ABSTRACT

Title of dissertation:	SIMULATION-BASED METHODS FOR STOCHASTIC CONTROL AND GLOBAL OPTIMIZATION
	Yongqiang Wang, Doctor of Philosophy, 2011
Dissertation directed by:	Professor Steven I. Marcus Department of Electrical and Computer Engineering
	Professor Michael C. Fu Department of Decision, Operations, and Information Technologies

Ideas of stochastic control have found applications in a variety of areas. A subclass of the problems with parameterized policies (including some stochastic impulse control problems) has received significant attention recently because of emerging applications in the areas of engineering, management, and mathematical finance. However, explicit solutions for this type of stochastic control problems only exist for some special cases, and effective numerical methods are relatively rare. Deriving efficient stochastic derivative estimators for payoff functions with discontinuities arising in many problems of practical interest is very challenging. Global optimization problems are extremely hard to solve due to the typical multimodal properties of objective functions. With the increasing availability of computing power and memory, there is a rapid development in the merging of simulation and optimization techniques. Developing new and efficient simulation-based optimization algorithms for solving stochastic control and global optimization problems is the primary goal of this thesis.

First we develop a new simulation-based optimization algorithm to solve a stochastic control problem with a parameterized policy that arises in the setting of dynamic pricing and inventory control. We consider a joint dynamic pricing and inventory control problem with continuous stochastic demand and model the problem as a stochastic control problem. An explicit solution is given when a special demand model is considered. For general demand models with a parameterized policy, we develop a new simulation-based method to solve this stochastic control problem. We prove the convergence of the algorithm and show the effectiveness of the algorithm by numerical experiments.

In the second part of this thesis, we focus on the problem of estimating the derivatives for a class of discontinuous payoff functions, for which existing methods are either not valid or not efficient. We derive a new unbiased stochastic derivative estimator for performance functions containing indicator functions. One important feature of this new estimator is that it can be computed from a single sample path or simulation, whereas existing estimators in the literature require additional simulations.

Finally we propose a new framework for solving global optimization problems by establishing a connection with evolutionary games, and show that a particular equilibrium set of the evolutionary game is asymptotically stable. Based on this connection, we propose a Model-based Evolutionary Optimization (MEO) algorithm, which uses probabilistic models to generate new candidate solutions and uses dynamics from evolutionary game theory to govern the evolution of the probabilistic models. MEO gives new insight into the mechanism of model updating in model-based global optimization algorithms from the perspective of evolutionary game theory. Furthermore, it opens the door to developing new algorithms by using various learning algorithms and analysis techniques from evolutionary game theory.

SIMULATION-BASED METHODS FOR STOCHASTIC CONTROL AND GLOBAL OPTIMIZATION

by

Yongqiang Wang

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
2011

Advisory Committee:
Professor Steven I. Marcus, Chair/Advisor
Professor Michael C. Fu, Co-Advisor
Professor P. S. Krishnaprasad
Professor André L. Tits
Professor Jeffrey W. Herrmann

Dedication

To my family.

Acknowledgments

First and foremost, I would like to express my sincerest thanks to my advisors Professor Steven I. Marcus and Professor Michael C. Fu, whose consistent guidance and support have made my graduate studies an unforgettable and exciting experience. I have learned and benefited enormously from not only their broad knowledge and insightful views of the research areas, but also their passion and dedication to high quality and integrity of research. I am grateful for their continuous support and guidance and I will always cherish my experience as a graduate students under their supervision.

I would also like to thank Professor P. S. Krishnaprasad and Professor André Tits for offering wonderful courses. I have learned a lot from Professor P. S. Krishnaprasad's passionate lecturing and deep understanding of the field and benefited a lot through discussion with him. I will always cherish the experience of being a teaching assistant with Professor Tits, who provided me with enormous guidance and I am grateful for his continuous support. I will also like to thank other Professors in ECE and Maryland, who have provided inspiring lectures that made my graduate life interesting and exciting.

I would like to thank Professor Jeffrey W. Herrmann for serving on my thesis committee and sparing his valuable time reviewing the manuscript.

