Importance Sampling for Least-Square Monte Carlo Methods

ILYAS AHMED
Importance Sampling for Least-Square Monte Carlo Methods

ILYAS AHMED

Master’s Thesis in Mathematical Statistics (30 ECTS credits)
Master Programme in Applied and Computational Mathematics (120 credits)
Royal Institute of Technology year 2016
Supervisor at Swedbank: Per Fust
Supervisor at KTH: Boualem Djehiche
Examiner: Boualem Djehiche

TRITA-MAT-E 2016:64
ISRN-KTH/MAT/E--16/64-SE

Royal Institute of Technology
SCI School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci
Abstract

Pricing American style options is challenging due to early exercise opportunities. The conditional expectation in the Snell envelope, known as the continuation value is approximated by basis functions in the Least-Square Monte Carlo-algorithm, giving robust estimation for the options price. By change of measure in the underlying Geometric Brownain motion using Importance Sampling, the variance of the option price can be reduced up to 9 times. Finding the optimal estimator that gives the minimal variance requires careful consideration on the reference price without adding bias in the estimator. A stochastic algorithm is used to find the optimal drift that minimizes the second moment in the expression of the variance after change of measure. The usage of Importance Sampling shows significant variance reduction in comparison with the standard Least-Square Monte Carlo. However, Importance Sampling method may be a better alternative for more complex instruments with early exercise opportunity.
Sammanfattning


Importance Sampling för Least-Square Monte Carlo-metoder
Acknowledgements

I take this opportunity to thank professor Boualem Djehiche at KTH for his excellent guidance and feedback. I would also like to thank Per Fust and Bengt Pramborg at Swedbank for their valuable guidance for writing this thesis. Finally, I would like to thank my family for their constant support for higher education.
# Contents

Abstract i

Sammanfattning ii

Acknowledgements iii

List of Figures vi

List of Tables vii

Abbreviations viii

1 Introduction 1
  1.1 Background ............................................. 1
  1.2 Previous Studies .......................................... 2
  1.3 Aim of this Thesis ...................................... 2

2 Mathematical Background 4
  2.1 American Options .......................................... 4
    2.1.1 Snell-envelope in discrete time ....................... 5
  2.2 Monte Carlo Methods .................................... 6
    2.2.1 Principal Aim .......................................... 7
    2.2.2 Importance Sampling .................................. 7
  2.3 Exponential Change of Measure .......................... 9

3 Least Square Monte Carlo 11
  3.1 The LSM Approach ........................................ 11
    3.1.1 Scenario Generation .................................. 12
  3.2 Basis Functions .......................................... 13
    3.2.1 Second order polynomial .............................. 14
    3.2.2 Laguerre Polynomials ................................ 15
  3.3 Convergence of the LSM algorithm ....................... 16

4 Implementing the IS-LSM Algorithm 17
  4.1 Expectation and the Likelihood Ratio .................... 17
4.2 IS-LSM Algorithm ............................................. 19
   4.2.1 Convergence of the IS-LSM algorithm ................. 20
   4.2.2 Stochastic Approximation for Optimal Drift ........... 21

5 Results .................................................................. 24

6 Discussion ............................................................ 34

A Essential supernum ............................................... 36

Bibliography ............................................................. 38
List of Figures

3.1 A realization of 200 stock rate paths using Euler-Maruyama scheme (left) and Millstein 1 scheme (right) over 251 trading days with $S_0 = 1, r = 0.02, \sigma = 20\%$ and $T = 1$. ............................... 12

3.2 The first 5 Laguerre polynomials ........................................... 15

4.1 A realization of 200 stock rate paths with drift $\theta = 0$ (left) and with drift $\theta = -1.6$ (right) over 251 trading days with $S_0 = 1, r = 0.02, \sigma = 20\%$ and $T = 1$. ............................... 18

5.1 A brute-force plot of $\frac{V_{IS-LSM}}{V_{LSM}}$ as a function of $\theta$ with $K = 50$ and $\sigma = 50\%$. ............................................. 25

5.2 Convergence of put option using standard LSM (left) and IS-LSM (right) with 2nd-order polynomials and sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 50, r = 0.02, \sigma = 10\%$ and $T = 1$. ............................. 28

5.3 Convergence of put option using standard LSM (left) and IS-LSM (right) with a sample size of with 2nd-order polynomials and sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 50, r = 0.02, \sigma = 50\%$ and $T = 1$. ............................. 28

5.4 Convergence of put option using standard LSM (left) and IS-LSM (right) with 2nd-order polynomials and sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 54, r = 0.02, \sigma = 10\%$ and $T = 1$. ............................. 29

5.5 Convergence of put option using standard LSM (left) and IS-LSM (right) with 2nd-order polynomials and sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 54, r = 0.02, \sigma = 50\%$ and $T = 1$. ............................. 29

5.6 Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 50, r = 0.02, \sigma = 10\%$ and $T = 1$. ............................. 30

5.7 Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 50, r = 0.02, \sigma = 50\%$ and $T = 1$. ............................. 30

5.8 Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 54, r = 0.02, \sigma = 10\%$ and $T = 1$. ............................. 31

5.9 Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50, K = 54, r = 0.02, \sigma = 50\%$ and $T = 1$. ............................. 31

5.10 Reference ratio vs $\theta \in [-1.17, -1.15]$ with $K = 56$ and $\sigma = 50\%$. ............................. 32
List of Tables

5.1 Strike price $K = 50$ with 2nd-order polynomial .................................. 25
5.2 Strike price $K = 52$ with 2nd-order polynomial .................................. 26
5.3 Strike price $K = 54$ with 2nd-order polynomial .................................. 26
5.4 Strike price $K = 56$ with 2nd-order polynomial .................................. 26
5.5 Strike price $K = 50$ with Hermite-polynomials .................................. 27
5.6 Strike price $K = 52$ with Hermite-polynomials ................................. 27
5.7 Strike price $K = 54$ with Hermite-polynomials .................................. 27
5.8 Strike price $K = 56$ with Hermite-polynomials .................................. 27
5.9 Optimal $\theta$ using second order polynomials .................................... 33
5.10 Optimal $\theta$ using using Hermite-polynomials ................................. 33
Abbreviations

DynP     Dynamic Programming
GBM      Geometric Brownian Motion
IS       Importance Sampling
LSM      Least-Squares Monte Carlo
IPA      Infinitesimal Perturbation Analysis
SA       Stochastic Approximation
Chapter 1

Introduction

Pricing methods of American style options has been well studied and there exists several different methods that are robust in a sense that they reflect the relevant price. Unlike the case of European options, where the price is deterministic by the Black & Scholes pricing formula, there is no closed form pricing model for American style options due to the early exercise opportunity. Pricing this kind of instruments requires numerical methods or Monte Carlo simulation in order to make relevant approximations of its price at time $t = 0$. This thesis will focus on efficient Monte Carlo methods for pricing an American put option. We will implement a modification of the well known algorithm developed by Longstaff & Schwartz by using importance sampling in order to get a variance reduction for the estimated price.

1.1 Background

Considering a one dimensional case of an underlying asset, the payoff function of the American option written on the asset is dependent on the strike price $K$ of the written contract. Unlike the European equivalents, the American style options have early exercise opportunity at some time $\tau < T$. These options with payoff function $f$ are priced as:

$$V_0 = \sup_{\tau \in [0,T]} \mathbb{E}[e^{-r\tau} f(\tau, S_\tau)].$$

The solution to (1.1) is an optimal stopping problem and has by usage of Dynamic Programming (DynP), a well-known solution. However, the outcome of the solution is not in closed form and $V_0$ needs to be priced by either numerical methods or Monte Carlo simulation, by the rules of the DynP solution. Longstaff & Schwartz introduced
a regression based approach to approximate (1.1). This is known as the Least-Square Monte Carlo algorithm or just simply LSM. It approximates the conditional expectations occurred in the optimal stopping problem with well behaved basis functions such that the price of the option at time $t = 0$ can be priced. However, the LSM-algorithm can give less accurate estimates if the strike prices and volatilities of the underlying asset are large. For this, efficient improvements of the LSM algorithm has to be done by means of change of measure in the underlying process.

1.2 Previous Studies

Studies of pricing American style options are well presented in several literature. Cox in [4] used the binomial tree methods for pricing these options. There are also trinomial trees and finite difference method for estimating the option price. The Longstaff & Schwartz method in [9] is based on Least-Square Monte Carlo method which is an efficient and yet simple method due to that it approximates the conditional expectation using basis functions. Moreni in [10] introduced an Importance Sampling (IS) approach for the LSM algorithm using American basket put options. The fact that LSM is only relevant for realizations that are in-the-money, the Monte Carlo estimate can deviate from the "true" value. The idea of implementing IS together with LSM is to make a change of drift in the underlying process and thereby getting more paths in-the-money. Moreni [10] used this technique with help of convergence ratio and European-derived criterion to find the optimal change of drift $\theta$ of the underlying process.

1.3 Aim of this Thesis

The aim of this thesis is to consider the LSM algorithm together with IS studied in Moreni [10], but with slightly different approaches. Here we focus on a simple American put option and try to find the optimal drift of the underlying process after change of measure. The method to be used is the exponential twisting method and we want to solve the optimization problem that minimizes the optimal drift $\theta$ as presented by Glasserman in [6]. The quantity of interest is that we have a convex function $H(\theta)$ which is the second moment of the IS estimator and want to solve:

$$H(\theta^*) = \min_{\theta \in \mathbb{R}} H(\theta)$$

(1.2)

We combine this study by using both stochastic approximation (SA) and infinitesimal perturbation analysis (IPA) as presented by Graham in [7] and Su in [11]. The goal is
to implement an algorithm that finds the value of the drift that gives a better variance reduction compared to standard LSM. The results of these methods then compared with findings in Moreni.

In Chapter 2, the mathematical background is introduced for setting up the necessary framework. In Chapter 3, the LSM algorithm is presented followed by the implementation of IS and the optimal drift in Chapter 4. Chapter 5 presents the results followed by conclusions in Chapter 6.
Chapter 2

Mathematical Background

This chapter presents the necessary mathematical background that is needed for later chapters when implementing the desired modified LSM algorithm. To begin with, we assume that our market model exists from time $t = 0$ to $t = T$ only. Introduce a probability space $(\Omega, \mathbb{P}, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]})$, where $\mathcal{F}$ is a $\sigma$-algebra with filtration $(\mathcal{F}_t)_{t \in [0,T]}$. The probability measure $\mathbb{P}$ is the standard measure that is relevant when changing to the risk-neutral measure during pricing.

