Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 222 (2008) 94–111
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

An irregular grid approach for pricing high-dimensional American options☆

S.J. Berridge[a,*], J.M. Schumacher[b]

[a] *Man Investments, Sugar Quay, Lower Thames Street, London EC3R 6DU, United Kingdom*
[b] *Department of Econometrics and Operations Research and Center for Economic Research (CentER), Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands*

Received 30 April 2006; received in revised form 9 April 2007

Abstract

We propose and test a new method for pricing American options in a high-dimensional setting. The method is centered around the approximation of the associated complementarity problem on an irregular grid. We approximate the partial differential operator on this grid by appealing to the SDE representation of the underlying process and computing the root of the transition probability matrix of an approximating Markov chain. Experimental results in five-dimensions are presented for four different payoff functions.

MSC: 35R35; 60G40; 65D15; 90C33

Keywords: American options; High-dimensional problems; Free boundary problems; Optimal stopping; Variational inequalities; Numerical methods; Unstructured mesh; Markov chain approximation

1. Introduction

The pricing of American options has always required numerical solution methods; in high-dimensional cases even the most sophisticated methods have difficulty in providing accurate solutions. Given the practical importance of such cases, it is of considerable interest to develop solution methods which are reliable and which provide accompanying exercise and hedging strategies.

Barraquand and Martineau [2] are perhaps the first to consider pricing high-dimensional American options specifically, proposing an algorithm based on the aggregation of paths with respect to the intrinsic value. The method is difficult to analyze and has a possible lack of convergence; Boyle et al. [3] demonstrate this and propose a modification of the algorithm which leads to a low-biased estimator.

☆ This paper is a considerably revised and extended version of the conference paper which was presented at the International Conference on Computational Science held in Amsterdam, 2002, and published in the special issue of Future Generation Computer Systems [S.J. Berridge, J.M. Schumacher, An irregular grid method for high-dimensional free-boundary problems in finance, Future Generation Computer Systems 20 (2004) 353–362] which was devoted to papers from this meeting.

* Corresponding author.
E-mail addresses: sberridge@maninvestments.com, Steffan.Berridge@virgin.net (S.J. Berridge), j.m.schumacher@uvt.nl (J.M. Schumacher).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2007.10.045

Broadie and Glasserman [4] use a stochastic tree algorithm to give both a low-biased and a high-biased estimator of the price, both asymptotically unbiased. They also argue that there exists no nontrivial unbiased estimator for the price. Their method requires an exponentially increasing amount of work in the number of exercise opportunities. In subsequent work [5] they present a related method based on a stochastic mesh which does not suffer from this problem, although this method has been found to be slow by several authors and to have a large finite-sample bias (see e.g. Fu et al. [6]).

The least squares Monte Carlo (LSM) method presented in [7] attempts to approximate the price of an American option using cross-sectional information from simulated paths. The optimal exercise strategy is successively approximated backwards in time on the paths by comparing the intrinsic values to the continuation values projected onto a number of basis functions over the states. Experimental success is reported for the LSM method, although in high dimensions the basis functions must be chosen carefully. Recently Clément et al. [8] and Stentoft [9] independently provided proofs of convergence for the LSM method, showing that the convergence rate is $n^{-1/2}$ in the number of paths used. The convergence behavior in the number of basis functions however has not been determined. Stentoft [9] and Glasserman and Yu [10] establish relationships between the paths and number of basis functions which are necessary for convergence; Stentoft finds that the number of paths should be greater than cubic in the number of basis functions to achieve convergence in probability, while Glasserman and Yu find the relationship should be exponential in the square for convergence on a worst case basis. Stentoft [11] and Moreno and Navas [12] test the LSM algorithm numerically. Stentoft suggests that basis functions up to order three are sufficient in five dimensions for arithmetical and geometrical average options, but not for minimum or maximum options. Moreno and Navas find that the method is sensitive to the choice of basis functions in five dimensions.

Tsitsiklis and Van Roy [13] propose a method similar to LSM where approximate value functions are projected onto an orthogonal set of basis functions, the orthogonality being with respect to a suitably chosen inner product which in general changes between time periods. They provide a proof of convergence but no empirical results. The method differs from LSM in that the projection is used to determine an approximate value function rather than an exercise rule.

Boyle et al. [14] recently extended the stochastic mesh method of Broadie and Glasserman [5] with their low discrepancy mesh (LDM) method. This involves generating a set of low-discrepancy interconnected paths and using a dynamic programming approach to find prices on the mesh.

An interesting alternative approach is proposed independently by Rogers [15] and Haugh and Kogan [16] and later developed by Jamshidian [17] and Kolodko and Schoenmakers [18]. They use a dual formulation of the problem in which a minimization is performed over martingales. The method is sensitive to the choice of basis martingales chosen to perform the minimization, and so requires the basis to be well chosen in order to give an accurate solution. The method gives a high-biased estimator.

In this paper, we propose a new approach to solving the American option pricing problem inspired by the success of numerical integration in high dimensions and related to the method of lines for solving partial differential equations (PDEs).

We first perform a discretization of the state space using quasi-Monte Carlo (QMC) points, the points being taken with respect to an importance sampling distribution related to the transition density of the process at expiry. We then propose an approximation to the partial differential operator on this grid by taking the logarithm of a transition probability matrix $P^{(T-t)}$ which approximates the joint density of the underlyings at the expiry of the option, $T-t$. This approximation is then used to formulate linear complementarity problems (LCPs) at successive time points, working back from the option expiry.

We propose an implementation of this method in which the matrix logarithm of $P^{(T-t)}$ does not need to be calculated explicitly, but instead a root of the matrix can be calculated. The root operation is cheaper than the logarithm, although the logarithm allows variation of the time step without recalculation. The computational elements of the method are thus the QMC trials, the generation of the matrix $P^{(T-t)}$, the matrix root and solving an LCP at each time step. For approximating the European option price this method amounts to performing a numerical integration with importance sampling, which is known to be an efficient method in high dimensions as long as the importance sampling distribution is chosen appropriately.

The remainder of the paper is organized as follows. In Section 2 we present a mathematical formulation of the problems to be solved numerically and in Section 3 we show how an irregular grid method can be used to solve the problem. We then present the experimental setup in Section 4, results in Section 5 and concluding remarks in Section 6.

2. Formulation

We consider a complete and arbitrage-free market described by state variable $X(s) \in \mathbb{R}^d$ for $s \in [t, T]$ which follows a Markov diffusion process

$$\mathrm{d}X(s) = \mu(X(s), s)\mathrm{d}s + \sigma(X(s), s)\mathrm{d}W(s) \tag{2.1}$$

with initial condition $X(t) = x_t$, and a derivative product on $X(s)$ with intrinsic value $\psi(X(s), s)$ at time s and value $V(s) = v(X(s), s)$ for some pricing function $v(x, s)$. The process $V(s)$ satisfies

$$\mathrm{d}V(s) = \mu_V(X(s), s)\mathrm{d}s + \sigma_V(X(s), s)\mathrm{d}W(s) \tag{2.2}$$

where μ_V and σ_V can be expressed in terms of μ and σ by means of Itô's lemma. The terminal value is given by $v(\cdot, T) = \psi(\cdot, T)$.

In such a market there exists a unique equivalent martingale measure under which all price processes are martingales. The risk-neutral process in this case is given by

$$\mathrm{d}X(s) = \mu_{RN}(X(s), s)\mathrm{d}s + \sigma(X(s), s)\mathrm{d}W(s) \tag{2.3}$$

where μ_{RN} is the risk-neutral drift.

