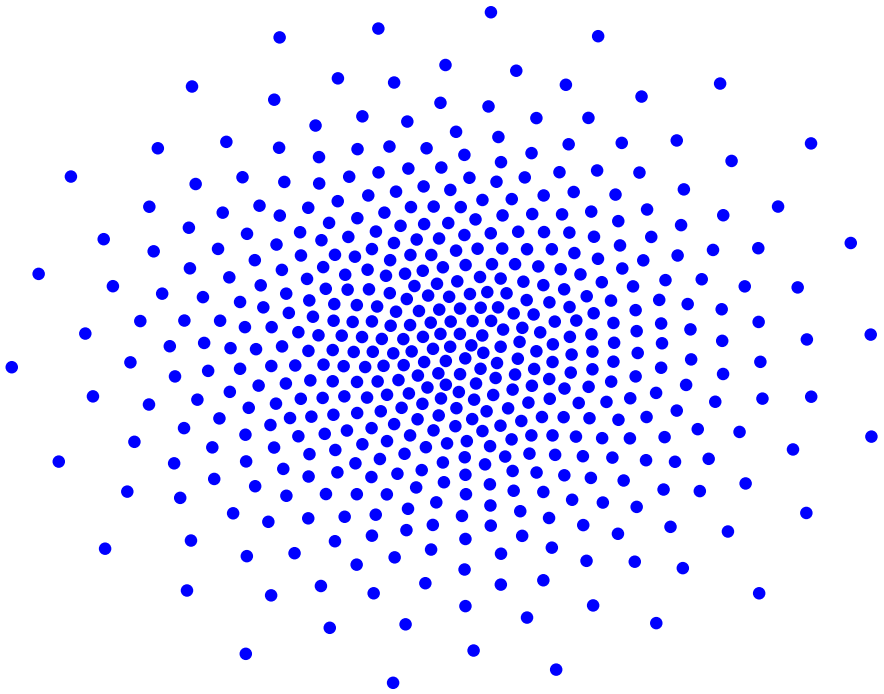
Irregular grid methods for pricing high-dimensional American options

Berridge, S.J.

Publication date:
2004

Irregular Grid Methods for Pricing High-Dimensional American Options



Irregular Grid Methods for Pricing High-Dimensional American Options

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof.dr. F.A. van der Duyn Schouten, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 18 juni 2004 om 16.15 uur door

STEFFAN JOHN BERRIDGE

geboren op 21 maart 1974 te Londen, Engeland.

PROMOTOR: PROF.DR. J.M. SCHUMACHER

to those with the courage to be irregular,
boe uiptf xip dbo xpsl pvu xibu uijt tbzt ...

Preface

This thesis consists of seven chapters, of which Chapters 2–4 are published in the CentER Discussion Paper series as [8, 10, 11]. Chapters 2–4 and 6 were written in cooperation with Hans Schumacher. Early versions of Chapter 2 are also published as [7, 9].

Acknowledgements

As my Tilburg era draws to a close I would like to thank those who have taught me how use laserjet ink to turn paper into papers, and those who made four years in Tilburg seem like one ride on a roller coaster.

The star of the official show was Hans, whose patience and encouragement motivated me at every meeting. His guidance has contributed in no small way to my inspiration for the contents of this thesis. Thank you Hans.

Thanks also go to the other members of the committee; Arun Bagchi, Gül Gürkan, Jong-Shi Pang, Antoon Pelsser, Dolf Talman and Bas Werker; both for reading the thesis and for providing valuable feedback.

In the last four years I have had contact with many people in the Faculty of Economics and Business Administration, who have helped and guided me in many ways. In particular I would like to thank the administrative staff at CentER and the Department of Econometrics and Operations Research, who have always been friendly and ready to help.

I have been motivated in my work through several other means. Participation in the Mathematical Research Institute graduate programme took me to the heart of current knowledge and research activities in mathematical finance, and I would like to thank the lecturers and students for providing motivation and interesting discussions. Attendance of many conferences and workshops put me in contact with important researchers and fellow students from other institutions. These valuable experiences have inspired me and given me perspective in the academic world.

On the nonacademic side of life, there are many people I would like to thank for making my time in Tilburg so enjoyable. The fellow inhabitants of Robert Fruinstraat - Giorgio, Pedro, Zhang, Rossella and the mouse - and the inhabitants of Mieredikhof - Charles, Vera, Jeroen, Pierre-Carl, Sabine and the mouse - have

always provided great company, interesting conversation and a range of culinary experiences. Thanks to those who helped set up the Mieredikhof Market Gardening enterprise, to the neighbours who seldom complained about this and other activities, and to one neighbour for so thoughtfully and frequently stating his outspoken perceptions on the odour of barbecues.

Of course any discussion about life in Tilburg cannot omit mention of the social genius of the Tilburg University International Club (Tunic) of which Aldo and Sandra have recently been the backbone, organising activities from pot luck dinners to kart racing. Then there's the people who selflessly attended the bierproeverij at Kandinsky on Sunday afternoons - these are too many to mention, but perhaps I should thank the waiters for always meticulously stating the alcohol percentage of the beer and its origin, and for never tiring of the question "Is that in Belgium?". I would also like to extend thanks the owner of the Villa for providing a meeting point for social activities and a great venue for celebrating various occasions in a modest but enjoyable fashion, and to the creator of the inspiring statue on the right inside the church on Heuvelplein.

Special thanks to those who helped in setting up the Graduate Students' Society. In particular the other founding board members Pierre-Carl, Rossella, Ania and Man-Wai. Also to the other students who selflessly gave their time and energy to be part of the society, to Niels van de Ven from the university bureau who provided much valuable advice on setting up the society, and to the faculty for providing funding for this initiative. My life would also not have been complete without GraDUAL magazine, with its dazzling feature articles and revealing student mini-autobiographies, so thanks to the people who made it possible (however "GraDUAL-ly" it comes out.)

On the sporting side, thanks to the guys and occasional girl that participated in the soccer games on Friday afternoons, and to the main organisers Bram, Jeroen and Mark-Jan. Thanks also to the people who encouraged me with running, mainly spearheaded by Bertrand Melenberg, and to those who joined on countless occasions running in the woods and along the picturesque Wilhelmina canal. Special thanks to Pierre-Carl for being far more attractive to large dogs than I was. On the culinary side, thanks to Tam Food for providing a great selection of inexpensive Chinese condiments. And how could I forget I Pin Ke, Bali and Minos Pallas which formed the hard core of good quality but reasonably-priced restaurants in

Tilburg.

My Tilburg experience has also been enriched by participating in the university students' choir Contrast, conducted by Peter van Aerts. Aside from some spectacular performances, including Carmina Burana and Mozart's Requiem, I have also enjoyed having a few quiet pints with fellow choir members at the Korenbloem on Monday evenings, and attending rehearsal weekends. Thanks Contrasters.

Finally, I would like to thank Vera, whose honesty, kindness and determination have added a new dimension to my life.

STEFFAN BERRIDGE
APRIL 8TH 2004, TILBURG

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The field of option pricing, and more generally mathematical finance, has represented one of the great triumphs of interdisciplinary research in the twentieth century. Louis Bachelier, aptly named the founder of mathematical finance, set the stage for its study in 1900 with his doctoral thesis "Théorie de la Spéculation" which proposed continuous-time stochastic processes, including Brownian motion, as a basis for the analysis of financial markets. Seventy-three years later Fischer Black, Myron Scholes and Robert Merton revolutionised the study of option pricing by devising closed-form expressions for European option prices based on a model in which the underling asset follows a geometric Brownian motion. In the same year the Chicago Board Options Exchange was founded and the rest, as they say, is history.

The study of option pricing centres on the modelling of financial market processes as simplified stochastic processes, the statistical analysis connecting the model with relevant data and the mathematical analysis of such stochastic models. The statistical analysis may involve the estimation of model parameters or the extraction of market information related to the model. The mathematical analysis typically starts with a particular stochastic model and attempts to draw relevant conclusions, for example regarding option prices, assuming that the model is correct. This field of study thus draws together, at least in the first instance, researchers from finance, econometrics, statistics and mathematics.

This thesis focuses entirely on the mathematical aspect of option pricing, and completely ignores the statistical aspect. Given a stochastic model for a number

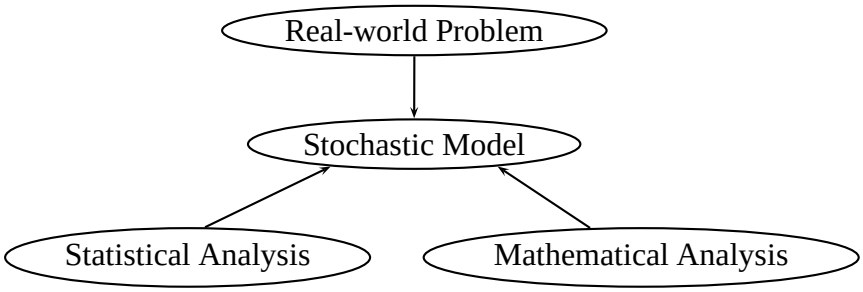


Figure 1.1: Aspects of a stochastic model.

of possibly interdependent financial market processes, the question of interest is how one can efficiently determine a no-arbitrage price for a derivative based on these processes. In particular we ask this question for American options written on multidimensional processes which follow correlated geometric Brownian motions.

Thus the empirical question of whether the models studied are realistic is not tackled. For the purposes of the research however, the models are believed to constitute an adequate representation of reality for many processes found in financial markets. On the other hand, it is an uncontestable fact that many superior models have been developed for various financial processes over the past three decades. Such models have not been considered here because simpler models provide a sufficiently rich testing ground on which to realise the main aims of the research.

As the title makes clear, this thesis tackles the problem of pricing American options when the number of underlying variables (the dimension) is large. The closely related problem of pricing of Bermudan options, where the number of exercise opportunities is finite, is included in the scope of the research. For “large” one may read “at least three”, since this is the dimension in which classical solution methods, in particular those based on regular grid discretisations, become cumbersome. The thesis further focuses on methods which use an irregular grid as a basis for calculations. The use of an irregular grid allows one to devise methods which are based on a tractable number of points, and which allow some freedom in placing more points in areas where the behaviour has more effect on the required solution.

The combination of the early exercise feature, offered by both American and Bermudan options, with the problem of high-dimensionality presents a consider-

able challenge for numerical analysis both in terms of the pricing problem and the associated hedging problem. The early exercise feature itself does not present a great challenge for one-dimensional problems; neither does high-dimensionality present a great challenge when pricing options without early exercise features; such problems can be solved relatively quickly using Monte Carlo or quasi-Monte Carlo integration methods. In combining these two complications however, one appears to require considerable computational resources.

An important question is whether the minimum amount of computational resources required to solve the problem to within a certain accuracy must increase exponentially with dimension; that is, are we facing a curse of dimensionality in the sense of approximating the solution? This question is explored in a related context by Rust [65], who shows that one can break the curse of dimensionality for certain classes of Markovian decision problems. The regularity assumptions are too strict for the types of option pricing problems considered in this thesis, and the question thus remains open for the latter.

The first widely-used numerical methods for pricing American options were those of Brennan and Schwartz [16] and Cox et al. [24], the former being an adaptation of the explicit finite difference method used for pricing European options, and the latter a binomial tree method. Later Wilmott et al. [74] showed how implicit finite difference methods for solving partial differential equations (PDEs) could be used for pricing American options by solving a linear complementarity problem (LCP) at each step. They used the projected successive overrelaxation (PSOR) method of Cryer [25] to solve the LCPs; more recently Huang and Pang [40] provide a review of state-of-the-art numerical methods for solving LCPs. Although these methods were all successful for solving one-dimensional problems, their reliance on regular grid discretisations rendered them unsuitable for high dimensions. Other relevant work included the analytic valuations of Geske and Johnson [32] and a number of methods involving approximations of the exercise boundary.

The first author to break ranks on the prevailing view that simulation techniques could not be used for pricing American options was Tilley [70], who used state space aggregation as a means for specifying the continuation values. This had important consequences for high-dimensional problems because Monte Carlo methods were at least able to make computations feasible in high dimensions, and were even able to break the curse of dimensionality for certain problems. Two

years later, Barraquand and Martineau [4] presented similar methods specifically aimed at the pricing of high-dimensional American securities.

The use of simulation for valuing American options was further developed by Broadie and Glasserman [18], who used a simulated tree structure to calculate confidence intervals for option values. The method is suitable for high-dimensional problems, although it suffers exponential computational complexity in the number of time steps. This method was extended by the same authors [19] using a stochastic recombining mesh which did not suffer the curse of dimensionality, at least in a computational sense.

Carrière [20] made an important development by showing that path simulations of the underlying process could be used simultaneously to approximate the optimal stopping time and to provide price estimates. This was done by using nonparametric techniques to estimate the continuation value, and using the implied stopping rule to determine the average value realised along the simulated paths. Longstaff and Schwartz [50] later showed this to be a feasible method for high-dimensional problems using least squares regression in place of nonparametrics to estimate the continuation value.

In related work, Tsitsiklis and Van Roy [71] use projection onto a set of "features" to perform value iteration. The projection used here is in principle no different to that performed by Longstaff and Schwartz, where the features are the basis functions used for the regression. The key difference between the two methods is in the use of information from the paths; Tsitsiklis and Van Roy perform value iteration using information only from the function approximation made for the previous time step, whereas Longstaff and Schwartz use the implied stopping rule to determine the realised value along each path. One can see this difference from another angle — namely the difference in the use of the functional approximation. Tsitsiklis and Van Roy use functional approximation directly to approximate the value function; Longstaff and Schwartz use functional approximation to specify a (hopefully near-optimal) stopping rule. This difference seems to weigh in favour of Longstaff and Schwartz in terms of accuracy; suggesting that the value of an American option is not very sensitive to the stopping rule used in the valuation.

Tsitsiklis and Van Roy prove convergence of their method in their paper, and a proof for the Longstaff and Schwartz method is undertaken by Clément et al. [22]. The question about the rate of convergence has been formally answered only with

respect to the number of paths in the Longstaff and Schwartz method, where the dependence is $n^{-1/2}$; more difficult is to determine the rate of convergence with respect to the number of basis functions. Given the use of functional approximation methods, which are known to suffer a curse of dimensionality, one may fear the worst. Recent analysis has focused on how the minimum number of paths required for convergence depends on the number of basis functions. Stentoft [69] lays a theoretical basis for requiring a cubic number of paths on an average case basis; Glasserman and Yu [34] prove that the requirement on a worst case basis is at least exponential.

This thesis contributes to the above literature by exploring new methods for the valuation of high-dimensional American options. The methods presented are different to those discussed above, except of course for their main aim which is to specify computationally efficient methods. Rather than using path simulations, we rather use an irregular grid as the means for storing information about the value function. In Chapters 2–4 the grid is assumed to be constant over time, thus constituting a method of lines approach when the problem is seen from the PDE point of view, or an approximating Markov chain approach when seen from the stochastic differential equation (SDE) point of view. A proof of convergence for the methods of Chapters 2–4 is presented in Chapter 6. The method presented in Chapter 5 differs fundamentally from the others in that the irregular grid is allowed to change over time, demanding a different approach for the use of information from previous time steps. A separate proof of convergence is provided for the method presented in that chapter.

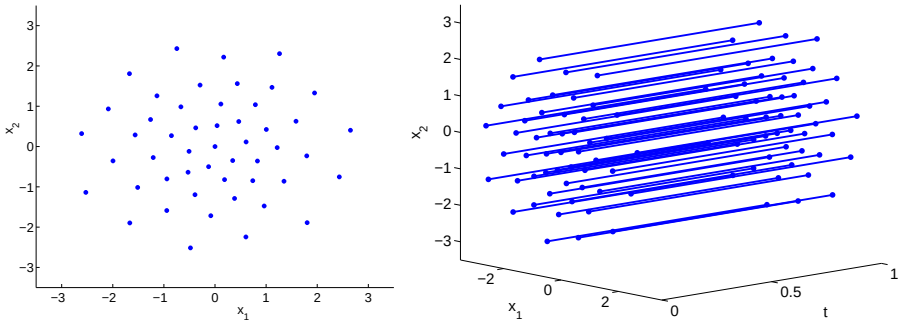
The use of an irregular grid is inspired by the use of similar methods for approximating integrals. The value of a European option is essentially the discounted integral of the payoff function with respect to the density of the process at expiry. This can be evaluated at least moderately efficiently using Monte Carlo (MC) methods, and in many cases more efficiently using quasi-Monte Carlo (QMC), or other low discrepancy- or low distortion-based methods. The use of simulation methods for pricing high-dimensional European options is treated in Boyle et al. [13]; another useful reference for MC integration is Evans and Swartz [28] and for QMC integration Niederreiter [58]. The historical succession of methods contains an interesting twist: firstly the deficiency of deterministic integration methods in high dimensions led to the development of MC methods, which are randomised;

secondly the slow rate of convergence observed for MC methods led to the development of the faster QMC methods, which are again deterministic. The question we hope to answer in this thesis is whether the success of MC and QMC methods can be combined with traditional PDE solution methods to form a framework for the numerical solution of optimal stopping problems, including the American option pricing problem.

There are two main reasons to believe that PDE methods may be preferable to MC methods for American option pricing. Firstly, PDE methods typically admit Taylor series error analyses for European problems, whereas simulation-based methods admit less optimistic probabilistic error analyses; in practise one indeed observes faster convergence rates for PDE methods. Secondly, the number of tuning parameters that must be used in PDE methods is much smaller than that required for the type of simulation-based techniques that have been suggested for American option pricing; for the latter one often faces the problem of having to choose a suitable set of basis functions, for which choice one must have sufficient a priori knowledge about the shape of the value function.

Like its European counterpart, the value of an American option is also a discounted integral; the integral in the latter case is rather performed with respect to the (unknown) optimal stopping boundary. Thus, for an American option one not only requires some representation of the process density at expiry, but also of its behaviour at intermediate times. The main challenge faced is thus the representation of the continuous process in a discrete state context, whether that be done in a continuous or discrete time setting. In the case of continuous time one works on a domain similar to that shown in Figure 1.2. In this case one requires an infinitesimal generator matrix to provide a representation of the original process, and in discrete time one requires a Markov transition matrix. In Chapter 2 the infinitesimal generator matrix is constructed by taking the logarithm of a matrix corresponding to transition probabilities at expiry; in Chapter 3, local quadratic approximations for the value function are used to build the generator matrix, and in Chapter 4 local consistency conditions similar to those of Kushner and Dupuis [46] are used.

One of the most difficult issues arising in irregular grid schemes turns out to be stability. Indeed one may specify transition probabilities using a number of seemingly consistent methods. The stability of such schemes is usually related to the eigenvalues of the generator or transition matrix, and is thus difficult to guarantee



(a) Grid (b) Time-state domain

Figure 1.2: Example of 50-point grid and corresponding time-state domain of approximating continuous-time Markov chain for methods of Chapters 2–4.

in many circumstances. Chapter 3 provides a case in point for this issue, where the connection between the weights and the eigenvalues is difficult to ascertain. Finding a way to force the eigenvalues into the stable region is thus very difficult, although experiments suggest that stability may be induced by considering relatively "smooth" grid specifications or by manipulating the neighbour configuration so that neighbours are distributed about each point more uniformly in an angular sense.

Chapters 2 and 4 admit simpler stability analyses. In the former, one obtains an infinitesimal generator by taking the logarithm of a nonsingular Markov transition matrix having eigenvalues in the unit interval, thus restricting the eigenvalues to the negative half line. In the latter, the infinitesimal generator is specified using local consistency conditions in combination with sign restrictions on the elements; this leads to a diagonally dominant matrix with negative diagonal thus restricting the eigenvalues to lie in the left half plane through Gershgorin's disk theorem.

A closely related strand of literature to Chapters 2–4 is the work on sparse grids initiated by Zenger [75], and continued by several German and Dutch researchers. Sparse grids attempt to deal in particular with the solution of high-dimensional PDE problems. The sparse grid approach differs in two important ways from the one taken here; first it assumes a rectilinear domain for the problem and second the sparse grid structure does not admit a simple stochastic interpretation. On the other hand, the irregular grid methods considered in this thesis are meant in

principle to form a basis for the approximation of unbounded stochastic processes, and the specification of transition intensities can be directly interpreted in terms of the local behaviour of the original stochastic process. One may also view the methods of Chapters 2–4 in a PDE context, facilitating the construction of implicit time stepping schemes.

Another important related work is that of Kushner and Dupuis [46], who formulate approximating Markov chain methods for solving optimal investment consumption problems. Although the problems they consider are mostly of low dimension, their ideas of local consistency extend naturally to higher dimensions. Their ideas have been used in Chapter 4 as a motivation for specifying transition intensities and probabilities in approximating Markov chains on irregular grids. Kushner and Dupuis do in fact present methods for approximating Markov chain dynamics on certain nonstandard grids, but they do not consider irregular grids as used in this thesis.

A proof of convergence related to the methods of Chapters 2–4 is presented in Chapter 6. The proof makes use of the variational inequality formulation for American option pricing, and follows the framework developed by Glowinski et al. [35], and the references therein, in proving convergence of numerical schemes for solving variational inequalities. The Glowinski et al. framework constitutes a powerful tool due to its abstractness; Jaillet et al. [42] and Zhang [76] have previously used these methods for proving convergence of one-dimensional American option pricing algorithms.

Chapter 5 presents a method which is fundamentally different from those found in the other chapters. The essential difference is that a different grid is permitted at each time step, thus allowing the grids to be more adapted to the process. Value iteration is performed, with information from the previous time step being conveyed through appropriate interpolation and quadrature operators. The interpolation operator extends information contained at grid points to define a value function on the entire state space, and the quadrature operator uses this interpolated value function as a basis for finding the expected value of continuation. This method is closely linked to the literature on mesh-free methods which studies interpolation methods that can be applied to irregular grids. Recent work by Levin [47] and Wendland [72] shows that moving least squares methods can produce interpolants which are accurate to high degrees, and work by Maz'ya and Schmidt [53] and Fasshauer [29]

hints at how one may obtain such interpolants in reasonable computation times. These latter methods have not been employed in the current work, but they show how the complexity may be reduced. Nearest neighbour interpolation is found to be much faster than moving least squares, and gives reasonable pricing results with the use of an inner control variate, i.e. the application of a control variate at each time step.

Chapter 2

A Method Using Matrix Roots

2.1 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 [4] 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 analyse and has a possible lack of convergence; Boyle et al. [12] demonstrate this and propose a modification of the algorithm which leads to a low-biased estimator.

Broadie and Glasserman [18] 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 [19] 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. [31]).

The least squares Monte Carlo (LSM) method presented by Longstaff and Schwartz [50] 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. [22] and Stentoft [69] independently provide proofs of convergence for the LSM method, showing that the convergence rate is $n^{-1/2}$ in the number of paths used. The convergence behaviour in the number of basis functions however has not been determined. Stentoft [69] and Glasserman and Yu [34] 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 [68] and Moreno and Navas [54] 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 sensitive to the choice of basis functions in five dimensions.

Tsitsiklis and Van Roy [71] 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 [19] 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 [64] and Haugh and Kogan [37] and later developed by Jamshidian [44] and Kolodko and Schoenmakers [45]. 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.

In this chapter, 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 discretisation 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 this chapter is organised as follows. In Section 2.2 we present a mathematical formulation of the problems to be solved numerically and in Section 2.3 we show how an irregular grid method can be used to solve the problem. We then present the experimental setup in Section 2.4, results in Section 2.5 and concluding remarks in Section 2.6.

2.2 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

$$dX(s) = \mu(X(s),s)ds + \sigma(X(s),s)dW(s) \tag{2.2.1}$$

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

$$dV(s) = \mu_V(X(s), s)ds + \sigma_V(X(s), s)dW(s) \tag{2.2.2}$$

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

$$dX(s) = \mu_{RN}(X(s), s)ds + \sigma(X(s), s)dW(s) \tag{2.2.3}$$

where μ_{RN} is the risk-neutral drift. Note that $W(s)$ is now a Brownian motion under the risk neutral measure, and is thus different to the $W(s)$ in (2.2.1).

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:

$$\tau : \mathbb{R}^d \times [t, T] \rightarrow \{0, 1\} \tag{2.2.4}$$

$$H : \mathbb{R}^d \times [t, T] \rightarrow \mathbb{R}^d. \tag{2.2.5}$$

In the following, we appeal to the complementarity formulation of the American option price which is presented for example in Jaillet et al. [42]. Let $\mathcal{L}$ be the related diffusion operator

$$\mathcal{L} = \tfrac{1}{2}\mathrm{tr}\sigma\sigma' \frac{\partial^2}{\partial x^2} + \mu_{RN}\frac{\partial}{\partial x} - r \tag{2.2.6}$$

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

$$\begin{cases} \frac{\partial v}{\partial t} + \mathcal{L}v & \leq & 0 \\ v - \psi & \geq & 0 \\ \left(\frac{\partial v}{\partial t} + \mathcal{L}v\right)(v - \psi) & = & 0 \end{cases} \tag{2.2.7}$$

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

2.3 Methodology

To solve the complementarity problem (2.2.7) we first form a semidiscrete complementarity problem by discretising the state space but leaving time continuous. This involves sampling the state space using QMC trials and finding a suitable approximation of $\mathcal{L}$. 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 summarise by providing a concise statement of the algorithm.

2.3.1 State space discretisation

We first consider a semidiscrete approximation to the complementarity problem (2.2.7) in which the state space is discretised and time left continuous. This is often called the method of lines. In the pricing problem this amounts to approximating (2.2.7) 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 discretised complementarity problems, the implicit scheme is the only time stepping method known to be unconditionally stable (see Chapter 6 and Glowinski et al. [35]).

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 2.3.2.

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

complementarity problem

$$\begin{cases} \frac{dv}{dt}(s) + Av(s) & \leq & 0 \\ v(s) - \psi & \geq & 0 \\ \left(\frac{dv}{dt}(s) + Av(s)\right)'(v(s) - \psi) & = & 0 \end{cases} \tag{2.3.1}$$

for $s \in [t, T]$ with terminal condition $v_i(T) = \psi(x_i)$ for each $i = 1, \ldots, n$. Note that $v(s)$ 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 Markov chain approximation to the optimal stopping problem. That is, the process $X(s)$ is approximated by a process restricted to the grid $\mathcal{X}$. The operator A gives transition intensities on this grid.

2.3.2 Approximating the partial differential operator

We now propose a method for specifying A in (2.3.1) 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 emphasised in Glasserman [33] as a favourable property of the stochastic mesh method presented by Broadie and Glasserman [19].

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

$$p_{ij}^{(T-t)} = \frac{1}{\sum_{k=1}^{n} \tilde{f}_{x_i,T-t}(x_k)} \cdot \tilde{f}_{x_i,T-t}(x_j) \tag{2.3.2}$$

where the weights are given by

$$\tilde{f}_{x_i,T-t}(x) = \frac{f_{x_i,T-t}(x)}{g(x)}. \tag{2.3.3}$$

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 $T - t$.

In the semidiscrete Markov chain setting, where A represents transition intensities, we note that the evolution of state probabilities is given by $p(s) = e^{A'(s-t)}p(t)$ where $p(t)$ is the initial probability distribution at time t, for example it may be a delta function in the case where the initial state is known. The matrix $P^{(T-t)}$ thus gives us access to an approximation A to $\mathcal{L}$ on $\mathcal{X}$ as follows:

$$A \triangleq \frac{1}{T-t} \log P^{(T-t)}. \tag{2.3.4}$$

The matrix logarithm of $P^{(T-t)}$ certainly exists and is unique if the matrix is diagonalisable and has positive eigenvalues. We find these two properties hold in our experiments; note however that $P^{(T-t)}$ is in general not symmetric. We shall see in Section 2.3.3 that instead of computing the matrix logarithm, one may alternatively compute the matrix root corresponding to the required time step.

2.3.3 Time discretisation

Let us now discretise (2.3.1) with respect to time. We denote the approximation at state x_i and time step t_k by $v_{i,k}$.

We use the θ-method, standard in the numerical solution to PDEs, to discretise (2.3.1). 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 $v^{(k)}$ of values at our grid points each at time t_k and discretise the first line of (2.3.1) as

$$\frac{v^{(k+1)} - v^{(k)}}{\delta t_k} + A\left((1-\theta)v^{(k+1)} + \theta v^{(k)}\right) \leq 0 \tag{2.3.5}$$

where $\delta t_k \triangleq t_{k+1} - t_k$. Thus (2.3.1) becomes

$$\begin{cases} \left(I + (1-\theta)A\delta t_k\right) v^{(k+1)} - \left(I - \theta A \delta t_k\right) v^{(k)} & \leq & 0 \\ v^{(k)} - \psi & \geq & 0 \\ \left(\left(I + (1-\theta)A\delta t_k\right) v^{(k+1)} - \left(I - \theta A \delta t_k\right) v^{(k)}\right)' \left(v^{(k)} - \psi\right) & = & 0. \end{cases} \tag{2.3.6}$$

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

$$M_L = \exp\left\{-\theta A \delta t_k\right\} \tag{2.3.7}$$

$$M_R = \exp\left\{(1-\theta) A \delta t_k\right\}. \tag{2.3.8}$$

The approximating complementarity problem to solve is then

$$\begin{cases} M_R v^{(k+1)} - M_L v^{(k)} & \leq & 0 \\ v^{(k)} - \psi & \geq & 0 \\ \left(M_L v^{(k)} - M_R v^{(k+1)}\right)' \left(v^{(k)} - \psi\right) & = & 0 \end{cases} \tag{2.3.9}$$

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

Numerically we must solve an LCP at each time step, for which the PSOR method of Cryer [25] 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 (2.3.9); for example, see Dempster and Hutton [26] for American option pricing using linear programming in the one-dimensional case.

An error analysis of the discretisation in (2.3.5) may be undertaken along the lines of Glowinski et al. [35] on variational inequalities or that of Kushner and Dupuis [46] on stochastic control. Chapter 6 formulates sufficient conditions for convergence using the framework of [35].

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

$$M_L = \left(P^{(T-t)}\right)^{-\theta \delta t_k/(T-t)} \tag{2.3.10}$$

$$M_R = \left(P^{(T-t)}\right)^{(1-\theta)\delta t_k/(T-t)}. \tag{2.3.11}$$

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 Golub and Van Loan [36]. 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 (2.3.12). We demonstrate the difference between the two constructions in Figure 2.3.1 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.

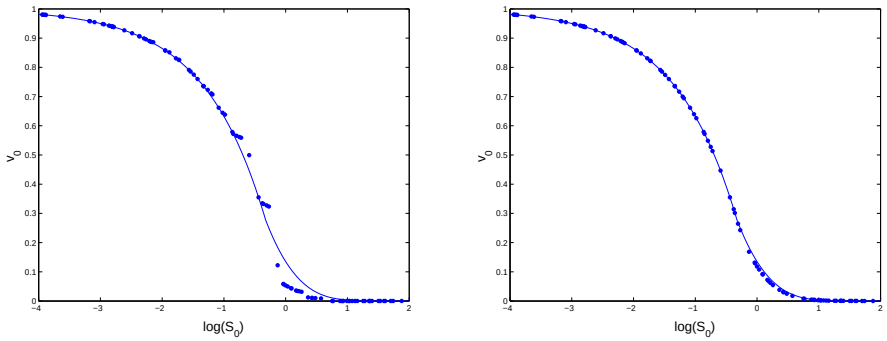


Figure 2.3.1: Random grid valuation of an American put option on a single asset with expiry 1, strike 1 and 100 asset points, using transition matrices $P^{(0.01)}$ and $\left(P^{(1)}\right)^{0.01}$, respectively (*dots*). The plots are in the log domain. Also plotted are values computed to high accuracy (*solid lines*) using a standard finite difference method.

Remark 2.3.1 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 2.3.2 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 realised as follows:

$$\begin{aligned} v_i(t) &= \left(\prod_{k=1}^{K} e^{-r\delta t} M_L^{-1} M_R \right) v^{(K)} \\ &= e^{-r(T-t)} \exp(A) v^{(K)} \\ &= e^{-r(T-t)} \sum_{j=1}^{n} \psi(x_j) p_{ij}^{(T-t)} \end{aligned} \tag{2.3.12}$$

where $v_i(t)$ is the value in state x_i at initial time t and $p_{ij}^{(T-t)}$ is defined in (2.3.2). 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, (2.3.12) holds only asymptotically as $\delta t \to 0$.

Equation (2.3.12) 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.

2.3.4 Randomisation

The QMC grids we have proposed are deterministic; however perturbing these points randomly allows us to observe the behaviour 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 Owen [59] for integration problems. The importance of randomised QMC is also emphasised in Glasserman [33].

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 realise randomised 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 realise the jth randomised Sobol' sequence as

$$\mathcal{S}_j = (s_{j,i})_{i \in \mathbb{N}} = (s_i + U_j \mod 1)_{i \in \mathbb{N}}. \tag{2.3.13}$$

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

2.3.5 Summary of procedure

We now present a concise statement of the proposed procedure as Algorithm 2.3.1. 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 2.3.1 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 (2.3.9)
  Let $\hat{v}_{i,j}$ 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

2.4 Experimental setup and details

We now use the algorithm presented in Section 2.3.5 to estimate prices of multi-asset options. We first present a detailed exposition of the setting, experimental procedure and various considerations. Numerical results are presented in the next section.

2.4.1 Specification of dynamics

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

$$dX = \left(r\mathbb{1} - \delta - \tfrac{1}{2}\,\mathrm{diag}(\Sigma)\right)dt + R'dW \tag{2.4.1}$$

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, $\Sigma = (\rho_{ij}\sigma_i\sigma_j)$ 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

$$\mathcal{L} = \tfrac{1}{2}\sum_{i,j=1}^{d} \rho_{ij}\sigma_i\sigma_j \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d}(r - \delta_i - \tfrac{1}{2}\sigma_i^2)\frac{\partial}{\partial x_i} - r. \tag{2.4.2}$$

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

$$X_0(s) = X(s) - (s-t)\mu, \tag{2.4.3}$$

where μ is the risk-neutral drift; for example in (2.4.1) we have $\mu = r\mathbb{1} - \delta - \frac{1}{2}\operatorname{diag}(\Sigma)$. The new process X_0 has zero drift and the covariance Σ is unchanged:

$$dX_0(s) = RdW(s). \tag{2.4.4}$$

The payoff under the reformulation is

$$\psi_0(x_i, s) = \psi\left(x_i + (s-t)\mu\right). \tag{2.4.5}$$

One may also eliminate a deterministic, time-dependent risk-neutral drift by subtracting $\int_t^s \mu(u)du$ in (2.4.3).

2.4.3 Grid specification

We consider normal RQMC grids as suggested in Section 2.3.4; 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 Evans and Swartz [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 minimisation 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 (2.4.3), 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.

Summarising, we suggest the parameters

$$\mu_g = x_t \tag{2.4.6}$$

$$\Sigma_g = \alpha\Sigma(T-t). \tag{2.4.7}$$

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

$$x_{i+1} = \mu_g + R_g' \left(\Psi^{-1}(s_{j,i,1}) \cdots \Psi^{-1}(s_{j,i,d})\right)' \tag{2.4.8}$$

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 Figure 2.4.1. It should be noted however that the advantage of using an irregular grid is realised in dimensions of at least three.

2.4.4 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.

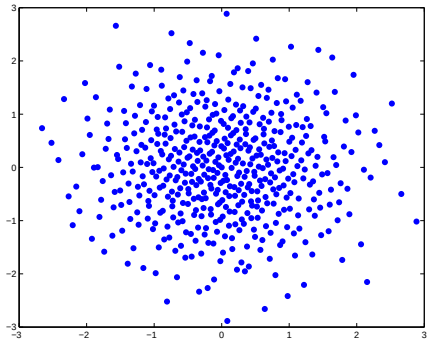


Figure 2.4.1: Example of normal QMC grid in 2 dimensions with 500 points.

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 2.4.1 *Suppose that $P^{(T-t)}$ is the transition matrix corresponding, through (2.3.2) and (2.3.3), 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*

$$\mathcal{Y} = \{y_1, \ldots, y_n\} = \{R'x_1, \ldots, R'x_n\}, \tag{2.4.9}$$

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

$$\tilde{P}^{(T-t)} = P^{(T-t)}. \tag{2.4.10}$$

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

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

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 normalised entries $p_{ij}^{(T-t)}$ and $\tilde{p}_{ij}^{(T-t)}$ obtained through (2.3.2) and (2.3.3) are also equal.

■

Remark 2.4.1 One may ask whether the weights g_i specified in Lemma 2.4.1 are indeed appropriate for the grid $\mathcal{Y}$. That is, whether (2.3.3) 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 $S \subset \mathbb{R}^d$,

$$P(x \in S) = \int_S g_X(x)dx.$$

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 $S \subset \mathbb{R}^d$,

$$\begin{aligned} P(y \in R'S) &= \int_{R'S} g_Y(y)dy \\ &= \int_S g_Y(R'x)\big||R|\big|dx \end{aligned} \tag{2.4.11}$$

by the multivariate substitution formula where $R'S = \{R'x : x \in S\}$ and $\big||R|\big| = \operatorname{abs}(\det R)$. But since $y = R'x$, $P(y \in R'S) = P(x \in S)$, and since (2.4.11) holds for all $S \subset \mathbb{R}^d$, we conclude that

$$g_X(x) = \big||R|\big|g_Y(R'x). \tag{2.4.12}$$

Finally, note that the averaging taking place in (2.3.3) 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 2.4.2 A time-dependent scaling of the covariance can also be incorporated by using the matrix logarithm, and constructing the time stepping matrices through (2.3.7) and (2.3.8) rather than (2.3.10) and (2.3.11).

2.4.5 Low-biased estimate

As is common practise 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 neighbour in the grid at the closest time. Specifically one may define the exercise rule for grid points to be

$$\tau(x_i, t_k) \triangleq \begin{cases} 1 & \text{if } v_i^{(k)} = \psi_i \\ 0 & \text{otherwise} \end{cases} \tag{2.4.13}$$

and for general points $x \in \mathbb{R}^d$ and times $s \in [t, T]$

$$\tau(x, s) \triangleq \begin{cases} 1 & \text{if } v_i^{(k)} = \psi_i \\ 0 & \text{otherwise} \end{cases} \tag{2.4.14}$$

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

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 could also implement weighted schemes with respect to x and t rather than using nearest neighbour rules.

2.4.6 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 Rogers [64], Haugh and Kogan [37]) in which one minimises the cost of hedging by minimising 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 practise, 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 2.5 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.

2.4.7 Benchmarks

There are few benchmark results for high-dimensional American options. Broadie and Glasserman [19] 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 [50] price the Broadie and Glasserman maximum options using the LSM method.

Stentoft [68] uses the binomial method of Boyle et al. [13] 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 [18] and Fu et al. [31] provide benchmark results for options over five assets with three exercise

opportunities. Finally, Rogers [64] and Haugh and Kogan [37] 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 (2.4.1), and the payoff function

$$\psi(s) = \left(K - \left(\prod s_i \right)^{1/d} \right)^+ \tag{2.4.15}$$

where x^+ denotes the positive part of x, 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

$$\begin{aligned} dY(s) &= \frac{1}{d} \sum_{i=1}^{d} dX_i(s) && (2.4.16) \\ &= \tilde{\mu} ds + \tilde{\sigma} dW(s). && (2.4.17) \end{aligned}$$

The parameters of the diffusion are given by

$$\begin{aligned} \tilde{\mu} &= r - \frac{1}{2d} \sum_{i=1}^{d} \sigma_i^2 && (2.4.18) \\ \tilde{\sigma}^2 &= \frac{1}{d^2} \sum_{i=1}^{d} \left(\sum_{j=1}^{d} R_{ij} \right)^2 . && (2.4.19) \end{aligned}$$

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%.

2.5 Experimental results

Our experiments are conducted in a Matlab environment and are based on the five-dimensional examples of Stentoft [68]. 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 2.3. 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 2.3.4 using the parameter values $\alpha = 1,\ 1.5$ and 2, respectively (these were found to give the best rates of convergence). The number of grids need not be so high in practise, depending on the accuracy required. The vector of initial stock prices x_0 was always included in the grid.

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

$$\begin{aligned} \psi_1(s) &= \left(K - \tfrac{1}{d}\textstyle\sum s_i\right)^+ & \psi_2(s) &= \left(K - (\textstyle\prod s_i)^{1/d}\right)^+ \\ \psi_3(s) &= \left(K - \max(s_i)\right)^+ & \psi_4(s) &= \left(K - \min(s_i)\right)^+. \end{aligned} \tag{2.5.1}$$

Figures 2.5.1–2.5.3 show the convergence behaviour of the irregular grid method where the implicitness parameter is $\theta = 0,\ \frac{1}{2}$ and 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 arithmetic and geometric 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 geometric 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 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 Figures 2.5.1–2.5.3 is achieved with α (the ratio of grid density to process density) being 1.5. We thus present in Tables 2.5.1 and 2.5.2 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 2.5.1 and 2.5.2 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 Stentoft [68], to obtain out-of-sample values for the LSM algorithm (these results are not given in [68]). Our LSM implementation also reproduced (up to a statistically insignificant difference) the in-sample LSM results given in [68]. For details of how out-of-sample paths are used in the LSM method to obtain low-biased estimators, we refer the reader to Longstaff and Schwartz [50].

