 0$, where

$$\mathcal{L}_{\mathrm{BS}}[V] \stackrel{\mathrm{def}}{=} \frac{\partial V}{\partial t^*} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - (r-q)S\frac{\partial V}{\partial S} + rV$$

Let $\{t_i^*\}_{i=0}^M$ denote the discretization in the time interval $[0,T]$ and $\{S_i\}_{i=0}^N$ denote the discretization in the spatial dimension, where S_N is sufficiently large. Let $\delta S_i = (S_{i+1} - S_{i-1})/2$ and $\delta S_{i+1/2} = S_{i+1} - S_i$. Using the finite-volume approach (Roache, 1972) and adopting the notation in Zvan, Vetzal and Forsyth (1997), a finite-difference approximation to the Black–Scholes partial differential operator can be described by

$$\begin{aligned}\mathcal{L}_{\mathrm{BS}}[V] \approx \frac{V_i^{j+1} - V_i^j}{\delta t^*} \quad &- \theta\Big(H_{i-1/2}^{j+1} - H_{i+1/2}^{j+1} + f_i^{j+1}\Big) \\ &- (1-\theta)\Big(H_{i-1/2}^{j} - H_{i+1/2}^{j} + f_i^{j}\Big)\end{aligned} \tag{4}$$

where $i = 1,\cdots,N-1$, $j = 0,\cdots,M-1$, and $0 \le \theta \le 1$ denotes the temporal weighting. In addition,

$$H_{i-1/2}^{j+1} = \frac{1}{\delta S_i}\left[\left(-\frac{1}{2}\sigma_i^{j\,2} S_i^2\right)\frac{V_i^{j+1} - V_{i-1}^{j+1}}{\delta S_{i-1/2}} - (r-q)S_i V_{i-1/2}^{j+1}\right]$$

$$H_{i+1/2}^{j+1} = \frac{1}{\delta S_i}\left[\left(-\frac{1}{2}\sigma_i^{j\,2} S_i^2\right)\frac{V_{i+1}^{j+1} - V_i^{j+1}}{\delta S_{i+1/2}} - (r-q)S_i V_{i+1/2}^{j+1}\right]$$

and

$$f_i^{j+1} = (-r)V_i^{j+1}$$

The approximation scheme is fully implicit when $\theta = 1$, and explicit when $\theta = 0$. The Crank–Nicolson method is obtained when $\theta = 1/2$. The central weighting scheme for the convection term corresponds to

$$V_{i+1/2}^{j+1} = \frac{V_{i+1}^{j+1} - V_i^{j+1}}{2} \tag{5}$$

If the grid spacing is uniform, this approximation has second-order accuracy and the above finite difference approximates the Black–Scholes operator by a linear function. For notational simplicity, we sometimes suppress the dependence on the time index j and denote $x = V - \Lambda$. Let $Bx + c$ be the linear function from the finite-difference approximation (4) to the Black–Scholes operator. Introducing an auxiliary variable y: $y = Bx + c$, the discretized problem (2) can be written as

$$\begin{cases} G(x,y) = Bx + c - y = 0 \\ Xy = 0 \\ x \geq 0, \quad y \geq 0 \end{cases} \tag{6}$$

where $X \stackrel{\text{def}}{=} \operatorname{diag}(x)$, and $G(x,y): \Re^{2(N-1)} \to \Re^{N-1}$ is a linear function. Problem (6) is a linear complementarity problem.

When the Black–Scholes partial differential inequality in (2) has little or no diffusion, the standard central weighting (5) for the convection term may produce oscillatory solutions. In addition, the standard finite-difference methods may be inaccurate due to excessive numerical diffusion (Barraquand and Pudet, 1996; Zvan, Vetzal and Forsyth, 1997). To overcome this, the following upstream weighting with the van Leer flux limiter has been suggested (Zvan, Vetzal and Forsyth, 1997):

$$V_{i+1/2}^{j+1} = V_{\text{up}}^{j+1} + \frac{\phi(q_{i+1/2})}{2}\left(V_{\text{down}}^{j+1} - V_{\text{up}}^{j+1}\right) \tag{7}$$

where

$$\text{up} = \begin{cases} i & \text{if } \; -(r-q)S_i \geq 0 \\ i+1 & \text{otherwise} \end{cases}$$

$$\text{down} = \begin{cases} i+1 & \text{if } \; -(r-q)S_i \geq 0 \\ i & \text{otherwise} \end{cases}$$

and $\phi \in [0, 2]$ is the weight which is a non-linear function of the option values:

$$\phi\left(q_{i+1/2}^{j+1}\right) = \frac{\left|q_{i+1/2}^{j+1}\right| + q_{i+1/2}^{j+1}}{1 + \left|q_{i+1/2}^{j+1}\right|}$$

$$q_{i+1/2}^{j+1} = \frac{V_{\text{up}}^{j+1} - V_{\text{2up}}^{j+1}}{S_{\text{2up}} - S_{\text{up}}} \bigg/ \frac{V_{\text{down}}^{j+1} - V_{\text{up}}^{j+1}}{S_{\text{up}} - S_{\text{down}}} \tag{8}$$

When $V_{\text{down}}^{j+1} = V_{\text{up}}^{j+1}$, the finite-difference approximation does not depend on the corresponding components of q and they can be eliminated. Note that when the weight $\phi \equiv 1$, the scheme becomes the central weighting scheme.

Using the van Leer flux limiter, the finite-difference approximation to the partial differential operator in (2) gives rise to non-linear functions. In addition, these non-linear functions are piecewise smooth due to presence of the absolute values in the weight. Standard mathematical programming methods do not apply directly to piecewise continuous functions. To deal with this non-differentiability, let $\bar{q}_i^{j+1} \geq 0$ denote the negative part of $q_{i-1/2}^{j+1}$ and $\hat{q}_i^{j+1}$ denote the positive part of the $q_{i-1/2}^{j+1}$, $i = 1, \cdots, N$. It is clear that

$$\left|q_{i-1/2}^{j+1}\right| = \hat{q}_i^{j+1} + \bar{q}_i^{j+1} \quad \text{and} \quad q_{i-1/2}^{j+1} = \hat{q}_i^{j+1} - \bar{q}_i^{j+1}$$

Hence the weight function becomes

$$\phi\left(q_{i+1/2}^{j+1}\right) = \frac{2\hat{q}_{i+1}^{j+1}}{1 + \hat{q}_{i+1}^{j+1} + \bar{q}_{i+1}^{j+1}}$$

and $\hat{G}$ denote the non-linear function using the upstream weighting with the van Leer limiter for the convection term – ie, for $i = 1, \cdots, N-1$, $\hat{G}$ denotes?

$$\hat{G}_i = \frac{V_i^{j+1} - V_i^j}{\delta t^*} - \left[\theta\left(H_{i-1/2}^{j+1} - H_{i+1/2}^{j+1} + f_i^{j+1}\right) + (1-\theta)\left(H_{i-1/2}^{j} - H_{i+1/2}^{j} + f_i^{j}\right)\right] \tag{9}$$

Let $\bar{G}$ be the following functions arising from the definition of the q:

$$\bar{G}_{i+1} = \frac{V_{\text{up}}^{j+1} - V_{\text{2up}}^{j+1}}{S_{\text{2up}} - S_{\text{up}}} - \frac{V_{\text{down}}^{j+1} - V_{\text{up}}^{j+1}}{S_{\text{up}} - S_{\text{down}}}\left(\hat{q}_{i+1}^{j+1} - \bar{q}_{i+1}^{j+1}\right), \quad i = 0, \cdots, N-1 \tag{10}$$

Let $G: \Re^{4N-2} \to \Re^{2N-1}$ denote

$$G(V^{j+1}, \hat{q}^{j+1}, w^{j+1}, \bar{q}^{j+1}) \stackrel{\text{def}}{=} (\hat{G} - w^{j+1}, \bar{G})$$

$x \stackrel{\text{def}}{=} (V - \Lambda, \hat{q}) \in \Re^{2N-1}$, and $y \stackrel{\text{def}}{=} (w, \bar{q}) \in \Re^{2N-1}$. The discretized American option problem, when the the upstream weighting with the van Leer flux limiter is used, can be formulated as the non-linear programming problem

$$\begin{cases} G(x,y) = 0 \\ Xy = 0 \\ x \geq 0, \quad y \geq 0 \end{cases} \tag{11}$$

where $G\colon \Re^{2n} \to \Re^n$, $x \in \Re^n$, and $y \in \Re^n$ with $n = 2N - 1$. Note that the formulation (6) is a special case of the formulation (11) where $G(x, y)$ is linear.

