ABSTRACT

Title of dissertation: STOCHASTIC OPTIMIZATION: ALGORITHMS AND CONVERGENCE

Xiaoping Xiong, Doctor of Philosophy, 2005

Dissertation directed by: Professor Michael C. Fu
Robert H. Smith School of Business

Stochastic approximation is one of the oldest approaches for solving stochastic optimization problems. In the first part of the dissertation, we study the convergence and asymptotic normality of a generalized form of stochastic approximation algorithm with deterministic perturbation sequences. Both one-simulation and two-simulation methods are considered. Assuming a special structure on the deterministic sequence, we establish sufficient conditions on the noise sequence for a.s. convergence of the algorithm and asymptotic normality. Finally we propose ideas for further research in analysis and design of the deterministic perturbation sequences.

In the second part of the dissertation, we consider the application of stochastic optimization problems to American option pricing, a challenging task particularly for high-dimensional underlying securities. For options where there are a finite number of exercise dates, we present a weighted stochastic mesh method that only

requires some easy-to-verify assumptions and a method to simulate the behavior of underlying securities. The algorithm provides point estimates and confidence intervals for both price and value-at-risk. The estimators converge to the true values as the computational effort increases.

In the third part, we deal with an optimization problem in the field of ranking and selection. We generalize the discussion in the literature to a non-Gaussian correlated distribution setting. We propose a procedure to locate an approximate solution, which can be shown to converge to the true solution asymptotically. The convergence rate is also provided for the Gaussian setting.

STOCHASTIC OPTIMIZATION: ALGORITHMS AND CONVERGENCE RESULTS

by

Xiaoping Xiong

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Commmittee:

Professor Michael C. Fu, Chair/Advisor
Professor Frank B. Alt
Professor Dilip B. Madan
Dr. I-Jeng Wang
Professor Tobias Von Petersdorff

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Professor Michael Fu, for his support, patience, and encouragement throughout my graduate studies. It is not often that one finds an advisor that always finds the time for listening to the little problems and roadblocks that unavoidably crop up in the course of performing research. His technical and editorial advice was essential to the completion of this dissertation and has taught me innumerable lessons and insights on the workings of academic research in general.

I am also grateful to Dr. I-Jeng Wang for his invaluable advice, support and thought-provoking ideas without which the stochastic approximation part of my work would have been Mission Impossible.

I am indebted to Professor Dilip Madan for his great help and advice in my job search.

I would like to express my deep appreciation to Professor Frank Alt and Professor Tobias Von Petersdorff for gladly serving on my dissertation committee.

The friendship of Dr. Xing Jin is also much appreciated and has led to many interesting and good-spirited discussions relating to this research.

Last, but not least, I would like to thank my parents with my deepest gratitude and love for their dedication and the many years of support during my undergraduate studies that provided the foundation for this work.

TABLE OF CONTENTS

Chapter 1

Introduction

Many realistic problems can be formulated as a stochastic optimization problem or be solved via algorithms involving stochastic optimization techniques. In this dissertation, we consider stochastic optimization in three different contexts: convergence analysis for a class of existing stochastic approximation algorithms; application to the pricing of American-style options; and application to the optimal allocation of a simulation computing budget. In the latter two cases, new provably convergent algorithms are proposed.

Stochastic approximation can be dated back to Robbins and Monro (1951). The authors put forward a new zero-location problem where a real-valued function defined on a continuous domain $M(x)$ is monotone and can only be estimated via noisy observations of some random variable $\xi(x)$. A new algorithm (RM) is proposed, which starts from an arbitrary constant x_1 and changes value recursively via $x_{n+1} = x_n - a_n\xi_n$, where ξ_n has the distribution of $\xi(x)$ given $x = x_n$, and $\{a_n\}$ is a fixed sequence of positive constants such that $0 < \sum_{1}^{\infty} a_n^2 < \infty$. With appropriate assumptions on function M and sequence $\{a_n\}$, x_n is shown to converge to the true zero in L^2 sense. This work spawned hundreds of papers in the following fifty years. Roughly speaking, almost all the work that has been done since then follows one of two main directions. One direction is to adapt the algorithm to solve more general problems or to solve problems more efficiently. The other involves theoretic analysis of the algorithms. Following these two directions, we will briefly review the development of stochastic approximation (SA) algorithms.

A natural extension of the original zero-location problem is to consider continuous stochastic optimization problems. Obviously a zero-location problem turns into a optimization problem if $M(x)$ can be expressed as derivative of a convex function $L(x)$. However, a pure RM algorithm is not applicable if we assume only $L(x)$, instead of $M(x)$, can be estimated via noisy observations given x. To solve this stochastic optimization problem, a new algorithm (KW) is proposed in Kiefer and Wolfowitz (1952). KW replaces ξ_n in RM's recursive formula with a finite-difference estimator $\frac{y_n^+ - y_n^-}{c_n}$, where y_n^+ and y_n^- are noisy observations of $L(x_n + c_n)$ and $L(x_n - c_n)$, respectively. Another step size $\{c_n\}$ is introduced because of finite difference approximation. To guarantee L^2 convergence of the algorithm, the two positive step sizes are assumed to satisfy $a_n, c_n \to 0$, $\sum a_n = \infty$, $\sum a_n c_n < \infty$ and $\sum a_n^2 c_n^{-2} < \infty$.

Both algorithms are extended to the multidimensional case in Blum (1954). Extension of the RM algorithm is relatively straightforward, whereas that of KW needs more work. The KW finite-difference $y_n^+ - y_n^-$ is replaced by a vector $Y_n \equiv [(y_n^1 - y_n), \cdots, (y_n^p - y_n)]$, where $y_n, y_n^1, \cdots, y_n^p$, are $p+1$ independent observations of $L(x_n), \text{Ł}(x_n + c_n u_1), \cdots, L(x_n + c_n u_p)$, respectively, and $\{u_1, \cdots, u_p\}$ represents the orthonormal set spanning $\mathbb{R}^p$. However, since Blum's extension of KW requires $p+1$ independent observations for each iteration, the computational burden could become prohibitive when the dimension of the problem is high, so various approaches have been proposed to circumvent the problem. One algorithm (random directions KW) estimates the directional derivatives along a sequence of randomized directions; see, for example, [23], [44] and [55]. Spall (1992) presents a simultaneous perturbation stochastic approximation (SPSA) algorithm using a simultaneous perturbation for gradient estimation. Both RDKW and SPSA require only two observations at each iteration. According to comparisons in Chen (1997), RDKW and SPSA give satisfactory approaches, both theoretically and practically, to the problem of searching optimizer via stochastic approximation.

Efforts to justify the algorithms are taken to establish two types of results: theoretical convergence of the algorithm, and asymptotic normality or convergence rate. In terms of theoretical convergence, much work has been done since the original papers on RM and KW algorithms; see, for example, Benveniste (1990), Kushner and Clark (1978), Ljung et al. (1992), and further references contained therein. Worth special mention is work in Wang et al. (1996, 1997) and Kulkarni (1996). They for the first time propose equivalent necessary and sufficient conditions on noise sequences for SA algorithms. Work on asymptotic normality begins with Chung (1954), who first gives results on the asymptotic distribution of RM and KW algorithms. Further work can be found in Burkholder (1956), Derman (1956) and Sacks (1958). Fabian (1968) gives a simple proof of asymptotic normality, and most of the current discussion is based on that result.

Both RDKW and SPSA algorithms randomly perturb all parameter components in two parallel simulations at each iteration for any $p-$ dimensional problem. An SPSA requiring only one simulation at each iteration has also been proposed in Spall (1997). These algorithms all rely on proper randomization to avoid the large number simulations required at each iteration, and at the same time move along the gradient descent direction on average. Similar in spirit to the use of low-discrepancy sequences in quasi-Monte Carlo integration (Niederreiter 1992), applications of deterministic sequences in randomized direction SA have been investigated recently with some success, including Sandilya and Kulkarni (1997) for a two-simulation RDKW algorithms and Bhatnagar et al. (2002) for two-timescale SPSA algorithms. The numerical simulations results reported in Bhatnagar et al. (2002) are particularly encouraging in that significant performance advantages over the random Bernoulli perturbation sequences were consistently observed. In Chapter 2, we present a generalized form of the stochastic approximation algorithm, of which SPSA and RDKW are just special cases. We then provide an asymptotic analysis, almost sure convergence and convergence rate of the generalized form with deterministic sequences, assuming a specified structure. Finally we discuss how to construct such a specified deterministic perturbation sequence.

The second field where we apply the stochastic optimization techniques is American options pricing. An option is a contract, or a provision of a contract, that

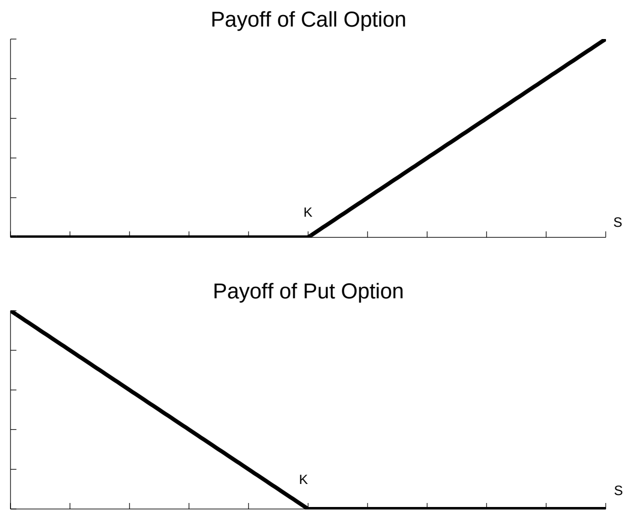

Figure 1.1: Payoff of Options with Strike K

gives one party (the option holder) the right, but not the obligation, to perform a specified transaction with another party (the option issuer or option writer). Option contracts take many forms. The two most common are call options, which provide the holder the right to purchase an underlier at a specified price, and put options, which provide the holder the right to sell an underlier at a specified price. The specified prices for call (put) options here are strike prices. Let S and K represent the underlier's price and the strike price at exercise day, respectively. Obviously the payoff on the exercise date will be $\max(S - K, 0)$ for the holder of a call option and $\max(K - S, 0)$ for the holder of a put option. Figure 1.1 illustrates the payoff functions. The last date on which an option can be exercised is called the expiration date. Options may allow for one of two main forms of exercise: with American exercise, the option can be exercised at any time up to the expiration date; with European exercise, the option can be exercised only on the expiration date.

Option pricing theory–also called Black-Scholes theory or derivatives pricing theory–traces its roots to Bachelier (1900), where Brownian motion is used to model options on French government bonds. Research picks up in the 1960s. Typical of efforts during this period is Samuelson (1965), who considers long-term equity options, and uses geometric Brownian motion to model the random behavior of the underlying stock. Based upon this, he models the random value of the option at exercise. Then Black and Scholes (1973) propose a completely new approach. They derive a partial differential equation for valuing claims contingent on a traded underlier. They obtain the famous option pricing formula by applying the boundary conditions for a European call option on a non-dividend-paying stock. Then the rigorous results for general options pricing theory are established in Harrison and Kreps (1979) and Harrison and Pliska (1981).

Today, the Black-Scholes and risk-neutral approaches are both widely used for pricing options and other derivative instruments. Although a closed-form pricing formula can often be obtained for European options by using these approaches, it is not the case for American options. Approximation methods are developed to price options when closed-form formulas are not available for some European options and all American options. Numerical methods have good computation performance when the state variables involved are in low-dimensional space. However, in derivative pricing we are often confronted with problems involving several state variables, such as an option written on several underlying assets or a pricing problem in which we allow some of the model parameters to become stochastic. In this case, pricing options with grid-based numerical methods becomes inefficient because of the curse of dimensionality– exponential growth in computation with the number of dimensions.

An alternative to grid-based methods is Monte Carlo simulation of the corresponding stochastic differential equation, first proposed for finance applications by Boyle (1977). Other research on analyzing option market via Monte Carlo simulation include Hull and White (1987), Johnson and Shanno (1987) and Scott (1987). Boyle et al. (1997) gives an overview of pricing using Monte Carlo simulation. Even if the European pricing problem can be solved in a high dimensional setting using this technique, pricing American options via Monte Carlo simulation still remains a very challenging problem, particularly in the high-dimensional case.

In general, most of the algorithms developed so far for American option pricing can be divided into two classes. The first class explores the structure of the optimal early exercise boundary by parameterizing the boundary and optimizing with respect to the parameters (e.g., Fu and Hu 1995, Wu and Fu 2000, Fu et al. 2000, Fu et al. 2001) or defining an estimator under an approximation of the boundary (e.g., Grant et al. 1996, 1997, Ben-Ameur et al. 2002). The other class estimates the price directly by a backward induction algorithm without assuming any knowledge on the structure of the exercise boundary (e.g., Longstaff and Schwartz, 2001, Broadie and Glasserman 1997, 2004).

In Chapter 3, we follow the route of the second approach and introduce a weighted stochastic mesh algorithm (WSM) for pricing high-dimensional American options that allow its holder to exercise at a fixed set of time points up to expiration. The algorithm extends the stochastic mesh (SM) algorithm introduced by Broadie and Glasserman (2004). SM has a simple structure, does not assume any knowledge of the exercise boundary, and requires computation effort that is only polynomial in the problem dimension as well as the number of exercise opportunity. All these features make SM an efficient algorithm compared to others. Retaining all of these desirable features, WSM requires milder assumptions that are also easier to verify than what is required by SM. The major advantage of WSM is that it does not require a closed-form expression for the transition density, and in fact the density function could be degenerate, in which case SM is not applicable. This generalization enables WSM to price, for example, American-style Asian options, which SM cannot handle.

Another important measure in the financial industry is Value at Risk (VaR). The early exercise feature of American options complicates the calculation of its VaR compared to its European counterpart. We also provide an estimator of VaR for American options. The convergence result is provided as well.

The last part of the dissertation deals with problems falling under a branch of statistics called ranking and selection and/or multiple comparison procedures. Suppose we will locate the best design among a finite number of choices, where the performance of each design can be only observed with uncertainty. The ranking and selection algorithms specify a level of correct selection first and then calculate the number of simulation replications required for each design to guarantee that level, whereas multiple comparison procedures provide confidence intervals on estimated performance differences between designs. Slightly different from these problems, our focus is on optimal allocation of given simulation budget. In particular, we maximize the probability of correct selection subject to the simulation budget constraints. We refer the readers to Chen et al. (1997, 2000), Chen and Kelton (2000), and Chick and Inoue (2001ab). In Chen (cf. Chen et al. 1997, 2000), this problem is called optimal computing budget allocation (OCBA) and discussed in the framework where all samples for different designs are independent and follow Gaussian distributions. In Fu et al. (2004), the discussion is generalized to correlated Gaussian distributions.

In Chapter 4 we further generalize the work to correlated non-Gaussian distributions. We derive optimal allocations for the setting of maximizing the probability of correct selection subject to a budget constraint on the total number of samples, when there is correlated sampling of the estimated design performances and the samples do not necessarily follow Gaussian distribution. We replace the original problem with an approximate problem and propose a solution procedure to the latter. Then we show the approximate solution converges to the true solution and establish the convergence rate as well.

In sum, the main contribution of the dissertation is two-fold:

- On the theoretical side, we discuss convergence in two different fields:
  - In the field of stochastic approximation, we study the convergence and asymptotic normality of a generalized form of stochastic approximation algorithm with deterministic perturbation sequences. Both one-simulation and two-simulation methods are considered. Assuming a special structure on the deterministic sequence, we establish sufficient conditions on the noise sequence for a.s. convergence of the algorithm. Construction

of such deterministic sequences follows the discussion of asymptotic normality.

  - In the field of ranking and selection and/or multiple comparison procedures, we propose an approximate solution to the so-called OCBA problem within a general framework where all simulation samples can be correlated non-Gaussian. The convergence rate of the approximate solution to the true solution is discussed and an exact order for the case of Gaussian distribution is obtained.

- On the practical side, we also present new methods in two different fields:

  - we present a weighted stochastic mesh method that only requires some easy-to-verify assumptions and a method to simulate the behavior of underlying securities. Our algorithm provides point estimates and confidence intervals for both options price and value-at-risk. The estimators converge to the true values as the computational effort increases.

  - In the field of ranking and selection and/or multiple comparison procedures, we propose a better way to allocate simulation budget when the simulation samples are drawn from correlated non-Gaussian distributions.

The layout of the dissertation is as follows. Chapter 2 (published in Xiong, Wang, and Fu 2002) discusses the asymptotic property of stochastic approximation algorithm with deterministic perturbation sequences. In Chapter 3, we present the weighted stochastic mesh algorithm for American option pricing. In Chapter 4, we propose a solution to the OCBA problem in correlated non-Gaussian setting and discuss the convergence issue as well. Some directions of future research are presented in Chapter 5.

Chapter 2

Stochastic Approximation with Deterministic Perturbation Sequences

2.1. Problem Setting

Throughout this chapter, we will consider the problem of locating minimum of a function $L : \mathbb{R}^p \to \mathbb{R}$. We assumes that L satisfies the following conditions.

(A1) The gradient of L, denoted by $g = \nabla L$, exists and is uniformly continuous.

(A2) There exist $\theta^* \in \mathbb{R}^p$ such that

- $f(\theta^*) = 0$; and
- for all $\delta > 0$, there exists $h_\delta > 0$ such that $\|\theta - \theta^*\| \geq \delta$ implies $f(\theta)^T(\theta - \theta^*) \geq h_\delta \|\theta - \theta^*\|^2$.

Before we advance to the asymptotic analysis, we present a generalized form of the stochastic approximation algorithm, of which SPSA and RDKW are just special cases. Let $\{d_n\}$ and $\{r_n\}$ are sequences on $\mathbb{R}^p$ and we denote the ith component of d_n and r_n as d_{ni} and r_{ni}, respectively. The recursive formulae of one-simulation and two-simulation forms are:

(1D)

$$\theta_{n+1} = \theta_n - a_n \frac{y_n^+}{c_n} r_n, \tag{2.1}$$

(2D)

$$\theta_{n+1} = \theta_n - a_n \frac{y_n^+ - y_n^+}{2c_n} r_n, \tag{2.2}$$

where y_n^+ and y_n^- are noisy samples obtained from simulations of the function L at perturbed points, defined by

$$\begin{aligned} y_n^+ &= L(\theta_n + c_n d_n) + e_n^+, \\ y_n^- &= L(\theta_n - c_n d_n) + e_n^-, \end{aligned}$$

with additive noise e_n^+ and e_n^-, respectively.

Obviously if $\{d_n\}$ and $\{r_n\}$ coincide, the two-simulation algorithm defined by (2.2) would reduce to the RDKW algorithm. SPSA is defined when $\{d_n\}$ and $\{r_n\}$ are related by

$$d_n = [\frac{1}{r_{n1}}, \cdots, \frac{1}{r_{np}}]^T.$$

Our goal is to find out an appropriate structure of $\{d_n\}$ and $\{r_n\}$ with which some desired asymptotic property can be obtained. The rest of the chapter is organized as follows. In section 2.2, with the deterministic sequence assuming a specified structure, we give sufficient conditions for a.s. convergence of both 1D and 2D. Also in section 2.2, asymptotic normality of both algorithms are discussed where the structure of deterministic is a little more specified. In section 2.3, we discuss how to construct such a specified deterministic perturbation sequence and the principle of defining parameters for practical simulation. Finally, section 2.4 offers some concluding remarks.

2.2. Almost Sure Convergence and Asymptotic Normality

Because our proofs of almost sure convergence rely mainly on a convergence theorem from Wang et al. (1996), Wang et al. (1997), and a lemma, we will introduce them first.

Theorem 2.1: Consider the stochastic approximation algorithm

$$\theta_{n+1} = \theta_n - a_n g(\theta_n) + a_n e_n + a_n b_n, \tag{2.3}$$

where $\{\theta_n\}$, $\{e_n\}$, and $\{b_n\}$ are sequences on $\mathbb{R}^p$, $g\colon \mathbb{R}^p \to \mathbb{R}^p$ satisfies Assumption (A2), $\{a_n\}$ is a sequence of positive real numbers satisfying $\lim_{n\to\infty} a_n = 0$, $\sum_{n=1}^{\infty} a_n = \infty$, and $\lim_{n\to\infty} b_n = 0$. Suppose that the sequence $\{g(\theta_n)\}$ is bounded. Then, for any θ_1 in $\mathbb{R}^p$, $\{\theta_n\}$ converges to θ^* if and only if $\{e_n\}$ satisfies any of the following conditions:

(B1)

$$\lim_{n\to\infty} \left(\sup_{n \le k \le m(n,T)} \left\| \sum_{i=n}^{k} a_i e_i \right\| \right) = 0$$

for some $T > 0$, where $m(n,T) \triangleq \max\{k : a_n + \cdots + a_k \le T\}$.

(B2)

$$\lim_{T\to 0} \frac{1}{T} \limsup_{n\to\infty} \left(\sup_{n\le k\le m(n,T)} \left\| \sum_{i=n}^{k} a_i e_i \right\| \right) = 0.$$

(B3) For any $\alpha, \beta > 0$, and any infinite sequence of non-overlapping intervals $\{I_k\}$ on $\mathbb{N}$ there exists $K \in \mathbb{N}$ such that for all $k \ge K$,

$$\left\| \sum_{n\in I_k} a_n e_n \right\| < \alpha \sum_{n\in I_k} a_n + \beta.$$

(B4) There exist sequences $\{f_n\}$ and $\{g_n\}$ with $e_n = f_n + g_n$ for all n such that

$$\sum_{k=1}^{n} a_k f_k \text{ converges, and } \lim_{n\to\infty} g_n = 0.$$

(B5) The weighted average $\{\bar{e}_n\}$ of the sequence $\{e_n\}$ defined by

$$\bar{e}_n = \frac{1}{\beta_n} \sum_{k=1}^{n} \gamma_k e_k,$$

converges to 0, where

$$\begin{aligned} \beta_n &= \begin{cases} 1 & n = 1, \\ \prod_{k=2}^{n} \frac{1}{1-a_k} & \text{otherwise}, \end{cases} \\ \gamma_n &= a_n \beta_n. \end{aligned}$$

Proof. See (Wang et al. 1996) for a proof for conditions (B1–4) and (Wang et al. 1997) for a proof for condition (B5). □

Lemma 2.2: Let $\{a_n\}$, $\{b_n\}$ and $\{e_n\}$ be sequences in $\mathbb{R}$ and $\{r_n\}$ in $\mathbb{R}^p$ such that:

(C1) $\lim_{n\to\infty} a_n = 0$, $\lim_{n\to\infty} \frac{a_n}{c_n} = 0, \sum_{n=1}^{\infty} a_n = \infty$;

(C2) $S_0 = \sup_{n,m} \left\| \sum_{i=n}^{m} r_i \right\| < \infty$, $E_0 = \sup_n ||e_n|| < \infty$;

(C3) $\sum_{n=1}^{\infty} |\frac{a_n}{c_n} - \frac{a_{n+1}}{c_{n+1}}| < \infty$ or $\lim_{n\to\infty} \frac{1}{c_n} - \frac{a_{n+1}}{a_n c_{n+1}} = 0$;

(C4) $\{\frac{||e_n - e_{n+1}||}{c_n}\}$ satisfies condition (B1-5).

Then $\{\frac{r_n e_n}{c_n}\}$ satisfies condition (B1).

Remarks: Lemma 2.2 still holds if $\{r_n\}$ and $\{e_n\}$ are in $\mathbb{R}^{p\times p}$ and $\mathbb{R}^p$, respectively. It is trivial to show that the first alternative of (C3) can be achieved by assuming $\frac{a_n}{c_n} \downarrow 0$.

Proof. Let $S_i \equiv \sum_{j=n}^{i} r_j$, $\forall i < n-1$ and $S_{n-1} = 0$. Then for all $n \le k \le m(n,T)$,

$$\begin{aligned} \left\|\sum_{i=n}^{k} \frac{a_i}{c_i} r_i e_i\right\| &= \left\|\sum_{i=n}^{k} \frac{a_i}{c_i}(S_i - S_{i-1})e_i\right\| = \left\|\frac{a_k}{c_k} S_k e_k + \sum_{i=n}^{k-1} S_i(\frac{a_i}{c_i} e_i - \frac{a_{i+1}}{c_{i+1}} e_{i+1})\right\| \\ &\le \left\|\frac{a_k}{c_k} S_k e_k\right\| + \sum_{i=n}^{k-1} \left\|S_i \frac{a_i}{c_i}(e_i - e_{i+1})\right\| + \sum_{i=n}^{k-1} \left\|(\frac{a_i}{c_i} - \frac{a_{i+1}}{c_{i+1}}) S_i e_{i+1}\right\| \\ &\le S_0 E_0 |\frac{a_k}{c_k}| + S_0 \sum_{i=n}^{k-1} \frac{a_i ||e_i - e_{i+1}||}{c_i} + S_0 E_0 \sum_{i=n}^{k-1} |\frac{a_i}{c_i} - \frac{a_{i+1}}{c_{i+1}}| \end{aligned} \tag{2.4}$$

1. The first term converges to 0 by assumption (C1).

2. Since $\{\frac{||e_n - e_{n+1}||}{c_n}\}$ satisfies condition (B4), we have $\{f_n\}$ and $\{g_n\}$ such that $\frac{||e_n - e_{n+1}||}{c_n} = f_n + g_n$, $\sum_{i=n}^{k-1} a_n f_n < \infty$ and $\lim_{n\to\infty} g_n = 0$, then we have

$$\begin{aligned} \sum_{i=n}^{k-1} \frac{a_i ||e_i - e_{i+1}||}{c_i} &= \sum_{i=n}^{k-1} a_i f_i + \sum_{i=n}^{k-1} a_i g_i \le \sum_{i=n}^{k-1} a_i f_i + \sup_{i\ge n} ||g_i|| \sum_{i=n}^{k-1} a_i \\ &\le \sum_{i=n}^{k-1} a_i f_i + T \sup_{i \ge n} ||g_i|| \to 0 \end{aligned}$$

.

3. $\sum_{n=1}^{\infty}|\frac{a_n}{c_n}-\frac{a_{n+1}}{c_{n+1}}|$ implies $\sum_{i=n}^{k-1}|\frac{a_i}{c_i}-\frac{a_{i+1}}{c_{i+1}}|\to 0$; while $\lim_{n\to\infty}\frac{1}{c_n}-\frac{a_{n+1}}{a_n c_{n+1}}=0$ yields

$$\sum_{i=n}^{k-1}|\frac{a_i}{c_i}-\frac{a_{i+1}}{c_{i+1}}|\leq \sup_{i\geq n}|\frac{1}{c_i}-\frac{a_{i+1}}{a_i c_{i+1}}|\sum_{i=n}^{k-1}a_i\leq T\sup_{i\geq n}|\frac{1}{c_i}-\frac{a_{i+1}}{a_i c_{i+1}}|\to 0$$

We are done since each term on RHS of (2.4) converges to zero when $n\to\infty$.

