5-2015

Monte Carlo Methods in Finance

Je Guk Kim
University of Tennessee - Knoxville, jkim74@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Je Guk Kim entitled "Monte Carlo Methods in Finance." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Management Science.

Russell Zaretski, Major Professor

We have read this dissertation and recommend its acceptance:

Xia Chen, Bogdan Bichescu, Xiaobing Feng

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Monte Carlo Methods in Finance

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Je Guk Kim

May 2015

Dedicated to My Parents

and My Teachers

Acknowledgement

I first would like to express my best gratitude to my advisor, Dr. Russell Zaretski. His generous understanding for student was a critical factor in completing this dissertation. I also would like to thank to my dissertation committee members; Dr. Xia Chen, Dr. Bogdan Bichescu, and Dr. Xiaobing Feng. Dr. Sihun Jo at Korea Institute for Advanced Study also has my sincere gratitude for fruitful discussion during preparation of the second part of this dissertation. I also appreciate Dr. Francis Clarke at Univerite Claude Bernard Lyon 1 for his invaluable help for the first part of the dissertation. Finally, I want to point out that I owe everything to my parents and aunt.

Abstract

Monte Carlo method has received significant consideration from the context of quantitative finance mainly due to its ease of implementation for complex problems in the field. Among topics of its application to finance, we address two topics: (1) optimal importance sampling for the Laplace transform of exponential Brownian functionals and (2) analysis on the convergence of quasi-regression method for pricing American option. In the first part of this dissertation, we present an asymptotically optimal importance sampling method for Monte Carlo simulation of the Laplace transform of exponential Brownian functionals via Large deviations principle and calculus of variations the closed form solutions of which induces an optimal measure for sampling. Some numerical tests are conducted through the Dothan bond pricing model, which shows the method achieves a significant variance reduction. Secondly, we study the convergence of a quasi-regression Monte Carlo method proposed by Glasserman and Yu (2004) that is a variant of least-squares method proposed by Longstaff and Schwartz (2001) for pricing American option. Glasserman and Yu (2004) showed that the method converges to an approximation to the true price of American option with critical relations between the number of paths simulated and the number of basis functions for two examples: Brownian motion and geometric Brownian motion. We show that the method surely converges to the true price of American option even under multiple underlying assets and prove a more promising critical relation between the number of basis functions and the number of simulations in the previous study holds. Finally, we propose a rate of convergence of the method.

Table of Contents

List of Tables

Part I.

INTRODUCTION

This dissertation studies sampling-based computational methods in the context of financial applications. Stochastic simulation has played a prominent role in computing quantities of interest across a variety of academic disciplines. The Monte Carlo method has, in particular, attracted attention from the researchers and practitioners in the area of quantitative finance because of its ease of implementation and stability to model perturbations for complex problems. The main concern is computing the expected value of a random variable defined on a proper probability space. Monte Carlo methods consist of two steps: (i) simulate the random quantity of interest and (ii) estimate the expected value via sample average of simulated variates.

Although Monte Carlo simulation is often easy to apply, it often displays a much slower convergence rate than traditional numerical methods. The variance of Monte Carlo estimates is $\frac{1}{N}$ where N represents the number of samples simulated. Then, since the size of the confidence interval is proportional to the standard deviation, the smaller variance means a more efficient implementation of the method. Therefore, it is worthwhile studying techniques of variance reduction for Monte Carlo simulation. The popular methods for variance reduction include antithetic variates, control variates, stratified sampling, variance reduction by conditioned sampling and importance sampling.

The first question of this dissertation is how to reduce variance and therefore accelerate the computing of Monte Carlo estimates of the Laplace transform of exponential Brownian functionals. These quantities play an important role in statistical physics and mathematical finance. We focus on the use of importance sampling. The intuition behind importance sampling is to make a change of probability measure so as to give high probabilities to events important in computing the quantity of interest. Although the idea is simple, it is never a trivial problem to make a proper change of measure resulting in variance reduction.

Consider a problem of estimating $E_P[F]$ where F is a random variable and the subscript P indicates the probability measure with respect to which the expected valued is computed. Importance sampling makes changes of both the probability P and the random variable F to reduce the variance while still estimating the quantity of interest. Specifically, if Q is a probability measure equivalent to P, then $E_P[F] = E_Q[F\frac{dP}{dQ}]$ via the Radon-Nikodym derivative $\frac{dP}{dQ}$. Thus, a good choice of Q has to reduce significantly the variance of $F\frac{dP}{dQ}$. Equivalently, noting

$$Var_Q(F\frac{dP}{dQ}) = E_P[F^2\frac{dP}{dQ}] - E_P[F]^2,$$

we seek to a Radon-Nikodym derivative $\frac{dP}{dQ}$ with which the expected value of $F^2\frac{dP}{dQ}$ is minimized.

It is easy to see that the Radon-Nikodym derivative $\frac{dQ}{dP} = \frac{F}{E_P[F]}$ results in the optimal change of measure since it gives the variance of zero. However, it includes the unknown quantity $E_P[F]$ to be estimated. It thus is of no use. Alternatively, one could try to solve a sub-optimization problem over a smaller class of equivalent probabilities excluding the optimal measure. While this attempt is intuitively natural it is often the case that the problem admits no closed-form solutions.

Rather than solving the sub-optimization problem, Glasserman et al. (1999) considered an asymptotic approximation of the second moment and minimizes the approximation over a set of probabilities equivalent to the original one in finite dimensional setting. Guasoni and Robertson (2008) extend the method to the infinite dimensional setting. We apply this method to the problem of Monte Carlo estimation of the Laplace transform of exponential Brownian functionals. Large deviations theory translates the asymptotic of the second moment into a calculus of variations problem. Fortunately, the calculus of variations problem admits closed-form solutions. Then, the Cameron-Martin theorem tells us that the solution for the calculus of variations problem gives us an asymptotically optimal importance sampling measure for Monte Carlo simulation for the Laplace transform of exponential Brownian functionals.

In addition to studying how to accelerate the speed of Monte Carlo simulation, we address the convergence of a Monte Carlo method for American option pricing. An American option differs from an European option in that the holder may select the time at which to exercise the option. A theoretical result says that one has to appeal to an optimal exercise policy for the fair price of an American option. Hence, the problem of pricing an American option reduces to solving an optimal stopping problem (see Glasserman (2004) in references in the first part).

The early-exercise feature of an American option makes this optimal stopping problem analytically intractable and puts an emphasis on numerical methods. Numerical methods include binomial trees and finite-difference methods. In low dimensional cases these methods are fast. That is, if the number of underlying assets are less than three, then the methods are effective. But many problems in real world have dimensions greater than three, which has sparked studies on methods that are not affected by the "curse of dimensionality". Monte Carlo simulation is well-known for its independence from dimension. However, it is not obvious how to apply Monte Carlo simulation in the context of the optimal stopping problem inherent in the course of American option pricing.

The algorithm by Longstaff and Schwartz (2001) may be the most popular method for Monte Carlo method for pricing American option, which is evidenced by the number of citations to this article. It builds on simulation, regression and backward induction. One can find the details of the method in many monographs on stochastic simulation for finance (for example, see Glasserman (2004) in the references of the second part).

While the implementation of the method is easy the analysis of convergence is difficult enough to motivate numerous authors to study it (see the references for the second part). Using Hermite polynomials and multiples of the powers as basis functions for Brownian motion and geometric Brownian motion respectively Glasserman and Yu (2004) showed that the method converges to an approximation of the true price of an American option under a critical relation between the number of basis functions and the number of Monte Carlo simulations if there is a single asset.

We note the method in Glasserman and Yu (2004) is a quasi-regression method, a variant of the least-squares method. In particular, the method differs from Longstaff and Schwartz method in how the sample paths of underlying assets are generated and the use of exact matrix in estimating coefficients of the basis functions. In this dissertation, we revisit the convergence of the quasi-regression method and address more promising results on its convergence. With help from polynomial chaos expansion for L^2 random variables, we show for the case of geometric Brownian motion that the quasi-regression method converges to the true value under a multi-asset environment with improved relation between the number of basis functions and the number of Monte Carlo simulations for the case of geometric Brownian motion. Further, the rate of convergence of the method is provided.

The dissertation is organized as follows. Part II addresses the problem of optimal importance sampling of the Laplace transform of exponential Brownian functional. Convergence of the quasi-regression Monte Carlo method is dealt with in the part III. Finally, the results of the thesis are summarized in part IV.

Part II.

OPTIMAL IMPORTANCE SAMPLING FOR LAPLACE TRANSFORMS OF EXPONENTIAL BROWNIAN FUNCTIONALS

2.1. Abstract

An asymptotically optimal importance sampling method for Monte Carlo simulation of the Laplace transform of exponential Brownian functionals is developed. We appeal to the theory of large deviations, which converts the problem of finding a measure for importance sampling into a calculus of variations problem and leads to a closed-form solution. Moreover, a path to test regularity of optimal drift which is necessary in implementing the proposed method is also addressed. Significant variance reduction in comparison to the crude Monte Carlo simulations is demonstrated through numerical tests in the Dothan bond pricing model.

2.2. Introduction

In this part we develop a Monte Carlo method for estimating the Laplace transform of exponential Brownian functional of the form

$$\int_0^T e^{\sigma W(t)+\rho\sigma^2 t/2}dt$$

where $\{W(s)\}_{x\in\mathbb{R}^+}$ is a standard Brownian motion, $\rho \in \mathbb{R}$ and $\sigma, T > 0$. Exponential Brownian functionals are an important quantities in many academic disciplines. The list of the disciplines includes mathematical finance and statistical physics of disordered systems (See Comet et al. (1998), Linetsky (2004), Pintoux and Privault (2011), and references therein). Computing the Laplace transform

$$L(a,T) = E[exp(-a\int_0^T e^{\sigma W(t)+\rho\sigma^2 t/2}dt)],\ a > 0,\ \rho \in \mathbb{R}$$

is in turn an important problem in such academic fields.

In the literature many papers are devoted to finding expressions such as integral representations for the Laplace transform which aim at facilitating numerical to calculate these quantities. There are two main applications in such studies: PDE and probability models (see Pintoux and Privault (2011) and Privault and Uy (2013) for detailed review). However, the representations are vulnerable to model perturbations and difficult to compute for some values of a (see Guasoni and Robertson (2008) and Privault and Uy (2013)). In such cases, Monte Carlo simulation may prove an easy solution.

However, Monte Carlo method suffers from its slow convergence rate. Thus, it is natural to study methods to make it more efficient, which amounts to studying variance reduction techniques. In Privault and Uy (2013) the author presented a Monte Carlo simulation method by choosing an integral representation where the random quantity further follows the generalized hyperbolic secant distribution with restricted $\rho > 0$. They also have presented importance sampling and control variate methods to reduce variance based on likelihood ratio with Beta function. However, numerical results there show that the performance of their methods offers improvement only in restricted regions of T and $\rho > 0$.

We present a more promising and convenient importance sampling method, *asymptocially optimal importance sampling*, which is free from the issues in Privault and Uy (2013). We apply the method proposed by Glasserman et al. (1999) for finite dimensional case and later extended to infinite dimensional case by Guasoni and Roberson (2008). The method builds on the theory of large deviations, Cameron Martin theorem, and calculus of variations. The large deviations principle results in a calculus of variations problem the solution of which in turn induces a probability measure for importance sampling.

The main contribution of this chapter is that the resulting Euler equation admits closed-form solutions. It dramatically improves the efficiency of Monte Carlo estimation in comparison to standard Monte Carlo with parsimonious computational overhead. Moreover, considering the fact that the regularity of optimal drift is important in implementing the method, a path to test its regularity is provided. Numerical tests applying the method to the Dothan bond pricing model demonstrate the method is highly effective in variance reduction in comparison to the crude Monte Carlo method.

2.3. Importance Sampling

Consider the general problem of estimating $c \equiv E[F(G)]$ with a probability space $(\Omega, \mathcal{F}, P)$ where $G : \Omega \to X$ is a random variable and $F : X \to \mathbb{R}$ is a measurable function. We assume that X is a topological space with Borel *sigma*-algebra. Suppose there exists a measure Q equivalent to P. Then, we have, through Radon-Nikodym derivative $\frac{dP}{dQ}$, $E_P[F(G)] = E_Q[F(G)\frac{dP}{dQ}]$ where the subscripts indicate the probability measure for the integral. The aim of importance sampling is to find an equivalent measure Q for which Monte Carlo estimation of c achieves a significant variance reduction in comparison to the original measure P, which therefore results in a shorter confidence interval and more accurate estimator. More specifically, an optimal importance distribution Q should minimize the variance under the measure Q itself,

$$Var_Q[F(G)\frac{dP}{dQ}] = E_P[F(G)^2\frac{dP}{dQ}] - E_P[F(G)]^2.$$

Noting the key factor to variance reduction is to make small the valued of $E_P[F(G)^2\frac{dP}{dQ}]$, Glassserman et al. (1999) proposed the use of a large deviations result for asymptotic integrals to identify an effective change of measure for estimating small perturbations of the second moment in the variance in finite dimensional setting. Guasoni and Robertson (2008), later, extended the method to infinite dimensional setting, which is also effective for the present chapter.

To be more precise, recall the quantity of interest,

$$E_P[exp\{-a\int_0^T e^{\sigma W(t)+\rho\sigma^2 t/2}\}].$$

The standard Brownian motion induces Wiener measure P on the space of continuous functions on $[0, T]$. Then, the Cameron-Martin theorem identifies the class of drifts that induce measures equivalent to the one by a Brownian motion. Specifically, denote by $\mathbb{W}_T$ the Wiener space

$$\mathrm{\mathbb{W}_T \equiv \{x \in C([0,T],\mathbb{R}) : x(0) = 0\}.}$$

This space is equipped with the sup-norm topology. The set of candidates for drifts inducing equivalent measures is the Cameron-Martin space

$$\mathbb{H}_T \equiv \{h \in AC[0,T] : h(0) = 0, \int_0^T h^{'}(t)^2 dt < \infty\}.$$

Then, Cameron-Martin theorem (see Peres and Morters (2010)) tells us that each $h \in \mathbb{H}_T$ induces a measure Q_h equivalent to P via Radon-Nikodym derivative

$$\frac{dQ^h}{dP}(W) = exp\{-\frac{1}{2}\int_0^T h^{'}(t)^2 dt + \int_0^T h^{'} dW\}.$$

Then, we have

$$\begin{aligned} &E_P[exp\{-a \textstyle\int_0^T e^{\sigma W(t)+\rho\sigma^2 t/2} dt\}] \\ &= E_{Q^h}[exp\{-a \textstyle\int_0^T e^{\sigma W(t)+\rho\sigma^2 t/2} dt\} \frac{dP}{dQ^h}] \\ &= E_{Q^h}[exp\{-a \textstyle\int_0^T e^{\sigma W(t)+\rho\sigma^2 t/2} dt - \int_0^T h^{'}(t) dW(t) + \frac{1}{2}\int_0^T h^{'}(t)^2 dt\}] \end{aligned}$$

Therefore, since $\tilde{W} = W - h$ is a standard Brownian motion under Q_h, the estimator, in implementing Monte Carlo simulation, is

$$exp\{-a \textstyle\int_0^T e^{\sigma(\tilde{W}(t)+h(t))+\rho\sigma^2 t/2} dt - \int_0^T h^{'}(t) d\tilde{W}(t) - \frac{1}{2}\int_0^T h^{'}(t)^2 dt\} \ .$$

In section 2.4 below, we choose a drift h for variance reduction via importance sampling for the Laplace transform of exponential Brownian functional through the method proposed in Guasoni and Robertson (2008), which will be detailed below.