When the upstream weighting with the van Leer flux limiter is used for the convection term, it is easy to verify that a European option value can be computed by solving

$$\begin{cases} G(x,y,z) = 0 \\ Xy = 0 \\ x \geq 0, \quad y \geq 0 \end{cases} \tag{12}$$

with $G\colon \Re^{2n+m} \to \Re^m$, $x \in \Re^n$, $y \in \Re^n$, and $z \in \Re^m$ for some positive integers n and m.

Problem (11) can be written as a non-linear complementarity problem when $V_{\text{down}} \neq V_{\text{up}}$, whereas problem (12) is not a standard non-linear complementarity problem. Problem (11) has $2n$ non-linear equality constraints. The simple bound constraints adds additional complexity. One may be tempted to compute the solution to (11) by solving the optimization problem

$$\min_{x,y} \| G(x,y) \|^2 + \| Xy \|^2 \quad \text{subject to } \; x \geq 0, \; y \geq 0 \tag{13}$$

using a standard optimization software. However, a solution of (13) may not be a solution to (11) because (13) may have more than one local minimizer. Moreover, since optimization problem (13) is degenerate at the solution of the discretized problem (11), even when problem (11) itself is not, this approach may encounter some numerical difficulty.

Next we propose a Newton-type interior-point algorithm that solves (11) directly; this algorithm can be easily modified to solve (12). In Section 4 we examine the continuity of the delta hedge factors using the complementarity approach, the binomial method, and the explicit payoff method. In addition, we illustrate the efficiency of our proposed algorithm for solving (11). Those not interested in the details of the computational method can go directly to Section 4 for the computational results.

3 A Newton-type interior-point method

A general linear complementarity problem is a NP-hard problem (Jaillet, Lamberton and Lapeyre, 1990). However, we observe that in the option pricing setting each discretized problem at t^*_{i+1} is endowed with a good approximation to its solution, ie, the solution to the problem at t^*_i. Moreover, this approximation becomes more accurate as the time-discretization parameter δt^* is reduced. This is the situation in which a Newton process brings fast convergence. In addition, the non-linear programming problem (11) is highly structured. We propose a Newton method that utilizes this initial point as well as the special structure of the discretized problem (11).

First, when $x \geq 0$ and $y \geq 0$, the inner product $x^T y \geq 0$ measures satisfaction of the constraints $Xy = 0$. Hence we consider the following auxiliary optimization problem:

$$\min_{x,y} \frac{1}{2} \| G(x,y) \|^2 + x^T y \qquad \text{subject to } \ x \geq 0, \ \ y \geq 0 \tag{14}$$

The objective function measures satisfaction of optimality of the original problem (11); it is simpler than that of (13) because it has the quadratic $x^T y$ term replacing the quartic $\|Xy\|^2$. The solution to the complementarity problem (11) is clearly always a global minimizer of (14). However, a local minimizer of (14) may not be a global minimizer and thus may not be a solution to (11). The possibility of computing a local minimizer which is not a solution can be alleviated through use of an additional penalty parameter; consider

$$\min_{x,y} f(x,y) \stackrel{\text{def}}{=} \frac{\rho}{2} \| G(x,y) \|^2 + x^T y \qquad \text{subject to } \ x \geq 0, \ \ y \geq 0 \tag{15}$$

where $\rho > 0$ is a penalty parameter. The use of this penalty parameter is motivated below.

The gradient of $f(x, y)$ is

$$\nabla f(x,y) = \begin{bmatrix} \rho \nabla G_x^T G + y \\ \rho \nabla G_y^T G + x \end{bmatrix}$$

If the Kuhn–Tucker condition of the minimization problem (15) is satisfied at a feasible point, then

$$\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} \rho \nabla G_x^T G + y \\ \rho \nabla G_y^T G + x \end{pmatrix} = 0 \tag{16}$$

where $Y \stackrel{\text{def}}{=} \text{diag}(Y)$. Hence, a local minimizer of (15) satisfying $G(x, y) = 0$ is a global minimizer of (15) since (16) implies that $Xy = 0$. Therefore a large penalty parameter $\rho > 0$ makes it more likely that a local minimizer of (15) will satisfy $G(x, y) = 0$.

Using an algorithm which monotonically reduces the objective function measuring optimality of the original problem (11) further alleviates the potential global versus local minimizer problem. Recall that the discretized problem (11) is always endowed with a good starting point. A monotonically decreasing optimization algorithm restricts the possible local minimizer that it computes; starting from a neighborhood of the global minimizer and insisting that the objective function values decrease at each iteration further increases the likelihood of the iterates converging to the solution. Indeed, if the solution (x^*, y^*) of the discretized problem (11) is the only point in the intersection of the local minimizers of (15) and the level set $\mathcal{L} = \{x : f(x, y) \le f(x^{(0)}, y^{(0)})\}$, then solving the minimization problem produces a solution to the discretized problem (11). The proposed algorithm below monotonically reduces the objective function $f(x, y)$; the problematic issue of a local minimizer of (15) that is not a global minimizer is not a practical concern in our experience.

We now consider a direct local Newton process for the discretized problem (11). Let

$$F(x, y) \stackrel{\text{def}}{=} \left[G(x, y); Xy\right]$$

$X \stackrel{\text{def}}{=} \operatorname{diag}(x)$ and $Y \stackrel{\text{def}}{=} \operatorname{diag}(y)$. The solution (x^*, y^*) to the discretized complementarity problem (11) can be considered as a feasibility problem: the solution satisfies the bound constraints $(x^*, y^*) \in \mathcal{F} = \{(x, y) : x \ge 0, y \ge 0\}$ and the system of non-linear equations $F(x, y) = 0$.

A Newton step $d = [d_x; d_y]$ for $F(x, y) = 0$ satisfies

$$\nabla F^T d = -[G; Xy] \tag{17}$$

where ∇F is the Jacobian matrix of F, ie,

$$\nabla F^T = \begin{bmatrix} \nabla G_x^T & \nabla G_y^T \\ Y & X \end{bmatrix}$$

The gradient of the objective function $f(x, y)$ can be written as

$$\nabla f = \begin{bmatrix} \rho \nabla G_x^T G + y \\ \rho \nabla G_y^T G + x \end{bmatrix} = \nabla F [\rho G; e_n]$$

If (x, y) is not the solution of the discretized problem (11), the Newton step d defined by (17) is clearly a descent direction for $f(x, y)$ since

$$\nabla f^T d = -\left[\rho G; e_n\right]^T \nabla F^T \nabla F^{-T} [G; Xy] = -\left(\rho \|G\|^2 + x^T y\right) < 0$$

Hence the Newton step (17) is compatible with the minimization problem (15). If the functions $F(x, y)$ were a general non-linear function, the Newton step from $F(x, y) = 0$ could conflict with the inequality constraints $x \ge 0$ and $y \ge 0$; a Newton step truncated to satisfy the non-negativity constraints may not be able to produce sufficient progress. Fortunately, since the second half of the non-linear constraints $Xy = 0$ is bilinear and intimately related to the non-negativity constraints, the non-negativity constraints $x \ge 0$ and $y \ge 0$ do not prevent a sufficiently large portion of the Newton step being taken as long as the components of x and y are not zero simultaneously. This can be seen from the equations

$$Yd_x = -X(y + d_y) \quad \text{and} \quad Xd_y = -Y(x + d_x)$$

If x_i and y_i are not zero simultaneously, the step size to the bound constraints along d_x eventually converges to unity, and fast convergence of the Newton process occurs.