Our objective is to provide approximations for the current value $v(x_t, t)$ of the derivative product and the corresponding optimal exercise and hedging strategies τ and H:

$$\tau : \mathbb{R}^d \times [t, T] \to \{0, 1\} \tag{2.4}$$

$$H : \mathbb{R}^d \times [t, T] \to \mathbb{R}^d. \tag{2.5}$$

In the following, we appeal to the complementarity formulation of the American option price which is presented, for example, in [19]. Let $\mathcal{L}$ be the related diffusion operator

$$\mathcal{L} = \frac{1}{2}\mathrm{tr}\,\sigma\sigma' \frac{\partial^2}{\partial x^2} + \mu_{RN}\frac{\partial}{\partial x} - r \tag{2.6}$$

where r is the risk-free rate and the prime denotes transpose. Then the option value is found by solving the complementarity problem

$$\begin{cases} \dfrac{\partial v}{\partial t} + \mathcal{L}v \le 0 \\ v - \psi \ge 0 \\ \left(\dfrac{\partial v}{\partial t} + \mathcal{L}v\right)(v - \psi) = 0 \end{cases} \tag{2.7}$$

for $(x, s) \in \mathbb{R}^d \times [t, T]$ with the terminal condition $v(\cdot, T) \equiv \psi(\cdot, T)$. In fact it turns out to be simpler in terms of the following discretization to work with the forward-price version $w(x, t) = e^{r(T-t)}v(x, t)$, which satisfies the complementarity problem

$$\begin{cases} \dfrac{\partial w}{\partial t} + \hat{\mathcal{L}}w \le 0 \\ e^{-r(T-t)}w - \psi \ge 0 \\ \left(\dfrac{\partial w}{\partial t} + \hat{\mathcal{L}}w\right)\left(e^{-r(T-t)}w - \psi\right) = 0 \end{cases} \tag{2.8}$$

where $\hat{\mathcal{L}} = \mathcal{L} + r$, again for $(x, s) \in \mathbb{R}^d \times [t, T]$ and with the terminal condition $w(\cdot, T) \equiv \psi(\cdot, T)$.

3. Methodology

To solve the complementarity problem (2.8) we first form a semidiscrete complementarity problem by discretizing the state space but leaving time continuous. This involves sampling the state space using QMC trials and finding a

suitable approximation of $\hat{\mathcal{L}}$. We then use standard time-stepping techniques to form a system of fully discrete LCPs. There are many methods for solving LCPs; examples include projected successive overrelaxation (PSOR) and linear programming.

We first present and motivate each step of the algorithm separately, and then summarize by providing a concise statement of the algorithm.

3.1. State space discretization

We first consider a semidiscrete approximation to the complementarity problem (2.8) in which the state space is discretized and time left continuous. This is often called the method of lines. In the pricing problem this amounts to approximating (2.8) by a system of ordinary differential equations with complementarity conditions.

The choice of a constant grid in the state space has the advantage that Crank–Nicolson and implicit solutions can be easily considered. This seems advantageous since, in the case of solving PDEs without complementarity conditions, the Crank–Nicolson method is known to have a convergence rate of δt^2 rather than δt for other first order schemes. Furthermore, when solving discretized complementarity problems, the implicit scheme is the only time-stepping method known to be unconditionally stable (see [20]).

The choice of grid begs importance sampling considerations. That is, in order to obtain a more accurate approximation, more grid points should be placed at states which are more likely to be visited by the process, and at locations where the value function has greater magnitude.

We denote the grid by $\mathcal{X} = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$, and the corresponding operator approximation by A. The construction of A will be considered in Section 3.2.

Assuming that $\mathcal{X}$ and A are given, we form the corresponding semidiscrete complementarity problem

$$\begin{cases} \dfrac{\mathrm{d}w}{\mathrm{d}t}(s) + Aw(s) \;\leq\; 0 \\ e^{-r(T-s)}w(s) - \psi \;\geq\; 0 \\ \left(\dfrac{\mathrm{d}w}{\mathrm{d}t}(s) + Aw(s)\right)' \left(e^{-r(T-s)}w(s) - \psi\right) \;=\; 0 \end{cases} \tag{3.1}$$

for $s \in [t, T]$ with terminal condition $w_i(T) = \psi(x_i)$ for each $i = 1, \ldots, n$. Note that $w(s)$ is now a time-dependent vector in $\mathbb{R}^n$ where n is the number of grid points.

It is also instructive to view the semi-discrete setting as a continuous-time Markov chain approximation to the optimal stopping problem. That is, the process $X(s)$ is approximated by a process restricted to the grid $\mathcal{X}$ with the operator A giving transition intensities on this grid. The latter can be seen as follows: since the stopped process $w(X_t, t)$ is a martingale, the vector $p(t)$ giving the probability of being in each state at time t must satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(p(t)'w(t)\right) = 0.$$

Substituting $\frac{\mathrm{d}w}{\mathrm{d}t}(t) = -Aw(t)$, we have

$$\left(\frac{\mathrm{d}p}{\mathrm{d}t}(t) - A'p(t)\right)' w(t) = 0.$$

Since this equation must hold for all w in the continuation region, we must have $\frac{\mathrm{d}p}{\mathrm{d}t} = A'p$. This connection between the differential operator and the evolution of probability densities will be used in Section 3.2 to specify transition intensities in the approximating Markov chain.

3.2. Approximating the partial differential operator

We now propose a method for specifying A in (3.1) for a given grid $\mathcal{X}$. The method is inspired by numerical integration, and in the European case the resulting method will reduce to numerical integration with importance sampling. This property is emphasized in [21] as a favorable property of the stochastic mesh method presented in [5].

Our approximation method proceeds via the stochastic representation of the pricing problem. Since we want to approximate the Black–Scholes partial differential operator $\mathcal{L}$ (or its forward-price version $\hat{\mathcal{L}}$), we may consider the European case. Suppose that we determine a continuous-time Markov chain that uses the grid $\mathcal{X}$ as a state space and that approximates the continuous-state process (2.1). Let A be the matrix of transition intensities of the chain, and let $p_i(t)$ denote the probability of finding the chain in state i at time t. As is well known, the vector of probabilities $p(t)$ satisfies the forward differential equation $(\mathrm{d}p/\mathrm{d}t)(t) = A'p(t)$. Let $w(x,t)$ be a solution of the forward-price Black–Scholes equation $\frac{\partial w}{\partial t} + \hat{\mathcal{L}}w = 0$. We know that such solutions are characterized in stochastic terms by the martingale property of the process $w(X_t, t)$. We therefore require that the approximating Markov chain should satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(p(t)'w(t)\right) = 0$$

where, as above, $w(t)$ denotes the state-discretized version of the function $w(x,t)$. Substituting $\frac{\mathrm{d}p}{\mathrm{d}t}(t) = A'p(t)$, we have

$$p(t)'\left(\frac{\mathrm{d}w}{\mathrm{d}t}(t) + Aw(t)\right) = 0.$$

Since this equation must hold in particular for all deterministic initial conditions, we must have $\frac{\mathrm{d}w}{\mathrm{d}t} = -Aw$. This shows that the matrix A may be taken as an approximation of the forward-price Black–Scholes operator $\hat{\mathcal{L}}$.

We assume that the grid $\mathcal{X}$ has been generated using random or QMC draws with respect to a certain density $g(x)$. We also assume that the joint density $f_{x,T-t}$ of the stochastic process is available for arbitrary initial points x and time horizons $T-t$, although in principle one could adapt the following construction to the case where the density is not known explicitly, but for example the process can be simulated.

Denote by $P^{(T-t)}$ the matrix with entries

$$p_{ij}^{(T-t)} = \frac{1}{\sum_{k=1}^{n} \tilde{f}_{x_i,T-t}(x_k)} \cdot \tilde{f}_{x_i,T-t}(x_j) \tag{3.2}$$

where the weights are given by

$$\tilde{f}_{x_i,T-t}(x) = \frac{f_{x_i,T-t}(x)}{g(x)}. \tag{3.3}$$

The matrix $P^{(T-t)}$ is a stochastic matrix, that is, a matrix with nonnegative entries and unit row sums. We think of the entries as giving transition probabilities between points in the grid over the horizon $T-t$. The matrix $P^{(T-t)}$ thus gives us access to an approximation A to $\hat{\mathcal{L}}$ on $\mathcal{X}$ as follows:

$$A \triangleq \frac{1}{T-t}\log P^{(T-t)}. \tag{3.4}$$

The matrix logarithm of $P^{(T-t)}$ certainly exists and is unique if the matrix is diagonalizable and has positive eigenvalues. We find these two properties hold in our experiments; note however that $P^{(T-t)}$ is in general not symmetrical. We shall see in Section 3.3 that instead of computing the matrix logarithm, one may alternatively compute the matrix root corresponding to the required time step. A direct method to construct an approximative transition intensity matrix, based on second-order matching of local transition densities, has been suggested in [22].