2.4. An Asymptotically Optimal Estimator and Large Deviations

We introduce basic definitions and theorems on the theory of large deviations needed to derive an optimal drift for importance sampling. We refer readers to Dembo and Zeitouni (1998) and Kalenberg (2002) for detailed development of the theory. The theory of large deviations concerns the asymptotic behavior of small probabilities on an exponential scale. The first rigorous result on large deviations was due to Cramer. He explored the problem of choosing premium for an insurance contract from the perspective of the insurance company. Equivalently, he explored the asymptotic behavior of probabilities of the tail of the empirical mean of $i.i.d$ random variables.

Later, his idea of the tail probabilities had generalized to the setting for the studies of the asymptotic behavior of probabilities of the rare events on an exponential scale. A typical setting for large deviations principle as follows; let $\mathcal{X}$ be a Hausdorff topological space equipped with its Borel σ-algebra $\mathcal{B}$ and $\mathcal{M}_1(\mathcal{X})$ be a space of probability measures on the measurable space. Consider a class $\{\mu_\epsilon\}_{\epsilon\in\mathbb{R}^+}$ of $\mathcal{M}_1(\mathcal{X})$. What we have interest in here is the limit behavior of the probabilities μ_ϵ assigns to an outcome $x \in \mathcal{X}$ on an exponential scale as ϵ goes smaller and smaller. However, it is often not the case that we assign a probability to an individual element x in the space of interest. To get around this impracticability one considers probabilities assigned to sets in the space while introducing $ratefunction$ to identify the impact of each outcome x on the asymptotic behavior. However, it is still problematic to consider all subsets of the space $\mathcal{X}$ for studying large deviations. The reasonable subclass of $\mathcal{X}$ for it is the set of all measurable sets. Here is the precise definition of an asymptotic behavior of small probabilities on an exponential scale.

Definition 2.1. Let $(X, \mathcal{B})$ be a metric space with its Borel σ-algebra, and consider a lower semicontinuous function $I : X \to [0, +\infty]$. A family of measures $\{\mu_\epsilon\}_{\epsilon\in(0,\delta)}$ satisfies a *large deviation principle with good rate function* I if

(i) $\{x \in X \ : \ I(x) \leq \alpha\}$ is compact for all $\alpha \in \mathbb{R}$,

(ii) For all sets $A \in \mathcal{B}$,

$$-\underset{x\in A^o}{inf}\, I(x) \le \underset{\epsilon\to 0}{liminf}\ \epsilon\ log\mu_\epsilon(A) \le \underset{\epsilon\to 0}{limsup}\ \epsilon\ log\mu_\epsilon(A) \le -\underset{x\in \bar{A}}{inf} I(x).$$

On the Wiener space the following asymptotic of probabilities holds.

Theorem 2.2. *(Schilder) Let $X = \mathbb{W}_T$ and μ_ϵ be the probability on $\mathbb{W}_T$ induced by the process $\sqrt{\epsilon}W$, where W is a standard Brownian motion. Then $(\mu_\epsilon)_{\epsilon\in(0,\delta)}$ satisfies a large deviation principle with good rate function*

$$I(x) = \begin{cases} \frac{1}{2}\int_0^T x'(t)^2 dt, & if\ x \in \mathbb{H}_T, \\ +\infty, & if\ x \in \mathbb{W}_T \setminus \mathbb{H}_T \end{cases}.$$

Here is another theorem for asymptotic of integrals.

Theorem 2.3. *(Varadhan) Let $(Z_\epsilon)_{\epsilon\in(0,\delta)}$ be a family of X-valued random variables, whose laws $\mu_\epsilon = Z_\epsilon(P)$ satisfy a large deviations principle with good rate function I. If $H : X \to \mathbb{R}$ is a continuous function which satisfies*

$$\underset{\epsilon\to 0}{limsup}\, \epsilon\, log E[exp(\frac{\alpha}{\epsilon}H(Z_\epsilon))] < \infty,$$

for some $\alpha > 1$, then

$$\underset{\epsilon\to 0}{lim}\, \epsilon\, log E[exp(\frac{1}{\epsilon}H(Z_\epsilon))] = \underset{x\in X}{sup}(H(x) - I(x)).$$

We introduce definitions proposed in Glasserman et al. (1999) with which we select a drift h for variance reduction. Glasserman et al. (1999) considered a more general problem of estimating

$$\alpha(\epsilon) = E_P[e^{F(\sqrt{\epsilon}W)/\epsilon}], \qquad \epsilon > 0$$

which includes the original problem as a specific case with $\epsilon = 1$. We will conduct an analysis of the limit behavior of the second moment in variance of estimators of $\alpha(\epsilon)$ when $\epsilon \to 0$

with help from Schilder and Varadhan' s theorems. Here is the first definition.

Definition 2.4. A family of estimators $\{\hat{\alpha}(\epsilon)\}$ is *asymptotically relatively unbiased* if

$$\frac{E_{Q(\epsilon)}[\hat{\alpha}(\epsilon)] - \alpha(\epsilon)}{\alpha(\epsilon)} \to 0 \; as \; \epsilon \to 0$$

where the subscript $Q(\epsilon)$ represents the measure for the expectations.

For comparisons among such estimators we utilize their second moments as in the following definition.

Definition 2.5. A family of asymptotically relatively unbiased estimators $\{\hat{\alpha}_0(\epsilon)\}$ is *asymptotically optimal if*

$$\underset{\epsilon \to 0}{limsup}\, \epsilon \, log E_{Q(\epsilon)}[\hat{\alpha}_0^2(\epsilon)] = \underset{\{\hat{\alpha}(\epsilon)\}}{inf} \; \underset{\epsilon \to 0}{limsup}\, \epsilon \, log E_{Q(\epsilon)}[\hat{\alpha}^2(\epsilon)],$$

the infimum taken over all $\{\hat{\alpha}(\epsilon)\}$ satisfying above definition.

Note the degenerate estimator $\hat{\alpha}(\epsilon) \equiv \alpha(\epsilon)$ is trivially asymptotically optimal and we get

$$\underset{\epsilon \to 0}{limsup}\, \epsilon \, log \alpha^2(\epsilon) = 2 \underset{\epsilon \to 0}{limsup}\, \epsilon \, log \alpha(\epsilon).$$

Thus, an asymptotically optimal estimator achieves twice the exponential rate of $\alpha(\epsilon)$ itself as pointed out in Glasserman et al. (1999). We specify the setting for our problem. Define $F : \mathbb{W}_T \to \mathbb{R}$ by

$$F(x) = -a \int_0^T e^{\sigma x(t) + \rho \sigma^2 t/2} dt,$$

and also define $F_h : \mathbb{W}_T \to \mathbb{R}$ by

$$F_h(x) = 2F(x) - \int_0^T h'(t) dx(t) + \frac{1}{2} \int_0^T h'(t)^2 dt.$$

Now, we address properties of F to be used later.

Theorem 2.6. *F is continuous on the Wiener space $\mathbb{W}_T$ and strictly concave.*

(Proof) Suppose $x_n \to x$ in sup-norm. Note

$$\begin{aligned}
\left|\int_0^T e^{\sigma x_n(s)+\rho\sigma^2 s/2}ds - \int_0^T e^{\sigma x(s)+\rho\sigma^2 s/2}ds\right| &= \left|\int_0^T (e^{\sigma x_n(s)} - e^{\sigma x(s)})e^{\rho\sigma^2 s/2}ds\right| \\
&\leq c\int_0^T \left|e^{\sigma x_n(s)} - e^{\sigma x(s)}\right| ds \\
&= c\int_0^T e^{\sigma x(s)} \left|e^{\sigma x_n(s)-\sigma x(s)} - 1\right| ds \\
&\leq \tilde{c}\int_0^T \left|e^{\sigma x_n(s)-\sigma x(s)} - 1\right| ds
\end{aligned}.$$

Thus, by dominated convergence theorem, F is continuous. Also, note that

$$\begin{aligned}
\int_0^T e^{\sigma(\frac{1}{2}x_1(s)+\frac{1}{2}x_2(s))+\rho\sigma^2 s/2}ds &= \int_0^T e^{\frac{1}{2}(\sigma x_1(s)+\rho\sigma^2 s/2)}e^{\frac{1}{2}(\sigma x_2(s)+\rho\sigma^2 s/2)}ds \\
&\leq \int_0^T \frac{1}{2}[e^{\sigma x_1(s)+\rho\sigma^2 s/2} + e^{\sigma x_2(s)+\rho\sigma^2 s/2}]ds \\
&= \frac{1}{2}\int_0^T e^{\sigma x_1(s)+\rho\sigma^2 s/2}ds + \frac{1}{2}\int_0^T e^{\sigma x_2(s)+\rho\sigma^2 s/2}ds
\end{aligned}.$$

Therefore, since F is continuous, F is also strictly concave.

Note F satisfies the conditions in Varadhan's theorem. Hence, by Schilder theorem, we have

$$2\underset{\epsilon\to 0}{limsup}\ \epsilon\ log\alpha(\epsilon) = 2\sup_{x\in\mathbb{H}_T}(F(x) - \frac{1}{2}\int_0^T x'(t)^2 dt).$$

Now we consider this optimization problem;

$$\sup_{x\in\mathbb{H}_T}(-2a\int_0^T e^{\sigma x(t)+\rho\sigma^2 t/2}dt - \int_0^T x'(t)^2 dt),$$

equivalently,

$$-\inf_{x\in\mathbb{H}_T}(\int_0^T 2ae^{\sigma x(t)+\rho\sigma^2 t/2}+x^{'}(t)^2 dt).$$

Let us consider this problem in larger set $AC[0,T]$;

$$\min_{AC[0,T]}\int_0^T 2ae^{\sigma x(t)+\rho\sigma^2 t/2}+x^{'}(t)^2 dt,\ \ x(0)=0,$$

equivalently,

$$\min_{AC[0,T]}\int_0^T 2ae^{\sigma x(t)+\rho\sigma^2 t/2}+x^{'}(t)^2 dt,\ \ x(0)=0,\ x^{'}(T)=0 \qquad (2.4.1)$$

by noting that any solution to the problem satisfies the transversality condition, $x^{'}(T)=0$. This problem of calculus of variations admits a unique and C^∞ solution h^*.

Theorem 2.7. *The problem (2.4.1) admits a unique C^∞ solution.*

(Proof) Let $\Lambda(t,x,\nu)=2ae^{\sigma x+\rho\sigma^2 t/2}+\nu^2$. Then, since Λ is coercive of degree 2, by Tonelli theorem, there exists a solution $h^*\in AC[0,T]$. Moreover, the function $G(x)=\int_0^T\Lambda(t,x(t),x^{'}(t))dt$ is strictly convex, h^* is a unique solution. Now we consider the regularity of the solution h^*. Note $\Lambda_x=2a\sigma e^{\sigma x+\rho\sigma^2 t/2}$ and $\Lambda_\nu=2\nu$. It is obvious h^* satisfies the Euler equation in integral form. Then, using $\theta(t)=t^2$ as a test function we see that Λ satisfies the Nagumo growth condition. Hence, h^* is Lipschitz. Finally, since Λ is of class C^∞ and $\Lambda_{\nu\nu}$ is positive definite, by the Hilbert-Weierstrass theorem, h^* is also of C^∞.

In the proof above we do not appeal to theorem 3.6 in Guasoni and Robertson (2008), which says nothing about the regularity of solutions at the cost of generality. However, we surely need a proper regularity of solutions to at least guarantee that its derivatives are of bounded variation. In fact, we need more regularity considering the nature of the calculus of variations problem. In particular, the path to regularity of optimal drift above will be helpful to deal with other specific problems that do not allow a closed-form solution and one must appeal to a numerical solution.

Note h^* is also the solution of the original problem and $(h^*)^{'}$ is of bounded variation. We also note for any $h^{'}$ of bounded variation and $\alpha>1$, F_h is continuous on $\mathbb{W}_T$ with respect to sup-norm and satisfies the condition in Varadhan's theorem (see lemma 7.6 in Guasoni and Robertson (2008)).

Let us consider the estimators

$$\hat{\alpha}_h(\epsilon) = exp\{\epsilon^{-1} F(\sqrt{\epsilon} W) - \int_0^T \frac{h^{'}(t)}{\sqrt{\epsilon}} dW(t) + \frac{1}{2}\int_0^T \frac{h^{'}(t)^2}{\epsilon} dt\}.$$

Denote $L(h)$ by

$$L(h) := \underset{\epsilon \to 0}{limsup}\ \epsilon\ log E_{Q^{h/\sqrt{\epsilon}}}[\hat{\alpha}_h^2(\epsilon)]$$

$$= \underset{\epsilon \to 0}{limsup}\ \epsilon\ log E_P[exp\{\frac{1}{\epsilon}(2F(\sqrt{\epsilon}W) - \int_0^T \sqrt{\epsilon} h^{'}(t) dW(t) + \frac{1}{2}\int_0^T h^{'}(t)^2 dt)\}].$$

Then, by Schilder's theorem, we have

$$L(h) = \sup_{x \in \mathbb{H}_T} (2F(x) + \frac{1}{2}\int_0^T (x^{'}(t) - h^{'}(t))^2 dt - \int_0^T x^{'}(t)^2 dt).$$

if $h^{'}$ is of bounded variation (see lemma 7.6 in Guasoni and Robertson (2008)).

We want to show that

$$L(h^*) = 2F(h^*) - \int_0^T (h^*)^{'}(t)^2 dt,$$

which means the family $\{\hat{\alpha}_{h^*}(\epsilon)\}$ is asymptotically optimal and, in this sense, h^* is an optimal drift for the change of measure for importance sampling of the Laplace transform of exponential Brownian functional.

Theorem 2.8. *The family of estimators* $\{\hat{\alpha}_{h^*}(\epsilon)\}$ *is asymptotically optimal.*

(Proof) We note that the problem

$$\sup_{x \in \mathbb{H}_T} (2F(x) + \frac{1}{2}\int_0^T (x^{'}(t) - h^{'}(t))^2 dt - \int_0^T x^{'}(t)^2 dt)$$

has a maximizer for all $h \in \mathbb{H}_T$ *(see theorem 3.6 in Guasoni and Robertson (2008)). Moreover, since* F *is concave and, for any* $r \geq 0$*, the set*

$$\{x \in \mathbb{H}_T : \|x^{'}\|_2^2 \leq 2r\}$$

is compact by Schilder theorem, utilizing minimax theorem, we have

$$L(h^*) = 2F(h^*) - \int_0^T (h^*)^{'}(t)^2 dt.$$

2.5. Optimal Drift

Now we determine h^* explicitly. Let us consider

$$\min_{AC[0,T]} \int_0^T 2ae^{bx(s)+\frac{b^2c}{2}s} + x^{'}(s)^2 ds,\ x(0) = 0,\ x^{'}(T) = 0.$$

Depending on c we have three cases.

Case 1) $c > 0$

We have Euler equation,

$$x^{''}(t) = abe^{bx(t)+\frac{b^2c}{2}t},\ x(0) = 0,\ x^{'}(T) = 0.$$

Note $x^{'}(0) < 0$. Let $y(t) := bx(t) + \frac{b^2c}{2}t$. Then, $y^{'} = bx^{'} + \frac{b^2c}{2}$, $y^{''} = bx^{''} = ab^2e^{bx(t)+\frac{b^2c}{2}t} > 0$, $y(0) = 0$, and $y^{'}(T) = \frac{b^2c}{2}$. Note $2y^{''}y^{'} = [(y^{'})^2]^{'} = 2abe^y y^{'} =]2ab^2e^y]^{'}$ and $(y^{'})^2 = 2ab^2e^y + d$. If $y^{'}(0) = 0$, then we have $x^{'}(0) = -\frac{bc}{2}$. Thus, we have three cases $-\frac{bc}{2} < x^{'}(0) < 0$, $x^{'}(0) < -\frac{bc}{2}$, and $x^{'}(0) = -\frac{bc}{2}$; $y^{'}(0) < 0$, $y^{'}(0) > 0$, and $y^{'}(0) = 0$.