2.1 American Options

Definition 2.1. An American put option with maturity $T$, strike price $K$ and payoff function $f(t, S(t)) = \max(K - S(t), 0)$ is a contingent claim that can be exercises at any time $t \leq T$.

At time to maturity $T$, the price of the option is $V(T, S(T)) = f(T, S(T))$ and in order to preclude arbitrage for the filtration $(\mathcal{F}_t)_{t \in [0,T]}$, the price has to satisfy the condition $V(t, S(t)) \geq f(t, S(t))$ for $t \in [0, T]$. Until the exercise moment, the option satisfies the partial differential equation:

$$rs \frac{\partial V(s, t)}{\partial s} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 V(s, t)}{\partial s^2} - rV(s, t) + \frac{\partial V(s, t)}{\partial t} = 0 \quad (2.1)$$

The general rule is that the holder should exercise the option as soon as the price $V(S(t), t) = f(S(t), t)$. During this time the equality in $(2.1)$ becomes an inequality due to the fact that $V(S(t), t)$ may not represent the value process of a strategy after the optimal exercise. The price of the American put satisfies the following boundary conditions:

$$V(S(T), T) = \max(K - S(T), 0) \quad (2.2)$$
\[
\lim_{s \to 0} V(s, t) = K
\]  
\[
\text{(2.3)}
\]

and

\[
\lim_{s \to \infty} V(s, t) = 0
\]  
\[
\text{(2.4)}
\]

Note that at \( t = T \), by equation (2.2) the price of the option is given by the standard Black & Scholes formula for European put (given that there are no dividend payments) option that is:

\[
V_{EU\text{put}} = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1)
\]

where \( \Phi \) is the standard normal cumulative distribution function and \( d_1 \) and \( d_2 \) are given by:

\[
d_1 = \frac{\log \left( \frac{S_0}{K} + T(r + \frac{\sigma^2}{2}) \right)}{\sigma \sqrt{T}}; \quad d_2 = d_1 - \sigma \sqrt{T}
\]

\[
\text{(2.5)}
\]

\[
\text{(2.6)}
\]

The extreme simplification by "elimination of risk" by the Black & Scholes formula does not require the knowledge on how the underlying asset behaves prior to time \( T \). This is however not fully applicable for the American style options due to early exercise opportunity. A concrete and relevant aspect of the American style options is presented by the Snell-envelope discussed next.

### 2.1.1 Snell-envelope in discrete time

In this thesis, we approximate the original stopping problem in continuous time with the stopping problem in discrete time as described in [2] and [5]. This results an optimal stopping time that differs from the original optimal stopping time. Let \((Z_j)_{j=0}^N\) be an adapted payoff process with \(Z_0, Z_1, \ldots, Z_N\) being square integrable random variables and let \( T = \{\tau_0, \tau_1, \ldots, \tau_N\} \cup \mathbb{R}^+ \) be set of optimal stopping times. As the objective is to compute equation (1.1), we have to solve the optimal stopping problem given by the Snell-envelope.

**Definition 2.2.** Let \((\Omega, \mathbb{P}, \mathcal{F})\) be a probability space. The adapted process \((V_j)_{j=0}^N\) is the Snell envelope on the process \((Z_j)_{j=0}^N\) if:

1. \( V \) is a super \( \mathbb{P} \)-martingale
2. \( V_j \geq Z_j \) for all \( j = 0, \ldots, N \)
3. If \( U = (U_j)_{j=0}^N \) is a super-\( \mathbb{P} \)-martingale dominating \( Z \), then \( U \) dominates \( V \)
In other words, the value process $V_j$ of an American option can be constructed with the Snell envelope using $N$ time steps with filtration $(\mathcal{F}_j)_{j \geq 0}$ such that:

$$V_j = \text{ess sup}_{\tau \in \mathcal{T}} \mathbb{E}[Z_\tau | \mathcal{F}_j], \quad \text{for } j = 0, \ldots, N$$

(2.7)

where ess sup denotes the *essential supremum* (see Appendix for details). The resulting DynP principle can be written as:

$$V_j = \begin{cases} 
Z_N, & j = N \\
\max\{Z_j, \mathbb{E}[V_{j+1} | \mathcal{F}_j]\}, & 0 \leq j \leq N - 1
\end{cases}$$

(2.8)

The sequence of stopping times is given by:

$$\begin{align*}
\tau_0 &= \inf\{k \geq 0 | V_k = Z_k\} \land N \\
\tau_j &= \inf\{k \geq \tau_{j-1} | V_k = Z_k\} \land N \\
\tau_{j+1} &\geq \tau_j
\end{align*}$$

(2.9)

which then obeys the DynP scheme (in discrete time) in terms of the optimal stopping times $\tau_j$:

$$\begin{align*}
\tau_N &= N \\
\tau_j &= j \mathbf{1}_{\{Z_j \geq \mathbb{E}[Z_{j+1} | \mathcal{F}_j]\}} + \tau_{j+1} \mathbf{1}_{\{Z_j < \mathbb{E}[Z_{j+1} | \mathcal{F}_j]\}}, \\
0 &\leq j \leq N - 1
\end{align*}$$

(2.10)

This formulation in terms of the optimal stopping rules place a central role in the least square regression method in the LSM algorithm. Note that $\tau_0 < \tau_1 < \ldots < \tau_N$ and that $\inf$ in (2.9) is indeed a min since we are dealing with discrete time. The adapted payoff process $(Z_j)_{j=0}^N$ is given by the American put option given in Definition 2.2 i.e:

$$Z_j = f(j, S_j)$$

(2.11)

where underlying model is assumed to be a $(\mathcal{F}_j)$-Markov chain for some Borel function $f(j, \cdot)$ (see [3] for more details). The expectation $\mathbb{E}[V_{j+1} | \mathcal{F}_j]$ given in (2.8) is of interest and is discussed in more detail in the next chapter. We now continue to discuss some important topics in Monte Carlo methods.

### 2.2 Monte Carlo Methods

Monte Carlo methods are a broad class of algorithms relying on repetitive random sampling for obtaining approximate numerical results. The algorithms are used in physical
and mathematical problems and are mainly useful when it is impossible to make any standard numerical approximations of a given problem.

### 2.2.1 Principal Aim

The principal aim for a Monte Carlo methods is related to compute an expectation of the form:

$$ V := \mathbb{E}(h(X)) $$

(2.12)

where $X \subseteq \mathbb{R}^d$, i.e. a random variable taking values in $d$-dimensional real line and the function $h : X \to \mathbb{R}$ is some objective function. The expectation under (2.12) can become complicated to calculate if the objective function has characteristics that implies a large variance. Assuming that this integral is complicated to evaluate both analytically and numerically, we can approximate it by generating a I.I.D. sequence of $X_i$’s using pseudo-random number generation and set:

$$ \frac{1}{N} \sum_{i=1}^{N} h(X_i) = \hat{V}_N $$

(2.13)

which is the standard Monte Carlo estimator. The estimator $\hat{V}_N$ is unbiased and strongly consistent as $N \to \infty$ meaning that:

$$ \hat{V}_N \to V \text{ with probability 1} $$

(2.14)

and for large $N$ we can construct a confidence interval for $\hat{V}_N$ by:

$$ \hat{V}_N \pm \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{\text{Var}(\hat{V}_N)}{N}} $$

where $\Phi^{-1}$ is the inverse of the cumulative standard normal distribution function.

### 2.2.2 Importance Sampling

Importance Sampling is attempted to reduce the variance of a standard monte carlo estimator by the change of measure. We know that under $\mathbb{P}$, the standard monte carlo estimator for a function $h(X)$ is:

$$ V_N = \mathbb{E}^\mathbb{P}(h(X)) \approx \frac{1}{N} \sum_{i=1}^{N} h(X_i) $$

(2.15)
where $E^P$ denotes that the expectation is taken under measure $P$. We now want to estimate the same expectation under a new measure $Q$ by constructing a sampling average of independent copies of

$$h(X) \frac{dP}{dQ}$$

such that $E^P[h(X)] = E^Q[h(X) \frac{dP}{dQ}]$. Assumptions that $P << Q$ is made such that the likelihood ratio exists. More concrete, if $X$ is a random element with probability density function $f$ under $P$-measure, then we introduce another probability density function $g$ under $Q$-measure such that:

$$f(x) > 0 \implies g(x) > 0 \text{ for all } x \in \mathbb{R}. \quad (2.16)$$

Here, we have that $dP = fdx$ and $dQ = gdx$. The importance sampling procedure changes the measure to try to give more weight to the important regions of the desired distribution and therefore increasing the efficiency of the sampling. $V$ can then in integral form be represented as:

$$V = \int h(x) \frac{f(x)}{g(x)} g(x) dx. \quad (2.17)$$

Expressing this integral as an expectation, we have the expectation with respect to the density $g$ (i.e. under measure $Q$) given by:

$$V = E^Q \left[ h(X) \frac{f(X)}{g(X)} \right] = E^Q[h(X)w(X)] \quad (2.18)$$

and its Importance Sampling (IS) estimator is given by:

$$V^IS_N = \frac{1}{N} \sum_{i=1}^{N} h(X_i)w(X_i). \quad (2.19)$$

The weight $w(X)$ is known as the Likelihood ratio, or by means of measure theory, the Radon-Nikodym derivative, i.e:

$$w(X) = \frac{dP}{dQ} =: L. \quad (2.20)$$

Following from equation (2.18), the importance sampling estimator is unbiased meaning that $E^Q \left[ \hat{V}^IS_N \right] = V$ which yields:

$$E^Q[h(X)w(X)] = E^P[h(X)]. \quad (2.21)$$
Due to this fact, the only quantity of interest is to compare the second moments of the IS and the standard Monte Carlo estimator:

\[
E_Q \left[ \left( h(X) \frac{f(X)}{g(X)} \right)^2 \right] = E^p \left[ h(X)^2 \frac{f(X)}{g(X)} \right]. \tag{2.22}
\]

It is desired, but not guaranteed that the variance in the standard Monte Carlo is larger, i.e:

\[
\text{Var} [h(X)] \geq \text{Var} [h(X)w(X)].
\]

Since IS can be used for rare-event simulation, the key usage will later on be on the stock-price realizations with more "weight" to important regions.