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 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

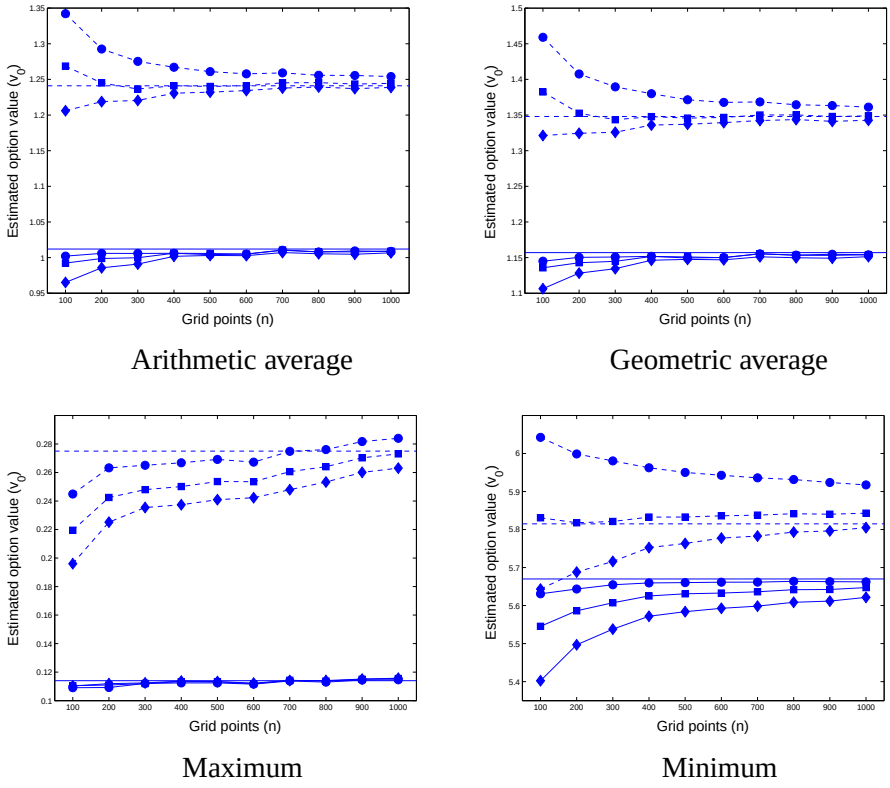


Figure 2.5.1: Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options over five assets using the explicit method ($\theta = 0.0$) and ten time steps. Stentoft's Bermudan LSM solutions are drawn as horizontal lines.

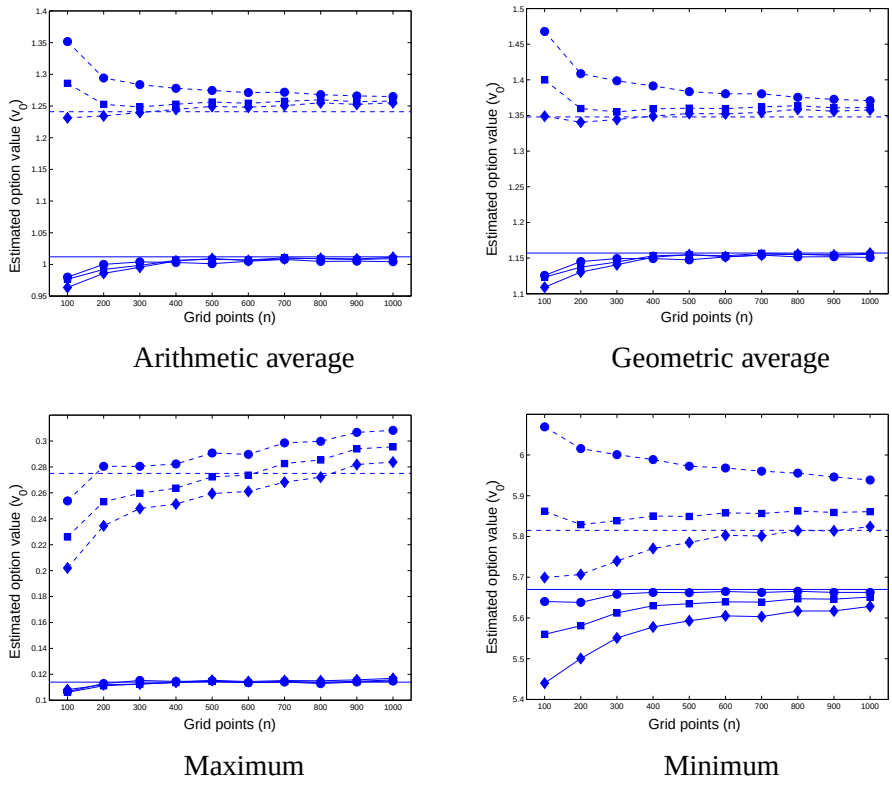


Figure 2.5.2: Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options over five assets using the Crank-Nicolson method ($\theta = 0.5$) and ten time steps. Stentoft's Bermudan LSM solutions are drawn as horizontal lines.

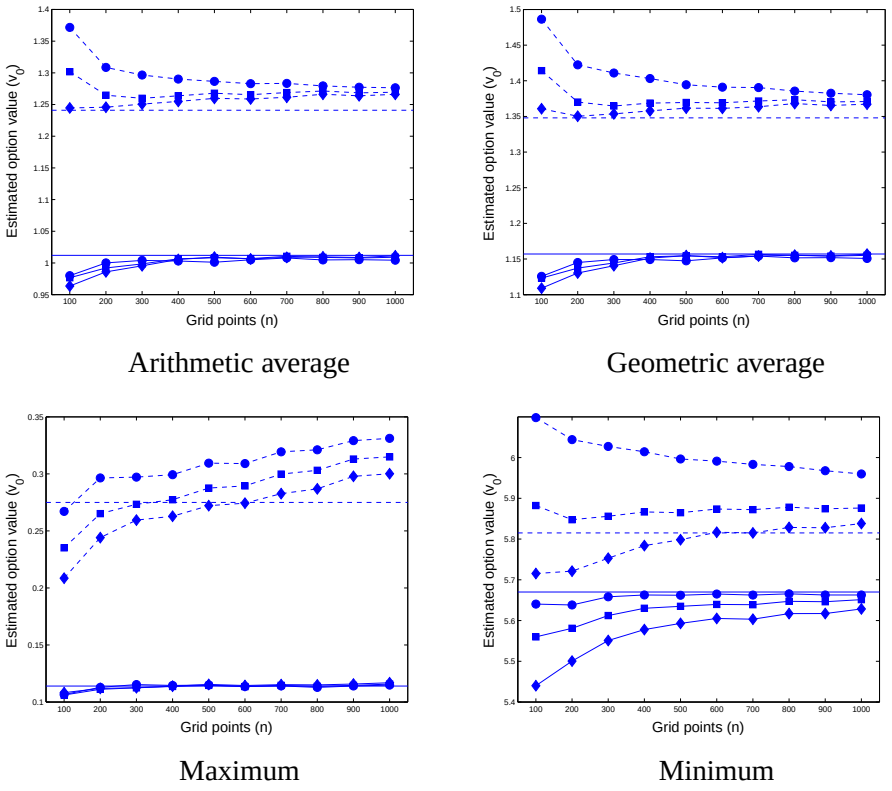


Figure 2.5.3: Average QMC grid valuation over 50 normal grids with $\alpha = 1.0$ (*circles*), $\alpha = 1.5$ (*squares*), $\alpha = 2.0$ (*diamonds*) of European (*solid lines*) and American (*dotted lines*) put options over five assets using the implicit method ($\theta = 1.0$) and ten time steps. Stentoft's Bermudan LSM solutions are drawn as horizontal lines.

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)

Table 2.5.1: Comparison of Bermudan price estimates ($\theta = 0$) with ten exercise opportunities. 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. [13] 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 [68] 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.

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)

Table 2.5.2: Comparison of American price estimates ($\theta = 0.5$). 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 2.4.6. 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.

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.

In Table 2.5.1 it is encouraging to see that the irregular grid prediction for the geometric average option is very accurate as compared to the benchmark. The exercise strategy performs well for the arithmetic and geometric average options, but not for the more problematic maximum and minimum payoffs.

As detailed in Section 2.4.6, 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 neighbours for the delta estimation, we obtain the results shown in the last column of Table 2.5.2. 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 866MHz machine are presented in Table 2.5.3. 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.

2.6 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.

Size P (n)	Memory full (MB)	Memory sparse (MB)	Time for P (sec)	Time for $P^{1/10}$ (sec)	Prop $P^{1/10}$ nonzero	Time-stepping (sec)
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

Table 2.5.3: Timings and storage requirements for the irregular grid method using Matlab 6.1 with a PIII 866MHz processor with 512 MB RAM, 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. 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 chapter, 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.

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 neighbouring 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 [68]. 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 [50] 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 [46], but may also take inspiration from the literature on nonparametric analysis. These and other possible refinements are reserved for future investigation.

Results relating to the convergence of this algorithm are provided in Chapter 6. Sufficient conditions are provided in that chapter, although it is still to be determined whether these conditions are satisfied for the method presented in this chapter.

Chapter 3

A Method Using Local Quadratic Approximations

3.1 Introduction

Recent literature tends to support the widely-held belief that pricing American options in a high-dimensional setting is a highly nontrivial task. Methods proposed by Carrière [20], Longstaff and Schwartz [50], Rogers [64] and Haugh and Kogan [37] give good results in many situations, but are sensitive to the choice of basis. Chapter 2 (also published as Berridge and Schumacher [8, 7, 9]) proposes a method which does not require the choice of a basis, as do Bally and Pagès [3].

The difficulty with primal methods in a high-dimensional problem is determination of the expected continuation value. This is essentially an integral with respect to the risk neutral process and the optimal stopping rule. The latter poses the most difficulty since it is itself a high-dimensional function which must be approximated.

In classical finite difference methods the expected continuation value is determined through an approximation to the operator on a regular grid. In the irregular grid setting this approximation becomes more difficult, mainly due to instabilities in the approximation. The construction of stable approximations will be investigated in this chapter through using a local polynomial representation of the value function.

Once a stable approximation is found, one has access to finite difference tech-

niques for time stepping, including the explicit, Crank-Nicolson and implicit methods. In general, these θ-methods lead to linear complementarity problems (LCPs), which can be solved using the projected successive overrelaxation (PSOR) method introduced by Cryer [25] or by linear programming, for example.

It is of interest to note the connection between the methods presented in this chapter and the literature on mesh-free methods (for an overview see Liu [48]). In particular the method of moving least squares provides an approach to partial differential equation (PDE) solution which is similar to that found in this chapter in that an irregular grid is used on which to calculate local quadratic approximations. The essential difference is that our approach does not use weak forms, and thus does not require integration of basis functions. It seems the method we present is more sensitive to the structure of the grid, producing unstable approximations in some circumstances. The MLS method is applied to option pricing in Chapter 5.

The remainder of this chapter is organised as follows. In Section 3.2 we introduce the setting, the problem and a review of irregular grid methodology from Chapter 2. Section 3.3 introduces the local quadratic approximation method; Sections 3.4 and 3.5 present results using this method for regular and irregular grids, respectively. Finally Section 3.6 presents conclusions and ideas for future research.

3.2 The pricing problem

3.2.1 Formulation

We consider an American option at time t on d underlying assets with values $X(s) = (X_1(s), \ldots, X_d(s))'$ at time $s \in [t, T]$, payoff $\psi(X(s), s)$ and expiry T. The assets follow the diffusion

$$dX_i(s) = \mu_i(X(s), s)ds + \sigma_i(X(s), s)dW(s) \tag{3.2.1}$$

for $i = 1, \ldots, d$ and where $X(t)$ is known, $dW(s)$ are the increments of a standard d-dimensional Brownian motion and the μ_i and σ_i are measurable with respect to the filtration generated by the Brownian motion.

The price of an American option, giving the long party the right to receive the payoff $\psi(X(s), s)$ at any time $s \in [t, T]$, is given in the primal formulation as

$$v(X(t), t) = \sup_{\tau \in \mathcal{T}} \mathbb{E}^{\mathbb{Q}}_{X(t)} \left(e^{-r(\tau - t)} \psi(X(\tau), \tau) \right) \tag{3.2.2}$$

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 $X(t)$.

This can be reformulated as the following complementarity problem

$$\begin{cases} \frac{\partial v}{\partial t} + \mathcal{L}v & \leq & 0 \\ v - \psi & \geq & 0 \\ \left(\frac{\partial v}{\partial t} + \mathcal{L}v\right)(v - \psi) & = & 0 \end{cases} \tag{3.2.3}$$

for $(x, s) \in \mathbb{R}^d \times [t, T]$ with the terminal condition $v(\cdot, T) \equiv \psi(\cdot, T)$. Here $\mathcal{L}$ is the Black-Scholes operator implied by the diffusion.

3.2.2 Discretisation

One way to discretise the complementarity formulation is to sample the state space. This is the approach taken in finite difference methods, however the traditional grid approach is not suitable for high-dimensional problems due to the curse of dimensionality. We thus consider an irregular sampling of the state space on which to approximate the problem.

Suppose now that we are given some sampling of the state space, $\mathcal{X} = (x_1, \ldots, x_n) \subset \mathbb{R}^d$. We do not concentrate on the properties of the sampling, but we may assume it is a sequence of low discrepancy in the sense of Niederreiter [58] or low distortion in the sense of Bally and Pagès [3]. These quantities may be best measured in terms of the terminal distribution of the process, which in our setting is a multivariate normal distribution.

Having taken such a sample we approximate the complementarity problem (3.2.3) by the new complementarity problem

$$\begin{cases} \frac{dv}{dt} + Av & \leq & 0 \\ v - \psi & \geq & 0 \\ \left(\frac{dv}{dt} + Av\right)'(v - \psi) & = & 0 \end{cases} \tag{3.2.4}$$

where v is an n-vector and the ith component of ψ is $\psi(x_i, t)$. The matrix A should approximate $\mathcal{L}$ on our grid $\mathcal{X}$ in that

$$(Av(t))_i \simeq (\mathcal{L}v)(x_i, t). \tag{3.2.5}$$

Let us write v as $v = (v^{(1)} \cdots v^{(n)})'$ where each $v^{(i)} : [t, T] \to \mathbb{R}$.

3.2.3 Nearest neighbours

Just as in traditional finite difference methods, a great deal of efficiency can be gained by only considering local interactions. We thus use nearest neighbour sets on which to construct local approximations to $\mathcal{L}$.

The kth-nearest neighbour function $N_{k,\mathcal{X}} : \{1,\dots,n\} \to \{1,\dots,n\}$ for some set of points $\mathcal{X}$ is then defined as

$$N_{k,\mathcal{X}}(i) \triangleq \{j : \|x - x_i\| \leq \|x_j - x_i\| \text{ for exactly } k \text{ different } x,\, x \in \mathcal{X}\}\,.$$

Note that $N_{1,\mathcal{X}}(i) = i$, that is x_i is the nearest neighbour of x_i in this definition. Further let $\mathcal{N}_i = \{N_{j,\mathcal{X}}(i)\}_{j=1,\dots,k}$ be the ordered set of the k nearest neighbours for each i. For brevity we denote the jth nearest neighbour of point x_i as

$$x_i^{(j)} \equiv N_{j,\mathcal{X}}(i).$$

In addition to considering other points as neighbours, we may also allow boundary points to be neighbours. Thus in some situations we use the extended nearest neighbour function $N_{k,\mathcal{X}} : \{1,\dots,n\} \to \{1,\dots,n+2d\}$ where the $2d$ extra points are projections of x_i onto the boundaries.

3.3 Methodology

3.3.1 Approximating the differential operator

We consider now the construction of a direct approximation to A with respect to a grid $\mathcal{X}$. Let us form this approximation by assuming that v is approximately locally quadratic about each grid point x_i. This is justified since we know that v is almost everywhere continuously differentiable, and convenient since we wish to approximate the effect of the second order operator $\mathcal{L}$.

We write $\mathcal{L}$ as

$$\mathcal{L} = \alpha_{0,0} + \sum_{j=1}^{d} \alpha_{j,0} \frac{\partial}{\partial x_j} + \sum_{j=1}^{d} \sum_{k=1}^{j} \alpha_{j,k} \frac{\partial^2}{\partial x_j \partial x_k}$$

for some $\alpha_{j,k} \in \mathbb{R}$.

Now let us introduce the quadratic interpolant $\bar{v}^{(i)} : \mathbb{R}^d \to \mathbb{R}$ at grid point x_i as

$$\bar{v}^{(i)}(x) = a^{(i)}_{0,0} + \sum_{j=1}^{d} a^{(i)}_{j,0} x_j + \sum_{j=1}^{d} \sum_{k=1}^{j} a^{(i)}_{j,k} x_j x_k \tag{3.3.1}$$

where the $a^{(i)}_{j,k}$ are chosen so that $\bar{v}^{(i)}(x^{(j)}_i)(s) = v^{(j)}(t)$ for all $s \in [t, T]$ whenever $j \in \mathcal{N}_i$. By finding this interpolant for each i we have an approximation for $v(x, t)$ in a neighbourhood of all our grid points. Letting $\eta = \frac{1}{2}\left(d^2 + 3d + 2\right)$ (the number of parameters in each $\bar{v}^{(i)}$), we can determine a unique quadratic interpolant for each i. Denoting the kth component of the $j$the neighbour of x_i by $x^{(j)}_{i,k}$, the coefficients of the ith interpolant are

$$\begin{aligned}
&a^{(i)}(t) \\
&\equiv \left(a^{(i)}_{0,0}, a^{(i)}_{1,0}, \ldots, a^{(i)}_{d,0}, a^{(i)}_{1,1}, a^{(i)}_{2,1}, \ldots, a^{(i)}_{d,d}\right)' \\
&= \begin{pmatrix}
1 & x^{(1)}_{i,1} & \cdots & x^{(1)}_{i,d} & \left(x^{(1)}_{i,1}\right)^2 & x^{(1)}_{i,2} x^{(1)}_{i,1} & \cdots & \left(x^{(1)}_{i,d}\right)^2 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
1 & x^{(\eta)}_{i,1} & \cdots & x^{(\eta)}_{i,d} & \left(x^{(\eta)}_{i,1}\right)^2 & x^{(\eta)}_{i,2} x^{(\eta)}_{i,1} & \cdots & \left(x^{(\eta)}_{i,d}\right)^2
\end{pmatrix}^{-1} \begin{pmatrix} v^{(i)}_1 \\ \vdots \\ v^{(i)}_\eta \end{pmatrix} \\
&\equiv \left(M^{(i)}\right)^{-1} v^{(i)}(t)
\end{aligned} \tag{3.3.2}$$

assuming that the matrix $M^{(i)}$ is nonsingular.

Now let us consider the effect of the operator $\mathcal{L}$ on $\bar{v}^{(i)}$. Denoting now the ith grid point by $x^{(i)}$, we note that

$$\begin{aligned}
\frac{\partial \bar{v}^{(i)}}{\partial x_j}(x^{(i)}) &= a^{(i)}_{j,0} + \sum_{k=1}^{d} a^{(i)}_{j,k}(1 + \delta_{j,k}) x^{(i)}_k \\
\frac{\partial^2 \bar{v}^{(i)}}{\partial x_j \partial x_k}(x^{(i)}) &= a^{(i)}_{j,k}(1 + \delta_{j,k})
\end{aligned}$$

where $\delta_{j,k}$ is the Kronecker delta function. Hence the effect is

$$(\mathcal{L}\bar{v}^{(i)})(x^{(i)}) = \left[\alpha_{0,0}\bar{v}^{(i)} + \sum_{j=1}^{d} \alpha_{j,0} \frac{\partial \bar{v}^{(i)}}{\partial x_j} + \sum_{j=1}^{d} \sum_{k=1}^{j} \alpha_{j,k} \frac{\partial^2 \bar{v}^{(i)}}{\partial x_j \partial x_k}\right](x^{(i)})$$

$$
\begin{aligned}
&= \alpha_{0,0}\left\{a_{0,0}^{(i)} + \sum_{j=1}^{d} a_{j,0}^{(i)} x_j^{(i)} + \sum_{j=1}^{d}\sum_{k=1}^{j} a_{j,k}^{(i)} x_j^{(i)} x_k^{(i)}\right\} \\
&+ \sum_{j=1}^{d} \alpha_{j,0}\left\{a_{j,0}^{(i)} + \sum_{k=1}^{d} a_{j,k}^{(i)}(1+\delta_{j,k}) x_k^{(i)}\right\} \\
&+ \sum_{j=1}^{d}\sum_{k=1}^{j} \alpha_{j,k}\left\{a_{j,k}^{(i)}(1+\delta_{j,k})\right\} \\
&= a_{0,0}^{(i)}\left\{\alpha_{0,0}\right\} + \sum_{j=1}^{d} a_{j,0}^{(i)}\left\{\alpha_{0,0}x_j^{(i)} + \alpha_{j,0}\right\} \\
&+ \sum_{j=1}^{d}\sum_{k=1}^{j} a_{j,k}^{(i)}\left\{\alpha_{0,0}x_j^{(i)}x_k^{(i)} + (1+\delta_{j,k})\alpha_{j,k} + \left(\alpha_{k,0}x_j^{(i)} + \alpha_{j,0}x_k^{(i)}\right)\right\}
\end{aligned}
$$

or in matrix form

$$
\begin{aligned}
(\mathcal{L}\bar{v}^{(i)})(x^{(i)}) &= \begin{pmatrix} \alpha_{0,0} \\ \alpha_{1,0} + \alpha_{0,0}x_1^{(i)} \\ \vdots \\ \alpha_{d,0} + \alpha_{0,0}x_d^{(i)} \\ \alpha_{0,0}\left(x_1^{(i)}\right)^2 + 2\alpha_{1,0}x_1^{(i)} + 2\alpha_{1,1} \\ \left(\alpha_{0,0}x_2^{(i)}x_1^{(i)} + \alpha_{2,1}\right) + \left(\alpha_{1,0}x_2^{(i)} + \alpha_{2,0}x_1^{(i)}\right) \\ \vdots \\ \alpha_{0,0}\left(x_d^{(i)}\right)^2 + 2\alpha_{d,d} + 2\alpha_{d,0}x_d^{(i)} \end{pmatrix}' \begin{pmatrix} a_{0,0}^{(i)} \\ a_{1,0}^{(i)} \\ \vdots \\ a_{d,0}^{(i)} \\ a_{1,1}^{(i)} \\ a_{2,1}^{(i)} \\ \vdots \\ a_{d,d}^{(i)} \end{pmatrix} \\
&= \beta(\alpha, x^{(i)})' a^{(i)}(t)
\end{aligned}
$$

and substituting for a and evaluating at $x^{(i)}$ we have

$$
\begin{aligned}
(\mathcal{L}\bar{v}^{(i)})(x^{(i)}) &= \beta(\alpha, x^{(i)})'\left(M^{(i)}\right)^{-1} v^{(i)}(t) \\
&= A^{(i)} v^{(i)}(t) \qquad (3.3.3)
\end{aligned}
$$

where $A^{(i)}$ is a row vector of length η which we can think of as containing the elements of the ith row of some matrix A which defines the constrained system of ordinary differential equations (3.2.4).

In order to unify these n equations into the form of Equation 3.2.4, introduce an operator $\mathcal{S}_{\mathcal{X}} : \mathbb{R}^{\eta} \to \mathbb{R}^{n}$ which stretches and rearranges the η-vector $v^{(i)}$ so that it becomes an n-vector with the entries placed in positions corresponding to the nearest neighbours of $x^{(i)}$.

$$\mathcal{S}_{\mathcal{X}}(x, i) \triangleq \left(0, \ldots, 0, x^{(j_1)}, 0, \ldots, 0, x^{(j_2)}, 0, \ldots, 0, x^{(j_\eta)}, 0, \ldots, 0\right)$$

where $x^{(j_k)}$ is in the $N_{k,\mathcal{X}}(i)$th position (in particular $x^{(i)}$ is in the ith position). Now let

$$A \triangleq \begin{pmatrix} \mathcal{S}_{\mathcal{X}}(A^{(1)}, 1) \\ \vdots \\ \mathcal{S}_{\mathcal{X}}(A^{(n)}, n) \end{pmatrix}. \tag{3.3.4}$$

3.3.2 Weighted least squares

As a simple extension to the above one may consider a least squares regression using $\xi > \eta$ nearest neighbours. As a further extension one may consider weighting the points in the regression according to their distance from x_i.

Letting $M^{(i)}$ be the matrix in (3.3.2), but now with ξ rows. Furthermore let Λ denote a diagonal weighting matrix, yet to be specified. The least squares criterion then gives

$$a^{(i)}(t) = \left(M^{(i)\prime}\Lambda^2 M^{(i)}\right)^{-1} M^{(i)\prime}\Lambda v^{(i)}(t). \tag{3.3.5}$$

Now, as in (3.3.3), we have the approximation

$$A^{(i)} \triangleq \beta(\alpha, x^{(i)})' \left(M^{(i)\prime}\Lambda^2 M^{(i)}\right)^{-1} M^{(i)\prime}\Lambda \tag{3.3.6}$$

and we define A from the $A^{(i)}$ as in (3.3.4).

3.3.3 Time stepping

We can now discretise time using a θ-method. Let $t_k = \frac{kT}{K}$ for some δt, $k \in \{0, \ldots, K\}$ and $\theta \in [0, 1]$, which can be thought of as the implicitness.

In the unconstrained case we form the finite difference equation

$$\frac{v_{k+1} - v_k}{\delta t} + (1 - \theta)Av_{k+1} + \theta Av_k = 0$$

where $v_k \equiv v(t_k)$. Then the set of equations we have to solve at each time step is

$$(I + \delta t(1-\theta)A)\, v_{k+1} = (I - \delta t\theta A)\, v_k$$

where the initial conditions are given by $v_K = \psi(x, T)$.

In the constrained case, we obtain the complementarity problem

$$\begin{cases} (I + (1-\theta)A\delta t)\, v_{k+1} - (I - \theta A\delta t)\, v_k & \leq & 0 \\ v_k - \psi & \geq & 0 \\ ((I + (1-\theta)A\delta t)\, v_{k+1} - (I - \theta A\delta t)\, v_k)'\, (v_k - \psi) & = & 0 \end{cases} \tag{3.3.7}$$

which can be solved using Cryer's PSOR [25] or linear programming, for example.

3.3.4 Stability

The stability of the time stepping algorithm depends crucially on the eigenvalues of the matrix A. In particular, the stability of the time stepping method in the unconstrained case requires that real eigenvalues of A must be nonpositive. The mapping of eigenvalues from the matrix A to the time stepping matrix is shown in Figure 3.3.1.

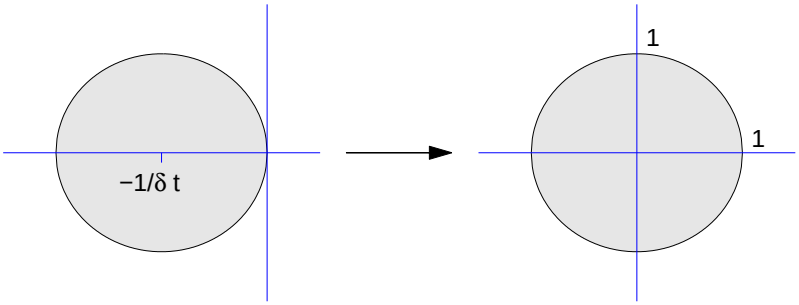
Algebraically, stability can be guaranteed through the diagonal dominance condition

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}| \tag{3.3.8}$$

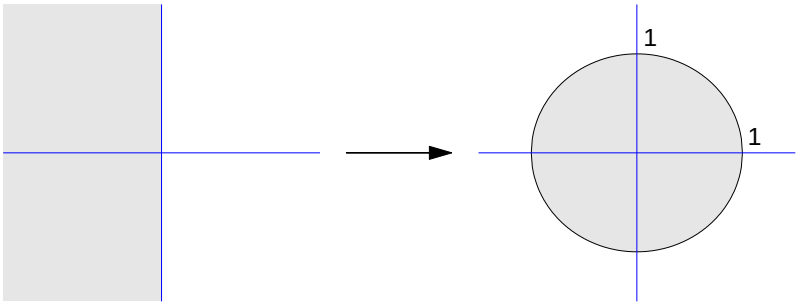
where $a_{ii} \leq 0$ for all i. This condition ensures the real parts of the eigenvalues of A are negative, a direct consequence of the Gershgorin disc theorem. Since row sums are zero, this also implies that off-diagonal entries must be nonnegative. The condition (3.3.8) is not necessary for stability however.

3.4 Application to regular grid

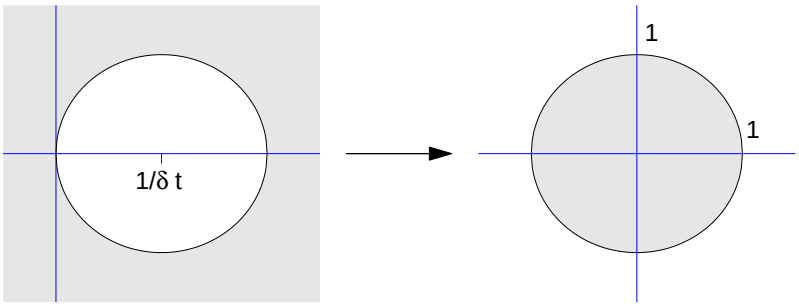
Before applying the discretisation method in its generality, we would first like to investigate its behaviour on regular grids. In particular it is of interest to compare the irregular grid method to standard finite difference methods. In the following analysis we consider only internal grid elements, and not boundary elements.



(a) Explicit scheme



(b) Crank-Nicolson scheme



(c) Implicit scheme

Figure 3.3.1: Mapping of eigenvalues from generator matrix A to time stepping matrix M in the explicit, Crank-Nicolson and implicit schemes. The shaded areas correspond to stable eigenvalues.

3.4.1 One dimension

The standard finite difference method on a regular grid approximates the first and second derivatives as

$$\frac{\partial v}{\partial x}(x_i) \simeq \frac{v_{i+1} - v_{i-1}}{2\delta x}, \qquad \frac{\partial^2 v}{\partial x^2}(x_i) \simeq \frac{v_{i+1} - 2v_i + v_{i-1}}{\delta x^2} \tag{3.4.1}$$

thus leading to the standard finite difference matrix A which has nonzero components in row i of A_i where

$$A_i = \left(\frac{\alpha_{11}}{\delta^2} - \frac{\alpha_{10}}{2\delta}, \alpha_{00} - 2\frac{\alpha_{11}}{\delta^2}, \frac{\alpha_{11}}{\delta^2} + \frac{\alpha_{10}}{2\delta}\right). \tag{3.4.2}$$

This implies the approximation

$$\left(\mathcal{L}v\right)(x_i) \simeq A_i \begin{pmatrix} v_{i-1} \\ v_i \\ v_{i+1} \end{pmatrix}. \tag{3.4.3}$$

Consider now the irregular grid approximation introduced in the previous section. Since we are using a regular grid, let us choose our x to be multiples of some δ and let $x_i = i\delta$ for all i. Making use of three nearest neighbours, we then have

$$M^{(i)} = \begin{pmatrix} 1 & (i-1)\delta & (i-1)^2\delta^2 \\ 1 & i\delta & i^2\delta^2 \\ 1 & (i+1)\delta & (i+1)^2\delta^2 \end{pmatrix} \tag{3.4.4}$$

and thus our approximation to $\mathcal{L}$ at x_i is

$$\begin{aligned} A^{(i)} &= \beta(\alpha, x^{(i)})' \left(M^{(i)}\right)^{-1} \\ &= \begin{pmatrix} \alpha_{00} \\ \alpha_{10} + \alpha_{00} i\delta \\ \alpha_{00} i^2\delta^2 + 2\alpha_{10} i\delta + 2\alpha_{11} \end{pmatrix}' \begin{pmatrix} \frac{1}{2}i(i+1) & 1 - i^2 & \frac{1}{2}i(i-1) \\ -\frac{1}{2\delta}(2i+1) & \frac{1}{2\delta}i & \frac{1}{2\delta}(1-2i) \\ \frac{1}{2\delta^2} & -\frac{1}{\delta^2} & \frac{1}{2\delta^2} \end{pmatrix} \\ &= \left(\frac{\alpha_{11}}{\delta^2} - \frac{\alpha_{10}}{2\delta}, \alpha_{00} - 2\frac{\alpha_{11}}{\delta^2}, \frac{\alpha_{11}}{\delta^2} + \frac{\alpha_{10}}{2\delta}\right). \end{aligned}$$

Hence we see that for a regular grid in one dimension the method of the previous section is equivalent to the standard finite difference method.

7	5	9
2	1	3
6	4	8

Table 3.4.1: Assignment of indices of v to neighbours.

3.4.2 Two dimensions

We now compare the irregular grid method to the standard finite difference method on a regular grid in two dimensions. For the finite difference scheme, in addition to the derivative approximations given above for one dimension we introduce the standard cross-derivative approximation (see for example Wilmott [73])

$$\frac{\partial^2 v}{\partial x_1 \partial x_2}(x_i, x_j) \simeq \frac{v_{i+1,j+1} - v_{i-1,j+1} - v_{i+1,j-1} + v_{i-1,j-1}}{4\delta x^2}. \tag{3.4.5}$$

In this case the ith row of the approximation matrix is

$$A_i = (\,\alpha_{00} - \frac{2(\alpha_{11}+\alpha_{22})}{\delta^2}, -\frac{\alpha_{10}}{2\delta} + \frac{\alpha_{11}}{\delta^2}, \frac{\alpha_{10}}{2\delta} + \frac{\alpha_{11}}{\delta^2}, \dots$$
$$-\frac{\alpha_{20}}{2\delta} + \frac{\alpha_{22}}{\delta^2}, \frac{\alpha_{20}}{2\delta} + \frac{\alpha_{22}}{\delta^2}, \frac{\alpha_{12}}{4\delta^2}, -\frac{\alpha_{12}}{4\delta^2}, -\frac{\alpha_{12}}{4\delta^2}, \frac{\alpha_{12}}{4\delta^2}\,)$$

implying the approximation

$$(\mathcal{L}v)(x_i) \simeq A_i \begin{pmatrix} v_{i,j} \\ v_{i-1,j} \\ v_{i+1,j} \\ v_{i,j-1} \\ v_{i,j+1} \\ v_{i-1,j-1} \\ v_{i-1,j+1} \\ v_{i+1,j-1} \\ v_{i+1,j+1} \end{pmatrix}. \tag{3.4.6}$$

Pictorially, we present in Table 3.4.1 the way in which entries in the vector v correspond to the points around $v_{i,j}$.

Nine neighbours

In order to compare the finite difference scheme to the irregular grid method, we consider nine nearest neighbours in the approximation. We thus use the least squares approach and denote $x_{i,j} = (i\delta, j\delta)$. The resulting approximation is

$$A^{(i)} = \begin{pmatrix} \frac{5}{9}\alpha_{00} - \dfrac{2(\alpha_{11}+\alpha_{22})}{3\delta^2} \\ \frac{2}{9}\alpha_{00} + \dfrac{2\alpha_{11} - \alpha_{10}\delta - 4\alpha_{22}}{6\delta^2} \\ \frac{2}{9}\alpha_{00} + \dfrac{2\alpha_{11} + \alpha_{10}\delta - 4\alpha_{22}}{6\delta^2} \\ \frac{2}{9}\alpha_{00} + \dfrac{2\alpha_{22} - \alpha_{20}\delta - 4\alpha_{11}}{6\delta^2} \\ \frac{2}{9}\alpha_{00} + \dfrac{2\alpha_{22} + \alpha_{20}\delta - 4\alpha_{11}}{6\delta^2} \\ -\frac{1}{9}\alpha_{00} + \dfrac{4(\alpha_{11}+\alpha_{22}) + 2(-\alpha_{20}-\alpha_{10})\delta + 3\alpha_{12}}{12\delta^2} \\ -\frac{1}{9}\alpha_{00} + \dfrac{4(\alpha_{11}+\alpha_{22}) + 2(\alpha_{20}-\alpha_{10})\delta - 3\alpha_{12}}{12\delta^2} \\ -\frac{1}{9}\alpha_{00} + \dfrac{4(\alpha_{11}+\alpha_{22}) + 2(-\alpha_{20}+\alpha_{10})\delta - 3\alpha_{12}}{12\delta^2} \\ -\frac{1}{9}\alpha_{00} + \dfrac{4(\alpha_{11}+\alpha_{22}) + 2(\alpha_{20}+\alpha_{10})\delta + 3\alpha_{12}}{12\delta^2} \end{pmatrix}' . \quad (3.4.7)$$

One can view this approximation as a weighted finite difference method, with the weightings being given in Tables 3.4.2–3.4.7. Note that the only case in which the weights are the same is for α_{12}; in all other cases the weights for the nine neighbour irregular grid scheme have less concentrated weights than in the finite difference method.

0	0	0
0	1	0
0	0	0

$$\frac{1}{9}$$

-1	2	-1
2	5	2
-1	2	-1

Table 3.4.2: Weights for α_{00}, for finite difference and irregular grid, respectively.

$\frac{1}{2\delta}$

0	0	0
-1	0	1
0	0	0

$\frac{1}{6\delta}$

-1	0	1
-1	0	1
-1	0	1

Table 3.4.3: Weights for α_{10}, for finite difference and irregular grid, respectively.

$\frac{1}{2\delta}$

0	1	0
0	0	0
0	-1	0

$\frac{1}{6\delta}$

1	1	1
0	0	0
-1	-1	-1

Table 3.4.4: Weights for α_{20}, for finite difference and irregular grid, respectively.

Six neighbours

Alternatively if we use six nearest neighbours, which is the minimum required to find a quadratic interpolant, we must make a choice between the diagonally located neighbours. Focusing on the points $x_{i,j}$,$x_{i,j-1}$,$x_{i,j+1}$,$x_{i-1,j}$,$x_{i+1,j}$ and $x_{i+1,j+1}$ we find that

$$A^{(i)} = \begin{pmatrix} \alpha_{00} - \frac{2(\alpha_{11}+\alpha_{22})-\alpha_{12}}{\delta^2} \\ -\frac{\alpha_{10}}{2\delta} + \frac{\alpha_{11}-\alpha_{12}}{\delta^2} \\ \frac{\alpha_{10}}{2\delta} + \frac{\alpha_{11}}{\delta^2} \\ -\frac{\alpha_{20}}{2\delta} + \frac{\alpha_{22}-\alpha_{12}}{\delta^2} \\ \frac{\alpha_{20}}{2\delta} + \frac{\alpha_{22}}{\delta^2} \\ \frac{\alpha_{12}}{\delta^2} \end{pmatrix}' . \tag{3.4.8}$$

We thus can find no equivalence between the irregular grid method and the standard finite difference method on a two dimensional regular grid, in contrast to the one dimensional case. One can however see the irregular grid method as a modified finite difference scheme in which different weights are used in the finite difference approximations.

We shall see in the Section 3.5 that the local quadratic approximation method

$\frac{1}{\delta^2}$

0	0	0
1	-2	1
0	0	0

$\frac{1}{3\delta^2}$

1	-2	1
1	-2	1
1	-2	1

Table 3.4.5: Weights for α_{11}, for finite difference and irregular grid, respectively.

$\frac{1}{\delta^2}$

0	1	0
0	-2	0
0	1	0

$\frac{1}{3\delta^2}$

1	1	1
-2	-2	-2
1	1	1

Table 3.4.6: Weights for α_{22}, for finite difference and irregular grid, respectively.

leads to a stable scheme when six neighbours are used as in (3.4.8), but an unstable scheme when nine neighbours are used as in (3.4.7).

3.5 Experimental results

In order to solve the complementarity problem related to the American option pricing problem, we must find a stable and convergent method for time stepping. As outlined in Section 3.3.4, a necessary condition for obtaining a stable time stepping matrix is that the real eigenvalues of A are nonpositive.

Having ascertained that A will lead to a stable time stepping scheme, it also remains to check the convergence conditions for the LCP solution method.

3.5.1 Grids on the unit cube in $\mathbb{R}^2$

We first consider grids on the region $\Omega = [0,1]^2 \subset \mathbb{R}^2$. This is a natural place to start since we know that finite difference schemes using regular grids lead to stable A matrices in this case. Boundaries are allowed to be neighbours in this setting.

We present in Figures 3.5.1–3.5.7 point sets of size approximately 500 and the eigenvalues of the corresponding A matrix obtained using the irregular grid method (plotted in the complex plane).

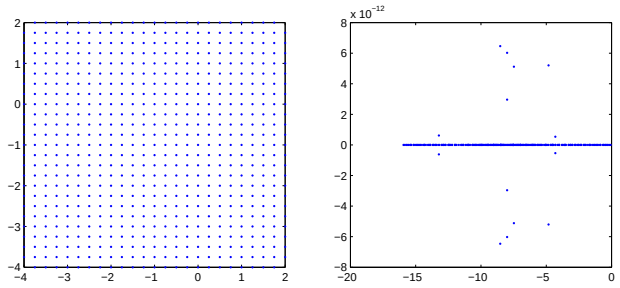


Figure 3.5.1: Points and eigenvalues of A for regular grid with 529 interior points and using 6 neighbours - stable. Note that the vertical scale in the eigenvalue plot is close to zero.