For the case of $y^{'}(0) < 0$, since $y^{'}(T) = \frac{b^2c}{2}$, there is a unique point $\tau \in (0, T)$ where $y^{'}(\tau) = 0$. It follows that $d < 0$: we write $d = -k^2$ where $k > 0$. On the interval $[\tau, T]$ we have $y^{'} = \sqrt{2ab^2e^y - k^2}$. This separable differential equation, by letting $u^2 = 2ab^2e^y - k^2$, can be integrated to get

$$y(t) = log\frac{k^2}{2ab^2} + log(1 + tan^2(\frac{k(t+\alpha)}{2}))\ or\ y(t) = log\frac{k^2}{2ab^2} + log(1 + tan^2(\frac{k(\alpha - t)}{2})).$$

Differentiating leads to

$$y^{'} = ktan(\frac{k(t+\alpha)}{2}) \; or \; y^{'} = ktan(\frac{k(t-\alpha)}{2}).$$

Thus, from $y^{'}(\tau) = 0$, we have $\alpha = -\tau$ or $\alpha = \tau$. In either case, on $[\tau, T]$, we have

$$y(t) = log\frac{k^2}{2ab^2} + log(1 + tan^2(\frac{k(t-\tau)}{2})).$$

By same analysis, on $[0, \tau]$, we have the same result. Thus, since $y(0) = 0$ and $y^{'}(T) = \frac{b^2c}{2}$ we have two equations for k and τ;

$$\begin{aligned} 1 &= \tfrac{k^2}{2ab^2}(1 + tan^2(\tfrac{k\tau}{2})) \\ \tfrac{b^2c}{2} &= ktan(\tfrac{k(T-\tau)}{2}) \end{aligned}.$$

Now we consider the case $y^{'}(0) > 0$. If $d = 0$ then we have

$$y^{'} = \sqrt{2ab^2e^y}.$$

By letting $u^2 = 2ab^2e^y$ and integrating, we have

$$y = -log2ab^2 - 2log(\frac{t}{2} + \alpha) \; or \; y = -log2ab^2 - 2log(-\frac{t}{2} + \alpha),$$

and,

$$y^{'} = -\frac{2}{t + 2\alpha} \; or \; y^{'} = -\frac{2}{t - 2\alpha}.$$

Hence, we have $\alpha = \pm\frac{1}{b\sqrt{2a}}$ from $y(0) = 0$. If $\alpha = -\frac{1}{b\sqrt{2a}}$, y is undefined in either case. Moreover, since $y^{'}(0) = -\frac{1}{\alpha}$ or $y^{'}(0) = \frac{1}{\alpha}$, from $y^{'}(0) > 0$, we have

$$y(t) = -log2ab^2 - 2log(-\frac{t}{2} + \frac{1}{b\sqrt{2a}}) \; and \; y^{'}(t) = -\frac{2}{t - \frac{2}{b\sqrt{2a}}}$$

where $T = \frac{2}{b\sqrt{2a}} - \frac{4}{b^2c}$ and $bc > 2\sqrt{2a}$ by $y^{'}(T) = \frac{b^2c}{2}$. If $d > 0$, then we have

$$(y^{'})^2 = 2ab^2e^y + d = 2ab^2e^y + k^2,\ k > 0.$$

Letting $u^2 = 2ab^2e^y + k^2$, we have

$$y = log\frac{2k^2}{ab^2} + k(t + 2\alpha) - 2log(1 - e^{k(t+2\alpha)})$$

or

$$y = log\frac{2k^2}{ab^2} + k(-t + 2\alpha) - 2log(1 - e^{k(-t+2\alpha)}),$$

and

$$y^{'} = k + \frac{2ke^{k(t+2\alpha)}}{1 - e^{k(t+2\alpha)}}\ or\ y^{'} = -k - \frac{2ke^{k(t+2\alpha)}}{1 - e^{k(-t+2\alpha)}}.$$

Since $y^{'}(0) > 0$ we have

$$y = log\frac{2k^2}{ab^2} + k(t + 2\alpha) - 2log(1 - e^{k(t+2\alpha)})\ and\ y^{'} = k + \frac{2ke^{k(t+2\alpha)}}{1 - e^{k(t+2\alpha)}}.$$

So, since $y(0) = 0$ and $y^{'}(T) = \frac{b^2c}{2}$, we have two equations for k and α;

$$0 = log\tfrac{2k^2}{ab^2} + 2k\alpha - 2log(1 - e^{2k\alpha})$$

,

$$\tfrac{b^2c}{2} = k + \tfrac{2ke^{k(T+2\alpha)}}{1-e^{k(T+2\alpha)}}$$

where $\alpha < \frac{T}{2}$ and $0 < k < \frac{b^2c}{2}$. For the case of $d < 0$, we have

$$y(t) = log\frac{k^2}{2ab^2} + log(1 + tan^2(\frac{k(t+\alpha)}{2}))\ or\ y(t) = log\frac{k^2}{2ab^2} + log(1 + tan^2(\frac{k(t-\alpha)}{2})).$$

Differentiating leads to

$$y^{'} = ktan(\frac{k(t+\alpha)}{2})$$

or,

$$y^{'} = ktan(\frac{k(t-\alpha)}{2}).$$

Since $y^{'}(0) > 0$ we have $\alpha > 0$ or $\alpha < 0$. Thus, we have

$$y(t) = log\frac{k^2}{2ab^2} + log(1 + tan^2(\frac{k(t+\alpha)}{2}))\ and\ y^{'} = ktan(\frac{k(t+\alpha)}{2}).$$

So, we have two equations for $k > 0$ and $\alpha > 0$;

$$\begin{array}{ccc} 1 & = & \frac{k^2}{2ab^2}(1 + tan^2(\frac{k\alpha}{2})) \\ & & . \\ \frac{b^2c}{2} & = & ktan(\frac{k(T+\alpha)}{2}) \end{array}$$

Finally, if $y^{'}(0) = 0$, then $d = -2ab^2$ and $k = b\sqrt{2a}$. Thus, we have

$$y(t) = log(1 + tan^2(\frac{b\sqrt{2a}}{2}t)),$$

where $T = \frac{2}{b\sqrt{2a}}tan^{-1}(\frac{bc}{2\sqrt{2a}})$ by $y^{'}(T) = \frac{b^2c}{2}$.

In conclusion, by the existence and uniquness, we must have, depending on $a, b, c, and\ T$, one of the cases below;

(i)

$$x(t) = \frac{1}{b}log\frac{k^2}{2ab^2} + \frac{1}{b}log(1 + tan^2(\frac{k(t-\tau)}{2})) - \frac{bc}{2}t$$

if $0 < \tau < T$ and $0 < \tau k < \pi$ satisfy

$$\begin{aligned} 1 &= \tfrac{k^2}{2ab^2}(1+tan^2(\tfrac{k\tau}{2})) \\ \tfrac{b^2c}{2} &= ktan(\tfrac{k(T-\tau)}{2}) \end{aligned}.$$

(ii)

$$x(t) = -\frac{1}{b}log2ab^2 - \frac{bc}{2}t - \frac{2}{b}log(-\frac{t}{2} + \frac{1}{b\sqrt{2a}})$$

if $T = \frac{2}{b\sqrt{2a}} - \frac{4}{b^2c}$ and $bc > 2\sqrt{2a}$.

(iii)

$$x(t) = \frac{1}{b}log\frac{2k^2}{ab^2} + \frac{k}{b}(t+2\alpha) - \frac{2}{b}log(1-e^{k(t+2\alpha)}) - \frac{bc}{2}t$$

if $\alpha < -\frac{T}{2}$ and $0 < k < \frac{b^2c}{2}$ satisfy

$$\begin{aligned} 0 = log\tfrac{2k^2}{ab^2} + 2k\alpha - 2log(1-e^{2k\alpha}) \\ \tfrac{b^2c}{2} = k + \tfrac{2ke^{k(T+2\alpha)}}{1-e^{k(T+2\alpha)}} \end{aligned}.$$

(iv)

$$x(t) = \frac{1}{b}log\frac{k^2}{2ab^2} + \frac{1}{b}log(1+tan^2(\frac{k(t+\alpha)}{2})) - \frac{bc}{2}t$$

if $\alpha > 0$ and $0 < \alpha k < \pi$ satisfy

$$\begin{aligned} 1 &= \tfrac{k^2}{2ab^2}(1+tan^2(\tfrac{k\alpha}{2})) \\ \tfrac{b^2c}{2} &= ktan(\tfrac{k(T+\alpha)}{2}) \end{aligned}.$$

(v)

$$x(t) = \frac{1}{b}log(1+tan^2(\frac{b\sqrt{2a}}{2}t)) - \frac{bc}{2}t$$

if $T = \frac{2}{b\sqrt{2a}} tan^{-1}(\frac{bc}{2\sqrt{2a}})$.

Case 2) $\mathbf{c < 0}$

For this case we first note that $y^{'}(T) = \frac{b^2c}{2} < 0$ and $y^{''} > 0$ imply $y^{'} < 0$. Thus, it is sufficient to consider the case of $y^{'}(0) > 0$ among sub-cases for the case $c > 0$ with proper changes. Following the reasoning there line by line we have three cases;

(vi)

$$x(t) = -\frac{1}{b} log 2ab^2 - \frac{2}{b} log(\frac{t}{2} + \frac{1}{b\sqrt{2a}}) - \frac{bc}{2} t$$

if $T = -\frac{2}{b\sqrt{2a}} - \frac{4}{b^2c}$ and $2\sqrt{2a} > -bc$.

(vii)

$$x(t) = \frac{1}{b} log \frac{2k^2}{ab^2} + \frac{k}{b}(-t + 2\alpha) - \frac{2}{b} log(1 - e^{k(-t+2\alpha)}) - \frac{bc}{2} t$$

if $\alpha < 0$ and $0 < k < -\frac{b^2c}{2}$ satisfy

$$0 = log \tfrac{2k^2}{ab^2} + 2k\alpha - 2log(1 - e^{2k\alpha})$$

.

$$\tfrac{b^2c}{2} = -k - \tfrac{2ke^{k(-T+2\alpha)}}{1-e^{k(-T+2\alpha)}}$$

(viii)

$$x(t) = \frac{1}{b} log \frac{k^2}{2ab^2} + \frac{1}{b} log(1 + tan^2(\frac{k(t+\alpha)}{2})) - \frac{bc}{2} t$$

if $k > 0$, $\alpha < 0$, and $-\pi < \alpha k < 0$ satisfy

$$1 = \tfrac{k^2}{2ab^2}(1 + tan^2(\tfrac{k\alpha}{2}))$$

.

$$\tfrac{b^2c}{2} = ktan(\tfrac{k(T+\alpha)}{2})$$

Case 3) $c = 0$

In this case, we first note that $y^{'}(T) = 0$ and $y^{'}(0) < 0$. Thus, by considering the case

$d < 0$ when $c > 0$ and $y^{'}(0) > 0$ with proper changes, we have the only case below;

(ix)

$$x(t) = \frac{1}{b} log \frac{k^2}{2ab^2} + \frac{1}{b} log(1 + tan^2(\frac{k(t-T)}{2}))$$

if $k > 0$ satisfies

$$1 = \frac{k^2}{2ab^2}(1 + tan^2(\frac{kT}{2})).$$

2.6. Numerical Results

We test the proposed method for the case of $\rho > 0$. Consider the Dothan model in Pintoux and Privault (2011) for bond pricing for the zero-coupon price

$$L(T, r_0) = E[e^{-\int_0^T r_s ds} \mid \mathcal{F}_t]$$

where $r_s = r_0 e^{\sigma W(s) + \rho\sigma^2 s/2}$, $s \in \mathbb{R}_+$ and T is the maturity of a bond. $L(T, r_0)$ is the bond price under the condition $L(0, r_0) = 1$.

For example, setting a volatility coeffieient of $\sigma = 30\%$ and a drift value of $\rho\sigma^2/2 = 0.045$ per year the approximate theoretical value of the Dothan model for a zero-coupon bond ten years before maturity via an integral representation from PDE approach is 0.43 when the underlying short term interest rate is of 6% (see page 512, 516 in Privault and Uy (2013)). The values of parameters in the model allow us to have type (i) as the formula for asymptotically optimal drift for the Monte Carlo simulation.

In Table 2.1 simulation results for crude and optimal drift cases are recorded. Simulations are performed with 100, 300, and 500 paths with the time-increment of $1/2520$ corresponding to one business day. As the number of paths increase the results of simulations approaches to the approximate theoretical value above.

Table 2.2 shows the performance of asymptotically optimal estimator in terms of variance reduction which is obtained by dividing the variance of crude Monte Carlo sample by the

Table 2.1.: Monte Carlo estimation of crude and optimal estimators

Sample size	Crude	Optimal
100	0.4549	0.4533
300	0.4431	0.4393
500	0.4376	0.4344

Table 2.2.: Variance reduction ratio across various parameters

r_0	σ	$\frac{\rho\sigma^2}{2}$	S.E.	Variance ratio	Type
0.03	0.2	0.02	0.0307	9.4210	(iv)
	0.3	0.045	0.0721	4.8846	(iv)
	0.4	0.08	0.1170	3.6642	(iv)
0.06	0.2	0.02	0.0337	10.400	(i)
	0.3	0.045	0.0980	3.5417	(i)
	0.4	0.08	0.1025	4.9143	(i)
1.00	0.2	0.02	0.0001	116.6425	(i)
	0.3	0.045	0.0004	97.81570	(i)
	0.4	0.08	0.0008	84.43210	(i)

Table 2.3.: Variance reduction ratio for other cases

T	a	σ	ρ	S.E.	Variance reduction	Type
0.4	2	3.16	4	0.0005	471.18	(v)
0.5	1	1	20	0.0004	166.95	(iii)
1	0.5	1	4	0.0648	7.8671	(ii)

variance of optimal sample. Each simulation is performed with 100,000 paths. The results show that the application of optimal drift significantly improves Monte Carlo estimate in variance reduction. In the table we observe a tendency that the variance reduction might be more and more significant as the underlying short term interest rate increases. Hence, the optimal drift method could be more effective when we use the Dothan model with a high short term interest rate. However, since we get at least about four-fold in variance reduction even at worst case, the optimal drift method is still effective for the case of a low short term interest rate.

Some simulation results for other cases (ii), (iii) and (v) are recorded in Table 2.3. The values of T are chosen to be small keeping in mind comparison to results in Table 2.2 where $T = 10$. The results still show the optimal drift method significantly improves Monte Carlo estimation by reducing dramatically sample variance in comparison to the sample variance of crude Monte Carlo simulation.

2.7. Concluding Remarks

An importance sampling method for Monte Carlo simulation of the Laplace transform of exponential Brownian functionals is developed. A Large deviation principle converts the problem of finding the asymptotically optimal change of drift into a calculus of variations problem. The solution of the variational problem is unique and is an element of C^∞. Moreover, the corresponding Euler equation allows us to have closed-form solutions. Hence, the overhead for implementing the method is negligible.

Numerical experiments were performed in the case of the Dothan bond pricing model.

The results shows that the proposed method demonstrates an effective variance reduction in comparison to the case of a standard crude Monte Carlo method.

References

Clarke, F. (2013): *Functional Analysis, Calculus of Variations and Optimal Control.* London: Springer.

Comet, A., C. Monthus, and M. Yor (1998): Exponential functionals of Brownian motion and disordered systems, J. *App. Probab.* 35(2), 255-271.

Dembo, A., and O. Zeitouni (1998): *Large deviations Techniques and Applications, 2nd ed.* New York: Springer.

Desmond, H. (2001): An algorithmic introduction to numerical simulation of stochastic differential equations, *SIAM Rev.* 43(3), 525-546.

Dufresne, D. (2001): The integral of geometric Brownian motion, *Adv. App. Probab.* 33(1), 223-241.

Kallenberg, O. (2002): *Foundations of Modern Probability, 2nd ed.* New York: Springer.