3.3. Time discretization

Let us now discretize (3.1) with respect to time. We denote the approximation at state x_i and time step t_k by $w_{i,k}$. We use the θ-method, standard in the numerical solution to PDEs, to discretize (3.1). For PDE solutions, $\theta = 0$ corresponds to the explicit method, $\theta = 1$ corresponds to the implicit method and $\theta = \frac{1}{2}$ corresponds to the Crank–Nicolson method. The latter has δt^2 convergence for European problems, whereas the explicit and implicit methods exhibit δt convergence.

To implement the θ-method, we consider the vector $w^{(k)}$ of values at our grid points each at time t_k and discretize the first line of (3.1) as

$$\frac{w^{(k+1)} - w^{(k)}}{\delta t_k} + A\left((1-\theta)w^{(k+1)} + \theta w^{(k)}\right) \leq 0 \tag{3.5}$$

where $\delta t_k \triangleq t_{k+1} - t_k$. Thus (3.1) becomes

$$\begin{cases} (I + (1-\theta)A\delta t_k)\, w^{(k+1)} - (I - \theta A\delta t_k)\, w^{(k)} & \leq & 0 \\ e^{-r(T-t_k)} w^{(k)} - \psi & \geq & 0 \\ \left((I + (1-\theta)A\delta t_k)\, w^{(k+1)} - (I - \theta A\delta t_k)\, w^{(k)}\right)' \left(e^{-r(T-t_k)} w^{(k)} - \psi\right) & = & 0. \end{cases} \tag{3.6}$$

Now note that $I + A\delta t_k = \exp(A\delta t_k) + o(\delta t_k)$. We thus define the matrices

$$M_L = \exp\{-\theta A\delta t_k\} \tag{3.7}$$

$$M_R = \exp\{(1-\theta)A\delta t_k\}. \tag{3.8}$$

The approximating complementarity problem to solve is then

$$\begin{cases} M_R w^{(k+1)} - M_L w^{(k)} & \leq & 0 \\ e^{-r(T-t_k)} w^{(k)} - \psi & \geq & 0 \\ \left(M_L w^{(k)} - M_R w^{(k+1)}\right)' \left(e^{-r(T-t_k)} w^{(k)} - \psi\right) & = & 0 \end{cases} \tag{3.9}$$

for $k = K-1, \ldots, 0$ where the inequalities are componentwise and $w^{(K)} = \psi$.

Numerically we must solve an LCP at each time step, for which the PSOR method of Cryer [23] has been used with much success in the past. Since the solution does not change greatly between time steps, a good starting guess for PSOR is the solution at the previous time step. Various other methods may be used for solving (3.9); for example, see [24] for American option pricing using linear programming in the one-dimensional case.

An error analysis of the discretization in (3.5) may be undertaken along the lines of Glowinski et al. [20] on variational inequalities or that of Kushner and Dupuis [25] on stochastic control.

It turns out that the matrix logarithm does not have to be calculated explicitly in our method; instead we may calculate roots of the matrix $P^{(T-t)}$ corresponding to the time step and implicitness parameters. In particular we have

$$M_L = \left(P^{(T-t)}\right)^{-\theta\delta t_k/(T-t)} \tag{3.10}$$

$$M_R = \left(P^{(T-t)}\right)^{(1-\theta)\delta t_k/(T-t)}. \tag{3.11}$$

We prefer to use the matrix root because we have found it to be a quicker and more robust operation in Matlab than the matrix logarithm. If one would choose to compute the logarithm, however, one would have access to a varying time step without performing any extra computations.

There are many methods available for evaluating matrix functions, as detailed in [26]. The general method suggested involves Schur decomposition in combination with Parlett's algorithm, which computes general functions of an upper triangular matrix. Matrix functions can also be computed using eigendecomposition, which is the method used by Matlab to compute general matrix powers. We note that the structure of the matrix $P^{(T-t)}$ may mean that more efficient methods are available for computing matrix roots and logarithms; it is not the purpose of the current research to investigate such methods however.

We now highlight the importance of using the matrix logarithm or root, as opposed to constructing $P^{(\delta t)}$ directly (the latter being more attractive computationally). The intuition for this importance is that $P^{(\delta t)}$ does not produce consistent transition probabilities over longer time horizons as in (3.12). We demonstrate the difference between the two constructions in Fig. 3.1 for a one-dimensional example and a random grid. In particular, when δt is too small compared to the separation of grid points, the solutions become distorted. This problem is more pronounced in higher dimensions due to the larger average separation between grid points.

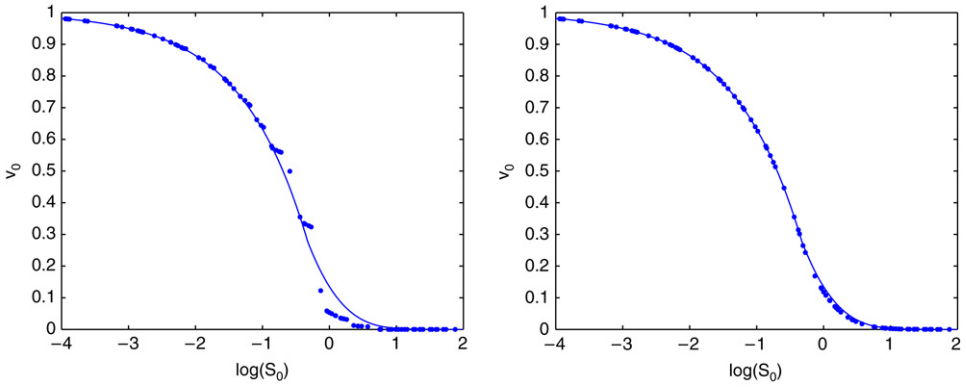

Fig. 3.1. Random grid valuation of an American put option on a single asset with expiry 1, strike 1 and 100 asset points, using transition matrices $P^{(0.01)}$ and $(P^{(1)})^{0.01}$ respectively (*dots*). The plots are in the log domain. Also plotted are values computed to high accuracy (*solid lines*) using a standard finite difference method.

Remark 3.1. It is clearly more efficient if the matrices M_L and M_R need be calculated only once; hence the choice of a constant time step $\delta t_k \equiv \delta t$ seems convenient. We also note that, given a small enough δt, M_L and M_R should be approximately sparse in that most elements can be set to zero without affecting the solution significantly. Using this observation can dramatically improve the efficiency of the solution procedure.

Remark 3.2. As already noted above, an important property of this construction in the European setting is that the time stepping reconstructs numerical integration with importance sampling function g. The reconstruction is realized as follows:

$$\begin{aligned} v_i(t) = e^{-r(T-t)} w_i(t) &= e^{-r(T-t)} \left(\prod_{k=1}^{K} M_L^{-1} M_R \right) w^{(K)} \\ &= e^{-r(T-t)} \exp(A)\psi \\ &= e^{-r(T-t)} \sum_{j=1}^{n} p_{ij}^{(T-t)} \psi(x_j) \end{aligned} \tag{3.12}$$

where $v_i(t)$ is the value in state x_i at initial time t and $p_{ij}^{(T-t)}$ is defined in (3.2). The last line of the equation is precisely QMC integration of the payoff ψ with importance sampling function equal to the grid density g. Note that in case M_L and M_R are constructed from the matrix logarithm, (3.12) holds only asymptotically as $\delta t \to 0$.

Eq. (3.12) also shows that the calculation of the European price on the grid may be carried out without time stepping, given that the transition probabilities $p_{ij}^{(T-t)}$ are available. Thus, using the European price as control variate is a faster operation than would normally be expected.

3.4. Randomization

The QMC grids we have proposed are deterministic; however, perturbing these points randomly allow us to observe the behavior of solutions for a random selection of QMC grids, and thus to obtain estimates of the bias and standard error of solutions. The use of such methods is surveyed in [27] for integration problems. The importance of randomized QMC is also emphasized in [21].

When using Sobol' points and a normal density for example, one first generates the Sobol' points, then applies the inverse normal distribution function to the points. In order to realize randomized QMC points, one perturbs the Sobol' points modulo one by a random factor before applying the inverse normal distribution function.