I would also like to acknowledge help and support from my former officemate Enlu Zhou and current officemates Kun Lin and James Ferlez. I will never forget the interesting discussions and debates we had and will always cherish the wonderful time we spent together.

Words cannot express my thanks to my parents for their unconditional support and love. I also want to thank my wife for her continuous love and support. Without them, this thesis wouldn't be possible.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Stochastic control has been widely applied to many different areas of engineering and management. There are many stochastic control problems that do not have analytical solutions due to the generality of the model, and existing numerical methods are either not applicable or not effective. Gradient estimation plays an important role in simulation optimization and sensitivity analysis, and deriving efficient stochastic derivative estimators for payoff functions with discontinuities arising in many problems of practical interest is very challenging. Global optimization problems are generally very hard to solve. Although there are many heuristic algorithms proposed to solve global optimization problems, many are very hard to analyze in general. Due to the increasing availability of computing power, simulation optimization is a very promising technique to handle difficult stochastic control and global optimization problems. In this dissertation, we propose new simulation optimization methodologies to solve stochastic control and global optimization problems. This chapter gives a brief introduction of the motivation, and then introduces simulation optimization and the idea of using game theory, particularly evolutionary game theory, for optimization, followed by research contributions and outline of the thesis.

1.1 Motivation

1.1.1 Stochastic Control Problems in Revenue Management

With the rapid expansion of the Internet and e-commerce, consumers have easier access to prices of products. Pricing is one of the most effective tools that can be manipulated to encourage or discourage demand, especially for price-sensitive customers. A high initial inventory level provides the seller incentive to lower the price in order to stimulate demand and reduce holding costs. However, as the inventory level changes, the behavior of customers also changes. In order to maximize the expected profit, the seller should incorporate all the information available and adjust the price accordingly. There is a substantial literature on dynamic pricing ([1, 27, 24, 13]), and various models have been proposed. The dynamic pricing models can be grouped into two categories: discrete-time models and continuous-time models.

For discrete-time models, independent demands arrive at discrete times ([27, 24, 21]). The demand is generally a function of a deterministic term and a random noise term, both of which might depend on the price, time, and the amount sold. Additive and multiplicative models are two types of models that are commonly used ([13]). Under the Markovian assumption, the dynamic pricing problem in the discrete-time case can be modeled as a Markov decision process (MDP), and can be solved by using standard value iteration or policy iteration algorithms ([8, 27, 24]). Under some mild conditions, [27] proved that a base-stock list price policy is optimal; they also showed the benefit of dynamic pricing versus static pricing.

For continuous-time models, the most common formulation assumes that the demand follows a Poisson process with a deterministic intensity that depends on the price

and time ([39, 40, 14, 28, 73]). In [39], sellers intend to sell a finite number of identical units over a finite horizon with the demand following a Poisson process, with independent, identically distributed reservation prices. The dynamic pricing problem is then formulated as a stochastic control problem. In the case of Poisson demand, when reservation prices is exponentially distributed, the optimal pricing strategy can be easily derived; however no analytical solutions exist for general models. [73] introduced uncertainties into the arrival rate of the demand model, formulated the dynamic pricing problem as a stochastic game, and gave an analytical solution for a special demand model. [26] incorporated customer choice into the model, and gave a heuristic pricing policy with upper and lower bounds.

Continuous stochastic diffusion processes have been used to model demands that arrive continuously; decisions can be made at any time for these models ([4, 25, 99, 10, 6]), as opposed to the Poisson demand model, in which demands arrive at discrete epochs, where decisions are made. [4] applied stochastic control theory to study an inventory control problem with a Brownian demand and explicitly characterized the optimal policy. [99] proved the optimality of an (s, S) policy for a one-dimensional diffusion inventory system. More recently, [6] considered an inventory control problem where the demand process is composed of a compound Poisson process and a Brownian motion and proved the optimality of an (s, S) policy. Other related works using Brownian demands include [85, 7, 19, 20], and [5]. All these works treating diffusion demand models focus on inventory control to characterize the optimal ordering policy, whereas works addressing the problem of *jointly* optimizing pricing and ordering policies under Brownian demands are relatively rare. [86] analyzed a dynamic pricing problem with demand uncertainty modeled by a Brownian motion without incorporating ordering decisions.

