No. 2002-99

AN IRREGULAR GRID APPROACH FOR PRICING HIGH-DIMENSIONAL AMERICAN OPTIONS

By S.J. Berridge, J.M. Schumacher

October 2002

ISSN 0924-7815

Discussion paper

An irregular grid approach for pricing high-dimensional American options

S.J. Berridge* and J.M. Schumacher
Department of Econometrics and Operations Research
and Center for Economic Research (CentER),
Tilburg University, Postbus 90153, 5000 LE Tilburg, The Netherlands.
{s.j.berridge,j.m.schumacher}@kub.nl

October 11, 2002

Abstract

We propose and test a new method for pricing American options in a high-dimensional setting. The method is centred around the approximation of the associated variational inequality on an irregular grid. We approximate the partial differential operator on this grid by appealing to the SDE representation of the stock process and computing the logarithm of the transition probability matrix of an approximating Markov chain. Experimental results in five dimensions are promising.

Keywords: American options, high-dimensional problems, free boundary problems, optimal stopping, variational inequalities, numerical methods, unstructured mesh, Markov chain approximation

MSC 2000: 35R35, 60G40, 65D15, 90C33

JEL Codes: C15, C61, C63

*Research support by the Dutch National Science Foundation (NWO)

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.

1.1 Literature

Barraquand and Martineau [1] are perhaps the first to consider pricing high-dimensional American options specifically. They propose an algorithm based on aggregation of paths with respect to the intrinsic value, which was an improvement over the earlier Tilley [20] method which required aggregation with respect to the state. The method is difficult to analyse and has a possible lack of convergence; Boyle et al. [2] demonstrate this and propose a modification of the algorithm which leads to a low-biased estimator.

Broadie and Glasserman [5] 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, but obtaining estimates for an increasing number of exercise opportunities sometimes allows one to use Richardson extrapolation to find a more accurate estimate. In a subsequent working paper [6] 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. [9]).

The "simple yet powerful" least squares Monte Carlo (LSM) method of Longstaff and Schwartz [14] 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. This is done for in-the-money paths only since out-of-the-money paths should never be exercised. Experimental success is reported for the LSM method, although in high dimensions the basis functions must be chosen carefully. Recently Clément et al. [7] and Stentoft [19] independently provide proofs of convergence for the LSM method, showing that the convergence is $\sqrt{n}$ in the

number of paths used. The convergence behaviour in the number of basis functions however has not been determined. Stentoft [18] and Moreno and Navas [15] test the LSM algorithm numerically. Stentoft suggests that basis functions up to order three are sufficient in five dimensions for arithmetic and geometric average options, but not for minimum or maximum options. Moreno and Navas find that the method is not robust to the choice of basis functions in five dimensions.

Tsitsiklis and Van Roy [22] 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. [4] recently extended the stochastic mesh method of Broadie and Glasserman [6] 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 [17] and Haugh and Kogan [11]. They use a dual formulation of the problem in which a minimisation is performed over martingales. The method is sensitive to the choice of basis martingales chosen to perform the minimisation, and so requires the basis to be well-chosen in order to give an accurate solution. The method gives a high-biased estimator.

The work of Glowinski et al. [10] on variational inequalities is also interesting to note in this context, providing a firm footing for analysing the American option pricing problem in complete markets. Jaillet et al. [12] consider American options using this framework, providing a convergence proof for the Brennan-Schwartz algorithm[1] in one dimension.

1.2 Our approach

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 PDEs.

We first perform a discretisation of the state space using QMC trials with respect to an importance sampling distribution related to the transition density of the process at expiry. We then propose an approximation to the

[1] constrained explicit finite difference method

partial differential operator (PDO) 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 starting from the option expiry.

In our method the matrix logarithm of $P^{(T-t)}$ does not need to be calculated explicitly, but instead a root of the matrix must be calculated. 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. Traub et al. [21] show this and also study the theoretical efficiency of many other problems in high dimensions.

2 Formulation

2.1 The market

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

$$dX(s) = \mu(X(s), s)ds + \sigma(X(s), s)dW(s) \tag{1}$$

with initial condition $X(t) = x_t$, and a derivative product on $X(s)$ with exercise 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

$$dV(s) = \mu_V(X(s), s)ds + \sigma_V(X(s), s)dW(s) \tag{2}$$

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 $v(\cdot, T) = \psi(\cdot, T)$.

The objective is to determine the current value $v(x_t, t)$ of the derivative product and the accompanying adapted exercise and hedging strategies τ and H:

$$\tau : \mathbb{R}^N \times [t, T] \rightarrow \{0, 1\} \tag{3}$$

$$H : \mathbb{R}^N \times [t, T] \rightarrow \mathbb{R}^N \tag{4}$$

Supposing that one has an estimate $\hat{V}(t)$ of the derivative price, one must also provide an exercise rule $\hat{\tau}$ or a hedging strategy $\hat{H}$ in order for the buyer or seller respectively to be able to realise the estimated price.

2.2 Pricing

2.2.1 The primal formulation

The value of the derivative product is formulated in the primal problem as a supremum over stopping times

$$v(x_t, t) = \sup_{\tau \in \mathcal{T}} \mathbb{E}^{\mathbb{Q}}_{x_t} \left(e^{-r(\tau - t)} \psi(X(\tau)) \right) \tag{5}$$

where $\mathcal{T}$ is the set of stopping times on $[t, T]$ with respect to the natural filtration, the expectation is taken with respect to the risk-neutral measure $\mathbb{Q}$, and the initial value is x_t.

2.2.2 The dual formulation

The dual formulation (see Rogers [17] or Haugh and Kogan [11]) forms a price by minimising the cost of the hedging strategy over equivalent martingale measures. Theorem 1 of [17] implies that the price is given by

$$v(x_t, t) = \inf_{M \in H_0^1} \mathbb{E}^{\mathbb{Q}}_{x_t} \left[\sup_{s \in [t,T]} \left(e^{-r(s-t)} \psi(X(s)) - M(s) \right) \right] \tag{6}$$

where H_0^1 is the space of martingales with $M(0) = 0$ and $\sup_{s \in [t,T]} |M(s)| \in L^1$. The infimum is attained by taking $M = M^*$.

2.2.3 The variational inequality formulation

Formulating the problem as a variational inequality invites implications from the large number of results that have been developed in this field, for example the work of Glowinski et al. [10]. Jaillet et al. [12] applied this approach to the analysis of American option pricing.