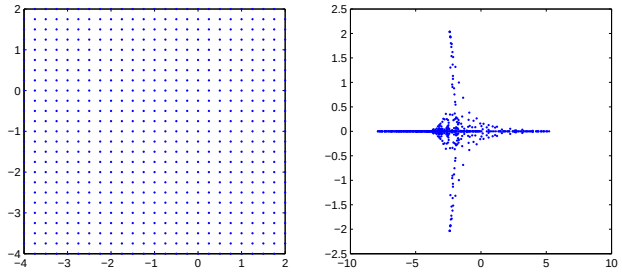


Figure 3.5.2: Points and eigenvalues of A for regular grid with 529 interior points and using 9 neighbours - unstable.

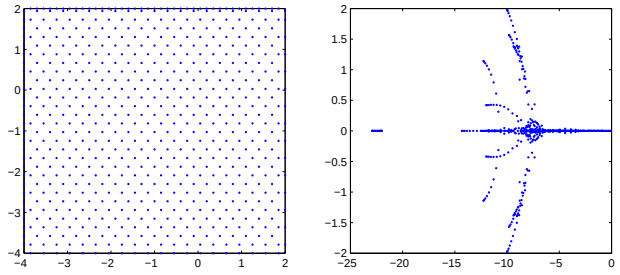


Figure 3.5.3: Points and eigenvalues of A for triangular grid with 546 interior points and using 6 neighbours - stable.

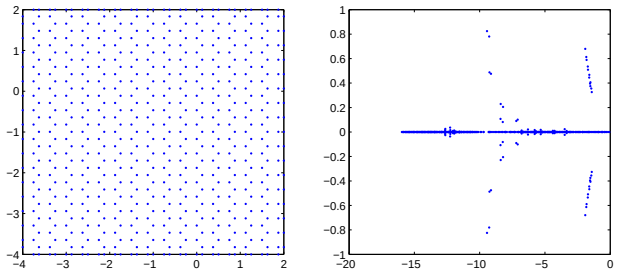


Figure 3.5.4: Points and eigenvalues of A for hexagonal grid with 512 interior points and using 6 neighbours - stable.

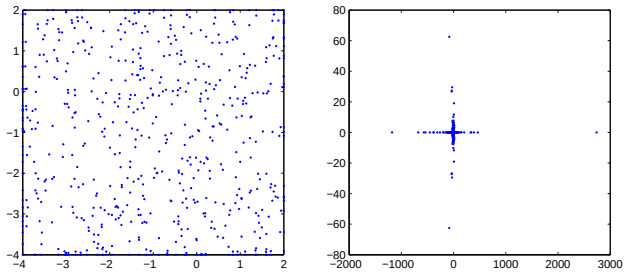


Figure 3.5.5: Points and eigenvalues of A for uniform pseudo-random grid with 500 interior points and using 6 neighbours - unstable.

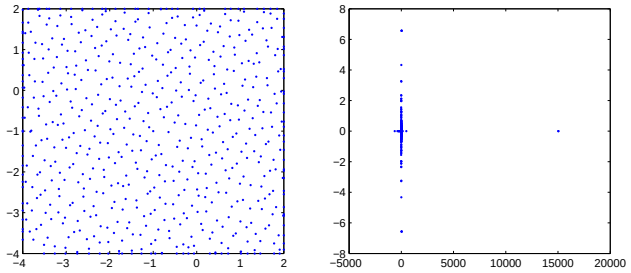


Figure 3.5.6: Points and eigenvalues of A for Sobol' grid with 500 interior points and using 6 neighbours - unstable.

$\frac{1}{4\delta^2}$

-1	0	1
0	0	0
1	0	-1

$\frac{1}{4\delta^2}$

-1	0	1
0	0	0
1	0	-1

Table 3.4.7: Weights for α_{12}, for finite difference and irregular grid, respectively.

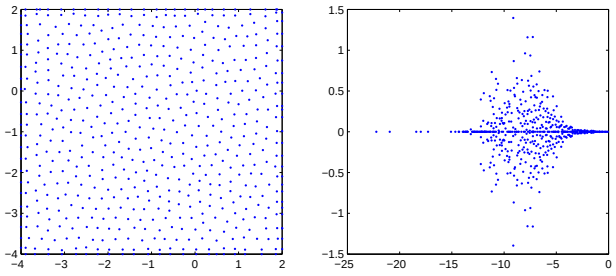


Figure 3.5.7: Points and eigenvalues of A for low distortion grid with 500 interior points and using 6 neighbours - stable.

In the case of the Sobol' grid one can in practise observe which points are causing instability by examining the eigenvector corresponding to the eigenvalues with positive real part. The instability can often be resolved by changing the neighbour configuration, in particular so that neighbours are well-distributed about the point. No systematic method was found to perform this stabilisation however.

Note that the least squares scheme using nine neighbours was not stable, despite the fact that it uses the same points as in the regular finite difference scheme. It was also found that none of the grids considered above leads to a stable A when considering local least squares fits over 7 neighbours.

3.5.2 Normally distributed grids in $\mathbb{R}^2$

Grids constructed on the unit cube are not optimal for the application under consideration. The main reason for this is that they are not representative of the regions of space that are likely to be visited by the stochastic process introduced in (3.2.1).

A further problem with using a grid on the unit cube is that neither Dirichlet

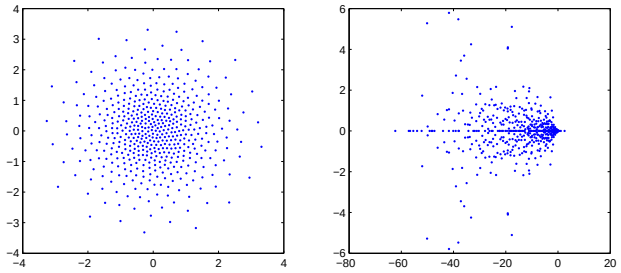


Figure 3.5.8: Points and eigenvalues of A for low distortion normal grid with 500 points and using $r = \infty$ and 6 neighbours - unstable, $\max(\Re(\lambda)) = 2.54$.

nor Neumann boundary conditions are known. Approximate conditions may be specified, thus adding an extra source of error to the computed solution.

A natural grid choice for our problem would be one that is related to the process, and in this case a normally distributed grid seems appropriate. This also alleviates the second problem mentioned, in that a normally distributed grid covers $\mathbb{R}^d$ asymptotically, and so boundary conditions have a vanishing effect on the solution.

Given the previous results of grids on the unit cube, we choose to focus on low distortion grids in the following. We generate a low distortion grid with respect to the standard normal density in $\mathbb{R}^2$, and apply the irregular grid method to it to obtain the matrix A. We then examine the eigenvalues of A.

Since there are no natural boundary points when dealing with normally distributed grids, we choose a radius outside which all points are considered to be boundary points. In particular we consider the radii $r = \infty$, 3.0, 2.5 and 2.0. In general one expects the eigenvalues to be different for different choices of r, and in particular that A should only be stable for smaller choices of r.

The results are presented in Figures 3.5.8 – 3.5.11. The maximum real parts of the eigenvalues are given in the captions. In this case the transition to stability occurs when r is between 3.0 and 2.5.

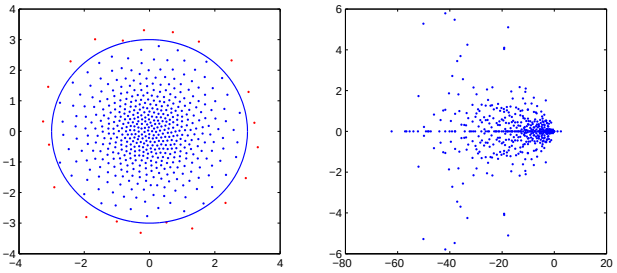


Figure 3.5.9: Points and eigenvalues of A for low distortion normal grid with 500 points and using $r = 3.0$ and 6 neighbours - unstable, $\max(\Re(\lambda)) = 2.51$.

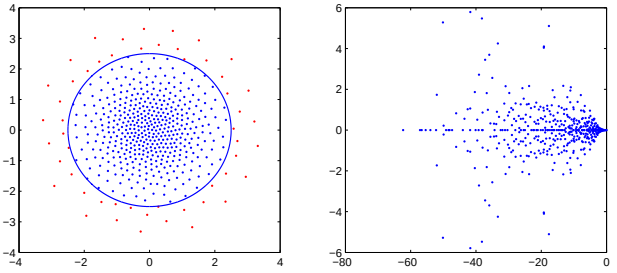


Figure 3.5.10: Points and eigenvalues of A for low distortion normal grid with 500 points and using $r = 2.5$ and 6 neighbours - stable, $\max(\Re(\lambda)) = 0$.

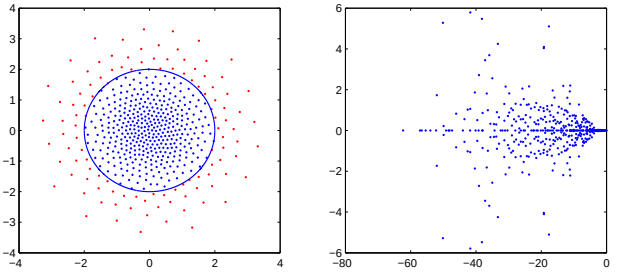


Figure 3.5.11: Points and eigenvalues of A for low distortion normal grid with 500 points and using $r = 2.0$ and 6 neighbours - stable, $\max(\Re(\lambda)) = 0$.

3.6 Conclusions

We presented a method for approximating a differential operator on an irregular grid. The method uses local polynomial interpolants to construct derivative approximations. We analysed the stability of the operator approximation using different grid and boundary configurations.

In one dimension, we showed that the method is equivalent to the standard finite difference method.

Our main finding in two dimensions was that grids with a regular local structure are more likely to lead to stable approximations. Thus square, triangular and hexagonal grids lead to stable approximations, but pseudo-random and quasi-random grids did not. Low distortion grids as used in Bally and Pagès [3] were also found to lead to stable approximations. We were able to induce stability in the case of a low discrepancy grid by altering neighbour configurations so that the neighbours were more uniformly distributed in an angular sense. For a low distortion grid adapted to the normal distribution, we found that the approximations constructed were unstable when the boundary radius was too large, but stable for smaller radii.

Summarised, this study indicates that instabilities in the approximation are a consequence of the local roughness of points in the grid, and of boundary effects.

The hurdle in extending this work to higher dimensions is stability, in particular more research is required either in the direction of determining sufficient conditions for stability on arbitrary grids or towards modifications of the approximation method. Such conditions are provided in Chapter 6; however the satisfaction of these conditions has not yet been investigated for the method presented in this paper.

The literature on mesh-free methods provides one solution in the form of the moving least squares method, where one attempts to integrate a particular interpolant, as opposed to working with derivatives of the interpolants. The moving least squares method seems to be less susceptible to instabilities than the present method; this is investigated further in Chapter 5.

Chapter 4

A Method Using Local Consistency Conditions

4.1 Introduction

The pricing of American options is a problem that has remained inaccessible to closed form solution. It was also long assumed to be inaccessible to Monte Carlo techniques, but Tilley quashed this belief in his 1993 paper [70]. Simulation techniques are of particular importance for higher dimensional problems where conventional discretisation methods become intractable.

Methods for solving American and Bermudan option pricing problems have become increasingly important with the widespread use of options and the development of more and more complex contracts. Examples of potentially high-dimensional options include basket options, swaptions and real options. We consider "high-dimensional" problems to be those where the number of stochastic factors is at least three or four, and thus conventional grid techniques become unmanageable.

Much progress has been seen in the past decade in the area of Monte Carlo techniques, through the work of Barraquand and Martineau [4], Broadie and Glasserman [19] and more recently Longstaff and Schwartz [50], Tsitsiklis and Van Roy [71], Rogers [64], Haugh and Kogan [37], Boyle et al. [14], and through the method proposed in Chapter 2 (also published as Berridge and Schumacher [8, 7, 9]).

Most techniques proposed have centred around path generations of the process.

This has the advantage that the points sampled are well adapted to the process, but the disadvantage that it is difficult to determine the expected value of continuation at each point. It is important to know the latter in order to make a stopping decision, and thus determine the early exercise premium.

The last method in the above list is the only one to consider a constant sampling of the state space over time. Since the method centres around an approximating Markov chain, it is simple to estimate continuation values on the grid using an appropriate Markov transition matrix. This method is thus more like a finite difference method, as opposed to the methods in [19, 50, 71, 14] which are more tree-like.

An important advantage of the irregular grid method proposed here is that the number of tuning parameters is small. Furthermore, convergence requires increasing only the number of grid points and the number of time steps, as with finite difference methods. In particular the method does not involve approximation of the value function or exercise region by basis functions.

We also note that using a constant grid allows implicit solutions to be easily obtained; for finite difference techniques this represents an increase in convergence speed from δt to δt^2 when considering European problems.

We proceed along the lines of Chapters 2 and 3 in that we approximate the value function on an irregular grid. We use a stable and more tractable method however for approximating the transition probabilities; instead of taking a root of a transition matrix, we directly construct the transition probabilities using local consistency conditions presented in Kushner and Dupuis [46] in the parabolic case and similar conditions to construct the infinitesimal generator in the elliptic case. This allows us to use much larger grids, and thus obtain more accurate solutions.

Using the root method in Chapter 2 the grid size was limited to 3000 on a desktop computer, and averaging was needed over several experiments to obtain accurate solutions. We can now deal with grid sizes in the hundreds of thousands, and solutions from a single experiment are of sufficient accuracy that randomisation is no longer required.

This chapter continues in Section 4.2 with a formulation of the problem of interest. Section 4.3 presents the proposed methodology, refinements are presented in Section 4.4 and experiments are carried out in Section 4.5. Section 4.6 concludes.

4.2 Formulation

4.2.1 The market

As in Chapter 2, 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

$$dX(s) = \mu(X(s), s)ds + \sigma(X(s), s)dW(s) \tag{4.2.1}$$

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

$$dV(s) = \mu_V(X(s), s)ds + \sigma_V(X(s), s)dW(s) \tag{4.2.2}$$

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 $V(\cdot, T) = \psi(\cdot, T)$, and intermediate values satisfy $V(\cdot, s) \geq \psi(\cdot, s)$, $s \in [t, T]$. It is assumed that $\mu(x, s)$ and $\sigma(x, s)$ satisfy suitable regularity conditions, such that the change of measure implied by (4.2.2) is well-defined.

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

$$dX(s) = \mu_{RN}(X(s), s)ds + \sigma(X(s), s)dW(s) \tag{4.2.3}$$

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

Our 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:

$$\tau : \mathbb{R}^d \times [t, T] \rightarrow \{0, 1\} \tag{4.2.4}$$

$$H : \mathbb{R}^d \times [t, T] \rightarrow \mathbb{R}^d. \tag{4.2.5}$$

Supposing that one has an estimate $\hat{V}(t)$ of the derivative price, it is often important to specify 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.

4.2.2 Pricing

The primal formulation

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

$$v(x_t, t) = \sup_{\tau \in \mathcal{T}} \mathbb{E}^{\mathbb{Q}}_{x_t} \left(e^{-r(\tau - t)} \psi(X(\tau), \tau) \right) \tag{4.2.6}$$

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 $X(t) = x_t$.

The dual formulation

The dual formulation (see Rogers [64] or Haugh and Kogan [37]) forms a price by minimising the cost of the hedging strategy over martingales. Theorem 1 of [64] implies that the price is given by

$$v(x_t, t) = \inf_{M \in H_0^1} \mathbb{E}^{\mathbb{Q}}_{x_t} \left[\sup_{s \in [t,T]} \left(e^{-r(s-t)} \psi(X(s), s) - M(s) \right) \right] \tag{4.2.7}$$

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 at a certain martingale $M = M^*$.

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. [35]. Jaillet et al. [42] 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

$$\mathcal{L} = \tfrac{1}{2} \mathrm{tr} \sigma \sigma' \frac{\partial^2}{\partial x^2} + \mu_{RN} \frac{\partial}{\partial x} - r \tag{4.2.8}$$

where r is the risk-free rate.

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

$$a(v, u) = \langle u, \mathcal{L} v \rangle . \tag{4.2.9}$$

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

$$\begin{cases} v(x,s) - \psi(x,s) & \geq & 0 \\ u \geq \psi \text{ a.e. } \Rightarrow a(v, u-v) - \left\langle u - v, \frac{\partial v}{\partial t} \right\rangle & \geq & 0 \qquad \text{a.e. } [t,T] \end{cases} \tag{4.2.10}$$

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

The complementarity formulation

The variational inequality formulation is not directly amenable to computation. For this reason it is convenient to reformulate it as a complementarity problem. Let $\mathcal{L}$ be the related diffusion operator; then the option value is found by solving the complementarity problem

$$\begin{cases} \frac{\partial v}{\partial t} + \mathcal{L}v & \leq & 0 \\ v - \psi & \geq & 0 \\ \left(\frac{\partial v}{\partial t} + \mathcal{L}v\right)(v - \psi) & = & 0 \end{cases} \tag{4.2.11}$$

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

Such a problem can be solved using standard PDE discretisation techniques, with some modifications to account for the inequalities.

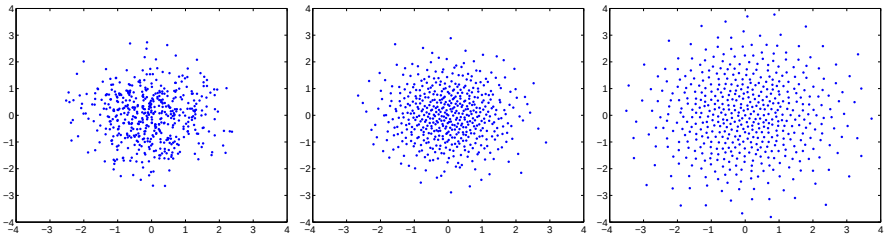
4.2.3 Consequences

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 (4.2.11) is active and the stopping rule says not to exercise. In the stopping region the second line of (4.2.11) 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.

4.3 Methodology

The basic methodology presented is similar to that of Chapter 2, with the exception of the manner in which the transition matrix is constructed. This is now done



(a) Random (b) Sobol' (c) Low Distortion

Figure 4.3.1: Grids with 500 points adapted to the normal density.

using the local consistency conditions presented in Kushner and Dupuis [46], and a modification of these conditions is used to find an approximation to the infinitesimal generator. These conditions ensure that the approximating Markov chain has a local mean and variance that match those of the continuous process.

4.3.1 Irregular grid

We first briefly review the irregular grid methodology presented in Chapter 2. We define an irregular grid to be a representative sampling of the state space

$$\mathcal{X} = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d. \tag{4.3.1}$$

The method of sampling is to be specified at a later stage, but one can think of it as a low discrepancy or low distortion set (see for example Bally and Pagès [60]) which is dense in the entire state space as $n \to \infty$.

Examples of possible grids in two dimensions are presented in Figure 4.3.1. As in the case of Monte Carlo integration, it is expected that low discrepancy (e.g. Sobol') and low distortion grids will lead to faster convergence than random grids. For results regarding integration see Evans and Swartz [28] and Pagès [60].

In order to simplify the analysis we now make the assumption that the risk-neutral process is a d-dimensional time homogeneous diffusion process

$$dX(s) = \mu_{RN}(X(s))ds + R(X(s))dW(s) \tag{4.3.2}$$

where $R'R$ is the Cholesky decomposition of the state-dependent covariance matrix $\Sigma(X(s))$ and X and W are of the same length d. This assumption is not necessary for the method to work; it merely simplifies some aspects and allows for a clearer exposition.

4.3.2 Approximation of Markov chain

We consider approximating the risk-neutral process (4.3.2) using a discrete state, discrete time Markov chain where the states are exactly the points in our irregular grid $\mathcal{X}$ and the time step is δt.

The Markov transition matrix P is constructed in such a way as to satisfy the local consistency conditions given in Kushner and Dupuis [46]. We require[1] for each state $i = 1, \ldots, n$

$$\begin{aligned} \Sigma(x_i)\delta t &= \textstyle\sum_{j=1}^{n}(x_j - x_i - \mu_{RN}(x_i)\delta t)(x_j - x_i - \mu_{RN}(x_i)\delta t)' p_{i,j} \\ \mu_{RN}(x_i)\delta t &= \textstyle\sum_{j=1}^{n}(x_j - x_i) p_{i,j} \\ 1 &= \textstyle\sum_{j=1}^{n} p_{i,j} \\ p_{i,j} &\geq 0 \end{aligned} \tag{4.3.3}$$

where $p_{i,j}$ is the (i,j)th entry of P.

One must solve for each state i a feasibility problem over the $p_{i,j}$. The number of equality constraints in the problem is given by $\eta_d + 1$ where

$$\eta_d = \frac{1}{2}d(d+3) \tag{4.3.4}$$

and the number of variables is n. In the problems we consider, η_d is much smaller than n.

In practise one can impose the extra condition that the transitions should only be allowed to close neighbours of each point. Computationally this means that we only need to consider a small number of transitions k where $\eta_d + 1 < k \ll n$, thus dramatically reducing the complexity of the problem.

It is also useful to specify a linear objective function to optimise the proximity of transitions. That is, to satisfy the local consistency conditions using points as close as possible to the mean. The linear objective function, to be minimised, should have a coefficient relating to point x_j which is an increasing function of the distance $||x_i - x_j||$. Let us denote the objective function by $f_i \cdot p_i$ where p_i is the ith row of P.

[1] The formulation in [46] is more general in that it allows $o(\delta t)$ terms to be added on the RHS of the first two conditions.

We thus pose for each point x_i a linear program $\min f_i \cdot p_i$ subject to (4.3.3). In experiments we found that a convenient specification for f is $f_j = k^3$ where x_j is the kth nearest neighbour of $x_i + \mu_{RN}\delta t$.

We note that the solution to the linear program will in general be a corner solution using as many zero variables as possible; the number of nonzero transition probabilities per point is the minimum number, $\eta_d + 1$. This is a consequence of Corollary 7.11 in Schrijver [66], and of the fact that the constraint matrix has $\eta_d + 1$ rows. Note that the points with positive weights are not necessarily the $\eta_d + 1$ nearest neighbours of $x_i + \mu_{RN}\delta t$, since these may not satisfy the feasibility conditions; the points form rather the closest possible feasible set (with respect to the objective function).

4.3.3 Approximation of infinitesimal generator

Rather than approximating transition probabilities, one may attempt to approximate the infinitesimal generator directly. This amounts to constructing a discrete space, continuous time approximation to the problem.

Constructing an approximation to the infinitesimal generator allows quick reconstruction of transition probabilities for arbitrary time steps δt, or for scaling the effect of the diffusion operator, through a first order approximation. Consequently this method is preferred over that of Section 4.3.2, provided we do not have a large state-dependent drift. We assume the latter in this section.

In the case of a non-state dependent drift, we refer the reader to Section 4.4.5 where we introduce a simple transformation of the continuous process to eliminate a risk-neutral drift that depends deterministically on time.

We start with the problem (4.3.3), and define

$$a_{i,j} = \frac{1}{\delta t}(p_{i,j}(\delta t) - \delta_{ij}) \tag{4.3.5}$$

where δ_{ij} is the Kronecker delta. As $\delta t \to 0$ in (4.3.5) we obtain elements of the infinitesimal generator matrix A.

Substituting (4.3.5) into (4.3.3) and letting $\delta t \to 0$ yields the new feasibility problem

$$\begin{aligned} \Sigma(x_i) &= \textstyle\sum_{j\neq i}(x_j - x_i)(x_j - x_i)' a_{i,j} \\ \mu_{RN}(x_i) &= \textstyle\sum_{j\neq i}(x_j - x_i) a_{i,j} \\ a_{i,j} &\geq 0 \end{aligned} \tag{4.3.6}$$

and $a_{i,i} = -\sum_{j \neq i}^{n} a_{i,j}$. Note that (4.3.6) now contains only η_d equality constraints, one less than (4.3.3).

The same considerations as in Section 4.3.2 are also applied in this case. We solve for each point x_i a linear program $\min f \cdot a_i$ subject to (4.3.6) and $a_{i,j} \geq 0$ where a_i is the ith row of A with the diagonal entry omitted. Following from the observation at the end of Section 4.3.2, we again expect a maximum of $\eta_d + 1$ nonzero entries per row of A.

We note that when a large drift term is present, one may be able to satisfy the local consistency conditions (4.3.6), but this may require using points x_j which are nonlocal to x_i. This method differs from the usual method of lines in that here we produce a stable system before checking for localness of the neighbours, whereas in the usual method one selects the neighbours a priori before building the equations and finally considering stability (see for example Hundsdorfer and Verwer [41]).

4.3.4 Time stepping

Given a transition matrix P, corresponding to time step δt, the option pricing problem can be solved using dynamic programming on the discretised Markov chain. Namely, one solves the problem

$$\begin{aligned} v(T) &= \psi \\ v(t_k) &= \max\left(\psi, e^{-r\delta t} P v(t_{k+1})\right) \end{aligned}$$

for $t_k = k\delta t$ and $k = K-1, \ldots, 0$ where K is the number of time steps considered, v is a vector of values at grid points and ψ is a vector of payoffs at grid points, here assumed to be constant over time. The resulting solution $v(x, 0)$ is an exact solution to the approximating Markov chain.

Given the infinitesimal generator A, one can form a first order approximation to the transition matrix $P \simeq I + A\delta t$ and proceed as above. Alternatively, it is possible to solve the problem to a higher order using the matrix exponential

$$\begin{aligned} v(T) &= \psi \\ v(t_k) &= \max\left(\psi, e^{-r\delta t} e^{A\delta t} v(t_{k+1})\right) \end{aligned}$$

for $k = K - 1, \ldots, 0$. Since A is sparse, the effect of the matrix exponential can

be calculated efficiently using Krylov subspace methods, see for example Druskin and Knizhnerman [27] or Hochbruck and Lubich [38].

The above time stepping methods are suitable for Bermudan pricing problems with δt being the period between exercise possibilities. We expect convergence to the Bermudan solution as $n \to \infty$, and convergence to the American solution as $\delta t \to 0$.

When considering a truly American problem, it is useful to consider Crank-Nicolson and implicit solutions. In particular the Crank-Nicolson method is known to converge at a rate δt^2 for the European problem as opposed to δt for the explicit and implicit methods, and implicit methods are known to be unconditionally stable for solving sequences of LCPs (see Chapter 6 and Glowinski et al. [35]).

The Crank-Nicolson method corresponding to the truly American problem is the following system with $\theta = \frac{1}{2}$

$$\begin{aligned} v(T) &= \psi \\ 0 \leq (v(t_k) - \psi) &\perp \left(e^{-\theta A \delta t} v(t_k) - e^{-r\delta t} e^{(1-\theta)A\delta t} v(t_{k+1}) \right) \geq 0 \end{aligned} \tag{4.3.7}$$

for $k = K-1, \ldots, 0$. The second line is a linear complementarity problem (LCP). There are many methods available for solving LCPs, including the projected successive overrelaxation (PSOR) method proposed in Cryer [25]. Another possible candidate is linear programming, which is used for example by Dempster and Hutton [26] to solve the one-dimensional American option pricing problem.

4.3.5 Summary of the algorithm

We present a concise statement of the proposed algorithm as Algorithm 4.3.1. The generation of the matrices A can be done in advance for a given grid $\mathcal{X}$, with obvious changes to the algorithm.

4.4 Fine tuning and extensions

We now mention some implementation issues and refinements of the method. These issues are not essential to the method, but may improve performance and allow quicker execution for a given required accuracy.

Algorithm 4.3.1 Proposed algorithm for solving high-dimensional American option pricing problems.

Choose the grid size n
Generate a QMC grid $\mathcal{X}$
Compute the generator matrix A
Choose the time step $\delta t > 0$ and implicitness $\theta \in [0, 1]$
Solve the linear complementarity problems (4.3.7)

4.4.1 Grid specification

In the presentation so far, we have taken the grid $\mathcal{X}$ to be given; we now consider ways one might specify the grid.

Taking inspiration from the literature on MC and QMC integration, we first suggest that the grid be constructed using low discrepancy (Niederreiter [58]) or low distortion (Pagès [60]) methods. Just as in the regular grid case, we expect the error to be related to the separation of grid points, more specifically the separation of grid points having positive weights in the generator matrix.

Importance sampling considerations tell us that the most efficient grid density is given by the density of the process itself. Given our suggestion of a constant grid (for efficiency reasons), 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 Evans and Swartz [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 minimisation of boundary effects on the solution. This suggests that the grid covariance should be larger than the covariance of the process.

In Chapter 2, where a root method was used to construct transition probabilities, and the process considered was a five-dimensional Brownian motion with drift, a grid covariance of 1.5 times the process covariance at expiry was found to

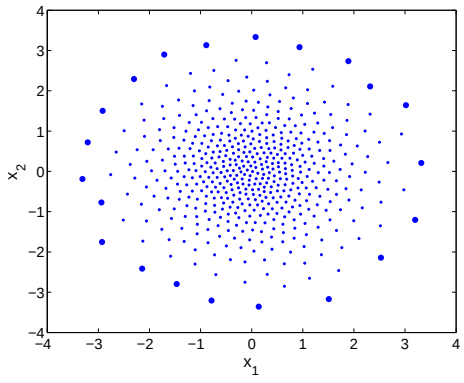


Figure 4.4.1: Interior points (small) and boundary points (large) on a normal low distortion grid for $d = 2$, $n = 500$.

give the best convergence rate when tested against grids with covariances of 1.0 and 2.0 times the covariance at expiry.

4.4.2 Boundary region and boundary conditions

It is clear that (4.3.3) and (4.3.6) may be infeasible for some i. In such a case we say that x_i is an implied boundary point, otherwise it is an implied interior point. Given nondegenerate Σ and a well-adapted grid, one expects that the implied boundary points will indeed lie at the extremities of the grid, and the implied interior points away from the extremities.

One may specify appropriate boundary conditions in this region to reflect the behaviour of the process. In the experiments we let these points be absorbing, which is appropriate for value functions having a linear behaviour at the boundary. One may also apply Dirichlet, Neumann or mixed conditions using neighbours in the grid.

It would be useful to know a priori which points are likely to be in the implied boundary, since we would like to avoid trying to solve infeasible linear programming problems. In practise however it is difficult to do this even for simple cases.

A plot of the boundary behaviour for a 500-point low distortion grid in two dimensions is given in Figure 4.4.1. Notice in this case that the number of infeasible points is 21, this being about 4% of the total.

If one assumes a distribution for the neighbours over which (4.3.3) or (4.3.6) is

to be solved, then one can quantify the probability of feasibility. Near the boundary of the grid, there may be a low density of points on the boundary side, and thus the probability of feasibility changes.

For example, if our grid consists of n independent standard normal draws, we can calculate the expected number of grid points in a halfspace away from the centre of the grid at some radius r. One can then say what the minimum number of points n is where the expected number of grid points in the halfspace away from the grid centre at radius r is less than some bound.

Let us set this bound to be $\frac{1}{2}\eta_d$, where η_d is given in (4.3.4), a very optimistic bound but useful to illustrate the approximate behaviour of the boundary. Requiring an expected number of $\frac{1}{2}\eta_d$ points in the halfspace away from the center implies a boundary radius of

$$r = \Phi^{-1}\left(1 - \frac{1}{2n}\eta_d\right) \tag{4.4.1}$$

where Φ is the cumulative normal distribution function. In order to find the expected number of boundary points we then note that the squared norm of a standard normal variable in d dimensions is a chi square random variable with d degrees of freedom. Thus, if the boundary region is defined by $\left\{x : \|x\|^2 \geq r^2\right\}$, then the expected numbers of interior and boundary points are

$$\mathbb{E}N_i = n\Psi(r^2, d) \tag{4.4.2}$$

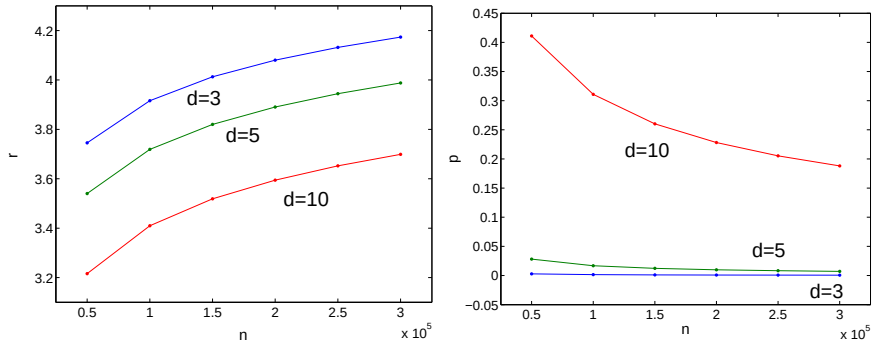
$$\mathbb{E}N_b = n\left(1 - \Psi(r^2, d)\right), \tag{4.4.3}$$

respectively, where Ψ is the chi square cumulative distribution function.

Plots of the radius and expected number of boundary points are presented in Figure 4.4.2 for $d = 3, 5, 10$ and n up to $300,000$.

Experimentally we find that (4.4.1) underestimates the implied radius for lower dimensions and overestimates it for higher dimensions (see Section 4.5 for numerical results). The latter is not surprising since one generally requires more than the minimum number of points η_d to satisfy the feasibility conditions (4.3.3) and (4.3.6).

Finally we mention that in estimating the boundary, we prefer an underestimate to an overestimate. An overestimate of the boundary may lead us to waste a considerable amount of computing time trying to solve infeasible linear programs. An underestimate on the other hand just results in the grid having extra boundary



(a) Radius (b) Proportion

Figure 4.4.2: Naive prediction for the radius of the boundary and the proportion of points which are in the boundary region for a standard normal grid.

points. The latter does not add a significant amount of overhead to the method, the effect being limited to a slight increase in complexity of the nearest neighbour problem and extra zero rows to the sparse generator matrix.

4.4.3 Parallelism

In the language of computer science, problems (4.3.3) and (4.3.6) are said to be embarrassingly parallel. This refers to the fact that a speedup linear in the number of processors can be trivially achieved. For example, having a large number n of linear programs to solve and m computers, we can reduce the time by a factor $1/m$ by solving n/m linear programs per computer, assuming $n > m$ and that the communication time between the computers is negligible. We make use of this point when conducting the experiments, using a distributed computing environment to solve the linear programs.

4.4.4 Control variates and Richardson extrapolation

To obtain more accurate solutions we consider variance reduction and extrapolation techniques.

Variance reduction is already used in the method in that the grids are constructed using points designed to cover the state space evenly according to the process density at expiry. In the current context the idea of control variates is also

very easy to apply since the European solution is usually highly correlated with the Bermudan and American solutions. Since the European price is easy to determine to a high degree of accuracy, it constitutes an ideal control variate.

The concept of extrapolation is also useful once we have an idea of how the error behaves with increasing n. In Section 4.5.5 below, experimental evidence is given which implies the estimates behave asymptotically as

$$\hat{v}_n = v + c_1 n^{c_2/d} \tag{4.4.4}$$

for some constants c_1, c_2, which may be estimated. Here we assume that the error is always of the same sign, which may be indicated for example by a monotone behaviour of the approximations.

4.4.5 Matrix reuse

Given that generating the transition and infinitesimal generator matrices is an expensive operation compared to the final time stepping procedure, it is of interest to know under which conditions these matrices can be reused for related problems. It is clear that a single matrix can be reused for as many different payoff functions as required; it can also be reused for processes with different risk-neutral drifts and covariances as follows.

Suppose that a transition or infinitesimal generator matrix has been constructed for a process with covariance matrix I and zero risk-neutral drift on the grid $\mathcal{X}$. Let us construct the grid $\mathcal{Y}$ where $y_i = R'x_i$, R being a Cholesky factor of the covariance matrix Σ. The implied covariance of the transition or infinitesimal generator matrix on $\mathcal{Y}$ is now Σ.

Suppose now that our process has covariance Σ, and constant (nonzero) risk-neutral drift μ. Consider now the time dependent grid $\mathcal{Y}_k$ where the subscript k corresponds to time $k\delta t$ and $y_k = x + k\mu\delta t$. The implied covariance of the transition or infinitesimal generator matrix remains Σ, but the implied drift is now μ.

Two simple extensions to the time homogeneous problem are those in which the risk-neutral drift is deterministically time-dependent and the covariance matrix is scaled over time,

$$dX(s) = \mu_{RN}(s)ds + \alpha(s)RdW(s). \tag{4.4.5}$$

The most convenient way to deal with the drift term is to incorporate the drift in the payoff function. This amounts to the change of variables

$$X_0(s) = X(s) - \int_0^s \mu_{RN}(u)du, \tag{4.4.6}$$

the new process having zero drift

$$dX_0(s) = \alpha(s)RdW(s) \tag{4.4.7}$$

and the payoff being

$$\psi_0(x_i, s) = \psi\left(x_i + \int_0^s \mu_{RN}(u)du, s\right). \tag{4.4.8}$$

The scaled covariance term can be accommodated by manipulating the time step. By using time step $\alpha(s)\delta t$ at time s in place of δt, we achieve a covariance of $\alpha(s)^2\Sigma$ as required.

4.4.6 Grid expansion

Grid expansion relates the size of the grid to the variance of the process. A convenient way to generate a grid is to sample the process at expiry; one thus obtains a grid $\mathcal{X}$ that becomes dense in the state space as $n \to \infty$. For a finite n however one can ask how well the process can be represented on $\mathcal{X}$. For example if we consider a standard d-dimensional Brownian motion on $s \in [0, 1]$, the process density at expiry is $\mathcal{N}(0, I)$. If the implied boundary begins at $r < 2$ for example, there is a nonnegligible chance of the discrete Markov process hitting the absorbing boundary before expiry, thus reducing the accuracy of the solution.

In this case we can set a lower limit r_0 for the implied boundary, for example $r_0 = 4$ for which the process has a negligible chance of hitting the boundary. This limit can be achieved by expanding the grid; to do this, one scales the grid points by a factor r_0/r and the generator matrix entries by a factor r/r_0, thus removing the boundary effects while preserving local consistency.

The grid expansion factor allows us to make a tradeoff between errors caused by the boundary and errors related to the discretisation. The higher the factor applied in the grid expansion, the lower the effect from the boundaries but the coarser the grid becomes and hence the higher the discretisation error.

4.4.7 Partially absorbing boundaries

Infeasibility of points in the boundary region is usually caused by a lack of points in the halfspace away from the center of the grid. If the grid boundary looks locally linear, as in a spherical grid, it is possible that the infeasibility is only in this direction, and not "along" the boundary.

In this case it may be useful to consider partially absorbing boundaries in which one only tries to satisfy local consistency conditions in the direction tangent to the boundary. In the case of a normal grid this amounts to requiring a zero variance along lines through the grid center for points in the boundary layer. This type of boundary condition has not been employed in the current study.

4.5 Experiments

A major hurdle in testing algorithms for pricing high-dimensional American options is the difficulty of verifying results. One common method is using out-of-sample paths to estimate the value of the exercise and hedging strategies implied by the model. Another, which we use here, is to use benchmark results from a special case that can be solved accurately. In the following we introduce benchmark results and then test the proposed method against those results.

4.5.1 Geometric average options

We choose to focus on geometric average options, since the pricing problem for these options can be reduced to a one-dimensional problem. The one-dimensional problems can be solved to a high degree of accuracy, thus providing benchmark results for the algorithm.

A geometric average put option written on d assets following the risk-neutral process (4.2.3) has payoff function

$$\psi(s) = \left(K - \left(\prod s_i \right)^{1/d} \right)^+ \tag{4.5.1}$$

where s is the asset value and K is the strike price of the option. Assuming a complete and arbitrage free market with the log asset prices following a multivariate Brownian motion with constant covariance Σ, we have a constant risk-neutral drift

$$\mu_{RN} = r\mathbb{1} - \frac{1}{2} \operatorname{diag} \Sigma. \tag{4.5.2}$$

4.5.2 Benchmarks

Using Itô's lemma with $Y = f(X) = \overline{X}$, we find that Y follows the risk-neutral process

$$dY(s) = \frac{1}{d}\sum_{i=1}^{d} dX_i(s) \tag{4.5.3}$$

$$= \tilde{\mu}ds + \tilde{\sigma}dW(s), \tag{4.5.4}$$

the parameters of the diffusion being given by

$$\tilde{\mu} = r - \frac{1}{2d}\sum_{i=1}^{d}\sigma_i^2 \tag{4.5.5}$$

$$\tilde{\sigma}^2 = \frac{1}{d^2}\sum_{i=1}^{d}\left(\sum_{j=1}^{d} R_{ij}\right)^2. \tag{4.5.6}$$

The option is thus equivalent to a standard put option on an asset with starting value $\exp\{\overline{X}_0\}$, strike price K, risk-free rate r and continuous dividend stream

$$\delta = \frac{1}{2}\left(\frac{1}{d}\sum_{i=1}^{d}\sigma_i^2 - \tilde{\sigma}^2\right). \tag{4.5.7}$$

In Table 4.5.1 we provide benchmark results for geometric put options written on up to ten assets, with starting asset values $S_i = 40$ for all i and strike price 40. The risk-free rate is taken as 0.06, the volatilities $\sigma_i = 0.2$ for all i, and correlations $\rho_{ij} = 0.25$, $i \neq j$.