When a price-sensitive demand is modeled by continuous diffusion process, the dy-

3

namic pricing problem can be formulated as a stochastic control problem. For a class of pricing policies that allow a limited number of price changes, for example the markdown pricing policy, the problem of jointly optimizing the prices and the initial inventory level can be formulated as a continuous stochastic control problem with the initial ordering as a decision variable.

The aforementioned stochastic control problem generally does not admit an analytical solution. [22] considered a special continuous demand model, where the arrival rate and the uncertainty term in the demand only depend on the price. In their formulation, the dynamic pricing problem is transformed into a deterministic optimization problem and is solved numerically. But their method cannot be generalized to solve dynamic pricing problems with general demand models. There are two major types of numerical methods for stochastic control problems ([83]). Purely deterministic approximation methods discretize the HJB equation by finite differences or a finite element method, and thus obtain an approximate value function at points on the space-time grid. Probabilistic methods include the Markov chain approximation method presented by [70]. Markov chain approximation and finite differences cannot be applied directly to the dynamic pricing problem with a limited number of price changes because of this constraint on the pricing policy. Moreover, it is desirable to understand how the uncertainties in the demand and the changing of other parameters, such as parameters in the arrival rate and holding cost, would affect the pricing policy and the profit. This sensitivity information cannot be obtained directly using Markov chain approximation.

In the literature of supply chain management, simulation-based algorithms have been applied to solve inventory control problems ([31, 34, 45, 66, 60]) and network revenue management problems ([9, 104]). In simulation-based algorithms, the key is to obtain

efficient gradient estimations. [9] proposed an innovative simulation-based method for computing protection levels in a virtual nesting control scheme, in which a finite differences estimator is derived, because a discrete model for capacity and demand is used. [104] analyzed a continuous version of the problem, which enabled the derivation of a gradient estimator using infinitesimal perturbation analysis (IPA). However, all these methods can only be applied to discrete-time models.

1.1.2   Gradient Estimation

For complex stochastic models requiring simulation, gradient estimates play an important role in both sensitivity analysis and gradient-based optimization. As indicated in chapter 7 of [44] for the application of derivative estimates in finance, "Whereas the prices themselves can often be observed in the market, their sensitivities cannot, so accurate calculation of sensitivities is arguably even more important than calculation of prices."

In the last three decades, gradient estimation has been studied extensively in the simulation literature. Infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method are two of the main techniques [53, 92]. Other techniques include smoothed perturbation analysis (SPA), which can be applied to performance functions containing discontinuities, and methods based on weak derivatives (WD) and Malliavin calculus.

Introduced in [52], IPA has been widely used in sensitivity analysis for discrete-event systems (please see applications of IPA in queueing systems and inventory control problems in [100, 43, 53, 35, 46, 72, 3]) as well as for financial derivatives [44, 17, 36]. IPA enables the sensitivity of a performance function to be estimated while observing a *single* sample path of a system and hence offers significant computational savings compared with the "brute force" finite differences method. However, IPA requires the performance function to be

5

almost surely continuous, which makes IPA not applicable in many cases. For example, second-order derivatives cannot be obtained by IPA for European call options [17].

Rather than differentiating a performance function as IPA does, the likelihood ratio (LR) method constructs derivative estimators from the derivatives of the probability measure associated with a simulation model. The method was proposed by [48], [87], and [91] to study discrete event systems and has also been used in financial applications [17, 44]. It is also called the score function (SF) method [92]. More exposition of the LR method can be found in [91, 90, 32]. The LR method does not require continuity of performance functions and hence is more widely applicable than IPA. However, parameters of interest have to be in probability density functions in order to apply LR, whereas in many cases they appear naturally in performance functions. The push in and push out method may be able to move parameters of interest out of the performance function and push them into a probability density function [92]; then the LR method can be applied. However the push in and push out method has only been demonstrated on some simple cases [92].