One must first define an elliptic operator $\mathcal{L}$ giving the diffusion of the process. This is given by

$$\mathcal{L} = \tfrac{1}{2} \mathrm{tr} \sigma \sigma' \frac{\partial^2}{\partial x^2} + (\mu - \sigma \lambda) \frac{\partial}{\partial x} - r \tag{7}$$

where r is the riskfree rate and $\lambda : [t,T] \times \mathbb{R}^N \to \mathbb{R}^N$ is the function satisfying

$$\mu_V - rv = \sigma_V \lambda. \tag{8}$$

One must also specify a function space in which to work. Briefly one defines an inner product $\langle \cdot, \cdot \rangle$ and a bilinear form $a(\cdot, \cdot)$ on the Hilbert space H^1 satisfying

$$a(v,u) = \langle u, \mathcal{L}v \rangle \tag{9}$$

The equivalent variational inequality formulation is then to find $v(x,t)$ such that

$$\begin{cases} v(x,s) - \psi(x,s) & \geqslant & 0 \\ u \geqslant \psi \text{ a.e.} \ \Rightarrow a(v, u-v) + \left\langle u - v, \frac{\partial v}{\partial t} \right\rangle & \geqslant & 0 \qquad \text{a.e. } [t,T] \end{cases} \tag{10}$$

for $(x,s) \in \mathbb{R}^N \times [t,T]$ with the terminal condition $v(\cdot,T) \equiv \psi(\cdot,T)$.

2.2.4 The complementarity formulation

The variational inequality formulation is not amenable to computation; fortunately it can be reformulated as a complementarity problem. Let $\mathcal{L}$ be the related diffusion operator; then the option value is found by solving the complementarity problem

$$\begin{cases} \frac{\partial v}{\partial t} + \mathcal{L}v & \leqslant & 0 \\ v - \psi & \geqslant & 0 \\ \left(\frac{\partial v}{\partial t} + \mathcal{L}v\right)(v - \psi) & = & 0 \end{cases} \tag{11}$$

for $(x,s) \in \mathbb{R}^N \times [t,T]$ with the terminal condition $v(\cdot,T) \equiv \psi(\cdot,T)$.

2.3 Consequences

In solving the pricing problem we divide the time-state space into two complementary regions: the continuation region where it is optimal to hold the option and the stopping region where it is optimal to exercise. In the continuation region the first line of (11) is active and the stopping rule says not to exercise. In the stopping region the second line of (11) is active and the stopping rule says to exercise.

In all formulations presented, high dimensionality poses a practical problem since functional approximation in a high-dimensional space is called for.

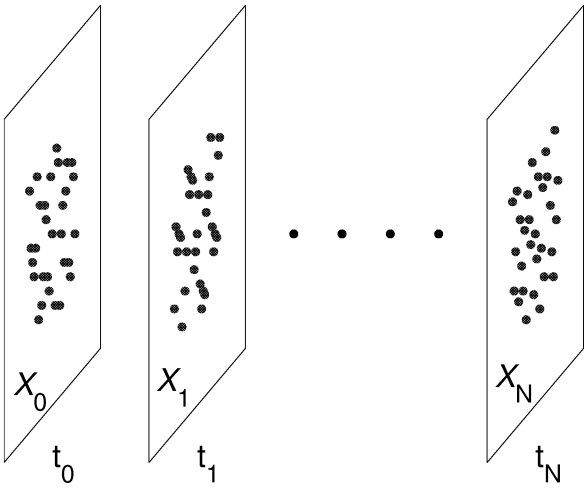

Figure 1: Generic grid framework for American option pricing schemes

3 Framework for pricing American options

If we are to solve the American option pricing problem numerically, it seems reasonable that we should first try to reduce the dimensionality of the problem. This is the approach taken by numerical methods for solving PDEs. For example finite difference methods approximate the value function at certain times and a finite number of states, and finite element methods use a finite number of localised basis functions.

We will approach the pricing problem from the finite difference point of view, that is we will choose a discretisation of the time-state space on which to form approximations of the value function. Traditional finite element methods, which are popular when complex geometries are involved, do not seem attractive in a high dimensional setting due to the difficulty in specifying tractable elements. The sparse grid literature initiated by Zenger [24] is nonetheless encouraging in this context.

We also emphasise that the solution to the American option pricing problem is not complete without an accompanying set of rules. For the holder of an option there must be an exercise rule and for the seller there must be a hedging strategy. Such strategies are intimately related to the resulting price.

Suppose we choose a set of timesteps $t = t_0 < \cdots < t_K = T$ and for each timestep t_k a set of states (grid) $\mathcal{X}_k = \{x_{k,1}, \ldots, x_{k,n}\} \subset \mathbb{R}^N$ on which to approximate the solution to (11). This setup is illustrated in Figure 1.

In a traditional finite difference method the grid is constant over time,

and in a finite difference adaptive grid method it changes. We can also accommodate the stochastic tree, LSM and LDM methods in this context where the grids contain realisations of sample paths at each timestep. The difference between the methods is the way in which the grids are generated and the way in which the continuation values are estimated.

In the stochastic tree method, the grid samples b paths forward in time from each node $(x_{k,m}, t_k)$, giving b^k nodes in $\mathcal{X}_k$. Continuation values are taken to be the discounted average of the observed continuation values along the paths emanating forward in time from each node. This is attractive empirically since at each node we are performing a numerical integration with b trials, but unattractive from the computational point of view.

In LSM the grids are simply generated from sample paths and the expected continuation values are estimated by regressing the discounted realised continuation values onto a set of basis functions over the current state space. The continuation values are determined from the times at which the paths are stopped, which means all future grids may be used in the calculation. This method is nice from a computational point of view, but it is not obvious how many basis functions are required as the dimension increases, or how those basis functions should be chosen.

In LDM the grids are generated using a low-discrepancy method where the density of nodes is a mixture of transition densities from each node. The continuation values are then estimated by taking a weighted average of the discounted continuation values over all paths at the next timestep.

In most methods the grid reflects the evolution of the stock process, in that more grid points are concentrated in the main support of the state transition density. Finite difference methods do not necessarily adhere to this rule since the solution region is freely specifiable only provided appropriate boundary conditions are available.

Accurate estimation of expected continuation values in LSM requires that the stopping times have been accurately calculated without using too much path-specific information. With the other methods accurate estimation relies on having enough points $x_{k+1,j}$ close to the current state $x_{t,i}$ (with respect to the transition density) to give a stable average. In the Broadie and Glasserman method this issue is addressed directly by placing b points close to the current point. In other methods using simulated paths there is clearly one point that is close, with other points being distributed according to the density of the process. Using these points to do numerical integration can be like using importance sampling with an ill-conditioned importance sampling density.

4 An Irregular Grid Method

In the framework of Section 3 we choose to work with a constant grid $\mathcal{X} \subset \mathbb{R}^N$ on which to solve the pricing problem. Traditionally constant grid methods have used regular grids, for example finite difference methods. However in a high-dimensional problem this is not practical due to the curse of dimensionality, or efficient in the sense that such a grid is not well adapted to the SDE.