The Newton process described thus far is local: the iterates may not converge to a local minimizer unless the initial point is sufficiently close. To resolve the potential difficulty of the components of x and y being (near) zero simultaneously, and to globalize the local Newton process, we combine the Newton step with the use of the affine-scaling steepest descent direction used in the trust region and affine-scaling method (Coleman and Li, 1996a). The scaled steepest descent direction is used in Coleman and Li (1996a) to obtain global convergence to a minimizer of (15); it is simply $-D^2\nabla f$, with $D = \text{diag}(h)$; h is defined below.

Definition 3.1 *The vector $h \in \Re^{2n}$ is defined: for each component $1 \le i \le 2n$:*
(i) *If $(\nabla f_x)_i \ge 0$, then $h_i \stackrel{def}{=} x_i$,*
(ii) *If $(\nabla f_y)_i \ge 0$, then $h_{i+n} \stackrel{def}{=} y_i$,*
(ii) *If $(\nabla f_x)_i < 0$, then $h_i \stackrel{def}{=} -1$,*
(iii) *If $(\nabla f_y)_i < 0$, then $h_{i+n} \stackrel{def}{=} -1$.*

To guarantee convergence, each new iterate needs to yield a sufficiently large decrease for the objective function $f(x, y)$. For a discretized option pricing problem (11), a Newton step typically results in a sufficient decrease because it has a good starting point. Since computational efficiency is crucial in pricing the American option, we want to fully exploit this fact. Although sophisticated globalization techniques, such as the trust region method (Coleman and Li, 1996a) can be used, we choose here a simple line search technique to avoid as much unnecessary computation as possible. A full description of the line search method is tedious and unenlightening; we only point out that the projection below is used to ensure strict feasibility when taking a Newton step $d^{(k)}$:

$$\left(x^{(k+1)}, y^{(k+1)}\right) = \max\left((x^{(k)}, y^{(k)}) + d^{(k)}, \varepsilon\right)$$

where $\varepsilon > 0$ is a small parameter about machine precision. The iteration proceeds to the next if it yields a sufficient reduction of the objective function value; for example, if the new iterate $(x^{(k+1)},y^{(k+1)})$ along the Newton direction satisfies

$$f\left(x^{(k+1)}, y^{(k+1)}\right) < f\left(x^{(k)}, y^{(k)}\right) - \gamma_0 f\left(x^{(k)}, y^{(k)}\right)$$

for some $0 < \gamma_0 < 1$, the decrease is sufficient. If the projected Newton step yields a sufficient decrease, the algorithm proceeds to the next iteration.

Otherwise, a subsequent line search is performed. A scaled steepest descent direction defined in Definition 3.1 for (15) is computed if sufficient decrease cannot be obtained along the Newton direction. Strict feasibility $(x, y) > 0$ is maintained at each iteration. For more details of the scaled steepest descent and the line search, see Coleman and Li (1996a, 1996b).

The proposed algorithm can be summarized as follows.

Initialization Let $\rho > 0$ and $(x^{(0)}, y^{(0)}) > 0$.

Step 1 Compute a Newton step $d^{(k)}$ satisfying

$$\nabla F^{(k)^T} d = -\left[G^{(k)}; X^{(k)} y^{(k)}\right]$$

Step 2 Perform the line search along the Newton direction to compute $(x^{(k+1)}, y^{(k+1)}) > 0$. If

$$f\left(x^{(k)}, y^{(k)}\right) - f\left(x^{(k+1)}, y^{(k+1)}\right) \leq \gamma_1 f(x^{(k)}, y^{(k)})$$

go to Step 1. Otherwise, continue to Step 3.

Step 3 Compute the scaled steepest descent direction; Perform the line search along the scaled steepest descent direction to compute $(x^{(k+1)}, y^{(k+1)}) > 0$ which satisfies sufficient decrease conditions.

Typically, the Newton step leads to sufficient decrease and the cost per iteration is roughly the cost of computing one Newton step, assuming that a line search is implemented efficiently. We discuss below how a Newton step can be computed to exploit the sparsity structure of the discretized problem (11).

First, we examine the case when the discretized problem (11) is a linear complementarity problem. In this case, $\nabla G_x = B$ and $\nabla G_y = -I$. Instead of solving the $2n$-by-$2n$ system (17), the Newton step can be computed by solving an n-by-n linear system:

$$(XB + Y)\, d_x = -X(G + Y) \tag{18}$$

and $d_y = G + Bd_x$. Notice that, typically, B is asymmetric. If finite differencing with central weighting is used, system (17) is tridiagonal and this band sparsity structure can be exploited for efficiency. Moreover, $XB + Y$ has the same sparsity structure as B. In addition, the matrix $XB + Y$ is diagonally dominant if B is diagonally dominant when $x, y > 0$. Thus, a linear system with the coefficient matrix $XB + Y$ can typically be solved as efficiently as the system $Bx = c$.

In addition, the equation $Xy = 0$ is bilinear. The Newton step can be further improved by computing a correction step $(\hat{d}_x, \hat{d}_y)$ satisfying

$$(x + \hat{d}_x)(y + \hat{d}_y) = 0$$

This can be approximately achieved by

$$(XB + Y)\hat{d}_x = -X(G + y) - \mathrm{diag}(d_x)\, d_y \tag{19}$$

and $\hat{d}_y = G + B\hat{d}_x$. Notice that the coefficient matrices of (17) and (19) are exactly the same; the LU factorization obtained when computing a Newton step can be used again to compute the correction step.

For the Newton step $\hat{d}$ with correction, we have

$$\begin{aligned}\nabla f^T \hat{d} &= -\left[\rho G; e_n\right]^T \nabla F^T \nabla F^{-T} \left[G; Xy + \mathrm{diag}(d_x)\, d_y\right] \\ &= -\left(\rho \|G\|^2 + x^T y\right) - d_x^T d_y\end{aligned}$$

Hence $(\hat{d}_x, \hat{d}_y)$ is a descent direction for $f(x, y)$ if

$$d_x^T d_y > -\left(\rho \|G\|^2 + x^T y\right)$$

In addition, when $G(x, y) = Bx + c - y$, $f(x, y)$ is a quadratic function and the first exact local minimizer of f along a descent direction $d^{(k)}$ in the feasible region $\mathcal{F} = \{x: x \geq 0\}$ can be computed easily when performing a line search. Therefore, the corrected Newton step and the exact line search, if desired, can be used when the finite difference leads to a linear complementarity problem.

Now we consider the case when $G(x, y)$ is non-linear in the discretized problem (11); this happens when the upstream weighting with the non-linear van Leer flux limiter is used. Here $-\nabla G_y$ does not equal to the identity matrix. In this case, the coefficient matrix ∇F for the Newton equation (17)

$$\nabla F^T = \begin{bmatrix} \nabla G_x^T & \nabla G_y^T \\ Y & X \end{bmatrix}$$

is still typically an asymmetric sparse matrix. The matrices ∇G_x and ∇G_y typically have band structures. A Newton step, in this case, can be computed using a sparse LU factorization. To exploit sparsity, it is important to use a column ordering to reduce fill-in. Moreover, for computational efficiency it is essential that the column ordering is only computed once at the first time step $t^* = \delta t^*$; the computed ordering can be used for subsequent time steps since the sparsity structure remains the same.

Under the non-degenerate assumptions, analysis similar to that in Coleman and Li (1994) can be used to show that the Newton process is locally quadratically convergent. A discretized problem (11) for American option pricing is typically near-degenerate in the sense that there are components with x_i and y_i simultaneously near zero. We illustrate computationally in Section 4 that this does not prohibit fast local convergence of the Newton process that is derived directly from the discretized problem (11).

The Newton process proposed here is related to the one used in the primal and dual interior-point methods for linear programming (Kojima, Mizuno and Yoshise, 1989). However, no barrier parameter is embedded in our method. This allows us to exploit fully the given initial approximation. The potential difficulty of approaching the boundary prematurely is dealt with using the scaled steepest descent direction for globalization.