Linetsky, V.: Spectral expansions for Asian (average price) options, *Oper. Res.* 52(6), 856-867.

Guasoni, P., and S. Robertson (2008): Optimal Importance Sampling with Explicit Formulas, *Finance and Stochastics* 12, 1-19.

Glasserman, P. (2004): *Monte Carlo Methods in Financial Engineering.* New York: Springer.

Glasserman, P., P. Heldelberger, and P. Shahabuddin (1999): Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options, *Math. Finance.* 9, 117-152.

Peres, Y., and P. Morters (2010): *Brownian Motion.* New York: Cambridge.

Pintoux, C., and N. Privault (2011): The Dothan pricing model revisited, *Math. Finance.* 21, 355-363.

Privault, N., and W. Uy (2013): Monte Carlo computation of the Laplace transform of exponential Brownian functionals, *Methodol. Comput. Appl. Probab.* 15, 511-524.

Schilder, M. (1966): Some asymptotic formulas for Wiener integrals, *Trans. Am. Math. Soc.* 125, 63-85.

Varadhan, S. (1966): Asymptotic probabilities and differential equations. *Commun. Pure. Appl. Math.* 19, 261-286.G

Part III.

ON THE CONVERGENCE OF QUASI-REGRESSION METHOD: POLYNOMIAL CHAOS AND REGULARITY

3.1. Abstract

An analysis of convergence of a quasi-regression Monte Carlo method for American option pricing proposed by Glasserman and Yu (2004) is conducted. In that paper the author showed that the method converges to an approximation of the true price of American option with a single underlying asset while establishing a critical relation between the number of basis functions and the number of Monte Carlo simulations, the method uses Hermite polynomials and multiples of the powers as basis functions for Brownian motion and geometric Brownian motion, respectively. We show that the method surely converges to the true price of American option even under multiple underlying assets. Moreover, we show the critical relation for Brownian motion case with single asset holds also for geometric Brownian motions with basis functions different from ones in Glasserman and Yu (2004) with multiple-underlying assets. A rate of convergence of the method is also provided by introducing regularity of value functions.

3.2. Introduction

In this chapter we study the convergence of a Monte Carlo method for pricing American options. American option differs from better known European option in that it gives the holder the right to exercise the option at any time before the expiration date. A standard theory of American option pricing tells us that the problem accompanies an optimal stopping problem (see, e.g., Glasserman (2004)). Hence, the next step for the problem is to seek numerical methods. Many papers are devoted to the topic of numerical methods for the problem. Among numerical methods the least-squares Monte Carlo method proposed by Longstaff and Schwartz (2001) has been one of the most popular one to both researchers and practitioners(see Stentoft (2004) for more detailed review and comparison among other approaches such as PDE or binomial tree methods).

The strong point of the least-squares Monte Carlo method is, needless to say, the ease of implementation. However, analysis of the convergence of the method is a difficult task. The first paper addressing this question was Clement, Lamberton and Protter (2002), which shows convergence of the method to an approximation to the true price under a fixed number of basis functions. Glasserman and Yu (2004) and Stentoft (2004) have studied the problem when the number of paths and number of basis functions increase at the same time. Gerhold (2011) extended Glasserman and Yu (2004) to the case where underlying asset follows several Levy processes. Stentoft (2004) used results about series estimator to show that the method achieves polynomial growth of the number of paths in the number of basis functions. Later, Egloff (2005), Egloff, Kohler and Todorovic (2007) appealed to statistical learning theory to study the convergence of the method under the assumption of boundedness of state space. Zanger (2013) later took a similar approach to the problem without the assumption.

However, it should be noted that Glasserman and Yu (2004) and Gerhold (2013) analyzed a quasi-regression method which is a variant of the standard least-squares Monte Carlo regression method proposed by Longstaff and Schwartz in two aspects: how paths are generated and the use of the exact matrix in calculation of coefficients of basis functions (see Glasserman and Yu (2004), page 2095).

The purpose of the present part is to improve the results in Glasserman and Yu (2004). To be specific, Glasserman and Yu (2004) have shown a quasi-regression method converges to an approximation of the true price of the American option with two examples of single underlying asset: Brownian motion and geometric Brownian motion. To prove convergence, they show that the number of basis functions (Hermite polynomials) K in a sample size N must grow at $O(logN)$ ($O(\sqrt{logN})$ for geometric Brownian motion using multiples of the powers x^k as basis functions) in order to get convergence. We show that, even in the case of multiple underlying assets, the algorithm converges to the true value using the asymptotic of the moments of Hermite polynomials. Further, we show that the critical value on relation between the number of basis functions and the number of paths simulated in Glasserman and Yu (2004) still holds for geometric Brownian motion with basis functions different from the multiples of powers, which points out the importance of proper choice of basis functions in implementing a quasi-regression method. To this end, we use polynomial chaos expansion to ensure the assumptions for Brownian motion case in Glasserman and Yu (2004) still holds for the case of geometric Brownian motion. Finally, considering the regularity of continuation value function, we present a rate of convergence of the algorithm.

In Section 3.3 we recall the general backward induction framework for pricing of American option. In section 3.4 we assume our market model. We also introduce some notations and results for polynomial chaos and regularity of functions in Section 3.5 and 3.6. We propose the algorithm to be analyzed in section 3.7. Section 3.8 presents two main results: (1) proof of the convergence of the algorithm and (2) a rate of convergence for the algorithm. We will end with some concluding remarks in section 3.9.

3.3. Pricing via Backward Induction

In this section, we present a general framework for pricing of an American option to which our algorithm in the subsequent sections applies. We follow the presentation of Glasserman and Yu (2004) since the goal of this chapter improves upon their results. One can find a more detailed and kind description about the formulation of the framework in many sources (see, e.g., Glasserman (2004), Korn et al. (2010)).

We assume a complete probability space $(\Omega, \mathcal{F}, P)$ where P is the risk-neutral measure. We deal with the problem in a discretized time setting: the option expires in m periods with T as the expiration date and set the early exercise points as $t_0 = 0 < t_1 < ... < t_m = T$. Hence, our problem can be considered as an approximation to the price of an American option in discretized time or the exact price of a Bermudan option. A theoretical value $V_{t_n}(x)$ of American option at t_n in state x is given by

$$V_{t_n}(x) = \sup_{\tau \in \Gamma_n} E[h(S_\tau) \mid S(t_n) = x]$$

where $h : \mathbb{R}^d \to \mathbb{R}$ is a payoff function for the option and $h \in L^2(\Omega, \mathcal{F}, P)$, Γ_n is the set of all stopping times taking values in $\{t_n, ..., t_m\}$ adapted to the filtration corresponding to a market model, $S(t) = (S_1(t), ..., S_d(t))$ where $S(t)$ is a given stochastic process.

The option value satisfies the backward induction equations:

$$V_{t_m}(x) = h_{t_m}(x)$$

and

$$V_{t_n}(x) = max\{h_{t_n}(x), E[V_{t_{n+1}}(S(t_{n+1}))|S(t_n) = x]\},$$

$n = 0, 1, ..., m-1$. We can rewrite these with respect to continuation values

$$C^*_{t_n}(x) = E[V_{t_{n+1}}(S(t_{n+1})) \mid S(t_n) = x], \quad n = 0, 1, ..., m-1,$$

as

$$C^*_{t_m}(x) = 0,$$

$$C^*_{t_n}(x) = E[max\{h_{t_{n+1}}(S(t_{n+1})), C^*_{t_{n+1}}(S(t_{n+1}))\} \mid S(t_n) = x],$$

$n = 0, 1, ..., m-1$. The option value satisfies

$$V_{t_n}(x) = max\{h_{t_n}(x), C^*_{t_n}(x)\}.$$

Therefore, we can calculate the value from the continuation values at least from a theoretical perspective. We note if $S(0)$ is a constant, then, $C^*_{t_0} = E[V_{t_1}]$. We, further, note that deterministic or stochastic discounting can be absorbed into h_n (see Glasserman and Yu (2004)).

3.4. Market Model

We assume that the underlying assets $\{S_i\}_{i=1}^d$ follow a correlated geometric Brownian motion with a fixed initial value $S_0 = \{s_1, ..., s_d\}$ under the risk-neutral measure:

$$dS_i(t) = rS_i(t)dt + \sigma_i S_i(t)dW_i(t), \;\; i = 1, ..., d,$$

equivalently,

$$S_i(t) = s_i e^{(r-\frac{1}{2}\sigma_i^2)t+\sigma_i W_i(t)}, \ \ i = 1, ..., d,$$

where $\{W_i\}_{i=1}^d$ is a correlated Brownian motion, r is the risk free rate, σ_i is the volatility of the i-th asset. The correlations between W_i and W_j are given by a $d \times d$ matrix ρ with $[\rho]_{ij} = \rho_{ij}$ where $\rho_{ii} = 1$ and $-1 \leq \rho_{ij} \leq 1$. Since, by Cholesky decomposition,

$$\rho = HH^T$$

where H is a lower triangular matrix,

$$\boldsymbol{W}(t) = H\boldsymbol{Z}(t),$$

where $\boldsymbol{Z}(t)$ is d-dimensional Brownian motion (see Glasserman (2004)). Then, in our discretized setting, for $i = 1, ..., d$

$$S_i(t_n) = s_i e^{(r-\frac{1}{2}\sigma_i^2)t_n+\sigma_i \sum_{j=1}^d h_{ij} Z_j(t_n)}, \ \ n = 0, 1, ..., m.$$

For notational convenience, we denote $(S_i(t_n))_{i=1}^d$ and $(Z_i(t_n)/\sqrt{t_n})_{i=1}^d$ by $\boldsymbol{S}_n$ and $\boldsymbol{\xi}_n$ for $n = 1, ..., m$. Note $\boldsymbol{\xi}_n$ is a random vector consisting of $i.i.d.$ random variables with standard normal distribution. The assumption below turns out to be useful:

Assumption (1): ρ is positive-definite.

Then, since H is invertible, the σ-algebras generated by $\boldsymbol{S}_n$ and $\boldsymbol{\xi}_n$ are equivalent:

$$\sigma(\boldsymbol{S}_n) = \sigma(\boldsymbol{\xi}_n).$$

Therefore, by the Doob-Dynkin lemma (see Ernst et al. (2012)), there exists a Borel-measurable function C_{t_n} from $\mathbb{R}^d$ to $\mathbb{R}$ such that

$$C^*_{t_n}(\boldsymbol{S}_n) = C_{t_n}(\boldsymbol{\xi}_n), \ \ n = 1, ..., m-1.$$

Thus, for each $n \in \{1, ..., m-1\}$, we have

$$C_{t_n}^*(\boldsymbol{S}_n) = C_{t_n}(\boldsymbol{\xi}_n) \in L^2(\Omega, \sigma(\boldsymbol{\xi}_n), P).$$

For notational simplicity, we write C_n for C_{t_n} for each n.

3.5. Polynomial Chaos Expansion

Here we introduce some results on polynomial chaos expansion of a function. This section is just an short summary of section 3 in Ernst et al. (2012). Thus, one can consult with Ernst et al. (2012) for more detailed and comprehensive development of the theory of generalized polynomial chaos expansion. An important issue in stochastic computation is to find a manageable representation of random object of interest. A popular approach for this is polynomial chaos expansion; a random variable is represented by a series of Hermite polynomials. We detail this approach in the context of the current application.

Define the normalized Hermite polynomials $\{\psi_k\}_{k\in\mathbb{N}_0}$ by

$$\psi_k(x) = \frac{1}{\sqrt{k!}} H_k(x),\ x \in \mathbb{R},$$

where $H_k(x) = (-1)^k e^{\frac{x^2}{2}} (\frac{d^k}{dx^k} e^{-\frac{x^2}{2}})$. From the results for ("probabilist's") Hermite polynomials (See Abramowits and Stegun (1972)), the normalized Hermite polynomials satisfy

$$\psi_k'(x) = \sqrt{k}\psi_{k-1}(x),\ k \geq 1, \tag{3.5.1}$$

and

$$\psi_{k+1}(x) = \frac{x}{\sqrt{k+1}}\psi_k(x) - \frac{\sqrt{k}}{\sqrt{k+1}}\psi_{k-1}(x),\ k \geq 1. \tag{3.5.2}$$

We further note

$$\int_{\mathbb{R}} \psi_m(x)\psi_n(x)\omega(x)dx = \delta_{mn}$$

where ω is the standard normal density function.

We consider $L^2(\Omega, \sigma(\xi), P)$ where ξ has a standard normal distribution F_ξ. For any $\varphi \in L^2(\Omega, \sigma(\xi), P)$, by Doob-Dynkin lemma, there exists a Borel-measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $\varphi = f(\xi)$. Then, since the normalized Hermite polynomials constitute an orthonormal system for $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_\xi(dx))$, the set $\{\psi_k(\xi)\}_{k\in\mathbb{N}_0}$ is also an orthonormal system of $L^2(\Omega, \sigma(\xi), P)$. The completeness of these systems amounts to the question of polynomials in an L^2-space to the unique solvability of a moment problem; one says that the moment problem is *uniquely solvable* for a probability distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, or that the distribution is *determinate* if the distribution function is uniquely defined by the sequence of its moments

$$\mu_k := \int_{\mathbb{R}} x^k F_\xi(dx), \quad k \in \mathbb{N}_0.$$

It is obvious that ξ satisfies Assumption 3.1. in Ernst et al. (2012): (i) each basic random variable ξ_m possesses finite moments of all orders and (ii) the distribution functions $F_{\xi_m}(x) := P(\xi_m \leq x)$ of the basic random variables are continuous. Now we need some results from Ernst et al. (2012).

Theorem 3.9. *(Theorem 3.3 in Ernst et al. (2012)) The sequence of orthogonal polynomials associated with a real random variable ξ, satisfying Assumption 3.1 is dense in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_\xi(dx))$ if and only if the moment problem is uniquely solvable for its distribution.*

Here is another one.

Theorem 3.10. *(Theorem 3.4 in Ernst et al. (2012)) If one of the following conditions for the distribution F_ξ of a random variable ξ satisfying Assumption 3.1 is valid, then the moment problem is uniquely solvable and therefore the set of polynomials in the random variable ξ is dense in the space $L^2(\Omega, \sigma(\xi), P)$.*

(a) The distribution F_ξ has compact support.

(b) The moment sequence $\{\mu_n\}_{n\in\mathbb{N}_0}$ of the distribution satisfies

$$\liminf_{n\to\infty} \frac{\sqrt[2n]{\mu_{2n}}}{2n} < \infty.$$

(c) The random variable is exponentially integrable, i.e., there holds

$$\int_{\mathbb{R}} exp(a \mid x \mid) F_\xi(dx) < \infty$$

for a strictly positive number a. An equivalent condition is the existence of a finite moment-generating function in a neighborhood of the origin.