Suppose $\mathcal{S} = (s_i)$ is our sequence of n Sobol' points, and U_j is a sequence of random variables uniformly distributed on the unit cube $[0, 1]^d$. We then realize the jth randomized Sobol' sequence as

$$\mathcal{S}_j = (s_{j,i})_{i\in\mathbb{N}} = \left(s_i + U_j \bmod 1\right)_{i\in\mathbb{N}}. \tag{3.13}$$

We refer to grids obtained in this way as randomized QMC (RQMC) grids.

3.5. Summary of procedure

We now present a concise statement of the proposed procedure as Algorithm 1. We let $\hat{v}_{i,j}$ denote the solution at initial time t and state x_i in the jth RQMC experiment. For the statement of the algorithm we assume a fixed number of grid points n and a constant time step $\delta t = (T-t)/K$ where K is the number of time steps.

Algorithm 1 The proposed irregular grid algorithm

for $j = 1, \ldots, J$ **do**
  Generate a RQMC grid $\mathcal{X}$
  Compute the transition matrix for expiry $P^{(T-t)}$
  Compute the matrix root $\left(P^{(T-t)}\right)^{1/2K}$ (Crank–Nicolson)
  Solve the LCPs (3.9) to obtain $\hat{w}_{i,j}$
  Let $\hat{v}_{i,j} = e^{-r(T-t)}\hat{w}_{i,j}$ be the solution at initial time t for state x_i
end for
for initial states of interest x_i **do**
  Estimate the solution as $\hat{v}_i = \frac{1}{J}\sum_j \hat{v}_{i,j}$
  Estimate the standard error as $\hat{\epsilon}_i = \left(\frac{1}{J-1}\sum_j (\hat{v}_{i,j} - \hat{v}_i)^2\right)^{1/2}$.
end for

4. Experimental setup and details

We now use the algorithm presented in Section 3.5 to estimate prices of multiasset options. We first present a detailed exposition of the setting, experimental procedure and various considerations. Numerical results are presented in the next section.

4.1. Specification of dynamics

Suppose our American option is based on d assets following a correlated geometrical Brownian motion where the risk-neutral dynamics in the log domain are given by

$$\mathrm{d}X = \left(r\mathbb{1} - \delta - \frac{1}{2}\mathrm{diag}(\Sigma)\right)\mathrm{d}t + R'\mathrm{d}W \tag{4.1}$$

and r is the risk-free rate, $\mathbb{1}$ is the d-vector of ones, $\delta = (\delta_1, \ldots, \delta_d)$ is the vector of dividend rates, $\Sigma = \left(\rho_{ij}\sigma_i\sigma_j\right)$ is the covariance matrix of the Brownian motions and R is any matrix such that $R'R = \Sigma$, determined for instance by Cholesky decomposition or principal vector decomposition. The operator $\mathcal{L}$ in this setting is just the multidimensional Black–Scholes operator given by

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{d} \rho_{ij}\sigma_i\sigma_j \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d}\left(r - \delta_i - \frac{1}{2}\sigma_i^2\right)\frac{\partial}{\partial x_i} - r. \tag{4.2}$$

4.2. Elimination of drift

In order to facilitate reuse of the matrix roots, we first reformulate the problem so that the process has zero drift. We introduce the change of variables

$$X_0(s) = X(s) - (s-t)\mu, \tag{4.3}$$

where μ is the risk-neutral drift; for example, in (4.1) we have $\mu = r\mathbb{1} - \delta - \frac{1}{2}\mathrm{diag}(\Sigma)$. The new process X_0 has zero drift and the covariance Σ is unchanged:

$$\mathrm{d}X_0(s) = R'\mathrm{d}W(s). \tag{4.4}$$

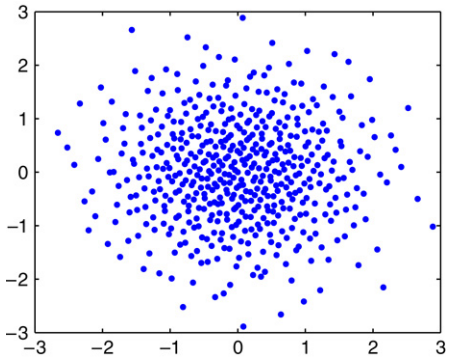

Fig. 4.1. Example of normal QMC grid in 2-dimensions with 500 points.

The payoff under the reformulation is

$$\psi_0(x_i, s) = \psi(x_i + (s - t)\mu). \tag{4.5}$$

One may also eliminate a deterministic, time-dependent risk-neutral drift by subtracting $\int_t^s \mu(u)\mathrm{d}u$ in (4.3).

4.3. Grid specification

We consider normal RQMC grids as suggested in Section 3.4; thus the grid density is multivariate normal. We now discuss parameter selection for the grid density.

Importance sampling considerations tell us that the most efficient sampling is given by the density of the process itself; thus using a constant grid we cannot provide the most efficient importance sampling at all times. However, given the restriction to a constant grid, we can still provide an acceptable importance sampling.

As outlined in [28], the rate of convergence for importance sampling of normal densities using normal importance sampling functions is most damaged when the variance of the importance sampling function is less than that of the true density. Conversely, convergence rates are not greatly affected when the variance of the importance sampling function is greater than that of the true density.

The situation we should try to avoid is that the process has a significant probability of lying in the "tails" of the grid density. A further consideration is the minimization of boundary effects on the solution. This suggests that the grid covariance should be larger than the covariance of the process.

These considerations lead us to set the grid mean to the initial state x_t and the grid covariance to be a multiple α of the grid density at expiry for some trial values $\alpha = 1.0, 1.5, 2.0$. Owing to the reformulation (4.3), this ensures that the grids are centered at the process mean for all times. We further ensure that the initial state is included in the grid.

Summarizing, we suggest the parameters

$$\mu_g = x_t \tag{4.6}$$

$$\Sigma_g = \alpha\Sigma(T - t). \tag{4.7}$$

The first grid point in the jth RQMC experiment is $x_1 = \mu_g$ and the $(i + 1)$th grid point is generated as

$$x_{i+1} = \mu_g + R_g' \left(\Psi^{-1}(s_{j,i,1}) \cdots \Psi^{-1}(s_{j,i,d}) \right)' \tag{4.8}$$

where Ψ^{-1} is the standard normal inverse function, $R_g' R_g$ is the Cholesky decomposition of Σ_g and $s_{j,i,k}$ is the kth component of $s_{j,i}$.

An example of a normal Sobol' grid in two-dimensions is shown in Fig. 4.1. It should be noted, however, that the advantage of using an irregular grid is realized in dimensions of at least three.

4.4. Reuse of roots for similar processes

Given that generating matrix roots is an expensive operation compared to the final time-stepping procedure; it is of interest to know under which conditions these matrix roots can be reused for related problems; for example, problems with different parameters.

Clearly a single matrix root can be reused for as many different payoff functions as required, but we also show how it can be reused for processes with different risk-neutral drifts and covariances. To answer this question for diffusion processes with zero drift we provide the following result.

Lemma 4.1. *Suppose that $P^{(T-t)}$ is the transition matrix corresponding, through (3.2) and (3.3), to the grid $\mathcal{X} = \{x_1, \ldots, x_n\}$, respective importance sampling weights $g_1, \ldots, g_n$, horizon $T-t$ and an d-dimensional Brownian motion with covariance I_d. Suppose further that $\tilde{P}^{(T-t)}$ is the transition matrix corresponding to the grid*

$$\mathcal{Y} = \{y_1, \ldots, y_n\} = \{R'x_1, \ldots, R'x_n\}, \tag{4.9}$$

importance sampling weights $g_1, \ldots, g_n$, horizon $T-t$ and an d-dimensional Brownian motion with positive definite covariance $\Sigma = R'R$.