$\square$

Now we are in a position to present our main results. Propositions 2.3 and 2.4 discuss a.s. convergence of $\{\theta_n\}$ defined by (2.1) and (2.2), respectively, and Propositions 2.5 and 2.6 give asymptotic normality of $\{\theta_n\}$ for both cases. Note we always assume $\lim_{n\to\infty}a_n=0$, $\lim_{n\to\infty}c_n=0$, $\lim_{n\to\infty}\frac{a_n}{c_n}=0$ and $\sum_n a_n=\infty$.

Proposition 2.3 (convergence of one-simulation algorithm): Suppose that the Assumptions (A1–2) hold, and

(D1) $\sum_{n=1}^{\infty}|a_n-a_{n+1}|<\infty$ or $\lim_{n\to\infty}\frac{a_n}{a_{n+1}}=1$

(D2) $\sum_{n=1}^{\infty}|\frac{a_n}{c_n}-\frac{a_{n+1}}{c_{n+1}}|<\infty$ or $\lim_{n\to\infty}\frac{1}{c_n}-\frac{a_{n+1}}{a_n c_{n+1}}=0$

(D3) $\{L(\theta_n)\}$ and $\{g(\theta_n)\}$ are bounded

(D4) both $\{d_n\}$ and $\{r_n\}$ are periodical with period M, $\sum_{n=1}^{M}r_n=0$ and $\frac{1}{M}\sum_{n=1}^{M}r_n d_n^T=\rho I$, where $\rho>0$

(D5) $\{\frac{a_n}{c_n^2}\}$ satisfies condition (B1–5), both $\{\frac{e_n^+ r_n}{c_n}\}$ and $\{\frac{a_n|e_n|}{c_n^2}\}$ satisfy condition (B1–5) a.s.

Then, $\{\theta_n\}$ defined by (2.1) converges to θ^* a.s.

Remarks: The boundedness condition on L and g are not very strong. Practically we often restrict $\{\theta_n\}$ to a compact set by doing projection. The uniform continuity and boundedness are implied by continuity. The assumption $\frac{1}{M}\sum_{n=1}^{M} r_n d_n^T = \rho I$ implies that $p = Rank(\sum_{n=1}^{M} r_n d_n^T) \leq \sum_{n=1}^{M} Rank(r_n d_n^T) = M$. Actually we can see from proof that $\{r_n\}$ and $\{d_n\}$ are not necessarily periodical, all we need is that

- The partial sum of $\{r_n\}$ is bounded;
- There exists a positive constant ρ such that the partial sum of $\{r_n d_n^T - \rho I\}$ is bounded.

Proof. By the mean value theorem, we can rewrite (2.1)as

$$\begin{aligned}\theta_{n+1} &= \theta_n - \rho a_n g(\theta_n) - a_n r_n d_n^T[g(\theta_n + \lambda_n c_n d_n) - g(\theta_n)] \\ &\quad -a_n[r_n d_n^T - \rho I]g(\theta_n) - \frac{a_n}{c_n} L(\theta_n) r_n - a_n \frac{e_n^+}{c_n} r_n, \end{aligned} \tag{2.5}$$

where $0 \leq \lambda \leq 1$.

1. Since $\lim_{n\to\infty} g(\theta_n + \lambda_n c_n d_n) - g(\theta_n) = 0$ by the uniform continuity of g and $\lim_{n\to\infty} c_n = 0$, $\{r_n d_n^T[g(\theta_n + \lambda_n c_n d_n) - g(\theta_n)]\}$ satisfies condition (B4). Also, we know $\{g(\theta_n + \lambda_n c_n d_n)\}$ is bounded.

2. Combining boundedness of both $\{g(\theta_n + \lambda_n c_n d_n)\}$ and $\{L(\theta_n)\}$ with assumption (D5), we can check (2.5) and show

$$\lim_{n\to\infty} \theta_n - \theta_{n+1} = 0 \text{ a.s.}$$

   Hence $\lim_{n\to\infty} g(\theta_n) - g(\theta_{n+1}) = 0$ by uniform continuity of g. $\{r_n d_n^T - \rho I]g(\theta_n)\}$ satisfies condition (B1) by letting $\{c_n\}$, $\{r_n\}$ and $\{e_n\}$ in Lemma 2.2 be $\{1\}$, $\{r_n d_n^T - \rho I\}$ and $\{g(\theta_n)\}$, respectively.

3. Applying mean value theorem to L, we have

$$|L(\theta_n) - L(\theta_{n+1})| = |g^T[\theta_n + \mu_n(\theta_n - \theta_{n+1})](\theta_n - \theta_{n+1})|,$$

   where $0 \le \mu_n \le 1$. $\lim_{n\to\infty} g[\theta_n + \mu_n(\theta_n - \theta_{n+1})] - g(\theta_n) = 0$ implies boundedness of $||g[\theta_n + \mu_n(\theta_n - \theta_{n+1})]$. Hence,

$$\frac{|L(\theta_n) - L(\theta_{n+1}|}{c_n} \le M_0 \frac{||\theta_n - \theta_{n+1}||}{c_n} \le M_0(M_1 \frac{a_n}{c_n^2} + M_2 \frac{a_n}{c_n^2}|e_n^+|),$$

   where the second inequality is obtained by applying (2.5) to $\theta_n - \theta_{n+1}$ and using some boundedness conditions; M's are positive constants. Since, by assumption (D5), the RHS of above formula satisfies condition (B1), it is trivial to prove the LHS also satisfies condition (B1). Hence we can let $\{e_n\}$ in Lemma 2.2 be $\{L(\theta_n\}$ and conclude $\{\frac{L(\theta_n)r_n}{c_n}\}$ satisfies condition (B1).

4. $\{\frac{e_n^+}{c_n} r_n\}$ satisfies condition (B1) by assumption (D5).

The proof completes by combining above arguments with Theorem 2.1. □

Proposition 2.4 (convergence of two-simulation algorithm): Suppose that the assumptions (A1–2, D1) hold, and

- $\{g(\theta_n)\}$ is bounded, $\lim_{n\to\infty} c_n = 0$;
- $\{d_n\}$ is periodical with period M, and $\frac{1}{M}\sum_{n=1}^{M} r_n d_n^T = \rho I$, where $\rho > 0$.

Then, $\{\theta_n\}$ defined by (2.2) converges to θ^* a.s. if and only if $\frac{e_n^+ r_n}{c_n}$ satisfies (B1–5) a.s.

Proof. The sufficiency proof completes by following the same arguments in the proof of Proposition 2.3, and the necessity proof is trivial. □

We denote the Hessian matrix and sth derivative of $L(\theta)$ as $H(\theta)$ and $L^{(3)}(\theta)$ respectively.

Proposition 2.5 (asymptotic normality of one-simulation algorithm): Suppose that the Assumptions (A1–2) hold and $\{\theta_n\}$ is defined by (2.1), and

(E1) $a_n = a/n^\alpha$ and $c_n = c/n^\gamma$ where a, c, α, $\gamma > 0$;

(E2) $\alpha \le 1$, $\beta = \alpha - 2\gamma > 0$, $3\gamma - \alpha/2 \ge 0$, $1 + 2\gamma < 2\alpha$;

(E3) both $\{d_n\}$ and $\{r_n\}$ are periodical with period M, $\sum_{n=1}^{M} r_n = 0$, $\sum_{n=1}^{M} r_n \otimes d_n \otimes d_n = 0$, $\frac{1}{M}\sum_{n=1}^{M} r_n d_n^T = \rho I$, where $\rho > 0$;

(E4) $Q \equiv M^{-1}\sum_{n=1}^{M} r_n r_n^T$ and orthogonal matrix P satisfies $P^T H(\theta^*) P = (a\rho)^{-1}$ $\mathrm{diag}(\lambda_1, \cdots, \lambda_p)$;

(E5) L, g, H and $L^{(3)}$ are all continuous and bounded;

(E6) $\lim_{n\to\infty} n^{-\beta} e_n^+ = 0$, $E(e_n^+|\mathcal{F}_n) = 0$ a.s. and $E((e_n^+)^2|\mathcal{F}_n) \to \sigma^2$ a.s., $\forall n$, where $\mathcal{F}_n \equiv \sigma(\theta_0, \theta_1, \cdots, \theta_n)$;

(E7) There exists $\delta > 0$ such that $\sup_n E|e_n^+|^{2+2\delta} < \infty$.

Then

$$n^{\beta/2}(\theta_n - \theta^*) \stackrel{dist}{\rightarrow} N(\mu, PXP^T), \qquad n \to \infty$$

where $X_{ij} = a^2 c^{-2} \sigma^2 [P^T Q P]_{ij} (\lambda_i + \lambda_j - \beta_+)^{-1}$ with $\beta_+ = \beta < 2\min_i \lambda_i$ if $\alpha = 1$ and $\beta_+ = 0$ if $\alpha < 1$,and

$$\mu = \begin{cases} 0 & \text{if } 3\gamma - \alpha/2 > 0, \\ (a\rho H(\theta^*) - \frac{1}{2}\beta_+ I)^{-1} T & \text{if } 3\gamma - \alpha/2 = 0, \end{cases}$$

where the lth element of T is

$$-\frac{ac^2}{6M}[L^{(3)}_{lll}(\theta^*)\sum_{n=1}^{M} d^3_{nl} r_{nl} + 3\sum_{i=1,i\neq l}^{M} L^{(3)}_{lii}(\theta^*)\sum_{n=1}^{M} d^2_{ni} d_{nl} r_{nl}$$
$$+6\sum_{i,j=1;i\neq j\neq l}^{M} L^{(3)}_{lij}(\theta^*)\sum_{n=1}^{M} d_{nl} d_{ni} d_{nj} r_{nl}].$$

Proof. It is easy to show both $\sum_{i=1}^{n} \frac{a_n}{c_n} e^+_n r_n$ and $\sum_{i=1}^{n} \frac{a_n^2}{c_n^2}(|e^+_n| - E_{\mathcal{F}_n}|e^+_n|)$are martingales with finite L^2 norm. Hence $\sum_{n=1}^{\infty} \frac{a_n}{c_n} e^+_n r_n < \infty$ and $\sum_{i=1}^{n} \frac{a_n^2}{c_n^2}|e^+_n| < \infty$ a.s. by L^2 convergence theorem for martingale. Then Proposition 2.3 guarantees the a.s. convergence of θ_n to θ^*. To show the asymptotic normality, we will check if conditions (2.2.1–3) of Fabian (1968) hold. We will use notation of Fabian (1968) as well. Let $0 \le \lambda_n, \eta_n \le 1$. Use the mean value theorem and rewrite (2.1):

$$\begin{aligned}\theta_{n+1} &= \theta_n - a_n r_n d_n^T g(\theta_n) - \frac{a_n}{c_n} L(\theta_n) r_n - a_n \frac{e^+_n}{c_n} r_n - \frac{1}{2} a_n c_n r_n d_n^T H(\theta_n) d_n \\ &\quad - \frac{1}{6} a_n c_n^2 r_n L^{(3)}(\theta_n + \lambda_n c_n d_n) d_n \otimes d_n \otimes d_n \end{aligned} \tag{2.6}$$

Use this formula M times, we have

$$\begin{aligned}\theta_{nM+M} - \theta^* &= (I - n^{-\alpha}\Gamma_n)(\theta_{nM} - \theta^*) + n^{-(\alpha+\beta)/2}\Phi_n V_n + \\ &\quad n^{-\alpha-\beta/2}(T_n^{(1)} + T_n^{(2)} + T_n^{(3)} + T_n^{(4)}),\end{aligned}$$

where

$$\begin{aligned}\Gamma_n &= aM^{-\alpha}\sum_{i=nM}^{nM+M-1}(\frac{i}{nM})^{-\alpha} r_i d_i^T H(\theta_{nM} + \eta_n(\theta_{nM} - \theta^*)) \overset{a.s.}{\to} aM^{1-\alpha}\rho H(\theta^*), \\ \Phi_n &= I \ , \ V_n = \frac{a}{c} M^{-\alpha+\gamma}\sum_{i=nM}^{nM+M-1}(\frac{i}{nM})^{-\alpha+\gamma} e^+_i r_i, \\ T_n^{(1)} &= -an^{\alpha/2-\gamma} M^{-\alpha}\sum_{i=nM}^{nM+M-1}(\frac{i}{nM})^{-\alpha} r_i d_i^T [g(\theta_i) - g(\theta_{nM})], \\ T_n^{(2)} &= -\frac{a}{c} n^{\alpha/2} M^{-\alpha+\gamma}\sum_{i=nM}^{nM+M-1}(\frac{i}{nM})^{-\alpha+\gamma} L(\theta_i) r_i,\end{aligned}$$

$$
\begin{aligned}
T_n^{(3)} &= -\frac{1}{2}acn^{\alpha/2-2\gamma}M^{-\alpha-\gamma}\sum_{i=nM}^{nM+M-1}(\frac{i}{nM})^{-\alpha-\gamma}r_i d_i^T H(\theta_i)d_i, \\
T_n^{(4)} &= -\frac{1}{6}ac^2 n^{\alpha/2-3\gamma}M^{-\alpha-2\gamma}\sum_{i=nM}^{nM+M-1}(\frac{i}{nM})^{-\alpha-2\gamma}r_i L^{(3)}(\theta_i+\lambda_i c_i d_i)d_i\otimes d_i\otimes d_i.
\end{aligned}
$$

To prove $T_n^{(2)} \xrightarrow{L^2} 0$, we have

$$
\begin{aligned}
T_n^{(2)} &= K_0 n^{\alpha/2}\sum_{i=nM}^{nM+M-1}((\frac{i}{nM})^{-\alpha+\gamma}-1)L(\theta_i)r_i + K_0 n^{\alpha/2}\sum_{i=nM}^{nM+M-1}L(\theta_i)r_i \\
&= O(n^{-\alpha/2+\gamma}) + K_0 n^{\alpha/2}\sum_{i=nM}^{nM+M-1}(L(\theta_i)-L(\theta_{nM}))r_i \\
&= o(1) + K_0 n^{\alpha/2}\sum_{i=nM}^{nM+M-1}(\theta_i-\theta_{nM})^T g(\theta'_{nM}) \\
&= o(1) + n^{\alpha/2}O(n^{-\alpha+\gamma}) \\
&= o(1).
\end{aligned}
$$

The second equality is by $(1+\frac{A}{n})^{-\alpha+\gamma}-1 = O(1/n)$ and $\sum\limits_{i=nM}^{nM+M-1} r_i = 0$; the third is by taking a Taylor series expansion and using the fact that θ'_{nM} is on the line segment between θ_i and θ_{nM}; the fourth is by applying (2.6) to $\theta_i - \theta_{nM}$. Of course boundedness of functions are required when necessary. Also, $o(\cdot)$ and $O(\cdot)$ are in terms of L^2 norm and K_0 is a constant.

We have shown that $T_n^{(2)} \xrightarrow{L^2} 0$. Actually similar argument can be used to show that $T_n^{(1)} \xrightarrow{L^2} 0$ and $T_n^{(3)} \xrightarrow{L^2} 0$. If $3\gamma-\alpha/2>0$, we can also show $T_n^{(4)} \xrightarrow{L^2} 0$. If $3\gamma-\alpha/2=0$, it is easy to show that $T_n^{(4)} \xrightarrow{a.s.} M^{1-\alpha-\beta/2}T$.

Obviously $E_{\mathcal{F}_n}V_n = 0$ and $E_{\mathcal{F}_n}V_nV_n^T \xrightarrow{L^2} \frac{a^2\sigma^2}{c^2}M^{1-2\alpha+2\gamma}Q$. To show

$$
\lim_{k\to\infty} E(\chi_{\|V_n\|^2\geq rn^\alpha}\|V_n\|^2) = 0, \forall r>0,
$$

we have

$$
\begin{aligned}
E(\chi_{\|V_n\|^2 \geq rn^\alpha} \|V_n\|^2) &\leq P(\|V_n\|^2 \geq rn^\alpha)^{\delta'/(1+\delta')} (E\|V_n\|^{2+2\delta'})^{1/(1+\delta')} \\
&\leq K_1 (\frac{E\|V_n\|^2}{rn^\alpha})^{\delta'/(1+\delta')} \leq K_2 n^{-\alpha\delta'/(1+\delta')} \to 0,
\end{aligned}
$$

where K_1 and K_2 are constants and $0 < \delta' < \delta$.

Since all the conditions (2.1.1–3) in Fabian (1968) are verified, we have

$$n^{\beta/2}(\theta_{nM} - \theta^*) \stackrel{dist}{\to} N(M^{-\beta/2}\mu, M^{-\beta}PXP).$$

That is,

$$(nM)^{\beta/2}(\theta_{nM} - \theta^*) \stackrel{dist}{\to} N(\mu, PXP).$$

For all $0 < i < M$, we can similarly prove

$$(nM + i)^{\beta/2}(\theta_{nM+i} - \theta^*) \stackrel{dist}{\to} N(\mu, PXP).$$

$\square$

Proposition 2.6 (asymptotic normality of two-simulation algorithm): Suppose that the Assumptions (A1–2, E1–3) hold and $\{\theta_n\}$ is defined by (2.2), and

- both $\{d_n\}$ and $\{r_n\}$ are periodical with period M, let $\frac{1}{M}\sum_{n=1}^{M} r_n d_n^T = \rho I$, where $\rho > 0$, and let orthogonal matrix P such that $P^T H(\theta^*) P = (a\rho)^{-1}\,\mathrm{diag}(\lambda_1, \cdots, \lambda_p)$;
- g and H bounded, $L^{(3)}$ is continuous at θ^*;
- $E(e_n^+ - e_n^- | \mathcal{F}_n) = 0$ a.s. and $E((e_n^+ - e_n^-)^2 | \mathcal{F}_n) \to 4\sigma^2$ a.s., $\forall n$, where $\mathcal{F}_n \equiv \sigma(\theta_0, \theta_1, \cdots, \theta_n)$;

- There exists $\delta > 0$ such that $\sup_n E|e_n^{(\pm)}|^{2+2\delta} < \infty$.

Then we have the same conclusion as Proposition 2.5.

Remark: If we let each component of r_n and d_n assume ± 1, then we get exactly the same result as Proposition 2 in (Spall 1992).

Proof. Proof completes by following the same arguments in the proof of Proposition 2.5. $\square$

The four propositions above show that deterministic perturbation can do at least as well as randomized perturbation asymptotically. Following two propositions will show that the former might have higher convergence rate than the latter.

Let s be an even integer, $m = s/2$, $0 < u_1 < \cdots < u_m \leq 1$, $U = \left\| u_j^{2i-1} \right\|_{i,j=1}^m$, and v is the first column of $\frac{1}{2}U^{-1}$, and v_i is ith component of v, we define a $s-$ simulation form as in (Fabian 1967):

(SD)

$$\theta_{n+1} = \theta_n - \frac{a_n r_n}{c_n} \sum_{i=1}^{m} v_i (y_{n,i}^+ - y_{n,i}^-), \tag{2.7}$$

where y_n^+ and y_n^- are noisy samples of the function L at perturbed points, defined by

$$\begin{aligned} y_{n,i}^+ &= L(\theta_n + c_n u_i d_n) + e_{n,i}^+, \\ y_{n,i}^- &= L(\theta_n - c_n u_i d_n) + e_{n,i}^-, \end{aligned}$$

with additive noise $e_{n,i}^+$ and $e_{n,i}^-$, respectively.

Proposition 2.7 (convergence rate of s-simulation algorithm): Suppose $\{\theta_n\}$ is defined by (2.7), and

- $g(\theta^*) = 0$, for all $\theta \in \mathbb{R}^p$, we have $g(\theta)^T(\theta - \theta^*) \geq C \|\theta - \theta^*\|^2$;

- $a_n = a/n^\alpha$ and $c_n = c/n^\gamma$ where $a > 0$, $c > 0$, $0 < \alpha \leq 1$, $0 < \gamma < \alpha/2$;

- both $\{d_n\}$ and $\{r_n\}$ are periodical with period M, let $\frac{1}{M}\sum_{n=1}^{M} r_n d_n^T = \rho I$, where $\rho > 0$;

- $\beta = min\{2s\gamma, \alpha - 2\gamma\}$ and $\beta < 2aM\rho C$, if $\alpha = 1$;

- g, H and $L^{(s+1)}$ are bounded;

- $E_{\mathcal{F}_n}(e^+_{n,i} - e^-_{n,i}) = 0$ a.s., $E_{\mathcal{F}_n}(e^+_{n,i} - e^-_{n,i})(e^+_{n,j} - e^-_{n,j}) = 0$ and $E_{\mathcal{F}_n}(e^+_{n,i} - e^-_{n,i})^2 \leq \sigma^2$a.s. , $\forall n$, i, $j \neq i$, where $\mathcal{F}_n \equiv \sigma(\theta_0, \theta_1, \cdots, \theta_n)$.

Then

$$E\|\theta_n - \theta^*\|^2 = O(n^{-\beta}).$$

Remark: A similar conclusion for randomized perturbations has been obtained in Gerencser (1999).

Proof. Almost sure convergence can be shown by using similar argument in the proof of Proposition 2.3. Then it suffices to show $\limsup n^\beta E\|\theta_n - \theta^*\|^2 < \infty$. By a Taylor series expansion and the definition of $\{u_i\}$ and $\{v_i\}$, we can rewrite (2.7):

$$\theta_{n+1} - \theta^* = \theta_n - \theta^* - a_n r_n d_n^T g(\theta_n) - a_n c_n^s r_n \xi_n - \frac{a_n r_n e_n}{c_n}, \tag{2.8}$$

where ξ_n is some bounded r.v. depending on $L^{(s+1)}(\cdot)$ and $e_n = \sum_{i=1}^{m} v_i(e^+_{n,i} - e^-_{n,i})$.

$$\begin{aligned} b_{n+1} &\equiv E\|\theta_{nM+M} - \theta^*\|^2 \\ &= E\left\|\theta_{nM} - \theta^* - \sum_{i=nM}^{nM+M-1}[a i^{-\alpha} r_i d_i^T g(\theta_i) + K_1 i^{-\alpha-s\gamma} r_i \xi_i + K_2 i^{-\alpha+\gamma} r_i e_i]\right\|^2 \\ &\leq b_n + O(n^{-2\alpha}) + O(n^{-2\alpha-2s\gamma}) + K_3 n^{-2\alpha+2\gamma} - U_n - V_n - W_n, \end{aligned} \tag{2.9}$$

where we use $(a+b+c)^2 \leq 9(a^2+b^2+c^2)$ and

$$\begin{aligned}
U_n &= 2E(\theta_{nM}-\theta^*)^T \sum_{i=nM}^{nM+M-1} K_2 i^{-\alpha+\gamma} r_i e_i \\
&= 2E[(\theta_{nM}-\theta^*)^T E_{\mathcal{F}_{nM}} \sum_{i=nM}^{nM+M-1} K_2 i^{-\alpha+\gamma} r_i e_i] \\
&= 0, \qquad (2.10) \\
V_n &= 2E(\theta_{nM}-\theta^*)^T \sum_{i=nM}^{nM+M-1} K_1 i^{-\alpha-s\gamma} r_i \xi_i \\
&\geq -K_4 (nM)^{-\alpha-s\gamma} E\left\|\theta_{nM}-\theta^*\right\| \\
&\geq -\eta (nM)^{-\alpha} b_n - \frac{K_4^2}{4\eta}(nM)^{-\alpha-2s\gamma}, \qquad (2.11) \\
W_n &= 2aE(\theta_{nM}-\theta^*)^T \sum_{i=nM}^{nM+M-1} i^{-\alpha} r_i d_i^T g(\theta_i) \\
&\geq -K_5 n^{-\alpha-1} E\left\|\theta_{nM}-\theta^*\right\| + 2aM\rho(nM)^{-\alpha} E(\theta_{nM}-\theta^*)^T g(\theta_{nM}) \\
&\quad +2a(nM)^{-\alpha} E(\theta_{nM}-\theta^*)^T \sum_{i=nM}^{nM+M-1} r_i d_i^T [g(\theta_i)-g(\theta_{nM})] \\
&\geq -K_5 n^{-\alpha-1}(b_n+1) - K_6 n^{-2\alpha}(b_n+1) + 2aM\rho C(nM)^{-\alpha} b_n, \qquad (2.12)
\end{aligned}$$

where $\eta = aM\rho C$ if $\alpha < 1$ and $0 < \eta < 2aM\rho C - \beta$ if $\alpha = 1$. The first inequality of W_n is by $(1+\frac{A}{n})^{-\alpha} - 1 = O(1/n)$, the second is by applying mean value theorem to $g(\theta_i) - g(\theta_{nM})$ and then applying (2.8) to $\theta_i - \theta_{nM}$.

By (2.9), (2.10), (2.11) and (2.12), we have

$$b_{n+1} \quad \leq \quad b_n(1 - A_n(nM)^{-\alpha}) + K_7 n^{-\alpha-\beta},$$

where $\lim A_n = 2aC$. Hence, we can show $\limsup n^{\beta} E\left\|\theta_{nM}-\theta^*\right\|^2 < \infty$ by Lemma 4.2 in (Fabian 1967). Similarly, we can show for all $0 < i < M, \limsup n^{\beta} E\left\|\theta_{nM+i}-\theta^*\right\|^2 < \infty$, which completes the proof.

Note: The $K's$ in the proof are positive constants. □

Combining Lemma 4.2 and 4.3 in (Fabian 1967) yields the following Lemma.

Lemma 2.8: Let b_n, A, B, D_n, α, β be real numbers, and

$$b_{n+1} \leq b_n(1 - An^{-\alpha}) + Bn^{-\alpha-\beta} + D_n n^{-\beta}$$

, where $0 < \alpha \leq 1$, $\beta > 0$, $B > 0$, $\sum_{n=1}^{\infty} < \infty$, $A > \beta$ if $\alpha = 1$.

Then $\limsup n^{\beta} b_n < \infty$.

Using Lemma 2.8, we can prove a higher rate of a.s. convergence holds. A quick corollary for two-simulation setting is the method with deterministic perturbation can achieve a convergence rate arbitrarily close to $\frac{1}{3}$, which is an upper limit given by Proposition 2.7.

Proposition 2.9: Suppose that Assumptions (A1–2) hold and $\{\theta_n\}$ is defined by (2.7), and

- $a_n = a/n^{\alpha}$ and $c_n = c/n^{\gamma}$ where $a > 0$, $c > 0$, $0 < \alpha \leq 1$, $0 < \gamma < \alpha/2$, $\alpha - \gamma > \frac{1}{2}$;
- both $\{d_n\}$ and $\{r_n\}$ are periodical with period M, let $\sum_{n=1}^{M} r_n d_n^T = \rho I$, where $\rho > 0$;
- g, H and $L^{(s+1)}$ are bounded, H is positive definite and continuous at θ^*, let λ be the smallest characteristic value of $H(\theta^*)$;
- $\beta_0 = \min\{2s\gamma, 2\alpha - 2\gamma - 1\}$ and $\beta_0 < 2a\rho\lambda$ if $\alpha = 1$;
- $E(e^+_{n,i} - e^-_{n,i}|\mathcal{F}_n) = 0$ a.s., $E((e^+_{n,i} - e^-_{n,i})(e^+_{n,j} - e^-_{n,j})|\mathcal{F}_n) = 0$ and $E((e^+_{n,i} - e^-_{n,i})^2|\mathcal{F}_n) \leq \sigma^2$a.s. , $\forall n$, i, $j \neq i$, where $\mathcal{F}_n \equiv \sigma(\theta_0, \theta_1, \cdots, \theta_n)$;

- $\sup_{n,i} E(e^+_{n,i} - e^-_{n,i})^{2+2\delta} < \infty$ where $\delta > 0$.