4 Ensuring continuity of delta

The discretized American option problem (11) includes a system of n equations $G(x, y) = 0$, complementarity conditions $x_i y_i = 0$, $i = 1, \cdots, n$, and non-negativity constraints on the variables x and y. Depending on the finite-difference method, the discretized problem can be a linear complementarity problem with a tridiagonal structure, a linear complementarity problem with a more general band structure, or a genuine non-linear complementarity problem.

An efficient computational method for solving the discretized problems (11) is crucial for American option pricing since (11) needs to be solved at each time step. If a discretized linear complementarity problem has a special tridiagonal coefficient matrix, the Brennan–Schwartz method (1977) can be used. In addition, two popular methods are the binomial method (Cox, Ross and Rubinstein, 1979) and the explicit payoff method (eg, Hull, 1997; Duffie, 1996); the latter approximates the American option value by taking the maximum of the continuation value and the payoff at each time step. For example, when the finite difference with central weighting (4) and (5) is used (which leads to linear complementarity problems), the explicit payoff method approximates the American option value:

$$\begin{array}{ll} \text{for} & j = 1, 2, \cdots, M \\ & B^j x + c^j = 0 \\ & V^j = \max(x(1:n), \Lambda) \\ \text{end} & \end{array} \tag{20}$$

where $Bx + c$ is the finite-difference approximation ((4), (5)) to the Black–Scholes operator.

When the upstream weighting with the van Leer flux limiter ((4), (7), (8)) is used, the continuation value at each time step can be computed by solving

$$\begin{cases} \hat{G}(V, \hat{q}, \bar{q}) = 0 \\ \bar{G}(V, \hat{q}, \bar{q}) = 0 \\ \hat{q}_i \bar{q}_i = 0, \ i = 1, \cdots, N \\ \hat{q} \geq 0, \ \bar{q} \geq 0 \end{cases} \tag{21}$$

where $\hat{G}$ and $\bar{G}$ are defined by (9) and (10), respectively. The explicit payoff method is used in Zvan, Vetzal and Forsyth (1997).

One difficulty with the explicit payoff method is that its convergence properties are unclear. Although the computed option value is known to converge using the binomial method, we do not know of any convergence result of the hedge factors computed from finite-difference approximation. On the other hand, convergence in both the option values and the hedge factors computed using the complementarity approach has been established (Jaillet, Lamberton and Lapeyre, 1990; Zhang, 1997).

Unlike the Brennan and Schwartz method (1977), the proposed algorithm does not have restriction on the type of the finite-difference method used; it is applicable to the discretized problem from the standard finite difference as well as more sophisticated approximation schemes, such as the upstream weighting with the van Leer flux limiter ((4), (7), (8)). Using the vanilla American option in a generalized Black–Scholes model as an example, we show that:

- The (implicit finite-difference) complementarity approach for American option pricing produces continuous delta hedge factors, whereas the popular binomial method and the explicit payoff method yield discontinuous delta factors.
- The Crank–Nicolson method, while stable for solving partial differential equations, can be unstable for the partial differential complementarity problem. The delta hedge factor computed using the Crank–Nicolson method with the complementarity approach is, typically, oscillatory.
- The proposed algorithm for solving the discretized problem (11) is computationally efficient; typically, it solves the discretized problem in about two to five iterations to an appropriate accuracy.

For the discussion that follows we consider two types of finite-difference approximation. First, the central weighting (4) and (5) for the convection term is used; the discretized problem (11) is a linear complementarity problem. Second, the upstream weighting scheme with the van Leer non-linear flux limiter ((4), (7), (8)) is used; the discretized problem (11) is a non-linear complementarity problem. In our experiments the finite region $[0, T] \times [0, 2S_0]$ is used to approximate the region $[0, T] \times [0, +\infty)$. The finite region $[0, T] \times [0, 2S_0]$ is then discretized by a uniform grid $\{(t_i^*, S_j)\}$, $i = 0, \cdots, M$ and $j = 0, \cdots, N$. Unless explicitly stated otherwise, we use the implicit finite-difference scheme, ie, $\theta = 1$ in (4). The number of steps in the state dimension, shown in the first column [of?], is chosen so that $\delta S = \sqrt{\delta t^*}$; hence the discretization error is $O(\delta t^*)$. Query.

The starting point is known to be crucial for an interior-point method. Although the solution for the discretized problem at t_i^* offers a good approximation to the solution for the discretized problem at t_{i+1}^*, this point is typically at the boundary of the feasible region $\mathcal{F} = \{(x, y) : x \geq 0, y \geq 0\}$. We introduce a small perturbation. When the discretized problem is a linear complementarity problem we set the starting point $(x^{(0)}, y^{(0)}) = \max((x, y), \min(0.001, \delta t^*))$ at t_{i+1}^*, where (x, y) is the solution at t_i^*. When the upstream weighting with the van Leer flux limiter is used, we also would like to start from a point such that

$x_i \neq x_{i+1}$, $i = 1, \cdots, N$, since the Newton step is not defined at such a point; a small random perturbation is then used to avoid $x_i = x_{i+1}$. The stopping criterion for the optimization algorithm is

$$\max\left({x^{(k)}}^T y^{(k)},\ \left\| G\left(x^{(k)}, x^{(k)}\right) \right\|_2 \right) < \epsilon_{\text{opt}}$$

where $\epsilon_{\text{opt}} > 0$ is an error tolerance. Since the computed $(x^{(k)}, y^{(k)})$ is strictly positive, the discretized option pricing problem (11) is solved with at least ϵ_{opt} accuracy.

4.1 Continuity of delta hedge factors

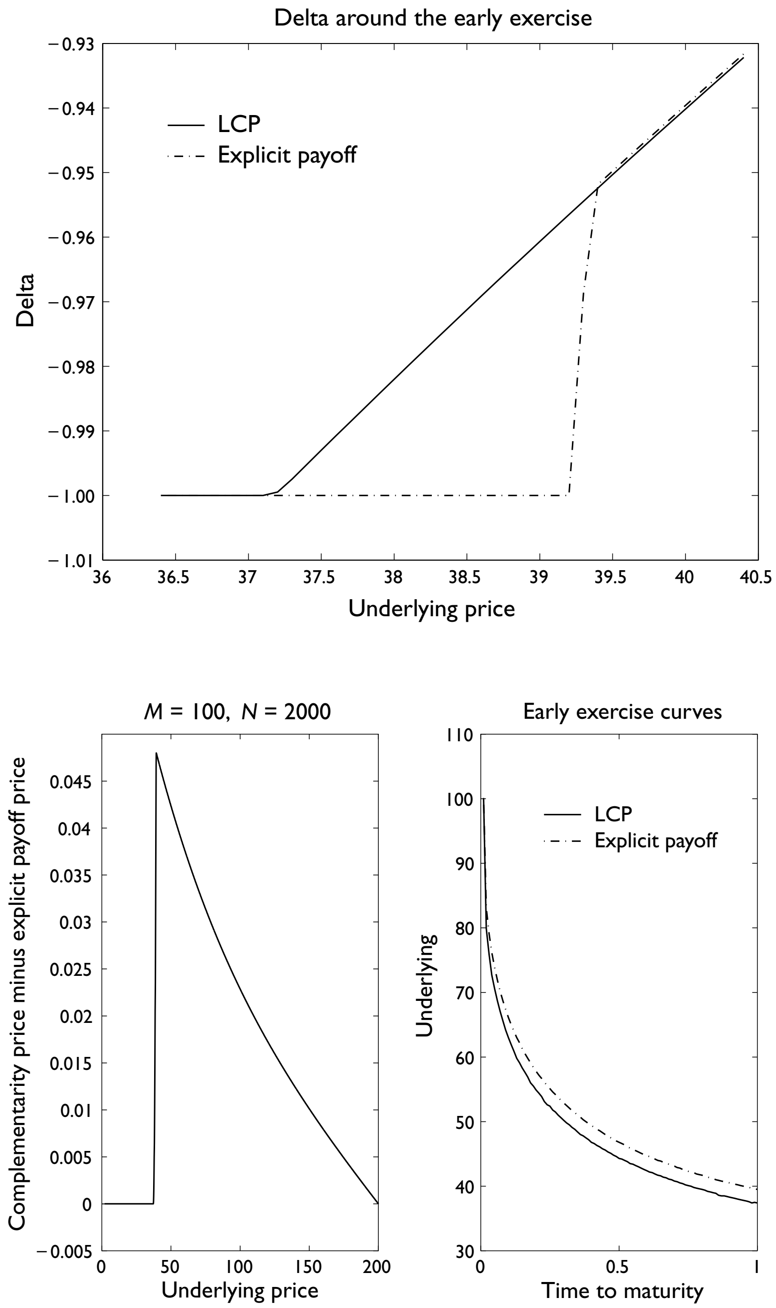
Continuity of the delta hedge factor is an arbitrage condition for the American option value. First we compare the computed delta hedge factors using the complementarity approach, the explicit payoff method, and the binomial method. The delta hedge factors are computed using finite difference; the delta factor at time $t = 0$, for example, is approximated as