### 2.3 Exponential Change of Measure

For a cumulative distribution function \( F \) on \( \mathbb{R} \), let \( M(\theta) = \int_{-\infty}^{\infty} e^{\theta x} dF(x) \) be the moment generating function of \( F \) and define the cumulant generating function:

\[
\psi(\theta) = \log \int_{-\infty}^{\infty} e^{\theta x} dF(x) \tag{2.23}
\]

Introduce \( \Theta = \{ \theta : \psi(\theta) < \infty \} \) as a non-empty set. Now, for each \( \theta \in \Theta \), set:

\[
F_\theta(x) = \int_{-\infty}^{x} e^{\theta u - \psi(\theta)} dF(u); \tag{2.24}
\]

where each \( F_\theta \) is a probability distribution and the set \( \{ F_\theta, \theta \in \Theta \} \) forms an exponential family of distributions. The transformation from distribution \( F \) to \( F_\theta \) is the exponential change of measure. Moreover if \( F \) has density \( f \) then \( F_\theta \) has the density:

\[
f_\theta(x) = e^{\theta x - \psi(\theta)} f(x) \tag{2.25}
\]

**Example 2.1.** Let \( X_1, \ldots, X_n \) be i.i.d. with distribution function \( F \). Applying the change of measure for which \( X_i \)'s becomes i.i.d. under distribution \( F_\theta \), the likelihood ratio of the transformation is:

\[
\prod_{i=1}^{n} \frac{dF(X_i)}{dF_\theta(X_i)} = \exp \left( -\theta \sum_{i=1}^{n} X_i + n\psi(\theta) \right). \tag{2.26}
\]

**Example 2.2.** Continuing from example 2.1, we can change the mean of a normal distribution. Let \( f \) be the standard normal distribution density and \( g \) normal distribution density with mean \( \mu \) and standard deviation 1. Using the same algebra as from the
previous example we simply get:

$$
\prod_{i=1}^{m} \frac{f(Z_i)}{g(Z_i)} = \exp \left( -\mu \sum_{i=1}^{m} Z_i + \frac{m}{2} \mu^2 \right). 
$$

(2.27)

If we introduce a grid $0 = t_0 < t_1 < ... < t_m$ and simulate a Brownian motion given by:

$$
W(t_n) = \sum_{i=1}^{n} \sqrt{t_i - t_{i-1}} Z_i
$$

(2.28)

then the likelihood ratio $L$ for the change of measure adding drift $\mu \sqrt{t_i - t_{i-1}}$ to the Brownian increment over $[t_{i-1}, t_i]$ is given by equation (2.27).

Comparing Example 2.1 and Example 2.2, we can conclude that for the standard normal distribution, $\psi(\theta) = \frac{\theta^2}{2}$, which is a generalization of change of mean for the normal distribution. With that said, we can now state the following theorem.

**Theorem 2.3. (Girsanov Theorem).** Consider the stochastic differential equation in $\mathbb{R}^d$ satisfying the Lipschitz coefficients and driven by the standard Brownian motion:

$$
X(t) = X(0) + \int_{0}^{t} b(X_s)ds + \int_{0}^{t} \sigma(X_s)dW_s
$$

(2.29)

Let $(\theta_s)_{0 \leq s \leq T}$ be an $\mathbb{R}^d$ valued measurable adapted process such that:

$$
\int_{0}^{T} |\theta_s|^2 ds < \infty, \quad a.s
$$

(2.30)

and

$$
\mathbb{E} \exp \left\{ \int_{0}^{T} \theta_s \cdot dW_s - \frac{1}{2} \int_{0}^{T} |\theta_s|^2 ds \right\} = 1
$$

(2.31)

The probability measure $\mathbb{P}^\theta$ on $(\Omega, \mathcal{F}_T)$ defined by the Radon-Nikodym derivative:

$$
\frac{d\mathbb{P}^\theta}{d\mathbb{P}} := \exp \int_{0}^{T} \theta_s \cdot dW_s - \frac{1}{2} \int_{0}^{T} |\theta_s|^2 ds
$$

(2.32)

is equivalent to $\mathbb{P}$. In addition, the process:

$$
W_t^\theta = W_t - \int_{0}^{t} \theta_s ds, \quad 0 \leq t \leq T,
$$

(2.33)

is an Brownian motion on $(\Omega, \mathcal{F}_T), \mathbb{P}^\theta$ and $\mathbb{P}^\theta$-a.s, for $0 \leq t \leq T$,

$$
X_t = X_0 + \int_{0}^{t} (b(X_s) + \sigma(X_s)\theta_s) ds + \int_{0}^{t} \sigma(X_s)dW_s^\theta.
$$

(2.34)
Chapter 3

Least Square Monte Carlo

This chapter goes through the most important aspects of the Longstaff & Schwartz algorithm. The procedure is well explained in [9] including a detailed example. This chapter goes briefly through the procedure of the LSM algorithm and emphasize the most important build-blocks and results of the algorithm.

3.1 The LSM Approach

As mentioned in the previous chapters, the expectation $E[V_{j+1} | F_j]$ in (2.8) is of interest. The holder of an American put option with maturity $T$ has to decide whether it is optimal to exercise the option at some time $t$ prior to $T$. Since the holder does not know the payoff at time $t+1$, the optimal stopping problem in terms of the expectation determines the expected value of the option at time $t+1$. This expectation is known as the continuation value and the LSM approach approximates this with a set of basis functions in which it then can easily be estimated by simple least square regression. Consider the standard Black & Scholes model that for which the bank account dynamics is given by:

$$dB(t) = rB(t)dt \text{ with } B(0) = 1, \quad (3.1)$$

and the stock price dynamics

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \quad (3.2)$$

Under the risk neutral measure $\mathbb{Q}$, the stock price has dynamics:

$$dS(t) = rS(t)dt + \sigma S(t)dW(t)^{\mathbb{Q}} \quad (3.3)$$
where $r \geq 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$. This model will be used as the underlying process in the LSM-algorithm. As Longstaff & Schwartz uses the argument of approximating $E[V_{j+1}|\mathcal{F}_j]$ using basis functions, we can simplify the problem to a least-square regression problem. This means that each continuation value in the Snell-envelope is indeed a least-squares approximation and we can therefore decide whether or not it is optimal to exercise at a given time $t = \tau$ prior to $t = T$.

### 3.1.1 Scenario Generation

Because we want to numerically simulate (3.3), we need a stochastic approximation for the solution of the underlying process. Considering the stochastic process $S$ driven by the stochastic differential equation:

$$dS(t) = a(S(t), t)dt + b(S(t), t)dW(t) \tag{3.4}$$

there are several schemes to simulate the above SDE. We set $\Delta_i = t_{i+1} - t_i$ and approximate the SDE by the Euler-Maruyama scheme and Millstein-1-scheme given by:

$$S_{t_{i+1}} = S_{t_i} + a_{t_i} \Delta_t + b_{t_i} \sqrt{\Delta_t} Z_i \tag{3.5}$$

and

$$S_{t_{i+1}} = S_{t_i} + a_{t_i} \Delta_t + b_{t_i} \sqrt{\Delta_t} Z_i + \frac{1}{2} b_{t_i}^2 \Delta_t (Z_i^2 - 1) \tag{3.6}$$

respectively, where $Z_i \sim N(0, 1)$. For the case of Geometric Brownian motion, eq (3.3) has $a(S(t), t) = rS(t)$ and $b(S(t), t) = \sigma S(t)$. Stock price realizations are shown in figure 3.1 below using both the Euler and Millstein scheme. The Euler scheme has convergence

![Figure 3.1](image-url)
order of $O(\Delta)$. Using Itô’s formula to any twice differentiable $C^2$ function $f(S(t), t)$ and limiting the expansions of the stochastic integrals, there will be another term in the approximation whose convergence is of order $O(\Delta)$. Therefore, the Euler Scheme will be incomplete and is compensated with the Millstein 1 Scheme in (3.6) which gives a better approximation of order one in $\Delta$ (see [2] for details).

### 3.2 Basis Functions

The goal is to approximate the conditional expectation appeared in the Snell envelope using basis function. If the payoff function $Z_t = f(S_t)$ of the option has its elements in the space of square-integrable finite-variance functions $L^2(\Omega, \mathcal{F}, \mathbb{Q})$, then the value of the option is the maximized values of the discounted cash-flows from the option over all $\mathcal{F}_t$ stopping times according to equation (2.9). The conditional expectation is the continuation value of the option, meaning that if it is larger than the current value at $t$ then we do not exercise at that point of time. In the original paper of Longstaff & Schwartz, the LSM algorithm represents the basis functions as:

$$ F(S_t) = \sum_{j=0}^{\infty} a_j e_j(S_t), \quad a_j \in \mathbb{R} $$

for some basis function $e(\cdot)$. To implement the LSM algorithm, the approximation of $F(S_t)$ is done using the first $m < \infty$ basis functions and is denoted as $\hat{F}_m(S_t)$. The choice of basis functions is entirely arbitrary and can be anything from Hermite, Legendre to Chebyshev polynomials. In this chapter, we will focus on simple second order polynomials and Laguerre polynomials. Since American style options can be approximated by *Bermuda*-style options, we will discretize the time steps to $n = 1, 2, ..., N$ number of time steps, meaning that the early exercise opportunities are restricted to a finite set. Of course here $M$ should be as large as possible. In the optimal stopping time given in equation (2.9), we let $Z_k = f(S_k)$.