Then $n^{\beta/2}(\theta_n - \theta^*) \to 0$ a.s. for every $\beta < \beta_0$.

Remark: The optimal γ for $s-$simulation is $\frac{2\alpha-1}{2+2s}$ and thus $\beta = \frac{s(2\alpha-1)}{s+1}$. When $\alpha = 1$, we achieve the upper limit of convergence rate given by Proposition 2.7.

Proof. Since $\theta_n \to \theta^*$ a.s. using (2.8), θ_n are a.s. bounded. By mean value theorem and continuity of H at θ, we have $\theta_n^T g(\theta_n) > (1-\eta)\lambda \|\theta_n\|^2$ for sufficiently large n. Then

$$\begin{aligned} \|\theta_{nM+M} - \theta^*\|^2 &= \left\| \theta_{nM} - \theta^* - \sum_{i=nM}^{nM+M-1} [a i^{-\alpha} r_i d_i^T g(\theta_i) + K_1 i^{-\alpha-s\gamma} r_i \xi_i + K_2 i^{-\alpha+\gamma} r_i e_i] \right\|^2 \\ &\le \|\theta_{nM} - \theta^*\|^2 + O(n^{-2\alpha}) + O(n^{-2\alpha-2s\gamma}) + K_3 n^{-2\alpha+2\gamma} \sum_{i=nM}^{nM+M-1} e_i^2 \\ &\quad -U_n - V_n - W_n, \end{aligned} \tag{2.13}$$

where

$$\begin{aligned} U_n &= 2a(\theta_{nM} - \theta^*)^T \sum_{i=nM}^{nM+M-1} i^{-\alpha} r_i d_i^T g(\theta_i) \\ &\ge -K_4 n^{-\alpha-1} \|\theta_{nM} - \theta^*\| + 2a\rho (nM)^{-\alpha} E(\theta_{nM} - \theta^*)^T g(\theta_{nM}) \\ &\quad +2a(nM)^{-\alpha}(\theta_{nM} - \theta^*)^T \sum_{i=nM}^{nM+M-1} r_i d_i^T [g(\theta_i) - g(\theta_{nM})] \\ &\ge O(n^{-2\alpha}) - K_5 n^{-2\alpha+\gamma} \sum_{i=nM}^{nM+M-1} e_i^2 + 2a\rho\lambda (nM)^{-\alpha} b_n, \end{aligned} \tag{2.14}$$

$$\begin{aligned} V_n &= 2E(\theta_{nM} - \theta^*)^T \sum_{i=nM}^{nM+M-1} K_1 i^{-\alpha-s\gamma} r_i \xi_i \\ &\ge -\eta (nM)^{-\alpha} \|\theta_{nM} - \theta^*\|^2 - K_6 (nM)^{-\alpha-2s\gamma}, \end{aligned} \tag{2.15}$$

$$\begin{aligned} W_n &= 2(\theta_{nM} - \theta^*)^T \sum_{i=nM}^{nM+M-1} K_2 i^{-\alpha+\gamma} r_i e_i \\ &\ge -K_7 n^{-\alpha+\gamma} \|\theta_{nM} - \theta^*\| \sum_{i=nM}^{nM+M-1} |e_i|. \end{aligned} \tag{2.16}$$

Hence, by (2.13), (2.14), (2.15) and (2.16), we have for almost all $\omega \in \Omega$

$$
\begin{aligned}
b_{n+1} &\equiv \left\|\theta_{nM+M} - \theta^*\right\|^2(\omega) \\
&\leq b_n(1 - (2a\rho\lambda - \eta)(nM)^{-\alpha}) + K_8 n^{-\alpha-\beta_0} + K_3 n^{-2\alpha+2\gamma} \sum_{i=nM}^{nM+M-1} (e_i^2 - \phi^2) \\
&\quad K_7 n^{-\alpha+\gamma} \left\|\theta_{nM} - \theta^*\right\| \sum_{i=nM}^{nM+M-1} |e_i|,
\end{aligned}
$$

where $\phi^2 = \sum_{i=1}^m u_i^2\sigma^2$. By the $L^{1+\delta}$ convergence theorem for martingale, we can show $\sum_{n=0}^{\infty} n^{-2\alpha+2\gamma+\beta} \sum_{i=nM}^{nM+M-1} (e_i^2 - \phi^2)$ converges, thus is bounded a.s. when $0 < \beta < \beta_0$. When $0 < \beta < \alpha - \gamma - \frac{1}{2}$, we have

$$
\sum_{n=0}^{\infty} E_{\mathcal{F}_n}(n^{-\alpha+\gamma+\beta} \left\|\theta_{nM} - \theta^*\right\| \sum_{i=nM}^{nM+M-1} |e_i|)^2 < \infty \tag{2.17}
$$

By the sharper form of the Borel-Cantelli lemma in Dubins (1965), we have, with probability one,

$$
\sum_{n=0}^{\infty} n^{-\alpha+\gamma+\beta} \left\|\theta_{nM} - \theta^*\right\| \sum_{i=nM}^{nM+M-1} |e_i| < \infty \tag{2.18}
$$

Then the condition of Lemma 2.8 holds for $\beta < \alpha - \gamma - \frac{1}{2}$. Hence, $\limsup n^{\beta/2}(\theta_n - \theta^*) \to 0$ a.s. when $\beta < \alpha - \gamma - \frac{1}{2}$. Apply this to (2.17) and (2.18) recursively, we can get $\frac{\beta}{\alpha-\gamma-1/2} < 1/2, 3/4, 7/8, \cdots$, which completes the proof. □

2.3. Construction of Deterministic Sequences

In this section, we present a general mechanism for construction of deterministic sequences $\{r_n\}$ and $\{d_n\}$ that satisfies conditions required for convergence of algorithms. Since stronger conditions required for convergence of one-simulation algorithms, we focus on constructions of sequences that satisfy the conditions stated in Proposition 2.3. The constructed sequences can be applied to two-simulation algorithms as well. We focus on sequences for RDKW and SPSA algorithms and consider the case where components of r_n and d_n take value from $\{\pm 1\}$. Note that in this case, the two classes of algorithms are identical. It is also clear that we only need to construct either $\{r_n\}$ or $\{d_n\}$, since they are identical as well.

Our constructions are based on the notion of orthogonal arrays (Hedayat et al. 1999). We claim that a desirable deterministic sequence in dimension p can be constructed from any binary (two-level) $N \times k$ orthogonal array with $k \geq p$. We first give the definition of orthogonal arrays:

Definition (Hedayat et al. 1999) An $N \times k$ array A with entries from $S = \{0, 1, \cdots, s\}$ is said to be an orthogonal array with s levels, strength t and index λ if every $N \times t$ subarray of A contains each t-tuple based on S exactly λ times as a row. We use the notation $OA(N, k, s, t)$ to denote such an array. For example, an $OA(8, 4, 2, 3)$ is given below

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}^T . \tag{2.19}$$

To construct a desired sequence $\{r_n\}$ in $\mathbb{R}^p$ from an $OA(N, k, 2, t)$ with $k \geq p$, we take the following simple steps:

1. Take any p columns from the orthogonal array to form a $N \times p$ array H.

2. Change all the zero entries in H into -1.

3. Use all the row vectors of H as one period for $\{r_n\}$.

For example, we can construct a sequence $\{r_n\}$ in $\mathbb{R}^4$ from (2.19) as

$$r1 = [-1,-1,-1,-1]^T, \quad r2 = [-1,-1,1,1]^T,$$

$$r3 = [-1,1,-1,1]^T, \quad r4 = [-1,1,1,-1]^T,$$

$$r5 = [1,-1,-1,1]^T, \quad r6 = [1,-1,1,-1]^T,$$

$$r7 = [1,1,-1,-1]^T, \quad r8 = [1,1,1,1]^T.$$

Orthogonal arrays have been applied in many areas including experiment designs, coding theory, and cryptography. A large body of literature exists on construction of orthogonal arrays. Hence the proposed construction provides a large set of deterministic sequences for use in stochastic approximation algorithms for optimization. A particular construction based on Hadamard matrices (Seberry and Yamada 1992) is presented in Bhatnagar et al. (2002).

2.4. Conclusion

In this chapter, we present a generalized form of the stochastic approximation algorithm of which SPSA and RDKW are special cases. We establish sufficient conditions on deterministic sequences for convergence of these algorithms. Asymptotic normality is established to show that deterministic sequences can at least achieve the same asymptotic performance with the random sequences. It remains to be shown theoretically that appropriately designed deterministic sequences can lead to faster convergence than the random sequences, which has been observed in experiments.

Chapter 3

A Weighted Stochastic Mesh Method for Pricing High-Dimensional American Options

3.1. Problem Setting

Let $S_t = (S_t^1, \ldots, S_t^n)$ denote a vector of m securities underlying the option, modeled as a Markov process on $R_+^n \equiv (0, \infty)^n$, with fixed initial state S_0 and discrete time parameter $t = 0, 1, \ldots, T$. The problem is to compute

$$Q = \max_{\tau} E[h(\tau, S_\tau)], \tag{3.1}$$

where τ is a stopping time taking values in the finite set $\{0, 1, \ldots, T\}$, and $h(t, x) \geq 0$ gives the payoff from exercise at time t in state x, discounted to time 0, with the possibly stochastic discount factor recorded in S_t. We can express the value starting at time t in state x recursively as

$$Q(t,x) = \begin{cases} \max(h(t,x), E[Q(t+1,S_{t+1})|S_t = x]) & t < T; \\ h(T,x) & t = T; \end{cases} \tag{3.2}$$

where the conditional expectation is with respect to the risk-neutral measure.

The rest of the chapter is organized as follows. Section 3.2 gives a description and theoretical analysis of the basic WSM algorithm for pricing American-style option. Section 3.3 applies the algorithm to the pricing of American-style Asian options. The estimators of value-at-risk are presented in Section 3.4. Section 3.5 contains concluding remarks.

3.2. The Weighted Stochastic Mesh Method (WSM)

From (3.2), we can see the major difficulty lies in the calculation of conditional expectation $E[Q(t+1,S_{t+1})|S_t = x]$ when we go backward from $t+1$ to t. The stochastic mesh method (SM) proposed by Broadie and Glasserman (2004) estimates this conditional expectation via a weighted sum of Q values at $t+1$. The idea can be illustrated via figure 3.1. For simplification, we let S_t be a two-dimensional vector and assume the option can be only exercised at t_0, t_1 and t_2, where t_2 is the maturity. We also assume only three nodes are generated for each time point after the starting time. In other words, the path of S_t starts from node α_0 at t_0 and may go through nodes $\{\alpha_1, \beta_1, \gamma_1\}$ at t_1 and $\{\alpha_2, \beta_2, \gamma_2\}$ at maturity. It is easy to get exact Q values for each node at maturity. Then the conditional expectation given $S_{t_1} = \alpha_1$ will be estimated by $W(\alpha_1\alpha_2)Q(\alpha_2) + W(\alpha_1\beta_2)Q(\beta_2) + W(\alpha_1\gamma_2)Q(\gamma_2)$,

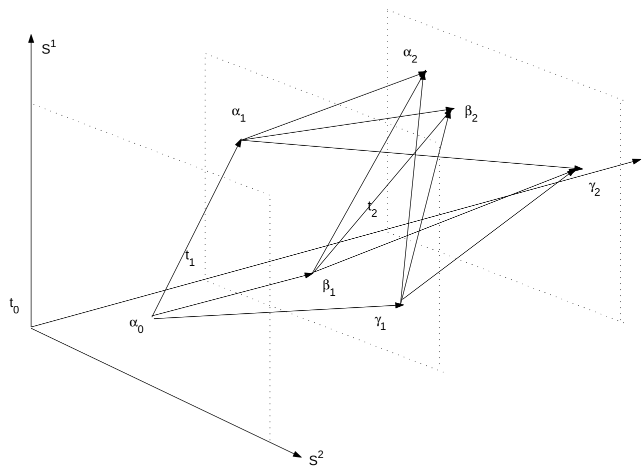

Figure 3.1: Stochastic Mesh Method

where $W(\alpha\beta)$ represents the weight assigned to nodes pair (α, β). The SM method uses an idea similar to importance sampling to define the weight $W(\alpha\beta)$, whereas the WSM method uses a different weight definition.

We use figure 3.2 to illustrate the idea behind the weight definition. First we define distance, $d(\alpha, \beta)$, between any two points of the state space at maturity. Then we split the entire space into three disjoint subsets $A(\alpha_2)$, $A(\beta_2)$ and $A(\gamma_2)$, such that each set collects all points to which the corresponding node is the closest node among all three nodes in terms of the distance we defined. In other words, we have $A(\alpha_2) = \{\rho \in (0, \infty)^2 \big| d(\alpha_2, \rho) <= d(\beta_2, \rho), d(\alpha_2, \rho) <= d(\gamma_2, \rho)\}$. So we can see that the entire space can be split with the two dashed lines in figure 3.2. Now we define $W(\alpha_1\rho) \equiv Prob(S_{t_2} \in A(\rho)|S_{t_1} = \alpha_1)$, where $\rho = \alpha_2,\ \beta_2,\ \gamma_2$. In general, Monte Carlo simulation is required to estimate these probabilities. Therefore, WSM method requires the generation of two mutually independent random sequences:

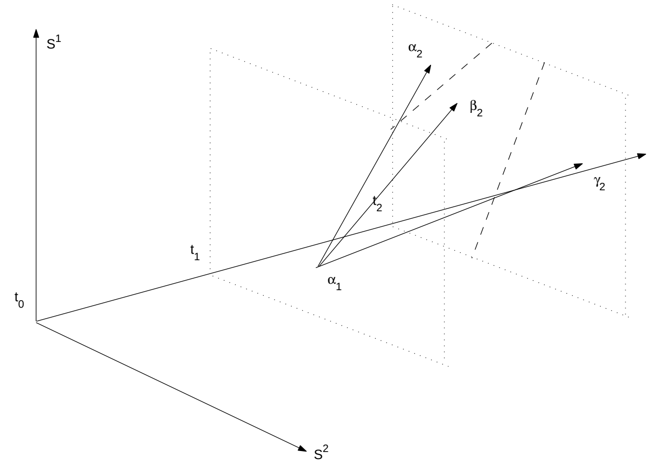

Figure 3.2: Weight Definition of Weighted Stochastic Mesh Method

- Node sequence of state vector $\{S_{t,i}; i = 1, \cdots, b, t = 0, \cdots, T\}$

b is the number of nodes at each epoch. $S_{0,i} \equiv S_0$ for $i = 1, \cdots, b$. For any fixed $t \in \{1, \cdots, T\}$, $\{S_{t,i}; i = 1, \cdots, b\}$ are generated as i.i.d. samples from a node generation density function, g_t.

For the case where the risk-neutral probability measure of S_t given S_0 is absolutely continuous, we restrict $g_t > 0$ on $G_t \subset R^n_+$, the open set of points where the risk-neutral density is positive. Now we consider the case where the measure is not absolutely continuous. We first split the state vector into $S^{(1)}_t$ and $S^{(2)}_t$, where $S^{(1)}_t = (S^1_t, \cdots, S^{n_1}_t)$ and $S^{(2)}_t = (S^{n_1+1}_t, \cdots, S^n_t)$ for some $1 \leq n_1 < n$. Then we assume (i) $S^{(2)}_t = f_t(S^{(1)}_t, S_0)$, where f_t is some known deterministic and continuous function; and (ii) the probability measure of $S^{(1)}_t$ given S_0 is absolutely continuous. Under the assumptions, we can generate $S^{(1)}_{t,i}$ via density g'_t, which is positive on $G'_t \subset R^{n_1}_+$, the open set of points where the risk-neutral density of $S_t^{(1)}$ given S_0 is positive. $S_{t,i}^{(2)}$ is obtained by letting $S_{t,i}^{(2)} = f_t(S_{t,i}^{(1)}, S_0)$. Hence, g_t does not exist for this case. We define G_t as $\{S_t : S_t^{(1)} \in G_t', S_t^{(2)} = f_t(S_t^{(1)}, S_0)\}$.

- Transition sequence of state vector $\{X_{t,i,m}; t = 0, \cdots, T-1, i = 1, \cdots, b, m = 1, \cdots, M\}$

  M is the number of transitions for each node at epochs prior to T. For any fixed $t \in \{0, \cdots, T-1\}$ and $i \in \{1, \cdots, b\}$, $\{X_{t,i,m}; m = 1, \cdots, M\}$ are generated as i.i.d. samples from the risk-neutral density of S_{t+1} given $S_t = S_{t,i}$. (Note that a closed-form expression for the density function is not necessarily known, e.g., the jump diffusion model proposed by Kou and Wang (2001). It suffices to know how to generate random variate with the density function.) Moreover, X is independent across t, but common random numbers shall be used to generate X across i for fixed t and m.

Then the estimator of the option value is defined recursively as follows:

$$\widehat{Q}(t, S_{t,i}) = \begin{cases} \max(h(t, S_{t,i}), \sum\limits_{j\in B} \widehat{w}(t, S_{t,i}, S_{t+1,j})\widehat{Q}(t+1, S_{t+1,j})) & t < T \\ h(T, S_{t,i}) & t = T \end{cases} \tag{3.3}$$

where

$$\widehat{w}(t, S_{t,i}, S_{t+1,j}) = \frac{1}{M}\sum_{m=1}^{M} 1(d_t(X_{t,i,m}, S_{t+1,j}) = \min_{k\in B} d_t(X_{t,i,m}, S_{t+1,k})) \tag{3.4}$$

and $1(\cdot)$ is the indicator function and $d_t(x, y)$ is a metric defined on R_+^n. If there are more than one $j \in B$, say $j_1 < j_2$, such that $d_t(X_{t,i,m}, S_{t+1,j}) = \min\limits_{k\in B} d_t(X_{t,i,m}, S_{t+1,k})$, we will let the indicator function take one only for j_1.

Now we give some additional notation. For any $t = 0, \cdots, T$, $D_t(f)$ represents the set of discontinuity of function $f(t,x)$ in x and $D_{t,r}(f) = \{x \in R_+^n : d_t(x, D_t(f)) \leq r\}$ for some $r > 0$. (If $D_t(f) = \phi$, $D_{t,r}(f) \equiv \phi$ for any $r > 0$.) For any $t = 0, \cdots, T$ and $x \in R_+^n$, we define the holding value of the option as

$$V(t,x) = \begin{cases} E[Q(t+1, S_{t+1})|S_t = x] & t < T \\ \\ 0 & t = T \end{cases} \tag{3.5}$$

To establish the convergence of the estimator, we first present a Lemma that requires the following three assumptions. We now write $\overline{A}$ as the closure of any set A, and write $V \Subset U$ when V and U are both open sets and the closure of V is compact and contained in U.

(A1) For any $t = 1, \cdots, T$, there exists a $one - to - one$ continuous map Φ_t that maps G_t onto $(-\infty, \infty)^n$ and the boundary of G_t onto that of $(-\infty, \infty)^n$.

(A2) $K \equiv \sup\{h(t,x); t = 0, \cdots, T, x \in R_+^n\} < \infty$, and $P(S_{t+1} \in \overline{D}_t(h)|S_t = x) = 0$ for any $x \in R_+^n$ and $t = 0, \cdots, T-1$.

(A3) $L_\lambda(x) \equiv E[e^{-\lambda S_{t+1}}|S_t = x]$ is continuous in $x \in R_+^n$ for any $\lambda \in [0, \infty)^n$ and $t = 0, \cdots, T$.

Lemma 3.1: Under assumptions (A1)-(A3), we have

(i) $V(t,x)$ is continuous in $x \in G_t$ for all $t = 0, \cdots, T$;

(ii) For any $\varepsilon \in (0,1)$ and any $A_t \Subset G_t$, where $t = 1, \cdots, T-1$, there exist $A_{t+1} \Subset G_{t+1}$ and $r_{t+1} > 0$ such that $\sup_{x \in \overline{A}_t} P(S_{t+1} \notin \overline{A}_{t+1}|S_t = x) < \varepsilon$ and $\sup_{x \in \overline{A}_t} P(S_{t+1} \in D_{t+1, r_{t+1}}(h)|S_t = x) < \varepsilon$, respectively.

Proof. We first prove the continuity of $V(t,x)$ in x by induction. Obviously there is nothing to prove for $t = T$. The induction hypothesis that $V(t+1,x)$ is continuous implies $D_{t+1}(Q) \subset D_{t+1}(h)$. Let $\{x_n\}$ be any sequence in G_t which converges to x. Let μ_n and μ be the probability measures of S_{t+1} when $S_t = x_n$ and $S_t = x$, respectively. Then assumption (A3) implies the weak convergence of μ_n to μ and (A2) implies $\mu(D_{t+1}(Q)) = 0$. Now Theorem 5.2(*iii*) in Billingsley (1968) gives $\int Q d\mu_n \to \int Q d\mu$, i.e., $V(t,x_n) \to V(t,x)$, which implies $V(t,x)$ is continuous in x.

Now we will prove the first part of (*ii*), i.e., the existence of A_{t+1} given A_t. Let x be arbitrary point in $\overline{A}_t$ and μ_x be the probability measure of S_{t+1} when $S_t = x$. We associate G_t with a metric ρ that is defined as $\rho(x,y) = |\Phi_t(x) - \Phi_t(y)|$ for any $x, y \in G_t$, where Assumption (A1) guarantees the existence of Φ_t. Then $\langle G_t, \rho \rangle$ is a polish space the closure of any sphere in this space is compact. If we go through the proof of Prohorov's Theorem, e.g., in Billingsley (1968), it suffices to show the family of measures $\{\mu_x : x \in \overline{A}_t\}$ is relatively compact. Let $\{x_n\}$ be any sequence in A_t, then there exists a subsequence, $\{x'_n\}$, that converges to some $x \in G_t$. Namely, for the sequence $\{\mu_{x_n}\}$, we can find a subsequence $\{\mu_{x'_n}\}$ that weakly converges to μ_x. Hence we prove the relative compactness.

Suppose the second part of (*ii*) does not hold. Then for $r_n \downarrow 0$, there exists $x_n \in A_t$ such that $\mu_{x_n}(D_{t,r_n}(h)) \geq \varepsilon$. Without loss of generality, we assume x_n converges to $x \in A_t$. Then μ_{x_n} converges to μ_x weakly. Since $D_{t,r_n}(h)$ is a closed set, for any n we have

$$\mu_x(D_{t,r_n}(h)) \geq \overline{\lim}_m \mu_{x_m}(D_{t,r_n}(h)) \geq \varepsilon.$$

However, $D_{t,r_n}(h) \downarrow \overline{D}_t(h)$ implies $\mu_x(D_{t,r_n}(h)) \downarrow \mu_x(\overline{D}_t(h)) = 0$. The contradiction gives the second part of (ii). $\square$

Now we make the dependence of $\widehat{Q}(t,x)$ on b and M explicit by denoting the estimator as $\widehat{Q}_{b,M}(t,x)$. The convergence of the estimator can be stated as:

Theorem 3.2: Under assumptions (A1)-(A3), for all $x \in R_+^n$ and $t = 0, \cdots, T$,

$$\lim_{b,M\to\infty} \left\| \widehat{Q}_{b,M}(t,x) - Q(t,x) \right\| = 0, \tag{3.6}$$

where $\|\cdot\|$ denotes the L^1 norm $E|\cdot|$.

Remarks:

- $b, M \to \infty$ is equivalent to saying $b^2 + M^2 \to \infty$. Indeed the proof guarantees the convergence of $\widehat{V}_{b,M}(t,x)$ to $V(t,x)$ as well.
- Assumption (A1) holds for many cases. For example, $G_t = R_+^n$, Φ_t may map $x \in R_+^n$ onto $y \in R^n$ with $y^i = \log x^i$.
- Assumption (A2) covers ordinary put options. Call options can be also considered by applying truncation. Obviously, options with discontinuous payoff function such as digital options are also covered. Indeed, this enables us to handle barrier options with some adjustment applied to the algorithm.
- Assumption (A3) holds if S_{t+1} can be expressed as $f(S_t, \xi)$, where f is a continuous in S_t and ξ represents the randomness. Hence multiplicative process, a process where $\log(S_{t+1}/S_t)$ is independent of S_t, is covered in this case. Also,

if S_t can be characterized with an SDE, $L_\lambda(x)$ will be the solution of a corresponding PDE and assumption (A2) will hold given some regularity condition on the drift and diffusion terms of the SDE.

- The computational effort in generating the mesh is proportional to $n \times b \times d$. The effort in the recursive pricing is proportional to $n \times M \times b \times d$. Hence the overall effort is polynomial in the problem dimension (n), the mesh parameter (b), the sample size of weight estimating (M), and the number of exercise opportunities ($d + 1$)

Proof. Given $\varepsilon > 0$, there exists $A'_1 \Subset G_1$ and $r_1 > 0$ such that $P[S_1 \notin \overline{A'_1}|S_0] < \varepsilon$ and $P[S_1 \in D_{1,r_1}(h)|S_0] < \varepsilon$. There exists A_1 such that $A'_1 \Subset A_1 \Subset G_1$. Suppose we have defined A'_t and A_t for $t = 1, \cdots, T-1$, Lemma 3.1 guarantees there exists $A'_{t+1} \Subset G_{t+1}$ and r_{t+1} such that $\sup_{x \in \overline{A}_t} P(S_{t+1} \notin \overline{A'_{t+1}}|S_t = x) < \varepsilon$ and $\sup_{x \in A_t} P(S_{t+1} \in D_{t+1,r_{t+1}}(h)|S_t = x) < \varepsilon$. Also, there exists A_{t+1} such that $A'_{t+1} \Subset A_{t+1} \Subset G_{t+1}$. Following this route, we can construct $\{A'_t, r_t, A_t; t = 1, \cdots, T\}$ recursively. Furthermore, we construct sequence $\{\varepsilon_t, \delta_t, \alpha_t; t = 1, \cdots, T\}$. Let $\varepsilon_t = d_t(\overline{A'_t}, A^c_t)$, where we write A^c_t as $\{x : x \in G_t, x \notin A_t\}$. Obviously, $\overline{A_t} \cap D^c_{t,r_t/2}(h)$ is a compact set on which $Q(t, x)$ is continuous thus uniform continuous in x. Hence, there exists $\delta_t \in (0, \varepsilon_t \wedge \frac{r_t}{2}]$ such that $|Q(t, x) - Q(t, y)| < \varepsilon$ whenever $d_t(x, y) \leq \delta_t$ and $x, y \in A_t \cap D^c_{t,r_t/2}(h)$. Then let $\alpha_t = \sup_{y \in A'_t} P_{g_t}[d_t(S_{t,1}, y) \geq \delta_t]$, where P_{g_t} is probability measure with density g_t. It is trivial to show that $\varepsilon_t > 0$ for any t. At last, we define $\alpha = \max\{\alpha_t; 1 \leq t \leq T\}$. Note $\alpha < 1$ is guaranteed by the property of g_t, or g'_t for the degenerate case.