(d) The moment sequence $\{\mu_n\}_{n\in\mathbb{N}_0}$ of the distribution satisfies

$$\sum_{n=0}^{\infty} \frac{1}{\sqrt[2n]{\mu_{2n}}} = \infty.$$

(e) If the distribution has a symmetric, differentiable and strictly positive density f_ξ and for a real number $x_0 > 0$ there holds

$$\int_{-\infty}^{\infty} \frac{-log f_\xi(x)}{1+x^2} dx = \infty \ \ and \ \ \frac{-x f_\xi'(x)}{f_\xi(x)} \nearrow \infty (x \to, x \geq x_0).$$

Note these two theorems imply $\{\psi_k(\xi)\}_{k\in\mathbb{N}_0}$ is a complete orthonormal system for $L^2(\Omega, \sigma(\xi), P)$:

$$\varphi = f(\xi) = \sum_{k=0}^{\infty} a_k \psi_k(\xi), \ in \ L^2,$$

where

$$a_k =< \varphi, \psi_k(\xi) >= \int_\Omega \varphi \psi_k(\xi) dP = \int_\Omega f(\xi)\psi_k(\xi) dP = \int_{\mathbb{R}} f(x)\psi_k(x) F_\xi(dx), \ k \in \mathbb{N}_0.$$

Next, we construct the same result for multi-dimensional case. Let us consider a random vector $\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$ where $\boldsymbol{\xi} = (\xi_1, ..., \xi_d)$ and $\{\xi_i\}_{i=1}^d$ is *i.i.d.* with standard normal distribution. First, denote by $\{p_j^{(m)}\}_{j\in\mathbb{N}_0}$, $m = 1, ..., d$, the sequence of polynomials orthonormal with respect to the distribution ξ_m. Then, the set of multivariate polynomials given by

$$p_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{m=1}^{d} p_{\alpha_m}^{(m)}(\xi_m), \quad \boldsymbol{\alpha}(\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d,$$

consititutes an orthonormal system of random variables in the space $L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$. Therefore the polynomials

$$p_{\boldsymbol{\alpha}} : \boldsymbol{x} \mapsto p_{\boldsymbol{\alpha}}(\boldsymbol{x}), \quad \boldsymbol{\alpha} \in \mathbb{N}_0^d,$$

form an orthonormal system in the image space $L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with the product probability measure $F_{\xi_1}(dx_1) \times \cdots \times F_{\xi_d}(dx_d)$. Now we address the multi-dimensional counterparts of theorem 3.9 and 3.10.

Theorem 3.11. *(Theorem 3.6 in Ernst et al. (2012)) Let $\boldsymbol{\xi} = (\xi_1, ..., \xi_M)$ be a vector of $M \in \mathbb{N}$ independent random variables satisfying Assumption 3.1 and $\{p_j^{(m)}\}_{j \in \mathbb{N}_0}$, $m = 1, ..., M$, the associated orthonormal polynomial sequences. Then the orthonormal system of random variables*

$$p_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{m=1}^{M} p_{\alpha_m}^{(m)}(\xi_m), \quad \boldsymbol{\alpha} \in \mathbb{N}_0^M,$$

is an orthonormal basis of the space $L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$ if and only if the moment problem is uniquely solvable for each random variable ξ_m, $m = 1, ..., M$. In this case any random variable $\eta \in L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$ can be expanded in an abstract Fourier series of multivariate orthonormal polynomials in the basic random variables, the generalized polynomial chaos expansion

$$\eta = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^M} a_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \quad \textit{with} \quad a_{\boldsymbol{\alpha}} = < \eta p_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) > .$$

Here is sufficient conditions for solvability.

Theorem 3.12. *(Theorem 3.7 in Ernst et al. (2012)) If the distribution function $F_{\boldsymbol{\xi}}$ of a random vector $\boldsymbol{\xi} = (\xi_1, ..., \xi_M)$ with continuous distribution and finite moments of all orders satisfies one of the following conditions, then the multivariate polynomials in $\xi_1, ..., \xi_M$ are* *dense in* $L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$*. In this case any random variable* $\eta \in L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$ *is the limit of its generalized polynomial chaos expansion, which converges in quadratic mean.*

(a) The distribution function $F_{\boldsymbol{\xi}}$ *has compact support.*

(b) The random vector is exponentially integrable, i.e., there exists $a > 0$ *such that*

$$\int_{\mathbb{R}^M} exp(a\|\boldsymbol{x}\|) F_{\boldsymbol{\xi}}(d\boldsymbol{x}) < \infty,$$

where $\|\cdot\|$ *denotes any norm on* $\mathbb{R}^M$*.*

Thus, denoting $\{\psi_j^{(i)}\}_{j\in\mathbb{N}_0}$ by the normalized Hermite polynomials corresponding to ξ_i, the set of multivariate tensor product of the polynomials given by

$$\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{m=1}^{d} \psi_{\alpha_m}^{(m)}(\xi_m), \quad \boldsymbol{\alpha} = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d,$$

is a complete orthonormal system of $L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$: For each $\varphi \in L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$ and a Borel-measurable function $f : \mathbb{R} \to \mathbb{R}$,

$$\varphi = f(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^d} a_{\boldsymbol{\alpha}} \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \; in \; L^2,$$

where

$$a_{\boldsymbol{\alpha}} =< \varphi, \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) >= \int_{\Omega} \varphi \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) dP = \int_{\Omega} f(\boldsymbol{\xi}) \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) dP = \int_{\mathbb{R}^d} f(\boldsymbol{x}) \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{x}) F_{\boldsymbol{\xi}}(\boldsymbol{dx})$$

and

$$F_{\boldsymbol{\xi}}(\boldsymbol{dx}) = F_{\xi^{(1)}}(dx) \times \cdots \times F_{\xi^{(d)}}(dx).$$

Although the multi-index representation is legitimate in theoretical development, it is impractical to use the multi-index representation for the purpose of finite-term approximation to a function in $L^2(\Omega, \sigma(\boldsymbol{\xi}), P)$. We therefore introduce a single-index that is more tractable in constructing a finite truncation of infinite-sum representation of a function in L^2. Among single-index schemes, we adopt the *graded lexicographic order*, which says that higher degree monomials are bigger and we use lexicographic order to break ties. By adopting the scheme,

$$\varphi = f(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} a_k \boldsymbol{\psi}_k(\boldsymbol{\xi}),$$

where $a_k =< \varphi, \boldsymbol{\psi}_k(\boldsymbol{\xi}) >,\ k \in \mathbb{N}_0$. Thus, we have

$$C_n(\boldsymbol{\xi}_n) = \sum_{k=0}^{\infty} a_k \boldsymbol{\psi}_k(\boldsymbol{\xi}_n),$$

where $a_k =< C_n(\boldsymbol{\xi}_n), \boldsymbol{\psi}_k(\boldsymbol{\xi}_n) >$. Note supposing $\boldsymbol{\alpha}(k) = (\alpha(k)_1, ..., \alpha(k)_d)$ is the multi-index with $|\boldsymbol{\alpha}(k)| = \sum_{m=1}^{d} \boldsymbol{\alpha}(k)_m$ corresponding to k we have $|\boldsymbol{\alpha}(k)| \leq |\boldsymbol{\alpha}(k+1)|$ and $| \boldsymbol{\alpha}(k) | \leq k$.

Now, we present an estimate for the fourth moments of Hermite polynomials useful in developing main results later.

Proposition 3.13. *Let ψ and $\boldsymbol{\psi}$ be Hermite polynomials and multi-dimensional Hermite polynomials, respectively. Then, the followings hold; (i) for sufficiently large k, there exist positive constants C and $\tilde{C}$ such that*

$$C\frac{3^{2k}}{k} \leq E[\psi_k^4] \leq \tilde{C}\frac{3^{2k}}{k},$$

and (ii) for $k \geq 1$, there exists a positive constant C such that

$$E[\boldsymbol{\psi}_k^4] \leq C 3^{2|\alpha(k)|}.$$

(Proof) From Theorem 2.1. in Lars (2002), it is obvious that (i) is true. For (ii) we note that

$$E[\boldsymbol{\psi}_k^4] = \prod_{j=1}^{d} E[\psi^4_{\boldsymbol{\alpha}(k)_j}].$$

Then, by (i), we have

$$E[\boldsymbol{\psi}_k^4] \leq C 3^{2|\boldsymbol{\alpha}(k)|},$$

which completes the proof.

Note C denotes a generic positive constant in this chapter.

3.6. Regularity Conditions

In addition to the theory of polynomial chaos, we need some facts about regularity. First, we introduce some basic notations for regularity which are common in PDE. We follow Section 2 in Guo (1999) and Guo and Xu (2003). Let $\Lambda = \{x \mid -\infty < x < \infty\}$ and $\omega(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Define

$$L^2_\omega(\Lambda) = \{v \mid v \textit{ is measurable and } \|v\|_{L^2_\omega(\Lambda)} < \infty\},$$

where $\|v\|_{L^2_\omega(\Lambda)} = (\int_\Lambda \mid v(x) \mid^2 \omega(x)dx)^{1/2}$. Further, let $\partial_x v = \frac{\partial v}{\partial x}$, and for a non-negative integer r,

$$H^r_\omega(\Lambda) = \{v \mid \partial^k_x v \in L^2_\omega(\Lambda),\ 0 \le k \le r\}.$$

The semi-norm and the norm of $H^r_\omega(\Lambda)$ are given by

$$|v|_{H^r_\omega(\Lambda)} = \|\partial^r_x v\|_{L^2_\omega(\Lambda)} \textit{ and } \|v\|_{H^r_\omega(\Lambda)} = (\sum_{k=0}^{r} |v|^2_{H^k_\omega(\Lambda)})^{1/2}.$$

Similarly, for d-dimensions, let

$$\begin{array}{rcl} \Lambda_i & = & \{x_i \mid -\infty < x_i < \infty\} \\ \Lambda^d & = & \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_d \\ x & = & (x_1, x_2, \ldots, x_d) \end{array} .$$

Also, let $\mid x \mid = (\sum_{i=1}^d x_i^2)^{1/2}$ and $\omega(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{|x|^2}{2}}$. Define

$$L^p_\omega(\Lambda^d) = \{v \mid v \textit{ is measurable and } \|v\|_{L^p_\omega(\Lambda^d)} < \infty\}.$$

Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_d)$ be a multi-index and

$$\partial_x^{\boldsymbol{\alpha}} v(x) = \frac{\partial^{|\boldsymbol{\alpha}|} v}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}}(x),$$

where $\mid \boldsymbol{\alpha} \mid = \sum_{j=1}^{d} \alpha_j$. For any non-negative integer r,

$$H_\omega^r(\Lambda^d) = \{v \mid \partial_x^{\boldsymbol{\alpha}} v \in L_\omega^2(\Lambda^d),\ 0 \leq |\boldsymbol{\alpha}| \leq r\}.$$

The semi-norm $\mid v \mid_{H_\omega^r(\Lambda^d)}$ and the norm $\|v\|_{H_\omega^r(\Lambda^d)}$ of $H_\omega^r(\Lambda^d)$ are the natural extension of one-dimensional case (see Adams (1975)).

3.7. Algorithm

We recall the quasi-regression algorithm to be analyzed that was proposed by Glasserman and Yu (2004) (see page 2094-2095):

Step 1. Set $\hat{C}_m = 0$ and $\hat{V}_m = max\{h_m, \hat{C}_m\} = h_m$.

Step 2. For each $n = 1, ..., m-1$, starting from $m-1$, we repeat the following: Generate N independent copies $\{\boldsymbol{S}_1^i, ..., \boldsymbol{S}_{n+1}^i\}$ of path $\{\boldsymbol{S}_1, ..., \boldsymbol{S}_{n+1}\}$, $i = 1, ..., N$, up to time t_{n+1}, independent of all previously generated paths. Set

$$\hat{\gamma}_{n,k} = \frac{1}{N} \sum_{i=1}^{N} \hat{V}_{n+1}(\boldsymbol{S}_{n+1}^i) \boldsymbol{\psi}_{n,k}(\boldsymbol{S}_n^i),\ k = 0, ..., K,$$

calculate the coefficients $\hat{\beta}_n = \Psi_n^{-1} \hat{\gamma}_n$ and set

$$\hat{C}_n = \sum_{k=0}^{K} \hat{\beta}_{n,k} \boldsymbol{\psi}_{n,k}\ and\ \hat{V}_n = max\{h_n, \hat{C}_n\}.$$

Step 3. Set $\hat{C}_{N,K,0}(\boldsymbol{S}_0) = \frac{1}{N} \hat{V}_1(\boldsymbol{S}_1^i)$ and $\hat{V}_0(\boldsymbol{S}_0) = max\{h_0(\boldsymbol{S}_0), \hat{C}_{N,K,0}(\boldsymbol{S}_0)\}$.

In this algorithm $\boldsymbol{S}_0$ is fixed and $\psi's$ are general basis functions . We note that step 2 is different from the algorithm in Longstaff and Schwartz (2001), which generates a single set of paths for all dates. Moreover, we also note that the present algorithm has another feature different from the algorithm in Longstaff and Schwartz (2001); in the regression process, we use the exact matrix

$$\Psi_n = E[\psi_n(S_n)\psi_n(S_n)^T]$$

instead of its sample counterpart

$$\frac{1}{N}\sum_{i=1}^{N}\psi_n(S_n^i)\psi_n(S_n^i)^T,$$

calculated from the simulated values themselves.

Our purpose is to analyze convergence of the algorithm for two concrete examples: Brownian motion and geometric Brownian motion. To this end we alter step 2 in the algorithm to be more convenient for our purpose. Specifically, we choose a proper transformation ϕ so to have $\phi(\boldsymbol{S}) = \boldsymbol{\xi}$ and take the composit $\psi \circ \phi$ as basis functions in the algorithm where $\psi's$ are Hermite polynomials. Considering the assumption and results in section 3 it is possible for one to have this kind of basis functions at least for correlated Brownian and geometric Brownian motion under assumption (1). The resulting modified version of step 2 is as followings;

Step 2. For each $n = 1, ..., m-1$, starting from $m-1$, we repeat the following: generate N independent copies $\{\boldsymbol{Z}_1^i, ..., \boldsymbol{Z}_{n+1}^i\}$ of path $\{\boldsymbol{Z}_1, ..., \boldsymbol{Z}_{n+1}\}$, $i = 1, ..., N$, up to time t_{n+1}, independent of all previously generated paths. Calculate

$$\boldsymbol{S}_{n+1}^i \; and \; \boldsymbol{\xi}_n^i$$

and

$$\hat{\beta}_{n,k} = \frac{1}{N}\sum_{i=1}^{N}\hat{V}_{n+1}(\boldsymbol{S}_{n+1}^i)\boldsymbol{\psi}_k(\boldsymbol{\xi}_n^i), \; k = 0, ..., K.$$

Set

$$\hat{C}_{N,K,n} = \sum_{k=0}^{K}\hat{\beta}_{n,k}\boldsymbol{\psi}_k \; and \; \hat{V}_n = max\{h_n, \hat{C}_{N,K,n}\}.$$

The meaning of additional subindices N, K will be clear in the next section.

We note Glasserman and Yu (2004) used the expectation of the weighted L^2-norm on functions $G : \mathbb{R} \to \mathbb{R}$ that slightly differs from the ordinary L^2-norm (see pp 2106 in Glasserman and Yu (2004)). We alter slightly step 3 so to use the ordinary L^2-norm to estimate errors in the analysis of convergence of the algorithm: For $n = 0$, generate N independent copies $\boldsymbol{S}_1^i$ of $\boldsymbol{S}_1$ independent of all previously generated paths. With these samples, we calculate $\hat{C}_{N,K,0}(\boldsymbol{S}_0)$. The overhead for this additional computational effort is negligible. Moreover, at the cost of adding this step, we gain a huge reward; there are a lot of assumptions in Glasserman and Yu (2004) but those assumptions are unnecessary, which will be clear in the next section.

Before going to the main results of this chapter we now reconsider single-period problems in Glasserman and Yu for $m = 2$ where the dimension of underlying asset is one. Glasserman and Yu (2004) proposed three assumptions (A1), (A2), and (A3) to get the desired result for the single-period problem;

$$\begin{array}{ll} (A1) & |\beta| = 1 \\ (A2) & h_2(S_{t_2}) = \sum_{k=0}^{K} a_k \psi_{2k}(S_{t_2}),\ for\ some\ constants\ a_k \\ (A3) & \psi_{nk}(S_n)\ are\ martingales,\ up\ to\ a\ deterministic\ function\ of\ time \end{array} .$$

The present algorithm with the step 2 altered is exactly same as one in 3.1 in (see pp 2098-2099 in Glasserman and Yu (2004)) where $\{S(t),\ 0 \le t \le T\}$ is a standard Brownian motion and basis functions are $\psi_{nk}(x) = \frac{1}{\sqrt{k!}} H_{e_k}(x/\sqrt{t_n})$. Therefore, for the single period problem, our results are identical to Theorem 1 in section 3.1 in Glasserman and Yu (2004).