4.5.3 Experimental details

Using the methodology proposed in Section 4.3, we conducted experiments to find the value of the geometric average put options given above.

We used six different grid sizes ranging from $50,000$ to $300,000$, and two types of grids consisting of normal Sobol' points and normal low distortion points with a covariance corresponding to 1.5 times the process covariance at expiry. The transition matrices were generated using distributed computing software in a Matlab environment. A maximum of $20\eta_d$ nearest neighbours were considered when trying to satisfy the local consistency conditions, where η_d is defined in (4.3.4).

d	$\tilde{\sigma}^2 \times 10^2$	$\delta \times 10^2$	European	Bermudan	American
1	4.000	0.000	2.0664	2.2930	2.3196
2	2.500	0.750	1.5553	1.7557	1.7787
3	2.000	1.000	1.3468	1.5380	1.5597
4	1.750	1.125	1.2318	1.4193	1.4392
5	1.600	1.200	1.1585	1.3421	1.3625
6	1.500	1.250	1.1077	1.2893	1.3094
7	1.429	1.286	1.0703	1.2504	1.2703
8	1.375	1.313	1.0416	1.2207	1.2404
9	1.333	1.333	1.0189	1.1971	1.2167
10	1.300	1.350	1.0004	1.1779	1.1974

Table 4.5.1: Benchmark results for geometric average options in dimensions 1–10. Also displayed are the variance $\tilde{\sigma}^2$ and continuous dividend δ for the equivalent one dimensional problem.

We consider the pricing problem for European options, Bermudan with ten exercise opportunities and true American where the option can be exercised at any time up to expiry. For the European and Bermudan problems we used the Crank-Nicolson method with 100 time steps. For solving the linear systems we used the conjugate gradients squared (CGS) and generalised minimum residual (GMRES) methods, the latter being slower but more robust. For the American problems we used projected successive overrelaxation (PSOR) to solve the linear complementarity problems, with 1000 time steps. While it is not necessary to use such a large number of time steps in practise, we wanted to focus on the error with respect to the space discretisation. Having a small enough δt causes the error resulting from time discretisation to be negligible in comparison, and thus allows a more accurate assessment of the error resulting from space discretisation.

4.5.4 Experimental results

We present results in Tables 4.5.2–4.5.4 for prices obtained using normal Sobol' grids for the Bermudan, American and European cases, respectively. The results for low distortion grids are presented in Tables 4.5.7–4.5.9.

Tables 4.5.5 and 4.5.6 show the results on normal Sobol' grids for Bermudan and American options when the European is used as control variate. Tables 4.5.10 and 4.5.11 show the same for low distortion grids.

Figures 4.5.1 and 4.5.2 present the results graphically for normal Sobol' grids. The results for low distortion grids are shown in Figures 4.5.3 and 4.5.4. We see that the error increases with dimension to about 5–10% for $d = 10$. The control variate improves the results dramatically, the error for $d = 10$ being now less than 1%.

When using the European control variate we see that the results are biased upwards, whereas the raw results are biased downwards. This is probably due to the upward bias introduced by the convexity of the max operator which appears in the Bermudan and American problems, but not in the European problem.

In one and two dimensions the generator matrix became numerically unstable for the grid sizes we consider; we have thus not presented results for these low dimensions here. This lack of convergence is due to the finite precision arithmetic, and not to instability in the sense that the generator matrix has unstable eigenvalues (i.e. eigenvalues having positive real part). The method has been found to work very well in one and two dimensions, but for smaller grid sizes.

4.5.5 Error behaviour

Drawing a parallel with regular grid methods, we expect the error to be related to δx, the distance between grid points with positive weights in A. In a regular grid with the same number of points N in each dimension we have $n = N^d$ points in total, and the distance to the nearest point is simply $n^{-1/d}$. The error when using a standard finite difference method is of order δx^2, or $n^{-2/d}$.

We thus propose modelling irregular grid errors as in the regular grid case, allowing for a scalar factor in the exponent as well as a multiplicative factor:

$$\log|\varepsilon| = c_1 + c_2 \frac{\log n}{d}. \tag{4.5.8}$$

In Figures 4.5.5 and 4.5.6 we present plots of the log absolute error versus $\log(n)/d$, and in Tables 4.5.12 and 4.5.13 the regression results. Referring to our assumption of error behaviour (4.5.8) we find that the complexity is accurately modelled by the given relationship in all three cases (for suitable c_1, c_2). The linear relationships

observed, on the log scale, strongly suggest that the algorithm has exponential complexity. We note that the behaviour in the Sobol' and low distortion cases is very similar, with the European and Bermudan prices showing about the same asymptotic relationship, and with American errors showing a slightly faster rate in the Sobol' case, although this is barely significant.

The convergence rate for finite difference methods used to solve PDE problems on regular grids is $1/\delta x^2$, or $n^{-2/d}$ which here translates to $c_2 = -2$. From this point of view our method seems to be slightly slower in convergence than the regular grid method, although this is barely significant. This may be due to the average δx being larger as a function of the grid size in the irregular grid case.

The given model for errors implies that the amount of work required to obtain solutions to a certain accuracy increases exponentially with dimension. This may seem pessimistic in that the curse of dimensionality is not broken; however the method we use has definite advantages over regular grid methodology in high dimensions. In particular we note that the number of grid points n can be chosen freely, the grid points can be adapted to the process density and the number of boundary points can be substantially reduced for unbounded problems. Regarding the last point, Section 4.5.7 provides a comparison between the number of boundary points found in regular and normally distributed grids. The results suggest that the proposed method can handle option pricing problems up to dimension ten, which sets it aside from traditional finite difference methods which start to become unwieldy in dimension three or four.

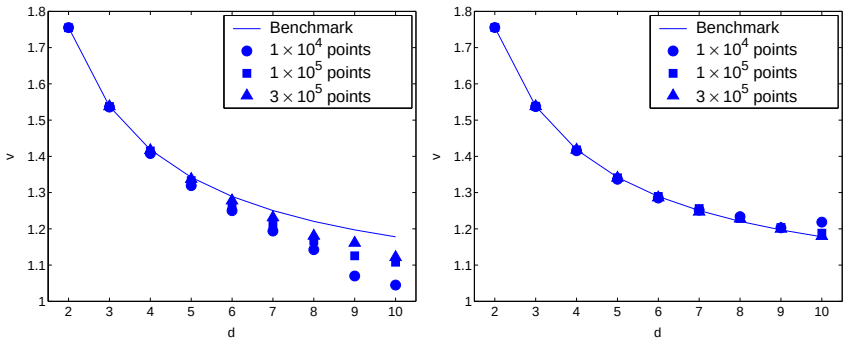


Figure 4.5.1: Bermudan pricing results for normal Sobol' grids presented raw (left) and using European price as control variate (right).

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5370	1.5375	1.5376	1.5376	1.5377	1.5377
4	1.4135	1.4147	1.4155	1.4161	1.4163	1.4166
5	1.3300	1.3329	1.3345	1.3360	1.3365	1.3371
6	1.2532	1.2630	1.2667	1.2757	1.2766	1.2780
7	1.1981	1.2133	1.2137	1.2305	1.2311	1.2313
8	1.1489	1.1664	1.1672	1.1891	1.1938	1.1807
9	1.1116	1.1255	1.1351	1.1530	1.1514	1.1612
10	1.0901	1.1080	1.1078	1.1129	1.1242	1.1218

Table 4.5.2: Results for Bermudan geometric average put options in dimensions 3-10 using normal Sobol' grids.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5584	1.5588	1.5590	1.5591	1.5592	1.5592
4	1.4332	1.4347	1.4357	1.4362	1.4365	1.4369
5	1.3489	1.3522	1.3537	1.3551	1.3557	1.3563
6	1.2721	1.2818	1.2858	1.2940	1.2951	1.2965
7	1.2182	1.2325	1.2331	1.2482	1.2491	1.2492
8	1.1693	1.1864	1.1870	1.2071	1.2114	1.1993
9	1.1316	1.1460	1.1549	1.1715	1.1700	1.1802
10	1.1102	1.1281	1.1267	1.1324	1.1433	1.1414

Table 4.5.3: Results for American geometric average put options in dimensions 3-10 on normal Sobol' grids.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.3461	1.3463	1.3465	1.3465	1.3465	1.3465
4	1.2274	1.2286	1.2293	1.2302	1.2304	1.2304
5	1.1482	1.1505	1.1520	1.1541	1.1545	1.1549
6	1.0716	1.0813	1.0849	1.0977	1.0984	1.0993
7	1.0156	1.0275	1.0318	1.0527	1.0541	1.0545
8	0.9624	0.9792	0.9848	1.0123	1.0151	0.9943
9	0.9231	0.9406	0.9507	0.9735	0.9755	0.9802
10	0.8966	0.9203	0.9277	0.9340	0.9418	0.9424

Table 4.5.4: Results for European geometric average put options in dimensions 3-10 on normal Sobol' grids.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5382	1.5382	1.5382	1.5379	1.5379	1.5379
4	1.4179	1.4178	1.4180	1.4177	1.4177	1.4179
5	1.3403	1.3409	1.3410	1.3404	1.3405	1.3407
6	1.2892	1.2893	1.2894	1.2857	1.2858	1.2863
7	1.2527	1.2560	1.2521	1.2481	1.2473	1.2470
8	1.2281	1.2288	1.2240	1.2184	1.2203	1.2279
9	1.2074	1.2038	1.2033	1.1984	1.1947	1.1999
10	1.1940	1.1881	1.1805	1.1793	1.1829	1.1799

Table 4.5.5: Results for Bermudan geometric average put options in dimensions 3-10 on normal Sobol' grids, using the European price as a control variate.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5595	1.5596	1.5596	1.5594	1.5594	1.5594
4	1.4376	1.4378	1.4382	1.4378	1.4379	1.4382
5	1.3592	1.3602	1.3603	1.3595	1.3597	1.3599
6	1.3082	1.3082	1.3085	1.3041	1.3044	1.3048
7	1.2728	1.2752	1.2716	1.2658	1.2653	1.2649
8	1.2484	1.2487	1.2437	1.2364	1.2379	1.2465
9	1.2274	1.2242	1.2231	1.2169	1.2133	1.2189
10	1.2141	1.2082	1.1994	1.1988	1.2020	1.1994

Table 4.5.6: Results for American geometric average put options in dimensions 3-10 on normal Sobol' grids, using the European price as a control variate.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5372	1.5375	1.5376	1.5377	1.5377	1.5378
4	1.4141	1.4155	1.4160	1.4163	1.4165	1.4166
5	1.3309	1.3338	1.3360	1.3364	1.3370	1.3371
6	1.2695	1.2729	1.2751	1.2777	1.2779	1.2796
7	1.2139	1.2249	1.2255	1.2292	1.2319	1.2321
8	1.1628	1.1773	1.1850	1.1898	1.1899	1.1863
9	1.1234	1.1397	1.1428	1.1548	1.1514	1.1588
10	1.1177	1.1008	1.1131	1.1103	1.1170	1.1242

Table 4.5.7: Results for Bermudan geometric average put options in dimensions 3-10 using low distortion grids.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5583	1.5587	1.5589	1.5590	1.5590	1.5591
4	1.4341	1.4355	1.4361	1.4364	1.4367	1.4369
5	1.3500	1.3528	1.3550	1.3554	1.3561	1.3564
6	1.2875	1.2912	1.2935	1.2961	1.2965	1.2981
7	1.2319	1.2432	1.2433	1.2474	1.2496	1.2502
8	1.1813	1.1952	1.2032	1.2082	1.2080	1.2042
9	1.1412	1.1580	1.1615	1.1730	1.1689	1.1774
10	1.1390	1.1206	1.1315	1.1288	1.1365	1.1434

Table 4.5.8: Results for American geometric average put options in dimensions 3-10 on low distortion grids.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.3460	1.3463	1.3464	1.3465	1.3465	1.3466
4	1.2287	1.2295	1.2299	1.2301	1.2304	1.2305
5	1.1501	1.1520	1.1535	1.1540	1.1544	1.1546
6	1.0904	1.0947	1.0965	1.0982	1.0987	1.0994
7	1.0394	1.0474	1.0497	1.0523	1.0545	1.0553
8	0.9877	1.0015	1.0078	1.0122	1.0137	1.0131
9	0.9405	0.9605	0.9654	0.9726	0.9729	0.9779
10	0.9080	0.9100	0.9247	0.9291	0.9322	0.9393

Table 4.5.9: Results for European geometric average put options in dimensions 3-10 using low distortion grids.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5379	1.5379	1.5379	1.5379	1.5380	1.5380
4	1.4172	1.4178	1.4179	1.4179	1.4179	1.4179
5	1.3393	1.3403	1.3410	1.3409	1.3410	1.3410
6	1.2867	1.2859	1.2863	1.2871	1.2868	1.2878
7	1.2448	1.2478	1.2461	1.2472	1.2477	1.2471
8	1.2167	1.2174	1.2187	1.2191	1.2178	1.2147
9	1.2017	1.1980	1.1964	1.2011	1.1974	1.1998
10	1.2101	1.1913	1.1888	1.1817	1.1853	1.1853

Table 4.5.10: Results for Bermudan geometric average put options in dimensions 3-10 using low distortion grids, using the European price as a control variate.

d	5×10^4	10×10^4	15×10^4	20×10^4	25×10^4	30×10^4
3	1.5590	1.5592	1.5592	1.5592	1.5593	1.5593
4	1.4372	1.4378	1.4380	1.4380	1.4381	1.4381
5	1.3584	1.3593	1.3601	1.3599	1.3602	1.3603
6	1.3048	1.3042	1.3047	1.3056	1.3054	1.3064
7	1.2628	1.2661	1.2639	1.2654	1.2654	1.2652
8	1.2352	1.2353	1.2369	1.2376	1.2359	1.2327
9	1.2196	1.2164	1.2150	1.2193	1.2149	1.2184
10	1.2315	1.2111	1.2072	1.2002	1.2048	1.2045

Table 4.5.11: Results for American geometric average put options in dimensions 3-10 on low distortion grids, using the European price as a control variate.

Option type	c_1	c_2	R^2
European	$-0.35(\pm 0.23)$	$-1.91(\pm 0.10)$	0.971
Bermudan	$-0.42(\pm 0.14)$	$-1.85(\pm 0.06)$	0.988
American	$-0.55(\pm 0.13)$	$-1.74(\pm 0.05)$	0.989

Table 4.5.12: Regression coefficients for the error behaviour on normal Sobol' grids (95% CI in parentheses).

Option type	c_1	c_2	R^2
European	$-0.49(\pm 0.12)$	$-1.94(\pm 0.05)$	0.992
Bermudan	$-0.59(\pm 0.08)$	$-1.84(\pm 0.03)$	0.997
American	$-0.83(\pm 0.08)$	$-1.65(\pm 0.03)$	0.995

Table 4.5.13: Regression coefficients for the error behaviour on low distortion grids (95% CI in parentheses).

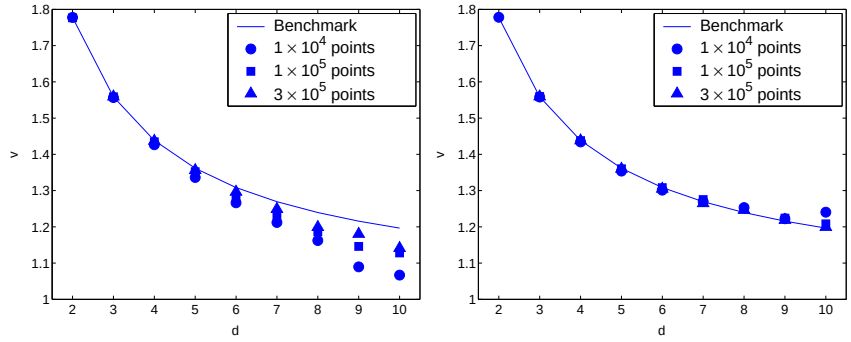


Figure 4.5.2: American pricing results for normal Sobol' grids presented raw (left) and using European price as control variate (right).

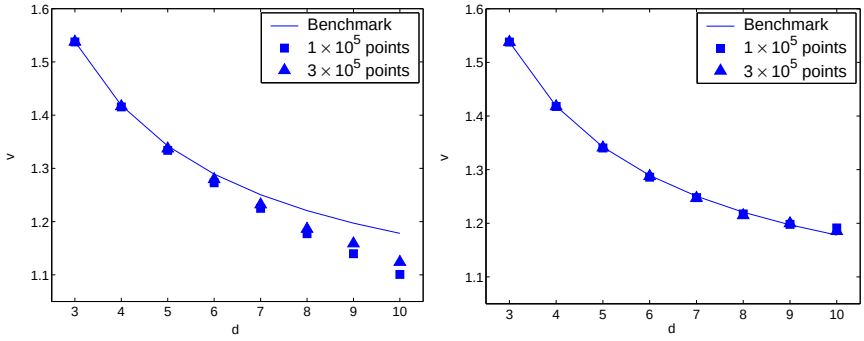


Figure 4.5.3: Bermudan pricing results for low distortion grids presented raw (left) and using European price as control variate (right).

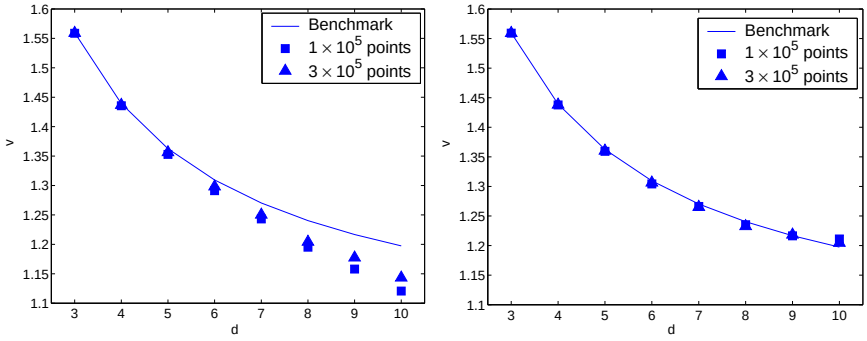
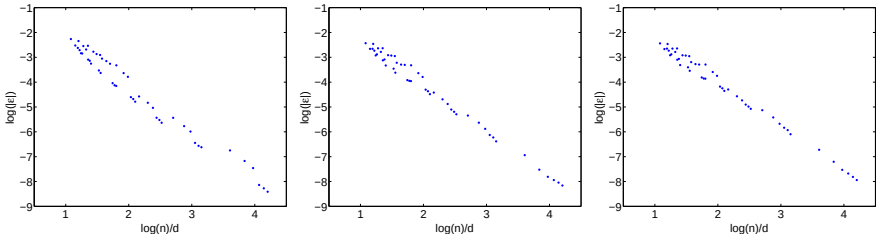


Figure 4.5.4: American pricing results for low distortion grids presented raw (left) and using European price as control variate (right).



(a) European (b) Bermudan (c) American

Figure 4.5.5: Log of absolute errors for European, Bermudan and American geometric average options plotted against $\log(n)/d$ for $d = 3, \ldots, 10$ for normal Sobol' grids. The points nearly lie in a straight line in all three cases, giving a clear indication of complexity. See Table 4.5.12 for regression results.

4.5.6 Timings

The irregular grid method presented in this chapter can be divided into two computationally intensive stages: obtaining the generator matrix and performing the time stepping. The first is the most expensive, but once a matrix has been obtained it can be reused for a wide range of related problems. We do not consider computing transition matrices here; it suffices to say that the situation is very similar to generator matrices.

Here we provide indications of the timings involved; as usual this depends heavily on the hardware and software used. The software aspect is emphasised here since there is a huge difference in the performance of different algorithms for solving the linear programming problem and for solving linear systems of equations. The experiments are carried out in Matlab on a 866MHz Pentium III under Windows 2000.

Generator matrix

In dimension d we are interested in solving a large number of linear programming problems with $\eta_d = d(d+3)/2$ equality constraints and where all variables are nonnegative. The number of variables needed is not known a priori, but it has been found that $5\eta_d$ is sufficient for points close to the center of the grid, and an increased number of $20\eta_d$ is needed closer to the boundary. The strategy is thus to order the points according to their norm and try $5\eta_d$ neighbours until a certain failure rate is reached, then to switch to $20\eta_d$ neighbours on the remaining points.

In two dimensions a single problem takes about 0.06s and is not sensitive to the number of variables changing from $5\eta_d$ to $20\eta_d$. This is probably due to the relatively large overhead involved in the Matlab routines. In five dimensions we see an increase from 0.07s for $5\eta_d$ neighbours to 0.10s for $20\eta_d$. In ten dimensions we see a corresponding increase from 0.31s to 1.90s per problem. It is thus clear that parallelisation is desirable to keep the computation times reasonable, especially for higher dimensional problems.

Time stepping

In dimension d and with n grid points we use a generator matrix with n rows each with $\eta_d + 1$ nonzero entries. The complexity of implicit time stepping should thus

be quadratic with dimension and linear with grid size.

For 300,000 points in five dimensions, explicit time steps take about 1.5s and implicit about 29s with CGS. For ten dimensions, explicit time steps take about 3.0s and implicit about 21s with CGS. The fact that implicit solutions can be faster in a higher dimension is due to the conditioning of the matrix, making it more amenable to solution even though it is more dense.

One can thus perform about 10-20 times more explicit than implicit time steps for the same running time. However there is a tradeoff since the latter generally give much better precision.

4.5.7 Boundaries

We now compare the observed boundaries presented in Figures 4.5.7 and 4.5.8 to the naive predictions in Section 4.4.2 and Figure 4.4.2.

The proportion of boundary points goes up quickly with dimension, as predicted in Section 4.4.2. A simple calculation reveals that the proportion of boundary points for a regular grid with $n^{1/d}$ steps per dimension is $1 - (1 - 2n^{-1/d})^d$. For example, for $d = 10$ one requires a grid size of about 5×10^{14} to bring the proportion of boundary points down to 0.5. Using the irregular grid method one needs about 3×10^5, as seen in Figure 4.5.7.

We cannot compare our results directly to the predictions since we used a maximum of $20\eta_d$ neighbours when trying to satisfy local consistency. A direct comparison would require that we used all points in the grid. It is clear that the observed boundaries lie at a smaller radius r than the predicted ones. This may be partially due to the small number of neighbours considered, but may also be caused by the optimism inherent in the predictions, namely that only the minimum number of neighbours is required to satisfy the local consistency conditions.

We finally note that the boundaries are not monotone with grid size in Figures 4.5.7 and 4.5.8. In the case of normal Sobol' grids, this may be attributed to the fact that new points in the grid are infeasible with respect to the closest $20\eta_d$ neighbours.

4.6 Conclusions

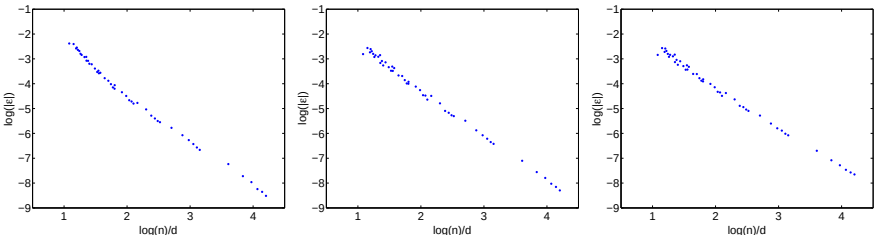
We proposed a method for pricing options with several underlying assets and an arbitrary payoff structure. The method was tested for geometric average options, which can be easily benchmarked, in dimensions three to ten with very accurate results.

We saw a decay in precision for increasing dimension, a phenomenon which can be attributed to the increasing distance between points in the approximating Markov chain, and to the increasing size of the boundary region. An analysis of the error implies that the method has exponential complexity with dimension, but the use of control variates was shown to reduce the error substantially. The use of extrapolation is also expected to provide accurate approximations, although this was not tested in the present work.

The computation of transition and generator matrices is expensive; however once generated these matrices can be reused for a large class of similar problems with time dependent parameters. Furthermore computations are cheap once the matrix is obtained.

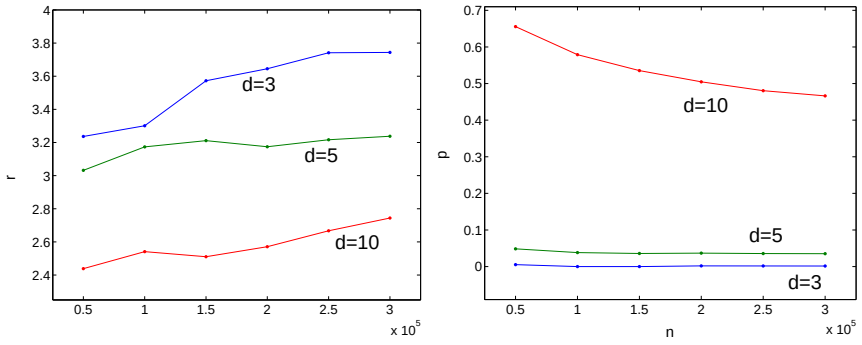
Interestingly we found little difference in complexity between the cases where Sobol' and low distortion grids were employed. The complexity observed was exponential in dimension, of approximately the same order as regular grid discretisations.

Although the method extends naturally in principle to arbitrary Markov processes with parameters depending on state and time, further extensions to the numerical procedures are required to make the proposed method computationally attractive in such cases. For example, this is of interest when considering Bermudan swaptions where the drift is state dependent.



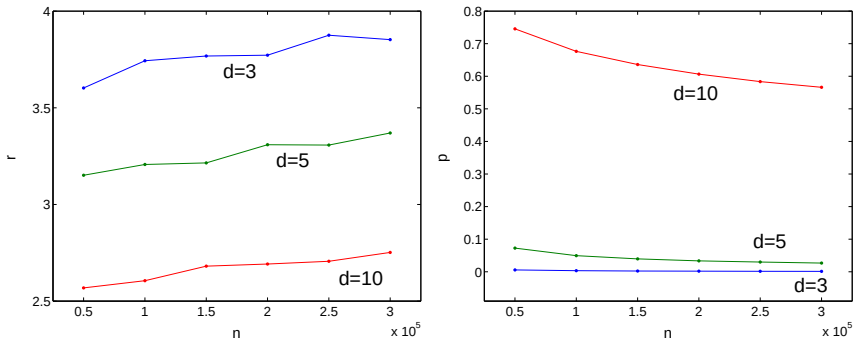
(a) European (b) Bermudan (c) American

Figure 4.5.6: Log of absolute errors for European, Bermudan and American geometric average options plotted against $\log(n)/d$ for $d = 3, \ldots, 10$ for low distortion grids. The points nearly lie in a straight line in all three cases, giving a clear indication of complexity. See Table 4.5.13 for regression results.



(a) Radius (b) Proportion

Figure 4.5.7: Smallest norms of points in normal Sobol' grids for which local consistency could not be satisfied and proportion of points in the boundary region with $20\eta_d$ nearest neighbours. Compare Figure 4.4.2.



(a) Radius (b) Proportion

Figure 4.5.8: Smallest norms of points in normal low distortion grids for which local consistency could not be satisfied and proportion of points in the boundary region with $20\eta_d$ nearest neighbours. Compare Figure 4.4.2.

Chapter 5

A Method Using Interpolation

5.1 Introduction

The prevalence of high-dimensional Bermudan-style derivative contracts in world markets, particularly those based on interest rates, has exposed a need for efficient algorithms which provide accurate price estimates. The Bermudan interest rate swap option (swaption) is an important practical example of such a high-dimensional problem.

While much focus in recent literature has been placed on dimensionality reduction, it has been argued for example in Longstaff et al. [49] that the application of single factor exercise strategies may result in significant losses. The necessity for multifactor models is also emphasised in Sidenius [67], who finds a strong sensitivity of derivative prices to the number of factors used in the model. He further argues that a large number of factors is required to obtain stationary estimates of the volatility term structure, which is what one typically expects to see in the market.

We therefore make the underlying assumption that we are dealing with a problem whose dimensionality cannot be reduced to a level which is manageable through the use of traditional methods. In practise this means we focus on Bermudan problems depending on at least three or four factors.

The value of such contracts, given the risk-neutral dynamics and payoff structure, is easily formulated as a dynamic programming problem using a continuous state space and a number of time steps corresponding to the number of exercise opportunities. In solving such problems on a digital computer, however, one faces the

challenge of approximating the continuous problem by a discrete one whose solution is close to the true solution and which can be solved in a reasonable amount of time.

There are several ways of performing such a reduction, but only a few are computationally attractive for high-dimensional problems. Carrière [20] and Longstaff and Schwartz [50] propose and test methods based on first estimating the optimal stopping rule. Once a stopping rule is found, the problem of finding the value becomes one of integration rather than optimisation, and an unbiased estimate of a lower bound on the option value (given that the approximate stopping rule is generally suboptimal) may be obtained through Monte Carlo simulation.

A dual approach, of which variants are proposed by Rogers [64], Haugh and Kogan [37] and developed further by Jamshidian [44] and Kolodko and Schoenmakers [45], approaches the problem from the hedging perspective. In this method one searches for an optimal martingale which, when viewed as the relative value of a hedging portfolio, minimises the cost of replicating the claim. Again, once such a martingale is found, the problem becomes linear, and now an unbiased estimate of an upper bound on the option value (given that the approximate hedging strategy is also generally suboptimal) may be obtained through simulation.

These methods are attractive in that they provide estimates with known biases, but both methods require optimisation over a functional space; the stopping rule being a function of the many underlying factors, and the hedging portfolio being a martingale based on these factors. Finding approximations to the optimal strategies usually involves searching in a finite-dimensional class parameterised by some basis, which must be chosen in a clever manner.

Our objective here is to estimate prices of Bermudan swaptions in the LIBOR market model (LMM). The LMM was developed by Jamshidian [43] and Brace et al. [15], and models the forward LIBOR rates which may be observed directly in the market. This model has been favoured over its sibling, the swap market model (SMM), because of the relative ease of pricing swap contracts in the LMM as opposed to pricing caps in the SMM. This is explained in some detail in Pietersz and Pelsser [63]. Geometric average options are also investigated, since accurate benchmarks are available through reduction to a one-dimensional problem.

Numerical methods for swaption pricing in the LMM have been treated by several authors, including Andersen [1] and Pedersen [61]. The predominant ap-

proach is in line with that of [20, 50], namely to first find an exercise rule and then to estimate the swaption value through simulation. We use the results of [1] as benchmarks in our experiments.

We propose an intuitive method based on a state space discretisation of the dynamic programming problem. Tsitsiklis and Van Roy [71] suggest such a method for high-dimensional Bermudan problems where the continuation values are determined using dynamic programming together with projections onto a space of features. Our approach differs however in that we do not specify features, but consider separate interpolation and approximate quadrature operators acting on the value function at each time step.

The approximate quadrature operator is constructed using standardised quasi-Monte Carlo (QMC) draws, as suggested in [21], for the geometric average options. The use of low discrepancy methods becomes more difficult in the LMM due to the time- and state-dependency of the parameters; in this case we propose antithetic simulations.

The interpolation operator seems to be a more difficult problem. We investigate two methods: nearest neighbour interpolation and local quadratic interpolation (linear interpolation was also investigated, but the results are not presented as the method produced very large biases).

Nearest neighbour interpolation is perhaps the simplest method, producing piecewise constant interpolants whose values may be determined quickly through the use of fast nearest neighbour searching techniques.

Local polynomial interpolation, or moving least squares, has been a topic of active research recently in the literature on mesh-free methods. Levin [47] and Wendland [72] show that such interpolants have a precision proportional to h^m where the fill distance h is a measure of the grid resolution and m is the order of the polynomials used. Unfortunately the method is very slow when applied directly due to the need for matrix inversion at each interpolation point. Maz'ya and Schmidt [53] propose an interesting matrix-free method for irregular grids which avoids the computational complexity, and much inspiration may be drawn from the literature on nonparametric statistics. We do not investigate such extensions here however.

Fasshauer et al. [30] and Hon [39] have applied moving least-squares methods to option pricing problems with some success; they do not however consider high-

dimensional problems.

We proceed to introduce the LMM and swaption pricing problem in Section 5.2. We present our methodology in Section 5.3 and experiments in Section 5.4. Conclusions are drawn in Section 5.5.

5.2 Bermudan swaption pricing in the LIBOR market model

In the LMM, forward rates are used to set up an arbitrage-free and easily calibrated formulation of the term structure. The pricing of swaption contracts in the LMM is by no means straightforward, especially when the contract allows early exercise.

5.2.1 Setup of the LMM

We introduce here the basic concepts of the LMM, referring the reader to Brace et al. [15] and Jamshidian [43] for the original presentation, and to Pelsser [62], Andersen and Andreasen [2] and Brigo and Mercurio [17] for more recent developments.

Consider the time steps $t_0 < \cdots < t_K$ (these are reset dates for the tenor structure), where each t_k may be thought of as referring to a swaption exercise opportunity. Let $\alpha_k = t_{k+1} - t_k$ be the day-count fraction between reset dates and $n(t)$ the next reset date function defined such that $t_{n(t)-1} \leq t < t_{n(t)}$.

Denote by $D_k(t)$ the price at time t of a discount bond maturing at t_k; the LIBOR rate corresponding to the period (t_k, t_{k+1}) is then

$$L_k(t) = \frac{1}{\alpha_k} \frac{D_k(t) - D_{k+1}(t)}{D_{k+1}(t)}. \tag{5.2.1}$$

One may also express the D_k in terms of the L_k :

$$D_k(t_j) = \prod_{i=j}^{k-1} (1 + \alpha_i L_i(t_j))^{-1}. \tag{5.2.2}$$

In the LMM, we assume that each $L_k(t)$ follows a diffusion law. All active rates L_k, $k = n(t), \ldots, K$, are brought under the terminal measure (the measure

under which L_K is a martingale) through a change of numeraire and application of Girsanov's theorem. This leads us to consider the correlated diffusion

$$\frac{dL_k(t)}{L_k(t)} = \mu_k(L(t),t)dt + \sigma_k(L(t),t)dW(t) \tag{5.2.3}$$

where $L(t) = (L_0(t), \ldots, L_K(t))'$, $W(t)$ is a Brownian motion under the terminal measure having dimension $d \leq K$ and σ_k are d-vectors giving the volatilities and correlations. The risk-neutral drift for each L_k is given by

$$\mu_k(L(t),t) = -\sum_{i=n(t)}^{K} \frac{\alpha_i \sigma_i(L(t),t) L_i(t)}{1+\alpha_i L_i(t)} \sigma_k(L(t),t), \tag{5.2.4}$$

and the σ_i are chosen appropriately, for example to fit market data within a parameterised setting.

5.2.2 Swaption pricing

An interest rate swap is a contract allowing the holder to exchange one set of interest rate payments for another. In the present work we limit the presentation to payer swaps, where one exchanges a floating set of payments for fixed payments. A European swaption is then a contract which gives the holder the right, but not the obligation, to enter into a swap contract at a certain future date, and a Bermudan swaption allows the holder to enter into a swap contract at any one of a prespecified set of dates (though not more than one).

We further limit our discussion to the case where the floating payments are determined by the LIBOR forward rate process. The value of a swap contract is determined by the discount bonds related to payment dates of the swap at time t_k as

$$v_k(t_k) = 1 - D_K(t_k) - K \sum_{i=k}^{K-1} \alpha_i D_{i+1}(t_k) \tag{5.2.5}$$

where t_K is the final maturity of the swap contract. Since bond prices may be determined from LIBOR rates through (5.2.2), the swap value may also be expressed through the LIBOR rates $L_k(t)$.

A European swaption expiring at t_k takes the value $\psi_k(t_k) = \max(v_k(t_k), 0)$ at time t_k, and at times $t < t_k$

$$v_k(t) = D_K(t)\mathbb{E}^K\left(\frac{\psi_k(t_k)}{D_K(t_k)}\right) \tag{5.2.6}$$

where the expectation is taken under the terminal measure. We thus see that in the European case the swaption value is an expectation of a known payoff with respect to an unknown density.

In the Bermudan case we add an extra complication to the problem in the form of multiple exercise opportunities. As in Andersen [1], we characterise the Bermudan contract with three dates: the first date where exercise is allowed, t_s (otherwise known as the lockout date); the last date where exercise is allowed, t_x; and the final swap maturity t_e. We assume that $t_0 < t_s < t_x < t_e$, and in the examples we only consider cases where $x = e - 1$. We now formulate the value of such a Bermudan swaption as

$$v_k(t) = D_K(t) \sup_{\tau \in \mathcal{T}_{s,x}} \mathbb{E}^K \left(\frac{\psi_k(\tau)}{D_K(\tau)} \right) \tag{5.2.7}$$

where $\mathcal{T}_{s,x}$ is the set of stopping times on $t_s, \ldots, t_x$ with respect to the natural filtration. The Bermudan swaption value is thus the expectation of a known payoff with respect to an unknown density, where the expectation is evaluated using the optimal stopping rule $\hat{\tau}$.

Neither the European nor the Bermudan case admits a closed form solution, and thus numerical methods must be employed to find approximate values.

5.3 Methodology

We focus on Bermudan options with a fixed set of exercise opportunities. Values are computed only at the exercise dates, the continuation value being approximated using numerical integration of the value function at the following exercise opportunity, which is extended to the entire state space using interpolation.

We first present a solution method which is computationally tractable for the kind of problem presented in Section 5.2. We prove convergence of this method under certain conditions.

5.3.1 Framework and assumptions

In order to simplify the analysis, we assume that any discounting is incorporated in the value functions, so that all value functions are given in time zero currency units.

We assume exercise is possible at times t_k for $k = 1, \ldots, K$ and where

$$0 = t_0 < t_1 < \cdots < t_K = T. \tag{5.3.1}$$

The case where exercise is not possible at $t = 0$ can be taken into account with obvious modifications in the following.

We consider problems where, at each exercise opportunity t_k, the value function can be represented as $v_k : \Omega_k \to \mathbb{R}$ where $\Omega_k \subseteq \mathbb{R}^{d_k}$ for some appropriate dimension d_k.

At each exercise opportunity we allow an arbitrary grid

$$\mathcal{X}_k = \{x_{k,1}, \ldots, x_{k,n_k}\} \subset \Omega_k \tag{5.3.2}$$

on which the value function is to be estimated, where the number of points in grid $\mathcal{X}_k$ is n_k. Our approximation to the value function at time t_k and state $x_{k,i}$ is denoted $\hat{v}_{k,i}$. Clearly the grids $\mathcal{X}_k$ should be chosen with importance sampling considerations in mind.

The most difficult part of implementing a dynamic programming algorithm for Bermudan options computationally is in estimating the continuation values. We now introduce the two operators which will be used to estimate these.

Let $\mathcal{Q}_{t_k,x}$ be the appropriate quadrature operator for time t_k and state x given by

$$\mathcal{Q}_{t_k,x} f = \int_{\Omega_{k+1}} f(y)\ dp(X(t_{k+1}) = y | X(t_k) = x) \tag{5.3.3}$$

where $X(t)$ is the stochastic process followed by the underlying variables and $p(\cdot|\cdot)$ is the conditional density implied by the process. Now define $\mathcal{Q}_{t_k,x,m}$ to be a numerical approximation to $\mathcal{Q}_{t_k,x}$ where m is a parameter affecting the precision of $\mathcal{Q}_{t_k,x,m}$. For notational compactness we also define the entire quadrature and its approximate version, respectively, as

$$\mathcal{Q}_k(x) = \mathcal{Q}_{t_k,x} \tag{5.3.4}$$

$$\mathcal{Q}_{k,m}(x) = \mathcal{Q}_{t_k,x,m}. \tag{5.3.5}$$

Further, let $\mathcal{I}_{\mathcal{X}}$ be an interpolation operator taking a vector of function values $\hat{f}$ defined on the grid $\mathcal{X}$, and returning a function in some class $\mathcal{C}_{\mathcal{I}}$, and defined on the entire state space,

$$\mathcal{I}_{\mathcal{X}} : \mathbb{R}^{\mathcal{X}} \to \mathcal{C}_{\mathcal{I}}(\Omega_k, \mathbb{R}). \tag{5.3.6}$$

For this purpose we define $\hat{v}_k$ as the vector having ith entry $\hat{v}_{k,i}$. Continuation value estimates are denoted $\hat{c}_{k,i}$, and $\hat{c}_k$ is the vector with ith entry $\hat{c}_{k,i}$.

We also define the linear operator $\mathcal{P}_{\mathcal{X}}$ to be the projection operator taking a function and returning a vector of values of the function at the grid points in $\mathcal{X}$:

$$\mathcal{P}_{\mathcal{X}} f = (f(x_1), \ldots, f(x_{n_k}))'. \tag{5.3.7}$$

For ease of notation we denote $\mathcal{P}_k = \mathcal{P}_{\mathcal{X}_k}$ and $\mathcal{I}_k = \mathcal{I}_{\mathcal{X}_k}$.

We note that the true value functions at the exercise opportunities are given by Algorithm 5.3.1. We propose solving the dynamic programming problem computationally as shown in Algorithm 5.3.2.

Algorithm 5.3.1 Exact algorithm for finding Bermudan option price where $v_k(\cdot)$ are the value functions, $c_k(\cdot)$ are the continuation values, $\psi_k(\cdot)$ are the payoff functions and $\mathcal{Q}_k$ are the entire quadrature operators, all for each time t_k.