Defining A_0 as the singleton $\{S_0\}$, now we will prove for any $\varepsilon > 0$, $x \in \overline{A}_t$, $b > 1$, $M > 1$ and $t = 0, \cdots, T$,

$$\Delta_{b,M}(t,x) \equiv \left\| \widehat{Q}_{b,M}(t,x) - Q(t,x) \right\| \le (T-t)\big[2K\big(3\varepsilon + b\alpha^b + \alpha^b + \frac{1}{\sqrt{M}}\big) + \varepsilon\big], \quad (3.7)$$

where K is defined in assumption (A2).

By induction, we proceed backwards from the terminal time. At T there is nothing to prove because $\widehat{Q}_{b,M}(T,\cdot) \equiv h(T,\cdot) \equiv Q(T,\cdot)$. Take as induction hypothesis that

$$\Delta_{b,M}(t+1,z) \le (T-t-1)\big[2K\big(3\varepsilon + b\alpha^b + \alpha^b + \frac{1}{\sqrt{M}}\big) + \varepsilon\big], \forall z \in A_{t+1}$$

Now we fix x in A_t and denote P_x as the probability measure given $S_t = x$. Using $|\max(a,b) - \max(a,c)| \le |b-c|$, we have

$$\begin{aligned}
\Delta_{b,M}(t,x) \ &\le \ \left\| \sum_{j\in B} \widehat{w}(t,x,S_{t+1,j})\widehat{Q}_{b,M}(t+1,S_{t+1,j}) - V(t,x) \right\| \\
&\le \ \left\| \sum_{j\in B} [\widehat{w}(t,x,S_{t+1,j}) - w(t,x,S_{t+1,j})]\widehat{Q}_{b,M}(t+1,S_{t+1,j}) \right\| + \\
&\quad\ \left\| \sum_{j\in B} w(t,x,S_{t+1,j})[\widehat{Q}_{b,M}(t+1,S_{t+1,j}) - Q(t+1,S_{t+1,j})] \right\| + \\
&\quad\ \left\| \sum_{j\in B} w(t,x,S_{t+1,j})Q(t+1,S_{t+1,j}) - V(t,x) \right\| \\
&= \ I_1 + I_2 + I_3
\end{aligned} \quad (3.8)$$

where $w(t,x,S_{t+1,j}) = P_x[d_{t+1}(S_{t+1},S_{t+1,j}) = \min_{k\in B} d_{t+1}(S_{t+1},S_{t+1,k}) | S_{t+1,j}; j \in B]$. We now bound I_1, I_2, and I_3 one by one.

We will first define $\{\eta_b^i; i = 1, \cdots, M\}$ as $\eta_b^i = \arg\min_{y\in\{S_{t+1,j};\ j\in B\}} d_{t+1}(y,\xi_i)$, where $\{\xi_i; i = 1, \cdots, M\}$ are i.i.d. sequence with the distribution of S_{t+1} given

$S_t = x$ and independent of $\{S_{t+1,j}; j \in B\}$. Let R be the $\sigma-$algebra generated by $\widehat{Q}_{b,M}$ and $\{S_{t+1,j}; j \in B\}$. Then we have

$$\begin{aligned}
\sum_{j \in B} \widehat{w}(t, x, S_{t+1,j}) \widehat{Q}_{b,M}(t+1, S_{t+1,j}) &= \frac{1}{M} \sum_{i=1}^{M} \widehat{Q}_{b,M}(t+1, \eta_b^i) \\
\sum_{j \in B} \widehat{w}(t, x, S_{t+1,j}) \widehat{Q}_{b,M}(t+1, S_{t+1,j}) &= E[\widehat{Q}_{b,M}(t+1, \eta_b^1)|R].
\end{aligned}$$

Hence,

$$\begin{aligned}
I_1 &= \left\| E[|\frac{1}{M} \sum_{i=1}^{M} \widehat{Q}_{b,M}(t+1, \eta_b^i) - E\widehat{Q}_{b,M}(t+1, \eta_b^1)||R] \right\| \\
&\leq \left\| E[\sqrt{\frac{1}{M} E|\widehat{Q}_{b,M}(t+1, \eta_b^1) - E\widehat{Q}_{b,M}(t+1, \eta_b^1)|^2} |R] \right\| \\
&\leq \frac{2K}{\sqrt{M}}, \qquad (3.9)
\end{aligned}$$

where the first step is by conditional expectation, the second by Cauchy-Schwartz inequality and the third by the boundedness in assumption (A2).

$$\begin{aligned}
I_2 &\leq \left\| \sum_{j \in B} w(t, x, S_{t+1,j}) E|\widehat{Q}_{b,M}(t+1, S_{t+1,j}) - Q(t+1, S_{t+1,j})| \right\| \\
&= \left\| \sum_{j \in B} w(t, x, S_{t+1,j}) \Delta_{b,M}(t+1, S_{t+1,j}) \right\| \\
&= \left\| \Delta_{b,M}(t+1, \eta_b^1)[1(\eta_b^1 \notin A_{t+1}) + 1(\eta_b^1 \in A_{t+1})] \right\| \\
&\leq 2KP_x(\eta_b^1 \notin A_{t+1}) + (T-t-1)[2K(3\varepsilon + b\alpha^b + \alpha^b + \frac{1}{\sqrt{M}}) + \varepsilon] \qquad (3.10)
\end{aligned}$$

where the first step is by conditional expectation, the second by the definition of $\Delta_{b,M}$ and the third by the definition of η_b^1 and conditional expectation. The last is by the induction hypothesis. Now we have to bound $P_x(\eta_b^1 \notin A_{t+1})$.

$$\begin{aligned}
P_x(\eta_b^1 \notin A_{t+1}) &\leq P_x(\xi_1 \notin A'_{t+1}) + P_x(\eta_b^1 \notin A_{t+1}, \xi_1 \in A'_{t+1}) \\
&\leq \varepsilon + bP_x(S_{t+1,1} \notin A_{t+1}, S_{t+1,1} = \eta_b^1, \xi_1 \in A'_{t+1}) \\
&\leq \varepsilon + bE[P_{g_{t+1}}(d_{t+1}(S_{t+1,j}, \xi_1) \geq \varepsilon_{t+1}; j \in B)|\xi_1 = y \in A'_{t+1}] \\
&\leq \varepsilon + b\alpha_{t+1}^b \leq \varepsilon + b\alpha^b \qquad (3.11)
\end{aligned}$$

where the third inequality is by the definition of ε_{t+1} and the last by the independence of $S_{t+1,j}$ and the definition of α_{t+1}.

Since $V(t,x)$ does not depend on $\{S_{t+1,j}; j \in B\}$, we can write

$$V(t,x) = EQ(t+1,\xi_1) = E[Q(t+1,\xi_1)|S_{t+1,j}, j \in B].$$

Substitute in I_3,

$$\begin{aligned}
I_3 &= \left\|E\{[Q(t+1,\eta_b^1) - Q(t+1,\xi_1)]|S_{t+1,j}, j \in B\}\right\| \\
&\leq \left\|Q(t+1,\eta_b^1) - Q(t+1,\xi_1)\right\| \\
&= \left\|[Q(t+1,\eta_b^1) - Q(t+1,\xi_1)][1(\xi_1 \notin A'_{t+1} \cap D^c_{t+1,r_{t+1}}(h)) + 1(\xi_1 \in A'_{t+1} \cap D^c_{t+1,r_{t+1}}(h))]\right\| \\
&\leq 4K\varepsilon + \left\|[Q(t+1,\eta_b^1) - Q(t+1,\xi_1)]1(\xi_1 \in A'_{t+1} \cap D^c_{t+1,r_{t+1}}(h), d_{t+1}(\eta_b^1, \xi_1) \leq \delta_{t+1})\right\| \\
&\quad + \left\|[Q(t+1,\eta_b^1) - Q(t+1,\xi_1)]1(\xi_1 \in A'_{t+1} \cap D^c_{t+1,r_{t+1}}(h), d_{t+1}(\eta_b^1, \xi_1) > \delta_{t+1})\right\| \\
&\leq (4K+1)\varepsilon + 2KE[P_{g_{t+1}}(d_{t+1}(S_{t+1,j}, \xi_1) > \delta_{t+1}; j \in B)|\xi_1 = y \in A'_{t+1}] \\
&\leq (4K+1)\varepsilon + 2K\alpha^b, \qquad (3.12)
\end{aligned}$$

where the first step is by the definition of η_b^1, the second by conditional expectation, the last three are by the definitions of A'_{t+1}, r_{t+1}, δ_{t+1} and α, respectively.

Then (3.7) is immediate from (3.8-3.12). For any given $x_0 \in R_+^n$, we could find sufficiently small ε_0 such that $x_0 \in A_t, \forall t$ whenever $\varepsilon < \varepsilon_0$. Hence, (3.7) implies $\lim\limits_{b,M\to\infty} \Delta_{b,M}(t,x_0) \leq (4K+1)T\varepsilon$. Since ε can be arbitrarily small, we complete our proof. □

Theorem 3.2 guarantees the asymptotic convergence of the weighted stochastic mesh estimator when the behavior of underlying securities can be simulated exactly. However, we usually do not know the solutions to the SDE characterizing the securities. For this case, we have to apply approximation techniques to generating the transition sequence required by WSM. A result similar to Theorem 3.2 will be introduced after we introduce some additional notations and definitions.

We consider the $n-$dimensional case with the process X satisfying the SDE

$$dX_t = a(X_t)dt + b(X_t)dW_t.$$

Let $\{A_t, t \geq 0\}$ be an increasing family of $\sigma-$algebras associated with X. We call $(\tau)_\Delta = \{\tau_n : n = 0, 1, \cdots, \frac{T-t_0}{\Delta}\}$, a time discretization of a bounded interval $[t_0, T]$ with $\tau_n = t_0 + n\Delta$. We call a process $Y = \{Y(t), t \geq 0\}$, which is right continuous with left hand limits, a time discrete approximation with step size Δ, if it is based on a time discretization $(\tau)_\Delta$ such that $Y(\tau_n)$ is $A_{\tau_n}-$measurable and $Y(\tau_{n+1})$ can be expressed as a function of both $Y(\tau_0), \cdots, Y(\tau_n), \tau_0, \cdots, \tau_{n+1}$ and a finite number of $A_{\tau_{n+1}}-$measurable random variables.

We shall say that a general time discrete approximation Y^Δ with step size Δ **converges strongly with order** $\gamma > 0$ at time T if there exists a positive constant

C, which does not depend on Δ, and a $\Delta_0 > 0$ such that for each $\Delta \in (0, \Delta_0)$

$$E|X_T - Y^{\Delta}(T)| \leq C\Delta^{\gamma}$$

We need one additional assumption.

(A4) Given $S_t = x$ and $x \in R_+^n$, there exists a strongly convergent time discrete approximation S_{t+1}^{Δ} of order γ such that the constant C only depends on r for any $x \in [e^{-r}, e^r]^n$.

Now we can use such approximation to generate $\{X_{t,i,m}^{\Delta}; m = 1, \cdots, M\}$, the second sequence required by WSM. Denoting the new estimator as $\widehat{Q}_{b,M,\Delta}(0, S_0)$, we have

Theorem 3.3: Under assumption (A1-4), for all $x \in R_+^n$ and $t = 0, \cdots, T$,

$$\lim_{b,M,\frac{1}{\Delta}\to\infty} \left\| \widehat{Q}_{b,M,\Delta}(t,x) - Q(t,x) \right\| = 0 \tag{3.13}$$

Remark: We will use the strong convergent approximations in Kloeden and Platen (1992), which can be shown to satisfy Assumption (A4).

Proof. In the proof, we will write $|A_1 - A_2|$ as $\equiv \min\{|x-y|; x \in A_1, y \in A_2\}$ for any two sets A_1 and A_2. Obviously, the continuity of $V(t,x)$ in x can be similarly shown.

We will construct sequence $\{A'_t, r_t, A_t; t = 1, \cdots, T\}$ as in the proof of Theorem 3.2. Furthermore, we construct sequence $\{A''_t, \varepsilon_t, \delta_t, \alpha_t; t = 1, \cdots, T\}$. Let A''_t be such that $A'_t \Subset A''_t \Subset A_t$ and $\varepsilon_t = \min(\frac{r_t}{4}, |\overline{A'_t} - A''^c_t|, |\overline{D}_{t,r_t/2}(h) - D^c_{t,r_t}(h)|, d_t(\overline{A''_t}, A^c_t))$. Obviously, $\overline{A_t} \cap D^c_{t,r_t/4}(h)$ is a compact set on which $Q(t,x)$ is continuous thus uniform continuous in x. Hence, there exists $\delta_t \in (0, \varepsilon_t]$ such that $|Q(t,x) - Q(t,y)| < \varepsilon$

whenever $d_t(x,y) \leq \delta_t$ and $x, y \in A_t \cap D^c_{t,r_t/4}(h)$. Then let $\alpha_t = \sup_{y \in A''_t} P_{g_t}[d_t(S_{t,1}, y) \geq \delta_t]$, where P_{g_t} is probability measure with density g_t. Plus, we define $\{C_t; t = 0, \cdots, T-1\}$ as the constant which only depends on r_t in assumption (A4). Note we still can show that $\varepsilon_t > 0$ and $\alpha_t < 1$ for any t. At last, let $\delta = \min\{\delta_t; t = 1, \cdots, T\}$, $\alpha = \max\{\alpha_t; 1 \leq t \leq T\}$ and $C = \max\{C_t; t = 0, \cdots, T-1\}$.

Also defining A_0 as the singleton $\{S_0\}$, now we will prove for any $\varepsilon > 0$, $x \in \overline{A}_t$, $b > 1$, $M > 1$ and $t = 0, \cdots, T$,

$$\begin{aligned} \Delta_{b,M,\Delta}(t,x) &\equiv \left\| \widehat{Q}_{b,M,\Delta}(t,x) - Q(t,x) \right\| \\ &\leq (T-t)[2K(5\varepsilon + \frac{4C}{\delta}\Delta^\gamma + b\alpha^b + \alpha^b + \frac{1}{\sqrt{M}}) + 2\varepsilon]. \end{aligned} \tag{3.14}$$

By induction, we proceed backwards from the terminal time. At T there is nothing to prove because $\widehat{Q}_{b,M,\Delta}(T,\cdot) \equiv h(T,\cdot) \equiv Q(T,\cdot)$. Take as induction hypothesis that

$$\Delta_{b,M,\Delta}(t+1,z) \leq (T-t-1)[2K(5\varepsilon + \frac{4C}{\delta}\Delta^\gamma + b\alpha^b + \alpha^b + \frac{1}{\sqrt{M}}) + 2\varepsilon], \forall z \in A_{t+1}$$

Now we fix x in A_t. Using $|\max(a,b) - \max(a,c)| \leq |b-c|$, we have

$$\begin{aligned} \Delta_{b,M,\Delta}(t,x) &\leq \left\| \sum_{j \in B} \widehat{w}(t,x,S_{t+1,j}) \widehat{Q}_{b,M,\Delta}(t+1,S_{t+1,j}) - V(t,x) \right\| \\ &\leq \left\| \sum_{j \in B} [\widehat{w}(t,x,S_{t+1,j}) - w^\Delta(t,x,S_{t+1,j})] \widehat{Q}_{b,M,\Delta}(t+1,S_{t+1,j}) \right\| + \\ &\quad \left\| \sum_{j \in B} w^\Delta(t,x,S_{t+1,j}) [\widehat{Q}_{b,M,\Delta}(t+1,S_{t+1,j}) - Q(t+1,S_{t+1,j})] \right\| + \\ &\quad \left\| \sum_{j \in B} w^\Delta(t,x,S_{t+1,j}) Q(t+1,S_{t+1,j}) - V(t,x) \right\| \\ &= I_1 + I_2 + I_3 \end{aligned} \tag{3.15}$$

where $w^{\Delta}(t, x, S_{t+1,j}) = P_x[d_{t+1}(S^{\Delta}_{t+1}, S_{t+1,j}) = \min_{k \in B} d_{t+1}(S_{t+1}, S_{t+1,k}) | S_{t+1,j}; j \in B]$. Note S^{Δ}_{t+1} has the approximate distribution assumed in Assumption (A4). Now we will also bound I_1, I_2 and I_3 one by one.

We will first define $\{\eta^i_{b,\Delta}; i = 1, \cdots, M\}$ as $\eta^i_{b,\Delta} = \arg\min_{y \in \{S_{t+1,j};\, j \in B\}} d_{t+1}(y, \xi^{\Delta}_i)$, where $\{\xi^{\Delta}_i; i = 1, \cdots, M\}$ are i.i.d. sequence with the distribution of S^{Δ}_{t+1} given $S_t = x$ and independent of $\{S_{t+1,j}; j \in B\}$. Then we still have

$$I_1 \leq \frac{2K}{\sqrt{M}}. \tag{3.16}$$

To bound I_2, we have

$$\begin{aligned} I_2 &\leq \left\| \Delta_{b,M,\Delta}(t+1, \eta^1_{b,\Delta})[1(\eta^1_{b,\Delta} \notin A_{t+1}) + 1(\eta^1_{b,\Delta} \in A_{t+1})] \right\| \\ &\leq 2K P_x(\eta^1_{b,\Delta} \notin A_{t+1}) + (T - t - 1)[2K(5\varepsilon + \frac{4C}{\delta}\Delta^{\gamma} + b\alpha^b + \alpha^b + \frac{1}{\sqrt{M}}) + 2\varepsilon] \end{aligned} \tag{3.17}$$

where the last step is by the induction hypothesis. To bound $P_x(\eta^1_{b,\Delta} \notin A_{t+1})$, we have

$$\begin{aligned} P_x(\eta^1_{b,\Delta} \notin A_{t+1}) &\leq P_x(\xi^{\Delta}_1 \notin A''_{t+1}) + P_x(\eta^1_{b,\Delta} \notin A_{t+1}, \xi^{\Delta}_1 \in A''_{t+1}) \\ &\leq P_x(\xi^{\Delta}_1 \notin A''_{t+1}) + b P_x(S_{t+1,1} \notin A_{t+1}, S_{t+1,1} = \eta^1_{b,\Delta}, \xi^{\Delta}_1 \in A''_{t+1}) \\ &\leq P_x(\xi_1 \notin A'_{t+1}) + P_x(\xi^{\Delta}_1 \notin A''_{t+1}, \xi_1 \in A'_{t+1}) \\ &\quad + bE[P_{g_{t+1}}(d_{t+1}(S_{t+1,j}, \xi^{\Delta}_1) \geq \delta_{t+1}; j \in B) | \xi^{\Delta}_1 = y \in A''_{t+1}] \\ &\leq \varepsilon + P_x(|\xi^{\Delta}_1 - \xi| \geq \varepsilon_{t+1}) + b\alpha^b_{t+1} \\ &\leq \varepsilon + b\alpha^b + \frac{C_{t+1}}{\delta_{t+1}}\Delta^{\gamma} \leq \varepsilon + b\alpha^b + \frac{C}{\delta}\Delta^{\gamma} \end{aligned} \tag{3.18}$$

where the forth inequality is by the definition of A'_{t+1} and α_{t+1}, and the last by Chebychev inequality and Assumption (A4).

To bound I_3, we have

$$\begin{aligned}
I_3 &= \big\| E\{[Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1)]|S_{t+1,j}, j \in B\}\big\| \\
&\leq \big\| Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1)\big\| \\
&\leq \big\| Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1^\Delta)\big\| + \big\| Q(t+1,\xi_1^\Delta) - Q(t+1,\xi_1)\big\| \\
&= I_4 + I_5. \qquad (3.19)
\end{aligned}$$

Furthermore,

$$\begin{aligned}
I_4 &\leq \big\|[Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1^\Delta)]1(\xi_1^\Delta \notin A''_{t+1} \cap D^c_{t+1,r_{t+1}/2}(h))\big\| \\
&\quad + \big\|[Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1^\Delta)]1(\xi_1^\Delta \in A''_{t+1} \cap D^c_{t+1,r_{t+1}/2}(h))\big\| \\
&\leq 2KP_x(\xi_1^\Delta \notin A''_{t+1} \cap D^c_{t+1,r_{t+1}/2}(h)) + \\
&\quad \big\|[Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1^\Delta)]1(\xi_1^\Delta \in A''_{t+1} \cap D^c_{t+1,r_{t+1}/2}(h))\big\| \\
&\leq 2KP_x(\xi_1^\Delta \notin A''_{t+1}) + 2KP_x(\xi_1^\Delta \in D_{t+1,r_{t+1}/2}(h)) \\
&\quad \big\|[Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1^\Delta)]1(\xi_1^\Delta \in A''_{t+1} \cap D^c_{t+1,r_{t+1}/2}(h), d_{t+1}(\eta^1_{b,\Delta},\xi_1^\Delta) \leq \delta_{t+1})\big\| + \\
&\quad \big\|[Q(t+1,\eta^1_{b,\Delta}) - Q(t+1,\xi_1^\Delta)]1(\xi_1^\Delta \in A''_{t+1} \cap D^c_{t+1,r_{t+1}/2}(h), d_{t+1}(\eta^1_{b,\Delta},\xi_1^\Delta) > \delta_{t+1})\big\| \\
&\leq 2K(\varepsilon + \frac{C}{\delta}\Delta^\gamma) + 2KP_x(\xi_1 \in D_{t+1,r_{t+1}}(h)) + \\
&\quad 2KP_x(\xi_1^\Delta \in D_{t+1,r_{t+1}/2}(h), \xi_1 \in D^c_{t+1,r_{t+1}}(h)) + \varepsilon + \\
&\quad 2KE[P_{g_{t+1}}(d_{t+1}(S_{t+1,j},\xi_1^\Delta) > \delta_{t+1}; j \in B)|\xi_1^\Delta = y \in A''_{t+1}] \\
&\leq 2K(2\varepsilon + \frac{C}{\delta}\Delta^\gamma + \alpha^b) + \varepsilon + 2KP_x(|\xi_1^\Delta - \xi_1| > \varepsilon_{t+1}) \\
&\leq 2K(2\varepsilon + \frac{2C}{\delta}\Delta^\gamma + \alpha^b) + \varepsilon, \qquad (3.20)
\end{aligned}$$

where the fourth step is by the argument used in (3.18) as well as uniform continuity, the fifth by the definition of r_{t+1} and α, the last by Chebychev inequality and assumption (A4).

For I_5, we have

$$
\begin{aligned}
I_5 \leq\ & \left\| [Q(t+1,\xi_1^\Delta) - Q(t+1,\xi_1)] 1(\xi_1 \notin A'_{t+1} \cap D^c_{t+1,r_{t+1}}(h)) \right\| + \\
& \left\| [Q(t+1,\xi_1^\Delta) - Q(t+1,\xi_1)] [1(\xi_1 \in A'_{t+1} \cap D^c_{t+1,r_{t+1}}(h), |\xi_1^\Delta - \xi_1| \leq \delta_{t+1}) \right\| + \\
& \left\| [Q(t+1,\xi_1^\Delta) - Q(t+1,\xi_1)] [1(\xi_1 \in A'_{t+1} \cap D^c_{t+1,r_{t+1}}(h), |\xi_1^\Delta - \xi_1| \geq \delta_{t+1}) \right\| \\
\leq\ & 2KP_x(\xi_1 \notin A'_{t+1}) + 2KP_x(\xi_1 \in D_{t+1,r_{t+1}}(h)) + \varepsilon + 2KP_x(|\xi_1^\Delta - \xi_1| \geq \delta_{t+1}) \\
\leq\ & 4K\varepsilon + \varepsilon + 2K\frac{C}{\delta}\Delta^\gamma, \qquad (3.21)
\end{aligned}
$$

where the second inequality is by the uniform continuity of Q, the third by the definition of A'_{t+1} and r_{t+1} and the last is by Chebychev inequality and assumption (A4).

Then (3.14) is immediate from (3.15-3.21). For any given $x_0 \in R^n_+$, we could find sufficiently small ε_0 such that $x_0 \in A_t, \forall t$ whenever $\varepsilon < \varepsilon_0$. Hence, (3.14) implies $\lim\limits_{b,M,\frac{1}{\Delta}\to\infty} \Delta_{b,M,\Delta}(t,x_0) \leq (6K+2)T\varepsilon$. Since ε can be arbitrarily small, we complete our proof. $\square$

In order to give a confidence interval for the option price Q, we generate N independent meshes with corresponding mesh estimates $\widehat{Q}^{(i)} = \widehat{Q}_b^{(i)}(0,S_0)$, $i = 1,\cdots,N$, and then the confidence interval can be calculated via the sample mean and sample variance.

3.3. Extension to Asian Options

We first consider Asian-style max options of the Bermudan flavor similar to that of Ben-Ameur et al. (2002). (We only consider max options here for simplification. Actually the arguments can be extended to other options like geometric average options.) Let $0 = t_0 \leq t_1 \leq \cdots \leq t_l = T$ be a fixed sequence of observation dates, where m^* is an integer satisfying $1 \leq m^* \leq l$. The exercise opportunities are at dates t_m, for $m^* \leq m \leq l$. If exercised at time t_m, we consider two types of payoff:

- $(\max\limits_{1\leq i\leq n} \overline{S}^i_{t_m} - K)^+$ where $\overline{S}^i_{t_m} = \frac{S^i_{t_1}+\cdots+S^i_{t_m}}{m}$ is the arithmetic average of the asset prices at the observation dates up to time t_m;
- $(\frac{\max\limits_{1\leq i\leq n}(S^i_{t_1})+\cdots+\max\limits_{1\leq i\leq n}(S^i_{t_m})}{m} - K)^+$.

For the case of Asian options, knowing the states of the underlying securities at one node is not sufficient to determine the exercise value or holding value at that node. We have to make some adjustment before WSM can be applied to pricing Asian options.