$$\left[\frac{\partial V}{\partial S}\right]_{t=0,\, S=S_i} \approx \frac{V_{i+1}^M - V_i^M}{\delta S}$$

Figure 1 shows the delta factors computed using the complementarity method and the explicit payoff method (20); the implicit finite difference with central weighting (4), (5) is used in these examples (hence the discretized problem is a linear complementarity problem). The top plot displays the delta factors of a specified put option, in the neighborhood of the early exercise curve, computed using the complementarity approach and the explicit payoff method. It is clear that the hedge factor given by the complementarity approach is continuous, while the hedge factor from the explicit payoff method has a jump near the early exercise curve. The plot at lower left graphs the price differences using the two methods; a relatively large difference can be observed around the early exercise curve. Notice that, although the difference in delta is localized, the price difference is propagated through a wide range of underlying values. The lower right plot presents the early exercise curves computed using the complementarity approach and the explicit payoff method; the early exercise curve for a put option is computed by locating, at each time step, the smallest underlying value S_i such that $|V(S_{i,t}) - \Lambda(S_i)| > 10^{-8}$.

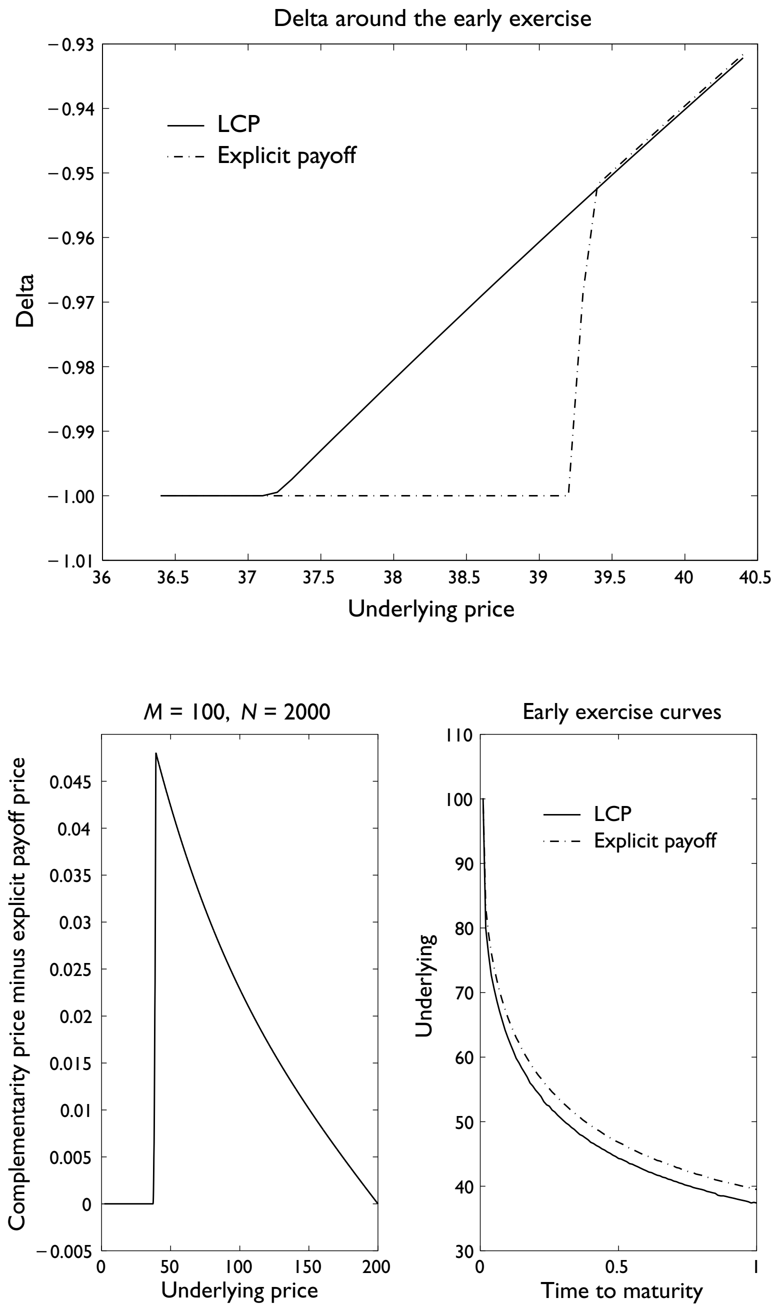
When the upstream weighting with the van Leer flux limiter is used, similar behavior is observed for the complementarity approach and the explicit payoff method (21). Figure 2 compares the delta factors computed using the complementarity approach and the explicit payoff method in this case. We note that the proposed method is easily adapted to compute a solution of the discretized problem (21) for the explicit payoff method (hence it can also be used to solve the discretized non-linear programming problem (12) for European options).

FIGURE 1 LCP comparisons



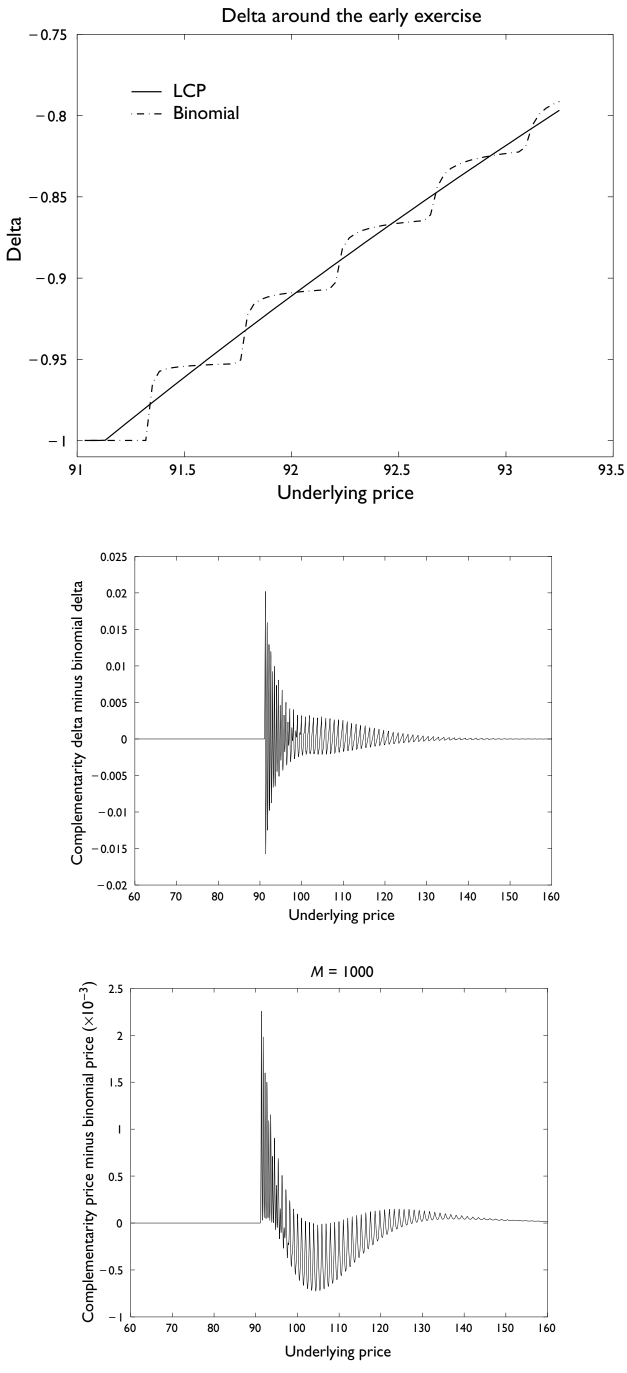
Conditions: $S_0 = 100$, $K = 100$, $T = 1$, $r = 0.1$, $q = 0$, $\sigma = 0.8$, $\epsilon_{opt} = 10^{-4}$.