$v_K(\cdot) := \psi_K(\cdot)$
for $k = K-1, \ldots, 0$ **do**
  $c_k(\cdot) := \mathcal{Q}_k \; v_{k+1}(\cdot)$
  $v_k(\cdot) := \max(\psi_k(\cdot), c_k(\cdot))$
end for

Algorithm 5.3.2 Numerical algorithm for finding Bermudan option price, where the $\hat{v}_k$ are the vectors of approximate values, $\hat{c}_k$ are the estimated continuation values, $\hat{\psi}_k$ are the vectors of (exact) payoffs and $\mathcal{Q}_{k,m}$ are the approximate entire quadrature operators, all for each time t_k.

$\hat{v}_K := \hat{\psi}_K$
for $k = K-1, \ldots, 0$ **do**
  $\hat{c}_k := \mathcal{P}_k \mathcal{Q}_{k,m} \; \mathcal{I}_{k+1} \; \hat{v}_{k+1}$
  $\hat{v}_k := \max(\hat{\psi}_k, \hat{c}_k)$
end for

Note that, since the value function at expiry v_K is known exactly, the interpolation operator in Algorithm 5.3.2 need not be applied at step $k = K-1$; instead $\mathcal{I}_{\mathcal{X}} \hat{v}_K$ is replaced by v_K. This implies further that the grid $\mathcal{X}_K$ need not be used in practise.

5.3.2 Convergence

The following elementary lemma states that a maximum operator of the sort applied in Algorithm 5.3.2 does not increase the accumulated error.

Lemma 5.3.1 *Let b, c and $\tilde{c}$ be real numbers. Define*

$$\begin{aligned} a &= \max(b, c) \\ \tilde{a} &= \max(b, \tilde{c}). \end{aligned}$$

Then

$$|a - \tilde{a}| \leq |c - \tilde{c}|. \tag{5.3.8}$$

Remark 5.3.1 *Lemma 5.3.1 merely states that the* $\max$ *function is Lipschitz in both of its arguments.*

Before stating the convergence theorem, we state some conditions that will be used. We assume that the operators $\mathcal{Q}_{k,m}$ and $\mathcal{I}_k$ are consistent in that, for every $\varepsilon_{\mathcal{Q},m},\ \varepsilon_{\mathcal{I}} > 0$ we can find $\bar{m},\ \bar{n}_k$ such that for all $m > \bar{m},\ n_k > \bar{n}_k$

$$|\mathcal{Q}_{k,m} v_{k+1} - \mathcal{Q}_k v_{k+1}|_\infty \leq \varepsilon_{\mathcal{Q},m} \tag{5.3.9}$$

$$|\mathcal{I}_k \mathcal{P}_k v_k - v_k|_\infty \leq \varepsilon_{\mathcal{I}} \tag{5.3.10}$$

for all k where v_k is defined in Algorithm 5.3.1. We also assume the operators are Lipschitz continuous in that there exist constants $c_{\mathcal{Q},m},\ c_{\mathcal{I}}$ such that

$$|\mathcal{Q}_{k,m} f|_\infty \leq c_{\mathcal{Q},m} |f|_\infty \tag{5.3.11}$$

$$|\mathcal{I}_k g|_\infty \leq c_{\mathcal{I}} |g|_\infty \tag{5.3.12}$$

for all k where $f \in \mathcal{C}_{\mathcal{I}} + \operatorname{linspan}\{v_{k+1}\}$ and g is any n_k-vector. We also assume $\mathcal{Q}_{k,m}$ to be linear in that

$$\mathcal{Q}_{k,m}(f + g) = \mathcal{Q}_{k,m} f + \mathcal{Q}_{k,m} g \tag{5.3.13}$$

for all $k,\ m$.

Theorem 5.3.1 *Under the conditions (5.3.9)–(5.3.12), the approximations $\hat{v}_{t_0,x_i}$ defined through Algorithm 5.3.2 converge uniformly to the exact solutions $v(t_0, x_i)$ of Algorithm 5.3.1 as $m,\ n_k \to \infty$.*

Proof. We first note that the error accumulated in the method arises solely from the interpolation and quadrature errors at previous time steps. We estimate the error as follows:

$$\begin{aligned} |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_k v_{k+1}|_\infty &\leq |\mathcal{Q}_{k,m}v_{k+1} - \mathcal{Q}_k v_{k+1}|_\infty \\ &\quad + |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1} - \mathcal{Q}_{k,m}v_{k+1}|_\infty \\ &\quad + |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_{k,m}\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1}|_\infty . \end{aligned} \tag{5.3.14}$$

The terms on the RHS of (5.3.14) are termed the integration error, interpolation error and accumulated error, respectively. By assumption, the integration error is bounded by $\varepsilon_{\mathcal{Q},m}$ for all v_k. The interpolation and accumulated errors are estimated, respectively, by

$$\begin{aligned} |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1} - \mathcal{Q}_{k,m}v_{k+1}|_\infty &= |\mathcal{Q}_{k,m}\left(\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1} - v_{k+1}\right)|_\infty \\ &\leq c_{\mathcal{Q},m}\,|\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1} - v_{k+1}|_\infty \\ &\leq c_{\mathcal{Q},m}\varepsilon_{\mathcal{I}} \end{aligned} \tag{5.3.15}$$

$$\begin{aligned} |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_{k,m}\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1}|_\infty &= |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\left(\hat{v}_{k+1} - \mathcal{P}_{k+1}v_{k+1}\right)|_\infty \\ &\leq c_{\mathcal{Q},m}\,|\mathcal{I}_{k+1}\left(\hat{v}_{k+1} - \mathcal{P}_{k+1}v_{k+1}\right)|_\infty \\ &\leq c_{\mathcal{Q},m}c_{\mathcal{I}}\,|\hat{v}_{k+1} - \mathcal{P}_{k+1}v_{k+1}|_\infty . \end{aligned} \tag{5.3.16}$$

We note that $\mathcal{P}_k v_k$ is just the vector of values at grid points, having ith entry $v(t_k, x_i)$. Hence, using the previous bounds and Lemma 5.3.1, the error bound for time step k is

$$\begin{aligned} |\hat{v}_k - \mathcal{P}_k v_k|_\infty &\leq |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_k v_{k+1}|_\infty \\ &\leq \varepsilon_{\mathcal{Q},m} + c_{\mathcal{Q},m}\varepsilon_{\mathcal{I}} + c_{\mathcal{Q},m}c_{\mathcal{I}}\,|\hat{v}_{k+1} - \mathcal{P}_{k+1}v_{k+1}|_\infty . \end{aligned} \tag{5.3.17}$$

Since we know the terminal conditions, we have $|\hat{v}_K - \mathcal{P}_K v_K|_\infty = 0$. Apply-

ing (5.3.17) recursively, we obtain

$$|\hat{v}_0 - v_0|_\infty \quad \leq \quad \left(1 + c_{\mathcal{Q},m}c_{\mathcal{I}} + \cdots + (c_{\mathcal{Q},m}c_{\mathcal{I}})^{K-1}\right)(\varepsilon_{\mathcal{Q},m} + c_{\mathcal{Q},m}\varepsilon_{\mathcal{I}})\,. \tag{5.3.18}$$

Hence choosing the m and n_k large enough allows us to achieve an arbitrarily small error.

■

Note that in the preceding theorem, one may reduce the error estimates by choosing operators which have small continuity constants $c_{\mathcal{Q},m},\ c_{\mathcal{I}}$.

The consistency condition on the interpolation operator (5.3.10) is very strong. In particular, due to our use of the L^∞ norm, (5.3.10) stipulates a uniform bound on the interpolation error over the entire state space. On the contrary, interpolation errors for areas of the state space which fall in the tails of the (unconditional) density should only have a minor effect on the accuracy of the solution, and thus greater errors may be tolerated in these tail areas. We now propose a framework which takes this into account by using measures which are adapted to the process density.

Consider now a sequence of measures μ_k relating to t_k for $k = 0, \ldots, K$, which will be used in estimating the accumulated error. We define discrete counterparts of these measures $\hat{\mu}_k$ which are related to the grids $\mathcal{X}_k$ in that there exists a constant $c_{\mathcal{P}}$ such that

$$|\mathcal{P}_k f|_{\hat{\mu}_k} \leq c_{\mathcal{P}}\,|f|_{\mu_k} \tag{5.3.19}$$

for all k where $\mathcal{P}_{\mathcal{X}}$ is the linear projection operator given in (5.3.7).

The approximate quadrature and interpolation operators are now assumed to be consistent in that for any $\varepsilon_{\mathcal{Q},m},\ \varepsilon_{\mathcal{I}} > 0$ we can find $\bar{m},\ \bar{n}_k$ such that for all $m > \bar{m},\ n_k > \bar{n}_k$

$$|\mathcal{Q}_{k,m}v_{k+1} - \mathcal{Q}_k v_{k+1}|_{\mu_k} \leq \varepsilon_{\mathcal{Q},m} \tag{5.3.20}$$

$$\left|\mathcal{I}_k\mathcal{P}_k v_k - v_k\right|_{\mu_k} \leq \varepsilon_{\mathcal{I}} \tag{5.3.21}$$

for all k. We further assume the existence of continuity constants $c_{\mathcal{Q},m},\ c_{\mathcal{I}}$ such that

$$|\mathcal{Q}_{k,m}f_{k+1}|_{\mu_k} \quad \leq \quad c_{\mathcal{Q},m}|f_{k+1}|_{\mu_{k+1}} \tag{5.3.22}$$

$$|\mathcal{I}_k g|_{\mu_k} \quad \leq \quad c_{\mathcal{I}}|g|_{\hat{\mu}_k} \tag{5.3.23}$$

for all k and where $f_{k+1} \in \mathcal{C}_{\mathcal{I}} + \text{linspan}\{v_{k+1}\}$ for each k, and g is any n_k-vector. As in (5.3.13) above, $\mathcal{Q}_{k,m}$ is again assumed to be linear.

Theorem 5.3.2 *Under the conditions (5.3.13), (5.3.19)–(5.3.23), the approximations $\hat{v}_{t_0,x_i}$ defined through Algorithm 5.3.2 converge in the sense of μ_k to the exact solutions $v(t_0, x_i)$ of Algorithm 5.3.1 as $m,\ n_k \to \infty$.*

Proof. The proof is similar to that of Theorem 5.3.1, except for the different norms and that the inequality in the first line of (5.3.17) is no longer valid in general. We now have

$$\begin{aligned} |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_k v_{k+1}|_{\mu_k} \quad \leq \quad & |\mathcal{Q}_{k,m}v_{k+1} - \mathcal{Q}_k v_{k+1}|_{\mu_k} \\ & + \ |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1} - \mathcal{Q}_{k,m}v_{k+1}|_{\mu_k} \\ & + \ |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_{k,m}\mathcal{I}_{k+1}\mathcal{P}_{k+1}v_{k+1}|_{\mu_k} . \end{aligned} \tag{5.3.24}$$

Using Lemma 5.3.1, conditions (5.3.13), (5.3.19)–(5.3.23) and (5.3.24) we have

$$\begin{aligned} |\hat{v}_k - \mathcal{P}_k v_k|_{\hat{\mu}_k} \quad &\leq \quad |\mathcal{P}_k \left(\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_k v_{k+1}\right)|_{\hat{\mu}_k} \\ &\leq \quad c_{\mathcal{P}} \, |\mathcal{Q}_{k,m}\mathcal{I}_{k+1}\hat{v}_{k+1} - \mathcal{Q}_k v_{k+1}|_{\mu_k} \\ &\leq \quad c_{\mathcal{P}} \left(\varepsilon_{\mathcal{Q},m} + c_{\mathcal{Q},m}\varepsilon_{\mathcal{I}} + c_{\mathcal{Q},m}c_{\mathcal{I}} \, |\hat{v}_{k+1} - \mathcal{P}_{k+1}v_{k+1}|_{\hat{\mu}_{k+1}}\right) . \end{aligned} \tag{5.3.25}$$

Applying (5.3.25) recursively and noting $\hat{v}_K - \mathcal{P}_K v_K \equiv 0$ gives

$$|\hat{v}_0 - v_0|_{\hat{\mu}_0} \quad \leq \quad c_{\mathcal{P}} \left(1 + c_{\mathcal{Q},m}c_{\mathcal{I}} + \cdots + (c_{\mathcal{Q},m}c_{\mathcal{I}})^{K-1}\right) \left(\varepsilon_{\mathcal{Q},m} + c_{\mathcal{Q},m}\varepsilon_{\mathcal{I}}\right) . \tag{5.3.26}$$

■

Remark 5.3.2 The consistency conditions (5.3.9), (5.3.10), (5.3.20) and (5.3.21) may be difficult to verify in practise. In fact these are joint conditions on the operators and the function class to which the solutions v_k belong; the conditions being easier to satisfy for more regular functions. As mentioned above, conditions

(5.3.9) and (5.3.10) are rather strict and are not expected to hold for many practical examples. Conditions (5.3.20) and (5.3.21) on the other hand may be expected to hold for a much wider class of functions, for given operators, and where the measures μ_k are chosen appropriately.

5.3.3 Approximating the quadrature operator

The quadrature operator as defined in (5.3.3) is an integral with respect to the conditional density of the process, which is in most cases not known explicitly. The most general case which yields tractability is where, given that $X(t_k) = x$, we may simulate values of $X(t_{k+1})$. We base our construction of the quadrature approximation $\mathcal{Q}_{t_k,m}$ on this case.

Note however that the LMM does not satisfy this assumption strictly, although one may obtain simulated values which are distributed arbitrarily well by decreasing the time step.

We consider the following quadrature rule

$$\mathcal{Q}_{t_k,x,m} f = \frac{1}{m} \sum_{i=1}^{m} f(X_i) \tag{5.3.27}$$

where the i.i.d. variables $X_i \sim p(X(t_{k+1}) = y | X(t_k) = x)$ may be evaluated through simulation. This quadrature approximation does not satisfy the assumptions of consistency (5.3.9) strictly, but for some $\varepsilon_{\mathcal{Q},m}$, m may be chosen such that the condition is satisfied with a probability close to 1. The continuity (5.3.22) and linearity (5.3.13) assumptions are clearly satisfied.

The literature on QMC methods suggests that in many cases the finite-sample precision of this quadrature operator may be dramatically improved by using low discrepancy numbers in the simulation procedure. The use of QMC or low distortion points [3] may also lead to deterministic error bounds in (5.3.27). If neither QMC nor low distortion methods are applicable, one is usually able to incorporate variance reduction techniques such as antithetic variables in simulations.

5.3.4 The interpolation operator $\mathcal{I}_{\mathcal{X}}$

The interpolation operator represents the manner in which we extend information from the grid to a function on the entire state space. Clearly such a function must

be well-behaved in relation to the value function, as expressed in the consistency and continuity conditions of Section 5.3.1.

The consistency condition (5.3.10) for Theorem 5.3.1 is perhaps the more difficult of these assumptions to satisfy, in particular due to the unbounded domain. The respective condition (5.3.21) for Theorem 5.3.2 is easier to satisfy. Consider for example the nearest neighbour interpolation

$$\left[\mathcal{I}_{\mathcal{X}}\hat{f}\right](x) = f_{argmin_i|x-x_i|} \tag{5.3.28}$$

which only satisfies (5.3.10) when the function to be approximated converges to a constant as $|x| \to \infty$, but which satisfies (5.3.21) so long as the function to be approximated is integrable with respect to μ_k, and the $\mathcal{X}_k$ are chosen appropriately.

Another candidate, which has received much attention recently in the literature on mesh free methods, is the local polynomial reproduction, or moving least-squares method. This is a class of methods using information from nearby points to form an interpolant, and has previously received attention in the option pricing literature in Hon [39] and Fasshauer et al. [30]. These papers do not however address the problem of high-dimensionality.

The method has received theoretical support in the case of irregularly-spaced points in the work of Levin [47] and Wendland [72]. To implement the local polynomial reproduction method on the grid $\mathcal{X} = \{x_1, \ldots, x_n\}$, one first chooses a class of polynomials; for example we may choose the class $\mathcal{R}_m^d$ of polynomials in d variables having degree less than or equal to m. We then look for weight functions w_j such that

$$\sum_j p(x_j)w_j(x) = p(x) \tag{5.3.29}$$

for all $p \in \mathcal{R}_m^d$. One then forms the (quasi-)interpolant

$$\left[\mathcal{I}_{\mathcal{X}}\hat{f}\right](x) = \sum_j w_j(x)f(x_j). \tag{5.3.30}$$

The extension operator in (5.3.30) is in general a quasi-interpolant because it does not necessarily satisfy the value reproduction condition $[\mathcal{I}_{\mathcal{X}}\hat{f}](x_j) = \hat{f}_j$. The weight functions $w_j(x)$ are assumed to be local by construction in that, for x far away from x_j, we have $w_j(x) = 0$.

We implement the local polynomial reproduction method for the class of quadratic functions $\mathcal{R}_2^d$ as follows. First, assume that $\xi_d = \dim \mathcal{R}_2^d \ll n$, that is the

number of grid points is much greater than the number of points required to fit quadratic functions. Now, for each point x the weight functions at that point are determined as follows: choose a number of nearest neighbours $\eta \geq \xi_d$ to use for fitting the local quadratic; only these η neighbours will receive nonzero weights in (5.3.29). The weights are calculated for $x \in \mathbb{R}^d$ through solving the underdetermined system

$$\begin{pmatrix} p_0(x_1) & \cdots & p_0(x_\eta) \\ \vdots & & \vdots \\ p_{\xi_d}(x_1) & \cdots & p_{\xi_d}(x_\eta) \end{pmatrix} \begin{pmatrix} w_1(x) \\ \vdots \\ w_\eta(x) \end{pmatrix} = \begin{pmatrix} p_0(x) \\ \vdots \\ p_{\xi_d}(x) \end{pmatrix} \tag{5.3.31}$$

for the $w_j(x)$ where $p_0, \ldots, p_{\xi_d}$ is a basis for $\mathcal{R}_2^d$. In our experiments, we use the basis $p_0 = 1, p_1 = x_1, \ldots, p_{\xi_d} = x_d^2$, and the nonzero $w_j(x)$ are determined by the columns of the matrix on the LHS having the largest norms. In this case we can thus conclude that the weight functions $w_j(x)$ are piecewise polynomial, where the pieces are exactly those regions of the state space having the same nearest neighbour set.

We note that this method is computationally intensive for moderate dimensions. In this respect promising developments are currently being made in the area of matrix-free methods for local polynomial interpolations, leading to much faster computing times (see for example [53, 29]). These methods have not been investigated for high dimensional problems however, so their effect on the current analysis is an open issue.

5.3.5 Variance reduction

There are a variety of variance reduction techniques which may be employed when using Algorithm 5.3.2. We comment on the techniques that will be employed for the experiments in Section 5.4.

For Bermudan option problems, there is almost always an effective control variate available in the form of a European counterpart, whose value can usually be found through simulation. To be effective, a control variate should have a high correlation with the problem at hand. For American options on stocks, the European option expiring at the same time as the American is often effective; for Bermudan swaptions, the European swaption expiring at the next reset date is often effective.

We consider here only *inner* control variates, as opposed to the more standard outer control variates. This means that the control variate is applied at each time step, and not only once at the end of the experiment.

For the quadrature operator $\mathcal{Q}_{t_k,x,m}$, a standard technique is to sample using a set of low discrepancy points. Asymptotically this leads to a convergence rate of close to n^{-1} rather than the usual $n^{-1/2}$ for standard Monte Carlo sampling. We have found for small samples that one can often improve the quadrature estimates markedly by normalising the first and second moments of the low discrepancy points, where this is possible. This technique is also suggested in Carrière [21], where it is found to improve finite sample behaviour in pricing a Bermudan put option.

For example, suppose we have generated the low discrepancy point set $\mathcal{X}_0 = \{x_{0,1}, \ldots, x_{0,m}\}$, to be used in the quadrature. We first normalise $\mathcal{X}_0$ with respect to the first moment; let $\mathcal{X}_1 = \{x_{1,1}, \ldots, x_{1,m}\}$ where

$$x_{1,i} = x_{0,i} - \frac{1}{m} \sum_j x_{0,j}. \tag{5.3.32}$$

The point set $\mathcal{X}_1$ now has mean zero. Let us now normalise $\mathcal{X}_1$ with respect to the second moment; let $\mathcal{X}_2 = \{x_{2,1}, \ldots, x_{2,m}\}$ where

$$x_{2,i} = R_{\mathcal{X}_1}^{-1} x_{1,i} \tag{5.3.33}$$

where $R_{\mathcal{X}_1}$ is the Cholesky factor of $\Sigma_{\mathcal{X}_1} = \frac{1}{m} \sum_j x_{1,j} x_{1,j}'$. The point set $\mathcal{X}_2$ now has zero first moment and unit second moment.

This method of normalisation is related to the ideas of local consistency presented in [46] and Chapter 4 (also published as [10]), in which transition probabilities to a given set of points are determined such that the first and second moments are matched. Here we first set the probabilities (namely equal weights adding to unity) and then select the points to be used in a locally consistent manner. Note that asymptotically the behaviour of this method is uncertain; we advocate its use only for small samples.

When standard low discrepancy point methods become less effective (e.g. for path dependent processes), we use antithetic variables.

As another possible method to improve accuracy in Bermudan problems, we consider breaking down the value function into the sum of the intrinsic value and

the early exercise premium. The former may be integrated directly, whereas the latter must first be interpolated.

5.3.6 Multiple grids

In order to further smooth the interpolation, one may consider using multiple grid solutions at each time step. This is expected to improve solution behaviour, especially in the case of nearest neighbour interpolation where the interpolant is most nonsmooth.

Briefly, one chooses a number of (different, possibly randomised) grids $\ell > 0$ to be used. At each time step, one forms approximations on each grid by applying the quadrature operator to the average of the interpolants at the previous time steps. The value function approximation is thus implicitly represented as an average of the ℓ grid solutions.

The reader is left to make the necessary adjustments to Algorithm 5.3.2.

5.3.7 Parallelism

It should be emphasised that all the methods suggested here are embarrassingly parallel in nature; that is, one can obtain an almost linear speedup by employing multiple processors. This is a result of the fact that the suggested operators can be evaluated in parallel when estimating the value function at separate points x_i.

5.4 Experiments

Numerical experiments are conducted for two examples. First we look at geometric average options in a Black-Scholes market with 1-10 dimensions. This is a useful test since, although geometric average options do not constitute an important application, accurate benchmarks are readily available for problems with arbitrary dimension. We refer the reader to Chapter 4 for derivations in the following setting.

Second, we look at Bermudan swaptions in a LIBOR setting. This is currently one of the most challenging applications in mathematical finance because of its high dimensionality and because of the time and state dependence of the parameters in the risk-neutral process. It is also an important application due to the widespread use of such contracts.

5.4.1 Bermudan geometric average options

The setting considered is the same as that in Chapter 4, to which the reader is referred for the benchmark calculations and discussion. We only present graphical results for this setting in the current work.

Briefly, we consider assets with values $S_i(t) = e^{X(t)}$ evolving according to a d-dimensional correlated geometric Brownian motion with corresponding risk neutral process

$$dX(t) = \left(r\mathbb{1} - \frac{1}{2}\operatorname{diag}\Sigma\right)dt + RdW(t) \tag{5.4.1}$$

for risk-free rate r and $R'R$ being the Cholesky decomposition of the covariance matrix $\Sigma = (\sigma_i\sigma_j\rho_{ij})$. We assume a time horizon T and payoff function

$$\psi(S) = \left(K - \left(\prod S_i\right)^{1/d}\right)^+ \tag{5.4.2}$$

where K is the strike price.

We assume starting asset values $S_i(0) = 40$ for all i and strike price 40. The risk-free rate is taken as 0.06, the volatilities $\sigma_i = 0.2$ for all i, and correlations $\rho_{ij} = 0.25$, $i \neq j$. Figures 5.4.1-5.4.4 show how the numerical method performs in predicting the early exercise premium, with increasing dimension.

We use a standardised, randomised low discrepancy quadrature operator, as discussed in Section 5.3.5. Ten trials are performed for each dimension and number of grid points, the results differing due to the randomisation involved in each case. We investigate the mean and range of the result obtained.

In Figure 5.4.1, we see the relative performance of using local quadratic and nearest neighbour interpolations; the randomisation in these cases being restricted to the quadrature operator and calculation of the control variate. For all cases, the deterioration of precision with dimension is clear. With local quadratic interpolation one sees reasonable results with 500 points, whether a control variate is used or not. With nearest neighbour interpolation we see a high bias for low dimensions and a low bias for higher dimensions, suggesting overall a combination of biases. Once the control variate is applied the results show a low bias for most cases, particularly in higher dimensions.

Figure 5.4.2 presents timings results, showing clearly the difference in computational complexity between the two interpolation methods. Indeed the nearest

neighbour method shows linear complexity, whereas the local quadratic method shows a much higher complexity. For 500 grid points in dimension ten, it takes nearly 25 minutes per experiment; for this reason we did not investigate local quadratic interpolation with 1000 points, and we restrict further experiments to the nearest neighbour method. We reserve the faster approximate methods of [53, 29] for future investigation.

Figure 5.4.3 shows results using the nearest neighbour method where the grids are randomised, and where multiple grids are used. These randomised grid methods show a high bias for higher dimensions, but when the control variate is applied, the bias is much reduced and the average result is accurate to within one cent for both 500 and 1000 point grids. The results become tighter when three grids are used as opposed to a single grid, although the average result is not affected greatly.

Figure 5.4.4 shows the effect of treating the early exercise premium and intrinsic value separately. Since interpolation is not required to integrate the intrinsic value, one may expect a performance increase in this case. In fact we see that the bias is reversed in this case, and the results become slightly tighter for the randomised cases; the improvement is not great however. The reason for the bias reversal can be explained in terms of the bias introduced through interpolation; in the original experiments the interpolated function is everywhere convex, whereas in Figure 5.4.4 the function has a downward kink at the exercise boundary.

5.4.2 Bermudan swaptions

For our experiments on the pricing of Bermudan payer swaptions, we follow the examples of Andersen [1]. Using the notation of [1], we recall the two volatility scenarios with one and two factors, respectively:

Scenario C:

$$\sigma_k(t) = 0.2 \quad \forall k,\ t \leq t_k; \tag{5.4.3}$$

Scenario D:

$$\sigma_k(t) = \left(0.15, 0.15 - \sqrt{0.009(t_k - t)}\right)' \quad \forall k,\ t \leq t_k. \tag{5.4.4}$$

We consider four of the contracts presented in [1]; the details are given in Table 5.4.1. The time between reset dates in the swaption is always $\alpha = 0.25$, and the starting LIBOR rates are assumed flat at $L_k(0) = 10\%$ for all $k = 0, \ldots, t_e - 1$.

Contract	t_s	t_x	t_e
I	0.25	1.25	1.50
II	1.00	2.75	3.00
III	1.00	5.75	6.00
IV	3.00	5.75	6.00

Table 5.4.1: Lockout time (t_s), final exercise time (t_x) and final swap maturity (t_e) of the Bermudan swap option contracts.

As in [1], we use four time steps between reset dates and an Euler discretisation in the simulation of the LIBOR process. The dimensionality of the contracts thus ranges from 5 for Contract I to 23 for Contracts III and IV.

For the control variate, the European swaption expiring at the next reset date was found to perform well.

We use 100 antithetic points and antithetic paths simulations in the quadrature. Antithetic paths were used due to the standardised QMC points losing their effectiveness with the time- and state-dependent parameters; knowledge of the first and second moments of the quadrature density could have helped in this respect. Further, 500 antithetic points were used in the inner control variate calculations.

To determine benchmarks for the early exercise premium in the experiments, we use the difference between the Monte Carlo results found in Andersen [1] for the European and Bermudan prices. The results for exercise strategy (I), as defined in [1], are used for the Bermudan.

Figures 5.4.5 and 5.4.6 show numerical results for Bermudan swaption pricing. Ten experiments were conducted for each contract, each grid size and each strike. One sees that results from a single experiment may not always be accurate, however on average the results are accurate for contracts I-II and slightly high-biased for contracts III-IV. The latter is in agreement with the geometric average pricing results, which found a high bias when integrating the value function directly. This also confirms that the longer-maturity contracts, which involve more state variables and more steps in the dynamic programming algorithm, are more difficult to price accurately.

The figures show that varying the number of grid points in the range 100-500 does not affect the results greatly, which seems surprising given that the amount of work varies five-fold over this range.

The timings presented in Figure 5.4.7 show that the work depends linearly on the number of grid points, as would be expected. We also see that the contracts with more exercise opportunities, in particular contract III, are computationally more intensive.

5.5 Conclusions

We have proposed a class of dynamic programming algorithms for solving Bermudan option pricing problems. The dynamic programming problem uses an irregular grid at each time step on which the value function is approximated. Continuation values are approximated by applying an interpolation and a quadrature operator. Convergence of the algorithm has been demonstrated provided that the operators satisfy certain boundedness, continuity and linearity conditions.

In the experiments, we find that using local quadratic approximations is very intensive computationally. This computational burden may possibly be alleviated through the use of "approximate approximations" as advocated by Maz'ya and Schmidt [53]; indeed these methods should improve the computational complexity significantly, but at an unknown cost in terms of accuracy.

In terms of complexity, it was noted that the algorithm presented is easily parallelisable; thus the computational time is limited by the number of processors available. Thus solutions may be computed much faster than what is indicated by the timings given in this chapter.

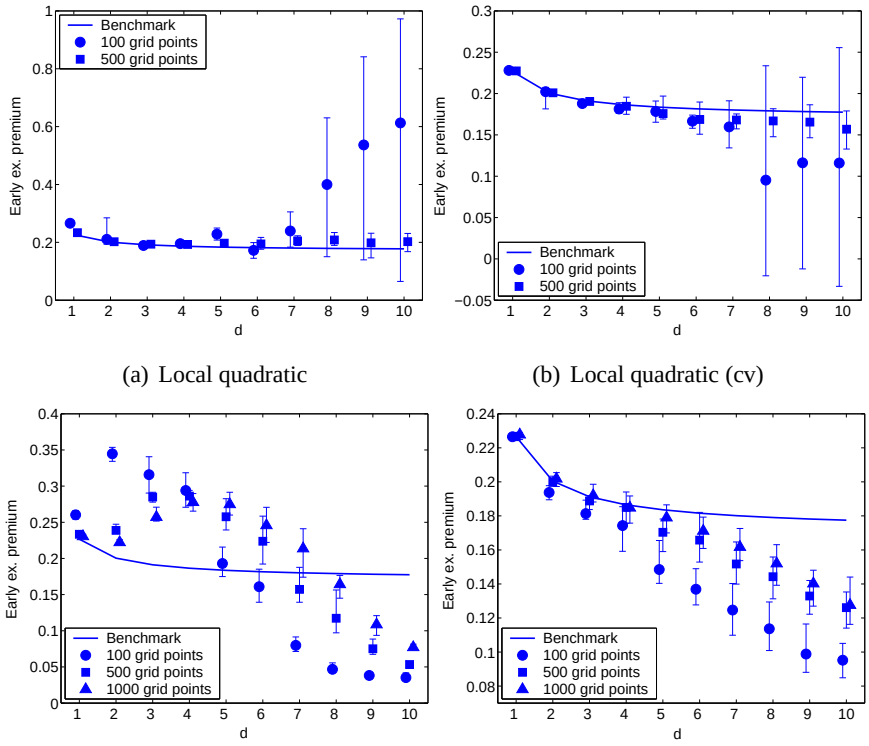
We find that using an inner control variate can improve results markedly, and indeed seems necessary to obtain reasonably unbiased estimates in our tests. We find that applying the interpolation operator directly to the value function leads to high-biased estimates; on the other hand, separating the value function into its intrinsic and early exercise components and applying the interpolation operator to the early exercise premium leads to low biased estimates. These biases are not surprising, since in the first case we are interpolating a convex function and in the second case a concave function.

What may be surprising is the relatively low number of grid points required to obtain an accurate solution. For example, the geometric average Bermudan put options, 500 points seemed sufficient with the use of an inner control variate; for the Bermudan swaptions there was little difference between the solutions from

using 100 and 500 points. In the latter case it is not possible to make a strong statement about accuracy due to the absence of benchmarks. In the former case the accuracy depends on the use of an inner control variate, as shown in the relevant figures.

A comparison with the results of Andersen [1] and Pedersen [61] shows agreement of the price estimates to within a small margin of error. Since no benchmarks are available, we cannot give a definitive report on the accuracy; we do know however that the solutions obtained in [1, 61] are low-biased estimates. Hence it is not clear whether many of our results, which tend to be higher than those of [1, 61], are high-biased or not. It would be useful in this case to have results from a dual method, such as those suggested by Rogers [64] and Haugh and Kogan [37], to give high-biased estimates for comparison.

The timings results are difficult to compare due to differing computing resources. Andersen [1] find low-biased estimates for the value of a swaption involving 16 factors in 20-60 seconds on a DEC Alpha 8400, depending on the accuracy required. This pricing problem would be roughly equivalent to Contracts III and IV in this chapter, which took anywhere from 50-700 seconds depending on the volatility scenario and the number of grid points.



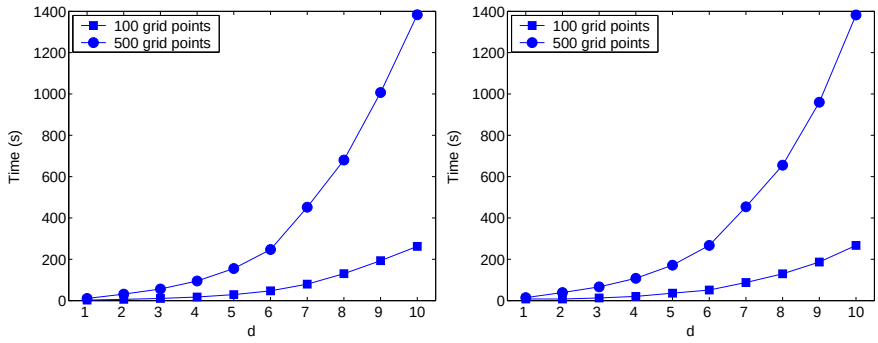
(a) Local quadratic

(b) Local quadratic (cv)

(c) Nearest neighbour

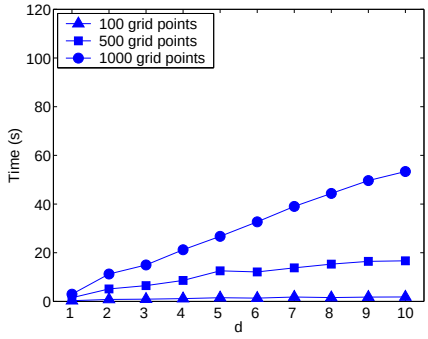
(d) Nearest neighbour (cv)

Figure 5.4.1: Bermudan geometric average option estimates based on standardised Sobol' quadrature operators, two different interpolation operators and normal Sobol' grids. Circles, squares and triangles are slightly displaced on the x-axis for clarity, and represent average results of ten experiments. Error bars give the ranges of the results.

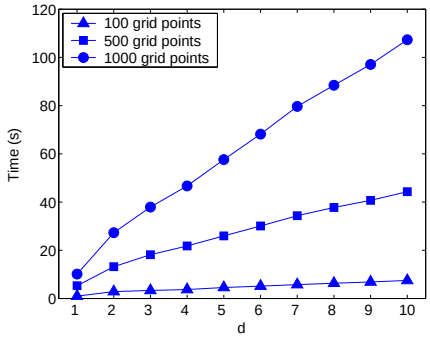


(a) Local quadratic

(b) Local quadratic (cv)

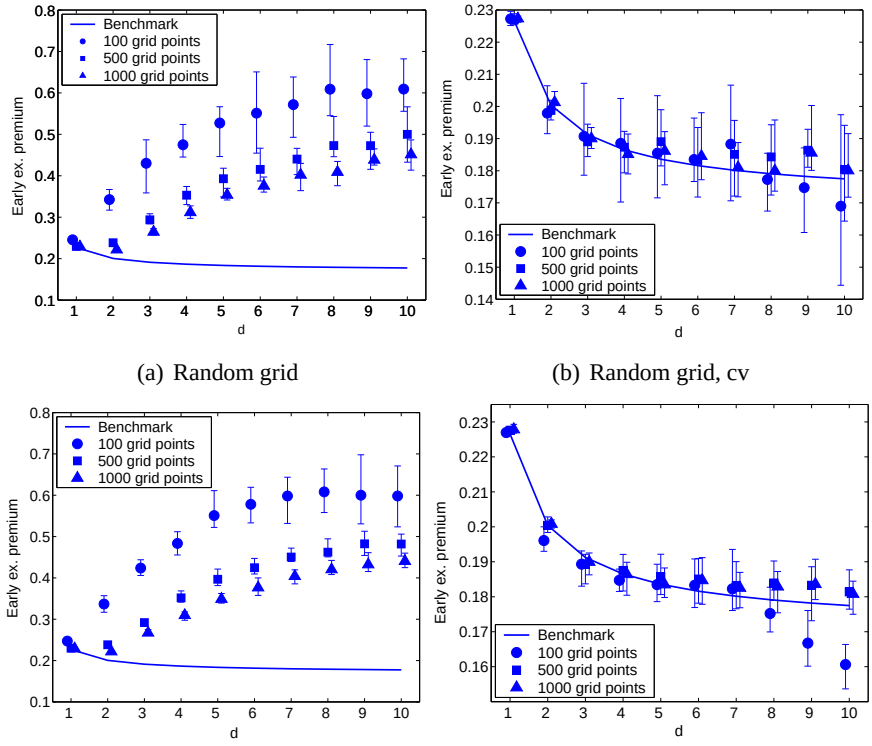


(c) Nearest neighbour



(d) Nearest neighbour (cv)

Figure 5.4.2: Timings for Bermudan geometric average option estimates based on standardised Sobol' quadrature operators, two different interpolation operators and normal Sobol' grids.



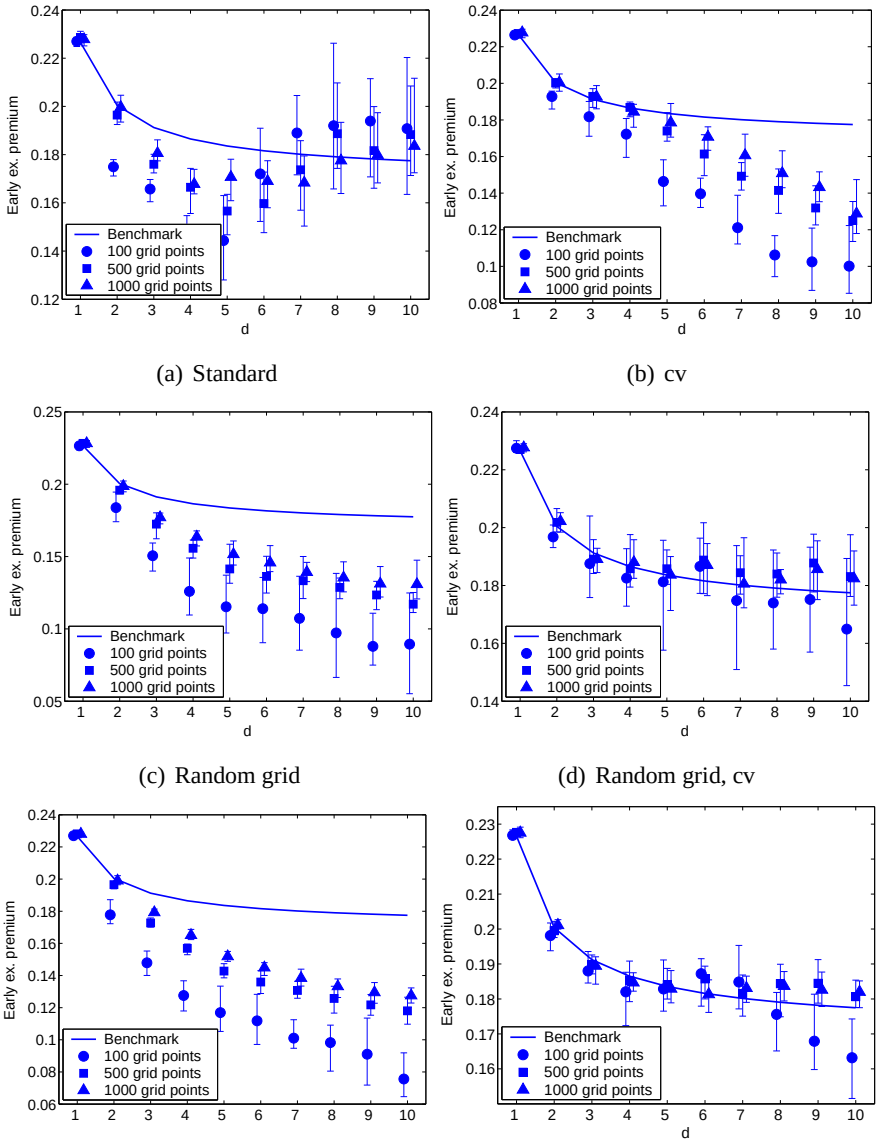
(a) Random grid

(b) Random grid, cv

(c) 3 random grids

(d) 3 random grids, cv

Figure 5.4.3: Bermudan geometric average option estimates based on standardised Sobol' quadrature and nearest neighbour interpolation operators and randomised normal Sobol' grids. Circles, squares and triangles are slightly displaced on the x-axis for clarity, and represent average results of ten experiments. Error bars give the ranges of the results.