For the first type of payoff, we expand the state vector to a $2n-$dimensional vector:

$$\widetilde{S}_{t_m} = (S^1_{t_m}, \cdots, S^n_{t_m}, \overline{S}^1_{t_m}, \cdots, \overline{S}^n_{t_m}),$$

while for the second type of payoff, we expand to a $(n+1)-$dimensional vector:

$$\widetilde{S}_{t_m} = (S^1_{t_m}, \cdots, S^n_{t_m}, \frac{\max\limits_{1\leq i\leq n}(S^i_{t_1}) + \cdots + \max\limits_{1\leq i\leq n}(S^i_{t_m})}{m}).$$

Now the holding value or exercise value at each node can be determined by the new state vector of that node. Obviously if we can simulate the behavior of S_{t+1}

given S_t, we can also simulate $\widetilde{S}_{t_{m+1}}$ given $\widetilde{S}_{t_m}$. Thus theorem 3.2 still holds with the state vector expansion. Suppose $t_1 = 0$ in the first type, we will have $G_{t_m} = \{x \in R_+^{2n} : x^{n+i} > \frac{1}{m}x^i + \frac{1}{m}S_0^i; i = 1, \cdots, n\}$ for $m > 2$ and a degenerate situation for $m = 1$. Generally G_t is no longer the whole space R_+^n as we encounter for the ordinary American options. Similarly, for Theorem 3.3 to apply for the state vector expansion, we give the following Lemma without proof.

Lemma 3.4: Given S_{t_m}, let S^{Δ} be a time discretization approximation with maximum step size Δ, which converges strongly to $S_{t_{m+1}}$ with order $\gamma > 0$. Then given $\widetilde{S}_{t_m}$, $\widetilde{S}^{\Delta}$ will be a time discretization approximation strongly converging to $\widetilde{S}_{t_{m+1}}$ with the same order, where the ith component of $\widetilde{S}^{\Delta}$ is given as follows:

$$\widetilde{S}^{\Delta,i} = \begin{cases} S^{\Delta,i} & i = 1, \cdots, n; \\ (1 - \frac{1}{m+1})\widetilde{S}_{t_m}^i + \frac{1}{m+1}S^{\Delta,i-n} & i > n, \text{ for first type;} \\ (1 - \frac{1}{m+1})\widetilde{S}_{t_m}^i + \frac{1}{m+1}\max\limits_{1 \le j \le n} S^{\Delta,j} & i = n+1, \text{ for second type.} \end{cases}$$

Now we turn to Asian-style max options of the continuously sampled flavor. For an option exercised at $t \in [0, T]$, we again consider two types of payoff as follows:

- $(\frac{1}{t}\max\limits_{0 \le i \le n}\int\limits_0^t S_u^i du - K)^+$;
- $(\frac{1}{t}\int\limits_0^t \max\limits_{0 \le i \le n} S_u^i du - K)^+$.

Also, we have to expand the status vector before WSM can be applied to this case. For the first type of payoff, the state vector is expanded to a $2n-$dimensional vector $\widetilde{S}_t = (S_t, \overline{S}_t)$, where $\overline{S}_t = (\int\limits_0^t S_u^1 du, \cdots, \int\limits_0^t S_u^n du)$; while the state vector for the second type of payoff is expanded to a $(n+1)-$dimensional vector $\widetilde{S}_t = (S_t, \overline{S}_t')$, where $\overline{S}'_t = \int_0^t \max_{0\le i\le n} S^i_u du$. Now the expanded state vector of each node will determine the exercise value, as well as the holding value at that node. However, we can no longer simulate the exact behavior of $\widetilde{S}_{t+1}$ even when we can simulate exactly S_{t+1}. Let us only focus on the case where S_t is characterized by an SDE and there exists a time discrete approximation which converges strongly with order γ to S_{t+1} given S_t. By rewriting the SDE that characterize $\widetilde{S}_{t+1}$ as follows, we can find a corresponding time discrete approximation that converges strongly with the same order to $\widetilde{S}_{t+1}$ given $\widetilde{S}_t$.

$$\begin{aligned} S_{t+1} &= S_t + \int_t^{t+1} a(S_u)du + \int_t^{t+1} b(S_u)dW_u \\ \overline{S}_{t+1} &= \overline{S}_t + \int_t^{t+1} S_u du \\ \overline{S}'_{t+1} &= \overline{S}'_t + \int_t^{t+1} \max_{0\le i\le n} S^i_u du \end{aligned}$$

Thus, assumption (A4) applies for the state vector expansion.

3.4. Value at Risk (VaR)

The measure of Value at Risk (VaR) has become an important measure in the financial industry. However, there are few papers that address the calculation of VaR for American-style options. Since VaR is nothing more than a quantile of the option's potential profit and loss over a given time period, the problem is to calculate the $r-$th quantile of Q at time $s \in (0, T]$, denoted by $\xi_{r,s}$ and defined by $F^{-1}(r) = \inf\{u : F(u) \ge r\}$, where $F(u) \equiv P(Q(\tau^* \wedge s, S_{\tau^*\wedge s}) \le u|S_0)$ and $\wedge$

denotes the minimum operator, s and r are pre-specified, τ^* is the optimal stopping time that solves problem in (3.1), and Q has been discounted to time 0. The major difference between American-style and European-style options in calculating the VaR now for some time $s > 0$ in the future is that an American-style option may already have been exercised by time s. Hence, we can not just base our calculation on the behavior of the underlying securities at s but require the entire path up to s.

The path estimator proposed in Broadie and Glasserman (2004) gives an idea for estimating VaR. We will simulate a trajectory of the underlying securities until the exercise region determined by the mesh or s is reached. Independent of the mesh points, the path $S_v = (S_{v,0}, \cdots, S_{v,s})$ will be simulated according to the risk-neutral density of S_t. Along the path, we will generate a sequence $\{X_{t,m}; m = 1, \cdots, M\}$ under the risk neutral density of S_{t+1} given $S_t = S_{v,t}$ for each $t \in \{0, \cdots, s \wedge (T-1)\}$. Note this sequence has to share the random numbers with the transition sequence generated by WSM. Then we can define an approximate stopping rule by letting $\widehat{\tau}(S_v) = \min\{t : h(t, S_{v,t}) = \widehat{Q}(t, S_{v,t})\}$, where

$$\widehat{Q}(t, S_{v,t}) = \begin{cases} \max(h(t, S_{v,t}), \sum\limits_{j \in B} \widehat{w}(t, S_{v,t}, S_{t+1,j}) \widehat{Q}(t+1, S_{t+1,j})) & t < T \wedge s \\ \\ h(T, S_{v,t}) & t = T \text{ if } s = T \end{cases} \tag{3.22}$$

$$\widehat{w}(t, S_{v,t}, S_{t+1,j}) = \frac{1}{M} \sum_{m=1}^{M} 1(d_t(X_{t,m}, S_{t+1,j}) = \min_{k \in B} d_t(X_{t,m}, S_{t+1,k})). \tag{3.23}$$

We might have two cases: if an exact simulation is available for S_v, $\widehat{Q} = \widehat{Q}_{b,M}$; otherwise, $S_{v,t}, S_v$ are replaced by $S^{\Delta}_{v,t}, S^{\Delta}_v$ and $\widehat{Q} = \widehat{Q}_{b,M,\Delta}$.

Generate n_v independent paths $\{S^i_{v,t}; i = 1, \cdots, n_v, t = 0, \cdots, s\}$ for each mesh and calculate $\widehat{Q}_v^i = \widehat{Q}_{b,M}(\widehat{\tau} \wedge s, S^i_{v,\widehat{\tau}\wedge s})$, where $\widehat{\tau} = \widehat{\tau}(S_v^i)$. Then $\widehat{\xi}_{r,s} = F_{n_v}^{-1}(r) = \inf\{u : F_{n_v}(u) \geq r\}$ will be the estimator of $r-$th VaR at time $s \in (0, T]$, where $F_{n_v}(u) = \frac{1}{n_v}\sum_{i=1}^{n_v} 1(\widehat{Q}_v^i \leq u)$. ($\widehat{Q}_v^i$ is replaced with $\widehat{Q}_v^{\Delta,i}$ when only approximate simulation exists.) For the case where we can simulate the exact behavior of the underlying securities, we make the dependence of $\widehat{\xi}_{r,s}$ on b, M and n_v explicit by denoting the estimator as $\widehat{\xi}_{r,s}(b, M, n_v)$. The convergence theorem requires two additional assumptions.

(B1) There exists $\varepsilon_0 > 0$ such that $F(u)$ is strictly increasing and continuous on $(\xi_{r,s} - \varepsilon_0, \xi_{r,s} + \varepsilon_0)$.

(B2) $P(h(t, S_t) = V(t, S_t)) = 0$ for all $t = 0, \cdots, T-1$

Theorem 3.5: Suppose Assumptions (A1-2) and (B1-2) hold, then for all $s \in (0, T-1]$ and $r \in (0, 1)$,

$$\lim_{b,M,n_v\to\infty} \left\| \widehat{\xi}_{r,s}(b, M, n_v) - \xi_{r,s} \right\| = 0 \tag{3.24}$$

Remark: The assumption (B2) implies that the exercise boundary will be thin in terms of probability measure. Assumption (B1) and (B2) hold for many cases, but further work seems to be required before we can verify them rigorously.

Proof. We will denote $\widehat{\xi}_{r,s}(b, M, n_v)$ as $\widehat{\xi}_{r,s}$ for short. Also, we will denote the randomness associated with the mesh as R. With the boundedness of $\widehat{\xi}_{r,s}$, we only have to show $P[|\widehat{\xi}_{r,s} - \xi_{r,s}| > \varepsilon] = P[\widehat{\xi}_{r,s} - \xi_{r,s} > \varepsilon] + P[\widehat{\xi}_{r,s} - \xi_{r,s} > -\varepsilon] \to 0$ for $\varepsilon \in (0, \varepsilon_0)$, where ε_0 is introduced in Assumption (B1). We will show $P[\widehat{\xi}_{r,s} - \xi_{r,s} > \varepsilon] \to 0$

below and similar arguments could be applied to showing $P[\widehat{\xi}_{r,s} - \xi_{r,s} > -\varepsilon] \to 0$. Let $\delta = F(\xi_{r,s} + \frac{\varepsilon}{2}) - r$ and $z_R = P(\widehat{Q}_v^1 \leq \xi_{r,s} + \varepsilon | R)$. We will denote the expectation conditioning on R as E_R.

$$\begin{aligned}
P[\widehat{\xi}_{r,s} > \xi_{r,s} + \varepsilon] &\leq P[r = F_{n_v}(\widehat{\xi}_{r,s}) \geq F_{n_v}(\xi_{r,s} + \varepsilon)] \\
&\leq P(z_R < r + \delta) + P[|F_{n_v}(\xi_{r,s} + \varepsilon) - z_R| \geq z_R - r \geq \delta] \\
&\leq P(z_R < r + \delta) + \frac{1}{\delta} E E_R |F_{n_v}(\xi_{r,s} + \varepsilon) - z_R| \\
&\leq P(z_R < r + \delta) + \frac{1}{\delta}\sqrt{\frac{2}{n_v}}, \qquad (3.25)
\end{aligned}$$

where the first inequality by the non-decreasing property of F_{n_v}, the third by Chebychev inequality and the last by Cauchy-Schwartz inequality and the definition of F_{n_v}. Now it suffices to show $P(z_R < r + \delta) \to 0$ whenever $b, M \to \infty$ because z_R does not depend on n_v. Since L_1 convergence implies convergence in distribution and $F(\xi_{r,s} + \varepsilon) > r + \delta$, it suffices to show $E|z_R - F(\xi_{r,s} + \varepsilon)| \to 0$. Note $Q_v^1 \equiv Q(\tau^* \wedge s, S^1_{v,\tau^* \wedge s})$ is independent of R thus $F(\xi_{r,s} + \varepsilon) = P(Q_v^1 \leq \xi_{r,s} + \varepsilon) = P(Q_v^1 \leq \xi_{r,s} + \varepsilon | R)$. We have

$$\begin{aligned}
E|z_R - F(\xi_{r,s} + \varepsilon)| &= E|E_R 1(\widehat{Q}_v^1 \leq \xi_{r,s} + \varepsilon) - E_R 1(Q_v^1 \leq \xi_{r,s} + \varepsilon)| \\
&\leq E|1(\widehat{Q}_v^1 \leq \xi_{r,s} + \varepsilon) - 1(Q_v^1 \leq \xi_{r,s} + \varepsilon)| \\
&\leq P(|Q_v^1 - \xi_{r,s} - \varepsilon| \leq \delta_0) + P(|\widehat{Q}_v^1 - Q_v^1| \geq |Q_v^1 - \xi_{r,s} - \varepsilon| > \delta_0) \\
&\leq P(|Q_v^1 - \xi_{r,s} - \varepsilon| \leq \delta_0) + \frac{1}{\delta_0} E|\widehat{Q}_v^1 - Q_v^1|, \qquad (3.26)
\end{aligned}$$

where δ_0 is arbitrary positive number. Since $F(u)$ is continuous at $\xi_{r,s} + \varepsilon$ by assumption (B1), the first term on the right hand side could be arbitrarily small if

δ_0 is sufficiently small. Now it suffices to show $E|\widehat{Q}_v^1 - Q_v^1| \to 0$.

$$
\begin{aligned}
E|\widehat{Q}_v^1 - Q_v^1| &= EE[|\widehat{Q}_v^1 - Q_v^1||S_{v,t}^1; t = 0, \cdots, s] \\
&\leq 2KEP(\widehat{\tau} \neq \tau^*|S_{v,t}^1; t = 0, \cdots, s) + \left\|\widehat{Q}_{b,M}(s, S_{v,s}^1) - Q(s, S_{v,s}^1)\right\| \\
&= 2KEP_v^1(\widehat{\tau} \neq \tau^*) + \left\|\widehat{Q}_{b,M}(s, S_{v,s}^1) - Q(s, S_{v,s}^1)\right\|, \qquad (3.27)
\end{aligned}
$$

where P_v^1 is the probability measure conditioning on $\{S_{v,t}^1; t = 0, \cdots, s\}$. The second term on the right hand side converges to zero by Theorem 3.2. If we can show $P_v^1(\widehat{\tau} \neq \tau^*) \to 0$ a.s., dominated convergence theorem will imply $EP_v^1(\widehat{\tau} \neq \tau^*) \to 0$. Since Assumption (B2) guarantees the exercise boundary is hit with probability zero, we can focus on $\{S_{v,t}^1; t = 0, \cdots, s\}$ that never hits the boundary. Thus there exists $\varepsilon' > 0$ such that $|h(t, S_{v,t}^1) - V(t, S_{v,t}^1)| > \varepsilon'$ for $t = 0, \cdots, s$. This yields

$$
\begin{aligned}
P_v^1(\widehat{\tau} \neq \tau^*) &\leq \sum_{t=0}^{s} P_v^1(\widehat{V}_{b,M}(t, S_{v,t}^1) \leq h(t, S_{v,t}^1) < V(t, S_{v,t}^1)) + \\
&\quad + \sum_{t=0}^{s} P_v^1(V(t, S_{v,t}^1) \leq h(t, S_{v,t}^1) < \widehat{V}_{b,M}(t, S_{v,t}^1)) \\
&\leq \sum_{t=0}^{s} P_v^1(|\widehat{V}_{b,M}(t, S_{v,t}^1) - V(t, S_{v,t}^1)| > \varepsilon') \qquad (3.28)
\end{aligned}
$$

In Theorem 3.2, we indeed have shown $E[|\widehat{V}_{b,M}(t, S_t) - V(t, S_t)||S_t; t = 0, \cdots, T] \to 0$, a.s., $t = 0, \cdots, T$. Since L_1 convergence implies convergence in probability, we have $P_v^1(|\widehat{V}_{b,M}(t, S_{v,t}^1) - V(t, S_{v,t}^1)| > \varepsilon') \to 0$, which implies $P_v^1(\widehat{\tau} \neq \tau^*) \to 0$ a.s. and completes our proof. $\square$

For the case where there only exists a strongly convergent time discrete approximation with step size Δ, we use it to generate n_v independent paths $\{S_{v,t}^{\Delta,i}; i = 1, \cdots, n_v, t = 0, \cdots, s\}$ for each mesh, i.e., for fixed $i = 1, \cdots, n_v$, we generate $S_{v,0}^{\Delta,i}$

given S_0 and generate $S_{v,1}^{\Delta,i}$ given $S_1 = S_{v,0}^{\Delta,i}$, and so on. Since the simulation error could accumulate due to the subsequent approximation, we have to further assume

(A5) The time sequence of discrete approximation $\{S_t^{\Delta}; t = 0, \cdots, s\}$ with step size Δ converges strongly to $\{S_t; t = 0, \cdots, s\}$ from 0 up to s if $\lim_{\Delta \to 0} E|S_t^{\Delta} - S_t| = 0$ for any $t = 1, \cdots, s$.

Then we can calculate $\widehat{Q}_v^{\Delta,i} = \widehat{Q}_{b,M,\Delta}(\widehat{\tau} \wedge s, S_{v,\widehat{\tau}\wedge s}^{\Delta,i})$ as well as $\widehat{\xi}_{r,s}$. By denoting the estimator as $\widehat{\xi}_{r,s}(b, M, n_v, \Delta)$, we have

Theorem 3.6: Suppose (A1-5) and (B1-2) hold, then for all $s \in (0, T-1]$ and $r \in (0, 1)$,

$$\lim_{b,M,n_v,\frac{1}{\Delta} \to \infty} \left\| \widehat{\xi}_{r,s}(b, M, n_v, \Delta) - \xi_{r,s} \right\| = 0 \tag{3.29}$$

Proof. By the existence of a time discrete approximation, we can treat $\{S_{v,t}^{\Delta,1}; t = 0, \cdots, s\}$ as an approximation to $\{S_{v,t}^1; t = 0, \cdots, s\}$ and the latter is the path with exact distribution of $\{S_t; t = 0, \cdots, s\}$ given S_0. Let $Q_v^1 = Q(\tau^* \wedge s, S_{v,\tau^*\wedge s}^1)$ and $\widehat{Q}_v^{\Delta,1} = \widehat{Q}_{b,M,\Delta}(\widehat{\tau} \wedge s, S_{v,\widehat{\tau}\wedge s}^{\Delta,1})$. With similar arguments in the proof of Theorem 3.5, it only suffices to show $E|\widehat{Q}_v^{\Delta,1} - Q_v^1| \to 0$ when $b, M, \frac{1}{\Delta} \to \infty$.

$$\begin{aligned} E|\widehat{Q}_v^{\Delta,1} - Q_v^1| &\leq E|\widehat{Q}_v^{\Delta,1} - Q_v^1|[1(\widehat{\tau} \neq \tau^*) + 1(\widehat{\tau} = \tau^*)] \\ &\leq 2KP(\widehat{\tau} \neq \tau^*) + \sum_{t=0}^{s} \left\| h(t, S_{v,t}^{\Delta,1}) - h(t, S_{v,t}^1) \right\| + \\ &\quad \left\| \widehat{V}_{b,M,\Delta}(s, S_{v,s}^{\Delta,1}) - V(s, S_{v,s}^1) \right\|. \end{aligned} \tag{3.30}$$

Let $D_t^{\delta} = \{x : |h(t,x) - v(t,x)| < \delta\}$ for $t = 1, \cdots, s$. We can see D_t^{δ} converges to the exercise boundary at time t as $\delta \downarrow 0$. Given $\varepsilon > 0$, the property of probability measure and assumption (B2) imply $\lim_{\delta\to 0} P(S_t \in D_t^\delta | S_0) = 0$ for $t = 1, \cdots, s$. Hence, there exists $\delta > 0$ such that $P(S_t \in D_t^\delta | S_0) < \varepsilon$ for all $t = 1, \cdots, s$.

$$
\begin{aligned}
P(\widehat{\tau} \neq \tau^*) &\leq \sum_{t=0}^{s} P[h(t, S_{v,t}^1) \leq V(t, S_{v,t}^1), h(t, S_{v,t}^{\Delta,1}) > \widehat{V}_{b,M,\Delta}(t, S_{v,t}^{\Delta,1})] + \\
&\quad \sum_{t=0}^{s} P[h(t, S_{v,t}^1) > V(t, S_{v,t}^1), h(t, S_{v,t}^{\Delta,1}) \leq \widehat{V}_{b,M,\Delta}(t, S_{v,t}^{\Delta,1})] \\
&\leq \sum_{t=0}^{s} [P(|h(t, S_{v,t}^1) - V(t, S_{v,t}^1)| < \delta) + P(|h(t, S_{v,t}^{\Delta,1} - h(t, S_{v,t}^1)| \geq \frac{\delta}{2}) + \\
&\quad P(|\widehat{V}_{b,M,\Delta}(t, S_{v,t}^{\Delta,1}) - V(t, S_{v,t}^1)| \geq \frac{\delta}{2}) \\
&\leq T\varepsilon + \frac{2}{\delta} \sum_{t=0}^{s} E[|h(t, S_{v,t}^{\Delta,1} - h(t, S_{v,t}^1)| + |\widehat{V}_{b,M,\Delta}(t, S_{v,t}^{\Delta,1}) - V(t, S_{v,t}^1)|] \\
&\leq T\varepsilon + \frac{2}{\delta} \sum_{t=0}^{s} E|h(t, S_{v,t}^{\Delta,1}) - h(t, S_{v,t}^1)| + \frac{2}{\delta} \sum_{t=0}^{s} E|V(t, S_{v,t}^{\Delta,1}) - V(t, S_{v,t}^1)| \\
&\quad \frac{2}{\delta} \sum_{t=0}^{s} E|\widehat{V}_{b,M,\Delta}(t, S_{v,t}^{\Delta,1}) - V(t, S_{v,t}^{\Delta,1})| \\
&= T\varepsilon + I_1 + I_2 + I_3, \qquad (3.31)
\end{aligned}
$$

where the third step is by the definition of δ and Chebychev inequality. Assumption (A5) and continuity of V imply $I_2 \to 0$ when $\Delta \to 0$ and Theorem 3.3 implies $I_3 \to 0$ when $b, M, \frac{1}{\Delta} \to \infty$. So the third term on the right hand side of (3.30) also converges to zero. Now, with (3.30) and (3.31), it is only left to show $E|h(t, S_{v,t}^{\Delta,1}) - h(t, S_{v,t}^1)| \to 0$, for $t = 1, \cdots, s$.

We only need to consider case where $D_t(h) \neq \phi$. For given t and $\varepsilon > 0$, Lemma 3.1 implies there exist A_t', A_t and $r_t > 0$ such that $A_t' \Subset A_t \Subset G_t$, $P(S_{v,t}^1 \notin A_t' | S_0) < \varepsilon)$ and $P(S_{v,t}^1 \in D_{t,r_t}(h) | S_0) < \varepsilon$. Let $\delta_t \in (0, |\overline{A_t'} \cap D_{t,r_t}^c(h) - A_t^c \cup \overline{D}_{t,r_t/2}(h)|)$ such that $h(t,x) - h(t,y) < \varepsilon$ whenever $|x - y| \leq \delta_t$ and $x, y \in A_t \cap D_{t,r_t/2}^c(h)$.

$$
\begin{aligned}
I_1' &\equiv E|h(t, S_{v,t}^{\Delta,1}) - h(t, S_{v,t}^1)| \\
&\leq 2KP(S_{v,t}^1 \notin A_t' \cap D_{t,r_t}(h)) + 2KP(|S_{v,t}^{\Delta,1} - S_{v,t}^1| > \delta_t) + \\
&\quad E|h(t, S_{v,t}^{\Delta,1}) - h(t, S_{v,t}^1)| 1[S_{v,t}^1 \in A_t' \cap D_{t,r_t}(h), |S_{v,t}^{\Delta,1} - S_{v,t}^1| \leq \delta_t] \\
&\leq 4K\varepsilon + \varepsilon + \frac{2K}{\delta_t} E|S_{v,t}^{\Delta,1} - S_{v,t}^1|. \qquad (3.32)
\end{aligned}
$$

Now assumption (A5) implies the limitation of I_1' is no greater than $4K\varepsilon + \varepsilon$. We complete our proof since ε can be arbitrarily small. □

As in section 3.2, we can calculate estimators for each mesh and then the sample mean and sample variance will give us a confidence interval for $\xi_{r,s}$.

3.5. Conclusion

This chapter proposes a new algorithm, weighted stochastic mesh method, for American-style options pricing. This algorithm does not assume any knowledge of the exercise boundary, and requires computational effort that is only polynomial in the problem dimension as well as the number of exercise opportunities. After some adjustment, this algorithm can be applied to pricing path-dependent options and calculating VaR. Asymptotic convergence is guaranteed under fairly mild and easy to verify assumptions. However, further work is necessary to make the algorithm computationally practical.

Chapter 4

Simulation Allocation with Correlated Non-Gaussian Sampling

4.1. Problem Setting

Let k be the number of designs and T be the total number of simulation replications (budget). Under the budget constraint, we will allocate N_i simulation replications to design i so that the probability of correct selection (PCS) is maximized. Here "correct selection" is defined as picking the best design, which we will take as the design having maximum mean. Without loss of generality, we assume design 1 is the best, i.e., $\mu_1 > \mu_i \; \forall i > 1$, where μ_i is mean for design i. Let $\tilde{J}_{im}, m = 1, \ldots, N_i$ represent the mth simulation replication for design i, and $\bar{J}_i = \frac{1}{N_i} \sum_{m=1}^{N_i} \tilde{J}_{im}$ represent the sample average for design i. Then our goal is to maximize $P(\bar{J}_1 - \bar{J}_i > 0, \; i = 2, \ldots, k)$ by determining the values of $N_1, N_2, \ldots, N_k$ subject to $N_1 + N_2 + \ldots + N_k = T$.

We will make following assumptions on the samples:

(A1) $\tilde{J}_{im}$ is independent of $\tilde{J}_{jn}$ when $m \neq n$ for all $1 \leq i, j \leq k$;

(A2) $Ee^{\lambda \tilde{J}_{im}} < \infty$ for $\lambda \in (-\infty, \infty)$;

(A3) $P(\tilde{J}_{im} > \tilde{J}_{1m}) > 0$.

Since the problem is analytically intractable, we will replace the original problem with an approximate one. To simplify notation, we introduce $\{\xi_i, 1 \leq i \leq k\}$, which have the same joint distribution as $\{\tilde{J}_{im} - \mu_i, 1 \leq i \leq k\}$ $\forall m$.