FIGURE 2 NCP comparisons



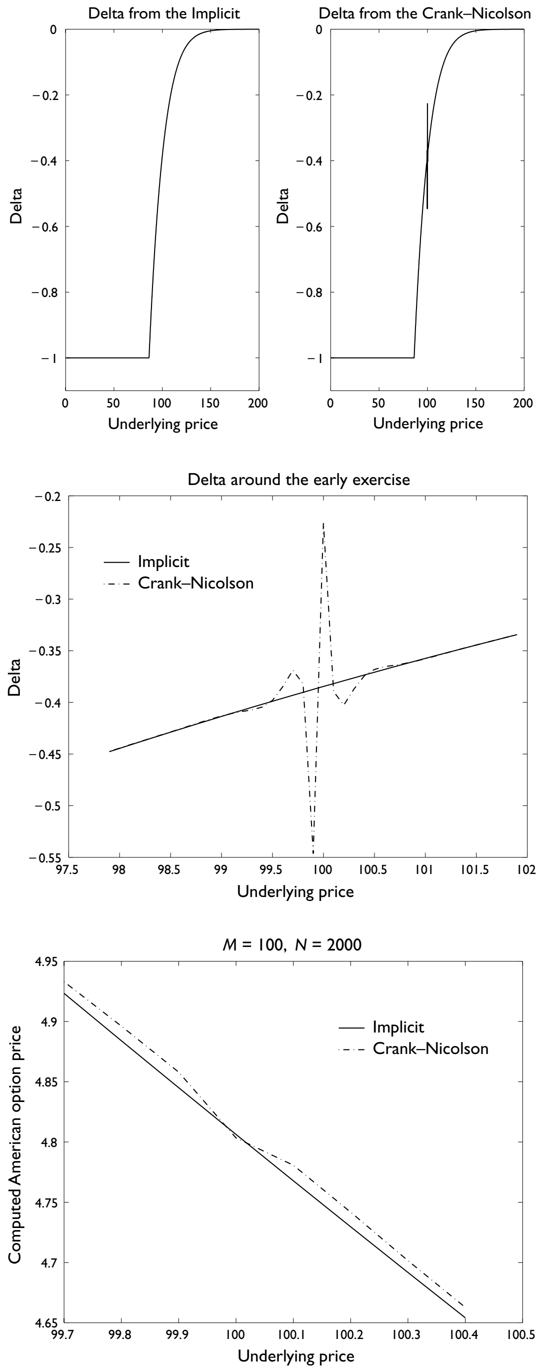
Conditions: $S_0 = 100$, $K = 100$, $T = 1$, $r = 0.1$, $q = 0$, $\sigma = 0.8$, $\epsilon_{opt} = 10^{-4}$, $\delta t^* = 0.01$.

FIGURE 3 Comparison with the binomial method



Conditions: $S_0 = 100$, $K = 100$, $T = 1$, $r = 0.1$, $q = 0$, $\sigma = 0.15$, $\epsilon_{opt} = 10^{-4}$.

FIGURE 4 Implicit and Crank–Nicolson



Conditions: $S_0 = 100$, $K = 100$, $r = 0.1$, $q = 0.0$, $\sigma = 0.2$, $\epsilon_{opt} = 10^{-4}$.

The computed delta hedge factors using the explicit payoff method are discontinuous; it is important to note that the computed option values given by the explicit payoff method do *not* solve the discretized complementarity problem (11). Indeed, the error, measured as $\max(-\min(0, Bx + c)) + x^T|Bx + c|$, of the solution obtained with the explicit payoff method is greater than 10 at each time step for the American put option with $S_0 = 100$, $T = 1$, $r = 0.1$, $\sigma = 0.1$, $q = 0.05$, $M = 2000$ and $N = 8943$.

We also investigate the continuity of delta when computed using the binomial method. Although convergence of the binomial method has been established, the convergence of the delta factor from the finite difference is unclear for the binomial method. Figure 3 compares the delta factors at $t = 0$ computed using the complementarity method and binomial method with 2,000 time steps. The discontinuity of the delta computed with the binomial method is clearly exhibited in the top plot. The middle plot shows that the size of the jump diminishes as S increases. The bottom plot illustrates the price difference.

Next we provide computational evidence to demonstrate that the Crank–Nicolson method is typically unstable for solving partial differential complementarity (American option pricing) problems. The Crank–Nicolson method has been a favorite finite-difference approximation in both European and American option pricing because its convergence rate is quadratic in time. For solving partial differential complementarity problems, however, it is noted in Jaillet, Lamberton and Lapeyre (1990) that the unconditional stability seems to be established only for the fully implicit scheme.

The instability of the Crank–Nicolson scheme is illustrated in Figure 4. We observe from the top right plot that the delta hedge factor computed using the method oscillates around the early exercise curve; this is demonstrated more clearly in the middle plot. The bottom plot graphs the option prices computed with the two methods. Moreover, we note that instability of the Crank–Nicolson method worsens as the volatility parameter becomes larger.

4.2 Efficiency of the proposed algorithm

We have demonstrated that the (implicit finite difference) complementarity approach produces continuous delta hedge factor while the popular binomial method and the explicit payoff method yield discontinuous delta. Next we show that the discretized complementarity problem (11) can be solved efficiently using the proposed algorithm, making the complementarity approach computationally feasible.

In practice, it is not necessary to solve the discretized option problem (11) to an accuracy that greatly surpasses the accuracy of the discretization. Accordingly, we set the stopping tolerance $\epsilon_{opt} = 10^{-4}$. Table 1 gives the option values computed with our proposed algorithm for discretized linear complementarity problems for two different volatility parameter settings, $\sigma = 10\%$ and $\sigma = 200\%$; the central weighting (5) is used for this example. The third and fifth

TABLE 1 Computed American put option values and average number of iterations.

M	Option value (σ = 0.1)	Iterations (σ = 0.1)	Option value (σ = 2)	Iterations (σ = 2)
100	2.3804	4.00	42.8952	5.48
200	2.3829	3.20	42.9190	5.21
300	2.3837	3.06	42.9270	4.91
400	2.3841	2.61	42.9309	4.63
500	2.3844	2.42	42.9333	4.48
600	2.3845	2.33	42.9349	4.95
700	2.3846	2.27	42.9361	4.94
800	2.3847	2.22	42.9370	4.96
900	2.3848	2.19	42.9376	4.95
1000	2.3849	2.16	42.9382	4.88
1100	2.3849	2.14	42.9386	4.76
1200	2.3849	2.12	42.9390	4.68
1300	2.3850	2.11	42.9393	4.63
1400	2.3850	2.10	42.9395	4.58
1500	2.3850	2.09	42.9398	4.56
1600	2.3850	2.08	42.9400	4.44
1700	2.3851	2.07	42.9402	4.48
1800	2.3851	2.07	42.9403	4.44
1900	2.3851	2.06	42.9405	4.38
2000	2.3851	2.06	42.9406	4.31

Conditions: $S_0 = 100$, $K = 100$, $T = 1$, $r = 0.1$, $q = 0.05$.

columns list the average number of iterations required at each time step when the proposed algorithm is used to solve a discretized problem (11). The second and the fourth columns list the respective computed option values.