Naturally, Theorem 2 for lognormal setting does not hold for the present algorithm because the basis functions used are different each other; in Glasserman and Yu (2004), the basis functions used are multiples of the powers x^k resulting in

$$\psi_k(S(t)) = e^{kW(t) - k^2 t/2},$$

under the geometric Brownian motion assumption. However, for the present algorithm the basis functions are, by the facts from section 3 and argument in the present section,

$$\psi_k(\phi_{t_n}(S_{t_n})) = \psi_k(\xi_{t_n}),$$

where $\phi_n(x) = \frac{logx + t_n/2}{\sqrt{t_n}}$. Then, since $\psi_k(\phi_{t_n}(S(t_n)) = \psi_k(W(t_n)/\sqrt{t_n})$, the assumption (A3) is satisfied (see page 2098 in Glasserman and Yu (2004)). For assumption (A2) we note that

$$h_2(S_{t_2}) = H(\xi_{t_2})$$

for some Borel-measurable function $H : \mathbb{R} \to \mathbb{R}$. Hence, by polynomial chaos expansion in section 4, we have

$$h_2(S_{t_2}) = \sum_{k=0}^{\infty} a_k \psi_k(\xi_{t_2}),$$

which is the motivating idea for the present chapter.

Gerhold (2011) has given some intuitive justification of the infinite series representation above and the interpretation of (A2) as a good approximation of the payoff at t_2 (see page 596 in Gerhold (2011)). However, in our setting, the intuitive justification turns into a rigorous one. It is now obvious that the argument in the proof for Theorem 1 in Glasserman and Yu (2004) is available for the single-period problem of geometric Brownian motion. Therefore, theorem 1 in Glasserman and Yu (2004) also holds for the case of geometric Brownian motion. We collect the above observations for the single-period problem with single-underlying asset as a proposition.

Proposition 3.14. *Set $c_\rho = 2log(2 + \sqrt{\rho})$ and $\rho = t_2/t_1$. Suppose (A1) holds. If $K = (1 - \delta)logN/c_\rho$ for some $\delta > 0$, then*

$$\lim_{N \to \infty} \sup_{|\beta|=1} E[| \beta - \hat{\beta} |^2] = 0.$$

If $K = (1 + \delta)logN/c_\rho$ for some $\delta > 0$, then

$$\lim_{N\to\infty} \sup_{|\beta|=1} E[| \beta - \hat{\beta} |^2] = \infty.$$

3.8. Main Results

In this section we present two main results: (1) the convergence of the algorithm when multiple underlying assets are considered and (2) a rate of convergence of the algorithm. To this end we introduce several artificial devices useful for the proof for the main results below. Define for $n \in \{1, 2, ..., m-1\}$

$$C_{K,n} = P_K V_{n+1} = \sum_{k=0}^{K} \beta_{n,k} \boldsymbol{\psi}_k,$$

where P_K is the orthogonal projection onto $span\{\psi_0, ..., \psi_K\}$ and $\beta_{n,k} = E[V_{n+1}\psi_k]$. Define an approximation to backward induction equations as follows: $\bar{C}_{K,m} \equiv 0$,

$$\bar{C}_{K,n} = P_K \bar{V}_{n+1} = \sum_{k=0}^{K} \bar{\beta}_{n,k} \psi_k, \; n \in \{1, ..., m-1\}$$

where $\bar{V}_{n+1} = max\{h_{n+1}, \bar{C}_{n+1}\}$ and $\bar{\beta}_{n,k} = E[\bar{V}_{n+1}\boldsymbol{\psi}_k]$. Finally, define for $n \in \{1, ..., m-1\}$

$$\tilde{C}_{N,K,n} = \sum_{k=0}^{K} \tilde{\beta}_{n,k} \boldsymbol{\psi}_k$$

where $\tilde{\beta}_{n,k} = \frac{1}{N} \sum_{i=1}^{N} \bar{V}_{n+1}(\boldsymbol{S}^i_{n+1}) \boldsymbol{\psi}_k(\boldsymbol{\xi}^i_n)$. Now, we address the single period problem where $m = 2$.

We introduce an assumption needed to derive the main results:

***Assumption* (2)** $E[h_n^4] < \infty$ for each n.

The Assumption (2) is less restrictive than the ones for fourth moment of h in the literature (see, e.g., (B3) in Glasserman and Yu (2004) and theorem 6 in Gerhold (2011)).

We now address the result for the case of the single-period and single underlying asset.

Theorem 3.15. *(i) If $K = \frac{(1-\delta)}{c} logN$ where $c = log3^2$ and $\delta \in (0, 1)$, the algorithm converges in L^2 as $N \to \infty$. (ii) If $K = \frac{(1+\delta)}{c} logN$, $\delta > 0$, the algorithm diverges to the* *infinite in L^2 as $N \to \infty$.*

(Proof) (i) First, we estimate $E[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2]$.

By independence and orthogonality,

$$E[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] = E[\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2].$$

Since $E[\hat{\beta}] = \beta$,

$$\begin{aligned} \textstyle\sum_{k=0}^{K} E[(\hat{\beta}_{1,k} - \beta_{1,k})^2] &= \textstyle\frac{1}{N}\sum_{k=0}^{K} Var(\psi_{1,k}(\xi_1)h_{t_2}(S_2)) \\ &\leq \textstyle\frac{1}{N}\sum_{k=0}^{K} E[\psi_{1,k}^2(\xi_1)h_{t_2}^2(S_2)]. \end{aligned}$$

Thus, by Cauchy-Schwartz and proposition 3.13,

$$E[\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2] \leq C\frac{(K+1)3^{2K}}{N}.$$

Therefore, since $C_{K,1} \to C_1$ in L^2, by triangle inequality, we have $\hat{C}_{N,K,1} \to C_1$ in L^2. Now, we show that $\hat{C}_0(S_0)$ converges to $C_0(S_0) = E[V_1]$ in L^2. Note

$$\begin{aligned} &E[(\hat{C}_0(S_0) - E[V_1])^2] \\ &= E[(\textstyle\frac{1}{N}\sum_{i=1}^{N}\hat{V}_1(S_1^i) - E[V_1])^2] \\ &\leq 2E[(\textstyle\frac{1}{N}\sum_{i=1}^{N}\hat{V}_1(S_1^i) - \frac{1}{N}\sum_{i=1}^{N}V_1(S_1^i))^2] + 2E[(\frac{1}{N}\sum_{i=1}^{N}V_1(S_1^i) - E[V_1])^2] \\ &\leq 2E[(\hat{C}_{N,K,1}(\xi_1^i) - C_1(\xi_1^i))^2] + 2\textstyle\frac{Var(V_1)}{N}. \end{aligned}$$

Then, since the coefficients of $\hat{C}_{N,K,1}$ are independent of ξ_1^i by alteration of step 3, we get the convergence as $N \to \infty$.

(ii) It is enough to address an example showing the divergence.

Let $h_{t_2}(S_2) = (\frac{t_2}{t_1})^{K/2}\psi_{2K}(\xi_2)$ via ϕ. By triangle inequality and the fact that $C_{K,1} \to C_1$

in L^2, it is sufficient to show $E[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2]$ diverges to the infinity. Note

$$\begin{aligned} E[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] &= E[\textstyle\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2] \\ &= \textstyle\sum_{k=0}^{K} Var(\hat{\beta}_k) \\ &= \tfrac{1}{N}\textstyle\sum_{k=0}^{K} E[(\tfrac{t_2}{t_1})^K \psi_{2K}^2(\xi_2)\psi_{1k}^2(\xi_1)] \\ &\quad -\tfrac{1}{N}\textstyle\sum_{k=0}^{K}(E[(\tfrac{t_2}{t_1})^{K/2}\psi_{2K}(\xi_2)\psi_{1k}(\xi_1)])^2. \end{aligned}$$

Then, by (28) in Lemma 1 in Glasserman and Yu (2004),

$$\begin{aligned} E[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] &= E[\textstyle\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2] \\ &= \tfrac{1}{N}\textstyle\sum_{k=0}^{K} E[(\tfrac{t_2}{t_1})^K \psi_{2K}^2(\xi_2)\psi_{1k}^2(\xi_1)] - \tfrac{1}{N} \\ &\geq \tfrac{1}{N}E[(\tfrac{t_2}{t_1})^K \psi_{2K}^2(\xi_2)\psi_{1K}^2(\xi_1)] - \tfrac{1}{N}. \end{aligned}$$

Now, we note that $E[t_2^{K/2}\sqrt{K!}\psi_{2K}(\xi_2)|\xi_1] = t_1^{K/2}\sqrt{K!}\psi_{1K}(\xi_1)$ (see page 2098-2099 in Glasserman and Yu (2004)). Then, by Jensen inequality,

$$\begin{aligned} E[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] &= E[\textstyle\sum_{k=0}^{K}(\hat{\beta}_{1,k} - \beta_{1,k})^2] \\ &\geq \tfrac{1}{N}E[(\tfrac{t_2}{t_1})^K \psi_{2K}^2(\xi_2)\psi_{1K}^2(\xi_1)] - \tfrac{1}{N}. \\ &= \tfrac{1}{N}E[\tfrac{1}{K!}(\tfrac{1}{t_1})^K \psi_{1K}^2(\xi_1) E[(t_2^{K/2}\sqrt{K!}\psi_{2K}(\xi_2))^2|\xi_1]] - \tfrac{1}{N} \\ &\geq \tfrac{1}{N}E[\psi_{1K}^4(\xi_1)] - \tfrac{1}{N}. \end{aligned}$$

Finally, by proposition 3.13, we have

$$E[(\hat{C}_{N,K,1}(\xi_1) - C_{K,1}(\xi_1))^2] \quad \geq \quad C\frac{3^{2K}}{NK} - \frac{1}{N},$$

which completes the proof.

We make a remark about the theorem. With $h_{t_2}(\boldsymbol{S}_2) = (\frac{t_2}{t_1})^{(\sum_j^d (\boldsymbol{\alpha}(K)_j)/2^d} \boldsymbol{\psi}_{2K}(\boldsymbol{\xi}_2)$, the proof for the theorem also holds for the case of multiple underlying assets by the independence of multi-dimensional Hermite polynomials with a generalized result $\mid \boldsymbol{\alpha}(K) \mid = O(logN)$. When $d = 1$ the result is exactly same as one in the above theorem. After Theorem 1 in Glasserman and Yu (2004) the author stated that "This results show rather precisely, from a sample size of N, the highest K for which coefficients of polynomials of order K can be estimated uniformly well is $O(logN)$.". This result is based on numerous unrealistic assumptions that are imposed on the quantities of interest to achieve a convergence to an approximation to the true value of an American option. However, we now states that the critical rate for the true value of American option is $O(logN)$ with the finiteness of fourth moment of payoff function as the only assumption.

We need two lemmas to deal with the multi-period problem.

Lemma 3.16. *(i) For $n = m-1$*

$$\sum_{k=0}^{K} E[(\hat{\beta}_{m-1,k} - \tilde{\beta}_{m-1,k})^2 = 0.$$

(ii) For each $n \in \{1, ..., m-2\}$,

$$\sum_{k=0}^{K} E[(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2] \leq 2^{m-n-1} A_K^{m-n-1} \sum_{l=1}^{m-n-1} \sum_{k=0}^{K} E[(\tilde{\beta}_{m-l,k} - \bar{\beta}_{m-l,k})^2],$$

where $A_K = (K+1)^2 \underset{0 \leq k \leq K}{max} E[\boldsymbol{\psi}_k^4]$.

(Proof) In this proof we drop the boldface notation for convenience. The idea of proof is same as the one for the case of Brownian motion in Glasserman and Yu (2004) although the details are different because of differences in the number of underlying assets and the assumptions on the fourth moments.

(i) It is obvious since $\hat{V}_m(S_m^i) = h_m(S_m^i) = \bar{V}_m(S_m^i)$.

(ii) By Cauchy-Schwartz,

$$
\begin{aligned}
(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2 &= (\tfrac{1}{N}\sum_{i=1}^{N} \hat{V}_{n+1}(S^i_{n+1})\psi_k(\xi^i_n) - \tfrac{1}{N}\sum_{i=1}^{N} \bar{V}_{n+1}(S^i_{n+1})\psi_k(\xi^i_n))^2 \\
&\leq \tfrac{1}{N}\sum_{i=1}^{N} \psi_k^2(\xi^i_n)(\hat{V}_{n+1}(S^i_{n+1}) - \bar{V}_{n+1}(S^i_{n+1}))^2.
\end{aligned}
$$

Then, by noting

$\mid \hat{V}_{n+1}(S^i_{n+1}) - \bar{V}_{n+1}(S^i_{n+1}) \mid$

$$
\begin{aligned}
&= \mid max\{h_{n+1}(S^i_{n+1}), \hat{C}_{N,K,n+1}(S^i_{n+1})\} - max\{h_{n+1}(S^i_{n+1}), \bar{C}_{N,K,n+1}(S^i_{n+1})\} \mid \\
&\leq \mid \hat{C}_{N,K,n+1}(S^i_{n+1}) - \bar{C}_{N,K,n+1}(S^i_{n+1}) \mid,
\end{aligned}
$$

and

$$
\begin{aligned}
(\hat{C}_{N,K,n+1}(S^i_{n+1}) - \bar{C}_{N,K,n+1}(S^i_{n+1}))^2 &= (\textstyle\sum_{k=0}^{K}(\hat{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})\psi_k(\xi^i_{n+1}))^2 \\
&\leq (K+1)\textstyle\sum_{k=0}^{K}(\hat{\beta}_{n+1,k} - \bar{\beta}_{n+1,k})^2\psi_k^2(\xi^i_{n+1}),
\end{aligned}
$$

we have

$$
\begin{aligned}
E[(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2] &\leq (K+1)E[\textstyle\sum_{l=0}^{K} \psi_k^2(\xi^i_n)\psi_l^2(\xi^i_{n+1})(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2] \\
&= (K+1)\textstyle\sum_{l=0}^{K} E[\psi_k^2(\xi^i_n)\psi_l^2(\xi^i_{n+1})]E[(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2] \\
&\leq (K+1)\textstyle\sum_{l=0}^{K} \sqrt{E[\psi_k^4(\xi^i_n)]}\sqrt{E[\psi_k^4(\xi^i_{n+1})]}E[(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2] \\
&\leq (K+1)\max_{0\leq k\leq K} E[\psi_k^4(\xi)]\textstyle\sum_{l=0}^{K} E[(\hat{\beta}_{n+1,l} - \bar{\beta}_{n+1,l})^2].
\end{aligned}
$$

Thus, letting $B_K = (K+1)\underset{0\le k\le K}{max} E[\psi_k^4(\xi)]$*, we have*

$$E[\sum_{k=0}^{K}(\hat{\beta}_{n,k}-\tilde{\beta}_{n,k})^2] \le B_K(K+1)\sum_{k=0}^{K} E[(\hat{\beta}_{n+1,k}-\bar{\beta}_{n+1,k})^2].$$

Let $A_K = (K+1)^2 \underset{0\le k\le K}{max} E[\psi_k^4(\xi)]$*. Then, since*

$$E[(\hat{\beta}_{n+1,k}-\bar{\beta}_{n+1,k})^2] \le 2E[(\hat{\beta}_{n+1,k}-\tilde{\beta}_{n+1,k})^2] + 2E[(\tilde{\beta}_{n+1,k}-\bar{\beta}_{n+1,k})^2],$$

we have

$$E[\sum_{k=0}^{K}(\hat{\beta}_{n,k}-\tilde{\beta}_{n,k})^2] \le 2A_K\sum_{k=0}^{K} E[(\hat{\beta}_{n+1,k}-\tilde{\beta}_{n+1,k})^2] + 2A_K\sum_{k=0}^{K} E[(\tilde{\beta}_{n+1,k}-\bar{\beta}_{n+1,k})^2].$$

By repeating the procedure, we reach

$$\begin{aligned} E[\textstyle\sum_{k=0}^{K}(\hat{\beta}_{n,k}-\tilde{\beta}_{n,k})^2] \quad &\le \quad (2A_K)^{m-n-1}\textstyle\sum_{k=0}^{K} E[(\hat{\beta}_{m-1,k}-\tilde{\beta}_{m-1,k})^2] \\ &\quad + \textstyle\sum_{l=1}^{m-n-1}(2A_K)^{m-n-l}\sum_{k=0}^{K} E[(\tilde{\beta}_{m-l,k}-\bar{\beta}_{m-l,k})^2]. \end{aligned}$$

Therefore, by (i), we finally have

$$E[\sum_{k=0}^{K}(\hat{\beta}_{n,k}-\tilde{\beta}_{n,k})^2] \le 2^{m-n-1}A_K^{m-n-1}\sum_{l=1}^{m-n-1}\sum_{k=0}^{K} E[(\tilde{\beta}_{m-l,k}-\bar{\beta}_{m-l,k})^2],$$

which completes the proof.