Then

$$\tilde{P}^{(T-t)} = P^{(T-t)}. \tag{4.10}$$

Proof. Let $f_{x,T-t}$ and $h_{x,T-t}$ be the densities at expiry from starting point x, expiry $T-t$ and corresponding to covariance I and Σ respectively. The densities from x_i to x_j in grid $\mathcal{X}$ and from $y_i = R'x_i$ to $y_j = R'x_j$ in grid $\mathcal{Y}$ are respectively

$$\begin{aligned} f_{x_i,T-t}(x_j) &= |2\pi(T-t)I_d|^{-1/2} \exp\left\{-\frac{1}{2(T-t)}(x_j - x_i)'(x_j - x_i)\right\} \\ h_{y_i,T-t}(y_j) &= |2\pi(T-t)\Sigma|^{-1/2} \exp\left\{-\frac{1}{2(T-t)}(x_j - x_i)'R\Sigma^{-1}R'(x_j - x_i)\right\} \\ &= |2\pi(T-t)\Sigma|^{-1/2} \exp\left\{-\frac{1}{2(T-t)}(x_j - x_i)'(x_j - x_i)\right\}, \end{aligned}$$

which are equal up to the constant factor $|\Sigma|^{1/2}$. Given the latter observation and that the weights are equal in both cases, we conclude that the normalized entries $p_{ij}^{(T-t)}$ and $\tilde{p}_{ij}^{(T-t)}$ obtained through (3.2) and (3.3) are also equal. □

Remark 4.1. One may ask whether the weights g_i specified in Lemma 4.1 are indeed appropriate for the grid $\mathcal{Y}$. That is, whether (3.3) leads to a standard importance sampling procedure for $\mathcal{Y}$ with these weights. We answer this question by comparing the grid densities.

Suppose that the density g_X was used to generate the grid $\mathcal{X}$ using random sampling. So that for every $S \subset \mathbb{R}^d$,

$$P(x \in S) = \int_S g_X(x)\mathrm{d}x.$$

Applying the transformation $x \mapsto y = R'x$ leads us to conclude that the grid $\mathcal{Y}$ consists of points generated randomly from some density g_Y satisfying, for each $S \subset \mathbb{R}^d$,

$$\begin{aligned} P(y \in R'S) &= \int_{R'S} g_Y(y)\mathrm{d}y \\ &= \int_S g_Y(R'x)||R||\mathrm{d}x \end{aligned} \tag{4.11}$$

by the multivariate substitution formula where $R'S = \{R'x : x \in S\}$ and $||R|| = \mathrm{abs}(\det R)$. But since $y = R'x$, $P(y \in R'S) = P(x \in S)$, and since (4.11) holds for all $S \subset \mathbb{R}^d$, we conclude that

$$g_X(x) = ||R||g_Y(R'x). \tag{4.12}$$

Finally, note that the averaging taking place in (3.3) implies that the weights $g_X(x)$, being proportional to $g_Y(R'x)$, are appropriate for importance sampling with respect to the grid $\mathcal{Y}$.

Remark 4.2. A time-dependent scaling of the covariance can also be incorporated by using the matrix logarithm, and constructing the time stepping matrices through (3.7) and (3.8) rather than (3.10) and (3.11).

4.5. Low-biased estimate

As is common practice in the American option pricing literature, a low-biased estimate may be obtained by taking an exercise rule implied by the pricing method and determining the expected value of using this rule using out-of-sample paths.

A natural approximation to the optimal exercise rule using the pricing results of the irregular grid method is to take the implied rule of the nearest neighbor in the grid at the closest time. Specifically one may define the exercise rule for grid points to be

$$\tau(x_i, t_k) \triangleq \begin{cases} 1 & \text{if } v_i^{(k)} = \psi_i \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

and for general points $x \in \mathbb{R}^d$

$$\tau(x, t) \triangleq \begin{cases} 1 & \text{if } v_i^{(k)} = \psi_i \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

where $k = \operatorname{argmin}_j |t - t_j|$ and $i = \operatorname{argmin}_j \left\{ \|x - x_j\| : x_j \in \mathcal{X}_k \right\}$.

This rule is easily implemented and can also be adapted to the case where we have several different grids. In this case we could base the exercise rule on a vote between grids. We could also implement weighted schemes with respect to x and t rather than using nearest neighbor rules.

4.6. High-biased estimate

Whereas applying an exercise rule to out-of-sample paths leads to a low-biased estimate of the option value, simulating the cost of a hedging strategy leads to a high-biased estimate. The latter may be seen as follows: the optimal hedging strategy enables the seller of the option, given a cash amount equal to the value of the option at the initial time t, to perfectly reproduce the payoff. A suboptimal strategy, however, will on average require a larger initial cash amount, thus the cost of a suboptimal hedging strategy is on average higher than the true option value.

A formal demonstration can be given in terms of the dual formulation for American option pricing (see [15,16]) in which one minimizes the cost of hedging by minimizing an objective function over martingales. Since the value of our hedging strategy is a martingale, it corresponds in general to a suboptimal martingale, and thus a high-biased estimate.

In practice, obtaining an upper bound in the way we suggest requires knowledge of the optimal exercise rule. Since we only have an estimate of this, the cost of the hedging strategy may not be purely upward biased. We find however that one can approximate the optimal exercise strategy far more accurately than one can approximate the optimal hedging strategy. We shall see in Section 5 that experimental results support this statement.

In the literature on American options there is little said about the practicalities of hedging in a high dimensional setting. The difficulty with using an approach such as LSM is that the method does not naturally form approximations to the value function from which derivatives can be estimated. One can form a hedging strategy by evaluating prices at states perturbed in each underlying; this demands the calculation of many additional option prices at each time step, each calculation being expensive in a high-dimensional setting. Furthermore one must be very careful with partial derivative estimates obtained from differencing stochastic point estimates; in particular the point estimates must be sufficiently accurate and the perturbations must be well chosen with respect to the (unknown) curvature of the value function.

A solution provided by the irregular grid method involves estimates of the price not only at the current state, but at all states in the grid. This allows one to extract derivative estimates using value information from nearby points in

the grid; for example, using partial derivatives implied by a local linear regression. Indeed the irregular grid method provides derivative information as a by-product.

4.7. Benchmarks

There are few benchmark results for high-dimensional American options. Broadie and Glasserman [5] provide 90% confidence intervals for American call options on the maximum of five assets with nine exercise opportunities and the geometric average of five and seven assets with ten exercise opportunities using their stochastic mesh method. Longstaff and Schwartz [7] price the Broadie and Glasserman maximum options using the LSM method.

Stentoft [11] uses the binomial method of Boyle et al. [29] and the LSM method to price put options on the arithmetical average, geometrical average, maximum and minimum of three and five assets. Broadie and Glasserman [4] and Fu et al. [6] provide benchmark results for options over five assets with three exercise opportunities. Finally, Rogers [15] and Haugh and Kogan [16] use the dual formulation to price a number of different American options.

A useful result involving options on the geometrical average of several assets is that this problem can be easily reduced to an option pricing problem in one dimension. Suppose that the risk-neutral dynamics in the log domain are given by (4.1), and the payoff function

$$\psi(s) = \left(K - \left(\prod s_i\right)^{1/d}\right)^+ \tag{4.15}$$

where K is the strike price and d is the number of assets. Then using Itô's lemma one finds that the price is the same as that of a vanilla put on the asset with log price Y where $Y(t) = \frac{1}{d}\sum_{i=1}^{d} X_i(t)$ and

$$\mathrm{d}Y(s) = \frac{1}{d}\sum_{i=1}^{d} \mathrm{d}X_i(s) \tag{4.16}$$

$$= \tilde{\mu}\mathrm{d}s + \tilde{\sigma}\mathrm{d}W(s). \tag{4.17}$$

The parameters of the diffusion are given by

$$\tilde{\mu} = r - \frac{1}{2d}\sum_{i=1}^{d}\sigma_i^2 \tag{4.18}$$

$$\tilde{\sigma}^2 = \frac{1}{d^2}\sum_{i=1}^{d}\left(\sum_{j=1}^{d} R_{ij}\right)^2. \tag{4.19}$$

Using this we find that an accurate price for the geometric average European option in the Stentoft setting is 1.159, and the Bermudan and American prices are 1.342 and 1.355 respectively. Note that the difference in early exercise premium between the Bermudan, which allows ten exercise opportunities, and American prices is about 6%.

5. Experimental results

Our experiments are conducted in a Matlab environment and are based on the five-dimensional examples of Stentoft [11]. Specifically we consider five stock processes driven by correlated Brownian motions for put options with four different payoff functions. The method we use for valuation is that of Section 3. Our programs are mostly script-based but some computationally intensive routines, for example, the PSOR code, have been implemented in C.