From Table 1 we see that the average number of iterations required by the proposed algorithm is in the range two to four when $\sigma = 10\%$. When the volatility is unusually high ($\sigma = 200\%$), the average number of iterations is slightly higher at four to six. In both cases the average number of iterations required decreases as the discretization becomes finer. The cost of American option pricing using the proposed interior-point method is roughly two to four times that of European option pricing with typical parameter settings. The performance of the proposed algorithm is similar when $G(x, y)$ is non-linear. Table 2 presents the computed American put option values and the average numbers of iterations required when upstream weighting with the van Leer flux limiter ((4), (7), (8)) is used ; the results obtained with central weighting are included for comparison.

To evaluate the efficiency of the proposed Newton method, we compare it with other computational methods based on the partial differential complementarity formulation, even though these methods are not as general: they are

TABLE 2 NCP: computed option values and average number of iterations.

M	NCP	Iterations	LCP	Iterations
100	2.3805	4.11	2.3804	4.00
200	2.3829	3.25	2.3829	3.20
300	2.3837	2.93	2.3837	3.06
400	2.3841	2.48	2.3841	2.61
500	2.3844	2.33	2.3844	2.42
600	2.3845	2.25	2.3845	2.33
700	2.3846	2.20	2.3846	2.27
800	2.3847	2.16	2.3847	2.22
900	2.3848	2.14	2.3848	2.19
1000	2.3849	2.12	2.3849	2.16
1100	2.3849	2.11	2.3849	2.14
1200	2.3850	2.09	2.3849	2.12
1300	2.3850	2.08	2.3850	2.11
1400	2.3850	2.07	2.3850	2.10
1500	2.3850	2.07	2.3850	2.09
1600	2.3851	2.06	2.3850	2.08
1700	2.3851	2.05	2.3851	2.07
1800	2.3851	2.05	2.3851	2.07
1900	2.3851	2.05	2.3851	2.06
2000	2.3851	2.04	2.3851	2.06

Conditions: $S_0 = 100$, $K = 100$, $T = 1$, $\sigma = 0.1$, $r = 0.1$, $q = 0.05$.

applicable only when the discretized problem is a linear complementarity problem. The most popular partial differential complementarity-based method is the projected successive over-relaxation (PSOR) method. However, as noted in Huang and Pang (1998), the convergence of a PSOR method for solving a discretized linear complementarity problem with an asymmetric positive definite matrix is not well understood and convergence requires highly restrictive conditions at best. In addition, Dempster and Hutton (1997) demonstrated computationally that the convergence speed of PSOR is very sensitive to such parameters as the volatility rate, dividend rate and interest rate. The simplex method for linear programming, which can be used to solve a discretized linear complementarity problem (6) under some conditions, is more reliable and roughly comparable to the PSOR method in computational time. In a subsequent paper Dempster and Hutton (1999) demonstrated that the simplex method is fast, accurate and parametrically robust for higher-dimensional American option valuation.

An increasingly popular method for solving linear programming problems is the interior-point method. This type of method is particularly attractive for large, simple, band-structured problems because the number of iterations required by an interior point is relatively insensitive to the size of the problem (the cost per

TABLE 3 Comparison of implicit and Crank–Nicolson schemes in American option pricing.

M	Implicit complementarity	Iterations	Crank–Nicolson complementarity	Iterations
100	2.3804	4.00	2.3836	3.61
200	2.3829	3.20	2.3852	3.27
300	2.3837	3.06	2.3853	3.03
400	2.3841	2.61	2.3853	2.77
500	2.3844	2.42	2.3853	2.59
600	2.3845	2.33	2.3853	2.49
700	2.3846	2.27	2.3853	2.41
800	2.3847	2.22	2.3853	2.37
900	2.3848	2.19	2.3853	2.32
1000	2.3849	2.16	2.3853	2.28
1100	2.3849	2.14	2.3853	2.25
1200	2.3849	2.12	2.3854	2.22
1300	2.3850	2.11	2.3853	2.21
1400	2.3850	2.10	2.3854	2.19
1500	2.3850	2.09	2.3854	2.17
1600	2.3850	2.08	2.3854	2.16
1700	2.3851	2.07	2.3854	2.14
1800	2.3851	2.07	2.3854	2.13
1900	2.3851	2.06	2.3854	2.12
2000	2.3851	2.06	2.3854	2.11

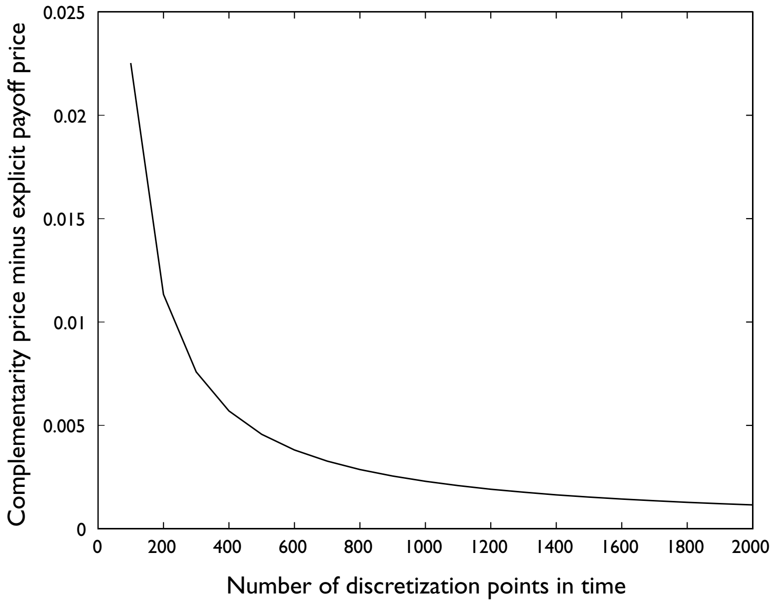
Conditions: $S_0 = 100$, $K = 100$, $T = 1$, $\sigma = 0.1$, $r = 0.1$, $q = 0.05$.

iteration becomes comparable for the simplex method and the interior-point method for band-structured problems). The number of iterations required by the simplex method for a discretized LCP problem, on the other hand, grows linearly with the number of discretization points, N, in the state dimension (Dempster and Hutton, 1999). For a linear complementarity problem with a tridiagonal matrix, for example, the cost of the simplex method is at least $3N$, whereas the dominant cost of computing a Newton step (18) is roughly $4N$. However, as illustrated by Table 2, the number of iterations required by the proposed Newton method seems to be independent of N, which suggests it is computationally more efficient than the simplex method (and, hence, the PSOR method), at least for sufficiently large N.

Table 3 compares the American option values computed using the complementarity approach and the implicit finite difference ($\theta = 1$) with the complementarity approach and the Crank–Nicolson method ($\theta = 1/2$). We note that the proposed interior-point method for the discretized American option value problem performs similarly when the Crank–Nicolson scheme is used.

In addition we note that, although the delta computed using the explicit payoff has a large jump, the computed option values are surprisingly close to those

FIGURE 5 Price difference.



Conditions: $\sigma = 2$, $r = 0.1$, $q = 0.05$, $T = 1$, $K = 100$, $S_0 = 100$, $\epsilon_{opt} = 10^{-4}$

computed using the complementarity approach. This, however, does not necessarily suggest that the option values computed with the explicit payoff method converge to the true values. Figure 5 plots the differences between the option values given by these two methods against the accuracy of discretization (the number of time steps). Note that the convergence of the (implicit) complementarity method is known (Jaillet, Lamberton and Lapeyre, 1990).