When using the LSM algorithm, the option holder chooses at every time $t_n$ whether to exercise or keeping the option alive by comparing the continuation value with the immediate value. Having the approximation of the conditional expectation in hand, the optimal stopping time ($\tau_j^{[m]}$) can be expressed recursively by using $N^*$ independent simulated paths with:

$$
\begin{cases}
\tau_N^{n,m,N^*} = N \\
\tau_j^{n,m,N^*} = j^*1_{\{Z_j^{(n)} \geq a_j^{(m,N^*)}e_m(X_j^{(n)})\}} + \tau_{j+1}^{n,m,N^*}1_{\{Z_j^{(n)} < a_j^{(m,N^*)}e_m(X_j^{(n)})\}}, & 0 \leq j \leq N - 1
\end{cases}
$$

(3.8)
The notion \( \cdot \) denotes the inner product in \( \mathbb{R}^m \), \( e^m \) is the vector valued function and \( \alpha_j^{(m,N^*)} \) is given by the estimator:

\[
\alpha_j^{(m,N^*)} = \arg \min_{a \in \mathbb{R}^m} \sum_{n=1}^{N^*} \left( Z_{n,m,N^*}^{ \tau_{j+1} } - a \cdot e^m(X_j^{(n)}) \right)^2 .
\] (3.9)

The procedure of the algorithm is presented by following pseudo-code:

**Algorithm 1 LSM-algorithm**

1: for \( n = 1, ..., N^* \) do
2: \hspace{1em} Generate realizations of the GBM with \( N \) time steps
3: \hspace{1em} Set \( \tau = N \)
4: end for
5: for \( t = N-1, ..., 1 \) do
6: \hspace{1em} for \( n = 1, ..., N^* \) do
7: \hspace{2em} Compute \( a_0, ..., a_M \)
8: \hspace{2em} Set \( \hat{F}(t, S^{(n)}(t)) = \sum_{l=0}^{M} \hat{a}_l e_l(S^{(n)}(t)) \)
9: \hspace{2em} if \( V(t, S^{(n)}(t)) \geq \hat{F}(t, S^{(n)}(t)) \) then
10: \hspace{3em} \( \tau^{(n)} = t \)
11: \hspace{2em} end if
12: end for
13: end for

After optimal stopping time is determined for each path, the price is determined by the averaging over the discounted payoffs. The LSM-algorithm only includes in-the-money paths for estimation which of course makes sense because the value of the payoff function is zero whenever it is out-the-money. However, if the volatility of the underlying process is large or the strike price of the option is high, the out-the-money paths will increase giving less precise estimation (see Chapter 4).

### 3.2.1 Second order polynomial

The conditional expectation can well be approximated by second order polynomials. For the simple case of American put option, implementations of the LSM-algorithm exists as a library for the \( \mathbb{R} \) statistical language. In this simple but effective algorithm, the approximation of the continuation value is given by \( E[Y|X] \approx a_0 + a_1 X + a_2 X^2 \).
3.2.2 Laguerre Polynomials

A well known basis function that also is introduced in Longstaff & Schwartz original paper are the Laguerre polynomials defined by:

\[ L_n(x) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \binom{n}{k} x^k \]  

and it is fortunately enough to only use the first four terms of the series [9], i.e:

\[
\begin{align*}
L_0(x) &= 1 \\
L_1(x) &= -x + 1 \\
L_2(x) &= \frac{1}{2} (x^2 - 4x + 2) \\
L_3(x) &= \frac{1}{6} (-x^3 + 9x^2 - 18x + 6)
\end{align*}
\]

Figure 3.2 shows the first 5 Laguerre polynomials plotted between \(-2 \leq x \leq 10\).

\textbf{Figure 3.2:} The first 5 Laguerre polynomials

In [9] it is emphasized that in order to avoid computational underflows during pricing, it is recommended to normalize the cash-flows by dividing all cash flows and prices by the strike price and weighting the Laguerre polynomials by \(\exp(-x/2)\) resulting in Hermite polynomials:

\[ H_n(x) = \exp(-x/2) \frac{e^x}{m!} \frac{d^n}{dx^n} (x^n e^{-x}). \]

With this specification, (3.7) can be represented as:

\[ F(S_t) = \sum_{j=0}^{\infty} a_j H_j(S_t), \quad a_j \in \mathbb{R} \]  

(3.11)
3.3 Convergence of the LSM algorithm

In the original paper by L&S, the convergence of the LSM estimator is stated. Letting $V_0$ be the value of the option at time 0 and by choosing $m$ number of relevant basis functions, one can state the convergence of the LSM algorithm as follows:

$$\hat{V}_0 \geq \lim_{N^* \to \infty} \frac{1}{N^*} \sum_{i=1}^{N^*} V(\omega_i; m, K)$$

(3.12)

where $\omega_i$ denotes path $i$, with each path having its own optimal stopping time $\tau$. It has been proven by Clément, Lamberton and Protter [8] that under some general hypothesis on the payoff process $Z$ and basis functions $e_k$’s, the following convergence results hold:

$$V_{0, N^*}^m \xrightarrow{a.s.} V_{0}^m, \quad \text{as } N^* \to \infty \quad \forall m \in \mathbb{N}$$

(3.13)

and

$$V_{0}^m \to V_0, \quad \text{as } m \to \infty$$

(3.14)

Moreover, it has been proven (see [3]) that under certain hypothesis the central limit theorem holds, meaning that for every $j = 1, \ldots, L - 1$, as $N^* \to \infty$ the vector:

$$\left( \frac{1}{\sqrt{N^*}} \sum_{n=1}^{N^*} \left( f(\tau_j^{m,N^*}, S_{\tau_j^{m,N^*}}^{(n)}) - \mathbb{E} \left[ f(\tau_j^{m}, S_{\tau_j^{m}}) \right] \right), \sqrt{N^*}(a_j^{(N^*)} - a_j) \right) \to G$$

(3.15)

meaning that it converges in distribution to a Gaussian vector $G$. 

Chapter 4

Implementing the IS-LSM Algorithm

Now that the optimal stopping time can be solved with LSM, this chapter presents the implementation of the Importance Sampling Least-Square Monte Carlo (IS-LSM) algorithm together with finding the optimal drift as discussed in the previous chapters.

4.1 Expectation and the Likelihood Ratio

Let us recall equation (1.1) as it is the aim for approximation:

\[
V_0 = \sup_{\tau \in \mathcal{T}_{[0,T]}} \mathbb{E}[e^{-r\tau} f(\tau, S_\tau)].
\]  

(1.1)

For a standard put option, the payoff function is given by \( f(\tau, S_\tau) = \max(K - S_\tau, 0) \) and to price the option, the expectation is taken under equivalent martingale measure \( Q \).

Assuming deterministic interest rate \( r \) and non-stochastic volatility \( \sigma \) under the whole filtration \( (\mathcal{F}_t)_{t \in [0,T]} \), the equation becomes:

\[
V_0 = \sup_{\tau \in \mathcal{T}_{[0,T]}} \mathbb{E}^Q \left[ e^{-r\tau} \max(K - S_\tau, 0) \right],
\]

(4.1)

where \( \mathcal{T} \) denotes a family of stopping times with values in \([0,T]\) and where, under the risk-neutral measure \( Q \), we assume \((S_t)_{t \geq 0}\) is a geometric Brownian motion:

\[
\frac{dS_t}{S_t} = rdt + \sigma dW_t^Q.
\]

(4.2)
Now, doing importance sampling for computing the price (4.1) by simulating paths and computing the expectation not under the risk-neutral measure $Q$, but rather, under a modified measure $Q^\theta$ with drift $\theta$ giving us (by the Girsanov theorem explained in [1]):

$$\frac{dS_t}{S_t} = (r + \theta) dt + \sigma dW_t^Q,$$  \hspace{1cm} (4.3)

which implicitly shapes the Radon-Nikodym derivative

$$\left. \frac{dQ}{dQ^\theta} \right|_{F_t} = \exp \left( -\theta W_t^Q - \frac{1}{2} \theta^2 t \right) := L(\theta, t).$$  \hspace{1cm} (4.4)

Figure 4.1 shows a comparison of a non-drifted and drifted realisation of stock rate paths using Millstein-1 scheme.

![Figure 4.1: A realization of 200 stock rate paths with drift $\theta = 0$ (left) and with drift $\theta = -1.6$ (right) over 251 trading days with $S_0 = 1$, $r = 0.02$, $\sigma = 20\%$ and $T = 1$.](image)

Now if we use a change of measure argument together with the optimal sampling theorem, we can state the following proposition.

**Proposition 4.1.** Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]})$ be a probability space and $W_t$ the standard Brownian motion. Then for all $\theta \in \mathbb{R}$ and for all stopping times $\tau \in \mathcal{T}_{[0, T]}$ the following equality holds:

$$\mathbb{E}^Q \left[ e^{-r\tau} \max(K - S_\tau, 0) \right] = \mathbb{E}^{Q^\theta} \left[ e^{-r\tau} \max(K - S_\tau, 0)L(\theta, \tau) \right]$$  \hspace{1cm} (4.5)

and

$$\sup_{\tau \in \mathcal{T}_{[0, T]}} \mathbb{E}^Q \left[ e^{-r\tau} \max(K - S_\tau, 0) \right] = \sup_{\tau \in \mathcal{T}_{[0, T]}} \mathbb{E}^{Q^\theta} \left[ e^{-r\tau} \max(K - S_\tau, 0)L(\theta, \tau) \right].$$  \hspace{1cm} (4.6)

**Proof.** Equation (4.5) is by knowledge of exponential martingale property, already proven and equation (4.6) is trivial since the set $\mathcal{T}_{[0, T]}$ remains unchanged in the change of probability measure. \hfill $\Box$
By proposition 4.1 we can rewrite the pricing equation (1.1) as:

\[ V_0 = \sup_{\tau \in \mathcal{T}[0,T]} E[f(\tau, S_\tau)] = \sup_{\tau \in \mathcal{T}[0,T]} E[f(\tau, S_\tau)L(\theta, \tau)] = V_0^\theta \]  

(4.7)

### 4.2 IS-LSM Algorithm

With the tools gathered from previous chapters and the knowledge of change of measure, we can state the steps for the IS-LSM algorithm:

1. Generate \( N \) stock price paths under the modified measure \( Q^\theta \) using the SDE (4.3).
2. Apply the LSM-algotihm to obtain, for each simulated path \( n = 1, \ldots, N \) the value of the discounted exercise value at the optimal stopping time \( \tau^{(n)} \)

\[ e^{-r\tau^{(n)}} \max(S_{\tau^{(n)}} - K, 0) \]

(Here as described in Chapter 3, the LSM algorithm identifies optimal stopping times by approximating continuation values through polynomial regressions relying on cross-sectional information and working backwards in time).