We are given initial stock prices $S_i(0) = 40$ for each i, the correlations between log stock prices are $\rho_{ij} = 0.25$, $i \neq j$, and volatilities are $\sigma_i = 0.2$ for all i, the risk-free interest rate is fixed at $r = 0.06$, the expiry is $T = 1$ and we use $K = 10$ time steps.

We generate 50 RQMC normal grids as detailed in Section 3.4 using the parameter values $\alpha = 1, 1.5, 2$ respectively (these were found to give the best rates of convergence). The number of grids need not be so high in practice, depending on the accuracy required. The vector of initial stock prices x_0 was always included in the grid.

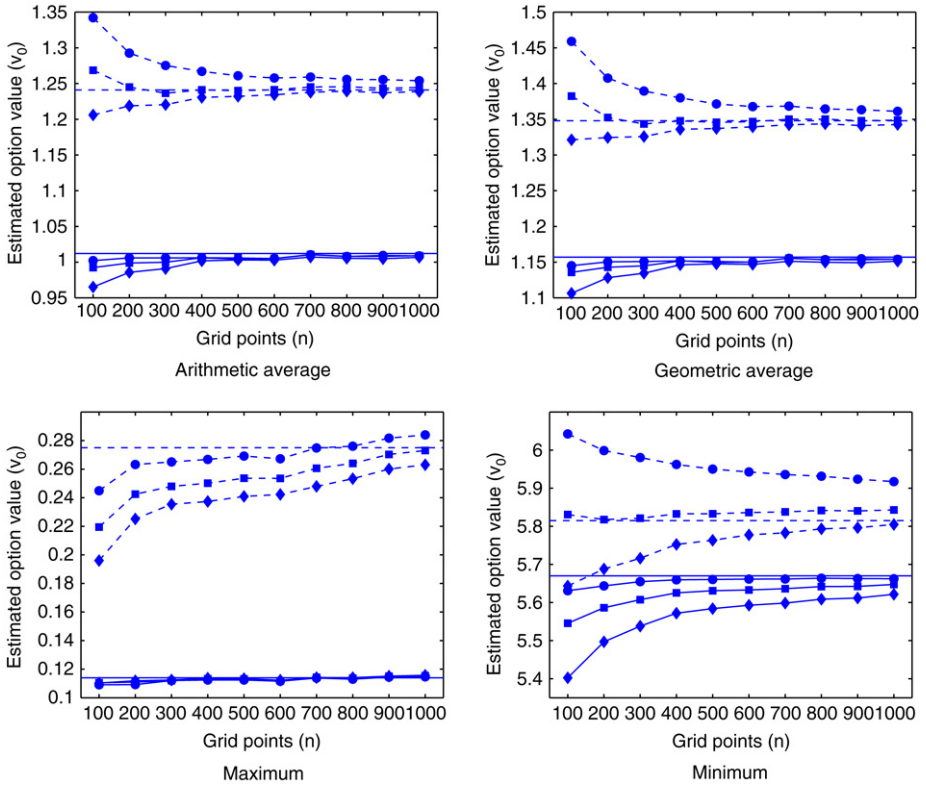

Fig. 5.1. Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options over five assets using the explicit method ($\theta = 0.0$) and ten time steps. Stentoft's Bermudan LSM solutions are drawn as horizontal lines.

The payoff functions considered correspond to put options on the arithmetical mean, geometrical mean, maximum and minimum respectively,

$$\begin{aligned} \psi_1(s) &= \left(K - \frac{1}{d}\sum s_i\right)^+ \quad \psi_2(s) = \left(K - \left(\prod s_i\right)^{1/d}\right)^+ \\ \psi_3(s) &= (K - \max(s_i))^+ \quad \psi_4(s) = (K - \min(s_i))^+ \end{aligned} \tag{5.1}$$

where x^+ denotes the positive part of x.

Figs. 5.1–5.3 show the convergence behavior of the irregular grid method where the implicitness parameter is $\theta = 0, \frac{1}{2}, 1$ respectively, and for grid sizes up to 1000. For the constrained solutions, we see that convergence is usually fastest for $\alpha = 1.5$, the algorithm reaching a fairly stable value for $n = 1000$ for all but the maximum option.

The solutions for the arithmetical and geometrical average options appear to converge to Stentoft's solutions for Bermudan options with ten exercise opportunities in the explicit case. For the Crank–Nicolson and implicit cases, the solutions appear to converge to a higher value.

The previous observation may be explained as follows. In the explicit case, our method calculates the price of a Bermudan option with ten exercise opportunities, just as in the case of Stentoft (provided we use ten equal time steps). This is because the explicit formulation takes the maximum of the intrinsic and continuation values at each exercise opportunity, and because we use exactly ten time steps. One can still see this Bermudan price as an approximation to the true American price, which we calculated previously to have an early exercise premium approximately 6% higher than the Bermudan price in the case of the geometrical average put option over five assets.

In the Crank–Nicolson and implicit cases, however, we cannot interpret the solution as approximating a Bermudan option due to the implicitness of the formulation. We can only say that as $\delta t \to 0$, the solution should converge to the American price. In the Crank–Nicolson case we suspect that the convergence is faster than in the implicit case (drawing a parallel with the unconstrained problem), and so we can think of our Crank–Nicolson solution as being

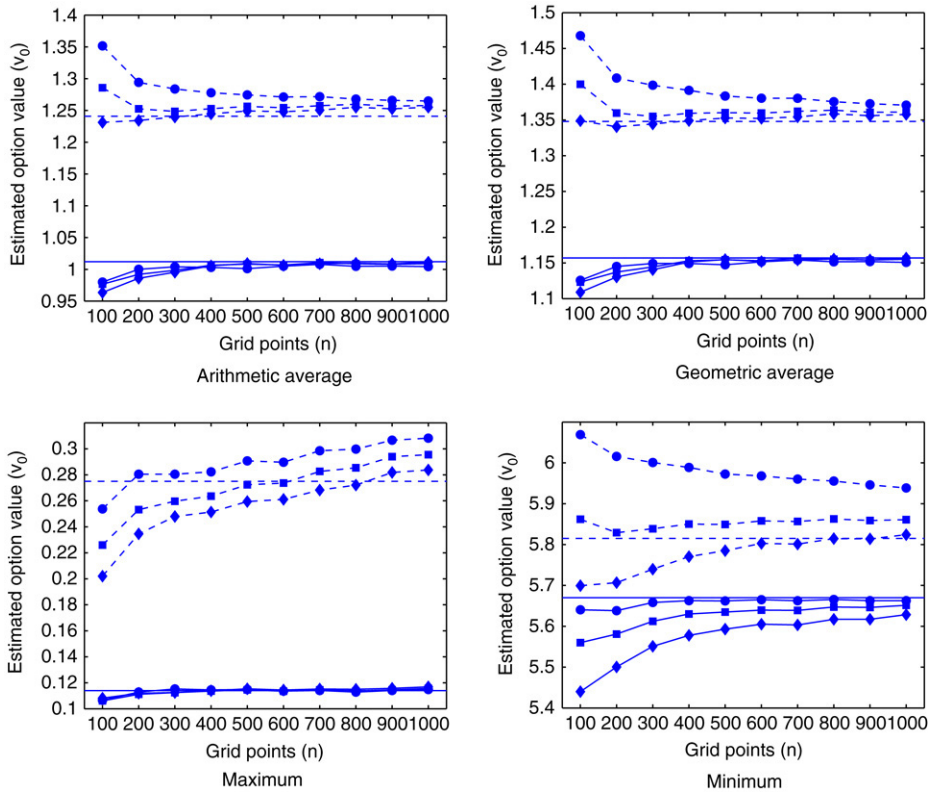

Fig. 5.2. Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options over five assets using the Crank–Nicolson method ($\theta = 0.5$) and ten time steps. Stentoft's Bermudan LSM solutions are drawn as horizontal lines.

close to our best-possible approximation to the true American option value, given that we use ten equal time steps. We thus stress that the convergence of the Bermudan price does not require $\delta t \to 0$, but the convergence of the American price does.

The fastest convergence rate in Figs. 5.1–5.3 is achieved with α (the ratio of grid density to process density) being 1.5. We thus present in Tables 5.1 and 5.2 some results and comparisons for Bermudan and American option prices respectively using the irregular grid method with a normal grid and $\alpha = 1.5$. Given the previous discussion, we take our explicit solutions to be approximations to the Bermudan problem, and the Crank–Nicolson solutions to be approximations for the American problem.