Specifically we work with a grid generated using MC or QMC trials in the state space. In particular if the grid is generated with trials distributed according to the transition density of the process, it is expected that the representation of the SDE should be accurate in the region of our initial vector of prices. It is also possible in this setting to directly control the number of points in the grid.

4.1 Semidiscrete setting

Let us first consider a discretisation of the state space, leaving time continuous. In the pricing problem this amounts to approximating the complementarity problem (11) by a system of ODEs with complementarity conditions. This semidiscrete setting, or method of lines, is well suited to our choice of a constant grid in the state space.

Recall we would like to solve the problem (11). Discretisation of the state space using grid $\mathcal{X}$ requires an approximation of the PDO $\mathcal{L}$ on $\mathcal{X}$; let us call this approximation A.

We now consider the complementarity problem represented by the constrained system of ODEs

$$\begin{cases} \frac{dv}{dt}(s) + Av(s) & \leqslant & 0 \\ v(s) - \psi & \geqslant & 0 \\ \left(\frac{dv}{dt}(s) + Av(s)\right)'(v(s) - \psi) & = & 0 \end{cases} \tag{12}$$

for $s \in [t, T]$ with terminal condition $v(T) = \psi$.

The semi-discrete setting can also be seen as a Markov chain approximation to the problem; the state is now constrained to move only between discrete states rather than contiuously.

4.2 Approximating the differential operator

We now propose a method for specifiying A in (12). The method is inspired by numerical integration, and in the European case the resulting method will be equivalent to numerical integration with importance sampling.

We assume that the joint density of the process $\tilde{f}_{x_t,t}(x)$ is available for arbitrary initial points x_t and time horizons t, although in principle one could adapt the following construction to the case where the density was not known explicitly, but for example the process could be simulated. Let $g(x)$ be the density used to generate the grid $\mathcal{X}$.

Denote by $P^{(T-t)}$ the transition probability matrix between points of $\mathcal{X}$ corresponding to the horizon of the option $T-t$ where the entries $\left(P^{(T-t)}\right)_{ij}$ are given by

$$p_{ij}^{(T-t)} = \frac{1}{\sum_{k=1}^{n} f_{x_i,T-t}(x_k)} \cdot f_{x_i,T-t}(x_j) \tag{13}$$

and the weights are given by

$$f_{x_i,T-t}(x) = \frac{\tilde{f}_{x_i,T-t}(x)}{g(x)} \tag{14}$$

We note that the evolution of state probabilities in the semidiscrete setting is given by $p(s) = e^{A'(s-t)}p(t)$ where $p(s)$ is the discrete probability distribution over our grid at time s and $p(t)$ is the initial probability distribution at time t which will in most cases be a delta function (since the initial state is known). The matrix $P^{(T-t)}$ thus gives us access to an approximation A to $\mathcal{L}$ as follows:

$$A \triangleq \frac{1}{T-t} \log P^{(T-t)}. \tag{15}$$

The matrix logarithm certainly exists and is unique if $P^{(T-t)}$ is diagonalisable and has positive eigenvalues, which we find to be the case in our experiments (it is not symmetric however). We shall see that it is not actually necessary to calculate A in our solution method; however it could be calculated in principle. A was introduced for the purposes of exposition.

In the European case we see that using this approximation reconstructs the method of numerical integration with importance sampling function g:

$$\begin{aligned} v_i(t) &= \int_{\mathbb{R}^n} \psi(x) f_{x_i,T-t}(x) dx \\ &\approx \sum_{j=1}^{n} \psi(X_j) \frac{1}{\sum_{k=1}^{n} f_{x_i,T-t}(X_k)} \cdot f_{x_i,T-t}(X_j) \\ &= \sum_{j=1}^{n} \psi(X_j) p_{ij}^{(T-t)} \end{aligned}$$

where v_i is the price in state x_i at initial time t, and X_j are iid random variables with density $g(x)$. Given the similarity to numerical integration

we are guided by the literature on importance sampling in our choice of grid density. In particular, the integration problem is most efficiently solved on a grid with the same density as the process itself.

Because the European problem is purely one of integration, it is not necessary to have an approximation of the diffusion at intermediate timesteps, and thus the matrix A need not be considered.

4.3 Fully discrete setting

Let us now discretise (12) along the the time axis. As in Section 3 we work with the discretisation $t = t_0 < \cdots < t_K = T$ and a constant grid $\mathcal{X}_k \equiv \mathcal{X}$ for all k. We now denote the approximation at state i and timestep k as $v_{i,k} \simeq v_i(t_k) \simeq v(x_i, t_k)$.

As with traditional finite difference methods for PDEs we are free to choose an implicitness parameter θ when discretising the time axis. Suppose we have obtained an approximate solution for $v_{i,k+1}$ for $i = 1, \ldots, n$, we then consider backward recursion methods allowing us to calculate $v_{i,k}$ for $i = 1, \ldots, n$.

The derivative with respect to time can be approximated as $\frac{dv}{dt} \simeq \frac{v_{k+1} - v_k}{t_{k+1} - t_k}$, and there is some freedom as to whether to substitute $v_{i,k}$ or $v_{i,k+1}$ for $v_i(t)$ in the recursion. A general method is given by using $\theta v_{i,k+1} + (1-\theta)v_{i,k}$ for $\theta \in [0,1]$. Choosing $\theta = 0$ 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 the best δt convergence in solving PDEs.

We discretise the first line of (12) as

$$\frac{v^{(k+1)} - v^{(k)}}{\delta t_k} + A\left((1-\theta)v^{(k+1)} + \theta v^{(k)}\right) \leqslant 0 \tag{16}$$

where $\delta t_k \triangleq t_{k+1} - t_k$. Thus (12) becomes

$$\begin{cases} \left(I + (1-\theta)A\delta t\right) v^{(k+1)} - \left(I - \theta A \delta t\right) v^{(k)} & \leqslant & 0 \\ v^{(k)} - \psi & \geqslant & 0 \\ \left(\left(I + (1-\theta)A\delta t\right) v^{(k+1)} - \left(I - \theta A \delta t\right) v^{(k)}\right)' \left(v^{(k)} - \psi\right) & = & 0 \end{cases} \tag{17}$$

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

$$M_L = \exp\{-\theta A \delta t_k\} \tag{18}$$

$$M_R = \exp\{(1-\theta)A\delta t_k\} \tag{19}$$

The approximating complementarity problem to solve is then

$$\begin{cases} M_R v^{(k+1)} - M_L v^{(k)} & \leqslant & 0 \\ v^{(k)} - \psi & \geqslant & 0 \\ \left(M_L v^{(k)} - M_R v^{(k+1)}\right)' \left(v^{(k)} - \psi\right) & = & 0 \end{cases} \tag{20}$$

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

An error analysis of the discretisation in (16) may be undertaken along the lines of Glowinski et al. [12] on variational inequalities or that of Kushner and Dupuis [13] on stochastic control. We do not carry this out here.