3. Now once that is clear, we can compute the Monte Carlo estimator:

\[ V_0 = \sup_{\tau \in \mathcal{T}[0,T]} E[Q^\theta_0 \left(e^{-r\tau} \max(S_\tau - K, 0)L(\theta, \tau)\right)] \]

\[ \approx \frac{1}{N} \sum_{n=1}^{N} \left(e^{-r\tau^{(n)}} \max(K - S_{\tau^{(n)}}, 0)\right) \times L(\theta, \tau^{(n)}) \]  

(4.8)

(4.9)

with 4.4

\[ L(\theta, \tau^{(n)}) = \exp \left(-\theta W^{Q^\theta}_{\tau^{(n)}} - \frac{1}{2} \theta^2 \tau^{(n)} \right) \]

and by definition of a Wiener process (i.e. standard Brownian motion)

\[ W^{Q^\theta}_{\tau^{(n)}} = \sum_{i:0 \leq t_i \leq \tau^{(n)}} Z_i \sqrt{t_{i+1} - t_i}, \quad Z_i \sim N(0,1) \text{ I.I.D.} \]

Note that the likelihood ratio \( L(\theta, n) \) should be computed for each individual path \( n = 1, \ldots, N \) depending on the optimal stopping time for that path \( \tau^{(n)} \). Algorithm 2 describes the total procedure in pseudocode.
Algorithm 2: IS-LSM-algorithm

1: for \( n = 1, ..., N \) do
2:     Generate realizations of the GBM using drift \( \theta \), with \( M \) time steps
3:     Set \( \tau = M \)
4: end for
5: for \( t = M-1, ..., 1 \) do
6: for \( n = 1, ..., N \) do
7:     Compute \( a_0, ..., a_M \)
8:     Set \( \hat{F}(t, S(n)(t)) = \sum_{l=0}^{M} \hat{a}_l e_l(S^n(t)) \)
9:     if \( V(t, S(n)(t)) \geq \hat{F}(t, S(n)(t)) \) then
10:        \( \tau^{(n)} = t \)
11: end if
12: end for
13: end for
14: \( V_0 = \frac{1}{N} \sum_{n=1}^{N} V(\tau^{(n)}, S(n)(\tau^{(n)})) \cdot L(\theta, \tau^{(n)}) \)

4.2.1 Convergence of the IS-LSM algorithm

Due to proposition 4.1, there exist a convergence result for the IS-LSM method. If the sequence of basis functions \( e_k((S_t_j))_{k \geq 1} \) is total in the space \( L^2 \) for \( j = 1, ..., M \), then \( \forall \theta \in \mathbb{R}, V_0^{\theta,m} \rightarrow V_0^{\theta} \), as \( m \rightarrow \infty \). Moreover if \( P(\alpha_j \cdot e(S_{t_j}) = Z_{t_j}) = 0 \) for \( j = 1, ..., M \), then \( \forall \theta \in \mathbb{R} \) and \( \forall m \in \mathbb{N}, V_0^{\theta,m,N} \rightarrow V_0^{\theta,m} \) as \( N \rightarrow \infty \). These statements are proven by Moreni in [10].

However, the fact that the non-drifted price \( V_0 \) is not unbiased, it is not clear whether the drifted price \( V_0^{\theta} \) will reflect the "true price". It can also be the case that for some \( \theta \), the price will deviate from the "true price" more than anticipated. Before tackling this problem, one can state that there exist a \( \theta^* \) that minimizes the second moment of the IS-LSM estimator. In order to find the optimal \( \theta \) such that the variance of \((4.9)\) is minimized, we take advantage of the second moment given in equation (2.22). The second moment for the LSM-IS estimator is a function of \( \theta \) given by \( H(\theta) \) such that:

\[
H(\theta) = \mathbb{E} \left[ (f(\tau, S_{\tau}))^2 \exp \left( -2\theta W_{\tau}^{Q^\theta} - \theta^2 \tau \right) \right] \quad (4.10)
\]

and we want to solve the problem:

\[
H(\theta^*) = \min_{\theta \in \mathbb{R}} H(\theta) \quad (4.11)
\]

Note that we require the function above be twice differentiable and the gradient \( \nabla H(\theta) \) needs to have a well behaved representation. The study by Su in [11] uses the infinitesimal perturbation analysis (IPA) for estimating \( \nabla H(\theta) \) which is shown to work well for pricing European options. Here, the same method is used but this time for American options.
It assumed that the likelihood ratio \( L(\theta) \) is piecewise differentiable on \( \Theta \), where \( \Theta \in \mathbb{R} \). By differentiating the expectation in (4.10), the IPA estimator is given by:

\[
f(\tau, S_{\tau}) \frac{\partial L(\theta)}{\partial \theta}
\]

hence,

\[
\frac{\partial L(\theta)}{\partial \theta} = \left( -W_{T(n)}^{Q^\theta} - \theta T(n) \right) e^{-\theta W_{T(n)}^{Q^\theta} - \frac{1}{2} \theta^2 T(n)} = \left( -W_{T(n)}^{Q^\theta} - \theta T(n) \right) L(\theta) = -W_{T(n)}^{Q} L(\theta).
\]

Since the underlying asset prices follows the geometric Brownian motion, the IPA estimator in (4.12) can be replaced with the IPA-Q estimator given by:

\[
f \left( \tau(n), S_{\tau(n)} \right)^2 \left( -W_{\tau(n)}^{Q} \right) L^2(\theta).
\]

To find the optimal \( \theta^* \) we will take usage of the stochastic approximation discussed the next section.

### 4.2.2 Stochastic Approximation for Optimal Drift

The task is to solve a well defined problem that features a parameter on which there is a reasonable control. Stochastic algorithms devises to iteratively update the parameter in question such that it can converge to a some desired value that is optimal in some specific way. This means that the structure of the algorithm and the precise sense of optimality (optimal drift \( \theta \) in this case) varies depending on the applications of the algorithm.

**Definition 4.2.** A *stochastic algorithm* is a random sequence \( \theta_n \in \mathbb{R}, n \geq 0 \) adapted to filtration \( \mathcal{F}_n \) and is written in the form:

\[
\theta_{n+1} = \theta_n + \gamma_{n+1} \left( F(\theta) + U_{n+1} \right), \quad n \geq 0
\]

for a real valued function \( F : \mathbb{R} \to \mathbb{R} \) and a deterministic real valued sequence \( \left( \gamma_n \right)_{n \geq 1} \) satisfying:

\[
\gamma_n > 0, \quad \lim_{n \to \infty} \gamma_n = 0, \quad \sum_{n \geq 1} \gamma_n = \infty
\]

\( U_n \) is an \( \mathcal{F}_n \)-adapted sequence such that \( \mathbb{E}[U_{n+1}|\mathcal{F}_n] = 0 \) for \( n \geq 0 \).

Su in [11], writes the iterative scheme expressed in Definition 4.2 in the form:

\[
\theta_{n+1} = \Pi_{\Theta}(\theta_n - \gamma_n g_n)
\]
where, \( g_n \) represents the gradient \( \nabla H(\theta) \) at \( \theta_n \) and \( \Pi_\Theta \) denotes a projection on \( \Theta \). The way of running this iterative scheme is described by the pseudo-code in Algorithm 3:

**Algorithm 3 Optimal Drift-algorithm**

1: Set \( \theta = \theta_0 \)
2: for \( n = 1, \ldots, N \) do
3:    for \( i = 1, \ldots, N_2 \) do
4:       - Generate a sample paths under (4.3) and store the \( \tau \)-sample path using LSM
5:       - Calculate IPA-Q\(_i\) estimator according to (4.13)
6:    end for
7:    \( g_n(\theta_n) = \frac{1}{N_2} \sum_{i=1}^{N_2} \text{IPA-Q}_i \)
8:    \( \theta_{n+1} = \theta_n - \gamma_n g_n(\theta_n) \)
9:    if \( |\gamma_n g_n(\theta_n) < \epsilon| \) then
10:       break
11: end if
12: Set \( \theta^* = \theta_{n+1} \)
13: end for
14: return \( \theta^* \)

The choice of \( \epsilon \) in the above algorithm is arbitrary but should be a small number in order to guarantee convergence such that \( |\gamma_n g_n(\theta_n) < \epsilon| \) holds for large \( n \) and small \( \epsilon \).

By combining the algorithms described in previous sections, we can state the pseudo-code for the whole pricing engine in two stages.
Algorithm 4 Pricing Engine

1: - STAGE 1
2: Set $\theta = \theta_0$
3: for $n = 1, ..., N_1$ do
4:    for $i = 1, ..., N_2$ do
5:      - Generate sample paths under (4.3) and store the $\tau$-sample path using LSM
6:      - Calculate IPA-Q$_i$ estimator according to (4.13)
7:    end for
8:  $g_n(\theta_n) = \frac{1}{N_2} \sum_{i=1}^{N_2}$ IPA-Q$_i$
9:  $\theta_{n+1} = \theta_n - \gamma_n g_n(\theta_n)$
10: if $|\gamma_n g_n(\theta_n)| < \epsilon$ then
11:    break
12: end if
13: $\theta^* = \theta_{n+1}$
14: end for
15: return $\theta^*$
16: 
17: - STAGE 2
18: for $n = 1, ..., N$ do
19:    Generate realizations of the GBM using drift $\theta^*$, with $M$ time steps
20: Set $\tau = M$
21: end for
22: for $t = M-1, ..., 1$ do
23:    for $n = 1, ..., N$ do
24:        Compute $a_0, ..., a_M$
25:        Set $F(t, S^{(n)}(t)) = \sum_{l=0}^{M} \hat{a}_l e_l(S^{(n)}(t))$
26:        if $V(t, S^{(n)}(t)) \geq F(t, S^{(n)}(t))$ then
27:            $\tau^{(n)} = t$
28:            $L(\theta^*, \tau^{(n)}) = L(\theta^*, t)$
29:        end if
30:    end for
31: end for
32: $V_0 = \frac{1}{N} \sum_{n=1}^{N} V(\tau^{(n)}, S^{(n)}(\tau^{(n)})) \cdot L(\theta^*, \tau^{(n)})$
Chapter 5

Results

Reference price and ratio

Due to the fact that no exact solution of the price is known, one has to determine a relevant "reference price" that could be compared to the price estimated after change of measure. One possible way to generate reference price is to:

1. Set $\theta = 0$ and price according to standard LSM.
2. Set $V_{0}^{0,m,N}$ to be the reference price.

Since the value of $\theta$ is sensitive in the sense that it may or may not give a correct price, one method to use is to compare the price ratio between the reference and drifted price such that the fraction $\frac{V_{0}^\text{IS-LSM}}{V_{0}^\text{LSM}} \approx 1$. This method gives a good overview over how the "true price" behaves with respect to the price estimated by IS-LSM. Figure 5.1 shows an example of the fraction brute forced for different values of $\theta$.