Using Bonferroni inequality and large deviation techniques, we have

$$\begin{aligned} P(\bar{J}_1 - \bar{J}_i > 0,\ i = 2, \ldots, k) &\geq 1 - \sum_{i=2}^{k} P(\bar{J}_i - \bar{J}_1 > 0) \\ &\geq 1 - (k-1) \max_{2 \leq i \leq k} P(\bar{J}_i - \bar{J}_1 > 0) \\ &= 1 - (k-1) \exp\{\max_{2 \leq i \leq k} \ln P(\bar{J}_i - \bar{J}_1 > 0)\} \\ &\geq 1 - (k-1) \exp\{\max_{2 \leq i \leq k} \inf_{\lambda \geq 0} \ln Ee^{\lambda N_i (\bar{J}_i - \bar{J}_1)}\}. \end{aligned} \quad (4.1)$$

For all $1 < i \leq k$, we define β_i as $\mu_i - \mu_1$ and a bi-variable function h_i as:

$$h_i(\lambda, x) \equiv \begin{cases} \beta_i x \lambda + (x-1) \ln E \exp(\lambda \xi_i) + \ln E \exp(\lambda \xi_i - \lambda x \xi_1), & x \geq 1 \\ \beta_i x \lambda + (1-x) \ln E \exp(-\lambda x \xi_1) + x \ln E \exp(\lambda \xi_i - \lambda x \xi_1), & x < 1 \end{cases} . \quad (4.2)$$

If we calculate the rate function on right hand side of (4.1) for $N_i \geq N_1$, we have

$$\begin{aligned} \ln Ee^{\lambda N_i (\bar{J}_i - \bar{J}_1)} &= \ln E \exp\{\lambda (\sum_{m=1}^{N_i} \tilde{J}_{im} - \frac{N_i}{N_1} \sum_{m=1}^{N_1} \tilde{J}_{1m})\} \\ &= \ln E \exp\{\lambda \sum_{m=1}^{N_1} (\tilde{J}_{im} - \frac{N_i}{N_1} \tilde{J}_{1m})\} + \ln E \exp\{\lambda \sum_{m=N_1+1}^{N_i} \tilde{J}_{im}\} \\ &= N_1 \ln E \exp\{\lambda [\xi_i + \mu_i - \frac{N_i}{N_1} (\xi_1 + \mu_1)]\} + (N_i - N_1) \ln E \exp\{\lambda (\xi_i + \mu_i)\} \\ &= N_i \lambda (\mu_i - \mu_1) + (N_i - N_1) \ln E \exp(\lambda \xi_i) + N_1 \ln E \exp(\lambda \xi_i - \lambda \frac{N_i}{N_1} \xi_1) \\ &= N_1 h_i(\lambda, \frac{N_i}{N_1}), \end{aligned}$$

where the second step uses assumption $(A1)$ and the third uses definition of ξ_i. We can verify that $\ln Ee^{\lambda N_i(\bar{J}_i - \bar{J}_1)} = N_1 h_i(\lambda, \frac{N_i}{N_1})$ also holds when $N_i < N_1$. Plugging into (4.1), we have

$$P(\bar{J}_1 - \bar{J}_i > 0,\ i = 2, \ldots, k) \geq 1 - (k-1)\exp\{N_1 \max_{2 \leq i \leq k} \inf_{\lambda \geq 0} h_i(\lambda, \frac{N_i}{N_1})\}.$$

Now we replace the original problem with a new one by maximizing the right hand side, not left hand side, of above inequality. Using notation $H_i(x) = \inf_{\lambda \geq 0} h_i(\lambda, x)$, we write this new problem as follows:

$$\min\{N_1 \max_{2 \leq i \leq k} H_i(\frac{N_i}{N_1}) | \sum_{i=1}^{k} N_i = T,\ N_i \geq 0\}. \tag{4.3}$$

From now on, $\{N_i^*, 1 \leq i \leq k\}$ and $\{\widetilde{N_i}, 1 \leq i \leq k\}$ will represent the optimal solutions to this new problem and the original problem with budget T, respectively. For arbitrary positive solution $\{N_i, 1 \leq i \leq k\}$, which satisfies the budget constraint, we can define $Y_i \equiv N_1 H_i(\frac{N_i}{N_1})$ for all $2 \leq i \leq k$ and $Y \equiv \max_{2 \leq i \leq k} Y_i$. We can also define Y_i^*, Y^*, $\widetilde{Y_i}$ and $\widetilde{Y}$, accordingly. Before we go to the following sections, we will point out two facts:

- The solutions and Y_is are functions of budget T;
- Y_is are always non-positive by noticing that $H_i(x) \leq h_i(0, x) = 0$ for all $x > 0$.

4.2. Solution Procedure to the Approximate Problem

In this section, we provide a solution procedure for problem (4.3). Throughout, we will always treat the variables to be optimized as continuous. For notational convenience, we define the set containing all indices except the best:

$$\Omega = \{2, \dots, k\}.$$

Let us denote σ_{ij} as covariance between paired simulations of design i and design j. We first consider three degenerate cases:

(i) $\sigma_{ii} = 0$ for some $i \in \Omega$.

Obviously we have $N_i^* = 1$ and thus we can eliminate that design i from the problem and reduce the budget by 1.

(ii) $\sigma_{ii}\sigma_{11} = \sigma_{i1}^2 > 0$ for some $i \in \Omega$.

We know the ith design and the best design are completely correlated in this case, we can eliminate design i by letting $N_i^* = 2$ and $N_1 > 2$.

(iii) $\sigma_{11} = 0$.

We have $N_1^* = 1$ accordingly and the problem (4.3) is reduced to

$$\min\{\max_{2 \le i \le k} H_i(N_i) | \sum_{i=2}^{k} N_i = T - 1, \ N_i \ge 0\} \tag{4.4}$$

We can show $H_i(N_i) = N_i H_i(1)$ and rewrite (4.4) as

$$\min_{\sum_{i=2}^{k} N_i = T-1} \max_{i \in \Omega} N_i H_i(1)$$

Since $\max_{i\in\Omega} N_i H_i(1) \geq N_i H_i(1)$ for each $i \in \Omega$, $\max_{i\in\Omega} N_i H_i(1) \geq \sum_{i=2}^{k} \beta_i N_i H_i(1)$, where $\{\beta_i\}$ can be any sequence of positive numbers summing to 1. Letting $\beta_i = \frac{1/H_i(1)}{\sum_{i=2}^{k} 1/H_i(1)}$, we have $\max_{i\in\Omega} N_i H_i(1) \geq \frac{T-1}{\sum_{i=2}^{k} 1/H_i(1)}$, a lower bound that can be achieved only if all $N_i H_i(1)$ are equal. Hence, the optimal allocation will be given by $N_i = (T-1)\beta_i$.

Henceforth, we only consider about non-degenerate cases. Before we propose our major results, we first state a few properties of function $H_i(\cdot)$ in Lemma 4.1.

Lemma 4.1

(i) For any $x > 0$ and $i \in \Omega$, $H_i(x) = h_i(\lambda_i(x), x)$, where $\lambda_i(x)$ is the unique solution to $\frac{\partial}{\partial\lambda} h_i(\lambda, x) = 0$.

(ii) $H_i(x)$ is decreasing on interval $(0, 1)$ for all $i \in \Omega$; $H_i(x)/x$ is increasing on interval $(1, \infty)$ for all $i \in \Omega$.

(iii) $P(\tilde{J}_{im} > \mu_1) \geq 0 \Longleftrightarrow H_i(0+) = 0$; $P(\tilde{J}_{1m} < \mu_i) \geq 0 \Longleftrightarrow H_i(\infty-) > -\infty$.

Proof. Letting $f(\lambda) = \ln E\exp(\lambda\eta)$, where η is any centralized non-trivial random variable, we claim

- $f(0) = 0$, $f'(0) = 0$, $f''(\lambda) > 0$, $f'(\lambda)$ has same sign with λ;
- $\lim_{\lambda\to\infty} f'(\lambda) = a(\eta) \equiv \sup\{a > 0 | P(\eta \geq a) > 0\}$.

The first claim is trivial. For the second one, we notice

$$\lim_{\lambda\to\infty} f'(\lambda) = E\eta e^{\lambda\eta}/Ee^{\lambda\eta} \leq a(\eta).$$

Now it suffices to show the limitation on the left hand side is above a when $P(\eta \geq a) > 0$. Let δ be a positive number. Then there exists $\varepsilon > 0$ such that

$$\frac{P(a-\varepsilon \leq \eta < a)}{P(\eta \geq a)} < \delta.$$

Hence, we have

$$\begin{aligned} f'(\lambda) &= \frac{E\eta e^{\lambda\eta}}{Ee^{\lambda\eta}} \geq \frac{aEe^{\lambda\eta}1(\eta \geq a)}{Ee^{\lambda\eta}1(\eta \geq a) + e^{\lambda a}P(a-\varepsilon \leq \eta < a) + e^{\lambda(a-\varepsilon)}} \\ &= a\big[1 + \frac{e^{\lambda a}P(a-\varepsilon \leq \eta < a) + e^{\lambda(a-\varepsilon)}}{Ee^{\lambda\eta}1(\eta \geq a)}\big]^{-1} \\ &\geq \frac{a}{1+\delta+e^{-\lambda\varepsilon}/P(\eta \geq a)}. \end{aligned}$$

Hence, $\lim_{\lambda\to\infty} f'(\lambda) \geq a/(1+\delta)$. Since δ can be arbitrarily small, we establish the second claim.

Now we will show part (i). For any fixed $x > 0$, we notice that $\frac{\partial}{\partial\lambda}h_i(\lambda, x)$ is increasing by using the first claim above. Also, we have $\frac{\partial}{\partial\lambda}h_i(0, x) = \beta_i x < 0$. So it suffices to show $\lim_{\lambda\to\infty} \frac{\partial}{\partial\lambda}h_i(\lambda, x) > 0$. Using the second claim above, we have

$$\begin{aligned} \lim_{\lambda\to\infty} \frac{\partial}{\partial\lambda}h_i(\lambda, x) &= \begin{cases} \beta_i x + a((x-1)\xi_i) + a(\xi_i - x\xi_1), & x > 1 \\ \\ \beta_i x + xa(-(1-x)\xi_1) + xa(\xi_i - x\xi_1), & x \leq 1 \end{cases} \\ &\geq x(\beta_i + a(\xi_i - \xi_1)), \end{aligned}$$

where notation $a(\cdot)$ is introduced in the second claim above and the inequality uses the fact that $a(\eta_1) + a(\eta_2) \geq a(\eta_1 + \eta_2)$ and $a(c\eta) = ca(\eta)$ for any positive constant c. Since assumption (A3) implies the right hand side of above formula is positive, we establish part (i).

By the definition of H_i and $\lambda_i(x)$, we have $H_i(x) = h_i(\lambda_i(x), x)$ and

$$H_i'(x) = \frac{\partial h_i(\lambda, x)}{\partial x}\Big|_{\lambda=\lambda_i(x)}.$$

Letting x fall in interval $(0,1)$ and using $\frac{\partial h_i(\lambda,x)}{\partial\lambda}|_{\lambda=\lambda_i(x)}=0$, we have

$$\begin{aligned}H_i'(x) &= [\frac{\partial h_i(\lambda,x)}{\partial x}-\frac{\partial h_i(\lambda,x)}{\partial\lambda}\frac{\lambda}{x}]|_{\lambda=\lambda_i(x)}\\ &= \ln Ee^{\lambda_i(x)\xi_i-\lambda_i(x)x\xi_1}-\ln Ee^{-\lambda_i(x)x\xi_1}-\lambda_i(x)\frac{E\xi_i e^{\lambda_i(x)\xi_i-\lambda_i(x)x\xi_1}}{Ee^{\lambda_i(x)\xi_i-\lambda_i(x)x\xi_1}}. \quad (4.5)\end{aligned}$$

Now we define function $G(\lambda)$ as $\ln Ee^{\lambda\xi_i+\eta}$, where η is any centralized non-degenerate random variable. We can show $G''(\lambda)>0$ for any λ. If we fix x and let $\eta=-\lambda_i(x)x\xi_1$, the right hand side of (4.5) can be rewritten as

$$\begin{aligned}H_i'(x) &= G(\lambda_i(x))-G(0)-G'(\lambda_i(x))\lambda_i(x)\\ &= -\frac{\lambda_i^2(x)}{2}G''(\theta\lambda_i(x))\le 0,\end{aligned}$$

where the second equality is by Taylor expansion of $G(0)$ around $G(\lambda_i(x))$ and θ is some positive number less than $\lambda_i(x)$. With this we establish the first claim of part (ii). Appplying similar arguments to $x>1$, we can show $xH_i'(x)-H(x)>0$ and complete part (ii).

Now we prove first claim of part (iii). For $0<x<1$, $\lambda_i(x)$ is the solution to

$$\beta_i-(1-x)\frac{E\xi_1 e^{-\lambda x\xi_1}}{Ee^{-\lambda x\xi_1}}+\frac{E(\xi_i-x\xi_1)e^{\lambda\xi_i-\lambda x\xi_1}}{Ee^{\lambda\xi_i-\lambda x\xi_1}}=0.$$

Suppose $\lambda_i(x)$ is bounded when x is close to 0, it is easy to see that $\lambda_i(0+)$ exists and solves

$$\beta_i+\frac{E\xi_i e^{\lambda\xi_i}}{Ee^{\lambda\xi_i}}=0.$$

We observe that the left hand side of this equation is monotone in λ and converges to $a(\xi_i+\beta_i)$ as $\lambda\to\infty$ by using the claim we proved in the beginning of this proof. Therefore this equation has a finite solution if and only if $P(\tilde{J}_{im}>\mu_1)>0$. We can verify $H_i(0+) = 0$ when the condition holds. Now if we suppose $\lambda_i(x)$ is unbounded in vicinity of 0, it is also easy to see that $\lambda_i(x)x$ has limit at 0 which solves

$$\beta_i + a(\xi_i) - \frac{E\xi_1 e^{-\lambda\xi_1}}{Ee^{-\lambda\xi_1}} = 0.$$

Similarly, this equation has solution if and only if $P(\tilde{J}_{im} > \mu_1) \leq 0$. We can verify $H_i(0+) = 0$ when $P(\tilde{J}_{im} > \mu_1) = 0$ whereas $H_i(0+) < 0$ when $P(\tilde{J}_{im} > \mu_1) < 0$. This establishes the first claim of part (iii). We can complete part (iii) by applying similar arguments to $x \to \infty$. $\square$

Letting $x_i = \frac{N_i}{N_1}$ for $i \in \Omega$, we can rewrite the budget constraint as $N_1(1 + \sum_{i\in\Omega} x_i) = T$, and problem (4.3) as:

$$\{\min \frac{\max_{i\in\Omega} H_i(x_i)}{1 + \sum_{i\in\Omega} x_i} | x_i > 0\}. \tag{4.6}$$

We observe that the optimal value of x_i may be zero when $H_i(0+) < 0$ whereas it may be infinity when $H_i(\infty-) = -\infty$. According to Lemma 4.1(iii), we can exclude such extreme cases by assuming

(A4) $P(\tilde{J}_{1m} < \mu_i) > 0$ and $P(\tilde{J}_{im} < \mu_1) > 0$ for all $i \in \Omega$.

Now we can derive some properties of the optimal solution $\{N_i^*\}$. First, we notice that all $H_i(\frac{N_i^*}{N_1^*}) = H_i(x_i^*)$ have to be equal. Suppose this is not true, assumption (A4) implies $x_i^* > 0$ for all $i \in \Omega$. Therefore we can always improve the objective value by reducing those x_i^* whose H_i values do not equate $\max_{i\in\Omega} H_i(x_i^*)$. Second, if we let the $z^* = H_i(\frac{N_i^*}{N_1^*})$, we have $\frac{N_i^*}{N_1^*} = H_i^{-1}(z^*)$, where $H_i^{-1}(z) \equiv \inf\{x > 0 | H_i(x) = z\}$. If this is not true, we can only have $\frac{N_i^*}{N_1^*} > H_i^{-1}(z^*)$. Then the objective value can be improved by replacing $\frac{N_i^*}{N_1^*}$ with $H_i^{-1}(z^*)$.

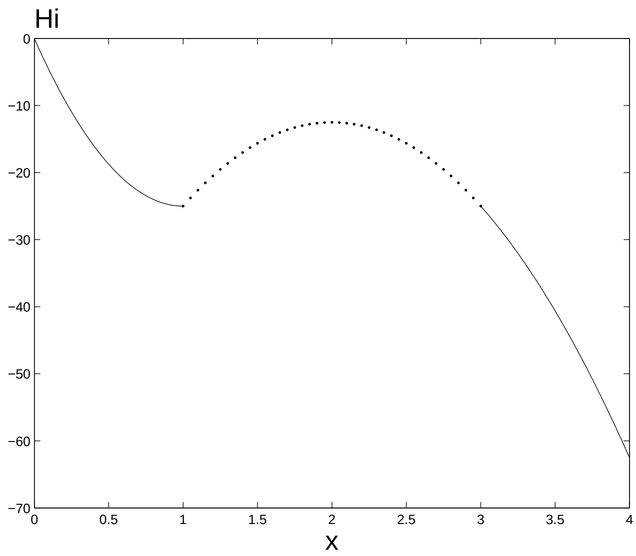

Figure 4.1: Function $H_i(x)$

Theorem 4.2 For any optimal solution $\{N_i^*\}$, there exists z^* such that $\frac{N_i^*}{N_1^*} = H_i^{-1}(z^*),\ \forall i \in \Omega$.

H_i^{-1} can be discontinuous when H_i is not monotone. In this situation, N_i/N_1 only can take values in a disconnected set. For example, if we assume $H_i(x)$ has S shape as in figure 1, N_i/N_1 can only take values on $(0,1) \cup (3,\infty)$. We notice from the figure that $H_i(\cdot)$ is always decreasing on this disconnected set. Actually Theorem 4.2 implies this is always true. More precisely, $H_i'(\frac{N_i^*}{N_1^*}) \leq 0$ when $\frac{N_i^*}{N_1^*} \neq 1$. Suppose this is not true, $H_i(\cdot)$ is increasing locally and there exists $0 < \widehat{x} < \frac{N_i^*}{N_1^*}$ such that $H_i(\widehat{x}) < z^* = H_i(\frac{N_i^*}{N_1^*}) < 0 = H_i(0)$. Continuity of H_i implies $H_i(x) = z^*$ has at least one root in between 0 and $\widehat{x}$, a contradiction with $\frac{N_i^*}{N_1^*} = H_i^{-1}(z^*)$. So we can write the following corollary, where part (ii) is straightforward using Lemma 4.1(ii):

Corollary 4.3

(i) $H_i'(\frac{N_i^*}{N_1^*}) \leq 0$ as $\frac{N_i^*}{N_1^*} \neq 1$, where H_i' is the first-order derivative of H_i;

(ii) $z^* \geq H_i(1) \Longleftrightarrow N_i^* \leq N_1^*$.

Let $H_i^b \equiv \inf_{x \geq 0} H_i(x)$. Obviously $z^* \geq H^b \equiv \max_{i \in \Omega} H_i^b$. Now we can rewrite problem (4.6) as

$$\max\{F(z)|\ H^b \leq z \leq 0\}, \text{ where } F(z) = \frac{1 + \sum\limits_{i=2}^{k} H_i^{-1}(z)}{z}. \tag{4.7}$$

Once we find optimizer of this problem, z^*, we can easily get $\{N_i^*\}$ via:

$$N_1^* = \frac{T}{1 + \sum_{i \in \Omega} H_i^{-1}(z^*)}, \text{ and } N_i^* = N_1^* H_i^{-1}(z^*). \tag{4.8}$$

Simple algebra yields

$$M(z) \equiv z^2 F'(z) = \sum_{i=2}^{k} \frac{z}{H_i'(H_i^{-1}(z))} - 1 - \sum_{i=1}^{k} H_i^{-1}(z). \tag{4.9}$$

If $F(\cdot)$ is smooth, we only need to locate all the roots of $M(z) = 0$ in interval $[H^b, 0]$. But $F(\cdot)$ is not differentiable at $H_i(1)$ when $\tilde{J}_{im}$ and $\tilde{J}_{1m}$ are not independent. Also we note that $F(\cdot)$ even may not be continuous when some $H_i(\cdot)$ is not monotone. So $M(z) = 0$ may not be satisfied or even well defined for each optimizer, z^*, to problem (4.7). Theorem 4.4 gives a necessary condition for z^*.

Theorem 4.4 Let $S_1 = \{H_i(1)|i \in \Omega\} \cap [H^b, 0)$ and $S_2 = \{z \in (H^b, 0)|F(\cdot)$ is discontinuous at $z\}$. If z^* is optimizer of problem (4.7), one of the following three conditions has to be satisfied:

(i) $F'(z^*) = 0$, $z^* \notin S_1$, $H^b < z^* < 0$;

(ii) $F'(z^*+) \leq 0$, $F'(z^*-) \geq 0$, $z^* \in S_1 \setminus S_2 \setminus \{H^b\}$;

(iii) $F'(z^*+) \leq 0$, $z^* \in S_1 \cap S_2$.

Proof. We first notice that $F(0-) = -\infty$, so 0 cannot be an optimizer either. Now we look at point H^b. If $H_i^{-1}(H^b) = \infty$ for some $i \in \Omega$, we have $F(H^b+) = -\infty$ and H^b can not be optimal. Now we assume $H_i^{-1}(H^b)$ is finite for all $i \in \Omega$ and $H^b \notin S_1$. By definition, H^b has to be minimum of some $H_i(\cdot)$. So $H_i'(H_i^{-1}(H^b)) = 0$ for some $i \in \Omega$. Then we have $F'(H^b+) = \infty$ because $\frac{z}{H_i'(H_i^{-1}(z))} - H_i^{-1}(z) \to \infty$ as $z \downarrow H^b$. Therefore, H^b cannot optimal as long as $H^b \notin S_1$.

We now assume $\omega \in S_2 \setminus S_1$. We observe that $F(\cdot)$ is right continuous at ω and $H_i'(H_i^{-1}(\omega)) = 0$ for some $i \in \Omega$. So we have $F'(\omega+) = \infty$, which means ω can not be a local optimizer to $F(\cdot)$. Now if we assume $z^* \in (H^b, 0) \setminus S_1$, we have $z^* \notin S_2$, i.e., $F(\cdot)$ is continuous and differentiable at z^*. So $F'(z^*) = 0$ has to be true. This is the first condition in the theorem. For the case $z^* \in S_1 \setminus S_2 \setminus \{H^b\}$, we know $F'(\cdot)$ has both left limit and right limit at z^*, so $F'(z^*+) \leq 0$ and $F'(z^*-) \geq 0$ have to hold. This is the second condition in the theorem. For the case $z^* \in S_1 \cap S_2$, we know $F'(\cdot)$ only has right limit, so $F'(z^*+) \leq 0$, which is the third condition in the theorem. $\square$

However, it still can be very hard to locate the global optimizer of problem (4.7) even with theorem 4.4. The first difficulty is to calculate $F(z)$, or more precisely, $H_i^{-1}(z)$, for any given $H_i^b \leq z < H_i(1)$. Obviously it is trivial to calculate $H_i^{-1}(z)$ via some numerical processes, say, bisectional method, as long as we know the monotonicity of $H_i(\cdot)$. Since it is hard to obtain knowledge on such monotonicity property, we will turn to a specified process. Theorem 4.5 proposes this process and guarantees it will yield $H_i^{-1}(z)$ for any $z < H_i(1)$. Here we implicitly extend definition of $H_i^{-1}(z)$ by letting $H_i^{-1}(z) = \infty$ as $z < H_i(1)$.

Theorem 4.5 Define sequence $\{x_1 = 1\ ; x_{n+1} = \frac{zx_n}{H_i(x_n)},\ \forall n \geq 1\}$ for any $i \in \Omega$ and $z < H_i(1)$, then x_n increasingly converges to $H_i^{-1}(z)$ as $n \to \infty$.

Proof. Note from Lemma 4.1(ii) that $H_i(x)/x$ is increasing in x. First we show $\{x_n\}$ is an increasing sequence by induction: $z < H_i(1) \Longrightarrow x_2 = \frac{z}{H_i(1)} > 1 = x_1$; $x_n > x_{n-1} \Longrightarrow x_{n+1} = \frac{zx_n}{H_i(x_n)} > \frac{zx_{n-1}}{H_i(x_{n-1})} = x_n$. Now we consider case $z \geq H_i^b$ and denote $x^* = H_i^{-1}(z) < \infty$. We can also show $x_n < x^*$ by induction: $x_1 < x^*$; $x_{n-1} < x^* \Longrightarrow x_n = \frac{zx_{n-1}}{H_i(x_{n-1})} < \frac{zx^*}{H_i(x^*)} = x^*$. Since $\{x_n\}$ is a bounded monotone sequence, it has a finite limitation, which has to be x^*. Now we handle the case $z < H_i^b$ by contradiction. Suppose x_n is bounded then it has a limit which satisfies equation $x = \frac{zx}{H_i(x)}$, i.e., $z = H_i(x)$. This contradicts with $z < H_i^b$. $\square$

Remark: Not knowing H^b in general means we can not tell if sequence $\{x_n\}$ will converge to a finite value or diverge for arbitrary $z < H_i(1)$. A solution is to define a good stopping rule which stops the iterative process whenever it goes across a bound and at the meantime does not risk stopping a slowly convergent sequence too early. Suppose we already get a value $F_1 = F(z_1)$ and we are testing if $F_2 = F(z_2)$ is a better value. Denote set I as $\{i \in \Omega | H_i(1) > z_2\}$. If I is empty, then $H_i^{-1}(z_2) \in (0, 1],\ \forall i \in \Omega$, and we can easily calculate $H_i^{-1}(z_2)$ since we know $H_i(\cdot)$ is decreasing on $(0, 1]$. Without loss of generality, we assume $I = \{2, 3, \cdots, |I| + 1\}$ and that we already calculate $H_i^{-1}(z_2)$ for all $1 < i < n$, where $2 \leq n \leq |I| + 1$. Letting the bound $BND = F_1 z_2 - \sum_{i=2}^{n-1} H_i^{-1}(z_2)$, we claim that $z^* > z_2$ given the iterative process crosses this bound. Obviously we only need to take care of the case where $z_2 > H_n^b$ and $H_i^{-1}(z_2) \geq BND$. Let z_3 fall in $(H_n^b, z_2]$ and notice that Lemma 4.1(ii) implies $H_i^{-1}(z)/z$ is increasing. We have

$$\begin{aligned} F(z_3) &= \frac{1+\sum_{i\in\Omega} H_i^{-1}(z_3)}{z_3} < \sum_{i\in\Omega} \frac{H_i^{-1}(z_3)}{z_3} \leq \sum_{i=2}^{n} \frac{H_i^{-1}(z_3)}{z_3} \\ &\leq \sum_{i=2}^{n} \frac{H_i^{-1}(z_2)}{z_2} \leq \sum_{i=2}^{n-1} \frac{H_i^{-1}(z_2)}{z_2} + \frac{BND}{z_2} = F_1. \end{aligned}$$

Now we propose a procedure to numerically solve the optimization problem (4.3)/(4.6)/(4.7):

- **Step 1.** Calculate $\{H_i(1),\ i = 2, \cdots, k\}$ and define $H_1(1) = 0$. Arrange all the elements in descending order so that we can write $H_{(1)}(1) = 0 > H_{(2)}(1) > \cdots > H_{(k)}(1) > -\infty$.

- **Step 2.** Fix a large number M and calculate $\Delta_i = \frac{H_{(i)}(1)-H_{(i+1)}(1)}{M}$ for all $1 \leq i < k$. Also define sequences $\{z_j, j = 0, \cdots, M(k-1)\}$ with $z_0 = 0$ and $z_j = z_{j-1} - \Delta_{\lceil j/M \rceil}$ for any $1 \leq j \leq M(k-1)$, where $\lceil j/M \rceil$ is the smallest integer no less than j/M.