Tables 5.1 and 5.2 also show out-of-sample results for LSM and the irregular grid methods. These are estimates of the expected value, under the risk-neutral measure, of using the implied exercise strategy. We implement the LSM method ourselves, as specified in [11], to obtain out-of-sample values for the LSM algorithm (these results are not given in [11]). Our LSM implementation also reproduced (up to a statistically insignificant difference) the in-sample LSM results given in [11]. For details of how out-of-sample paths are used in the LSM method to obtain low-biased estimators, we refer the reader to [7].

We remark that the values obtained from the irregular grid method are higher than those produced by the LSM algorithm, although this is not statistically significant except in the case of the minimum option. The OS (out-of-sample) results are also higher for all but the maximum option.

For the more problematic cases of the maximum and minimum options, we see that convergence is slower. In the case of the maximum it is not clear with 1000 grid points what an appropriate estimate should be. It is also not clear whether the convergence in our case for the explicit method agrees with the value obtained by Stentoft. These are cases where the grid could be adapted to the payoff function as well as to the process itself; such extensions are left for future investigation.

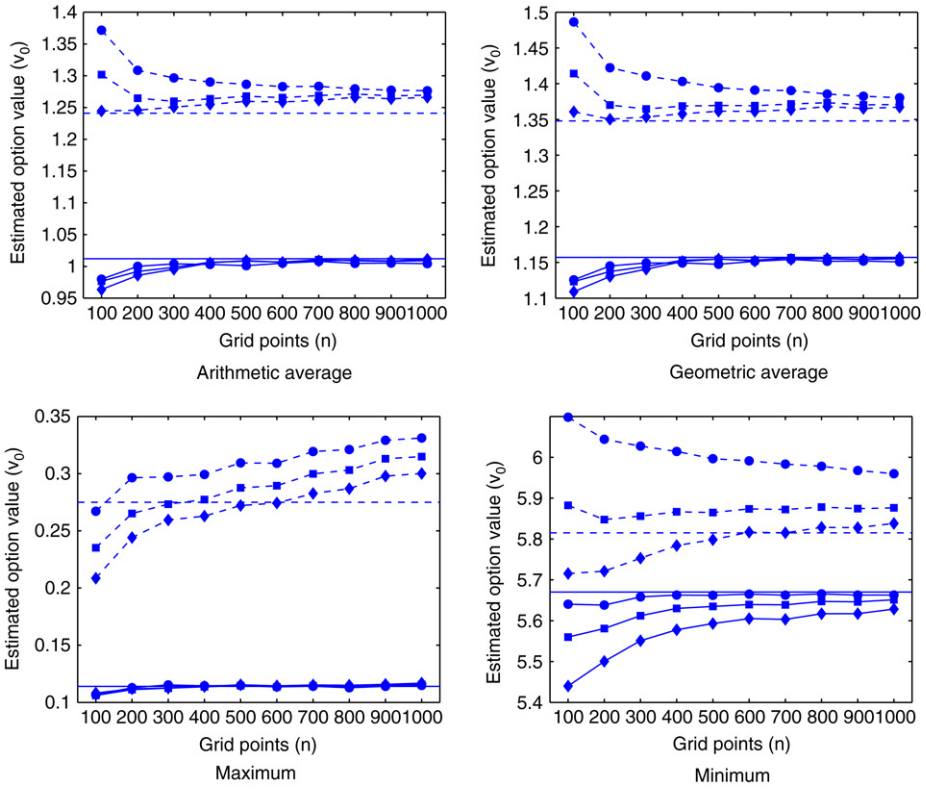

Fig. 5.3. Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options over five assets using the implicit method ($\theta = 1.0$) and ten time steps. Stentoft's Bermudan LSM solutions are drawn as horizontal lines.

Table 5.1
Comparison of Bermudan price estimates ($\theta = 0$) with ten exercise opportunities

Option type	Exact	Binomial	LSM	LSM OS (low-biased)	Normal grid	Normal grid OS (low-biased)
Arith. average	–	1.235	1.241 (0.0006)	1.231 (0.0006)	1.246 (0.004)	1.238 (0.005)
Geom. average	1.342	1.340	1.348 (0.0006)	1.335 (0.0007)	1.350 (0.004)	1.345 (0.005)
Maximum	–	0.230	0.275 (0.0004)	0.268 (0.0004)	0.276 (0.008)	0.233 (0.002)
Minimum	–	5.841	5.815 (0.0012)	5.816 (0.0014)	5.847 (0.009)	5.821 (0.013)

The grid estimates are the average price taken over 50 normal RQMC grids with size 1000, with $\alpha = 1.5$ and using ten time steps. The binomial method of Boyle et al. [29] was used with Richardson extrapolation. The OS (out-of-sample) columns give the value of the exercise strategy implied by the 50 grid solutions, calculated by taking the mean value over 100,000 simulated paths. The binomial and LSM prices are given in [11] and the OS prices for LSM are computed by running the LSM method 20-times each with 100,000 out-of-sample paths. The exact price given in the first column is the numerical solution to the equivalent one-dimensional problem. Standard errors are shown in brackets.

In Table 5.1 it is encouraging to see that the irregular grid prediction for the geometrical average option is very accurate as compared to the benchmark. The exercise strategy performs well for the arithmetical and geometrical average options, but not for the more problematic maximum and minimum payoffs.

Table 5.2
Comparison of American price estimates ($\theta = 0.5$)

Option type	Exact American	Normal grid American	Normal grid American OS (low-biased)	Hedged American OS (high-biased)
Arith. average	–	1.257 (0.004)	1.243 (0.004)	1.363 (0.004)
Geom. average	1.355	1.360 (0.004)	1.348 (0.005)	1.462 (0.004)
Maximum	–	0.295 (0.009)	0.267 (0.002)	0.504 (0.006)
Minimum	–	5.862 (0.009)	5.789 (0.012)	6.355 (0.010)

The grid estimates in the third and fourth columns are the average price taken over 50 normal RQMC grids with size 1000, with $\alpha = 1.5$ and using ten time steps. The OS (out-of-sample) column gives the estimated value of the implied exercise strategy, calculated by taking the mean value over 100,000 simulated paths and using 50 time steps. The hedged column gives the average cost of the hedging strategy obtained as a by-product of a single price computation; it is implemented using the results of a single grid solution, 50 time steps and computes the hedge as detailed in Section 4.6. In particular note that we have used a different time step in the OS exercise and hedging simulations than in computing the grid solutions. The exact price given in the first column is the numerical solution to the equivalent one-dimensional problem. Standard errors are shown in brackets.

Table 5.3
Timings and storage requirements for the irregular-grid method using Matlab 6.1 with a PIII 866 MHz processor with 512 MB RAM, matrix entries stored in double precision (8 bytes per entry)

Size P (n)	Memory full (MB)	Memory sparse (MB)	Time for P (s)	Time for $P^{1/10}$ (s)	Prop $P^{1/10}$ nonzero	Time-stepping (s)
500	2.0	0.6	1	22	0.190	0.5
1000	8.0	1.8	5	200	0.147	1.3
1500	16.0	3.3	12	750	0.123	2.0
2000	32.0	5.1	22	2000	0.106	2.9
2500	50.0	7.1	37	4000	0.094	3.8
3000	72.0	9.1	55	7200	0.084	4.9

The sparse matrices are formed by eliminating all entries smaller than 5×10^{-4} and renormalizing. The time stepping column gives the total time to complete 10 time steps, using the sparse matrix and the explicit method. Note that sparse matrices were not used for any experiments in this paper, the information provided rather serves to illustrate the complexity of the method as n increases. The second-to-last column gives the proportion of nonzero entries in the sparse matrices, an important consideration for computational complexity. Note that MB denotes 10^6 bytes in this context.

As detailed in Section 4.6, our method yields a hedging strategy as a by-product; thus simulation of a hedging strategy can be done quickly and efficiently. Using the implied hedging strategy of a single grid, and taking 20 nearest neighbors for the delta estimation, we obtain the results shown in the last column of Table 5.2. It is clear that the hedging errors are much larger than the exercise errors; this may be expected given that the exercise rule is a function having only two possible values, whereas the hedging rule takes values in $\mathbb{R}^d$. The hedging strategy used is naive in that the results of only a single-grid solution are used. It could probably be improved by using information from different grid solutions.