For the ordinary LSM case, the variance was estimated using the sample variance. To make the reference price $V_{0}^{0,m,N}$ as robust possible, a "bucket-simulation" was introduced by simulating $N$ paths and pricing in bucket of $N'$ times, meaning that the total number of simulations was $N \cdot N'$ for some large values of $N$ and $N'$.

Pricing Results

In the first case, second order polynomial of the form $Y = a_0 + a_1 S_t + a_2 S_t^2$ was used as basis functions. These polynomials were already implemented in R’s statistical package extension LSMonteCarlo. In the second case, the Hermite polynomials were used
of the form \( L_n(x) = \exp(-x/2)\frac{d^n}{dx^n}(x^n e^{-x}) \) for \( n = 0, ..., 3 \). The American option was approximated with Bermudan style option with 50 exercise opportunities between 0 \( \leq t \leq T \). Simulations of \( N = 2000 \) were run with \( N' = 100 \) buckets giving a total of \( N \cdot N' = 200,000 \) simulations. The variance reduction was calculated by the fraction 
\[
\frac{\text{Var}(\hat{V}_0)}{\text{Var}(\hat{V}_{IS0})}.
\]

Tables 5.1-5.4 shows the pricing results for LSM and IS-LSM using second order basis functions with spot price \( S_0 = 50 \), volatility \( \sigma \) extending from 10% to 80% and strike prices \( K = 50, 52, 54, 56 \) respectively. The risk free interest was set to \( r = 0.02 \).

On tables 5.5-5.8 the pricing results using Hermite polynomials \textit{ceteris paribus} are shown.

<table>
<thead>
<tr>
<th>( \sigma ) (%)</th>
<th>( \hat{V}_0 ) (SEK)</th>
<th>( \text{Var}(\hat{V}_0) )</th>
<th>( \hat{V}^{IS}_0 ) (SEK)</th>
<th>( \text{Var}(\hat{V}^{IS}_0) )</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.628318</td>
<td>3.502502</td>
<td>1.625061</td>
<td>0.5341099</td>
<td>6.557643</td>
</tr>
<tr>
<td>20</td>
<td>3.584137</td>
<td>19.18572</td>
<td>3.615153</td>
<td>2.579238</td>
<td>7.438523</td>
</tr>
<tr>
<td>30</td>
<td>5.527152</td>
<td>37.72148</td>
<td>5.539492</td>
<td>5.40498</td>
<td>6.979023</td>
</tr>
<tr>
<td>40</td>
<td>7.516779</td>
<td>68.46675</td>
<td>7.576484</td>
<td>9.554525</td>
<td>7.165898</td>
</tr>
<tr>
<td>60</td>
<td>11.33032</td>
<td>124.2308</td>
<td>11.32603</td>
<td>19.85705</td>
<td>6.256257</td>
</tr>
<tr>
<td>70</td>
<td>13.25487</td>
<td>131.6177</td>
<td>13.24744</td>
<td>25.76047</td>
<td>5.109290</td>
</tr>
<tr>
<td>80</td>
<td>15.05618</td>
<td>182.8255</td>
<td>15.09098</td>
<td>31.90902</td>
<td>5.729587</td>
</tr>
</tbody>
</table>

Figures 5.2 and 5.3 shows the convergence behaviour of standard LSM and IS-LSM using 2\(^{nd}\)-order polynomials with \( N = 5000 \) simulations, and pricing parameters \( S_0 = 50 \),

**Table 5.1:** Strike price \( K = 50 \) with 2\(^{nd}\)-order polynomial

![Figure 5.1: A brute-force plot of \( \frac{\hat{V}^{IS}_{LSM}}{\hat{V}_0} \) as a function of \( \theta \) with \( K = 50 \) and \( \sigma = 50\% \).](image)
Table 5.2: Strike price \( K = 52 \) with 2\textsuperscript{nd}-order polynomial

<table>
<thead>
<tr>
<th>( \sigma ) (%)</th>
<th>( \hat{V}_0 ) (SEK)</th>
<th>( \text{Var}(\hat{V}_0) )</th>
<th>( \hat{V}_{0IS} ) (SEK)</th>
<th>( \text{Var}(\hat{V}_{0IS}) )</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2.742791</td>
<td>5.500418</td>
<td>2.741661</td>
<td>0.5341099</td>
<td>3.519454</td>
</tr>
<tr>
<td>20</td>
<td>4.678061</td>
<td>23.58803</td>
<td>4.670319</td>
<td>2.579238</td>
<td>6.078116</td>
</tr>
<tr>
<td>30</td>
<td>6.649863</td>
<td>48.43375</td>
<td>6.623648</td>
<td>5.40498</td>
<td>6.786830</td>
</tr>
<tr>
<td>40</td>
<td>8.639865</td>
<td>67.74878</td>
<td>8.674506</td>
<td>9.554525</td>
<td>5.329002</td>
</tr>
<tr>
<td>50</td>
<td>10.61356</td>
<td>102.1687</td>
<td>10.60917</td>
<td>13.87287</td>
<td>6.910649</td>
</tr>
<tr>
<td>60</td>
<td>12.55942</td>
<td>144.4666</td>
<td>12.54883</td>
<td>19.85705</td>
<td>6.431053</td>
</tr>
<tr>
<td>70</td>
<td>14.48066</td>
<td>165.9735</td>
<td>14.48003</td>
<td>25.76047</td>
<td>5.619173</td>
</tr>
<tr>
<td>80</td>
<td>16.36053</td>
<td>183.049</td>
<td>16.37605</td>
<td>31.90902</td>
<td>5.070108</td>
</tr>
</tbody>
</table>

Table 5.3: Strike price \( K = 54 \) with 2\textsuperscript{nd}-order polynomial

<table>
<thead>
<tr>
<th>( \sigma ) (%)</th>
<th>( \hat{V}_0 ) (SEK)</th>
<th>( \text{Var}(\hat{V}_0) )</th>
<th>( \hat{V}_{0IS} ) (SEK)</th>
<th>( \text{Var}(\hat{V}_{0IS}) )</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4.210878</td>
<td>3.812905</td>
<td>4.217694</td>
<td>3.041304</td>
<td>1.253707</td>
</tr>
<tr>
<td>20</td>
<td>5.936218</td>
<td>21.59282</td>
<td>5.971708</td>
<td>6.066479</td>
<td>3.559366</td>
</tr>
<tr>
<td>30</td>
<td>7.938367</td>
<td>53.66102</td>
<td>7.926904</td>
<td>9.451961</td>
<td>5.677237</td>
</tr>
<tr>
<td>40</td>
<td>9.86214</td>
<td>82.51658</td>
<td>9.870942</td>
<td>14.98291</td>
<td>5.507380</td>
</tr>
<tr>
<td>50</td>
<td>11.89708</td>
<td>110.3234</td>
<td>11.89167</td>
<td>19.19366</td>
<td>5.749708</td>
</tr>
<tr>
<td>60</td>
<td>13.81342</td>
<td>149.0316</td>
<td>13.81165</td>
<td>25.8025</td>
<td>5.775859</td>
</tr>
<tr>
<td>70</td>
<td>15.77979</td>
<td>202.9246</td>
<td>15.78673</td>
<td>36.79025</td>
<td>5.515717</td>
</tr>
<tr>
<td>80</td>
<td>17.64294</td>
<td>206.3911</td>
<td>17.65848</td>
<td>36.40016</td>
<td>5.670060</td>
</tr>
</tbody>
</table>

Table 5.4: Strike price \( K = 56 \) with 2\textsuperscript{nd}-order polynomial

<table>
<thead>
<tr>
<th>( \sigma ) (%)</th>
<th>( \hat{V}_0 ) (SEK)</th>
<th>( \text{Var}(\hat{V}_0) )</th>
<th>( \hat{V}_{0IS} ) (SEK)</th>
<th>( \text{Var}(\hat{V}_{0IS}) )</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6.004848</td>
<td>2.782589</td>
<td>6.008509</td>
<td>5.702775</td>
<td>6.557643</td>
</tr>
<tr>
<td>20</td>
<td>7.330201</td>
<td>25.71255</td>
<td>7.337774</td>
<td>7.855498</td>
<td>3.273192</td>
</tr>
<tr>
<td>40</td>
<td>11.18193</td>
<td>83.24592</td>
<td>11.19667</td>
<td>16.43322</td>
<td>5.065710</td>
</tr>
<tr>
<td>50</td>
<td>13.17314</td>
<td>129.8683</td>
<td>13.18626</td>
<td>25.09518</td>
<td>5.054189</td>
</tr>
<tr>
<td>60</td>
<td>15.10658</td>
<td>141.4808</td>
<td>15.11283</td>
<td>25.11283</td>
<td>9.361635</td>
</tr>
<tr>
<td>70</td>
<td>17.07919</td>
<td>205.1197</td>
<td>17.07342</td>
<td>37.29034</td>
<td>5.500612</td>
</tr>
<tr>
<td>80</td>
<td>19.03812</td>
<td>202.4958</td>
<td>19.03074</td>
<td>43.91871</td>
<td>4.610696</td>
</tr>
</tbody>
</table>

\( K = 50, r = 0.02, \sigma = 10\% \) and \( T = 1 \). In figures 5.4 and 5.5 the same convergence behaviour is shown but with \( K = 54 \).

Figures 5.6 and 5.7 shows the same convergence behaviour of standard LSM and IS-LSM but this time with Hermite-polynomials. In figures 5.8 and 5.9 the same convergence behaviour is shown but with \( K = 54 \).
Table 5.5: Strike price $K = 50$ with Hermite-polynomials.