1.1.3   Global Optimization

Global optimization aims at finding the global optimal solutions for problems with many local optimal solutions. Due to the presence of possible multiple local extrema, global optimization problems are typically extremely difficult to solve. Various models and heuristics have been proposed to solve global optimization problems. According to the criteria in [110], global optimization algorithms can be grouped as *instance-based* methods and *model-based* methods. Instance-based methods generate new candidate solutions that explicitly depend on the the previous solutions. Among instance-based methods are simulated annealing [63], genetic algorithm [97], tabu search ([47]), and nested parti-

tion [93, 94]. Different from instance-based methods, model-based methods generate new candidate solutions based on models, which are updated by using the previously generated solutions. Some of the well known model-based methods include cross-entropy (CE) method [15, 75], model-based reference adaptive search (MRAS) [59], and estimation of distribution algorithms (EDAs) [80]. Recently, [109] formulated the global optimization problem as a filtering problem and presented a new particle filtering-based framework to solve global optimization problems. Bayes updating is used to guide the evolution of the probability models that are used to generate new candidate solutions.

Many global optimization algorithms need to generate and maintain a group of candidate solutions. How to update and maintain this group of samples is the key in these global optimization algorithms. Most of the model-based algorithms use single mode probabilistic models to generate candidate solutions, which fails to capture the multimodal properties of global optimization problems.

With the increasing availability of computing power and memory, there has been a rapid development in the merging of simulation and optimization techniques [33]. Simulation-based methods have become very promising when dealing with the above challenges.

1.2   Simulation Optimization

Generally speaking, simulation optimization covers a large collection of optimization techniques that are developed using Monte Carlo simulation to solve decision making problems in many different areas. Typically there are three classes of simulation optimization algorithms: sample average approximation (SAA) ([64]), gradient-based algorithms, and random search. The basic idea of SAA is to approximate expectations using the corresponding sample average function and then solve the resulting sample average optimization

problem. Then existing linear programming and nonlinear optimization techniques can be applied is one of the major advantages of SAA. Here we give a brief introduction of stochastic approximation, which belongs to the type of gradient-based algorithms.

1.2.1 Stochastic Approximation

Stochastic approximation (SA) is a gradient-based optimization method for stochastic systems. The SA method was first introduced by the pioneering paper [88]. The classical SA algorithm solves the following stochastic optimization problem

$$\sup_{\theta \in \Theta} E[H(\theta)], \tag{1.1}$$

where Θ is the domain of the parameter θ, by mimicking the gradient ascent method. Denote $g(\theta) = \frac{\partial E[H(\theta)]}{\partial \theta}$ as the gradient of $E[H(\theta)]$ with respect to the parameter θ. Let $\hat{g}(\theta)$ be an estimator of the gradient $g(\theta)$; the SA algorithm generates iterates by the formula

$$\theta_{k+1} = \Pi_\Theta(\theta_k + a_k \hat{g}(\theta_k)),$$

where θ_k is the value of the parameter at the beginning of iteration k, $\{a_k\}$ is a sequence of positive step sizes, and Π_Θ is a projection onto the parameter set Θ. SA has several desirable properties that make it attractive for adaptive schemes. It is a simulation-based algorithm, which only needs estimates of the gradient information, not the exact gradient. It usually has low computational and memory requirements for each iteration.

There are various conditions on the sequence of step sizes that ensure the convergence of the algorithm [68]. SA has been successfully applied in many areas such as adaptive signal processing, adaptive control, artificial intelligence, and operations research

8

[16]. Glasserman et al. [45] applied SA to solve an inventory control problem. Bertsimas et al. [9] and van Ryzin et al. [104] used stochastic approximation to find the optimal virtual nesting control policy for a network revenue management problem. Topaloglu [102] applied SA to find the optimal bid prices for a network revenue management problem. Kunnumkal et al.[66] computed the optimal base-stock levels for some inventory control problems using SA. However in all the aforementioned applications, they assume that demands follow some discrete-time models. To the best of our knowledge SA has not been applied to solve the continuous stochastic control problem formulated out of the dynamic pricing problem in Section 1.1.1 because of the difficulty of precisely simulating the evolution of the inventory level and the difficulty of obtaining good gradient estimators.