Lemma 3.17. *For each* $n \in \{1, ..., m-1\}$*,*

$$\sum_{k=0}^{K} E[(\tilde{\beta}_{n,k}-\bar{\beta}_{n,k})^2] \le C\frac{(K+1)^3}{N}(\max_{0\le k\le K}\sqrt{E[\boldsymbol{\psi}_k^4]} + \max_{0\le k\le K} E[\boldsymbol{\psi}_k^4]).$$

(Proof) Note $E[\tilde{\beta}_{n,k}] = \bar{\beta}_{n,k}$*. Then, for* $n \in \{1, ..., m-2\}$*,*

$$\begin{aligned}
\sum_{k=0}^{K} E[(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2] &= \sum_{k=0}^{K} \frac{1}{N} Var(\psi_k(\xi_n)\bar{V}_{n+1}(S_{n+1})) \\
&\leq \sum_{k=0}^{K} \frac{1}{N} E[\psi_k^2(\xi_n)\bar{V}_{n+1}^2(S_{n+1})] \\
&\leq \sum_{k=0}^{K} \frac{1}{N} E[\psi_k^2(\xi_n) max\{h_{n+1}^2(S_{n+1}), \bar{C}_{n+1}^2(S_{n+1})\}] \\
&\leq \sum_{k=0}^{K} \frac{1}{N} E[\psi_k^2(\xi_n)(h_{n+1}^2(S_{n+1}) + \bar{C}_{n+1}^2(S_{n+1}))] \\
&= \frac{1}{N} \sum_{k=0}^{K} E[\psi_k^2(\xi_n) h_{n+1}^2(S_{n+1})] + \frac{1}{N} \sum_{k=0}^{K} E[\psi_k^2(\xi_n) \bar{C}_{n+1}^2(S_{n+1})].
\end{aligned}$$

Thus, by Cauchy-Schwartz,

$$\begin{aligned}
\sum_{k=0}^{K} E[(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2] &\leq \frac{1}{N} \sum_{k=0}^{K} \sqrt{E[\psi_k^4(\xi)]} \sqrt{E[h_{n+1}^4(S_{n+1})]} \\
&\quad + \frac{1}{N} \sum_{k=0}^{K} E[\psi_k^2(\xi_n)(\sum_{l=0}^{K} \bar{\beta}_{n+1,l}\psi_l(\xi_{n+1}))^2] \\
&\leq \frac{1}{N}(K+1) \max_{1 \leq n \leq m} \sqrt{E[h_n^4(S_n)]} \max_{0 \leq k \leq K} \sqrt{E[\psi_k^4(\xi)]} \\
&\quad + \frac{1}{N}(K+1) \sum_{k=0}^{K} E[\psi_k^2(\xi_n)(\sum_{l=0}^{K} \bar{\beta}_{n+1,l}^2 \psi_l^2(\xi_{n+1}))].
\end{aligned}$$

Then, by noting $\sum_{k=0}^{K} \bar{\beta}_{n+1,k}^2 \leq E[(P_K \bar{V}_{n+2})^2] \leq E[\bar{V}_{n+2}^2]$ *by Plancherel,*

$$
\begin{aligned}
\sum_{k=0}^{K} E[(\tilde{\beta}_{n,k}-\bar{\beta}_{n,k})^2] \quad &\leq \quad \frac{1}{N}(K+1)\max_{1\leq n\leq m}\sqrt{E[h_n^4(S_n)]}\max_{0\leq k\leq K}\sqrt{E[\psi_k^4(\xi)]} \\
&\quad + \frac{1}{N}\left\|\bar{V}_{n+2}\right\|_{L^2}^2 (K+1)\sum_{k=0}^{K} E[\psi_k^2(\xi_n)(\sum_{l=0}^{K}\psi_l^2(\xi_{n+1}))] \\
&\leq \quad \frac{1}{N}(K+1)\max_{1\leq n\leq m}\sqrt{E[h_n^4(S_n)]}\max_{0\leq k\leq K}\sqrt{E[\psi_k^4(\xi)]} \\
&\quad + \frac{1}{N}E[\bar{V}_{n+1}^2](K+1)^3\max_{0\leq k\leq K}E[\psi_k^4(\xi)] \\
&\leq \quad C(K+1)^3\frac{1}{N}(\max_{0\leq k\leq K}\sqrt{E[\psi_k^4(\xi)]}+\max_{0\leq k\leq K}E[\psi_k^4(\xi)])
\end{aligned}
$$

for some constant $C > 0$. Since $max\{h_m^2(S_m), \bar{C}_m^2(S_m)\} = h_m^2(S_m)$ the estimate also holds for $n = m-1$, which completes the proof.

We are now in position to address the result for the case of multiple underlying assets.

Theorem 3.18. *Suppose assumption (1) and (2) hold. Then, if $K = \frac{(1-\delta)}{c}logN$ where $c = log3^{2m}$ and $\delta \in (0,1)$, the algorithm converges to the true value of American option in L^2 as $N \to \infty$.*

(Proof) Note for each $n \in \{1, ..., m-1\}$,

$E[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2]$

$$
\begin{aligned}
&\leq \quad 4(E[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n)-\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2]+E[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n)-\bar{C}_{K,n}(\boldsymbol{\xi}_n))^2] \\
&\quad + E[(\bar{C}_{K,n}(\boldsymbol{\xi}_n)-C_{K,n}(\boldsymbol{\xi}_n))^2]+E[(C_{K,n}(\boldsymbol{\xi}_n)-C_n(\boldsymbol{\xi}_n))^2]).
\end{aligned}
$$

We estimate each term of right-hand-side of inequality above:

(i) $E[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - \tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2]$

Note, by independence,

$$E[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - \tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2] = E[(\textstyle\sum_{k=0}^{K}(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})\psi_k(\boldsymbol{\xi}_n))^2]$$

$$= E[\textstyle\sum_{k=0}^{K}(\hat{\beta}_{n,k} - \tilde{\beta}_{n,k})^2].$$

Then, by lemmas and proposition,

$E[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - \tilde{C}_{N,K,n}(\boldsymbol{\xi}_n))^2]$

$$\leq 2^{m-n-1} A_K^{m-n-1} \textstyle\sum_{l=1}^{m-n-1} \sum_{k=0}^{K} E[(\tilde{\beta}_{m-l,k} - \bar{\beta}_{m-l,k})^2]$$

$$\leq C\tfrac{A_K^{m-n-1}(K+1)^3}{N}(\max_{0\leq k\leq K}\sqrt{E[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})]} + \max_{0\leq k\leq K} E[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})])$$

$$\leq C\tfrac{(K+1)^{2m-2n+1}}{N}3^{2(m-n)K}$$

$$\leq C\tfrac{(K+1)^{2m}3^{2m|\boldsymbol{\alpha}(K)|}}{N}.$$

(ii) $E[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n) - \bar{C}_{K,n}(\boldsymbol{\xi}_n))^2]$

Since $E[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n) - \bar{C}_{K,n}(\boldsymbol{\xi}_n))^2] = E[\sum_{k=0}^{K}(\tilde{\beta}_{n,k} - \bar{\beta}_{n,k})^2]$*, by lemma and proposition,*

$$E[(\tilde{C}_{N,K,n}(\boldsymbol{\xi}_n) - \bar{C}_{K,n}(\boldsymbol{\xi}_n))^2] \leq C(K+1)^3\tfrac{1}{N}(\max_{0\leq k\leq K}\sqrt{E[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})]} + \max_{0\leq k\leq K} E[\boldsymbol{\psi}_k^4(\boldsymbol{\xi})])$$

$$\leq C\tfrac{(K+1)^3 3^{2|\boldsymbol{\alpha}(K)|}}{N}.$$

(iii) $E[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2]$ *and* $E[(C_{K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2]$

Note

$$E[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] = \textstyle\sum_{k=0}^{K}(\bar{\beta}_{n,k} - \beta_{n,k})^2$$

$$= \textstyle\sum_{k=0}^{K}(E[(\bar{V}_{n+1}(\boldsymbol{S}_{n+1}) - V_{n+1}(\boldsymbol{S}_{n+1}))\boldsymbol{\psi}_k(\boldsymbol{\xi}_n)])^2.$$

Then, by Plancherel (see Folland (1999) or Weiss and Mcdonald (2012))

$$\begin{aligned}
E[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] &= E[(P_K(\bar{V}_{n+1}(\boldsymbol{S}_{n+1}) - V_{n+1}(\boldsymbol{S}_{n+1}))^2] \\
&\leq E[(\bar{V}_{n+1}(\boldsymbol{S}_{n+1}) - V_{n+1}(\boldsymbol{S}_{n+1}))^2] \\
&\leq E[|\ max\{h_{n+1}(\boldsymbol{S}_{n+1}), \bar{C}_{K,n+1}(\boldsymbol{\xi}_{n+1})\} \\
&\quad -max\{h_{n+1}(\boldsymbol{S}_{n+1}), C_{n+1}(\boldsymbol{\xi}_{n+1})\}\ |^2] \\
&\leq E[(\bar{C}_{K,n+1}(\boldsymbol{\xi}_{n+1}) - C_{n+1}(\boldsymbol{\xi}_{n+1}))^2] \\
&\leq 2E[(\bar{C}_{K,n+1}(\boldsymbol{\xi}_{n+1}) - C_{K,n+1}(\boldsymbol{\xi}_{n+1}))^2] \\
&\quad +2E[(C_{K,n+1}(\boldsymbol{\xi}_{n+1}) - C_{n+1}(\boldsymbol{\xi}_{n+1}))^2].
\end{aligned}$$

Thus, by repeating the procedure, we have

$$\begin{aligned}
E[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] &\leq 2^{m-n-1}E[(\bar{C}_{K,m-1}(\boldsymbol{\xi}_{m-1}) - C_{K,m-1}(\boldsymbol{\xi}_{m-1}))^2] \\
&\quad + \textstyle\sum_{l=1}^{m-n-1} 2^{m-n-l}E[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].
\end{aligned}$$

Since $\bar{V}(\boldsymbol{S}_m) = h_m(\boldsymbol{S}_m) = V_m(\boldsymbol{S}_m)$,*we have*

$$E[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] \leq 2^{m-n-1} \sum_{l=1}^{m-n-1} E[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].$$

Thus,

$E[(\bar{C}_{K,n}(\boldsymbol{\xi}_n) - C_{K,n}(\boldsymbol{\xi}_n))^2] + E[(C_{K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2]$

$$\leq \quad 2^{m-n-1} \sum_{l=1}^{m-n} E[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].$$

By (i), (ii) and (iii), we finally reach

$$E[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n) - C_n(\boldsymbol{\xi}_n))^2] \quad \leq \quad C_1 \frac{(K+1)^{2m} 3^{2m|\boldsymbol{\alpha}(K)|}}{N}$$

$$+ C_2 \sum_{l=1}^{m-n} E[(C_{K,m-l}(\boldsymbol{\xi}_{m-l}) - C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].$$

Therefore, as $N \to \infty$, $\hat{C}_{N,K,n}(\boldsymbol{\xi}_n)$ *converges to* $C_n(\boldsymbol{\xi}_n)$ *in* L^2 *for each* $n \in \{1, ..., m-1\}$, *which completes the proof.*

We make several remarks about the theorem. For the case of multi-periods and single underlying asset, we observe that the continuation value function at t_{m-1} is same as the one in single-period problem. Thus, the critical relation $O(logN)$ for single-period problem still holds for the multi-periods problem. Since the observation is also true for the case of multiple underlying assets, the critical relation $O(logN)$ still hods for this case. Therefore, for any case, the critical relation is $O(logN)$ for geometric Brownian motion. Furthermore, we note that the proof still holds for correlated Brownian motion by using proper transformation ϕ. Therefore, we conclude that the highest K to achieve convergence is $O(logN)$ for any case.

Next, we present a rate of convergence of the algorithm considering the regularity of continuation value function C_n. To this end we need a lemma. First, we deal with the one-dimensional case.

Lemma 3.19. *For any positive integer* r, *if* $v \in H^r_\omega(\Lambda)$, *then, for sufficiently large* K,

$$\|v - P_K v\|_{L^2_\omega(\Lambda)} \leq \frac{1}{\sqrt{(K+1)K \cdots (K-r+2)}} \|v\|_{H^r_\omega(\Lambda)}.$$

(Proof) We note, by Plancherel,

$$\|v - P_K v\|^2_{L^2_\omega(\Lambda)} = \sum_{l=K+1}^{\infty} a_l^2,$$

where $a_l = \int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)dx,\ \omega(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. *Now, by (3.5.1), (3.5.2) and integration by parts,*

$$\begin{aligned}\int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)dx &= \int_{\mathbb{R}} v(x)[\tfrac{x}{\sqrt{l}}\psi_{l-1}(x) - \tfrac{\sqrt{l-1}}{\sqrt{l}}\psi_{l-2}(x)]\omega(x)dx \\ &= \tfrac{1}{\sqrt{l}}\int_{\mathbb{R}} xv(x)\psi_{l-1}(x)\omega(x)dx - \tfrac{1}{\sqrt{l}}\int_{\mathbb{R}} v(x)\psi'_{l-1}(x)\omega(x)dx \\ &= \tfrac{1}{\sqrt{l}}\int_{\mathbb{R}} v'(x)\psi_{l-1}(x)\omega(x)dx.\end{aligned}$$

By repeating the calculation, we have

$$\int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)dx = \frac{1}{\sqrt{l}\sqrt{l-1}\cdots\sqrt{l-r+1}}\int_{\mathbb{R}} v^{(r)}(x)\psi_{l-r}(x)\omega(x)dx.$$

Thus,

$$\left|\int_{\mathbb{R}} v(x)\psi_l(x)\omega(x)dx\right| \leq \frac{1}{\sqrt{(K+1)K\cdots(K-r+2)}}\left|\int_{\mathbb{R}} v^{(r)}(x)\psi_{l-r}(x)\omega(x)dx.\right|$$

Therefore,

$$\|v - P_K v\|^2_{L^2_\omega(\Lambda)} \leq \frac{1}{\sqrt{(K+1)K\cdots(K-r+2)}}\left\|v^{(r)}\right\|^2_{L^2_{\omega(\Lambda)}},$$

which completes the proof.

If C_n is in $H^r_\omega(\Lambda)$ for each r, the error $E[(C_{K,n}-C_n)^2]$ converges faster than any polynomial order and we may expect exponential decay of the error in L^2. One can find a same result for $\omega(x) = e^{-x^2}$ in Guo (1999). However, the proof there is not applicable here.

We now address a convergence rate for the case where $d = 1$.