<table>
<thead>
<tr>
<th>$\sigma$ (%)</th>
<th>$\hat{V}_0$ (SEK)</th>
<th>Var($\hat{V}_0$)</th>
<th>$\hat{V}_0^{IS}$ (SEK)</th>
<th>Var($\hat{V}_0^{IS}$)</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.642307</td>
<td>3.684851</td>
<td>1.67942</td>
<td>0.5719566</td>
<td>6.442536</td>
</tr>
<tr>
<td>20</td>
<td>3.61541</td>
<td>5.367093</td>
<td>3.687093</td>
<td>2.549649</td>
<td>6.037588</td>
</tr>
<tr>
<td>30</td>
<td>5.598096</td>
<td>34.8271</td>
<td>5.613499</td>
<td>5.340667</td>
<td>6.521114</td>
</tr>
<tr>
<td>40</td>
<td>7.543088</td>
<td>76.23854</td>
<td>7.623854</td>
<td>9.891331</td>
<td>6.763538</td>
</tr>
<tr>
<td>60</td>
<td>11.38096</td>
<td>104.6874</td>
<td>11.403</td>
<td>19.70618</td>
<td>5.312415</td>
</tr>
<tr>
<td>70</td>
<td>13.25539</td>
<td>164.7672</td>
<td>13.28357</td>
<td>22.00076</td>
<td>7.489159</td>
</tr>
<tr>
<td>80</td>
<td>15.1085</td>
<td>189.4788</td>
<td>15.1103</td>
<td>29.7741</td>
<td>6.363880</td>
</tr>
</tbody>
</table>

Table 5.6: Strike price $K = 52$ with Hermite-polynomials.

<table>
<thead>
<tr>
<th>$\sigma$ (%)</th>
<th>$\hat{V}_0$ (SEK)</th>
<th>Var($\hat{V}_0$)</th>
<th>$\hat{V}_0^{IS}$ (SEK)</th>
<th>Var($\hat{V}_0^{IS}$)</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2.757571</td>
<td>6.208106</td>
<td>2.744517</td>
<td>2.744517</td>
<td>2.262003</td>
</tr>
<tr>
<td>20</td>
<td>4.729887</td>
<td>25.268483</td>
<td>4.711123</td>
<td>4.075958</td>
<td>6.199945</td>
</tr>
<tr>
<td>30</td>
<td>6.715419</td>
<td>35.679909</td>
<td>6.689882</td>
<td>7.853771</td>
<td>4.543029</td>
</tr>
<tr>
<td>40</td>
<td>8.686990</td>
<td>72.393169</td>
<td>8.685366</td>
<td>12.09995</td>
<td>5.982931</td>
</tr>
<tr>
<td>50</td>
<td>10.670997</td>
<td>110.528897</td>
<td>10.68087</td>
<td>17.51208</td>
<td>6.311580</td>
</tr>
<tr>
<td>60</td>
<td>12.584007</td>
<td>131.801013</td>
<td>12.56325</td>
<td>21.5633</td>
<td>6.112284</td>
</tr>
<tr>
<td>70</td>
<td>14.535953</td>
<td>153.505541</td>
<td>14.54229</td>
<td>30.16194</td>
<td>5.089379</td>
</tr>
<tr>
<td>80</td>
<td>16.392707</td>
<td>213.818094</td>
<td>16.42989</td>
<td>38.5391</td>
<td>5.545951</td>
</tr>
</tbody>
</table>

Table 5.7: Strike price $K = 54$ with Hermite-polynomials.

<table>
<thead>
<tr>
<th>$\sigma$ (%)</th>
<th>$\hat{V}_0$ (SEK)</th>
<th>Var($\hat{V}_0$)</th>
<th>$\hat{V}_0^{IS}$ (SEK)</th>
<th>Var($\hat{V}_0^{IS}$)</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4.228625</td>
<td>4.975201</td>
<td>4.210871</td>
<td>3.202642</td>
<td>1.553468</td>
</tr>
<tr>
<td>20</td>
<td>5.987346</td>
<td>28.069843</td>
<td>5.982083</td>
<td>5.187637</td>
<td>5.410911</td>
</tr>
<tr>
<td>30</td>
<td>7.911848</td>
<td>53.079082</td>
<td>7.899397</td>
<td>9.846293</td>
<td>5.390768</td>
</tr>
<tr>
<td>40</td>
<td>9.982095</td>
<td>74.236456</td>
<td>9.990763</td>
<td>14.10173</td>
<td>5.264351</td>
</tr>
<tr>
<td>50</td>
<td>11.927152</td>
<td>114.319966</td>
<td>11.94185</td>
<td>19.91554</td>
<td>5.740239</td>
</tr>
<tr>
<td>60</td>
<td>13.889657</td>
<td>155.462079</td>
<td>13.88364</td>
<td>23.89334</td>
<td>6.506503</td>
</tr>
<tr>
<td>70</td>
<td>15.828050</td>
<td>172.637771</td>
<td>15.80695</td>
<td>31.29834</td>
<td>5.515876</td>
</tr>
<tr>
<td>80</td>
<td>17.723791</td>
<td>228.226868</td>
<td>17.72223</td>
<td>38.06168</td>
<td>5.996237</td>
</tr>
</tbody>
</table>

Table 5.8: Strike price $K = 56$ with Hermite-polynomials.

<table>
<thead>
<tr>
<th>$\sigma$ (%)</th>
<th>$\hat{V}_0$ (SEK)</th>
<th>Var($\hat{V}_0$)</th>
<th>$\hat{V}_0^{IS}$ (SEK)</th>
<th>Var($\hat{V}_0^{IS}$)</th>
<th>Variance Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6.017468</td>
<td>1.189414</td>
<td>5.99501</td>
<td>6.266122</td>
<td>0.1898166</td>
</tr>
<tr>
<td>20</td>
<td>7.379858</td>
<td>35.090213</td>
<td>7.384168</td>
<td>7.872534</td>
<td>4.4572958</td>
</tr>
<tr>
<td>30</td>
<td>9.248224</td>
<td>68.370972</td>
<td>9.271849</td>
<td>12.81123</td>
<td>5.3368000</td>
</tr>
<tr>
<td>40</td>
<td>11.226492</td>
<td>87.284022</td>
<td>11.19545</td>
<td>17.65529</td>
<td>4.9437886</td>
</tr>
<tr>
<td>50</td>
<td>13.244236</td>
<td>134.337202</td>
<td>13.27871</td>
<td>25.90216</td>
<td>5.1863320</td>
</tr>
<tr>
<td>60</td>
<td>15.208831</td>
<td>166.926922</td>
<td>15.25854</td>
<td>35.36346</td>
<td>4.7203221</td>
</tr>
<tr>
<td>70</td>
<td>17.180185</td>
<td>206.949380</td>
<td>17.33694</td>
<td>35.91842</td>
<td>5.7616504</td>
</tr>
<tr>
<td>80</td>
<td>19.112274</td>
<td>227.486597</td>
<td>19.17055</td>
<td>45.69645</td>
<td>4.9782116</td>
</tr>
</tbody>
</table>
Figure 5.2: Convergence of put option using standard LSM (left) and IS-LSM (right) with 2nd-order polynomials and sample size of \( n = 5000 \). The pricing parameters here are \( S_0 = 50, K = 50, r = 0.02, \sigma = 10\% \) and \( T = 1 \).

Figure 5.3: Convergence of put option using standard LSM (left) and IS-LSM (right) with a sample size of with 2nd-order polynomials and sample size of \( n = 5000 \). The pricing parameters here are \( S_0 = 50, K = 50, r = 0.02, \sigma = 50\% \) and \( T = 1 \).
Figure 5.4: Convergence of put option using standard LSM (left) and IS-LSM (right) with 2nd-order polynomials and sample size of \( n = 5000 \). The pricing parameters here are \( S_0 = 50 \), \( K = 54 \), \( r = 0.02 \), \( \sigma = 10\% \) and \( T = 1 \).

Figure 5.5: Convergence of put option using standard LSM (left) and IS-LSM (right) with 2nd-order polynomials and sample size of \( n = 5000 \). The pricing parameters here are \( S_0 = 50 \), \( K = 54 \), \( r = 0.02 \), \( \sigma = 50\% \) and \( T = 1 \).
Figure 5.6: Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50$, $K = 50$, $r = 0.02$, $\sigma = 10\%$ and $T = 1$.

Figure 5.7: Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50$, $K = 50$, $r = 0.02$, $\sigma = 50\%$ and $T = 1$. 
Figure 5.8: Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50$, $K = 54$, $r = 0.02$, $\sigma = 10\%$ and $T = 1$.

Figure 5.9: Convergence of put option using standard LSM (left) and IS-LSM (right) with Hermite-polynomials and a sample size of $n = 5000$. The pricing parameters here are $S_0 = 50$, $K = 54$, $r = 0.02$, $\sigma = 50\%$ and $T = 1$. 
Results for optimal drift $\theta^*$

The optimal $\theta^*$ was found by running Algorithm 3 which estimates the IPA-Q estimator of the second moment. By empirical methods, the starting values for the iterative scheme given by \((4.16)\) was determined by looking at the reference ratio where the fraction is approximately equal to 1.

Figure 5.10 shows an the reference ratio $\frac{V_{\text{JS-LSM}}}{V_{\text{LSM}}}$ for $\theta \in [-1.17, -1.15]$ with $K = 56$ and $\sigma = 50\%$. This interval corresponds to the reference ratio approximately to one for all $\theta$. Hence, an arbitrary starting value for $\theta$ is chosen between the interval $[-1.17, -1.15]$. The sequence $a_n$ in algorithm 4 is set according to the decreasing sequence $\gamma_n = \frac{0.01}{n}$ for $n = 1, \ldots, n_{\text{MAX}}$. For all simulations $n_{\text{max}} = 10,000$ and the threshold was set to $\epsilon = 10^{-7}$.

![Figure 5.10: Reference ratio vs $\theta \in [-1.17, -1.15]$ with $K = 56$ and $\sigma = 50\%$.](image)

Tables 5.9 and 5.10 shows the optimal values for $\theta^*$ for both 2nd-order polynomials and Hermite-polynomials over extending $\theta$ and $K$. 
Table 5.9: Optimal $\theta$ using second order polynomials.