The most time-consuming operation in the irregular-grid method is the computation of the matrix root. Some timings for computing matrix roots in Matlab 6.1 on a PIII 866 MHz machine are presented in Table 5.3. It should be noted that the time does not depend strongly on the order of the root, so that square root and tenth root operations, for example, take about the same amount of time. The time taken for the construction of the matrix $P^{(T-t)}$ is seen to be small compared to the root operation.

Although the matrix root operation is time consuming for large values of n, it should be noted that once a root has been computed for a single normally distributed grid, it can subsequently be used for valuing options on a large class of diffusion processes with arbitrary payoff functions without the need for recomputation.

6. Conclusions

We have proposed a new method for finding the value of American and Bermudan options in a high-dimensional setting. Central to this method is the use of an irregular grid over the state space and an approximation of the partial differential operator on this grid.

In our analysis we allow any grid which is generated using MC or QMC trials with respect to a known density function. Once the Markov chain approximation has been obtained, we use the transition probability matrix to form a semidiscrete approximation to the partial differential operator corresponding to this Markov chain. This is done through taking a logarithm of the transition probability matrix; however solving the fully discrete problem only requires computing a certain root of the matrix related to the time step and implicitness parameters, at the cost of an extra approximation error.

An important aspect of the proposed method is the absence of any requirement to specify basis functions for approximating the value function or exercise strategy. Indeed the only specification needed is a grid density, although asymptotically even this choice is not critical. Furthermore, convergence in the Bermudan case should require asymptotics in only one parameter, namely the number of grid points. In the American case one also requires $\delta t \to 0$. These aspects set the root method aside from the LSM method where the specification of basis functions plays a critical role in the success of the method, and convergence involves asymptotics in two parameters in the Bermudan case, namely the number of basis functions and the number of paths, these two parameters producing opposite biases.

The irregular grid solution gives price estimates at all points in the grid. This is useful if one requires partial derivative information, for example, when hedging. Partial derivatives can be easily estimated from the solution by preforming a linear regression using values from neighboring points.

Our experiments suggest that the irregular grid method has very good convergence properties, especially when the grid density is related to the density of the process itself. In particular, the grid density should have a larger variance than the process; for a geometric Brownian motion process in five dimensions it was found that a ratio of 1.5 gave a good rate of convergence, although (slower) convergence was also observed for ratios of 1.0 and 2.0. Convergence of estimates for the maximum option was not clear with grids of up to 1000 points.

The numerical results obtained largely agree with those of Stentoft [11]. We find that the early exercise premium is increased by about 6% for the examples he considers when allowing a continuum of exercise opportunities rather than only ten. We also find that the exercise strategies implied by the LSM method produce significantly lower values (statistically) than the LSM price implies, except in the case of the minimum option; this is an indication that out-of-sample paths should be used in simulation methods — in this way the price obtained corresponds directly to the average value of the implied exercise strategy. This suggests that one should be careful in higher dimensions when applying the recommendation of Longstaff and Schwartz [7] to save time by only using in-sample paths.

A possible variance reduction technique is to adjust the transition probabilities according to the empirical density of the grid points rather than the density used for generation of the grid. Adjustment may be done after constructing the transition matrix, for example, using quadratic programming to improve local consistency in the sense of Kushner and Dupuis [25], but may also take inspiration from the literature on nonparametric analysis. These and other possible refinements are reserved for future investigation.

For further reading

[1].

Acknowledgement

The first author's research was supported by the Netherlands Organization for Scientific Research (NWO).

References

[1] S.J. Berridge, J.M. Schumacher, An irregular grid method for high-dimensional free-boundary problems in finance, Future Generation Computer Systems 20 (2004) 353–362.

[2] J. Barraquand, D. Martineau, Numerical valuation of high dimensional multivariate American securities, Journal of Financial and Quantitative Analysis 30 (1995) 383–405.

[3] P. Boyle, M. Broadie, P. Glasserman, Monte Carlo methods for security pricing, Journal of Economic Dynamics and Control 21 (1997) 1267–1321.
[4] M. Broadie, P. Glasserman, Pricing American-style securities using simulation, Journal of Economic Dynamics and Control 21 (1997) 1323–1352.
[5] M. Broadie, P. Glasserman, A stochastic mesh method for pricing high-dimensional American options, Journal of Computational Finance 7 (2004) 35–72.
[6] M.C. Fu, S.B. Laprise, D.B. Madan, Y. Su, R. Wu, Pricing American options: A comparison of Monte Carlo simulation approaches, Journal of Computational Finance 4 (2001) 39–88.
[7] F.A. Longstaff, E.S. Schwartz, Valuing American options by simulation: A simple least squares approach, Review of Financial Studies 14 (2001) 113–147.
[8] E. Clément, D. Lamberton, P. Protter, An analysis of a least squares regression method for American option pricing, Finance and Stochastics 6 (2002) 449–471.
[9] L. Stentoft, Convergence of the least squares Monte-Carlo approach to American option valuation, Management Science 50 (2004) 1193–1203.
[10] P. Glasserman, B. Yu, Number of paths versus number of basis functions in American option pricing, Annals of Applied Probability 14 (2004) 2090–2119.
[11] L. Stentoft, Assessing the least squares Monte-Carlo approach to American option valuation, Review of Derivatives Research 7 (2004) 129–168.
[12] M. Moreno, J.F. Navas, On the robustness of least-squares Monte Carlo (LSM) for pricing American derivatives, Review of Derivatives Research 6 (2003) 107–128.
[13] J.N. Tsitsiklis, B. Van Roy, Regression methods for pricing complex American-style options, IEEE Transactions on Neural Networks 12 (2000) 694–703.
[14] P.P. Boyle, A.W. Kolkiewicz, K.S. Tan, An improved simulation method for pricing high-dimensional American derivatives, Mathematics and Computers in Simulation 62 (2003) 315–322.
[15] L. Rogers, Monte Carlo valuation of American options, Mathematical Finance 12 (2002) 271–286.
[16] M.B. Haugh, L. Kogan, Pricing American options: A duality approach, Operations Research 52 (2004) 258–270.
[17] F. Jamshidian, Minimax optimality of Bermudan and American claims and their Monte-Carlo upper bound approximation, Working Paper, NIB Capital Bank.
[18] A. Kolodko, J. Schoenmakers, Upper bounds for Bermudan style derivatives, Monte Carlo Methods and Applications 10 (2004) 331–343.
[19] P. Jaillet, D. Lamberton, B. Lapeyre, Variational inequalities and the pricing of American options, Acta Applicandae Mathematicae 21 (1990) 263–289.
[20] R. Glowinski, J. Lions, R. Trémolières, Numerical Analysis of Variational Inequalities, North-Holland, Amsterdam, 1981.
[21] P. Glasserman, Monte Carlo Methods in Financial Engineering, Springer, New York, 2003.
[22] S. Berridge, J. Schumacher, Pricing high-dimensional American options using local consistency conditions, in: John Miller, David Edelman, John Appleby (Eds.), Numerical Methods for Finance, Chapman & Hall/CRC, 2007.
[23] C.W. Cryer, The solution of a quadratic programming problem using systematic overrelaxation, SIAM Journal of Control 9 (1971) 385–392.
[24] M. Dempster, J. Hutton, Pricing American stock options by linear programming, Mathematical Finance 9 (1999) 229–254.
[25] H.J. Kushner, P.G. Dupuis, Numerical Methods for Stochastic Control Problems in Continuous Time, Springer-Verlag, New York, 1992.
[26] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, 1996.
[27] A.B. Owen, Randomly permuted (t, m, s)-nets and (t, s)-sequences, in: H. Niederreiter, P. J.-S. Shiue (Eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, 1995, pp. 299–317.
[28] M. Evans, T. Swartz, Approximating Integrals via Monte Carlo and Deterministic Methods, Oxford University Press, Oxford, 2000.
[29] P.P. Boyle, J. Evnine, S. Gibbs, Numerical evaluation of multivariate contingent claims, Review of Financial Studies 2 (1989) 241–250.