Numerically we must solve a complementarity problem at each timestep, for which the PSOR method of Wilmott et al. [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 timestep. Various other methods may be used for solving (20), for example see Dempster and Hutton [8] for American option pricing using linear programming in the one-dimensional case.

It is important to note that the matrix logarithm does not have to be calculated explicitly in our method; instead we must calculate roots of the matrix $P^{(T-t)}$ corresponding to the timestep and implicitness parameters. In particular we have

$$M_L = \left[P^{(T-t)}\right]^{-\theta \delta t_k/(T-t)} \tag{21}$$

$$M_R = \left[P^{(T-t)}\right]^{(1-\theta)\delta t_k/(T-t)} \tag{22}$$

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 seems convenient.

It is clear that, given a small enough δt and large enough distance between grid points, M_L and M_R should be approximately sparse in that most elements can be set to zero without affecting the solution significantly.

Note that in the European case (20) still reconstructs numerical integration with importance sampling when M_L and M_R are specified in this way.

5 Irregular grid for multiasset options

We now present an algorithm for pricing multiasset options using the framework of the previous section.

5.1 Setting

Suppose the assets follow a correlated geometric Brownian motion where the risk-neutral dynamics in the log domain are given by

$$dX = \left(r\mathbb{1} - \delta - \frac{1}{2}\mathrm{diag}(\Sigma) \right) dt + R' dW \tag{23}$$

where r is the risk-free rate, $\mathbb{1}$ is the N-vector of ones, $\delta = (\delta_1, \ldots, \delta_N)$ is the vector of dividend rates, $\Sigma = (\rho_{ij}\sigma_i\sigma_j)$ is the covariance matrix of the Brownian motions and $R'R$ is its Cholesky decomposition. The operator $\mathcal{L}$ in this setting is just the multidimensional Black-Scholes operator given by

$$\mathcal{L} = \tfrac{1}{2} \sum_{i,j=1}^{N} \rho_{ij}\sigma_i\sigma_j \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{N} (r - \delta_i - \tfrac{1}{2}\sigma_i^2) \frac{\partial}{\partial x_i} - r. \tag{24}$$

Two grid types are discussed, namely grids with uniform and normal distributions. We will only consider grid generation using QMC trials, since this offers considerable efficiency gains over MC. However we do consider randomised QMC where a number of random QMC grids are generated using independent MC trials.

Suppose the current vector of prices is given by x_t and we wish to calculate $v(x_t, t)$. Let the evolution of log prices be given by (23) where Σ is known. Let us choose a parameter α to specify the size of the grid in relation to the density of the underlying process.

In both cases we consider the grid to be centred about

$$x_t + \left(r\mathbb{1} - \delta - \frac{1}{2}\mathrm{diag}(\Sigma) \right) (T - t) \tag{25}$$

which is the most appropriate for numerical integration of the payoff function at expiry. We would have to be careful if the risk-neutral process were convection-dominated in which case the SDE before expiry may not be well represented; a soiution for this problem is proposed in Section 5.5. The original state is included as an extra point in the grid.

The generation of $\mathcal{X} = \{x_1, \ldots, x_n\}$ is specified in terms of the Sobol' draws $\{s_1, \ldots, s_n\}$ of dimension N, which are elements of the unit cube. It is not necessary to use Sobol' points; rather any low discrepancy sequence in the N-dimensional unit cube is sufficient.

5.2 Grid generation

First we consider the case where the grid density is chosen to be uniform on the region $[a, b] \subset \mathbb{R}^N$ where

$$a = x_t + \left[r\mathbb{1} - \delta - \frac{1}{2}\text{diag}(\Sigma) - \alpha \left(\sigma_1 \cdots \sigma_n\right)' \right] (T - t) \tag{26}$$

$$b = x_t + \left[r\mathbb{1} - \delta - \frac{1}{2}\text{diag}(\Sigma) + \alpha \left(\sigma_1 \cdots \sigma_n\right)' \right] (T - t) \tag{27}$$

and $\sigma_i = \sqrt{\sigma_{ii}}$, r is the risk-free rate, δ is the vector of continuous dividend rates and $\mathbb{1}$ is the N-vector of ones. A sufficiently large value for α in this case would be 3. Note that in a high dimension the grid does not represent the process (23) well in that it is very unlikely for the paths to enter the corners of the grid, and thus the convergence rate is expected to be slow. This can be likened to numerical integration using a badly-adapted importance sampling function.

The ith grid point is set to be

$$x_i = a + s_i \otimes (b - a) \tag{28}$$

where s_i is the ith Sobol' point and $\otimes$ denotes componentwise multiplication. One of the disadvantages of using a uniform grid is that the support of the transition density is not contained in the support of the grid density; thus asymptotics also require $\alpha \to \infty$.

Second we consider the case where the grid density is normal with respective mean and covariance

$$\mu_g = x_t + \left[r\mathbb{1} - \delta - \frac{1}{2}\text{diag}(\Sigma) \right] (T - t) \tag{29}$$

$$\Sigma_g = \alpha\Sigma(T - t) \tag{30}$$

The parameter α should be at least 1 in this case for the grid to be well-adapted. In any dimension this grid specification is well-adapted to the risk-neutral process (23) so long as α is not too large and the process is not convection-dominated. The latter case is addressed in Section 5.5.

The ith grid point is set to be

$$x_i = \mu_g + R_g' \left(\Psi^{-1}(s_{i,1}) \cdots \Psi^{-1}(s_{i,N}) \right)' \tag{31}$$

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

Examples of the uniform and normal grids in two dimensions are shown in Figure 2. It should be noted that the advantage of using an irregular grid is realised in dimensions of at least three.

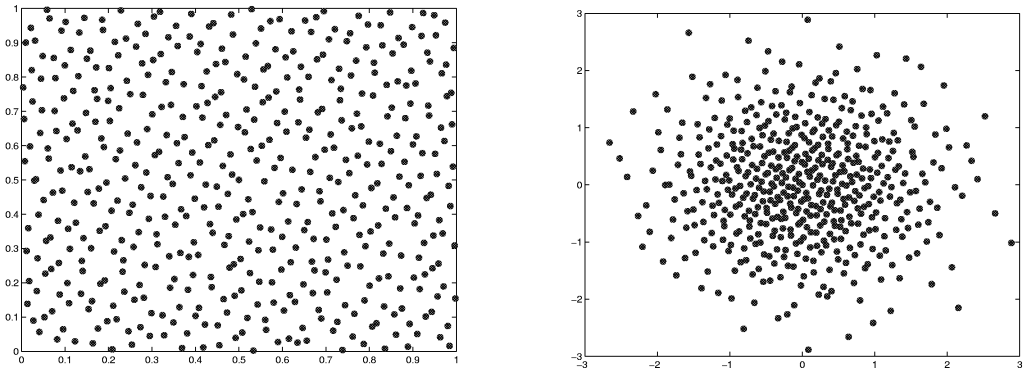

Figure 2: Examples of QMC uniform and normal grids respectively in two dimensions with 500 points

5.3 European option pricing

When pricing American options it also helps to price the European couterpart. This forms a crude lower bound for the price of an American option, and also provides a control variate for variance reduction.