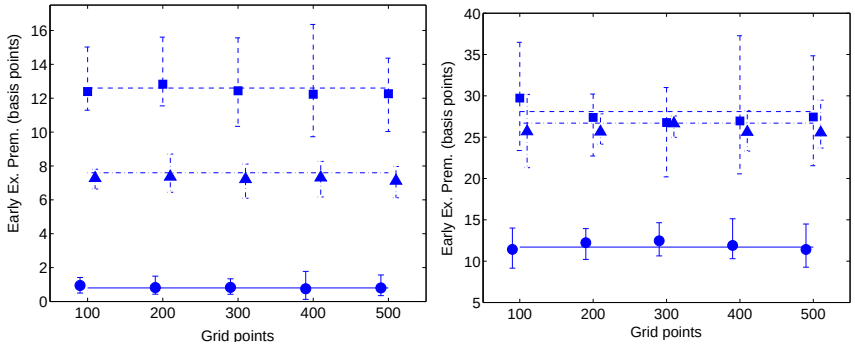


(a) Standard (b) cv

(c) Random grid (d) Random grid, cv

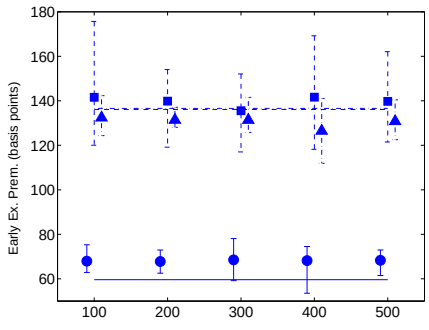
(e) 3 random grids (f) 3 random grids, cv

Figure 5.4.4: Bermudan geometric average option estimates based on standardised Sobol' quadrature and nearest neighbour interpolation operators, randomised normal Sobol' grids and where the integration is applied to the early exercise premium. Circles, squares and triangles are slightly displaced on the x-axis for clarity, and represent average results of ten experiments. Error bars give the ranges of the results.

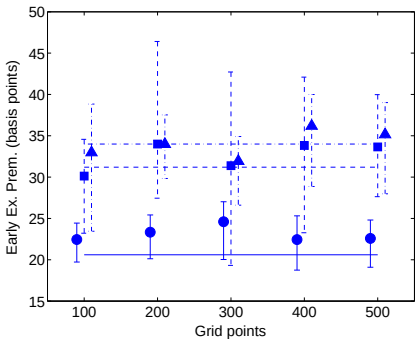


(a) Contract (I)

(b) Contract (II)

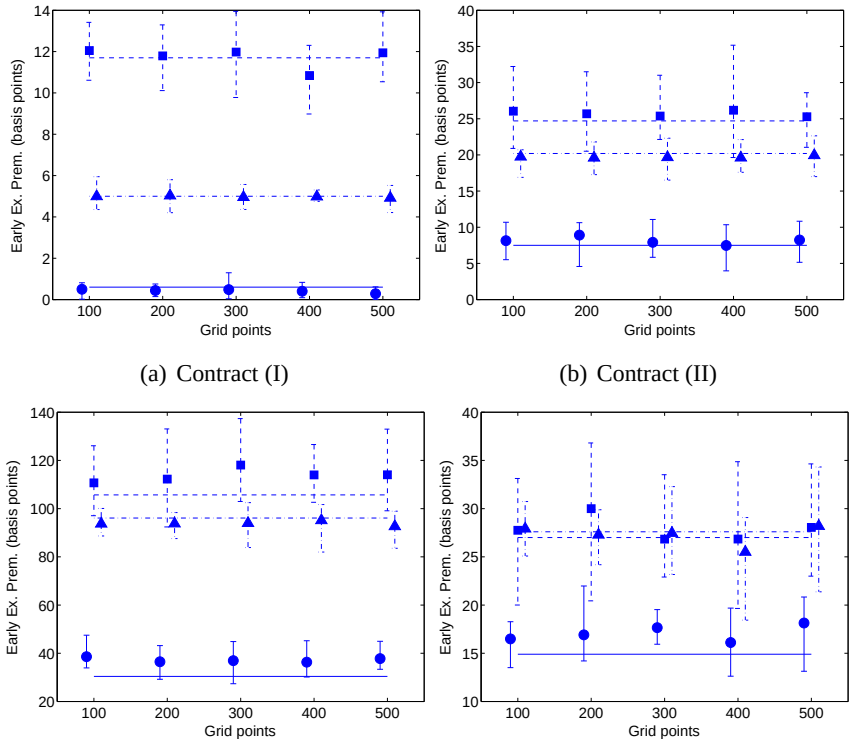


(c) Contract (III)



(d) Contract (IV)

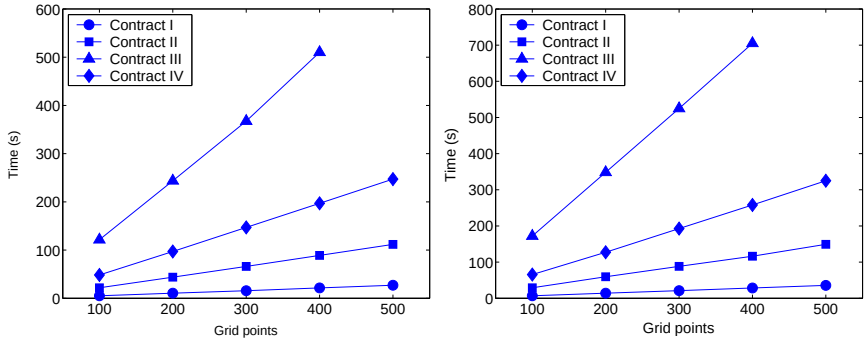
Figure 5.4.5: Swaption estimates for contracts I–IV and volatility scenario C, based on antithetic Monte Carlo quadrature and nearest neighbour interpolation operators and normal Sobol' grids. Circles and solid lines give average results of ten experiments and the benchmark, respectively, for K=8%, squares and dashed lines the same for K=10% and triangles and dash-dot lines the same for K=12%. Error bars give the ranges of the results. All results use control variates.



(a) Contract (I) (b) Contract (II)

(c) Contract (III) (d) Contract (IV)

Figure 5.4.6: Swaption estimates for contracts I–IV and volatility scenario D, based on antithetic Monte Carlo quadrature and nearest neighbour interpolation operators and normal Sobol' grids, and using an inner control variate. Circles and solid lines give average results of ten experiments and the benchmark, respectively, for K=8%, squares and dashed lines the same for K=10% and triangles and dash-dot lines the same for K=12%. Error bars give the ranges of the results. All results use control variates.



(a) Volatility scenario C (b) Volatility scenario D

Figure 5.4.7: Timings for Bermudan swaption estimates based on quadrature and nearest neighbour interpolation operators and normal Sobol' grids. The timing results for contract III with 500 points are omitted due to shortage of memory in both cases.

Chapter 6

Convergence Results

6.1 Introduction

Recent work within the mathematical finance community has seen the development of several algorithms aimed at solving high-dimensional optimal stopping problems. Such algorithms may be used for the pricing and hedging of American- and Bermudan-style derivative securities whose intrinsic value may be based on a large numbers of factors. Important practical examples of such derivatives are Bermudan swaptions, for which the outstanding notional amount worldwide runs into the trillions of euros, and real options, which constitute a growing research area of considerable practical importance.

Algorithms for pricing high-dimensional American options have largely been developed in a path simulation context, starting with the works of Tilley [70] and Barraquand and Martineau [4]. The ensuing methods have included the stochastic tree and stochastic mesh approaches of Broadie and Glasserman [18, 19], in which confidence intervals are constructed using high- and low-biased estimators, and the regression-based methods of Carrière [20], who uses nonparametric techniques to estimate the optimal stopping region, and Tsitsiklis and Van Roy [71] who use regression in the context of value iteration. Longstaff and Schwartz [50] showed that Carrière's method could succeed in a high-dimensional setting using parametric regressions onto a carefully chosen set of basis functions.

The new algorithms proposed in Chapters 2–4 (also published as [8, 7, 9, 10]) constitute a fundamentally different approach in that they approximate the con-

tinuous problem using a quasi-random or low distortion state-space discretisation, followed by time stepping. They do not thus use path simulations as a basis for estimating the continuation value, but instead employ an approximating Markov chain to represent the dynamics in a manner which is amenable to computation. Alternatively, one can see these methods as irregular grid versions of finite difference methods for partial differential equations (PDEs). Chapters 2 and 4 demonstrate the convergent behaviour of the respective algorithms experimentally, the estimates for ten-dimensional problems in the latter agreeing closely with benchmarks using 100,000 points and a control variate. No formal proof of convergence is provided in those chapters.

In a similar setting, Glowinski et al. [35] prove convergence of numerical schemes for variational inequalities. They make a rather strong coercivity assumption which, as noted by Zhang [76, 77], is not satisfied in general by the Black-Scholes operator. The operator rather satisfies the weaker Gårding inequality assumption, which Zhang uses to prove convergence of a numerical scheme for pricing American options on jump-diffusion processes, and based on a localised regular grid in one dimension. Jaillet et al. [42] also appeal to the Glowinski et al. framework to prove convergence of the Brennan-Schwartz algorithm, which amounts to an explicit finite difference method, in one dimension.

The irregular grid scheme differs from other implementations for American option pricing, such as those of Jaillet et al. [42] and Zhang [76], in two ways. First, the irregular grid scheme does not assume, a priori, a specification of the grid structure. This is a useful property in a high-dimensional space because, for a regular grid, the curse of dimensionality implies an exponential increase of the number of grid points with an increasing number of stochastic factors. The ability to use an unstructured grid thus allows specifications for which the computation of an approximate solution is a tractable operation, even if the error related to this solution may still be subject to the curse of dimensionality. The second difference is that no localisation is performed before discretisation; instead, the irregular grid framework considers state space discretisations which become dense in the state space as the discretisation parameter tends to infinity. This means that convergence should depend on only two parameters: the time step and the state space discretisation parameter.

Proof of convergence for the irregular grid scheme is carried out in this work

using the variational inequality framework of Glowinski et al. [35]. We relax the coercivity assumption to a Gårding inequality assumption, as in Zhang [76], but in contrast to the latter we treat the multidimensional case and use a discrete version of Gronwall's inequality to establish stability of the approximate solutions. The discrete Gronwall inequality leads to a better understanding of the stability behaviour through more explicit expressions for the stability bounds.

Finally, we investigate how the convergence assumptions are satisfied in the local consistency approach to irregular grid solutions presented in Chapter 4. Given the inherent complexity of operator approximations on irregular grids, we are unable to offer an analytic investigation of this; indeed, the convergence assumptions are related to the eigenvalues of the operator approximations, which are difficult to analyse. Instead we provide experimental evidence that certain conditions for stability are satisfied.

The remainder of this chapter is arranged as follows. In Section 6.2 we provide formulations of the mathematical problem to be solved. In Section 6.3, we review the local consistency approach to approximating process dynamics on an irregular grid. We then provide a proof of convergence in Section 6.4 and in Section 6.5 we investigate satisfaction of the sufficient conditions for convergence in the case of the local consistency approach. Finally Section 6.6 draws conclusions.

6.2 Formulation

6.2.1 Stopping time formulation

We consider processes of the form

$$dX(t) = \mu(X(t), t)\ dt + \sigma(X(t), t)\ dW(t) \tag{6.2.1}$$

on the time interval $[0, T]$ where the initial value $X(0)$ is known almost surely, and μ and σ are measurable with respect to the natural filtration of $dW(t)$. For example, in option pricing problems we would set μ to be the risk-neutral drift, and σ the instantaneous volatility.

An optimal stopping problem on the process (6.2.1) involves finding the value

$$v(x_0, 0) = \sup_{\tau \in \mathcal{T}} \mathbb{E}^{\mathbb{Q}}_{x_0} \left[\psi(X(\tau), \tau)\right] \tag{6.2.2}$$

where T is the time horizon, $\mathcal{T}$ is the set of stopping times on $[0, T]$ with respect to the natural filtration, the initial value is $X(0) = x_0$ and $\psi(X, \tau)$ is the value achieved by stopping in state X at time τ. The expectation is taken with respect to the process (6.2.1), which is assumed to be the risk neutral process. When applied to problems where discounting is applicable, we assume that this is included in ψ, that is, all values are in time zero euros.

Additionally, one may consider the numerical problem of finding the optimal stopping time itself

$$\hat{\tau} = \operatorname{argsup}_{\tau \in \mathcal{T}} \mathbb{E}^{\mathbb{Q}}_{x_0} \big[\psi(X(\tau), \tau) \big], \tag{6.2.3}$$

which may be useful for example in finding a low-biased estimate of the solution to (6.2.2).

6.2.2 Variational inequality formulation

Demonstrating the convergence of solution schemes for optimal stopping problems is often simplified by appealing to the variational inequality formulation of the problem. The connection between the two problem classes is established in Chapter 3 of Bensoussan and Lions [5], and results concerning the convergence of numerical schemes may be found in Glowinski et al. [35].

In contrast to the optimal stopping formulation for the value function, the variational inequality formulation is usually treated in reverse time, with the option payoff as initial condition. In the following, we work in reverse time to maintain consistency with [35] and with the usual analysis of initial boundary value problems.

We define an operator $\mathcal{A}$ giving the diffusion of the process; the construction of $\mathcal{A}$ from the coefficients of (6.2.1) is given in [5]. For example, in the case of the multidimensional extension of the Black-Scholes model we have

$$\mathcal{A} = -\tfrac{1}{2} \sum_{i,j=1}^{d} \sigma_{ij}(x,t) \frac{\partial^2}{\partial x_i \partial x_j} + \tfrac{1}{2} \sum_{i=1}^{d} \sigma_{ii}(x,t) \frac{\partial}{\partial x_i} \tag{6.2.4}$$

where we assume that the risk-free rate is zero as mentioned above.

As in [42], we introduce the weighted Sobolev spaces

$$W^{m,p,\mu}(\mathbb{R}^d) = \left\{ v \in L^p\left(\mathbb{R}^d, e^{-\mu|x|}\right) : D^\alpha v \in L^p\left(\mathbb{R}^d, e^{-\mu|x|}\right),\ |\alpha| \leq m \right\} \tag{6.2.5}$$

where D^α is the partial derivative corresponding to the multiindex α. We let $H_\mu = W^{0,2,\mu}$ and $V_\mu = W^{1,2,\mu}$; further we denote the H_μ norm and inner product by $|\cdot|_\mu$, $(\cdot,\cdot)_\mu$ and the V_μ norm and inner product by $\|\cdot\|_\mu$, $((\cdot,\cdot))_\mu$. So that

$$|v|_\mu^2 = \int_{\mathbb{R}^d} v(x)^2 e^{-\mu|x|} dx \tag{6.2.6}$$

$$\|v\|_\mu^2 = |v|_\mu^2 + \sum_{|\alpha|=1} |D^\alpha v|_\mu^2 . \tag{6.2.7}$$

The introduction of weighted spaces facilitates the consideration of intrinsic functions ψ which may not be integrable in a non-weighted space. This is the case with call options for example.

The variational inequality equivalent to the optimal stopping problem (6.2.2) is to find $v \in H_\mu$ such that

$$\begin{cases} v(\cdot,T) \equiv \psi(\cdot,T) \\ v(x,s) \geq \psi(x,s) \\ \left(\dfrac{\partial v}{\partial t} - \mathcal{A}v, u - v\right)_\mu \leq 0 \qquad \text{a.e. } \forall u \geq \psi,\ t \in [0,T] \end{cases} \tag{6.2.8}$$

for $(x,s) \in \mathbb{R}^d \times [0,T]$.

6.2.3 General variational inequality formulation

In Section 6.3 we will consider numerical methods for the solution of general parabolic time-dependent inequalities of type I, as defined in Chapter 6 of Glowinski et al. [35]. We follow the setting of [35], except that we assume a Gårding inequality for the bilinear form in place of strict coercivity.

We employ two Hilbert spaces H and V, with respective inner products $(\cdot,\cdot)$ and $((\cdot,\cdot))$, respective norms $|\cdot|$ and $\|\cdot\|$, and where V is a dense subset of H with continuous injection. Typically H will be an L^2 space, and V the Sobolev space of degree one, i.e. where all first order derivatives are in H.

Example 6.2.1 As mentioned in Section 6.2.2, it will often be advantageous to work with weighted spaces so that the obstacle or payoff remains integrable. Hence, for such problems we may use $H = H_\mu$ and $V = V_\mu$ as defined above.

□

From the above continuity requirement we can conclude that, for $v \in V$

$$|v| \leq c\|v\| \tag{6.2.9}$$

for some constant c. We denote by V' the dual of V with respect to H, thus obtaining the Gelfand triple

$$V \hookrightarrow H \hookrightarrow V' \tag{6.2.10}$$

with dense embeddings.

To accommodate the constraints we introduce a convex set $K \subset V$ in which the solution should lie for each t, and we assume the initial condition $v_0 \in K$. We consider a linear operator $\mathcal{A} : V \to V'$ (for example (6.2.4)) and its associated bilinear form $a(v, v) = (\mathcal{A}v, v)$ which satisfies a Gårding inequality

$$a(v, v) + \rho|v|^2 \geq \alpha\|v\|^2 \tag{6.2.11}$$

for some $\alpha > 0, \rho \geq 0$. In the case where $\rho = 0$ and $\alpha > 0$, this becomes a strict coercivity condition. As noted in Zhang [76], the Black-Scholes operator does not in general satisfy a strict coercivity condition.

In order to fully define the problem, we must define the dual norm and introduce some abstract spaces. For a full justification of the following steps the reader is referred to Glowinski et al. [35]. We first introduce a space for the solutions of the time-dependent problem. V' is first equipped with the norm

$$\|w\|_* = \sup_{v \in V, \|v\|=1} (w, v) \tag{6.2.12}$$

and we define the Hilbert space and subspace

$$W([0,T]) = \left\{ v \in L^2([0,T]; V) : \frac{dv}{dt} \in L^2([0,T]; V') \right\} \tag{6.2.13}$$

$$W_0([0,T]) = \left\{ v \in W([0,T]) : v(0) = v_0 \in H \right\} \tag{6.2.14}$$

where $L^2([0,T];V)$ denotes the space of functions $t \to f(t)$ which are measurable from $[0,T] \to V$ such that

$$\|f\|_{L^2([0,T];V)} = \left(\int_0^T \|f(t)\|_V^2 dt \right)^{1/2} < +\infty. \tag{6.2.15}$$

Finally we introduce the convex spaces

$$\mathcal{K} = \left\{ v \in W([0,T]) : v(t) \in K \text{ a.e. } [0,T] \right\} \tag{6.2.16}$$

$$\mathcal{K}_0 = \left\{ v \in W_0([0,T]) : v(t) \in K \text{ a.e. } [0,T] \right\}. \tag{6.2.17}$$

The problem to be solved is then to find $v \in \mathcal{K}_0$ such that

$$\int_0^T \left(\frac{\partial v}{\partial t} + \mathcal{A}v - f, u - v \right) dt \geq 0 \qquad \forall u \in \mathcal{K} \tag{6.2.18}$$

for some appropriate L^2 function f (the zero function in our case), or equivalently

$$\left(\frac{\partial v}{\partial t} + \mathcal{A}v - f, u - v \right) \geq 0 \qquad \forall u \in K, \text{ a.e. } [0,T]. \tag{6.2.19}$$

Remark 6.2.1 The problem formulation introduced accommodates a large class of parabolic variational problems, of which optimal stopping problems are just one example.

Example 6.2.2 We recall the American option pricing problem from Example 6.2.1, which is an initial value problem in reverse time in the current setting. We let

$$K = K_\mu = \{ v \in V_\mu : v \geq \psi \} \subset V_\mu. \tag{6.2.20}$$

In this case we set $v_0 = \psi \in K_\mu$.

Assuming the covariance matrix $\Sigma = (\sigma_{ij})$ is symmetric positive definite, we can set $\alpha \in (0, \frac{1}{4}\lambda)$ in (6.2.11) where λ is the smallest eigenvalue of Σ. This can

be seen as follows, using the operator defined in (6.2.4):

$$\begin{aligned}
a(v,v) &= (\mathcal{A}v, v) \\
&= -\tfrac{1}{2}\int \left(\sum_{i,j} \sigma_{ij} \frac{\partial^2 v}{\partial x_i \partial x_j} v - \sum_i \sigma_{ii} \frac{\partial v}{\partial x_i} v \right) e^{-\mu|x|} dx \\
&= \tfrac{1}{2}\int \left(\sum_{i,j} \sigma_{ij} \frac{\partial v}{\partial x_i}\frac{\partial v}{\partial x_j} + \sum_i \left(\sigma_{ii} - \mu \sum_j \frac{x_j}{|x|}\sigma_{ij} \right) \frac{\partial v}{\partial x_i} v \right) e^{-\mu|x|} dx \\
&\geq \tfrac{1}{2}\int \left((\nabla v)' \Sigma (\nabla v) + \sum_i \left(\sigma_{ii} - \mu \sum_j \sigma_{ij} \right) \left| \frac{\partial v}{\partial x_i} \right| |v| \right) e^{-\mu|x|} dx \\
&\geq 2\alpha \sum_i \left| \frac{\partial v}{\partial x_i} \right|_\mu^2 - \tfrac{1}{2} \sum_i \left| \sigma_{ii} - \mu \sum_j \sigma_{ij} \right| \left| \frac{\partial v}{\partial x_i} \right|_\mu |v|_\mu .
\end{aligned}$$

Now

$$\begin{aligned}
&a(v,v) + \rho |v|_\mu^2 \\
&\geq 2\alpha \sum_i \left| \frac{\partial v}{\partial x_i} \right|_\mu^2 - \tfrac{1}{2} \sum_i \left| \sigma_{ii} - \mu \sum_j \sigma_{ij} \right| \left| \frac{\partial v}{\partial x_i} \right|_\mu |v|_\mu + \rho |v|_\mu^2 \\
&= \alpha \|v\|_\mu^2 + \sum_i \left(\alpha \left| \frac{\partial v}{\partial x_i} \right|_\mu^2 + \frac{\rho - \alpha}{d} |v|_\mu^2 - \tfrac{1}{2} \left| \sigma_{ii} - \mu \sum_j \sigma_{ij} \right| \left| \frac{\partial v}{\partial x_i} \right|_\mu |v|_\mu \right) .
\end{aligned}$$

It is now possible to choose ρ large enough such that each term in the sum is positive, thus showing that the bilinear form satisfies a Gårding inequality.

□

6.3 Solution framework

We now propose a solution framework for solving variational inequalities in a general Hilbert space setting. Numerical schemes for solving the variational inequality (6.2.8) can be developed in this setting, for which relevant examples will be presented. The setting follows that of Glowinski et al. [35], except that we do not assume strict coercivity of the diffusion operator or its discrete counterpart.

The framework involves the following steps:

1. discretisation of space and time,
2. approximation of constraints and operators in the discretised setting,
3. numerical solution of the discretised problem.

In this section we will often reuse a constant c for various different bounds and continuity constants. Convergence results will be presented in Section 6.4.

6.3.1 Discretisation of space and time

We consider separate discretisations of space and time. Such discretisations are convenient for solving initial value problems, since time stepping methods allow division of the problem into a sequence of smaller problems. Having a constant grid in the state space also allows implicit solutions to be more easily considered.

For the space discretisation, we first introduce the auxiliary Hilbert space Φ to be used in relating V to its discretisation; Φ in effect is intended to contain the extra information required to know whether for some $v \in H$ we also have $v \in V$. In particular, when $H = W^{0,2,\mu}$ and $V = W^{1,2,\mu}$ as defined in (6.2.5), Φ would correspond to first order derivatives. We now denote

$$F = H \times \Phi \tag{6.3.1}$$

and the corresponding extension isomorphism

$$\sigma : V \to F \tag{6.3.2}$$

which we will always take to be the identity plus first derivative information

$$\sigma v = \left(v, \frac{\partial v}{\partial x_1}, \ldots, \frac{\partial v}{\partial x_d} \right). \tag{6.3.3}$$

Consider now finite dimensional Hilbert spaces H_n and V_n approximating H and V respectively where n denotes the dimension of the spaces. The properties required for these approximations are now studied. We denote the respective norms of H_n and V_n by $|\cdot|_n$ and $\|\cdot\|_n$, the respective inner products by $(\cdot,\cdot)_n$ and $((\cdot,\cdot))_n$, and introduce linear extension operators

$$q_n : V_n \to H \tag{6.3.4}$$

$$p_n : V_n \to F \tag{6.3.5}$$

relating elements of the discrete Hilbert spaces to those of the continuous spaces, where we assume for $v \in V_n$

$$|q_n v| = |v|_n \tag{6.3.6}$$

$$\|p_n v\| \leq c\|v\|_n \tag{6.3.7}$$

where c does not depend on n. These operators are related in that $p_n v = (q_n v, \ldots)$ and p_n is assumed to be convergent in that for each $v \in V$, there exists a bounded sequence $v_n \in V_n$ such that

$$p_n v_n \to \sigma v \qquad \text{strongly in } F. \tag{6.3.8}$$

The following relationships are assumed between the norms in the finite dimensional spaces for some constant c and (increasing) function $s(n)$:

$$|v|_n \leq c\|v\|_n \tag{6.3.9}$$

$$\|v\|_n \leq s(n)|v|_n. \tag{6.3.10}$$

Note that the first inequality is the discrete version of (6.2.9). The function $s(n)$, which in general increases without bound as $n \to \infty$, will appear in stability and convergence conditions; we will see in particular that specifying $s(n)$ to be as small as possible leads to more favourable estimates.

Example 6.3.1 Consider a generic grid, or set of states

$$\mathcal{X}_n = \{x_1, \ldots, x_n\} \tag{6.3.11}$$

where $\mathcal{X}_n$ becomes dense in $\mathbb{R}^d$ as $n \to \infty$. For example, one may consider generating random or quasi-random states from $\mathbb{R}^d$ according to an appropriate density. A regular grid may also be used, although in this case either the problem must first be localised or the span of the grid must increase while the resolution becomes simultaneously finer.

We now proceed to build an approximating Hilbert space based on indicator functions of the Voronoi cells of $\mathcal{X}_n$. Define the ith Voronoi cell of a grid $\mathcal{X}$ as the set

$$C_{\mathcal{X}}(i) = \left\{ x \in \mathbb{R}^d : |x - x_i| = \min_j |x - x_j| \right\} \tag{6.3.12}$$

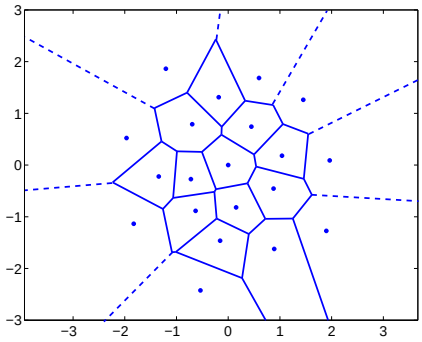


Figure 6.3.1: Voronoi cells of a grid $\mathcal{X}$ in dimension 2 containing 20 points. Distances are with respect to the Euclidean norm and dashed lines continue to infinity.

and the Voronoi index of $x \in \mathbb{R}^d$ with respect to $\mathcal{X}$ as the positive integer

$$J_{\mathcal{X}}(x) = \operatorname{argmin}_j |x - x_j|. \tag{6.3.13}$$

Further, define the (possibly multivalued) k-nearest neighbour mapping for $k \leq n$ as

$$N_{\mathcal{X},k}(x) = \Big\{ j_1, \ldots, j_k : j_1 = \operatorname{argmin}_j |x - x_j|,\ j_2 = \operatorname{argmin}_{j \neq j_1} |x - x_j|, \\ \ldots,\ j_k = \operatorname{argmin}_{j \notin \{j_1, \ldots, j_{k-1}\}} |x - x_j| \Big\}, \tag{6.3.14}$$

giving the ordered indices of the k nearest neighbours of x in $\mathcal{X}$. Figure 6.3.1 provides an example of the Voronoi cells for a grid $\mathcal{X}$ containing 20 points.

We now introduce the Hilbert spaces

$$H_n = V_n = \mathbb{R}^{\mathcal{X}} \\ \cong \left\{ v \in L^2 : v(x) = \sum_{i=1}^{n} c_i I_{C_{\mathcal{X}}(i)}(x),\ c_i \in \mathbb{R} \right\} \tag{6.3.15}$$

with norms to be specified, and the isomorphically equivalent space given will form the image of the extension operators q_n and p_n. We see that this image is not contained in V; the extension operator p_n will thus constitute an exterior approximation as noted in [35]. On the other hand, an interior approximation may result for example from using an extension operator related to a finite element discretisation.

Let us now define the extension operator q_n in the obvious manner for $v \in H_n$

$$[q_n v](x) = v_{J_{\mathcal{X}}(x)}. \tag{6.3.16}$$

This leads us through condition (6.3.6) to the norm definition for H_n

$$\begin{aligned} |v|_n^2 &= \int [q_n v](x)^2 e^{-\mu|x|} dx \\ &= \sum_{i=1}^{n} \theta_i v_i^2 \end{aligned} \tag{6.3.17}$$

where

$$\theta_i = \int_{C_{\mathcal{X}}(i)} e^{-\mu|x|} dx. \tag{6.3.18}$$

The space H_n is thus a weighted discrete L^2 space where the weights are decreasing with the norm of x_i, but increasing with the size of the Voronoi cell containing x_i.

In this setting the operator p_n should give information about the derivatives of $v \in V_n$. We set

$$[p_n v](x) = ([q_n v](x),\ [\delta_1 v](x),\ \ldots,\ [\delta_d v](x)) \tag{6.3.19}$$

where the discrete derivative operators δ_i are defined at $x \in \mathbb{R}^d$ through using first order Taylor series expansions at the $d+1$ nearest neighbours of x as follows: let $(j_1, \ldots, j_{d+1})$ be the (ordered) $d+1$ nearest neighbours of x_i, we then solve the equations

$$\begin{aligned} v(x_{j_1}) &= v(x_i) + \sum_{j=1}^{d} (x_{j_1} - x_i) \frac{\partial v}{\partial x_j}(x_i) \\ &\vdots \\ v(x_{j_{d+1}}) &= v(x_i) + \sum_{j=1}^{d} (x_{j_{d+1}} - x_i) \frac{\partial v}{\partial x_j}(x_i) \end{aligned}$$

for the $\frac{\partial v}{\partial x_j}$, which leads us to define

$$[\delta_i v](x) = \left(v_{j_1} \cdots v_{j_{d+1}}\right) A_x^{-1} e_{i+1} \tag{6.3.20}$$

where

$$A_x = \begin{pmatrix} 1 & \cdots & 1 \\ (x_{j_1} - x) & \cdots & (x_{j_{d+1}} - x) \end{pmatrix} \tag{6.3.21}$$

and e_i is the ith canonical unit vector. The norm in V_n is defined analogously:

$$\|v\|_n^2 = |v|_n^2 + \sum_{i=1}^{d} |\delta_i v|_n^2. \tag{6.3.22}$$

This allows us to set $c = 1$ in (6.3.7) and (6.3.9). To find a suitable bound $s(n)$ in (6.3.10), we consider the worst-case effect of the matrices A_x in (6.3.20). We propose the bound

$$s(n) = 1 + d\lambda(n)^{-2}\tilde{\eta} \tag{6.3.23}$$

where $\tilde{\eta}$ is the greater of the maximum number of neighbours per point and the maximum number of times a point is a neighbour (with respect to the derivative estimates), and $\lambda(n)$ is the minimum λ_j where, for each j, λ_j^{-2} is the smallest eigenvalue of $A_j A_j'$ where $A_j \equiv A_{x_j}$ as defined above. We demonstrate the validity of the bound (6.3.23) as follows:

$$\begin{aligned} |\delta_i v|_n^2 &= \sum_{j=1}^{n} \left(\left(v_{j_1} \cdots v_{j_{d+1}} \right) A_j^{-1} e_{i+1} \right)^2 \theta_j \\ &= \sum_{j=1}^{n} \left(\sum_{k=1}^{d+1} v_{j_k} e_k' A_j^{-1} e_{i+1} \right)^2 \theta_j \\ &\leq \sum_{j=1}^{n} \sum_{k=1}^{d+1} v_{j_k}^2 \lambda_j^{-2} \theta_j \\ &\leq \lambda^{-2} \tilde{\eta} |v|_n^2. \end{aligned}$$

In the third line we have used the following matrix inequalities

$$\begin{aligned} \left| e_k A_j^{-1} e_{i+1} \right| &\leq \|A_j^{-1}\|_\Delta \\ &\leq \|A_j^{-1}\|_2 \\ &= \left(\text{max eigenvalue of } \left(\left(A_j^{-1} \right)' A_j^{-1} \right) \right)^{1/2} \\ &= \left(\text{min eigenvalue of } (A_j' A_j) \right)^{-1/2} \\ &=: \lambda_j^{-1} \end{aligned}$$

where we use the matrix norm $\|A\|_\Delta = \max_{i,j} |a_{ij}|$. The inequality $\|A\|_\Delta \leq \|A\|_2$ can be seen as follows:

$$\begin{aligned}\|A\|_\Delta &= |a_{ij}| \qquad\qquad \text{for some } i,j \\ &\leq \left(\sum_k a_{kj}^2\right)^{1/2} = \frac{\|Ae_j\|_2}{\|e_j\|_2} \leq \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2.\end{aligned}$$

Note that one could use other matrix norms than $\|A_j^{-1}\|_2$ in the estimation; however these other norms are somewhat more difficult to interpret in terms of the properties of A_j itself.

Finally

$$\begin{aligned}\|v\|_n^2 &= |v|_n^2 + \sum_{i=1}^d |\delta_i v|_n^2 \\ &\leq |v|_n^2 + d\lambda(n)^{-2}\tilde{\eta}|v|_n^2 \\ &= \left(1 + d\lambda(n)^{-2}\tilde{\eta}\right)|v|_n^2. \end{aligned} \tag{6.3.24}$$

We thus see that the factor $s(n) = \left(1 + d\lambda(n)^{-2}\tilde{\eta}\right)^{1/2}$ bounding the V_n norm involves the dimension d, the maximum number of neighbours $\tilde{\eta}$ used for the derivative estimations and the term $\lambda(n)$ which is a function of the neighbour distances and their conditioning.

For example, for a regular grid in one dimension $\mathcal{X} = \{x_1, x_1 + \delta x, \ldots, x_n\}$, and using the δ_i as defined in (6.3.20), we have for each $j = 2, \ldots, n-1$

$$A_j = \begin{pmatrix} 1 & 1 \\ 0 & \pm\delta x \end{pmatrix}. \tag{6.3.25}$$

Assuming $\delta x < 1$, we have $\lambda_j^2 \in \rho(A_j' A_j) = 1 + \frac{1}{2}\delta x^2 \pm \left(1 + \frac{1}{4}h^4\right)^{1/2} \simeq \{\frac{1}{2}\delta x^2, 2\}$. Alternatively, for the δ_i proposed in [35] and [76] we have

$$A_j = \begin{pmatrix} 1 & 1 \\ -\frac{1}{2}\delta x & \frac{1}{2}\delta x \end{pmatrix}, \tag{6.3.26}$$

in which case $\lambda_j^2 \in \rho(A_j' A_j) = \{\frac{1}{2}\delta x^2, 2\}$ (the same as above up to $O(\delta x^4)$), and $\lambda(n)^{-2} = 2/\delta x^2$. Noting that in this case $\tilde{\eta} = 2$, we have

$$s(n) = \left(1 + \frac{4}{\delta x^2}\right)^{1/2} \tag{6.3.27}$$

as found in [76]. Note that $s(n)$ depends on n through $\delta x = (x_n - x_1)/(n-1)$.

□

We assume that time is discretised using K steps

$$t = t_0 < \cdots < t_K = T \tag{6.3.28}$$

and denote the time step sizes by $\delta t_k = t_{k+1} - t_k$ for $k = 0, \ldots, K-1$. As $K \to \infty$, we require

$$\max_k \delta t_k \to 0. \tag{6.3.29}$$

Given approximate solutions $v_k^{(n)}(x)$ corresponding to an n-dimensional space discretisation and $t = t_k$, $k = 0, \ldots, K$, we consider the (timewise) piecewise constant extension

$$v^{(n,\delta t)}(x,t) = \sum_{k=0}^{K} v_k^{(n)}(x) \mathbb{1}(\{k = \operatorname{argmin}_j |t - t_j|\}). \tag{6.3.30}$$

Example 6.3.2 The regularly spaced discretisation

$$t_k = \frac{k}{K} T \qquad k = 0, \ldots, K \tag{6.3.31}$$

is often the simplest to work with, although it cannot be expected that this discretisation leads to an optimal convergence rate in general. For example, the time step size in certain numerical solution algorithms for PDEs is determined adaptively depending on the shape of the computed solutions.

□

6.3.2 Approximation of constraints and operators

We now discuss approximation of the constraints and operators in the discrete setting introduced in Section 6.3.1. For operators we refer to the partial differential operators $\frac{\partial}{\partial t}$ and $\mathcal{A}$, and for constraints we refer to the convex spaces K in (6.2.19).

Let us first consider approximation of the constraints. The requirement $v \in K$ for each time t is approximated by the requirement

$$v_k^{(n)} \in K_n \subseteq V_n \tag{6.3.32}$$

for each time index $k = 1, \ldots, K$ where K_n is closed and convex. The K_n are assumed to be consistent in that, if for $v_n \in K_n$, $p_n v_n \to \xi \in F$ weakly then $\xi \in \sigma K$. We assume further that K_n are convergent in that for all $v \in K$ there exists a bounded sequence $v_n \in K_n$ such that $p_n v_n \to \sigma v \in F$ strongly.

Note that the extension of K_n to H, based on the extension operators p_n and q_n, is not in general required to be a subset of K, or even to be contained in V itself. In particular this cannot be expected in the case of exterior approximations (see page 135).

Example 6.3.3 In the case of American options, following from Example 6.3.1, a simple specification of K_n is

$$K_n = \{v \in V_n : v \geq \psi_n\} \tag{6.3.33}$$

where $\psi_n(x) = \psi(x_{\mathcal{J}_{\mathcal{X}_n}(x)})$ and $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ is the nth grid.

□

We now consider the operator $\frac{\partial}{\partial t}$, for which we specify a standard first order approximation for t_k, $k = 0, \ldots, K-1$:

$$\frac{\partial v}{\partial t}(x, t_k) \simeq \frac{1}{\delta t_k}\left(v^{(n,k)}(x, t_{k+1}) - v^{(n,k)}(x, t_k)\right). \tag{6.3.34}$$

This definition may be extended to $t \in [0, T]$ if required; however we shall only require approximations at times t_k.

We finally consider an approximation A_n to the operator $\mathcal{A}$ in the discretisation introduced above and the corresponding bilinear form $a_n(u, v) = (A_n u, v)_n$. As in the continuous problem, the approximation is assumed to satisfy a Gårding inequality

$$a_n(v, v) + \rho |v|_n^2 \geq \alpha \|v\|_n^2 \tag{6.3.35}$$

for some constants $\alpha > 0$, $\rho \geq 0$ not depending on n. These constants are not in general the same as those found in the continuous Gårding inequality formulation (6.2.11).

The bilinear form is also assumed to be continuous in that

$$|a_n(u, v)| \leq c \|u\|_n \|v\|_n \tag{6.3.36}$$

for all $u, v \in V_n$ where $c > 0$ is a constant independent of n. Further, the bilinear form is assumed to converge in the following sense: suppose $p_n v_n \to \sigma v$ weakly in F and $p_n w_n \to \sigma w$ strongly in F, then $a_n(v_n, w_n) \to a(v, w)$, $a_n(w_n, v_n) \to a(w, v)$ and $\liminf_n a_n(v_n, v_n) \geq a(v, v)$.

Example 6.3.4 For American option pricing problems, which constitute variational inequalities defined on an unbounded domain, most authors perform a localisation before discretising. The localised domain is assumed to be large enough to guarantee a suitably small truncation error. For a regular rectilinear grid on a localised domain, a sequence A_n satisfying the above conditions can be constructed using standard finite difference methods, as shown in Glowinski et al. [35] and Zhang [76]. In particular, no weighting is needed in the case of a localised domain to make a call payoff integrable.

For multidimensional problems, such a rectilinear domain suffers from the curse of dimensionality, and may not be appropriate from importance sampling considerations. For these reasons, irregular grid methods have been suggested in Chapters 2–4 (also published as [8, 7, 9, 10]) for the solution of multidimensional problems. Two stable constructions of A_n are suggested in those chapters, namely the method of Chapter 2 (also [8, 7, 9]) involving the logarithm or root of a transition matrix corresponding to importance sampling weights and the method of Chapter 4 (also [10]) involving the solution of a large number of small linear programming problems. Given the unbounded domain and the fact that these methods do not involve localisation, one must consider weighted norms in the discretisation in order to allow non-L^2 payoffs, as suggested in Example 6.3.1.