- **Step 3.** Let $j = 0$, $F_0 = -\infty$, $F_0^- = -\infty$, $z^* = 0$ and $F^* = -\infty$.

- **Step 4.** Let $j = j + 1$. If $j <= M$, go to step 6. If $j > M(k-1)$, let $\Delta = \min_{1\leq i<k} \Delta_i$ and $z_j = z_{j-1} - \Delta$.

- **Step 5.** Let $n = \lceil \frac{j}{M} \rceil$, $BND_1 = F^* z_j$, and $\alpha = 1$.

  - **Step 5.1** Let $\alpha = \alpha + 1$. Go to step 6 if $\alpha > n$; otherwise go to step 5.2.

- **Step 5.2** Use process described in Theorem 2(v) to locate $V_\alpha = H^{-1}_{(\alpha)}(z_j)$. If $V_\alpha >= BND_\alpha$, go to the step 7; otherwise, calculate $BND_{\alpha+1} = BND_\alpha - V_\alpha$ and go to step 5.1.

- **Step 6.** Calculate $F_j = F(z_j)$, $F_j^+ = F'(z_j+)$, and $F_j^- = F'(z_j-)$. If $F_j > F^*$, replace F^* and z^* with F_j and z_j. If $F_j^+ > 0$ and $F_{j-1}^- < 0$, locate $\overline{z}$ such that $z_j < \overline{z} < z_{j-1}$ and $F'(\overline{z}) = 0$, then replace F^* and z^* with $F(\overline{z})$ and $\overline{z}$ when $F^* < F(\overline{z})$. Go to step 4.

- **Step 7.** Calculate N_1^* from z^* via (4.8) and $N_i^* = N_1^* H_i^{-1}(z^*)$ for all $i \in \Omega$.

- **Return** $\{N_i^*\}$.

To explain why this procedure works, we look at the conditions in Theorem 4.4. The case where either condition (ii) or (iii) holds is easy to solve by calculating $F(z)$ of all $z \in S_1$ with process described in Theorem 4.5, and comparing those F values. But when condition (i) holds, we need to solve a nonlinear equation $F'(z) = 0$ of which we do not know about convexity or monotonicity. The procedure here first calculates objective values at lattice points and then seeks zeros of $F'(z)$ when $F'(\cdot)$ changes sign over two adjacent points. Therefore, the solution z^* from this procedure will be the true solution only if any two adjacent points are sufficiently close to each other, or M introduced at step 2 is sufficiently large.

The procedure we propose here is for general cases where we do not know about convexity or monotonicity of $H_i(\cdot)$. Now we discuss some special cases where H_i is always decreasing and convex. We recall set $S_1 = \{H_i(1)|i \in \Omega\} \cap [H^b, 0)$ introduced in Theorem 4.4 and let $s_j,\ 1 \le j \le |S_1|$ be jth largest element of set S_1. We further write $s_1 = 0$ and $s_{|S_1|+1} = H^b$. Then we have $0 = s_1 > s_2 > \cdots > s_{|S_1|+1} = H^b$. For these special cases, we can show, in Theorem 4.6, $F(\cdot)$ of problem (4.7) is smooth and $F'(\cdot)$ has at most one zero on interval $(s_{j+1}, s_j),\ \forall\ 0 \le j \le |S_1|$.

Theorem 4.6 Let us define $F_i(u,v)$ as

$$F_i(u,v) \equiv \frac{E\xi_i\xi_1 e^{u\xi_i - v\xi_1}}{Ee^{u\xi_i - v\xi_1}} - \frac{E\xi_i e^{u\xi_i - v\xi_1}}{Ee^{u\xi_i - v\xi_1}} \frac{E\xi_1 e^{u\xi_i - v\xi_1}}{Ee^{u\xi_i - v\xi_1}}.$$

Then H_i is decreasing on $(0,\infty)$ and (4.9) has at most one root in interval $[s_{j+1}, s_j]$ for any $0 \le j \le |S_1|$, if either

(i) $\tilde{J}_{im}$ is independent of $\tilde{J}_{1m}$ for all $i \in \Omega$, or

(ii) $\tilde{J}_{im}$ and $\tilde{J}_{1m}$ are negatively correlated for all $i \in \Omega$, $\frac{\partial F_i(u,v)}{\partial u} \le 0$ and $\frac{\partial F_i(u,v)}{\partial v} \le 0$ for all $u, v \ge 0$ and $i \in \Omega$.

Remark:

1. H_i^{-1} has no discontinuities when H_i is decreasing on $(0,\infty)$. Hence $|S_1| \le k$ for this case. Especially $|S_1| = 1$ for the independent case in (i).

2. To give sense on the assumption made in (ii) on monotonicity of F_i, we can consider the case where the copula between the best design and any of the other designs is linear, i.e., $\xi_i - \rho_i\xi_1$ is independent of ξ_1 for some constant ρ_i. Then the monotonicity assumption in (ii) is equivalent to $f_1'''(\lambda) \le 0$ when $\lambda \le 0$, where $f_1(\lambda) = \ln(Ee^{\lambda\xi_1})$.

Proof. When condition (i) holds, it is easy to show that H_i is always decreasing. Now we assume condition (ii) holds. Since we already show H_i is decreasing on $(0,1)$ in Lemma 4.1, we only have to look at its monotonicity on interval $(1,\infty)$.

Letting $x > 1$ and using $\frac{\partial}{\partial\lambda}h_i(\lambda_i(x),x) = 0$, we have

$$\begin{aligned}
H_i'(x) &= \partial_x h_i(\lambda,x)\big|_{\lambda=\lambda_i(x)} = [\partial_x h_i(\lambda,x) - \frac{\lambda}{x}\partial_\lambda h_i(\lambda,x)]\big|_{\lambda=\lambda_i(x)} \\
&= \big[\big(\ln Ee^{\lambda\xi_i} - \lambda\frac{E\xi_i e^{\lambda\xi_i}}{Ee^{\lambda\xi_i}}\big) + \frac{\lambda}{x}\big(\frac{E\xi_i e^{\lambda\xi_i}}{Ee^{\lambda\xi_i}} - \frac{E\xi_i e^{\lambda\xi_i-\lambda x\xi_1}}{Ee^{\lambda\xi_i-\lambda x\xi_1}}\big)\big]\big|_{\lambda=\lambda_i(x)} \\
&\le \frac{\lambda}{x}\big(\frac{E\xi_i e^{\lambda\xi_i}}{Ee^{\lambda\xi_i}} - \frac{E\xi_i e^{\lambda\xi_i-\lambda x\xi_1}}{Ee^{\lambda\xi_i-\lambda x\xi_1}}\big)\big|_{\lambda=\lambda_i(x)} \\
&= \lambda^2 F_i(\lambda,\lambda x\theta_1)\big|_{\lambda=\lambda_i(x)}, \qquad (4.10)
\end{aligned}$$

where $\theta_1 \in [0,1]$. The inequality uses convexity of $\ln Ee^{\lambda\xi_i}$, and the last step uses mean value theorem. Now we notice that condition (ii) actually implies for all $u, v \ge 0$, we have $F_i(u,v) \le F_i(0,0) = E\xi_i\xi_1 \le 0$. Combining this with (4.10), we establish H_i's monotonicity.

Now we show that (4.9) has at most one root when either of the two conditions holds. It suffices to show $M(z)$ is increasing, or $M'(z) > 0$. Simple algebra yields

$$M'(z) = \sum_{i=2}^{k} \frac{zH_i''(H_i^{-1}(z))}{(H_i'(H_i^{-1}(z)))^3}.$$

Since $z < 0$ and $H_i'(H_i^{-1}(z)) < 0$, it suffices to show $H_i''(x) \ge 0$ for all $i \in \Omega$ and $x > 0$. Using $\frac{\partial}{\partial\lambda}h_i(\lambda_i(x),x) = 0$ again, we have

$$H_i''(x) = \big[\partial_{xx}h_i(\lambda,x) - \frac{(\partial_{\lambda x}h_i(\lambda,x))^2}{\partial_{\lambda\lambda}h_i(\lambda,x)}\big]\big|_{\lambda=\lambda_i(x)}.$$

It only requires us to look at the sign of $G(\lambda,x) \equiv \partial_{xx}h_i\partial_{\lambda\lambda}h_i - (\partial_{\lambda x}h_i)^2$. Now we introduce notation $\langle X\rangle_{u,v}$ and $\langle X,Y\rangle_{u,v}$ for $u,v \ge 0$ and any random variable X

and Y as

$$\langle X\rangle_{u,v} \equiv \frac{EXe^{u\xi_i - v\xi_1}}{Ee^{u\xi_i - v\xi_1}} \text{ and } \langle X, Y\rangle_{u,v} = \langle XY\rangle_{u,v} - \langle X\rangle_{u,v}\langle Y\rangle_{u,v}.$$

Then we have

$$\partial_{\lambda\lambda} h_i = \begin{cases} (x-1)\langle \xi_i, \xi_i\rangle_{\lambda,0} + \langle \xi_i - x\xi_1, \xi_i - x\xi_1\rangle_{\lambda,\lambda x} & x > 1 \\ x^2(1-x)\langle \xi_1, \xi_1\rangle_{0,\lambda x} + x\langle \xi_i - x\xi_1, \xi_i - x\xi_1\rangle_{\lambda,\lambda x} & x < 1 \end{cases};$$

$$\partial_{xx} h_i = \begin{cases} \lambda^2\langle \xi_1, \xi_1\rangle_{\lambda,\lambda x} & x > 1 \\ 2\lambda\langle \xi_1\rangle_{0,\lambda x} + \lambda^2(1-x)\langle \xi_1, \xi_1\rangle_{0,\lambda x} - 2\lambda\langle \xi_1\rangle_{\lambda,\lambda x} + x\lambda^2\langle \xi_1, \xi_1\rangle_{\lambda,\lambda x} & x < 1 \end{cases};$$

$$\partial_{\lambda x} h_i = \begin{cases} \frac{1}{x}(\langle \xi_i\rangle_{\lambda,0} - \langle \xi_i\rangle_{\lambda,\lambda x}) - \lambda\langle \xi_i - x\xi_1, \xi_1\rangle_{\lambda,\lambda x} & x > 1 \\ x\langle \xi_1\rangle_{0,\lambda x} + \lambda x(1-x)\langle \xi_1, \xi_1\rangle_{0,\lambda x} - \lambda x\langle \xi_i - x\xi_1, \xi_1\rangle_{\lambda,\lambda x} & x < 1 \end{cases}.$$

Suppose condition (i) holds, we have

$$G(\lambda, x) = x\lambda^2\langle \xi_1, \xi_1\rangle_{\lambda,\lambda x}\langle \xi_i, \xi_i\rangle_{\lambda,0} > 0.$$

We now assume condition (ii) holds and check the case where $x > 1$. We have

$$\begin{aligned} G(\lambda, x) &\geq \lambda^2\langle \xi_i - x\xi_1, \xi_i - x\xi_1\rangle_{\lambda,\lambda x}\langle \xi_1, \xi_1\rangle_{\lambda,\lambda x} - \lambda^2\langle \xi_i - x\xi_1, \xi_1\rangle_{\lambda,\lambda x} \\ &\quad - \frac{1}{x}(\langle \xi_i\rangle_{\lambda,0} - \langle \xi_i\rangle_{\lambda,\lambda x})\big[\frac{1}{x}(\langle \xi_i\rangle_{\lambda,0} - \langle \xi_i\rangle_{\lambda,\lambda x}) - 2\lambda\langle \xi_i - x\xi_1, \xi_1\rangle_{\lambda,\lambda x}\big] \\ &\geq -\frac{1}{x}(\langle \xi_i\rangle_{\lambda,0} - \langle \xi_i\rangle_{\lambda,\lambda x})\big[\frac{1}{x}(\langle \xi_i\rangle_{\lambda,0} - \langle \xi_i\rangle_{\lambda,\lambda x}) - 2\lambda\langle \xi_i - x\xi_1, \xi_1\rangle_{\lambda,\lambda x}\big] \\ &= -\lambda F_i(\lambda, \lambda x\theta_2)\big[\frac{1}{x}(\langle \xi_i\rangle_{\lambda,0} - \langle \xi_i\rangle_{\lambda,\lambda x}) - 2\lambda\langle \xi_i - x\xi_1, \xi_1\rangle_{\lambda,\lambda x}\big], \end{aligned} \tag{4.11}$$

where the second step uses Cauchy-Schwartz inequality and the third step uses mean value theorem. Using a Taylor expansion, we have

$$\langle \xi_i\rangle_{\lambda,0} = \langle \xi_i\rangle_{\lambda,\lambda x} + \lambda x\langle \xi_i, \xi_1\rangle_{\lambda,\lambda x} - \frac{\lambda^2 x^2}{2}\partial_v F_i(\lambda, \lambda x\theta_3).$$

Plugging into (4.11) and noticing that $\langle \xi_i, \xi_1 \rangle_{\lambda, \lambda x} = F_i(\lambda, \lambda x) \leq 0$, we can show $H_i''(x) > 0$ when $x > 1$. Since the case of $x < 1$ can be similarly proved, we establish part (ii). $\square$

4.3. Asymptotic Analysis

Since the procedure described in previous section gives a solution to an approximate problem instead of our original problem, we need to know the magnitude of gap between the optimal solution to the approximate problem and to the original problem. We recall that $\{N_i^*\}$ and $\{\widetilde{N_i}\}$ represent the optimal solutions to the approximate problem and the original problem, respectively. In this section, we will show the gap between the two solutions is in the order of $o(T)$, the budget. Furthermore, we make the order more precise for the case of Gaussian distribution. Throughput this section, we assume that (A1-4) hold and exclude degenerate cases described in the beginning of previous section. The argument before Theorem 4.2 actually tells us $\{N_i^*\}$ is linear in budget T, or $\{N_i^*/T\}$ are constants independent of T. Then we can show

Lemma 4.7 Let $\{N_i, 1 \leq i \leq k\}$ be a solution such that $a_i \equiv \lim_{T\to\infty} \frac{N_i}{T}$ exists for all $1 \leq i \leq k$. Also, we assume $(a_1, \cdots, a_k) \neq (\frac{N_1^*}{T}, \cdots, \frac{N_k^*}{T})$. Then if either (i) $a_i > 0$ for all $1 \leq i \leq k$; or (ii) $P(\tilde{J}_{1m} < \mu_i) > 0$ and $P(\tilde{J}_{im} > \mu_1) > 0$ for all $i \in \Omega$, we have

$$\lim_{T\to\infty} \ln(\frac{1-P}{1-P^*})/T > 0, \tag{4.12}$$

where P and P^* represent the PCS associated with $\{N_i\}$ and $\{N_i^*\}$, respectively.

Proof. First we will show $\lim_{T\to\infty} \frac{P(\bar{J}_i-\bar{J}_1\geq 0)}{T} = a_1 H_i(\frac{a_i}{a_1})$ for all $i \in \Omega$ when $a_i > 0$ and $a_1 > 0$. Note that, with large deviation technique,

$$\ln P(\bar{J}_i - \bar{J}_1 \geq 0) \leq \inf_{\lambda\geq 0} \ln E e^{\lambda N_i(\bar{J}_i-\bar{J}_1)} = N_1 H_i(\frac{N_i}{N_1}).$$

We only need to show $\liminf_{T\to\infty} \frac{\ln P(\bar{J}_i-\bar{J}_1\geq 0)}{T} \geq \frac{N_1}{T} H_i(\frac{N_i}{N_1})$. Now we assume $N_i \geq N_1$ and the case of $N_i < N_1$ can be similarly shown. Then we have

$$N_i(\bar{J}_i - \bar{J}_1) = \sum_{m=1}^{N_i} \tilde{J}_{im} - N_i \sum_{m=1}^{N_1} \frac{\tilde{J}_{1m}}{N_1} = \sum_{m=1}^{N_1} (\tilde{J}_{im} - \alpha_i \tilde{J}_{1m}) + \sum_{m=N_1+1}^{N_i} \tilde{J}_{im},$$

where $\alpha_i = \frac{N_i}{N_1}$. Since we already show, in the proof of lemma 4.1, $\lim_{\lambda\to\infty} h_i(\lambda, x) > 0$ for any $x > 0$. So we can fix a positive number δ such that $\delta < \lim_{\lambda\to\infty} h_i(\lambda, \frac{a_i}{a_1})$. Let τ and ν represent the probability measures of $J_{im} - \alpha_i J_{1m} - \delta$ and $J_{im} - \delta$, respectively. We define two new probability measures as follows:

$$\frac{d\widehat{\tau}}{d\tau}(x) = e^{\lambda x}/E_\tau e^{\lambda x}, \text{ and } \frac{d\widehat{\nu}}{d\nu}(x) = e^{\lambda x}/E_\nu e^{\lambda x},$$

where λ will be specified later on. Letting $\varepsilon < \delta$ be some positive number, we have

$$\begin{aligned}
P(\bar{J}_i - \bar{J}_1 \geq 0) &\geq P(|\bar{J}_i - \bar{J}_1 - \delta| \leq \varepsilon) \\
&= \int_{|\sum_{m=1}^{N_i} x_m| \leq N_i\varepsilon} \tau(dx_1)\cdots\tau(dx_{N_1})\nu(dx_{N_1+1})\cdots\nu(dx_{N_i}) \\
&\geq e^{-\lambda N_i \varepsilon} \int_{|\sum_{m=1}^{N_i} x_m| \leq N_i\varepsilon} e^{\lambda \sum_{m=1}^{N_i} x_m} \tau(dx_1)\cdots\tau(dx_{N_1})\nu(dx_{N_1+1})\cdots\nu(dx_{N_i}) \\
&= e^{-(\varepsilon+\delta)\lambda N_i + N_1 h_i(\lambda,\ \alpha_i)} \int_{|\sum_{m=1}^{N_i} x_m| \leq N_i\varepsilon} \widehat{\tau}(dx_1)\cdots\widehat{\tau}(dx_{N_1})\widehat{\nu}(dx_{N_1+1})\cdots\widehat{\nu}(dx_{N_i}) \\
&= e^{-(\varepsilon+\delta)\lambda N_i + N_1 h_i(\lambda,\ \alpha_i)} P(\Big|\sum_{m=1}^{N_i} X_m\Big| \leq N_i\varepsilon), \qquad (4.13)
\end{aligned}$$

where $\{X_m, 1 \leq m \leq N_1\}$ are i.i.d. random variables of law $\widehat{\tau}$ and $\{X_m, N_1 + 1 \leq m \leq N_i\}$ are i.i.d. random variables of law $\widehat{\nu}$. Simple algebra yields

$$E\Big[\frac{\sum_{m=1}^{N_i} X_m}{N_i}\Big] = \frac{1}{\alpha_i} E_{\widehat{\tau}}[X] + (1 - \frac{1}{\alpha_i}) E_{\widehat{\nu}}[X] = \frac{1}{\alpha_i}(\frac{\partial h_i}{\partial \lambda}(\lambda, \alpha_i) - \delta). \qquad (4.14)$$

Using similar argument in the proof of lemma 4.1(i) and the fact that $\delta < \lim_{\lambda\to\infty} h_i(\lambda, \frac{a_i}{a_1})$, we can show there exists unique solution λ_i^δ to equation $\frac{\partial h_i}{\partial\lambda}(\lambda, \frac{a_i}{a_1}) = \delta$. If we specify $\lambda = \lambda_i^\delta$ above, $E\big[\frac{\sum_{m=1}^{N_i} X_m}{N_1}\big] \to 0$ will be implied by the continuity of $\frac{\partial h_i}{\partial\lambda}$ in both variables. It is easy to check that the variance of X_m is uniformly bounded, which, combining with Chebychev inequality, yields

$$\lim_{T\to\infty} P(|\sum_{m=1}^{N_i} X_m| \le N_i\varepsilon) = 1. \tag{4.15}$$

Combining (4.13) with (4.15), we have

$$\lim\inf_{T\to\infty} \frac{\ln P(\bar{J}_i - \bar{J}_1 \ge 0)}{T} \ge -(\varepsilon+\delta)\lambda_i^\delta a_i + a_1 h_i(\lambda_i^\delta, \frac{a_i}{a_1}).$$

By letting $\varepsilon \to 0$ first and $\delta \to 0$ second, we prove

$$\lim_{T\to\infty} \frac{\ln P(\bar{J}_i - \bar{J}_1 \ge 0)}{T} = \frac{N_1}{T} H_i(\frac{N_i}{N_1}).$$

Using inequality $\max_{2\le i\le k} P(\bar{J}_i - \bar{J}_1 \ge 0) \le 1 - P \le (k-1)\max_{2\le i\le k} P(\bar{J}_i - \bar{J}_1 \ge 0)$, we have

$$\lim_{T\to\infty} \frac{\ln(1-P)}{T} = \frac{Y}{T}, \text{ and } \lim_{T\to\infty} \frac{\ln(1-P^*)}{T} = \frac{Y^*}{T}.$$

Now we can complete part (i) by noticing that N_i^* is the minimizer of problem (4.3).

In order to prove part (ii), we only need to show (4.12) holds when $a_i = 0$ for some $1 \le i \le k$. We now consider the case where $a_1 = 0$. Without loss of generality, we can further assume $a_2 > 0$. The condition $P(\tilde{J}_{1m} < \mu_i) > 0$ actually implies there exists $\varepsilon > 0$ such that $P(\tilde{J}_{1m} \le \mu_2 - \varepsilon) > 0$. Then we have

$$\begin{aligned}
\ln P(\bar{J}_2 - \bar{J}_1 \ge 0) &= \ln P(\sum_{m=1}^{N_1}(\frac{\tilde{J}_{2m}}{\alpha_2} - \tilde{J}_{1m}) + \sum_{m=N_1+1}^{N_2} \frac{\tilde{J}_{2m}}{\alpha_2} \ge 0) \\
&\ge \ln P\big[\sum_{m=N_1+1}^{N_2} \frac{\tilde{J}_{2m}}{\alpha_2} \ge N_1(\mu_2 - \varepsilon)\big] + \ln P\big[\sum_{m=1}^{N_1}(\tilde{J}_{1m} - \frac{\tilde{J}_{2m}}{\alpha_2}) \le N_1(\mu_2 - \varepsilon)\big] \\
&\ge \ln P\big[\sum_{m=N_1+1}^{N_2} \frac{\tilde{J}_{2m}}{N_2} \ge \mu_2 - \varepsilon\big] + N_1 \ln P\big[\tilde{J}_{11} - \frac{\tilde{J}_{21}}{\alpha_2} \le \mu_2 - \varepsilon\big].
\end{aligned}$$

Using strong law of large number, we can show the first term on the right hand side converges to zero and

$$\ln P\big[\tilde{J}_{11} - \frac{\tilde{J}_{21}}{\alpha_2} \le \mu_2 - \varepsilon\big] \to \ln P(\tilde{J}_{11} \le \mu_2 - \varepsilon) > -\infty.$$

Hence, we have

$$\liminf_{T\to\infty} \frac{\ln P(\bar{J}_2 - \bar{J}_1 \ge 0)}{T} \ge a_1 \ln P(\tilde{J}_{11} \le \mu_2 - \varepsilon) = 0,$$

which then implies

$$\lim_{T\to\infty} \frac{\ln(1-P)}{T} = 0 > \frac{Y^*}{T} = \lim_{T\to\infty} \frac{\ln(1-P^*)}{T}.$$

So we establish part (ii) for the case where $a_1 = 0$. Indeed the case where $a_i = 0$ for some $i \in \Omega$ and $a_1 \neq 0$ can be similarly treated by switching the roles of N_i and N_1 in the arguments above. Hence, we complete part (ii). □

Roughly speaking, Lemma 4.7 tells us $\{N_i^*\}$ can dominate any asymptotically linear solution when budget T is large enough. This characterization leads to the following convergence result:

Theorem 4.8 If $P(\tilde{J}_{1m} < \mu_i) > 0$ and $P(\tilde{J}_{im} > \mu_1) > 0$ for all $i \in \Omega$,

$$\lim_{T\to\infty} \frac{\widetilde{N_i} - N_i^*}{T} = 0, \ \forall 1 \le i \le k.$$

Proof. We will denote $\widetilde{P}$ as the PCS associated with solution $\widetilde{N_i}$. The proof is straightforward. Suppose it does not hold. Then we can find an increasing and divergent sequence $\{T_n, n = 1, \cdots\}$ such that $a_i = \lim_{n\to\infty} \frac{\widetilde{N_i}(T_n)}{T_n}$ exists for all $1 \le i \le k$ and $(a_1, \cdots, a_k) \neq (\frac{N_1^*}{T}, \cdots, \frac{N_k^*}{T})$. Then lemma 2 implies that

$$\lim_{n\to\infty} \frac{\ln(1 - \widetilde{P}(T_n))}{T_n} > \lim_{n\to\infty} \frac{\ln(1 - P^*(T_n))}{T_n},$$

which is a contradiction with the fact that $\widetilde{P}$ is always no less than P^*. $\square$

Theorem 4.8 tells us $\widetilde{N_i} - N_i^* = o(T)$. Actually when optimal solution to problem (4.3) is not unique, the theorem means $\{\widetilde{N_i}\}$ is close to one solution $\{N_i^*\}$ and gap is also in order of $o(T)$.

In order to give a more precise estimate of order $o(T)$ above, we will assume all the samples are drawn from non-degenerate Gaussian distributions, where non-degenerate conditions require $\sigma_{ii} > 0$ for $1 \le i \le k$ and $\sigma_{11}\sigma_{ii} > \sigma_{1i}^2$ for $i \in \Omega$. We recall that the optimizer z^* to problem (4.7) has to satisfy one of three conditions in Theorem 4.4. Now we define z^* as **Zero Fit**, if and only if either (1) $F'(z^*) = 0$ and $z^* \in (H^b, 0) \setminus S_1$; or (2) $F'(z^*-)F'(z^*+) = 0$ and $z^* \in S_1 \setminus S_2 \setminus \{H^b\}$; or (3) $F'(z^*+) = 0$ and $z^* \in S_1 \cap S_2$. It can be shown the distance is either bounded by a constant or in the order of $O(\sqrt{T})$, depending on whether the optimizer is Zero Fit or not. More precisely, we have

Theorem 4.9 Assuming all samples $\tilde{J}_{im}$ are drawn from non-degenerate Gaussian distributions,

(i) If z^* is not Zero Fit, $\sup_{T>0} |\widetilde{N_i} - N_i^*| < \infty$ for any $1 \le i \le k$;

(ii) If z^* is Zero Fit, $\sup_{T>0} |\widetilde{N_i} - N_i^*|/\sqrt{T} < \infty$ for any $1 \le i \le k$.

Proof. Before we prove the theorem, we first establish some auxiliary results.

Claim 1: $\sup_{T>0} |\widetilde{Y} - Y^*| < \infty$.