Finally, we demonstrate that the proposed algorithm can solve the discretized problem to high accuracy, if necessary. Table 4 illustrates the asymptotic behavior of the proposed algorithm when the central weighting (4), (5) is used for the convection term, and Table 5 demonstrates the convergence behavior when the discretized problem (11) is a non-linear complementarity problem. In both cases we give the optimality value $\frac{1}{2}\|G(x, y)\|^2 + x^T y$ at each iteration for the first three time steps $t^* = \delta t^*$, $2\delta t^*$, and $3\delta t^*$. In our experience this asymptotic behavior is typical. The stopping tolerance, ϵ_{opt}, is set to 10^{-7} in these examples.

The discretized linear complementarity problems are, typically, nearly degenerate since the two equations

$$V(S, t) - \Lambda(S) = 0 \quad \text{and} \quad \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - q) S \frac{\partial V}{\partial S} - rV = 0$$

TABLE 4 Asymptotic behavior for an American put option: $G(x, y)$ is linear

	$\frac{1}{2}\|\|G(x^{(k)}, y^{(k)})\|\|^2 + x^{(k)T}y^{(k)}$		
k	$t = t_1^*$	$t = t_2^*$	$t = t_3^*$
0	7.947819e+02	5.004075e+04	4.667865e+04
1	3.564160e+00	1.318422e−01	6.598802e−02
2	2.343340e−01	2.679068e−03	2.423917e−03
3	3.252325e−02	1.148443e−04	1.730893e−04
4	4.531208e−03	2.647152e−06	1.139226e−05
5	6.322716e−04	3.392189e−07	5.378067e−07
6	8.820751e−05	4.947867e−08	4.395356e−08
7	1.232954e−05		
8	1.727045e−06		
9	2.455198e−07		
10	3.851848e−08		

Conditions: $S_0 = 100$, $K = 100$, $T = 1$, $\sigma = 0.1$, $r = 0.05$, $q = 0$, $\epsilon_{opt} = 10^{-7}$, $\delta t^* = 0.01$, $\delta S = 0.1$.

TABLE 5 Asymptotic behavior for an American put option: $G(x, y)$ is non-linear

	$\frac{1}{2}\|\|G(x^{(k)}, y^{(k)})\|\|^2 + x^{(k)T}y^{(k)}$		
k	$t = t_1^*$	$t = t_2^*$	$t = t_3^*$
0	4.468042e+08	4.069523e+08	3.759767e+08
1	4.481993e+05	1.124695e+00	5.422102e−01
2	3.893335e+03	1.573953e−02	1.437938e−02
3	3.963988e+02	2.778598e−03	2.524083e−03
4	3.741587e−05	3.597233e−04	4.196310e−04
5	9.382774e−06	2.765571e−05	6.150641e−05
6	2.880286e−06	2.958813e−06	1.023294e−05
7	1.373253e−06	1.257544e−06	1.928311e−06
8	9.978982e−07	8.415812e−07	8.534015e−07

Conditions as in Table 4.

are simultaneously approximately satisfied around the early exercise curve. The results in Table 4 show that this near-degeneracy does not cause visible computational difficulty; superlinear convergence is observed for a couple of iterations up to a fairly high accuracy. After this point convergence becomes fast linear.

Discretized non-linear complementarity problems generated from the upstream weighting with the van Leer non-linear flux limiter have additional degeneracy: the Jacobian of the constraints $\bar{G}(x, y) = 0$ can be singular when $V_{down} = V_{up}$, as can be seen from the definition of $\bar{G}$ given in (10). We observe that, although the proposed method requires a few more iterations when $G(x, y)$ is non-linear, convergence is still rapid.

5. Concluding remarks

American option pricing using finite-difference approximation can be formulated as a finite-dimensional complementarity problem. Depending on the finite-difference approximation used, this complementarity problem can be a linear complementarity with a tridiagonal coefficient matrix, a general linear complementarity problem, or a non-linear complementarity problem. As can be seen from (12), even the discretized problem for the European option becomes a non-linear programming problem with complementarity constraints when the upstream weighting with the van Leer flux limiter is used for the convection term.

We propose a Newton-type interior-point for solving these discretized non-linear programming problems arising from American option pricing; the close relationship between the option values of the consecutive time steps is exploited by the Newton process. We demonstrate that a small number of iterations, typically about two to five, is required to compute a sufficiently accurate price. More importantly, the number of iterations does not increase with the number of discretization points in the spatial dimension; it actually decreases as the time discretization becomes finer. This makes the proposed method particularly suitable for higher-dimension problems.

We investigate the computed (finite-difference) delta hedge factors using the complementarity method, the binomial method, and the explicit payoff method. We show that, whereas the (implicit finite-difference) complementarity approach produces continuous delta hedge factors, the binomial method and the explicit payoff method yield discontinuous delta. Therefore, the early exercise curve obtained with the last two methods can be inaccurate. Finally, we demonstrate computationally that the Crank–Nicolson method, which is stable for European option valuation, can yield an oscillatory delta hedge factor.

REFERENCES

Amin, K., and Khanna, A. (1994). Convergence of American option values from discrete- to continuous-time financial models. *Mathematical Finance* **4,** 289–304.

Barraquand, J., and Pudet, T. (1996). Pricing of American path-dependent contingent claims. *Mathematical Finance*, 17–51.

Brennan, M., and Schwartz, E. (1977). The valuation of American put options. *Journal of Finance* **32**, 449–62.

Broadie, M., and Detemple, J. (1996). American option valuation: new bounds, approximations, and a comparison of existing methods. *Review of Financial Studies* **9**, 1211–50.

Broadie, M., and Detemple, J. (1997). Recent advances in numerical methods for pricing derivative securities. In *Numerical Methods in Finance*, L. C. G. Rogers and D. Talay, eds, Publication of the Newton Institute, 43–66.

Coleman, T. F., and Li, Y. (1994). On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds. *Mathematical Programming* **67**, 189–224.

a, b added to distinguish

Coleman, T. F., and Li, Y. (1996a). An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization* **6**, 418–45.

Coleman, T. F., and Li, Y. (1996b). A reflective Newton method for minimizing a quadratic function subject to bounds on the variables. *SIAM Journal on Optimization* **6**, 1040–58.

Cox, J., Ross, S., and Rubinstein, M. (1979). Option pricing: a simplified approach. *Journal of Financial Economics*, 229–63.

Dempster, M. A. H., and Hutton, J. P. (1997). Fast numerical valuation of American, exotic and complex options. *Applied Mathematical Finance* **4**, 1–20.

Dempster, M. A. H., and Hutton, J. P. (1999). Pricing American stock option by linear programming. *Mathematical Finance* **9**, 229–54.

Dewynne, J. N., and Wilmot, P. (1995). Asian options as linear complementarity problems: analysis and finite difference solutions. *Advances in Futures and Options Research* **8**, 145–73.

Duffie, D. (1996). *Dynamic Asset Pricing Theory*. Princeton.

Huang, J. (1999). American options and complementarity problems. Ph.D. thesis, Johns Hopkins University.

Huang, J., and Pang, J. (1998). Option pricing and linear complementarity. *Journal of Computational Finance* **2**, 31–60.

Hull, J. (1997). *Options, Futures, and Other Derivatives*. Prentice-Hall.

Jaillet, P., Lamberton, D., and Lapeyre, B. (1990). Variational inequalities and pricing of American options. *Acta Applicanda Mathematicae* **21**, 263–89.

Kojima, M., Mizuno, S., and Yoshise, A. (1989). A primal–dual interior point method for linear programming. In *Progress in Mathematical Programming Interior-point and Related Methods*, N. Megiddo, ed., Springer-Verlag, 29–47.

Roache, P. (1972). *Computational Fluid Dynamics*. Hermosa, Albuquerque, New Mexico.

Wilmott, P., Dewynne, J., and Howison, S. (1993). *Option Pricing: Mathematical models and computation*. Oxford Financial Press.

Zhang, X. L. (1997). Valuation of American options in a jump–diffusion model. In *Numerical Methods in Finance*, L. C. G. Rogers and D. Talay, eds., Publication of the Newton Institute, 93–114.

Zvan, R., Vetzal, K., and Forsyth, P. (1997). Robust numerical methods for pde models of Asian options. *Journal of Computational Finance*, 39–78.