For the latter purpose we compute the European value using the same grid as for the American case. This gives us access to an approximation to the early exercise premium which can be added to the European value computed using a much larger grid to obtain a variance reduced American price.

The European price can be calculated on very large grids since it requires only numerical integration. The European price is calculated through

$$v(x_t, t) = \int_{\mathbb{R}^N} \psi(x) f_{x_t, T-t}(x) dx \tag{32}$$

$$\simeq \frac{1}{n} \sum_{i=1}^{n} \psi(X_i) \tag{33}$$

where the X_i are QMC points distributed according to the density $f_{x_t,T-t}$.

5.4 Randomised QMC

The QMC grids we use are deterministic; however perturbing the Sobol' points randomly in parallel allows us to observe the behaviour of solutions for a number of different QMC grids. Thus approximate error estimates can be obtained by treating the QMC solution as a random variable. The use of such methods is surveyed in Owen [16] for integration problems.

In our setting we generate both uniform and normally distributed grids. In order to randomise the Sobol' points in (28) and (31) we add the same random term to all points in the Sobol' sequence, modulo 1. Suppose $\mathcal{S} = (s_i)$ is our sequence of Sobol' points, then the jth randomised Sobol' sequence is given by

$$\mathcal{S}_j = \{s_i + U_j \mod 1\} \tag{34}$$

where U_j are independent uniform random vectors on $[0,1]^N$ and $x \mod 1$ denotes the fractional part of x.

The randomised QMC grids can then be generated using the framework of Section 5.2 with different randomised Sobol' sequences.

5.5 Reuse of $P^{(T-t)}$

The most computationally intensive task in the irregular grid method is the matrix root operation. In order to improve efficiency we propose a method where the matrix root can be reused for a number of trials. The method can be seen as a variance reduction technique.

First note that we can reformulate the problem so that the process has zero drift, but the drift is incorporated in the payoff function. That is, the price in the stopping time formulation is given by

$$v(x,t) = \sup_{\tau \in \mathcal{T}} \mathbb{E}^0_{x_t} \left(e^{-r(\tau - t)} \psi \left(X(\tau) + (r\mathbb{1} - \delta - \tfrac{1}{2}\mathrm{diag}(\Sigma)) \right) \right) \tag{35}$$

where the expectation is taken under the risk neutral measure without the drift term. That is, $\mathbb{E}^0_{x_t} X(s) = x_t$ for all $s \in [t,T]$.

Now given a grid $\mathcal{X} = \{x_1, \ldots, x_N\}$ we note that all rotations of $\mathcal{X}$ about $x = x_0$ and reflections in hyperplanes through x_0 will leave it invariant with respect to the driftless measure. The solutions from the different grids can be combined to provide a variance-reduced estimate at little extra cost.

Care must be taken that the payoff function is not invariant under the grid transformations. The idea of applying transformations is that solutions obtained using transformed grids represent new, although not independent, approximations to the price.

5.6 Exercise rule

A natural approximation to the optimal exercise rule using the pricing results of one grid is to take the implied rule of the nearest time and nearest neighbour in the grid.

Specifically we define the exercise rule for grid points to be

$$\tau(x_i, t_k) \triangleq \begin{cases} 1 & \text{if } v_i^{(k)} = \psi_i \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

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

$$\tau(x, t_k) \triangleq \begin{cases} 1 & \text{if } v_i^{(k)} = \psi_i \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

where $i = \text{argmin}\,(||x - x_i|| : x_i \in \mathcal{X}_k)$.

This rule is easily implemented and can also be adapted to the case where we have several different grids. In this case one could base the exercise rule on a vote between grids.

One can also extend the rule to time points between grid timesteps by taking the grid at the nearest time step or by taking a weighted average of the rules implied by the two neighbouring grids.

5.7 Hedging

Whereas applying an exercise rule to out-of-sample paths gives a low-biased estimate of the option value, applying a hedging rule gives a high-biased estimate. In order to hedge in a complete market setting we must be able to estimate partial derivatives of the value function.

In the literature there is little said about the practicalities of hedging in a high dimensional setting. The problem 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 an online hedging strategy by evaluating prices at states perturbed in each underlying; however simulations using such a method are expensive computationally because a large number of option values must be computed during the simulation. 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

affine approximation. The extra work involved in computing values at perturbed points is thus avoided; indeed the irregular grid method provides derivative information as a by-product.

To make this more formal, we define the multi-index-valued function

$$m(x, \mathcal{X}, k) \triangleq \{i_1, \ldots, i_k\} \tag{38}$$

where i_j is the jth nearest neighbour of x with respect to a Euclidean norm.

Now the partial derivative approximations are estimated by projecting the grid values at points $x_{m(x,\mathcal{X},k)}$ onto the space of affine functions and using the partial derivatives implied by the projection.

6 Experimental Results

6.1 Benchmarks

There are few benchmark results for high-dimensional American options. Broadie and Glasserman [6] 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 [14] price the Broadie and Glasserman maximum options using the LSM method.

Stentoft [18] uses the binomial method of Boyle et al. [3] and the LSM method to price put options on the arithmetic average, geometric average, maximum and minimum of three and five assets.

Broadie and Glasserman [5] and Fu et al. [9] provide benchmark results for options over five assets with three exercise opportunities.

Finally, Rogers [17] and Haugh and Kogan [11] use the dual formulation to price a number of different American options.

A useful result involving options on the geometric 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 (23), and the payoff function $\psi(s) = \left(K - (\prod s_i)^{1/N}\right)^+$ where K is the strike price and N 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 $Y(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t)$ and

$$\begin{aligned} dY(s) &= \frac{1}{N}\sum_{i=1}^{N} dX_i(s) && (39) \\ &= \tilde{\mu} ds + \tilde{\sigma} dW(s) && (40) \end{aligned}$$

The parameters of the diffusion are given by

$$\tilde{\mu} = r - \frac{1}{2N}\sum_{i=1}^{N}\sigma_i^2 \tag{41}$$

$$\tilde{\sigma}^2 = \frac{1}{N^2}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}R_{ij}\right)^2 \tag{42}$$

Using this we find that an accurate price for the geometric average American option in the Stentoft setting is 1.355, and for the Bermudan version with ten exercise opportunities it is 1.343; this represents an increase of about 6% in the early exercise premium.