We thus form the weighting matrix $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n)$ where the weights θ_i are given in (6.3.18), and consider the weighted bilinear form

$$a_n(u, v) = (A_n u, v)_n = v' \Theta A_n u. \tag{6.3.37}$$

In Section 6.5 we will investigate how the conditions on a_n are satisfied for the local consistency method as proposed in Chapter 4.

□

Finally, we assume that f is continuous and approximated in a (timewise) step-

wise constant manner

$$f^{(n,k)}(x,t) = \sum_{k=0}^{K} f_k^{(n)}(x) \mathbb{1}(\{k = \operatorname{argmin}_j |t - t_j|\}) \tag{6.3.38}$$

where $f_k^{(n)}(x) = f(t_k, x)$. We are now in a position to form the system of equations whose solution approximates that of (6.2.19):

$$\left(\frac{v_n^{i+1} - v_n^i}{k} + A_n v_n^{i+\theta}, u - v_n^{i+1} \right)_n \geq 0 \quad v_n^{i+1} \in K_n, \ \forall u \in K_n \tag{6.3.39}$$

for $i = K-1, \ldots, 0$ where $v_n^0 = \psi_n$ and $v_n^{i+\theta} = (1-\theta)v_n^i + \theta v_n^{i+1}$ is the θ-weighted value approximation for the ith iteration. In the case where $K_n = \{v \in V_n : v \geq \psi\}$ for some obstacle ψ, the discrete variational inequality (6.3.39) may be reformulated as a sequence of linear complementarity problems

$$0 \leq v^{i-1} - \psi^{i-1} \perp M_L v^{i-1} - M_R v^i \geq 0 \tag{6.3.40}$$

for $i = K-1, \ldots, 0$ where $v_n^0 = \psi_n$, $M_L = I - A_n \theta \delta t$ and $M_R = I + A_n(1-\theta)\delta t$.

6.4 Convergence of the discretised problems

Having discussed the functional setting of the discretised problem and proposed conditions for convergence in Section 6.3, we now demonstrate the convergence of the discretised solutions (6.3.39) to the true solutions (6.2.19).

In Section 6.4.1 we establish conditions under which the solution to the discretised system (6.3.39) is stable using an M-matrix assumption. Using the Gårding inequality assumption we prove in Section 6.4.2 the stability of the solutions along with some related quantities. We then place this in the framework of Glowinski et al. to achieve a theorem of convergence.

For simplicity we now assume a constant time step, although the following can be easily modified to accommodate a variable time step.

6.4.1 Stability under M-matrix assumption

We now provide conditions under which the solutions of the discretised variational inequalities (6.3.39) are stable. This is done through the complementarity representation of the variational inequalities.

We remind the reader of some matrix classes which will be used in the following analysis. For a full treatment of matrix classes we refer the reader to Berman and Plemmons [6] or Cottle, Pang and Stone [23].

Definition 6.4.1 *A real square matrix is said to be a Z-matrix if its off-diagonal entries are nonpositive.*

Definition 6.4.2 *A real square matrix is said to be an M-matrix if it is a Z-matrix with nonnegative diagonal entries.*

Definition 6.4.3 *A real square matrix is said to be a P-matrix if all of its principal minors are positive.*

The explicit case

The complementarity problems are

$$0 \leq v^{i-1} - \psi^{i-1} \perp v^{i-1} - Mv^i \geq 0 \tag{6.4.1}$$

for $i = K, \ldots, 1$. The payoff functions ψ^i are given and are allowed to vary over time. The matrix M is formed from the infinitesimal generator as follows

$$M = (1 - r\delta t)I + A\delta t \tag{6.4.2}$$

and has the property of being a P-matrix and having row sums $1 - r\delta t$. In the explicit case the complementarity problems reduce to the simpler problem

$$v^{i-1} = \max\left(\psi^{i-1}, Mv^i\right) \tag{6.4.3}$$

where $\max$ gives the pointwise maximum.

Lemma 6.4.1 *Suppose that A is an M-matrix with zero row sums and $r \geq 0$. Then under the stability condition*

$$\delta t \leq \frac{1}{\|A - rI\|_\Delta} \tag{6.4.4}$$

where $\|A\|_\Delta = max_{i,j}|a_{ij}|$, the solution at time index $i = 0$ of the explicit system of complementarity problems (6.4.1) satisfies

$$\|v^0\|_\infty \leq \max_{i=0,\ldots,K}(1 - r\delta t)^i \|\psi^i\|_\infty. \tag{6.4.5}$$

Proof. From (6.4.3) we have

$$\begin{aligned}\|v^{i-1}\|_\infty &\le \max\left(\|\psi^{i-1}\|_\infty, \|Mv^i\|_\infty\right)\\ &\le \max\left(\|\psi^{i-1}\|_\infty, \|M\|_\infty\|v^i\|_\infty\right)\\ &= \max\left(\|\psi^{i-1}\|_\infty, (1-r\delta t)\|v^i\|_\infty\right).\end{aligned}$$

Applying this inequality recursively with $v^K \equiv \psi^K$ gives the required result.

■

The stability bound on δt is presented in terms of the maximum absolute entry of the matrix M. We now investigate this bound and its asymptotic properties.

Lemma 6.4.2 *Suppose that A is an M-matrix with zero row sums and $r \ge 0$, and that A is locally consistent for a process having covariance matrix $\Sigma = (\sigma_{ij})$ (denote $\sigma_i^2 \equiv \sigma_{ii}$). Suppose further that the Euclidean distances between connected points in the grid $\mathcal{X}$ lie in the interval $[\varepsilon_1, \varepsilon_2]$. Then the norm appearing in Lemma 6.4.1 is bounded by*

$$\frac{1}{\varepsilon_2^2}\sum \sigma_i^2 + r \le \|A - rI\|_\Delta \le \frac{1}{\varepsilon_1^2}\sum \sigma_i^2 + r. \tag{6.4.6}$$

Proof. From the feasibility conditions we have

$$\begin{aligned} a_{ij_1}\delta x_{ij_i1}^2 + \cdots + a_{ij_\eta}\delta x_{ij_\eta 1}^2 &= \sigma_1^2\\ &\vdots\\ a_{ij_1}\delta x_{ij_id}^2 + \cdots + a_{ij_\eta}\delta x_{ij_\eta d}^2 &= \sigma_d^2.\end{aligned}$$

The hypothesis regarding the distance between connected points implies that, for each δx_{ij_k},

$$\varepsilon_1^2 \le \delta x_{ij_k1}^2 + \cdots + \delta x_{ij_kd}^2 \le \varepsilon_2^2.$$

Summing the above equations and applying the inequality gives

$$\varepsilon_1^2 \sum_k a_{ij_k} \le \sum \sigma_i^2 \le \varepsilon_2^2 \sum_k a_{ij_k}. \tag{6.4.7}$$

Now the off-diagonal entries of $A - rI$ are exactly the a_{ij_k} in (6.4.7), which are nonnegative and satisfy the bounds

$$\frac{1}{\varepsilon_2^2}\sum \sigma_i^2 \le |a_{ij_k}| \le \frac{1}{\varepsilon_1^2}\sum \sigma_i^2.$$

The diagonal entries of $A - rI$ are $-\sum a_{ij_k} - r$, which are equal in absolute value to $\sum a_{ij_k} + r$, provided $r \geq 0$, hence they satisfy the bounds

$$\frac{1}{\varepsilon_2^2}\sum \sigma_i^2 + r \leq \left| -\sum a_{ij_k} - r \right| \leq \frac{1}{\varepsilon_1^2}\sum \sigma_i^2 + r.$$

Putting these bounds together completes the proof.

■

Lemmas 6.4.2 and 6.4.1 are now combined to give a stability condition in terms of the ratio between the time step and the minimum point separation ε_1.

Lemma 6.4.3 *Under the conditions of Lemma 6.4.2, the stability condition in Lemma 6.4.1 holds provided*

$$\frac{\delta t}{\varepsilon_1^2} \leq \frac{1 - r\delta t}{\sum \sigma_i^2}. \tag{6.4.8}$$

It is clear that $\delta t/\varepsilon^2$ must be less than $1/\sum \sigma_i^2$ to guarantee stability as $\delta t \to 0$.

The fully implicit case

The stability condition in the explicit case can be rather restrictive, especially in a low dimension where the point separation decreases more rapidly with grid size. Since implicit methods often exhibit greater stability, we now investigate their properties.

The complementarity problems are now

$$0 \leq v^{i-1} - \psi^{i-1} \perp Mv^{i-1} - v^i \geq 0 \tag{6.4.9}$$

for $i = K, \ldots, 1$. Again the payoff functions ψ^i are given and are allowed to vary over time. The matrix M is now given by

$$M = (1 + r\delta t)I - A\delta t. \tag{6.4.10}$$

Lemma 6.4.4 *Suppose that A is an M-matrix with zero row sums and $r \geq 0$. Then the solution at time index $i = 0$ of the system of implicit complementarity problems (6.4.9) satisfies*

$$\|v^0\|_\infty \leq \max_{i=0,\ldots,K} (1 + r\delta t)^i \|\psi^i\|_\infty. \tag{6.4.11}$$

Proof. Consider the vector

$$v = \max\left((1+r\delta t)^{-1}\|v^i\|, \|\psi^{i-1}\|_\infty\right)\mathbb{1}. \tag{6.4.12}$$

We claim that v is feasible for (6.4.9), though it does not necessarily satisfy the complementarity conditions. For the first inequality in (6.4.9) this is obvious. For the second we have, recalling that A has zero row sums,

$$\begin{aligned} Mv - v^i &= (1+r\delta t)v - v^i \\ &= \max\left(\|v^i\|, (1+r\delta t)\|\psi^{i-1}\|_\infty\right)\mathbb{1} - v^i \\ &\geq 0. \end{aligned}$$

The solution v being feasible but not necessarily satisfying the complementarity conditions implies, according to Theorem 3.11.6 in [23] and noting that an M-matrix is also a Z-matrix, that v dominates (pointwise) the true solution v^{i-1}. Thus

$$\begin{aligned} \|v^{i-1}\|_\infty &\leq \|v\|_\infty \\ &= \max\left((1+r\delta t)^{-1}\|v^i\|, \|\psi^{i-1}\|_\infty\right). \end{aligned}$$

Applying this inequality recursively with $v^K \equiv \psi^K$ gives the desired result.

■

Remark 6.4.1 The fully implicit method is thus unconditionally stable, unlike the explicit method where a stability condition was imposed in Lemma 6.4.3. We shall see shortly that θ case time stepping methods admit a similar stability condition to the explicit case. The implicit method is thus the only first order time stepping scheme we can prove to be unconditionally stable, as also found in Chapter 6 of Glowinski et al. [35].

The θ case

The θ case lies "between" the explicit and implicit problems. The complementarity problems are

$$0 \leq v^{i-1} - \psi^{i-1} \perp M_L v^{i-1} - M_R v^i \geq 0 \tag{6.4.13}$$

for $i = K, \ldots, 1$. Again the payoff functions ψ^i are given and are allowed to vary over time. The matrices M_L and M_R are given by

$$M_L = (1 + r\theta\delta t)I - A\theta\delta t \tag{6.4.14}$$

$$M_R = (1 + r(1-\theta)\delta t)I + A(1-\theta)\delta t. \tag{6.4.15}$$

It is simple to extend the stability conditions presented in the explicit case to the θ case. The conditions may be weaker depending on the value of θ.

Lemma 6.4.5 *Suppose A is an M-matrix with zero row sums and $r \geq 0$. Then under the stability condition*

$$\delta t \leq \frac{1}{(1-\theta)\|A - rI\|_\Delta} \tag{6.4.16}$$

where $\|A\|_\Delta = max_{i,j}|a_{ij}|$, the solution at time index $i = 0$ of the θ-case system of complementarity problems (6.4.13) satisfies

$$\|v^0\|_\infty \leq \max_{i=0,\ldots,K} \left(\frac{1 - r(1-\theta)\delta t}{1 + r\theta\delta t} \right)^i \|\psi^i\|_\infty. \tag{6.4.17}$$

Lemma 6.4.6 *Under the conditions of Lemma 6.4.2, the stability condition in Lemma 6.4.5 holds provided*

$$\frac{\delta t}{\varepsilon^2} \leq \frac{1 - r(1-\theta)\delta t}{(1-\theta)\sum \sigma_i^2}. \tag{6.4.18}$$

We require that $\delta t/\varepsilon^2$ must be less than $2/(1-\theta)\sum \sigma_i^2$ to guarantee stability as $\delta t \to 0$, which is indeed weaker than the condition for stability in the explicit case.

6.4.2 Stability and convergence under Gårding inequality assumption

Following the framework proposed in Glowinski et al. [35], we investigate convergence of the discretised variational inequalities. We refer to [35] for a detailed treatment of the numerical solution to variational inequalities in an abstract setting.

In addition to stability, the results presented below will establish convergence of irregular grid schemes. The stability result relates not only to the computed solution, but also to the partial derivatives of its extension under p_n. This is a stronger

result than that of Section 6.4.1 where only stability of the value function was established. The convergence result relies crucially on the satisfaction of a Gårding inequality in the discretised setting. Previously Glowinski et al. [35] provided the same result under a strict coercivity assumption (a Gårding inequality with $\rho = 0$) and Zhang [76] proved stability of a localised regular grid discretisation method for pricing American options on a one-dimensional jump-diffusion process, and under a Gårding inequality assumption. Jaillet et al. [42] prove convergence of the Brennan-Schwartz algorithm (the explicit finite difference method in one dimension) using the framework of [35]. Matache et al. investigate convergence of a wavelet discretisation method for pricing European options [52] and American options [51] on single assets following one-dimensional Lévy processes.

All convergence results mentioned for American option pricing problems consider only one-dimensional numerical schemes. Although these results may well be generalised to higher dimensions, their generalisations would involve regular grid constructions involving intractable numbers of grid points. Our investigation of convergence for irregular grid schemes provides an extension to higher dimensions which is tractable computationally. Note that the approximation error may still be subject to the curse of dimensionality.

Our proof of stability mainly follows the arguments of Glowinski et al. [35] and Zhang [76], except that, in contrast to the former we assume a Gårding inequality instead of strict coercivity, and in contrast to the latter we do not concentrate on the one-dimensional case, but remain in a multidimensional setting. We also use a discrete version of Gronwall's lemma rather than the integral form. This allows us to quantify the form of the error more precisely.

Theorem 6.4.1 *Let u_n^k be the solution to the discretised variational inequality (6.3.39) at iteration k corresponding to $t = k/K = k\delta t$. Assume that the discrete Gårding inequality (6.3.35) holds for some $\alpha, \rho > 0$, then the quantities*

$$v_n^k, \quad p_n v_n^k, \quad \sum_{i=1}^{k} |v_n^{i+1} - v_n^i|_n^2 \tag{6.4.19}$$

are uniformly bounded for $0 \leq k\delta t \leq T$ with n and δt satisfying the stability assumption

$$1 - 2\delta t(1-\theta)\frac{c^2 s(n)^2}{\alpha} \geq \beta > 0 \tag{6.4.20}$$

for some β and where c is the continuity constant appearing in (6.3.36).

Remark 6.4.2 Note that the stability condition is similar to that of Lemma 6.4.6. In one dimension, recalling from (6.3.27) that $s(n) \simeq 1/\delta x^2$, it is clear that each of these conditions bounds the size of $\delta t/\delta x^2$ and moreover the bounds both become infinite as the implicitness parameter $\theta \to 1$.

Before proving Theorem 6.4.1, we introduce the following two lemmas. The first establishes a discrete Gronwall inequality, and the second a limit for the estimate implied by the Gronwall inequality.

Lemma 6.4.7 *Suppose a_i are nonnegative quantities for $i \in \mathbb{N}$ and*

$$a_k \leq d_0 + d_1 \sum_{i=0}^{k-1} a_i + d_2 \sum_{i=0}^{k-1} b_i, \tag{6.4.21}$$

holds for all $k \in \mathbb{N}$ where b_i,d_i are nonnegative constants. Then the inequality

$$a_k \leq d_0(1+d_1)^{k-1} + d_1(1+d_1)^{k-1}a_0 + d_1 \sum_{i=0}^{k-1} (1+d_1)^{k-i-1} b_i \tag{6.4.22}$$

holds for all $k \in \mathbb{N}$.

Proof. We proceed to prove the inequality (6.4.22) by induction. The inequality (6.4.22) certainly holds for $k = 1$. Assume now that it holds for all $k \leq m$ for

some $m > 1$. We have then for a_{m+1}

$$\begin{aligned}
a_{m+1} &\leq d_0 + d_1 \sum_{i=0}^{m} a_i + d_2 \sum_{i=0}^{m} b_i \\
&= d_0 + d_1 a_0 + d_2 \sum_{i=0}^{m} b_i \\
&\quad + d_1 \sum_{i=1}^{m} \left(d_0 (1+d_1)^{i-1} + d_1 (1+d_1)^{i-1} a_0 + d_2 \sum_{j=0}^{i-1} (1+d_1)^{i-j-1} b_j \right) \\
&= d_0 \left(1 + d_1 \sum_{i=1}^{m} (1+d_1)^{i-1} \right) + d_1 \left(1 + d_1 \sum_{i=1}^{m} (1+d_1)^{i-1} \right) a_0 \\
&\quad + d_2 \left(b_0 + \sum_{i=1}^{m} \left[b_i + d_1 \sum_{j=0}^{i-1} (1+d_1)^{i-j-1} b_j \right] \right) \\
&= d_0 (1+d_1)^m + d_1 (1+d_1)^m a_0 + d_2 \sum_{i=0}^{m} (1+d_1)^{m-i} b_i
\end{aligned}$$

where the last equality can be shown using standard summation formulae.

■

Lemma 6.4.8 *Assume that $b_i = b(i\delta t)$ for some function $b(\cdot)$ and $\delta t > 0$, and write $a_{\delta t}(t) = a_{[t/\delta t]}$ where $[\cdot]$ is the integer part and the a_k satisfy the inequality*

$$a_k \leq c_0 + c_1 \delta t \sum_{i=0}^{k} a_i + c_2 \delta t \sum_{i=0}^{k-1} b_i \tag{6.4.23}$$

for all $n \in \mathbb{N}$. Then as $\delta t \to 0$, $k \to \infty$ and $t = k\delta t$ remaining fixed, we have

$$\limsup_{k,\delta t} a_{\delta t}(t) \leq c_0 e^{c_1 t} + c_2 \int_0^t e^{c_1 (t-s)} b(s) ds. \tag{6.4.24}$$

Proof. We first reformulate (6.4.23) as

$$a_k \leq \frac{c_0}{1 - c_1 \delta t} + \frac{c_1 \delta t}{1 - c_1 \delta t} \sum_{i=0}^{k-1} a_i + \frac{c_2 \delta t}{1 - c_1 \delta t} \sum_{i=0}^{k-1} b_i. \tag{6.4.25}$$

Using the notation of Lemma 6.4.7, we set

$$d_0 = \frac{c_0}{1 - c_1\delta t}, \quad d_1 = \frac{c_1\delta t}{1 - c_1\delta t}, \quad d_2 = \frac{c_2\delta t}{1 - c_1\delta t}$$

and we note that as $\delta t \to 0$

$$\begin{aligned} d_1 &\to 0 \\ 1 + d_1 &\to 1 \\ (1 + d_1)^k = \left(1 - \frac{c_1 t}{k}\right)^{-k} &\to e^{c_1 t} \\ \sum_{i=0}^{k} \delta t (1 + d_1)^{k-i} b_i &\to \int_0^t e^{c_1(t-s)} b(s) ds \end{aligned}$$

where the limits are taken as $\delta t \to 0$, $k \to \infty$ and holding $t = k\delta t$ constant. Applying Lemma 6.4.7, and taking the limit on the RHS of (6.4.22) gives

$$\begin{aligned} d_0(1 + d_1)^{k-1} + d_1(1 + d_1)^{k-1} a_0 + d_1 \sum_{i=0}^{k} (1 + d_1)^{k-i} b_i \\ \longrightarrow\ c_0 e^{c_1 t} + c_2 \int_0^t e^{c_1(t-s)} b(s) ds \end{aligned}$$

as required.

■

Proof of Theorem 6.4.1. We first recall some identities:

$$v^{i+\theta} = \theta v^{i+1} + (1 - \theta) v^i \tag{6.4.26}$$

$$2(a - b, a) = |a|^2 - |b|^2 + |a - b|^2 \tag{6.4.27}$$

$$2(a, b) \le 2|a||b| \le \frac{1}{\varepsilon}|a|^2 + \varepsilon|b|^2, \ \ \varepsilon > 0. \tag{6.4.28}$$

We will also make use of the discrete Gårding inequality (6.3.35) and the continuity assumption (6.3.36).

Starting from the discretised variational inequality (6.3.39), and noting that $\psi_n \in K_n$, we have

$$\begin{aligned} \frac{1}{\delta t}(v_n^{i+1} - v_n^i, \psi_n - v_n^{i+1})_n + a_n(v_n^{i+\theta}, \psi_n - v_n^{i+1}) \\ \ge (f_n^{i+\theta}, \psi_n - v_n^{i+1})_n. \end{aligned} \tag{6.4.29}$$

Now using (6.4.27) with $a = v_n^{i+1} - \psi_n$, $b = v_n^i - \psi_n$, we have

$$(v_n^{i+1} - v_n^i, \psi_n - v_n^{i+1})_n = |v_n^{i+1} - \psi_n|_n^2 - |v_n^i - \psi_n|_n^2 + |v_n^{i+1} - v_n^i|_n^2.$$

Using the last equality, the Gårding inequality (6.3.35) and writing

$$a_n(v_n^{i+\theta}, v_n^{i+1}) = \theta a_n(v_n^{i+1}, v_n^{i+1}) + (1-\theta)a_n(v_n^i, v_n^i) + (1-\theta)a_n(v_n^i, v_n^{i+1} - v_n^i)$$

we find that

$$\begin{aligned}
&\frac{1}{2\delta t}\left(|v_n^{i+1} - \psi_n|_n^2 - |v_n^i - \psi_n|_n^2 + |v_n^{i+1} - v_n^i|_n^2\right) \\
&\qquad\qquad + \alpha\left(\theta\|v_n^{i+1}\|_n^2 + (1-\theta)\|v_n^i\|_n^2\right) \\
&\le (f_n^{i+\theta}, \psi_n - v_n^{i+1})_n + a_n(v_n^{i+\theta}, \psi_n) - (1-\theta)a_n(v_n^i, v_n^{i+1} - v_n^i) \\
&\qquad\qquad + \rho\theta|v_n^{i+1}|_n^2 + \rho(1-\theta)|v_n^i|_n^2. \quad (6.4.30)
\end{aligned}$$

To estimate the first three terms on the RHS, we note that

$$|(f_n^{i+\theta}, \psi_n - v_n^{i+1})_n| \le |f_n^{i+\theta}|_n^2 + \tfrac{1}{2}|v_n^i|_n^2 + \tfrac{1}{2}|\psi_n|_n^2,$$

$$\begin{aligned}
|a_n(v_n^{i+\theta}, \psi_n)| &\le \theta|a_n(v_n^{i+1}, \psi_n)| + (1-\theta)|a_n(v_n^i, \psi_n)|_n \\
&\le c\theta\|v_n^{i+1}\|_n\|\psi_n\|_n + c(1-\theta)\|v_n^i\|_n\|\psi_n\|_n \\
&\le \tfrac{1}{2}\theta\alpha\|v_n^{i+1}\|_n^2 + \tfrac{1}{4}(1-\theta)\alpha\|v_n^i\|_n^2 + (1-\tfrac{1}{2}\theta)\frac{c^2}{\alpha}\|\psi_n\|_n^2,
\end{aligned}$$

$$\begin{aligned}
|a_n(v_n^i, v_n^{i+1} - v_n^i)| &\le c\|v_n^i\|_n\|v_n^{i+1} - v_n^i\|_n^2 \\
&\le \tfrac{1}{4}\alpha\|v_n^i\|_n^2 + \frac{c^2}{\alpha}\|v_n^{i+1} - v_n^i\|_n^2 \\
&\le \tfrac{1}{4}\alpha\|v_n^i\|_n^2 + \frac{c^2 s(n)^2}{\alpha}|v_n^{i+1} - v_n^i|_n^2
\end{aligned}$$

where the second and third estimates use relation (6.4.28) variously with $\varepsilon = \frac{c}{\alpha}$ and $\varepsilon = \frac{2c}{\alpha}$. Incorporating these estimates in (6.4.30) we have

$$\begin{aligned}
&|v_n^{i+1} - \psi_n|_n^2 - |v_n^i - \psi_n|_n^2 + (1-\beta)|v_n^{i+1} - v_n^i|_n^2 \\
&\qquad\qquad + \alpha\delta t\left(\theta\|v_n^{i+1}\|_n^2 + (1-\theta)\|v_n^i\|_n^2\right) \\
&\le \delta t\Big(2|f_n^{i+\theta}|_n^2 + |\psi_n|_n^2 + (2-\theta)\frac{c^2}{\alpha}\|\psi_n\|_n^2 \\
&\qquad\qquad + (2\rho\theta + 1)|v_n^{i+1}|_n^2 + 2\rho(1-\theta)|v_n^i|_n^2\Big) \quad (6.4.31)
\end{aligned}$$

where we have used β from the stability assumption (6.4.20).

Now summing (6.4.31) from $i = 0, \ldots, k-1$ and noting $v_n^0 = \psi_n$, we have

$$
\begin{aligned}
&|v_n^k - \psi_n|_n^2 + \beta \sum_{i=0}^{k-1} |v_n^{i+1} - v_n^i|_n^2 + \alpha\delta t \sum_{i=0}^{k-1} \left(\theta \|v_n^{i+1}\|_n^2 + (1-\theta)\|v_n^i\|_n^2\right) \\
&\leq 2\delta t \Bigg(\sum_{i=0}^{k-1} |f_n^{i+\theta}|_n^2 + (\tfrac{1}{2} + \rho\theta) \sum_{i=0}^{k-1} |v_n^{i+1}|_n^2 + \rho(1-\theta) \sum_{i=0}^{k-1} |v_n^i|_n^2 \\
&\qquad\qquad + k(1 - \tfrac{1}{2}\theta) \frac{c^2}{\alpha} \|\psi_n\|_n^2 + \tfrac{1}{2} k |\psi_n|_n^2 \Bigg). \quad (6.4.32)
\end{aligned}
$$

Using the relation $\frac{1}{2}|a|^2 - |b|^2 \leq |a-b|^2$ for the first term on the LHS, collecting the $|v_n^i|$ terms and rearranging we have

$$
\begin{aligned}
&|v_n^k|_n^2 + 2\beta \sum_{i=0}^{k-1} |v_n^{i+1} - v_n^i|_n^2 + 2\alpha\delta t \sum_{i=0}^{k-1} \left(\theta \|v_n^{i+1}\|_n^2 + (1-\theta)\|v_n^i\|_n^2\right) \\
&\leq 2(1 + k\delta t)|\psi_n|_n^2 + 4\delta t \sum_{i=0}^{k-1} |f_n^{i+\theta}|_n^2 + 2\delta t(1+2\rho) \sum_{i=0}^{k} |v_n^i|_n^2 \\
&\qquad\qquad + 2k\delta t(2-\theta) \frac{c^2}{\alpha} \|\psi_n\|_n^2. \quad (6.4.33)
\end{aligned}
$$

Now we use Lemma 6.4.8 with the first term on the LHS. Letting

$$
\begin{aligned}
c_0 &= \sup_n 2|\psi_n|_n^2 \\
c_1 &= 2(1+2\rho) \\
c_2 &= 1 \\
b(t) &= \sup_n 4|f(t)|_n^2 + 2|\psi_n|_n^2 + 2(2-\theta)\frac{c^2}{\alpha}\|\psi_n\|_n^2
\end{aligned}
$$

we have

$$
\limsup |v_n^k|_n^2 \leq c_0 e^{c_1 t} + c_2 \int_0^t e^{c_1(t-s)} b(s) ds,
$$

and therefore boundedness of v_n^k in the $|\cdot|_n$ norm for $0 \leq k\delta t \leq T$ and for n and δt satisfying the stability condition (6.4.20). This implies a uniform stability estimate

$$
|v_n^k|_n^2 \leq c_0 e^{c_1 t} + c_2 \int_0^t e^{c_1(t-s)} b(s) ds + c_3
$$

for some $c_3 > 0$.

We now note that this leads to a uniform asymptotic bound on the RHS of (6.4.33), and thus the other terms on the LHS of (6.4.33) are also uniformly bounded.

■

6.4.3 Convergence

Theorem 6.4.2 *Suppose that (6.2.19) has a unique solution, the stability condition (6.4.20) holds and that $v^{(n,\delta t)}$ are the solutions to the discretised systems (6.3.39). Then*

$$p_n v^{(n,\delta t)} \to \sigma v \tag{6.4.34}$$

as $n \to \infty$, $\delta t \to 0$ provided that $(1-\theta)\delta t s(n)^2 \to 0$.

Proof. We refer the reader to Chapter 6, Section 3 of Glowinski et al. [35] for the details, with the stability of the relevant quantities being ensured through Theorem 6.4.1.

■

6.5 Approximating the bilinear form

In Example 6.3.4 we suggested using the methods presented in Chapter 4 for approximation of the bilinear form on an irregular grid. We now investigate properties of these methods and in particular the manner in which the conditions presented in Section 6.3.2 are satisfied for these approximations.

6.5.1 Local consistency method

The local consistency method presented in Chapter 4 uses considerations proposed in Kushner and Dupuis [46] to construct the operator approximations A_n. Namely, to construct A_n on a grid $\mathcal{X} = \{x_1, \ldots, x_n\}$ one solves for each row i of A_n,

relating to point x_i, the feasibility problem

$$\sum_{j \neq i} (x_j - x_i)(x_j - x_i)' a_{i,j} = \Sigma \tag{6.5.1}$$

$$\sum_{j \neq i} (x_j - x_i) a_{i,j} = \mu \tag{6.5.2}$$

$$a_{i,j} \geq 0 \tag{6.5.3}$$

for the closest possible set of neighbouring points[1]. One then sets $a_{i,i} = -\sum_{j \neq i}^{n} a_{i,j}$ in order to maintain zero row sums in A_n.

One may arrive at the conditions (6.5.1) through considering the Taylor series expansions of the value function

$$\begin{aligned} v(x_i + \delta x_0) &= v(x_i) + \sum_{j=1}^{d} \delta x_{0,j} \frac{\partial v}{\partial x_j}(x_i) + \frac{1}{2} \sum_{j,k=1}^{d} \delta x_{0,j} \delta x_{0,k} \frac{\partial^2 v}{\partial x_j \partial x_k}(x_i) \\ &\quad + O\left(\sum_{j,k,l=1}^{d} |\delta x_{0,j} \delta x_{0,k} \delta x_{0,l}| \right) \\ &\vdots \\ v(x_i + \delta x_\eta) &= v(x_i) + \sum_{j=1}^{d} \delta x_{\eta,j} \frac{\partial v}{\partial x_j}(x_i) + \frac{1}{2} \sum_{j,k=1}^{d} \delta x_{\eta,j} \delta x_{\eta,k} \frac{\partial^2 v}{\partial x_j \partial x_k}(x_i) \\ &\quad + O\left(\sum_{j,k,l=1}^{d} |\delta x_{\eta,j} \delta x_{\eta,k} \delta x_{\eta,l}| \right) \end{aligned}$$

around a feasible x_i and for neighbouring points $x_1, \ldots, x_\eta$ and where $\delta x_{k,j}$ is the jth component of $x_k - x_i$ and $\eta = d(d+3)/2$ is the number of nonzero entries in each feasible row of A_n. We assume here that $\delta x_0 = 0$, so that the closest neighbour of x is the point x itself and the first row above is reduced to a trivial statement.

Using the weighted norms of Example 6.3.4, the discrete Gårding inequality is

[1]One may consider many measures of the "closeness" of a set of points; in Chapter 4, linear weightings were considered, and linear programming was used to find the weights.

equivalent to the matrix

$$\rho I - \Theta^{-1/2} \left(\tfrac{1}{2}(\Theta A + A'\Theta) + \alpha\Theta + \alpha \sum_{i=1}^{d} \delta_i' \Theta \delta_i \right) \Theta^{-1/2} \tag{6.5.4}$$

having nonnegative eigenvalues. One may use for example the first derivative approximations suggested in (6.3.20) for the δ_i. It suffices to check that the matrix

$$\Theta^{-1/2} \left(\tfrac{1}{2}(\Theta A + A'\Theta) + \alpha \sum_{i=1}^{d} \delta_i' \Theta \delta_i \right) \Theta^{-1/2} \tag{6.5.5}$$

has eigenvalues bounded above, as $n \to \infty$ and for some $\alpha > 0$. The Gårding inequality is then satisfied where ρ is sufficiently large such that $\rho - \alpha$ is at least equal to this eigenvalue bound.

6.5.2 Experiments

Due to the difficulty in analysing the maximum eigenvalues of (6.5.5) analytically, we consider experimental evidence for their behaviour. We also conduct experiments to directly investigate stability of the quantities on the LHS of (6.4.33).

The eigenvalue investigation is conducted in dimensions 1–8, and involves computing the maximum eigenvalue of (6.5.5) numerically using the Matlab sparse matrix analysis routines. The experiments involved the following steps:

1. constructing the grid $\mathcal{X}$,
2. constructing the infinitesimal generator A and first derivative matrices δ_i,
3. constructing the weights matrix Θ,
4. computing the maximum eigenvalue.

This procedure was repeated for various grids, and for various specifications of the parameters α and μ. In all cases the dynamics involve zero drift and unit covariance per unit time.

The weights matrix Θ is calculated according to (6.3.18), where the entries are estimated using quasi-Monte Carlo integration as follows. Let f be an appropriate importance sampling density (we used the standard normal density in d dimensions), and let X_j, $j = 1, \ldots, N$, be quasi-Monte Carlo points corresponding to f.

Now let N_i be the number of sample points in the Voronoi cell containing the grid point x_i, and denote by $X_{i,j}$, $j = 1, \ldots, N_i$, the jth point in this cell. Then

$$\begin{aligned}\theta_i &= \int_{C_{\mathcal{X}}(i)} e^{-\mu|x|} dx \\ &= \int_{C_{\mathcal{X}}(i)} \frac{e^{-\mu|x|}}{f(x)} f(x) dx \\ &\simeq \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{e^{-\mu|X_{i,j}|}}{f(X_{i,j})} \frac{N_i}{N} \\ &= \frac{1}{N} \sum_{j=1}^{N_i} \frac{e^{-\mu|X_{i,j}|}}{f(X_{i,j})}.\end{aligned}$$

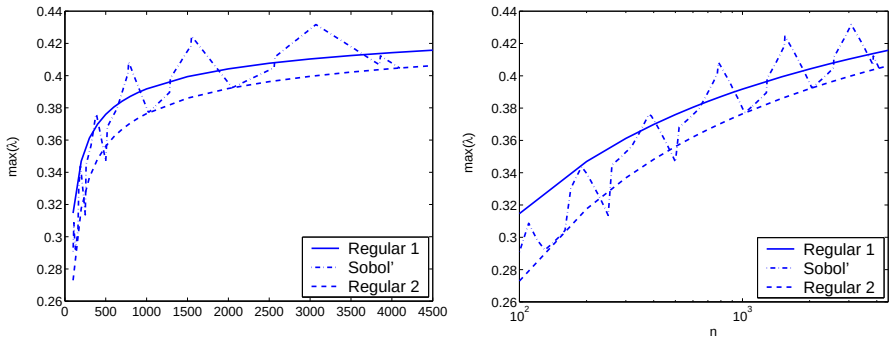
In this way, the θ_i may all be calculated from a single quasi-Monte Carlo sampling.

We found that the parameter μ made little difference to the qualitative behaviour of the eigenvalues, although the results were generally shifted vertically. It was also found that $\alpha = 0.01$ was sufficiently small to allow a stable behaviour of the eigenvalues in most cases.

The results are plotted in Figures 6.5.1–6.5.10. Since the main question involves the boundedness of the eigenvalues with grid size, the maximum eigenvalue is plotted against the grid size.

In Figure 6.5.1 results are shown for grids constructed from inverse normal transformations of regular and Sobol' grids in one dimension. One sees that the maximum eigenvalues of the inverse normal regular grids show a monotone sub-log behaviour, while the normal Sobol' grids show an oscillating behaviour which seems to increase at a slightly faster rate than the inverse normal regular grid, at least on average.

Two types of inverse normal regular grid are considered, where the regular grids are denoted type 1 and type 2. A regular grid of type 1 with n points contains a point at the center of each interval $[k/n, (k+1)/n]$ for $k = 0, \ldots, n-1$, thus containing the points $1/2n, 3/2n, \ldots, (2n-1)/2n$. A regular grid of type 1 with n points contains the points $1/(n+1), \ldots, n/(n+1)$. It should be noted that the Sobol' grid coincides with the regular grid type 2 for grid sizes $n = 2^m - 1$ where m is an integer. The Sobol' grid never coincides with the regular grid of type 1 for grid sizes $n > 1$.



(a) Natural scale (b) Log scale

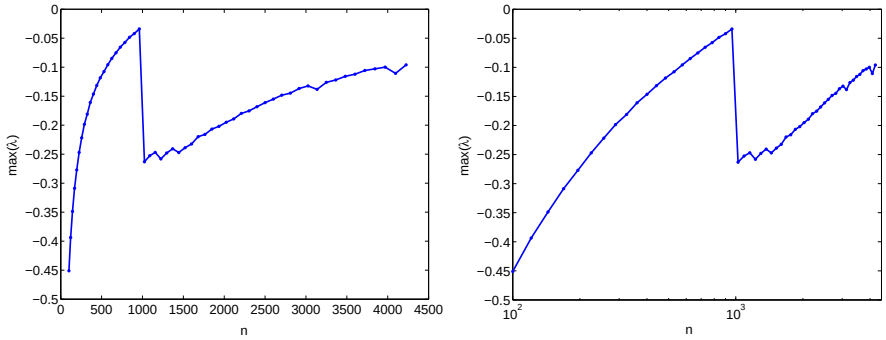
Figure 6.5.1: Maximum eigenvalues of (6.5.5) for inverse normal regular and Sobol' grids in one dimension plotted against grid size n and for $\mu = 2$, $\alpha = 0.01$. The regular grid type 1 of size n contains points $1/2n, 3/2n, \ldots, (2n-1)/2n$, and the regular grid type 2 of size n contains points $1/(n+1), \ldots, n/(n+1)$.

In two dimensions we present results for inverse normal regular grids in Figure 6.5.2 and normal Sobol' and low distortion grids in Figure 6.5.3. The maximum eigenvalue for the inverse normal regular grids shows a monotone sub-log behaviour to start with, similar to that of the one-dimensional inverse normal regular grids. At about $n = 1000$ however, the maximum eigenvalue drops and starts to behave in a slightly less regular fashion. This behaviour is puzzling, and probably has to do with the configuration of neighbours used for the infinitesimal generator.

For dimensions 2–8 we consider the use of normal Sobol' and low distortion grids. Sobol' grids are constructed using Sobol' sequences, which are examples of low discrepancy sequences as described by Niederreiter [58]. Low distortion grids are a separate class of point sets which aim to minimise a distortion function related to the relevant density, in this case the standard normal. These point sets are usually generated using stochastic descent algorithms, as described by Pagès [60]; we used 10^5 iterations and a step size of $k^{-0.4}$ to generate the grids (where k is the iteration number), along with some multiple sampling and quasi-Monte Carlo adaptations.

The generator matrix is constructed as detailed above, where we attempt to satisfy the local consistency conditions using 5η nearest neighbours.

The results for normal Sobol' and low distortion grids in 2–8 dimensions are



(a) Natural scale (b) Log scale

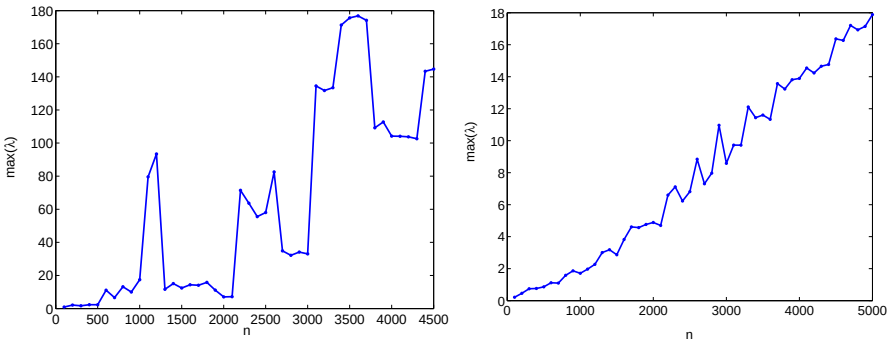
Figure 6.5.2: Maximum eigenvalues of (6.5.5) for inverse normal regular grids in two dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.

presented in Figures 6.5.3–6.5.9. The low distortion plots, also shown together in Figure 6.5.10, show a largely stable monotone and convex behaviour, with the exception of the two-dimensional case which does not offer evidence of stability. This may be due to the stochastic nature of the grid generation, which may be less efficient at smoothing the points in two dimensions.

We also observe less regular behaviour in higher dimensions, which may be due to the number of points being very low compared to the dimension. This has the effect that the local consistency conditions may not be satisfied for a large number of points in the grid, leading to greater asymmetry and a larger number of zero entries in the generator matrix.