Theorem 3.20. *Suppose* $K = \frac{(1-\delta)}{c}logN$. *Then, if* $C_n \in H^r_\omega(\Lambda)$ *for some positive integer* r *and each* $n = 1, ..., m-1$, *then the algorithm converges at least as fast as* $O((logN)^{-r/2})$ *in*

L^2.

(Proof) By the proof of theorem 3.18,

$$E[(\hat{C}_{N,K,n}(\xi_n) - C_n(\xi_n))^2] \leq C_1 \frac{K^{2m}3^{2mK}}{N} + C_2 \sum_{l=2}^{m-n} E[(C_{K,m-l}(\xi_{m-l}) - C_{m-l}(\xi_{m-l}))^2].$$

Then, assumptions and lemma 3.19 complete the proof.

We now consider the multi-dimensional case. To this end, we add one more condition to our multi-index scheme, the graded lexicographic order; given an expansion order L, we use a truncated basis $\{\boldsymbol{\psi}_k : |\alpha(k)| \leq L\}$. With this new scheme we thus have $\sum_{k=0}^{K} a_k \boldsymbol{\psi}_k$ where $1 + K = \frac{(d+L)!}{d!L!}$ for $L = 0, 1, 2, ...$

Lemma 3.21. *For any positive integer r, if $v \in H^r_\omega(\Lambda^d)$, for sufficiently large K,*

$$\|v - P_K v\|_{L^2_\omega(\Lambda^d)} \leq \frac{1}{\sqrt{(L/d)(L/d-1)\cdots(L/d-(r-1))}} \|v\|_{H^r_\omega(\Lambda^d)}.$$

(Proof) We note, by Plancherel,

$$\|v - P_L v\|^2_{L^2_\omega(\Lambda)} = \sum_{|\alpha(k)|>L} a^2_{\alpha(k)},$$

where $a_{\alpha(k)} = \int_{\mathbb{R}^d} v(x)\psi_{\alpha(k)_1}(x_1)\cdots\psi_{\alpha(k)_d}(x_d)\omega(x)dx$. For $|\alpha(k)| = L+1$, there exists at least one component $\alpha(k)_i$ such that $\alpha(k)_i \geq \frac{|\alpha(k)|}{d}$. Suppose, for some k, $\alpha(k)_1 \geq (L+1)/d$. Then, by lemma 3.19, we have

$\int_{\mathbb{R}^d} v(x)\psi_{\alpha(k)_1}(x_1)\cdots\psi_{\alpha(k)_d}(x_d)\omega(x)dx$

$$\leq \frac{1}{\sqrt{(\frac{L+1}{d})(\frac{L+1}{d}-1)\cdots(\frac{L+1}{d}-(r-1))}} \int_{\mathbb{R}^d} \frac{\partial^{(r)}v(x)}{\partial x_1}\psi_{\alpha(x)_1-r}(x_1)\psi_{\alpha(k)_2}(x_2)\cdots\psi_{\alpha(k)_d}(x_d)\omega(x)dx.$$

Hence, we have

$$\|v - P_L v\|^2_{L^2_\omega(\Lambda^d)} \leq \frac{1}{(\frac{L}{d})(\frac{L}{d}-1)\cdots(\frac{L}{d}-(r-1))} \sum_{i=1}^{d} \|\frac{\partial^{(r)}v(x)}{\partial x_i}\|^2_{L^2_{\omega(\Lambda^d)}},$$

which completes the proof.

We address a rate of convergence of the algorithm for multi-dimensional case.

Theorem 3.22. *Suppose* $L = \frac{(1-\delta)}{c} logN$*. Then, if* $C_n \in H^r_\omega(\Lambda^d)$ *for some positive integer* r *and each* $n = 1, ..., m-1$*, then the algorithm converges at least as fast as* $O((logN)^{-r/2})$ *in* L^2*.*

(Proof) By the proof of theorem 3.18,

$$E[(\hat{C}_{N,K,n}(\boldsymbol{\xi}_n)-C_n(\boldsymbol{\xi}_n))^2] \leq C_1 \frac{(\frac{(d+L)!}{d!L!})^{2m} 3^{2mL}}{N} + C_2 \sum_{l=2}^{m-n} E[(C_{K,m-l}(\boldsymbol{\xi}_{m-l})-C_{m-l}(\boldsymbol{\xi}_{m-l}))^2].$$

Then, assumption and lemma 3.21 complete the proof.

3.9. Concluding Remarks

We improved the results in Glasserman and Yu (2004) in three different ways. First, we prove the L^2-convergence of the quasi-regression Monte Carlo method to the true price of American option under the multiple underlying assets where the number of paths and number of basis functions increase together. Second, we show, given N simulated paths, the highest possible number of basis functions necessary for obtaining convergence is $O(logN)$ in both Brownian motion and geometric Brownian motion cases. Furthermore, this holds even in the case of multiple underlying assets. This implies the importance of the proper choice of basis functions in implementing the method. Finally, we propose a rate of convergence considering the regularity of the continuation value function.

References

ABRAMOWITZ, M. and I.A. STEGUN (1972): *Handbook of Mathematical Functions*, *9th ed.*, New York:Dover Publications.

ADAMS, R. A. (1975): *Sobolev Spaces*, New York:Academic Press.

CLEMENT, E., D. LAMBERTON, and P. PROTTER (2002): An analysis of a least squares regression method for American option pricing, *Finance Stoch.* 6, 449-471.

DURRET, R. (2010): *Probability*: *Theory and Examples*, *Fourth ed.*, Cambridge: Cambridge University Press.

EGLOFF, D. (2005): Monte Carlo algorithms for optimal stopping and statistical learning, *Ann. Appl. Probab.* 15, 1396-1432.

EGLOFF, D., M. KOHLER, and N. TODOROVIC (2007): A dynamic look-ahead Monte Carlo algorithmfor pricing Bermudan options, *Ann. Appl. Probab.* 17, 1138-1171.

ERNST, O. G., A. MUGLER., H. J. STARKLOFF, and E. ULLMANN (2012): On the convergence of generalized polynomial chaos expansions, *ESIM*: *Math. Modell. and Num. Analy.* 46, 317-339.

FOLLAND, G. B. (1999): *Real analysis*: *Modern techniques and their applications.* Pure and Applied Mathematics, Second ed., New York:John Wiley \&Sons Inc.

GERHOLD, S. (2011): The Longstaff-Schwartz algorithm for Levy models: Results on fast and slow convergence, *Ann. Appl. Probab.* 21, 589-608.

GLASSERMAN, P. (2004): *Monte Carlo Methods in Financial Engineering*, New York : Springer.

GLASSERMAN, P. and B. YU (2004): Number of paths versus number of basis functions in American option pricing, *Ann. Appl. Probab.* 14, 2090-2119.

GUO, B. Y. (1999): Error estimation of Hermite spectral method for nonlinear partial differential equations, *Mathematics of Computation*, 68, 1067-1078.

GUO, B. Y. and C. L. XU (2003): Hermite spectral and pseudospectral methods for nonlinear partial differential equations, *Computational and Applied Mathematics* 22, 167-193

JANSON, O. (1997): *Gaussian Hilbert Spaces*, Cambridge : Cambridge University Press.

KORN, R., E. KORN, and T. A. KROISANDIT (2010): *Monte Carlo Methods and Models in Finance and Insurance*: CRC.

LARS, L. C. (2002): L^p-norms of Hermite polynomials and an extremal problem on Wiener chaos, *Ark. Mat.*, 40, 133-144.

LONGSTAFF, F. A. and E.S. SCHWARTZ (2001): Valuing American options by simulation: A simple least-squares approach, *Rev. Financial Stud*, 14, 113-148.

SEPPALAINEN, T. (2012): *Basics of Stochastic Analysis*, University of Wisconsin-Madison.

STENTOFT, L. (2004): Convergence of the least squares Monte Carlo Approach to American option valuation, *Manag. Sci.* 50, 1193-1203.

WEISS, N. A. and J. N. MCDONALD (2012): A *Course in Real Analysis*, *2nd ed.*, Oxford :Academic Press.

XIU, D. (2010): *Numerical Methods for Stochastic Computations*: A *Spectral Method Approach*, Princeton: Princeton University Press.

ZANGER, D. Z. (2013): Quantitative error estimates for a least-squares Monte Carlo algorithm for American option pricing, *Finance Stoch.* 17, 503-534.

Part IV.

CONCLUSION AND SUMMARY

In this dissertation we studied stochastic simulation from the perspective of financial applications. The first topic was development of an optimal importance sampling scheme for the Laplace transform of exponential Brownian functional that is an important quantity in many disciplines such as statistical physics and mathematical finance.

To this end, we first specify a class of probabilities over which we choose a measure for importance sampling. For this purpose, we utilize the Cameron-Martin theorem that says each element of Cameron-Martin space induces a distribution equivalent to the distribution on the Wiener space induced by a Brownian motion. Then, we focus on the asymptotic of the second moment in variance rather than considering the sub-optimization problem of it over the specified class. That is, we generalize the optimization problem to one of finding a class of asymptotically optimal estimators for the expected values of random quantities slightly perturbed from the original one.

Schilder 's theorem tells us that the distributions on Wiener space induced by the small perturbations of a Brownian motion satisfy large deviations principle with a good rate function. Then, by Varadhan's theorem, we have an asymptotic result of the second moment when the perturbations are small. Noting the class of degenerate estimators achieves the lowest limit of asymptotics, our problem finally reduces to finding an element of Carmeron-Martin space inducing a class of estimators which achieves the same asymptotic behavior as the degenerate one. Indeed, the minimax theorem allows us to have such an element in Carmeron-Martin space. Moreover, the calculus of variations problem induced by the asymptotic of the degenerate estimator admits a closed-form solution. We summarize the results:

- We develop a Monte Carlo method for the Laplace transform of exponential Brownian functionals.
- The scheme utilizes large deviations theory to derive a calculus of variations problem the solution of which results in asymptotically optimal importance measure to be used to sample for Monte Carlo simulations.
- The main point is that the calculus of variations problem admits closed-form solutions.

Thus, the computational effort for the method is parsimonious.

- We also address a path to the test of regularity of optimal drift which plays an important role in implementing the proposed method.
- Numerical tests are provided through the Dothan bond pricing model, which shows the method is significantly effective in variance reduction.

The second topic is to improve the results on the convergence of quasi-regression Monte Carlo method for pricing of American option. In Glasserman and Yu (2004) the authors showed that a variant of least-squares Monte Carlo method for American option pricing by Longstaff and Schwartz (2001) converges to an approximation to the true price of American option and the number of simulation paths grows exponentially in the number of basis functions to obtain convergence for Brownian motion (however, faster for geometric Brownian motion) under single underlying asset.

In this dissertation we improve the results in Glasserman and Yu (2004) in three directions:

- We provide a more general proof that the method indeed converges to the true price of American option under multiple underlying assets.
- We show, again under simplified assumptions, that the highest possible number of basis functions for N paths is $O(logN)$ in both Brownian motion and geometric Brownian motion cases even under multiple underlying assets.
- A rate of convergence is provided considering the regularity of the continuation value function.

To achieve first two results in the list above we introduce a polynomial chaos expansion of L^2-functions, which says that the assumptions for Brownian motion case in Glasserman and Yu (2004) still holds for geometric Brownian motion case via a proper choice of basis functions. Then, we proved the desired results with help from an asymptotic result for

Hermite polynomials. Moreover, by presenting L^2-error bounds for orthogonal projection of value functions to span of Hermite polynomials depending on its regularities, we derived the last result in the list.

For further research, one question is to find a sharper convergence rate than one proposed in this paper. It amounts to finding a sharper bound on error between finite truncation of the continuation value function and the true function. It could be a challenging problem. Another question is an extension of the results in Gerhold (2011) in the same manner here. The critical difficulty could be how to get L^p-asymptotics on the basis function used in Gerhold (2011) like one in Appendix B.1.

Appendix

A. Calculus of Variations

A.1. Tonelli Theorem

Theorem A.1. *(Tonelli) Let the Lagrangian* $\Lambda(t, x, \nu)$ *be continuous, convex in* ν*, and coercive of degree* $r > 1$*: for certain constants* $\alpha > 0$ *and* β *we have*

$$\Lambda(t, x, \nu) \geq \alpha \mid \nu \mid^r + \beta \; \forall (t, x, \nu) \in [a, b] \times \mathbb{R}^n \times \mathbb{R}^n.$$

Then the probelm: minimize $J(x) : x \in AC[a, b]$*,* $x(a) = A$*, x(b)=B.*

A.2. Nagumo Growth Condition

Definition A.2. We say Λ has Nagumo growth along x_* if there exists a functon $\theta : \mathbb{R}_+ \to \mathbb{R}$ satisfying $lim_{t \to \infty} \theta(t)/t = +\infty$ such that

$$t \in [a, b], \; \nu \in \mathbb{R}^n \Longrightarrow \Lambda(t, x_*(t), \nu) \geq \theta(\mid \nu \mid).$$

Theorem A.3. *Let* Λ *admit gradients* Λ_x*,* Λ_ν *which, along with* Λ*, are continuous in* (t, x, v)*. Suppose further that for every bounded set* S *in* $\mathbb{R}^n$*, there exist a constant* c *and a summable functon* d *such that, for all* $(t, x, \nu) \in [a, b] \times S \times \mathbb{R}^n$*, we have*

$$\mid \Lambda_x(t, x, \nu) \mid + \mid \Lambda_\nu(t, x, \nu) \mid \leq c(\mid \nu \mid + \mid \Lambda(t, x, \nu) \mid) + d(t).$$

If $\Lambda(t, x, \nu)$ *is convex in* ν *and has Nagumo growth along* x_**, then* x_* *is Lipschitz.*

A.3. Hilbert-Weierstrass Theorem

Theorem A.4. *(Hilbert-Weierstrass) Let* $x_* \in Lip[a, b]$ *satisfy the integral Euler equation, where* Λ *is of class* $C^m(m \geq 2)$ *and satisfies*

$$t \in [a, b],\ \nu \in \mathbb{R}^n \Longrightarrow \Lambda_{\nu\nu}(t, x_*(t), \nu) > 0(positive\ definite).$$

Then x_* *belongs to* $C^m[a, b]$.

B. Other Mathematical Results

B.1. Asymptotics of Hermite Polynomials

Theorem B.1. *As $k \to \infty$,*

$$(\int_{\mathbb{R}} \mid \psi_k(x) \mid^4 \omega(x)dx)^{1/4} = \frac{C}{k^{1/4}} 3^{k/2}(1 + O(\frac{1}{k}))$$

where $C = (\frac{2}{\pi})^{1/4}(\frac{3}{4})^{3/8}$.

B.2. Minimax Theorem

Theorem B.2. *Let $\mathcal{K}$ be a compact convex subset of a Hausdorff topological vector space $\mathcal{X}$ and C be a convex subset of a vector space $\mathcal{Y}$. Let f be a real-valued function defined on $\mathcal{K} \times C$ such that*

(i) $x \mapsto f(x, y)$ is convex and lower-semicontinuous for each y.

(ii) $y \mapsto f(x, y)$ is concave for each x.

Then

$$\inf_{x \in \mathcal{K}} \sup_{y \in C} f(x, y) = \sup_{y \in C} \inf_{x \in \mathcal{K}} f(x, y).$$

Vita

Je Guk Kim was born in Seoul, Republic of Korea, on April 13, 1976, the son of Kim Dong Soo and Kim Bok Hee. He completed the degree of Bachelor of Business Administration at Hongik University in Seoul on February 2002. He served military service for two years (1998-2000). In March 2003, he entered the graduate school in the Department of Mathematics at Yonsei University in Seoul. After receiving the degree of Master of Science in Mathematics he moved to University of Texas at Austin for graduate study in economics. He completed there the degree of Master of Science in Economics on May 2009. After that he moved to the University of Tennessee at Knoxville to continue his graduate study toward a doctoral degree in Management Science at the Department of Business Analytics and Statistics. He received a PhD in Management Science in May 2015.