If we let P represent the PCS associated with solution $\{N_i\}$ and denote as $\Phi(\cdot)$ the cumulative distribution function of normal distribution, we will have

$$\begin{aligned} P &= P(\bar{J}_1 - \bar{J}_i > 0,\ i = 2, \dots, k) \\ &\le \min_{i\in\Omega} P(\bar{J}_1 - \bar{J}_i > 0) = 1 - \max_{i\in\Omega} P(\bar{J}_1 - \bar{J}_i \le 0) \\ &= 1 - \max_{i\in\Omega} \Phi(-\sqrt{-2Y_i}) = 1 - \Phi(-\sqrt{-2Y}), \end{aligned} \tag{4.16}$$

where the last step uses assumption that all samples are drawn from normal distribution. On the other side, we have

$$\begin{aligned} P &= P(\bar{J}_1 - \bar{J}_i > 0,\ i = 2, \dots, k) \\ &\ge 1 - \sum_{i=2}^{k} P(\bar{J}_1 - \bar{J}_i \le 0) \\ &\ge 1 - (k-1) \max_{i\in\Omega} P(\bar{J}_1 - \bar{J}_i \le 0) \\ &= 1 - (k-1) \max_{i\in\Omega} \Phi(-\sqrt{-2Y_i}) = 1 - (k-1)\Phi(-\sqrt{-2Y}). \end{aligned} \tag{4.17}$$

If we apply inequalities (4.16) and (4.17) to $\widetilde{P}$ and P^* and use the following inequalities

$$\frac{-2}{a} e^{-a^2/2} \ge \int_{-\infty}^{a} e^{-x^2/2} dx \ge \frac{-a}{1+a^2} e^{-a^2/2} \text{ for } a < 0,$$

we will have

$$\begin{aligned} \ln(1 - \widetilde{P}) &\ge \widetilde{Y} - \frac{1}{2}\ln(-\widetilde{Y}) + \ln(\frac{-2\widetilde{Y}}{1 - 2\widetilde{Y}}) + C_1, \text{ and} \\ \ln(1 - P^*) &\le Y^* - \frac{1}{2}\ln(-Y^*) + C_2, \end{aligned}$$

where C_1 and C_2 are constants. Obviously Theorem 4.8 holds and implies $\frac{\widetilde{Y}}{T} \to \frac{Y^*}{T}$. Now $\widetilde{Y} \to -\infty$ implies that $\ln(-2\widetilde{Y}/(1-2\widetilde{Y}))$ is bounded by some constant. Therefore, we have

$$Y^* \leq \widetilde{Y} \leq Y^* + \frac{1}{2}\ln(\frac{\widetilde{Y}}{Y^*}) + C_3 \leq Y^* + C_4. \tag{4.18}$$

where the last step uses boundedness of $\ln(\widetilde{Y}/Y^*)$ which is implied by $\frac{\widetilde{Y}}{T} \to \frac{Y^*}{T}$.

Claim 2: $\sup_{T>0} |\widetilde{Y}_i - Y_i^*| < \infty,\ \forall i \in \Omega$.

Since $\widetilde{Y} \geq \widetilde{Y}_i$ and $Y_i^* = Y^*$ for all $i \in \Omega$, $\widetilde{Y}_i - Y_i^*$ is bounded from above by (4.18). So we only need to show that $\widetilde{Y}_i - Y_i^*$ is bounded from below for any $i \in \Omega$. Now we fix $i \in \Omega$ and define constant $C_5 = \max_{j\in\Omega} H_j(\frac{N_j^*}{N_1^*})/2 < 0$. We also define a new allocation $\{N_j\}$ with budget T as follows:

$$N_j = \begin{cases} \widetilde{N_j} - \frac{C_4}{C_5}\frac{\widetilde{N_j}}{\widetilde{N_1}}, & j \neq i \\ \widetilde{N_i} + \frac{C_4}{C_5}\frac{T-\widetilde{N_i}}{\widetilde{N_1}}, & j = i \end{cases}. \tag{4.19}$$

Obviously, $\{N_j\}$ satisfies the budget constraint. Theorem 4.8 implies $\frac{\widetilde{N_j}}{\widetilde{N_1}} \to \frac{N_j^*}{N_1^*}$. So given sufficiently large T, $\{N_j\}$ is positive and $H_j(\widetilde{N_j}/\widetilde{N_1}) < C_5$ for all $j \in \Omega$. Now we have

$$Y_j = \widetilde{Y}_j - \frac{C_4}{C_5}H_j(\frac{\widetilde{N_j}}{\widetilde{N_1}}) < \widetilde{Y}_j - C_4 \leq Y^*,\ \forall j \neq i, \tag{4.20}$$

where the last step uses (4.18).

Since Y^* is the optimal value of problem (4.3), we have

$$\widetilde{Y}_i - Y_i^* \geq \widetilde{Y}_i - Y_i = \widetilde{N_i}\big[H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}}) - H_i(\frac{N_i}{N_1})\big] - \frac{C_4}{C_5}\frac{T-\widetilde{N_i}}{\widetilde{N_1}}H_i(\frac{N_i}{N_1}) < \infty,$$

where the last step uses $\frac{\widetilde{N_j}}{\widetilde{N_1}},\ \frac{N_j}{N_1} \to \frac{N_j^*}{N_1^*}$ and continuity of H_i.

Claim 3: If we denote $\widetilde{Y}_i - Y_i^*$ as ΔY_i for $i \in \Omega$, we have

$$\sup_{T>0} |\Delta N_1| < \infty \Longrightarrow \sup_{T>0} |\Delta N_i| < \infty, \ \forall i \in \Omega. \tag{4.21}$$

The mean value theorem yields

$$\begin{aligned} \widetilde{Y}_i - Y_i^* &= H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}})(\widetilde{N_1} - N_1^*) + H_i'(V_i)(\frac{N_1^*\widetilde{N_i}}{\widetilde{N_1}} - N_i^*) \\ &= \big[H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}}) - \frac{\widetilde{N_i}}{\widetilde{N_1}} H_i'(V_i)\big]\Delta N_1 + H_i'(V_i)\Delta N_i, \end{aligned} \tag{4.22}$$

where $V_i = \theta_i \frac{\widetilde{N_i}}{\widetilde{N_1}} + (1-\theta_i)\frac{N_i^*}{N_1^*}$ for some $\theta_i \in [0,1]$. Here H_i' is right (left) derivative if $\frac{\widetilde{N_i}}{\widetilde{N_1}}$ is greater (less) than $\frac{N_i^*}{N_1^*}$. Notice the coefficients of ΔN_1 and ΔN_i converge to $H_i(\frac{N_i^*}{N_1^*}) - \frac{N_i^*}{N_1^*} H_i'(\frac{N_i^*}{N_1^*})$ and $H_i'(\frac{N_i^*}{N_1^*})$, respectively. Also claim 2 implies that $\widetilde{Y}_i - Y_i^*$ is bounded for $i \in \Omega$.

Suppose ΔN_1 is bounded, then $H_i'(V_i)\Delta N_i$ has to be bounded from (4.22). If $H_i'(\frac{N_i^*}{N_1^*}) \neq 0$, ΔN_i is also bounded. Now suppose $H_i'(\frac{N_i^*}{N_1^*}) = 0$ and $H_i'(V_i) \to 0$ for some i. It is easy to know this can be true only if $\sigma_{ii} = 2\sigma_{i1}$ and $V_i \geq \frac{N_i^*}{N_1^*} = 1$. Thus we only might have $\sup_{T>0} \Delta N_i = \infty$. This leads to $\sup_{T>0} \sum_{i=1}^k \Delta N_i = \infty$, which contradicts with $\sup_{T>0} \sum_{i=1}^k \Delta N_i = 0$ implied by the budget constraint. Hence, we establish claim 3.

Now we will show part (i) by considering two cases:

I. $H_i'(\frac{N_i^*}{N_1^*}) = 0$ for some $i \in \Omega$;

II. $H_i'(\frac{N_i^*}{N_1^*}) \neq 0$ for all $i \in \Omega$.

By claim 3, we only have to look at the boundedness of ΔN_1. As we point out above, $H_i'(\frac{N_i^*}{N_1^*}) = 0$ only if $\sigma_{ii} = 2\sigma_{1i}$ and $N_i^* = N_1^*$. In this case, $H_i(x)$ is a constant when $x \geq 1$. $H_i'(V_i) \to 0$ implies $H_i'(V_i) = 0$. Hence, (4.22) implies $\big[H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}}) - \frac{\widetilde{N_i}}{\widetilde{N_1}} H_i'(V_i)\big]\Delta N_1$ is bounded. Since $H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}}) - \frac{\widetilde{N_i}}{\widetilde{N_1}} H_i'(V_i) \to \frac{Y_i^*}{N_1^*} \neq 0$, ΔN_1 is bounded. Therefore, by claim 3, all ΔN_i is bounded for case I.

For case II, we need condition that $\frac{Y^*}{N_1^*}$ is not Zero Fit. (4.22) implies boundedness of

$$[\frac{H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}})}{H_i'(V_i)} - \frac{\widetilde{N_i}}{\widetilde{N_1}}]\Delta N_1 + \Delta N_i, \ \forall i \in \Omega. \tag{4.23}$$

Hence, also bounded is

$$\sum_{i=2}^{k} [\frac{H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}})}{H_i'(V_i)} - \frac{\widetilde{N_i}}{\widetilde{N_1}}]\Delta N_1 + \sum_{i=2}^{k} \Delta N_i = \big(\sum_{i=2}^{k} [\frac{H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}})}{H_i'(V_i)} - \frac{\widetilde{N_i}}{\widetilde{N_1}}] - 1\big)\Delta N_1. \tag{4.24}$$

Checking the coefficient of ΔN_1 above, we have

$$\begin{aligned}\widetilde{B} \equiv \sum_{i=2}^{k} [\frac{H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}})}{H_i'(V_i)} - \frac{\widetilde{N_i}}{\widetilde{N_1}}] - 1 \quad &\to \quad \sum_{i=2}^{k} \frac{\frac{Y_i^*}{N_1^*}}{H_i'(\frac{N_i^*}{N_1^*})} - \frac{T}{N_1^*} = \sum_{i=2}^{k} \frac{\frac{Y^*}{N_1^*}}{H_i'(H_i^{-1}(\frac{Y^*}{N_1^*}))} - \frac{T}{N_1^*} \\ &= \quad \sum_{i=2}^{k} \frac{\frac{Y^*}{N_1^*}}{H_i'(H_i^{-1}(\frac{Y^*}{N_1^*}))} - 1 - \sum_{i=2}^{k} H_i^{-1}(\frac{Y^*}{N_1^*}) \equiv B^*,\end{aligned}$$

where the first equality uses $Y_i^* = Y^*$ and the second uses $\frac{N_i^*}{N_1^*} = H_i^{-1}(\frac{Y^*}{N_1^*})$.

It is easy to check that $B^* = 0$ if and only if $\frac{Y^*}{N_1^*}$ is Zero Fit. So we establish boundedness of ΔN_1 and ΔN_i for case II. This completes our proof for part (i).

Now we assume $\frac{Y^*}{N_1^*}$ is Zero Fit, i.e. $B^* = 0$. Note that

$$H_i'(V_i) = \big(H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}}) - H_i(\frac{N_i^*}{N_1^*})\big)/(\frac{\widetilde{N_i}}{\widetilde{N_1}} - \frac{N_i^*}{N_1^*}). \tag{4.25}$$

If we use second order Taylor expansion of $H_i(\frac{\widetilde{N_i}}{\widetilde{N_1}})$ around $\frac{N_i^*}{N_1^*}$, we have

$$H_i'(V_i) - H_i'(\frac{N_i^*}{N_1^*}) = \frac{1}{2}(\frac{\widetilde{N_i}}{\widetilde{N_1}} - \frac{N_i^*}{N_1^*})H_i''(U_i), \tag{4.26}$$

where $U_i = \theta_i' \frac{\widetilde{N}_i}{\widetilde{N}_1} + (1-\theta_i')\frac{N_i^*}{N_1^*}$ for some $\theta_i' \in [0,1]$. Now we can rewrite (4.24) as

$$
\begin{aligned}
\widetilde{B}\Delta N_1 &= (\widetilde{B} - B^*)\Delta N_1 \\
&= \sum_{i=2}^{k} \Big[\frac{H_i(\frac{\widetilde{N}_i}{\widetilde{N}_1})}{H_i'(V_i)} - \frac{H_i(\frac{N_i^*}{N_1^*})}{H_i'(\frac{N_i^*}{N_1^*})} - (\frac{\widetilde{N}_i}{\widetilde{N}_1} - \frac{N_i^*}{N_1^*})\Big]\Delta N_1 \\
&= \sum_{i=2}^{k} \Big[\frac{H_i(\frac{N_i^*}{N_1^*})}{H_i'(V_i)} - \frac{H_i(\frac{N_i^*}{N_1^*})}{H_i'(\frac{N_i^*}{N_1^*})}\Big]\Delta N_1 \\
&= \sum_{i=2}^{k} \Big[H_i'(\frac{N_i^*}{N_1^*}) - H_i'(V_i)\Big] \frac{H_i(\frac{N_i^*}{N_1^*})}{H_i'(V_i)H_i'(\frac{N_i^*}{N_1^*})}\Delta N_1 \\
&= -\frac{1}{2}\sum_{i=2}^{k} \Big[\frac{\widetilde{N}_i}{\widetilde{N}_1} - \frac{N_i^*}{N_1^*}\Big] \frac{H_i(\frac{N_i^*}{N_1^*})H_i''(U_i)}{H_i'(V_i)H_i'(\frac{N_i^*}{N_1^*})}\Delta N_1 \\
&= -\frac{1}{2}\sum_{i=2}^{k} \frac{\Delta N_1}{\widetilde{N}_1}\Big[\Delta N_i - \frac{N_i^*}{N_1^*}\Delta N_1\Big] \frac{H_i(\frac{N_i^*}{N_1^*})H_i''(U_i)}{H_i'(V_i)H_i'(\frac{N_i^*}{N_1^*})} \\
&= -\frac{1}{2}\sum_{i=2}^{k} \frac{\Delta N_1}{\widetilde{N}_1}\Big[\Delta N_i + (\frac{H_i(\frac{\widetilde{N}_i}{\widetilde{N}_1})}{H_i'(V_i)} - \frac{\widetilde{N}_i}{\widetilde{N}_1})\Delta N_1\Big] \frac{H_i(\frac{N_i^*}{N_1^*})H_i''(U_i)}{H_i'(V_i)H_i'(\frac{N_i^*}{N_1^*})} + \\
&\quad \frac{(\Delta N_1)^2}{2N_1^*}\sum_{i=2}^{k} \Big[\frac{N_1^*}{\widetilde{N}_1}(\frac{H_i(\frac{\widetilde{N}_i}{\widetilde{N}_1})}{H_i'(V_i)} - \frac{\widetilde{N}_i}{\widetilde{N}_1} + \frac{N_i^*}{N_1^*})\frac{H_i(\frac{N_i^*}{N_1^*})H_i''(U_i)}{H_i'(V_i)H_i'(\frac{N_i^*}{N_1^*})}\Big],
\end{aligned}
$$

where the third step uses (4.25) and the fifth uses (4.26). Using boundedness of (4.23), $\frac{\Delta N_1}{\widetilde{N}_1} \to 0$, and $\frac{\widetilde{N}_i}{\widetilde{N}_1} \to \frac{N_i^*}{N_1^*}$, we notice the first term on the right hand side converges to zero. So boundedness of $\widetilde{B}\Delta N_1$ implies boundedness of the second term. Actually the summation in the second term converges to

$$
(\frac{Y^*}{N_1^*})^2 \sum_{i=2}^{k} \frac{H_i''(\frac{N_i^*}{N_1^*})}{(H_i'(\frac{N_i^*}{N_1^*}))^3} = \frac{Y^*}{N_1^*} M'(\frac{Y^*}{N_1^*}),
$$

where $M(\cdot)$ is defined by (4.9). It can be shown $M(z)/z^2$ is strictly increasing when $z < 0$. This implies that $M'(\frac{Y^*}{N_1^*}) \neq 0$ when $\frac{Y^*}{N_1^*}$ is Zero Fit. So we can conclude that $\frac{(\Delta N_1)^2}{2N_1^*}$ is bounded, or $\Delta N_1 = O(\sqrt{T})$. We can also conclude that $\Delta N_i = O(\sqrt{T})$ by using boundedness of (4.23). Part (ii) is established. $\square$

4.4. Conclusion

This chapter proposes an algorithm to solve an OCBA problem under non-Gaussian setting. The discussion here can be taken as a generalized version to the Gaussian setting. Although the solution obtained from this algorithm is an approximate solution, we show that this solution is better than any other linear solution when the budget is large enough. The convergence rate is given when the setting is simplified to Gaussian. Our further direction will be to combine Bayesian analysis with our algorithm.

Chapter 5

Future Research

The following are some future research directions we wish to pursue:

- For the stochastic approximation algorithm with deterministic perturbations, we will consider more general cases and attempt to relax the assumptions required for proving convergence of the algorithm. To relax the boundedness assumption, we will apply the framework of projection to the case of deterministic perturbations and explore conditions required for convergence.
- For the American option pricing algorithm, we will apply it to estimating gradients of American options. A possible direction is to combine our technique with Malliavin Calculus.
- For the OCBA algorithm, we will attempt to relax the negative correlation assumption, so that the algorithm can be applied to more general cases. Also we wish to introduce Bayesian analysis into the algorithm, so that exact knowledge of the sampling distribution is not required.

BIBLIOGRAPHY

[1] Bachelier, L., "Theory of speculation," *The Random Character of Stock Prices*, Cootner, P.H., (editor), Cambridge: MIT. Translated from the 1900 doctoral thesis, 1964.

[2] Ben-Ameur, H., Breton M., and L'Ecuyer, P., "A dynamic programming procedure for pricing American-style Asian options," *Management Science*, **48**, 625-643, 2002.

[3] Benveniste, A., Metivier, M., and Priouret, P., *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, New York, 1990.

[4] Bhatnagar, S., Fu, M.C., Marcus, S.I. and Wang, I-J., "Two-timescale simultaneous perturbation stochastic approximation using deterministic perturbation sequences," Submitted to ACM Transactions on Modeling and Computer Simulation, 2002.

[5] Billingsley, P., *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[6] Black, F., and Scholes, M.S. “The pricing of options and corporate liabilities,” *Journal of Political Economy*, **81**, 637-654, 1973.

[7] Blum, J.R., “Approximation methods which converge with probability one,” *Ann. Math. Stat.*, **25**, 382-386, 1954a.

[8] Blum, J.R., “Multidimensional stochastic approximation procedures,” *Ann. Math. Stat.*, **25**, 737-744, 1954b.

[9] Boyle, P.P., “Options: a Monte Carlo approach,” *Journal of Finan- cial Economics*, **4**, 323-338, 1977.

[10] Broadie, M., and Glasserman, P., “Pricing American-style securities using simulation,” *Journal of Economic Dynamics and Control*, **21**, 1323-1352, 1997.

[11] Broadie, M., and Glasserman, P., “A stochastic mesh method for pricing high-dimensional American options”, *Journal of Computational Finance*, **7**, 35-72, 2004.

[12] Boyle, P.P., Broadie, M., and Glasserman, P., “Simulation methods for security pricing,” *Journal of Economic Dynamics and Control*, **21**, 1267-1321, 1997.

[13] Burkholder, D.L., “On a class of stochastic approximation procedures,” *Ann. Math. Stat.*, **27**, 1044-1059, 1956.

[14] Chen, C.H., Lin, J., Yücesan, E., and Chick, S.E., “Simulation budget allocation for further enhancing the efficiency of ordinal optimization,” *Discrete Event Dynamic Systems: Theory and Applications*, **10**, 251-270, 2000.

[15] Chen, D.C., "Comparative study of stochastic algorithms for system optimization based on gradient approximations," *IEEE Trans. Syst., Man, Cybern.*, **27**, 1997.

[16] Chen, H.C., Chen, C.H., Dai, L., and Yücesan, E., "New development of optimal computing budget allocation for discrete event simulation," *Proceedings of the Winter Simulation Conference*, 334-341, 1997.

[17] Chen, E.J., and Kelton, W.D., "An enhanced two-stage selection procedure," *Proceedings of the Winter Simulation Conference*, 727-735, 2000.

[18] Chick, S.E., and Inoue, K., "New two-stage and sequential procedures for selecting the best simulated system," *Operations Research*, **49**, 732-743, 2001a.

[19] Chick, S.E., and Inoue, K., "New procedures to select the best simulated system using common random numbers," *Management Science*, **47**, 1133-1149, 2001b.

[20] Chung, K.L., "On a stochastic approximation method," *Ann. Math. Stat.*, **25**, 463-483, 1954.

[21] Derman, C., "An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure," *Ann. Math. Stat.*, **27**, 529-532, 1956.

[22] Dubins, L.E., and Freedman, D.A., "A sharper form of the Borel-Cantelli lemma and the strong law," *Ann. Math. Stat.*, **36**, 800-807, 1965.

[23] Ermoliev, Y., "On the method of generalized stochastic gradients and quasi-Fejer sequences," *Cybernetics*, **5**, 208-220, 1969.

[24] Fabian, V., "Stochastic approximation of minima with improved asymptotic speed," *Ann. Math. Stat.*, **38**, 191-200, 1967.

[25] Fabian, V., "On asymptotic normality in stochastic approximation," *Ann. Math. Stat.*, **39**, 1327-1332, 1968.

[26] Fu, M.C., and Hu, J.Q., "Sensitivity analysis for Monte Carlo simulation of option pricing," *Probability in the Engineering and Informational Sciences*, **9**, 417-446, 1995.

[27] Fu, M.C., and Hu, J.Q., Chen, C.H., and Xiong, X., "Simulation allocation for determining the best design in the presence of correlated sampling," working paper.

[28] Fu, M.C., Laprise, S.B., Madan, D.B., Su, Y., and Wu, R., "Pricing American options: a comparison of Monte Carlo simulation approaches," *Journal of Computational Finance*,**4(3)**, 39-88, 2001.

[29] Fu, M.C., Wu, R., Gürkan, G., and Demir, A.Y., "A note on perturbation analysis estimators for American-style options," *Probability in the Engineering and Informational Sciences*, **14**, 385-392, 2000.

[30] Gerencser, L., "Convergence rate of moments in stochastic approximation with simultaneous perturbation gradient approximation and resetting," *IEEE Trans. Auto. Contr.*, **44(5)**, 894-905, 1999.

[31] Goldsman, D., and Nelson, B.L., "Comparing systems via simulation," In *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, 273-306, John Wiley and Sons, 1998.

[32] Grant, D., Vora, G., and Weeks, D., "Simulation and the early-exercise option problem," *Journal of Financial Engineering*, **5(3)**, 211-227, 1996.

[33] Grant, D., Vora, G., and Weeks, D., "Path-dependent options: extending the Monte Carlo simulation approach," *Management Science*, **43**, 1589-1602, 1997.

[34] Harrison, J.M., and Kreps, D., "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, **20**, 381-408, 1979.

[35] Harrison, J. M. and Pliska, S.R., "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and their Applications*, **11**, 215-260, 1981.

[36] Hedayat, A.S., Sloane, N.J.A. and Stufken, J., *Orthogonal Arrays: Theory and Applications*, Springer Verlag, New York, 1999.

[37] Hull, J.C., and White, A., "The pricing of options on assets with stochastic volatilities," *Journal of Finance*, **42**, 281-300, 1987.

[38] Johnson, H., and Shanno, D., "Option pricing when the variance is changing," *Journal of Financial and Quantitative Analysis*, **22**, 143-151, 1987.

[39] Kallenberg, O., *Foundations of Modern Probability*, Springer-Verlag, New York, 1997.

[40] Kiefer, J. and Wolfowitz, E.L., "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, **23**, 462-466, 1952.

[41] Kloeden, P., and Platen, E., *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York, 1992.

[42] Kloeden, P., Platen, E., and Schurz, H., *Numerical Solution of SDE Through Computer Experiments*, Springer-Verlag, New York, 1991.

[43] Kulkarni, S.R., and Horn, C., "An alternative proof for convergence of stochastic approximation algorithms," *IEEE Transactions on Automatic Control*, **41(3)**, 419-424, 1996.

[44] Kushner, H.K. and Clark, D.S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer, 1978.

[45] Ljung, L., Pflug, G., and Walk, H., *Stochastic Approximation and Optimization of Random Systems*, Birkhauser, Germany, 1992.

[46] Longstaff, F., and Schwartz, E., "Pricing American options by simulation: a simple least-squares approach," *The Review of Financial Studies*, **14**, 113-147, 2001.

[47] Niederreiter, H., *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

[48] Robbins, H. and Monro, S., "A stochastic approximation method," *Ann. Math. Stat.*, **22**, 400-407, 1951.

[49] Sacks, J., "Asymptotic distribution of stochastic approximation procedures," *Ann. Math. Stat.*, **29**, 373-405, 1958.

[50] Samuelson, P.A., "Rational theory of warrant pricing," *Industrial Management Review*, **6**, 13-31, 1965.

[51] Scott, L.O., "Option pricing when the variance changes randomly: theory, estimation, and an application," *Journal of Financial and Quantitative Analysis*, **22**, 419-438, 1987.

[52] Seberry, J., and Yamada, M., "Hadamard matrices, sequences, and block designs," In *Contemporary Design Theory – A Collection of Surveys*, (eds. D.J. Stinson and J. Dintiz), 431-560, John Wiley and Sons, 1992.

[53] Spall, J.C., "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automat. Contr.*, **37**, 1992.

[54] Spall, J.C., "A one-measurement form of simultaneous perturbation stochastic approximation," *Automatica*, **33**, 109-112, 1997.

[55] Styblinski, M.A. and Tang, T.S., "Experiments in nonconvex optimization: stochastic approximation with function soothing and simulated annealing," *Neural Networks*, **3**, 467-483, 1990.

[56] Wang, I-J., Chong, E.K.P., and Kulkarni, S.R., "Equivalent necessary and sufficient conditions on noise sequences for stochastic approximation algorithms," *Advances in Applied Probability*, **28**, 784-801, 1996.

[57] Wang, I-J., Chong, E.K.P., and Kulkarni, S.R., "Weighted averaging and stochastic approximation," *Mathematics of Control, Signals, and Systems*, **10(1)**, 41-60, 1997.

[58] Wang, I-J., and Chong,E.K.P., "A deterministic analysis of stochastic approximation with randomized directions," *IEEE Transactions on Automatic Control*, **43(12)**, 1745-1749, 1998.

[59] Wolfowitz, E.L., "On the stochastic approximation method of Robbins and Monro," *Ann. Math. Stat.*, **23**, 457-461, 1952.

[60] Wolfowitz, E.L., "On stochastic approximation methods," *Ann. Math. Stat.*, **27**, 1151-1155, 1956.

[61] Wu, R., and Fu, M.C., "Optimal exercise policies and simulation-based valuation for American-Asian options," *Operations Research*, **51(1)**, 52-66, 2003.

[62] Xiong, X., Wang, I-J., and Fu, M.C., "Randomized-direction stochastic approximation algorithms using deterministic sequences," *Proceedings of the 2002 Winter Simulation Conference*, 285-291, 2002.