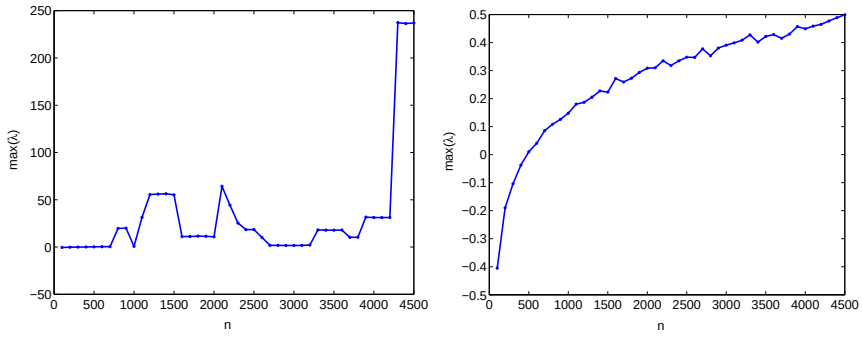
The results for normal Sobol' grids in Figures 6.5.3–6.5.9 are not conclusive regarding stability. We tend to see a stable minimum trend with occasional large deviations which may persist for a number of successive grids. The latter behaviour is not unexpected for the normal Sobol' grids since they are related in that each grid is a superset of the preceding ones. We conjecture that the large eigenvalues are a result of local roughness in the grids which persists as a result of this relationship.

In addition to the eigenvalue investigations we now provide a direct investigation into the behaviour of the stability quantities on the LHS of (6.4.33). The experiments conducted involved pricing a geometric average option on one, two and five assets. The parameters used were identical to those used in Chapter 4, and we used $\mu = 2$ in the weighting matrix. The stability quantities investigated were



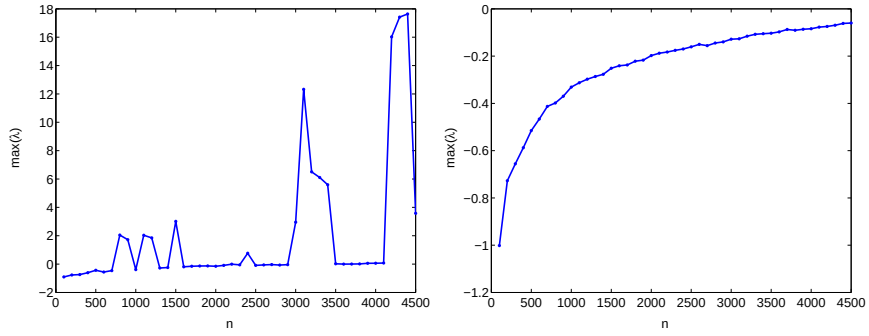
(a) Sobol' (b) Low distortion

Figure 6.5.3: Maximum eigenvalues of (6.5.5) for normal Sobol' and low distortion grids in two dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.



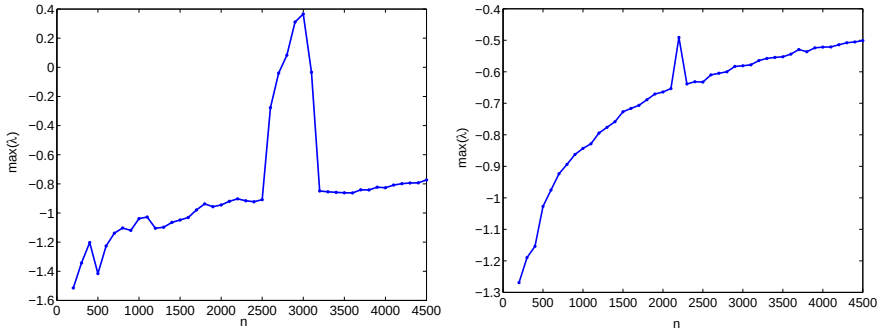
(a) Sobol' (b) Low distortion

Figure 6.5.4: Maximum eigenvalues of (6.5.5) for normal Sobol' and low distortion grids in three dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.



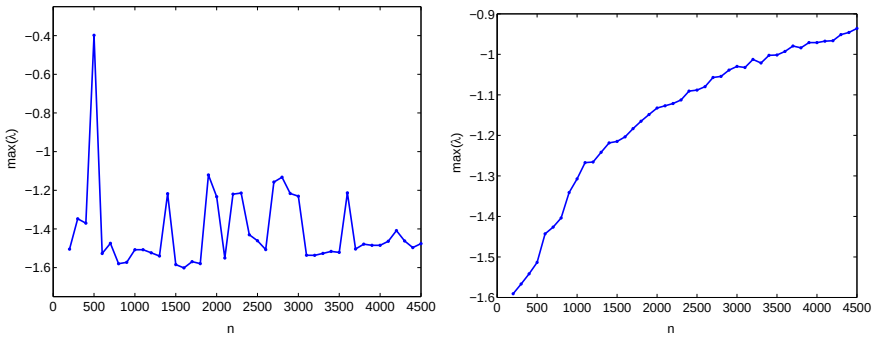
(a) Sobol' (b) Low distortion

Figure 6.5.5: Maximum eigenvalues of (6.5.5) for normal Sobol' and low distortion grids in four dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.



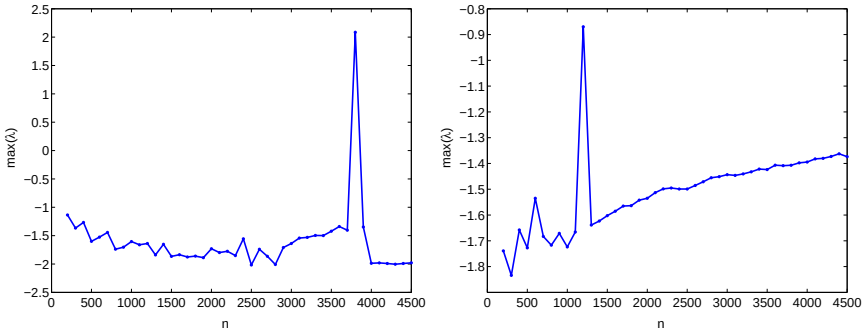
(a) Sobol' (b) Low distortion

Figure 6.5.6: Maximum eigenvalues of (6.5.5) for normal Sobol' and low distortion grids in five dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.



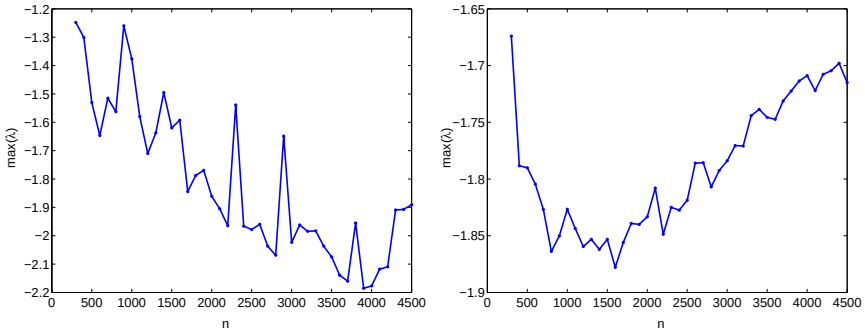
(a) Sobol' (b) Low distortion

Figure 6.5.7: Maximum eigenvalues of (6.5.5) for normal Sobol' and low distortion grids in six dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.



(a) Sobol' (b) Low distortion

Figure 6.5.8: Maximum eigenvalues of (6.5.5) for normal Sobol' and low distortion grids in seven dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.



(a) Sobol' (b) Low distortion

Figure 6.5.9: Maximum eigenvalues of (6.5.5) for normal Sobol' and low distortion grids in eight dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.

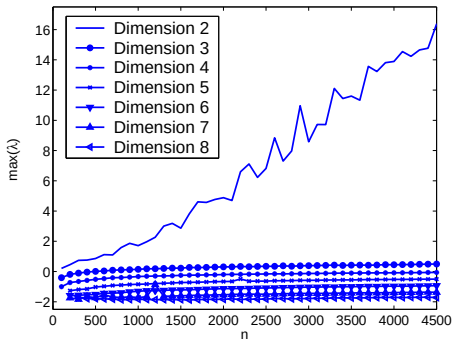


Figure 6.5.10: Maximum eigenvalues of (6.5.5) for low distortion grids in 2–8 dimensions plotted against grid size n with $\mu = 2$, $\alpha = 0.01$.

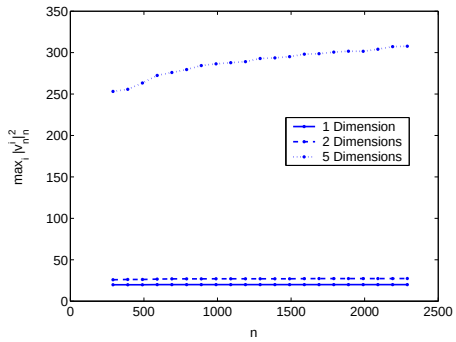


Figure 6.5.11: Plot of the first stability quantity $\max_i |v_n^i|_n^2$ versus grid size for normal Sobol' grids in dimensions 1, 2 and 5. We used $\delta t = 0.1$ and $\mu = 2$.

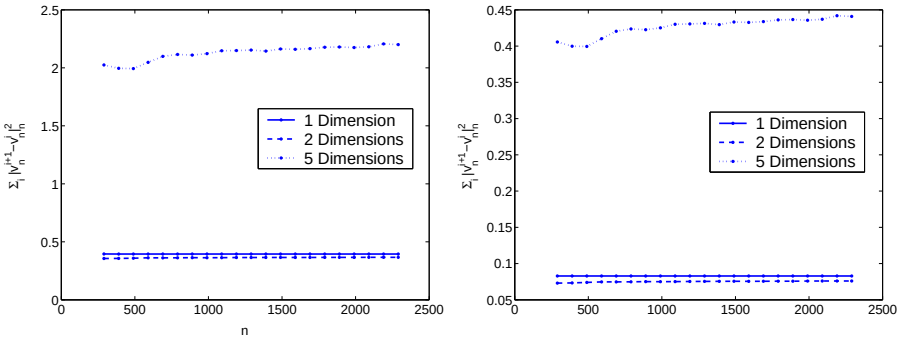
1. $\max_i |v_n^i|_n^2$,
2. $\sum_i |v_n^{i+1} - v_n^i|_n^2$,
3. $\delta t \sum_i \left(\theta \|v_n^{i+1}\|_n^2 + (1-\theta)\|v_n^i\|_n^2\right)$.

Figures 6.5.11–6.5.13 show the behaviour of the stability parameters plotted against grid size. In addition to the information displayed, we found that the time step had negligible effect on the first and third parameters, and only affected the level for the second parameter as shown in Figure 6.5.12. Again, altering the weighting parameter μ only affected the level of the results, and not the qualitative behaviour.

The graphs do not show any particularly unstable behaviour, except for the third parameter for low grid sizes in five dimensions. This does not seem to be a cause for concern since the grid sizes concerned are very low for a five-dimensional problem, and the quantity seems to be stable for larger grid sizes.

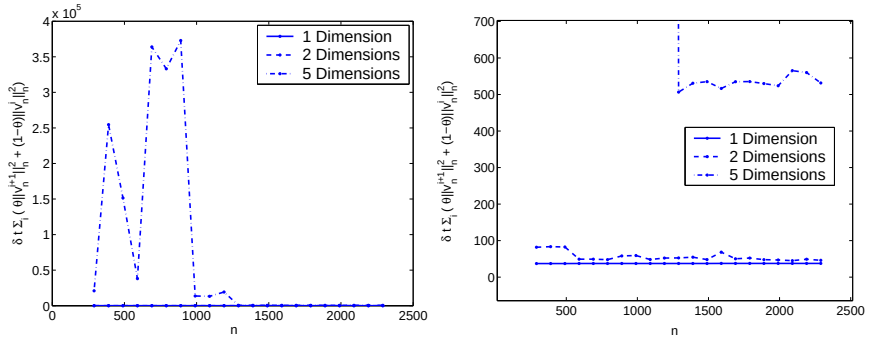
6.6 Conclusions

In this chapter we have used the variational inequality framework of Glowinski et al. [35] to prove convergence of numerical schemes for solving high-dimensional optimal stopping problems, including the American option problem. The proof may be applied in particular to the schemes suggested in Chapters 2–4 (also published as [8, 7, 9, 10]).



(a) $\delta t = 0.10$ (b) $\delta t = 0.02$

Figure 6.5.12: Plot of the second stability quantity $\sum_i |v_n^{i+1} - v_n^i|_n^2$ versus grid size for normal Sobol' grids in dimensions 1, 2 and 5 and two different time steps. We used $\mu = 2$.



(a) Full plot (b) Zoomed

Figure 6.5.13: Plot of the third stability quantity $\delta t \sum_i \left(\theta \|v_n^{i+1}\|_n^2 + (1-\theta)\|v_n^i\|_n^2\right)$ versus grid size for normal Sobol' grids in dimensions 1, 2 and 5. The right hand plot is zoomed to better show the behaviour for dimensions 1, 2 and for larger grid sizes in dimension 5. We used $\delta t = 0.1$ and $\mu = 2$.

The approach is similar to that of Jaillet et al. [42] and Zhang [76], but in contrast to these authors we consider a high-dimensional setting in which an irregular grid is used. The use of such a grid allows one to use a tractable number of points in the discretisation, as opposed to regular grid schemes where the number of points grows exponentially with dimension. For example, one can compute approximate solutions to ten dimensional problems using an irregular grid, whereas for a regular grid this would be nearly impossible using current computer technology.

The Glowinski et al. [35] framework is sufficient for our problem, except that we relax the coercivity assumption to a Gårding inequality assumption; this affects the proof of stability in their framework. Zhang [76] also assumes a Gårding inequality for proving convergence of a numerical method for pricing American options on a one-dimensional jump-diffusion process. Our proof rather treats the multidimensional case, and uses a discrete version of Gronwall's inequality to demonstrate stability. The latter provides a clearer understanding of the stability conditions.

A number of sufficient conditions are stipulated in the convergence proof which are difficult to check for the suggested schemes. Instead we check certain conditions experimentally for the numerical scheme of Chapter 4 for various specifications of the relevant parameters.

We find that the maximum eigenvalues related to the matrix (6.5.5) for inverse normal regular grids in one dimension exhibit a monotone increasing but sub-log behaviour. The maximum eigenvalues for normal Sobol' grids in one dimension show an oscillating behaviour whose mean appears to behave in a similar fashion to the regular grid case. For the two-dimensional case we observe unusual behaviour both for the regular and low distortion grids; further investigations are required here.

In higher dimensions the maximum eigenvalue behaviour is less regular, although for low distortion grids in dimensions 3–8 we observe a mostly monotone increasing convex dependence on the grid size with some slight aberrations. For the normal Sobol' grids it is more difficult to draw conclusions since a lot of variation is observed in the maximum eigenvalue behaviour.

Finally, a direct investigation of the stability quantities on the LHS of (6.4.33) does not indicate unbounded behaviour of these quantities with increasing grid size for dimensions 1, 2 and 5.

Chapter 7

Conclusions and Future Research

This thesis has presented several new numerical methods aimed at the pricing of high-dimensional American options, and more generally aimed at the efficient solution of high-dimensional optimal stopping problems. The numerical methods presented are based on irregular grid discretisations of the state space and most have been shown to be effective for problems of up to ten dimensions.

Chapters 2–4 present methods which use a discrete space Markov chain approximation to create a related but tractable optimal stopping problem. One may also see these methods as providing a numerical approach for solving the associated high-dimensional PDE problems. The experiments presented in these chapters show that the computed solutions are very accurate, as compared to available benchmarks, with the application of a simple control variate. Chapter 6 shows how one may prove the convergence of such schemes using the variational inequality framework developed by Glowinski et al. [35].

Chapter 5 presents a scheme based on value iteration, and using a different irregular grid at each time step. The method is found to work well when an inner control variate is applied, that is, when a suitable control variate is applied at each time step. The method is tested in up to ten dimensions and produces results consistent with other authors for the prices of Bermudan swaptions in a LIBOR market model setting. Surprisingly, the results are quite accurate even when only 100 grid points are used, and do not improve noticeably when up to 500 points are used.

The key ingredients of the methods presented are the use of randomisation and the absence of parametric functional approximation methods. The use of random-

isation includes both Monte Carlo and quasi-Monte Carlo methods, which have been used with great effect for finding numerical solutions to high-dimensional integration problems. This thesis shows how one may extend these methods to find numerical solutions for high-dimensional optimal stopping problems. The absence of parametric functional approximation sets the methods in this thesis apart from other methods used for solving American option pricing problems. Methods such as those suggested by Longstaff and Schwartz [50] and Tsitsiklis and Van Roy [71] require a clever choice to be made in the selection of basis functions for functional approximation; the methods presented in this thesis may thus be preferable when it is not possible to make such a choice.

One of the most difficult aspects of constructing numerical methods based on irregular grid discretisation turns out to be the stability analysis. One would naturally like to construct methods which are stable, but the connection between the specification of the method and its stability is often tenuous mathematically. We have seen two methods in this thesis where stability could be guaranteed on a constant irregular grid, namely those of Chapters 2 and 4. Stability was guaranteed in the former because the infinitesimal generator was constructed as a root of a diagonalisable matrix with eigenvalues in the unit interval, and in the latter through the use of linear programming in combination with an application of the Gershgorin disk theorem. We saw in Chapter 3 however that, even though the method presented is defined in a seemingly consistent manner, one may lack a tractable mathematical approach for analysing the stability; this makes such a method difficult to apply in practise.

A key question which remains open is whether the curse of dimensionality can be beaten for problems of the type considered in this thesis. Evidence is provided in Chapter 4 that the method used in that chapter suffers from the curse of dimensionality in that, in order to obtain an approximation with a certain accuracy, one requires an exponentially increasing amount of computational work with dimension. Rust [65] provides evidence that certain optimal control problems admit randomised numerical solution methods that do not suffer the curse of dimensionality. Such problems have a discrete action space, which is certainly the case for American options, but also require a compact state space. One may reformulate the American option problem on a compact state space, but it seems that performing such a transformation would violate the conditions specified for the dynamics in

[65].

In Chapter 5 we saw that the local quadratic interpolation, or moving least squares method, is very slow in a high-dimensional setting. Recent work by Maz'ya and Schmidt [53] and Fasshauer [29] suggest that this interpolation may be done much faster, with a small error as penalty, using so-called matrix-free methods to avoid the matrix inversion required in the interpolation. Such methods, if successful in a high-dimensional setting, would be useful to make the interpolation method viable for the type of problem considered in this thesis.

A simple question which remains open is whether one can in practise use an irregular grid to approximate types of high-dimensional stochastic process other than the ones considered in this thesis. Given their recent popularity, it would be of great interest to know for example whether multidimensional Lévy processes could be approximated consistently and efficiently in an irregular grid framework. Related work by Matache et al. [51, 52] shows how one can efficiently price both European and American options based on Lévy processes using a regular grid discretisation in one dimension.

The convergence analysis presented in Chapter 6 builds on the variational inequality framework introduced by Glowinski et al. [35] to provide a proof of convergence for the type of irregular grid methods introduced in Chapters 2–4. Sufficient conditions are provided for convergence of such methods; however, given the mathematical complexity of the methods, it turns out to be difficult to formulate conditions which can be checked analytically. We check certain conditions experimentally for the method presented in Chapter 4. It is of great interest to make stronger statements, including the specification of more easily verifiable conditions, regarding convergence of the proposed methods. Another approach that may be considered for constructing a proof of convergence is that of Kushner and Dupuis [46], who use an approximating Markov chain framework to prove convergence of numerical methods for stochastic optimal control problems.

A natural extension of the methods presented would be to the solution of optimal control problems. In related work, Munos and Moore [55, 56, 57] present adaptive discretisation methods for solving deterministic optimal control problems. Such methods use irregular discretisations for the purpose of obtaining more accurate representations of the problem in regions of the state space which are more turbulent, or likely to have a more significant effect on the solution. Though not

concentrating in particular on high-dimensional problems, their adaptive methods include the use of sparse and low discrepancy grids and they do report being able to solve stochastic optimal control problems in six dimensions; this is not expanded on in the above papers however. Their work demonstrates that irregular grid methodologies are applicable in deterministic, and possibly stochastic, optimal control problems. The results contained in this thesis concur, at least for the special case of optimal stopping problems.

Appendix A

Software

The software used for conducting experiments in this thesis was entirely Matlab-based. Matlab versions 5.3, 6.1 and 6.5 were used at various stages, and use was made both of the scripting capabilities of Matlab and the ability to compile parts of the code as `mex` files, using the C language.

The linear programming problems in Chapter 4 were solved using Tomlab/MINOS v4.1, which contains an implementation of the MINOS optimisation library developed at Stanford. These routines were found to give faster and more accurate results than those included in the Matlab optimisation toolbox.

The fast nearest neighbour searching required in Chapters 3–6 was carried out using the TSTOOL package developed at DPI Göttingen by Christian Merkwirth, Ulrich Parlitz, Immo Wedekind and Werner Lauterborn. This package is released under the GNU General Public License, and available from the DPI Göttingen web site.

Parallel computing was achieved in Matlab using the Parmatlab package developed by Lucio Andrade, and available online through the Matlab Central File Exchange. This package also makes use of the TCP/IP toolbox v1.2.3 by Peter Rydesater. The Parmatlab package was extended to function in Matlab 6 with version 2.0.2 of the TCP/IP toolbox, and various extra features were added such as more detailed output information and a timeout capability.

Bibliography

[1] Leif Andersen. A simple approach to the pricing of Bermudan swaptions in the multifactor LIBOR market model. *Journal of Computational Finance*, 3:5–32, 2000.

[2] Leif Andersen and Jesper Andreasen. Volatility skews and extensions of the Libor market model. *Applied Mathematical Finance*, 7:1–32, 2000.

[3] Vlad Bally and Gilles Pagès. A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli*, 6:1003–1049, 2003.

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

[5] A. Bensoussan and J.L. Lions. *Applications of Variational Inequalities in Stochastic Control*. North Holland, Amsterdam, 1982.

[6] Abraham Berman and Robert J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York, 1979.

[7] S.J. Berridge and J.M. Schumacher. An irregular grid method for solving high-dimensional problems in finance. In Sloot et al., editor, *International Conference on Computational Science (2)*, pages 510–519. Springer, Berlin, 2002.

[8] S.J. Berridge and J.M. Schumacher. An irregular grid approach for pricing high-dimensional American options. *CentER Discussion Paper, Tilburg University*, 18, 2004.

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

[10] S.J. Berridge and J.M. Schumacher. Pricing high-dimensional American options using local consistency conditions. *CentER Discussion Paper, Tilburg University*, 19, 2004.

[11] S.J. Berridge and J.M. Schumacher. Using localised quadratic functions on an irregular grid for pricing high-dimensional American options. *CentER Discussion Paper, Tilburg University*, 20, 2004.

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

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

[14] Phelim P. Boyle, Adam W. Kolkiewicz, and Ken Seng Tan. An improved simulation method for pricing high-dimensional American derivatives. *Mathematics and Computers in Simulation*, 62:315–322, 2003.

[15] A. Brace, M. Gątarek, and M. Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7:127–155, 1987.

[16] M. Brennan and E. Schwartz. The valuation of American put options. *Journal of Finance*, 32:449–462, 1977.

[17] Damiano Brigo and Fabio Mercurio. *Interest Rate Models, Theory and Practice*. Springer-Verlag, Berlin, 2001.

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

[19] Mark Broadie and Paul Glasserman. A stochastic mesh method for pricing high-dimensional American options. *Working Paper, Paine Webber Series, Columbia Business School*, 1997.

[20] Jacques F. Carrière. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30, 1996.

[21] Jacques F. Carrière. Linear regression and standardized quasi-Monte Carlo for approximating optimal stopping times. *Working Paper, University of Alberta*, 2003.

[22] 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.

[23] Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. *The Linear Complementarity Problem*. Academic Press, Boston, 1992.

[24] J. Cox, S. Ross, and M. Rubenstein. Option pricing: a simplified approach. *Journal of Financial Economics*, 7:229–263, 1979.

[25] Colin W. Cryer. The solution of a quadratic programming problem using systematic overrelaxation. *SIAM Journal of Control*, 9:385–392, 1971.

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

[27] Vladimir Druskin and Leonid Knizhnerman. Extended Krylov subspaces: approximation of the matrix square root and related functions. *SIAM Journal of Matrix Analysis and Applications*, 19:755–771, 1998.

[28] M. Evans and T. Swartz. *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford University Press, Oxford, 2000.

[29] Gregory E. Fasshauer. Approximate moving least squares approximation: a fast and accurate multivariate approximation method. In *Curve and Surface Fitting, Saint-Malo*, pages 1–10, 2002.

[30] Gregory E. Fasshauer, Abdul Q. M. Khaliq, and Dave A. Voss. Using meshfree approximation for multi-asset American option problems. *Working paper, Illinois Institute of Technology*, 2003.

[31] 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.

[32] R. Geske and H. Johnson. The American put option valued analytically. *Journal of Finance*, 39:1511–1524, 1984.

[33] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2003.

[34] Paul Glasserman and Bin Yu. Number of paths versus number of basis functions in American option pricing. *Working Paper, Columbia University*, 2003.

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

[36] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1996.

[37] Martin B. Haugh and Leonid Kogan. Pricing American options: a duality approach. *Working Paper, The Wharton School, University of Pennsylvania*, 2001.

[38] Marlis Hochbruck and Christian Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM Journal of Numerical Analysis*, 34:1911–1925, 1997.

[39] Y. C. Hon. A quasi-radial basis functions method for American option pricing. *Computers & Mathematics with Applications*, 43:513–524, 2002.

[40] J. Huang and J.-S. Pang. Option pricing and linear complementarity. *Journal of Computational Finance*, 2:31–60, 1998.

[41] Willem Hundsdorfer and Jan Verwer. *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*. Springer-Verlag, Berlin, 2003.

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

[43] Farshid Jamshidian. LIBOR and swap market models and measures. *Finance and Stochastics*, 1:293–330, 1987.

[44] Farshid Jamshidian. Minimax optimality of Bermudan and American claims and their Monte-Carlo upper bound approximation. *Working Paper, NIB Capital Bank*, 2003.

[45] Anastasia Kolodko and John Schoenmakers. An efficient dual Monte Carlo upper bound for Bermudan style derivatives. *Preprint 877, Weierstrass Institute*, 2003.

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

[47] David Levin. The approximation power of moving least-squares. *Mathematics of Computation*, 67(224):1517–1531, 1998.

[48] G.R. Liu. *Mesh Free Methods: Moving Beyond the Finite Element Method*. CRC Press, Boca Raton, 2002.

[49] Francis A. Longstaff, Pedro Santa-Clara, and Eduardo S. Schwartz. Throwing away a billion dollars: The cost of suboptimal exercise strategies in the swaptions market. *Journal of Financial Economics*, 62:39–66, 2001.

[50] 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.

[51] A.-M. Matache, P.-A. Nitsche, and C. Schwab. Wavelet Galerkin pricing of American options on Lévy driven assets. *ETHZ Seminar for Applied Mathematics (SAM) Research Report 2003-06*, 2003.

[52] A.-M. Matache, T. von Petersdorff, and C. Schwab. Fast deterministic pricing of options on Lévy driven assets. *ETHZ Seminar for Applied Mathematics (SAM) Research Report 2002-08*, 2002.

[53] Vladimir Maz'ya and Gunther Schmidt. On quasi-interpolation with non-uniformly distributed centers on domains and manifolds. *WIAS Preprint No. 422, Weierstrass Institute*, 1998.

[54] Manuel Moreno and Javier F. Navas. On the robustness of least-squares Monte Carlo (LSM) for pricing American derivatives. *Review of Derivatives Research*, 6:107–128, 2003.

[55] Rémi Munos and Andrew Moore. Influence and variance of a Markov chain: application to adaptive discretization in adaptive control. *IEEE Conference on Decision and Control*, 1999.

[56] Rémi Munos and Andrew Moore. Variable resolution discretization for high-accuracy solutions of optimal control problems. *International Joint Conference on Artificial Intelligence*, 1999.

[57] Rémi Munos and Andrew Moore. Variable resolution discretization in optimal control. *Machine Learning*, 49:291–323, 2002.

[58] Harald Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, 1992.

[59] Art B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.-S. Shiue, eds.), pages 299–317, 1995.

[60] Gilles Pagès. A space quantization method for numerical integration. *Journal of Computational and Applied Mathematics*, 89:1–38, 1997.

[61] Morton Bjerregaard Pedersen. Bermudan swaptions in the LIBOR market model. *SimCorp Financial Research Working Paper*, 1999.

[62] Antoon Pelsser. *Efficient Methods for Valuing Interest Rate Derivatives*. Springer-Verlag, London, 2000.

[63] Raoul Pietersz and Antoon Pelsser. Risk managing Bermudan swaptions in the BGM model. *ERIM Working Paper, Erasmus University*, 2002.

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

[65] John Rust. Using randomization to break the curse of dimensionality. *Econometrica*, 65(3):487–516, 1997.

[66] Alexander Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, Chichester, 1998.

[67] Jakob Sidenius. LIBOR market models in practice. *Journal of Computational Finance*, 3:5–26, 2000.

[68] Lars Stentoft. Assessing the least squares Monte-Carlo approach to American option valuation. *CAF Working Paper No. 90, Aarhus School of Business*, 2001.

[69] Lars Stentoft. Convergence of the least squares Monte-Carlo approach to American option valuation. *CAF Working Paper No. 113, Aarhus School of Business*, 2002.

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

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

[72] Holger Wendland. Local polynomial reproduction and moving least squares approximation. *IMA Journal of Numerical Analysis*, 21(1):285–300, 2000.

[73] Paul Wilmott. *Derivatives: The Theory and Practice of Financial Engineering*. John Wiley & Sons, Chichester, 1998.

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

[75] Christoph Zenger. Sparse grids. *Parallel Algorithms for Partial Differential Equations: Proceedings of the Sixth GAMM-Seminar Kiel, Notes on Numerical Fluid Mechanics*, 31:241–251, 1990.

[76] Xiaolan Zhang. Analyse numérique des options Américaines dans un modèle de diffusion avec des sauts. *PhD Thesis, CERMA-École Nationale des Ponts et Chaussées*, 1994.

[77] Xiaolan Zhang. Valuation of American options in a jump-diffusion model. In *Numerical Methods in Finance,* L.C.G. Rogers and D. Talay, eds., Cambridge University Press, pages 93–114, 1997.

Samenvatting

Dit proefschrift behandelt het probleem van de waardering van Amerikaanse opties in modellen waarbij de beslissing over al of niet uitoefening afhankelijk is van drie tot tien toestandsvariabelen. In het onderzoek is ook aandacht besteed aan het verwante probleem van de waardering van Bermudaanse opties, die slechts op een beperkt aantal tijdstippen uitgeoefend kunnen worden. Modellen met drie of meer toestandsvariabelen worden "hoogdimensionaal" genoemd omdat klassieke discretisatiemethoden, gebaseerd op regelmatige roosters, in deze context lastig toepasbaar zijn. In dit proefschrift worden discretisatiemethoden toegepast met onregelmatige roosters. Voordelen van het gebruik van een onregelmatig rooster zijn onder meer vrijheid in het kiezen van het aantal roosterpunten, en vrijheid in het plaatsen van de roosterpunten in gebieden die belangrijke invloed hebben op de te bepalen oplossing. Verder blijkt er een gunstig effect te zijn van onregelmatige roosters op een typisch probleem dat zich voordoet in hoogdimensionale situaties, namelijk het toenemend aandeel van randpunten in het totaal aantal roosterpunten. Een belangrijk nadeel van het gebruik van onregelmatige rooosters is dat het discretiseren van de differentiaaloperatoren niet meer op een voor de hand liggende manier kan gebeuren. In het proefschrift worden hiervoor verschillende oplossingen gesuggereerd.

De methoden die in het proefschrift worden ontwikkeld zijn niet alleen van toepassing op Amerikaanse opties; ze kunnen meer in het algemeen worden gebruikt voor stochastische beslissingsproblemen met binaire keuze. De voornaamste impuls voor het ontwikkelen van geavanceerde numerieke methoden voor hoogdimensionale problemen komt echter uit de financiële wereld. In de moderne financiële markten wordt een toenemend aantal contracten verhandeld waarvan de waarde kan afhangen van een groot aantal onderliggende variabelen. Bovendien di-

enen er niet alleen prijzen, maar ook afdekkings- en uitoefeningsstrategieën bepaald te worden. Producten die op rentestanden gebaseerd zijn, zoals Bermudaanse swaptions, vormen een groot deel van zulke contracten; andere voorbeelden zijn reële opties, die toegepast worden in investeringsanalyses, en multi-asset-opties waarvan de uitbetaling kan afhangen van bijvoorbeeld een aantal verschillende aandelen.

De mogelijkheid van vervroegde uitoefening, die zich voordoet zowel bij Amerikaanse als bij Bermudaanse opties, levert op zichzelf nog geen grote problemen op in één-dimensionale modellen. In hoogdimensionale modellen is de waardering van opties zonder mogelijkheid van vervroegde uitoefening al evenmin lastig: zulke problemen kunnen relatief snel opgelost worden met behulp van Monte Carlo of quasi-Monte Carlo integratiemethoden. Het is de combinatie van hoogdimensionaliteit met de mogelijkheid van vervroegde uitoefening die tot een zware opgave leidt voor de numerieke analyse, zowel wat betreft het waarderingsprobleem als wat betreft het bijbehorende afdekkingsprobleem. Er zijn geen oplosmethoden bekend die niet aanzienlijke rekenkracht vereisen.

Hoofdstukken 2–4 presenteren methoden die gebruik maken van een benaderende Markovketen om een verwant maar toch hanteerbaar beslissingsprobleem te verkrijgen. Deze methoden kunnen ook beschouwd worden als numerieke benaderingen van de overeenkomstige partiële differentiaalvergelijkingen. De experimenten in deze hoofdstukken tonen aan dat de methoden nauwkeurig zijn, in vergelijking met bekende methoden, als een geschikte stuurvariabele (control variate) wordt gebruikt. Hoofdstuk 6 laat zien hoe de convergentie van zulke methoden bewezen kan worden door middel van de analyse van numerieke methoden voor de benadering van variationele ongelijkheden die is ontwikkeld door Glowinski, Lions en Trémolières [35].

Hoofdstuk 5 presenteert een methode die gebaseerd is op waarde-iteratie, en die gebruik maakt van verschillende onregelmatige roosters op verschillende tijdstippen. Deze methode werkt goed als een inwendige stuurvariabele wordt toegepast, hetgeen wil zeggen dat voor ieder tijdstip afzonderlijk een geschikte stuurvariabele wordt gebruikt. De nauwkeurigheid van de methode wordt getoetst in problemen tot dimensie 10; de resultaten komen overeen met die van andere auteurs voor de waardering van Bermudaanse swaptions in een LIBOR-marktmodel. Opmerkelijk genoeg zijn de resultaten erg nauwkeurig, zelfs als er maar 100 pun-

ten gebruikt worden, en verbeteren ze niet merkbaar wanneer het aantal punten verhoogd wordt tot 500.

Essentiële ingrediënten van de gepresenteerde methoden zijn het gebruik van randomisatiemethoden, en het vermijden van benaderingen in geparametriseerde functieruimtes. Onder randomisatiemethoden worden hier zowel Monte Carlo als quasi-Monte Carlo methoden verstaan. De QMC methoden zijn zeer effectief gebleken voor het bepalen van numerieke oplossingen van hoogdimensionale integratie problemen. Dit proefschrift toont aan hoe deze methoden uitgebreid kunnen worden om numerieke oplossingen te vinden voor problemen die betrekking hebben op het kiezen van een optimaal tijdstip van beëindiging. De afwezigheid van benaderingen in geparametriseerde functieruimtes onderscheidt de methoden in dit proefschrift van andere methoden die gebruikt worden voor de waardering van Amerikaanse opties. Methoden zoals voorgesteld door Longstaff en Schwartz [50] en Tsitsiklis en Van Roy [71] vereisen een goede keuze van basisfuncties voor functiebenaderingen; de methoden in dit proefschrift zijn te prefereren als het lastig is zo'n keuze te maken.

Een van de moeilijkste aspecten van de constructie van op onregelmatige roosters gebaseerde numerieke methoden blijkt de analyse van de stabiliteit te zijn. Natuurlijk is het belangrijk om methoden te construeren die stabiel zijn, maar het verband tussen de specificatie van de methode en de stabiliteit is vaak lastig te bepalen. In hoofdstuk 3 wordt getoond dat, zelfs als de gepresenteerde methode op een schijnbaar consistente manier gedefinieerd is, een hanteerbare manier om de stabiliteit te analyseren daarmee niet gegeven hoeft te zijn. Dit maakt het moeilijk zo'n methode toe te passen in de praktijk. Het proefschrift geeft twee methoden waarmee stabiliteit gegarandeerd kan worden op basis van een constant onregelmatig rooster, namelijk in hoofdstukken 2 en 4. In hoofdstuk 2 wordt stabiliteit gegarandeerd door de infinitesimale generator te construeren als de wortel van een diagonaliseerbare matrix met eigenwaarden in de eenheidscirkel. In hoofdstuk 4 wordt stabiliteit verkregen door het gebruik van lineaire programmering in combinatie met de cirkelsstelling van Gershgorin.

Summary

This thesis addresses the problem of pricing American options where the decision on whether or not to exercise depends on between three and ten state variables. The closely related problem of pricing Bermudan options, where the number of exercise opportunities is finite, is included in the scope of the research. For “large” one may read “at least three”, since this is the dimension in which classical solution methods, in particular those based on regular grid discretisations, become cumbersome. The thesis further focuses on methods which use an irregular grid as a basis for calculations. Advantages of using an irregular grid are that one has freedom in choosing the number of grid points, and freedom in placing more points in areas where the behaviour has a greater effect on the required solution. A further advantage of using an irregular grid in high-dimensional situations is that the number of boundary points increases less quickly as a proportion of the total number of grid points. An important disadvantage of using an irregular grid is the lack of obvious methods for discretising the differential operator. This thesis suggests several methods for performing this discretisation.

In fact the methods in this thesis are not only applicable to American option pricing, but can also be used for general optimal stopping problems. The prime motivation for the study of high dimensional problems comes from the financial world. In modern financial markets one observes an increasing number of contracts whose values may each depend on a large number of underlying variables. Moreover it is of interest not only to determine the prices, but also the hedging and exercise strategies. Interest rate products such as Bermudan swaptions constitute a large class of such contracts; other examples are real options which are applied in investment analysis, and multiasset options where the payoff may depend for example on a number of stocks.

The early exercise feature, offered by both American and Bermudan options, does not present a great challenge for one-dimensional problems. Neither does the valuation of European options in a high-dimensional setting present a great challenge; such problems can be solved relatively quickly using Monte Carlo or quasi-Monte Carlo integration methods. In combining high-dimensionality with the early exercise feature however, one appears to require considerable computational resources both for the pricing problem and the associated hedging problem.

Chapters 2–4 present methods which use a discrete-space Markov chain approximation to create a related but tractable optimal stopping problem. One may also see these methods as providing a numerical approach for solving the associated high-dimensional PDE problems. The experiments presented in these chapters show that the computed solutions are very accurate, as compared to available benchmarks, with the application of a simple control variate. Chapter 6 shows how one may prove the convergence of such schemes using the variational inequality framework developed by Glowinski, Lions and Trémolières [35].

Chapter 5 presents a scheme based on value iteration, and using a different irregular grid at each time step. The method is found to work well when an inner control variate is applied, that is, when a suitable control variate is applied at each time step. The method is tested in up to ten dimensions and produces results consistent with other authors for the prices of Bermudan swaptions in a LIBOR market model setting. Surprisingly, the results are quite accurate even when only 100 grid points are used, and do not improve noticeably when up to 500 points are used.

The key ingredients of the methods presented are the use of (quasi-)randomisation and the absence of parametric functional approximation methods. The use of randomisation includes both Monte Carlo and quasi-Monte Carlo methods, which have been used with great effect for finding numerical solutions to high-dimensional integration problems. This thesis shows how one may extend these methods to find numerical solutions for high-dimensional optimal stopping problems. The absence of parametric functional approximation sets the methods in this thesis apart from other methods used for solving American option pricing problems. Methods such as those suggested by Longstaff and Schwartz [50] and Tsitsiklis and Van Roy [71] require a clever choice to be made in the selection of basis functions for functional approximation; the methods presented in this thesis may thus be preferable when it is not possible to make such a choice.

One of the most difficult aspects of constructing numerical methods based on irregular grid discretisation turns out to be the stability analysis. One would naturally like to construct methods which are stable, but the connection between the specification of the method and its stability is often tenuous mathematically. In Chapter 3 we see that even though method may be defined in a seemingly consistent manner, one may lack a tractable mathematical approach for analysing the stability. This makes such a method difficult to apply in practise. The thesis presents two methods where stability could be guaranteed on a constant irregular grid, namely those of Chapters 2 and 4. Stability was guaranteed in the former because the infinitesimal generator was constructed as a root of a diagonalisable matrix with eigenvalues in the unit interval, and in the latter through the use of linear programming in combination with an application of the Gershgorin disk theorem.