6.2 Experiments

Our experiments are conducted in a MATLAB environment and are based on the five-dimensional examples of Stentoft [18]. Specifically we consider stock processes driven by correlated Brownian motions for put options with four different payoff functions. The setting we use for valuation is identical to that presented in Section 4.

Our programs are mostly script-based but some computationally intensive routines, for example the PSOR code, have been written in C as `mex` functions.

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[2] 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 uniform and normal grids in a randomised QMC framework as detailed in Section 4 using the parameter values $\alpha = 3$ and $\alpha = 1, 1.5, 2$ respectively; these were found to give good rates of convergence. The vector of initial stock prices x_0 was always included as the first grid point.

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

$$\begin{aligned} \psi_1(s) &= \left(K - \tfrac{1}{n}\textstyle\sum s_i\right)^+ & \psi_2(s) &= \left(K - \left(\textstyle\prod s_i\right)^{1/n}\right)^+ \\ \psi_3(s) &= \left(K - \max(s_i)\right)^+ & \psi_4(s) &= \left(K - \min(s_i)\right)^+ \end{aligned} \tag{43}$$

where x^+ denotes the positive part of x.

[2] In Stentoft's paper the volatilities are misprinted as $\sigma_i^2 = 0.2$.

Option type	Exact	Binomial	LSM	LSM (OS)	Normal grid	Normal grid (OS)
Arith. Average	-	1.235	1.241 (0.0006)	1.231 (0.0006)	1.246 (0.004)	1.238 (0.005)
Geom. Average	1.343	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)

Table 1: Comparison of Bermudan price estimates with ten exercise opportunities. The grid estimates are an average of 50 grids with size 1000; the normal grids had $\alpha = 1.5$. The binomial method of Boyle et al. [3] was used with Richardson extrapolation. The OS (out of sample) columns give the result of 100,000 trials using the exercise strategy implied by the method. The binomial and LSM prices are given by Stentoft [18] and the OS values for LSM are computed using 20 trials. The exact price given is the numerical solution to the equivalent one-dimensional problem. Standard errors are shown in brackets.

Tables 1 and 2 show results and comparisons for Bermudan and American option prices respectively using the irregular grid method with a normal grid and $\alpha = 1.5$.

We remark that the values obtained from the irregular grid method are higher than those produced by the LSM algorithm. This is particularly interesting in the out-of-sample results since this is the average value of the exercise strategy implied by the pricing algorithm.

6.3 Uniform grid

In the uniform grid case, only a single QMC grid was generated for each payoff function. Figure 3 shows the resulting convergence behaviour.

It was found that taking averages over several grids did not affect the convergence pattern greatly, and severe downward bias for small samples for both European and American prices was always observed, this bias seemingly caused by the poorly adapted grid. It can be seen however that the

Option type	Exact American	Normal grid American	Normal grid American (OS)	Hedged American (OS)
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)

Table 2: Comparison of American price estimates. The grid estimates in the third and fourth columns are an average of 50 grids with size 1000; the normal grids had $\alpha = 1.5$ and used ten time steps. The OS (out of sample) column gives the result of 100,000 trials using the exercise strategy implied by the 50 grid solutions, and is computed 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 at 50 times steps and uses the grid solution at the nearest grid time to compute the hedge. The exact price given is the numerical solution to the equivalent one-dimensional problem. Standard errors are shown in brackets.

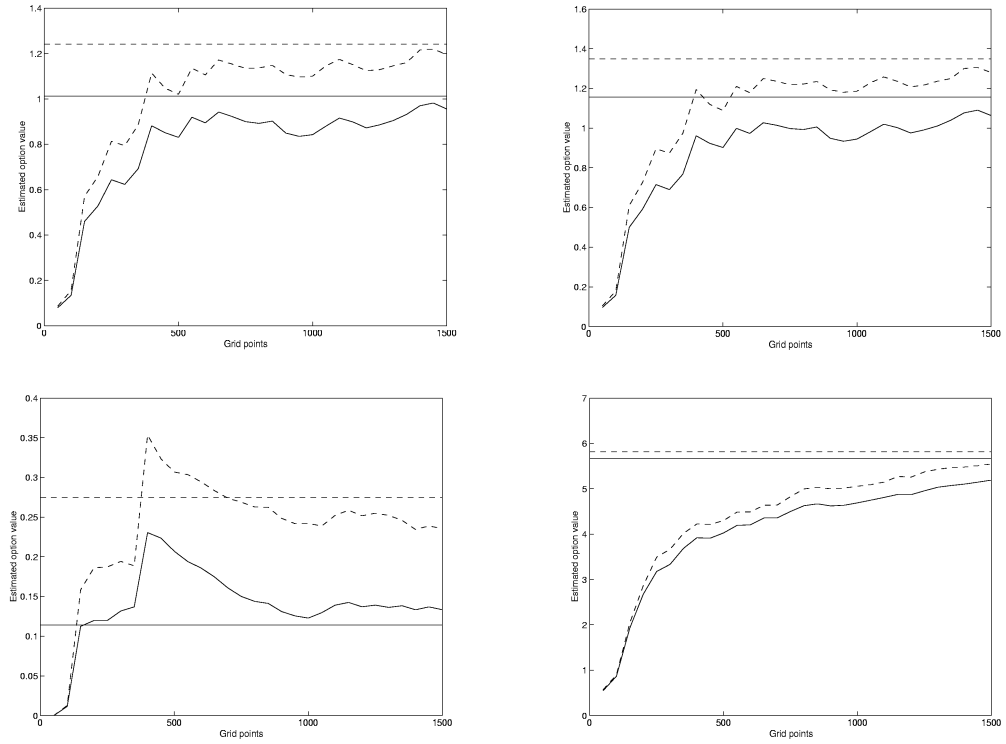

Figure 3: QMC grid valuation on a uniform grid with $\alpha = 3$ of European (*solid lines*) and American (*dotted lines*) put options on the arithmetic average, geometric average, maximum and minimum respectively over five assets. Grids use the same Sobol' sequence. Stentoft's LSM solutions are drawn as horizontal lines.

uniform grid solution is converging, albeit at a very slow rate.

We also see that the convergence patterns in the American and European cases are very similar, and the convergence of prices for arithmetic and geometric average payoffs are similar. The former is not surprising if we consider the similarities between the American and European problems, and the latter can be explained by the similarities in shapes of the arithmetic and geometric payoff functions.

6.4 Normal grid

In the normal grid case 50 randomised QMC grids were used for each payoff function, although such a high number should not be needed in practise.