<table>
<thead>
<tr>
<th>$\sigma$ (%)</th>
<th>$K = 50$</th>
<th>$K = 52$</th>
<th>$K = 54$</th>
<th>$K = 56$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta^*$</td>
<td>$\theta^*$</td>
<td>$\theta^*$</td>
<td>$\theta^*$</td>
</tr>
<tr>
<td>10</td>
<td>-1.61</td>
<td>-1.33</td>
<td>-1.09</td>
<td>-0.86</td>
</tr>
<tr>
<td>20</td>
<td>-1.49</td>
<td>-1.38</td>
<td>-1.25</td>
<td>-1.16</td>
</tr>
<tr>
<td>30</td>
<td>-1.43</td>
<td>-1.35</td>
<td>-1.28</td>
<td>-1.12</td>
</tr>
<tr>
<td>40</td>
<td>-1.34</td>
<td>-1.29</td>
<td>-1.25</td>
<td>-1.2</td>
</tr>
<tr>
<td>50</td>
<td>-1.31</td>
<td>-1.28</td>
<td>-1.22</td>
<td>-1.16</td>
</tr>
<tr>
<td>60</td>
<td>-1.25</td>
<td>-1.21</td>
<td>-1.17</td>
<td>-1.15</td>
</tr>
<tr>
<td>70</td>
<td>-1.19</td>
<td>-1.16</td>
<td>-1.11</td>
<td>-1.1</td>
</tr>
<tr>
<td>80</td>
<td>-1.13</td>
<td>-1.1</td>
<td>-0.94</td>
<td>-1.06</td>
</tr>
</tbody>
</table>

Table 5.10: Optimal $\theta$ using using Hermite-polynomials.

<table>
<thead>
<tr>
<th>$\sigma$ (%)</th>
<th>$K = 50$</th>
<th>$K = 52$</th>
<th>$K = 54$</th>
<th>$K = 56$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta^*$</td>
<td>$\theta^*$</td>
<td>$\theta^*$</td>
<td>$\theta^*$</td>
</tr>
<tr>
<td>10</td>
<td>-1.59</td>
<td>-1.33</td>
<td>-1.09</td>
<td>-0.83</td>
</tr>
<tr>
<td>20</td>
<td>-1.49</td>
<td>-1.37</td>
<td>-1.28</td>
<td>-1.16</td>
</tr>
<tr>
<td>30</td>
<td>-1.43</td>
<td>-1.34</td>
<td>-1.27</td>
<td>-1.2</td>
</tr>
<tr>
<td>40</td>
<td>-1.35</td>
<td>-1.3</td>
<td>-1.25</td>
<td>-1.2</td>
</tr>
<tr>
<td>50</td>
<td>-1.28</td>
<td>-1.25</td>
<td>-1.21</td>
<td>-1.16</td>
</tr>
<tr>
<td>60</td>
<td>-1.24</td>
<td>-1.22</td>
<td>-1.18</td>
<td>-1.11</td>
</tr>
<tr>
<td>70</td>
<td>-1.21</td>
<td>-1.15</td>
<td>-1.15</td>
<td>-1.1</td>
</tr>
<tr>
<td>80</td>
<td>-1.15</td>
<td>-1.1</td>
<td>-1.1</td>
<td>-1.06</td>
</tr>
</tbody>
</table>
Chapter 6

Discussion

This thesis presents a variance reduction method for pricing American options using GBM. The LSM method is studied using two different but yet fundamental basis functions for the backward iterative least-square regression step. By implementing the importance sampling procedure in the LSM method resulting as IS-LSM method, the reduction of variance compared to standard LSM-method is well noticeable.

Comparing the usage of second order polynomials with Hermite-polynomial did not give any significant difference in the price. In this thesis, the underlying asset is a simple one-dimensional GBM, meaning that the difference in price estimation between these two choices of basis functions is not very remarkable. As Longstaff & Schwartz uses more complex instruments in their original paper such as pricing interest rate swaps, the choice of "good" basis functions gets more critical for estimating the continuation value, without adding bias to the pricing step. When looking at the variance for different strike prices and different volatilities, it is clear that the dominating factor for higher variance is a higher volatility. The value of $\sigma$ for which the variance of the put price estimator gets very high is between $\sigma \in [70\%, 80\%]$ for both second order polynomials and Hermite-polynomials.

When simulating the drifted GBM, the choice of a good reference price of the option in terms of the reference ratio is indeed a good method for knowing that the option is priced correctly. A challenge however was the large amount of time consumption in finding the reference ratio for which the fraction is very close to 1. Figure 5.10 shows clearly that it is difficult to choose the right $\theta$ corresponding to the "correct" ratio $\frac{V_{\text{IS-LSM}}}{V_{\text{LSM}}} \approx 1$. In fact, this was computationally challenging since pricing had to be done several times over different values for $\theta$. When running Algorithm 3 there was a deviation for which the sequence $\theta_{n+1} = \theta_n - \gamma_n g_n(\theta_n)$ converged to $\theta^*$. For some runs, $\theta^*$ deviated very much from the empirical $\theta^*$ for which the starting value was chosen. The reason for this may
rely on the fact that each simulated scenario is dependent on the optimal stopping time $\tau^*$, and cashflows in time $0 \leq t \leq T$ can be very different depending on $\tau^*$.

Nevertheless, the difficulty in finding $\theta^*$, the estimated price after running IS-LSM gave a much lesser variance in comparison to LSM. At most, the reduction of variance was up to the factor of 9 time less than the original LSM. But for most cases, the reduction was around 5 times less compared to standard LSM. Both tables 5.9 and 5.10 indicate that for a larger $\sigma$ and $K$, the smaller the value of the drift $\theta$ is needed. As $S_0 = 50$, for strike price $K = 50$, the value of $\theta$ over higher $\sigma$ gets a significant decrease in comparison to cases with $K = 52, ..., 56$.

The results presented by Moreni paper are not comparable in the sense that he uses constant volatility when pricing over different strike prices. Moreover the choice of interest rate is different and the models are multidimensional. However, for higher $K$ the drift $\theta$ gets smaller similarly to tables 5.9 and 5.10.

The way of implementing a good pricing engine is relevant. For this thesis all the implementation was written in $\mathcal{R}$ which is optimized for vector calculation and is not very efficient in handling $<$for$>$ loops. Since some parts of the LSM-algorithm requires using $<$for$>$ loops, the total pricing engine for finding the optimal drift $\theta$ increased the computational time dramatically which limited the efficiency of both time saving and CPU usage. For this reason a high-level programming language is recommended.

Even if in this thesis there is no comparison presented for the choice of discretization method for scenario generation, the usage of Millstein-1 scheme compared to Euler-Maryama during test-simulations was slower. This is due to the extra term given in (3.6). This should have been considered since Euler-Maryama scheme would most likely be equivalent but more efficient.

Important sampling is an efficient method for reducing variance of an estimator. This thesis presented the change of drift in the underlying GBM by means of the Girsanov theorem. The pricing of American put option on a stock price using IS-LSM is efficient. However, this method may even be a better alternative for more complex instruments with early exercise opportunity rather than a single stock based on GBM.
Appendix A

Essential supernum

This explanation is taken from a short course in optimal stopping problem by Damien Lamberton in [8], where the details about the Snell envelope is also well explained. Here we only take into account the essential supernum which occurs in equation (2.7).

It is well known that if \((X_n)_{n \in \mathbb{N}}\) is a sequence of real valued random variables, \(\sup_{n \in \mathbb{N}} X_n\) is a random variable (with values in \(\mathbb{R} \cup \{+\infty\}\)). When uncountable families of random variables have to be considered, as occurs in the theory of optimal stopping, the notion of essential upper bound is needed.

**Theorem A.1.** Let \((X_i)_{i \in I}\) be a family of real valued random variables (with a possibly uncountable index set \(I\)). There exists a random variable \(\bar{X}\) with values in \(\bar{\mathbb{R}}\), which is unique up to null events, such that

1. For all \(i \in I\), \(X_i \leq X\) a.s.
2. If \(X\) is a random variable with values in \(\bar{\mathbb{R}}\) satisfying \(X_i \leq X\) a.s., for all \(i \in I\), then \(\bar{X} \leq X\) a.s.

Moreover, there is a countable subset \(J\) of \(I\) such that \(\bar{X} = \sup_{i \in J} X_i\) a.s.

The random variable \(\bar{X}\) is called the essential upper bound (or essential supremum) of the family \((X_i)_{i \in I}\) and is denoted by \(\text{ess sup}_{i \in I} X_i\).

**Proof.** By using a one-to-one increasing mapping from \(\bar{\mathbb{R}}\) onto \([0, 1]\), we can assume that the \(X_i\)'s take on values in \([0, 1]\). Now, given a countable subset \(J\) of \(I\), set

\[\bar{X}_J = \sup_{i \in J} X_i.\]
This defines, for each $J$, a random variable with values in $[0, 1]$. Denote by $\mathcal{P}_0$ the set of all countable subsets of $I$ and let:

$$
\alpha = \sup_{J \in \mathcal{P}_0} \mathbb{E}\bar{X}_J.
$$

Consider a sequence $(J_n)_{n \in \mathbb{N}}$ of elements in $\mathcal{P}_0$ such that $\lim_{n \to \infty} \mathbb{E}\bar{X}_{J_n} = \alpha$. The union $J^* = \bigcup_{n \in \mathbb{N}} J_n$ is a countable subset of $I$ and we have $\alpha = \mathbb{E}\bar{X}_{J^*}$. It can now be proved that the random variable $\bar{X} = \bar{X}_{J^*}$ satisfies the required conditions. First fix $i \in I$. The set $J^* \cup \{i\}$ is a countable subset of $I$ and $\bar{X}_{J^* \cup \{i\}} = \max(\bar{X}, X_i)$. Therefore, $\mathbb{E}\max(\bar{X}, X_i) \leq \mathbb{E}\bar{X}$ and $\max(\bar{X}, X_i) = \bar{X}$ a.s., which means that $X_i \leq \bar{X}$ a.s.. Now consider a random variable $X$ such that $X \geq X_i$ for all $i \in I$. Since $J^*$ is countable, we have that $X \geq \sup_{i \in J^*} X_i = \bar{X}$ a.s.. \hfill \Box
Bibliography