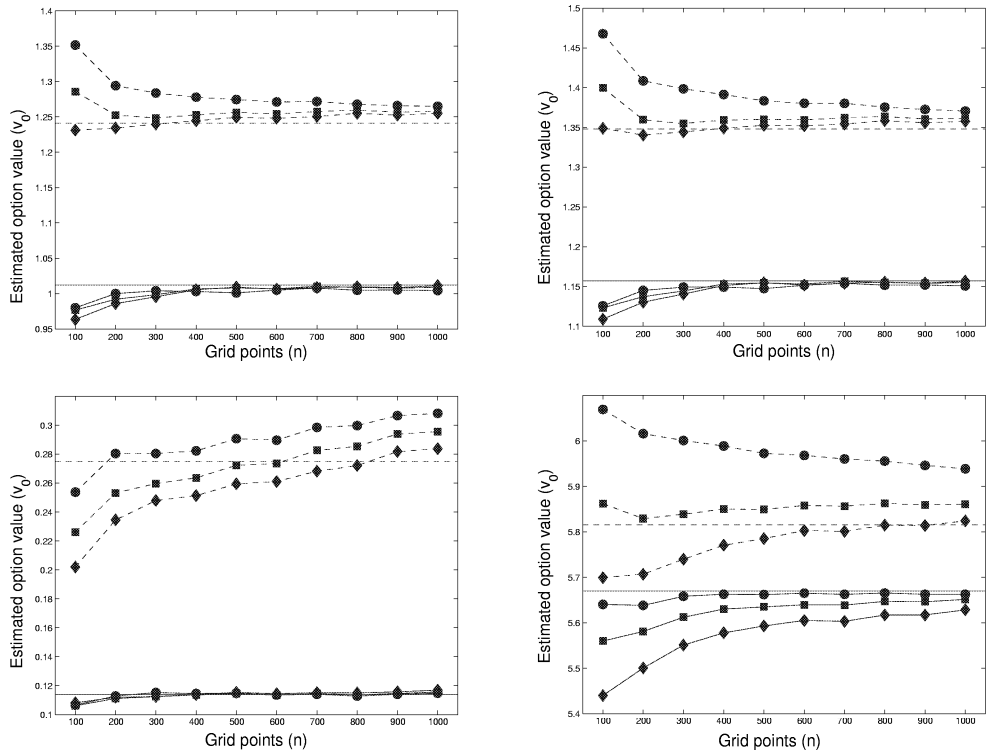

Figure 4: Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options on the arithmetric average, geometric average, maximum and minimum respectively over five assets using $\theta = 0.5$. Stentoft's Bermudan LSM solutions are drawn as horizontal lines.

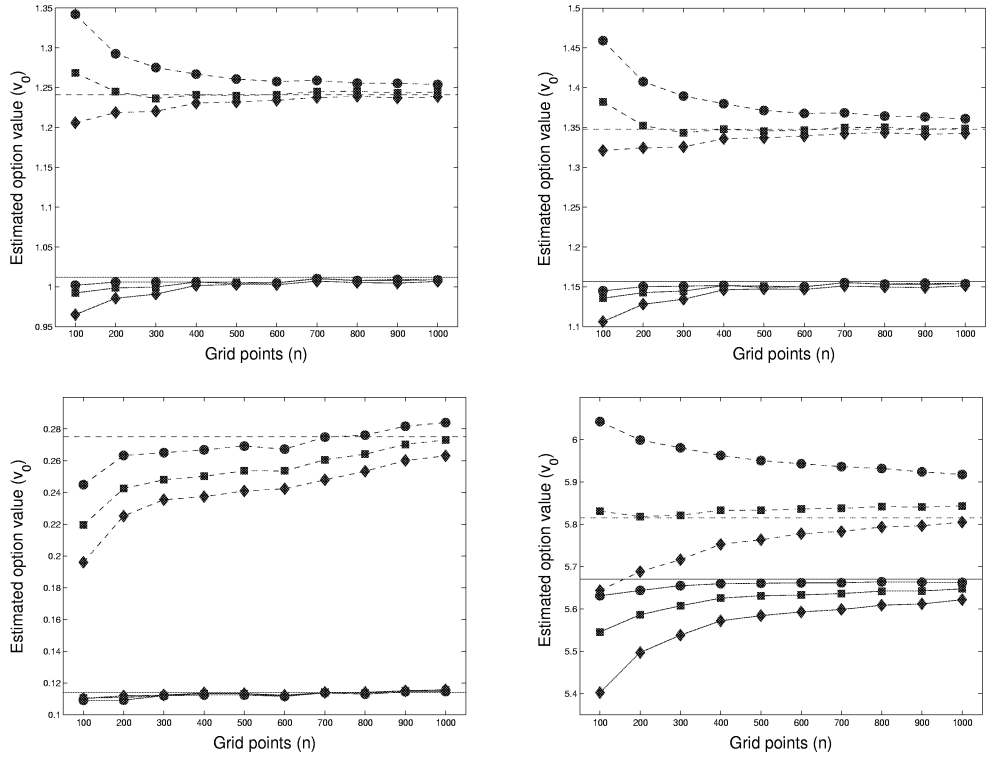

Figure 5: Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options on the arithmetic average, geometric average, maximum and minimum respectively over five assets using $\theta = 0$. Stentoft's LSM solutions are drawn as horizontal lines.

Figure 4 shows the convergence behaviour of the irregular grid method for QMC uniform and normal grids respectively where the implicitness parameter is $\theta = \frac{1}{2}$. The convergence of the normal grid solution is much faster than for the uniform grid, as expected.

We see that convergence is very quick for the arithmetic and geometric average options. It seems that the convergence is always to a value slightly higher than Stentoft's Bermudan prices, the increase in early exercise premium of about 5% representing the extra value of a true American contract.

Figure 5 shows the convergence behaviour for a normal grid where the explicit method ($\theta = 0$) is used. This can be directly compared to the results of Stentoft since the explicit method solves the Bermudan pricing problem.

For the more problematic cases of the maximum and minimum options, the convergence is much slower. In the case of the maximum it is not clear even 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. This is a case where the grid could be adapted to the payoff function as well as to the process itself; such extensions are reserved for future investigation.

In Table 1 it is encouraging to see that the irregular grid prediction for the geometric average option is very accurate as compared to the exact price obtained by solving an equivalent one-dimensional problem. The exercise strategy performs well in the average rate options, but not for the more problematic maximum and minimum payoffs.

6.5 Hedging

We have proposed an algorithm for pricing American options which, as detailed in Section 5.7 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 neighbours for the delta estimation, we obtain the results of Table 2. It is clear that the hedge errors are much larger than the exercise errors.

The strategy used is naive however in that the results of only a single grid solution are used. If could be improved by using information from different solutions or by recomputing the hedging strategy at each timestep.

6.6 Timings

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 866MHz machine are presented in Table 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 take about the same amount of time for example.

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 diffusions with arbitrary payoff functions and parameters without the need for recomputation. For options with different payoffs the matrix can be reused in the obvious way. When different parameters are considered, grid transformations and time changes can be used to adjust the implied

Size P (n)	Mem. full (1E6 b)	Mem. sparse (1E6 b)	Prop. $\neq 0$	Time for $P^{1/10}$ (sec)	Time-stepping (sec)
500	2.0	0.6	0.190	23	0.5
1000	8.0	1.8	0.147	200	1.3
1500	16.0	3.3	0.123	750	2.0
2000	32.0	5.1	0.106	2000	2.9
2500	50.0	7.1	0.094	4000	3.8
3000	72.0	9.1	0.084	7200	4.9

Table 3: Timings and storage requirements for the irregular grid method using MATLAB 6.1 with a PIII 866MHz processor, matrix entries stored in double precision (8 bytes per entry). The sparse matrices are formed by eliminating all entries smaller than 5×10^{-4} and renormalising. Timestepping is total over 10 timesteps, 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.

covariance of the transition probabilities, and the payoff functions can be manipulated in order to account for a different risk-neutral drift.

7 Conclusions

We have proposed a new method for finding the value of American 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 (PDO) 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 PDO corresponding to this Markov chain. This is done through taking a logarithm of the transition probability matrix, although in practise one must only compute a root of the matrix.

Most importantly we remove the requirement to specify basis functions to be used in approximations of the value function or exercise strategy. Indeed the only specification needed is a grid density, although asymptotically even

this choice is not critical. This sets our method aside from methods such as LSM where the specification of basis functions plays a critical role in the success of the method.

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 easily be estimated from the solution by projecting values from nearby points onto the space of affine functions.

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 the maximum option was not clear with a grid of 1000 points.

In the uniform grid case, although the convergence rate was not as fast, it could be seen that the convergence patterns of the European and American estimates were very similar. Given that the European estimate amounts to the tractable method of Monte Carlo integration with importance sampling, this similarity is encouraging.

The numerical results obtained largely agree with those of Stentoft [18]. We find that the early exercise premium is increased by about 5% 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 give 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 [14] to save time by only using in-sample paths.

This emphasises the need to specify an exercise (or hedging) strategy when pricing American options; without such an accompanying strategy the price is not meaningful to the buyer (or seller).

In terms of future research possibilities there are a number of areas which can be explored, in terms of both the numerical and theoretical properties of the algorithm.

The irregular grid method extends naturally to the valuation of other derivatives including those related to the term structure, and determination of the corresponding exercise and hedging strategies. This is of considerable

practical interest given the amount of money involved in such markets and the need for reliable algorithms.

A further variance reduction technique would be to adjust the transition probabilities according to the empirical density of points; this should result in even better convergence properties. 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 [13], but may also take inspiration from the literature on nonparametric analysis. These and other possible refinements are reserved for future investigation.

Further work is also needed to obtain theoretical justification of the proposed method. Such a justification may follow the analysis of Glowinski et al. [10] and Jaillet et al. [12] or the stochastic control methods of Kushner and Dupuis [13].

References

[1] Jérôme Barraquand and Didier Martineau. Numerical valuation of high dimensional multivariate American securities. *Journal of Financial and Quantitative Analysis*, 30:383–405, 1995.

[2] Phelim Boyle, Mark Broadie, and Paul Glasserman. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21:1267–1321, 1997.

[3] Phelim P. Boyle, Jeremy Evnine, and Stephen Gibbs. Numerical evaluation of multivariate contingent claims. *Review of Financial Studies*, 2:241–250, 1989.

[4] Phelim P. Boyle, Adam W. Kolkiewicz, and Ken Seng Tan. Pricing American style options using low discrepancy mesh methods. *Forthcoming, Mathematics and Computers in Simulations.*

[5] Mark Broadie and Paul Glasserman. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, 21:1323–1352, 1997.

[6] Mark Broadie and Paul Glasserman. A stochastic mesh method for pricing high-dimensional American options. *Working paper*, 1997.

[7] E. Clément, D. Lamberton, and P. Protter. An analysis of a least squares regression method for American option pricing. *Finance and Stochastics*, 6:449–471, 2002.

[8] M.A.H. Dempster and J.P. Hutton. Pricing American stock options by linear programming. *Mathematical Finance*, 9:229–254, 1999.

[9] Michael C. Fu, Scott B. Laprise, Dilip B. Madan, Yi Su, and Rongwen Wu. Pricing American options: a comparison of Monte Carlo simulation approaches. *Journal of Computational Finance*, 4:39–88, 2001.

[10] R. Glowinski, J.L. Lions, and R. Trémolières. *Numerical Analysis of Variational Inequalities.* North-Holland, 1981.

[11] Martin B. Haugh and Leonid Kogan. Pricing American options: a duality approach. *Working Paper*, 2001.

[12] Patrick Jaillet, Damien Lamberton, and Bernard Lapeyre. Variational inequalities and the pricing of American options. *Acta Applicandae Mathematicae*, 21:263–289, 1990.

[13] Harold J. Kushner and Paul G. Dupuis. *Numerical Methods for Stochstic Control Problems in Continuous Time.* Springer-Verlag, 1992.

[14] Francis A. Longstaff and Eduardo S. Schwartz. Valuing American options by simulation: a simple least squares approach. *Review of Financial Studies*, 14:113–147, 2001.

[15] Manuel Moreno and Javier F. Navas. On the robustness of least-squares Monte Carlo (LSM) for pricing American options. *Working paper*, 2001.

[16] Art B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, 1995.

[17] L.C.G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12:271–286, 2002.

[18] Lars Stentoft. Assessing the least squares Monte-Carlo approach to American option valuation. *CAF Working paper No. 90*, 2001.

[19] Lars Stentoft. Convergence of the least squares Monte-Carlo approach to American option valuation. *CAF Working paper No. 113*, 2002.

[20] James A. Tilley. Valuing American options in a path simulation model. *Transactions of the Society of Actuaries*, 45:499–520, 1993.

[21] Joseph F. Traub, G.W. Wasilkowski, and H. Wozniakowski. *Information-based Complexity.* Academic Press, 1988.

[22] John N. Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12:694–703, 2000.

[23] Paul Wilmott, Jeff Dewynne, and Sam Howison. *Option Pricing: Mathematical Models and Computation.* Oxford Financial Press, 1993.

[24] Christoph Zenger. Sparse grids. *Parallel Algorithms for Partial Differential Equations: Proceedings of the Sixth GAMM-Seminar*, 1990.