1.2.2 Random Search and Evolutionary Game Theory

The class of random search algorithms includes nested partitions ([93]), and algorithms adapted from metaheuristics for deterministic optimization problems such as genetic algorithms ([97]), tabu search ([47]), and simulated annealing ([63]). As introduced in Section 1.1.3, model-based algorithms such as EDAs ([80]), cross-entropy ([15, 75]), and MRAS ([59]) are also random search algorithms proposed recently. As a general tool to study strategic interaction of players using different strategies, game theory can also be used to study the interaction and evolving of samples in random search. We now give a brief introduction to the literature of game theory and evolutionary game theory and their application to random search, learning, and optimization.

Game theory studies the strategic interaction of players using different strategies; it has been applied in many areas such as economics, engineering, and biology [38, 95]. A Nash equilibrium is a set of strategies, one for each player, such that no one has an incentive

to unilaterally deviate from his own strategy. In order to obtain Nash equilibria in games, many learning algorithms have been designed. Some of the well know learning algorithms include fictitious play [18], stochastic fictitious play [38], and regret matching [50]. These learning algorithms have also been successfully applied in other areas, such as multi-agent learning [95] and optimization [71, 41]. Recently, Lambert et al. [71] and Garcia et al. [42] have applied game theory to solve discrete optimization problems, where they model the optimization problem as a potential game. Fictitious play and joint fictitious play are adopted to obtain the Nash equilibrium and two sampled version of fictitious and joint fictitious play are also proposed in [71, 42]. For a potential game, although fictitious play has been proven to converge to a mixed strategy Nash equilibrium, the mixed strategy equilibrium might not be a feasible solution for the optimization problem. The algorithms in [42, 71] only work for discrete optimization problems with a finite solution space, and moreover the Nash equilibrium obtained by fictitious play might only be a local optimal solution.

Evolutionary game theory applies game theory to study the evolution of the number of players playing different strategies in a population setting. After being introduced by biologist Maynard Smith [76], evolutionary game theory has become popular in biology and is attracting increasing interest from researchers in other areas such as engineering and economics. Different from static games, replicator dynamics is introduced in evolutionary games, and in this scheme the growth rate of the proportion of players using a certain strategy is equal to the difference between the average payoff of that strategy and the average payoff of the whole population. Replicator dynamics can also be used as a learning algorithm to study the behavior of multiple agents ([103]).

1.3 Research Contributions

The goal is to develop efficient simulation-based algorithms to solve stochastic control problems derived from a revenue management problem, and to solve global optimization problems. We are also interested in constructing efficient gradient estimators for payoff functions with discontinuities. We have conducted our research along the following three lines.

First, we develop a simulation-based algorithm for solving a stochastic control problem with parameterized policies in revenue management. The algorithm is motivated by the fact that the combined dynamic pricing and inventory control problem in revenue management generally does not have analytical solutions, and numerical methods such as Markov Chain approximations and finite differences cannot be applied directly to solve the problem. We model the dynamic pricing problem as a stochastic control problem and give a theoretical solution for a special case when there are no constraints on the number of price changes during the selling process. For pricing policies that allow a limited number of price changes, a new algorithm is developed to simulate the evolution of the inventory level; a novel sample path based gradient estimator for stopping times with respect to parameters of interest is constructed. We also derive a new SPA gradient estimator to overcome the difficulty of differentiating a performance function with discontinuous sample paths. We have studied the unbiasedness property of the gradient estimators and the convergence of the stochastic approximation algorithm.

Second, we derive a new computationally efficient derivative estimator for payoff functions with indicator functions containing parameters of interest. Motivated by the push out method, we circumvent the difficulty of differentiating the indicator function by a change of random variables. Thus we change the underlying probability measure with

which we run simulations and move parameters of interest out of the indicator function. The support of the new random variable does not depend on any parameters of interest. Then inspired by IPA and LR, we derive a new derivative estimation technique called the Support independent unified Likelihood Ratio and Infinitesimal Perturbation Analysis (SLRPA), which has the following desirable properties:

- SLRPA connects and generalizes IPA and LR, which can be viewed as special cases of the SLRPA estimators.

- SLRPA estimators need no additional simulations, i.e., sensitivities with respect to various parameters can be obtained by a *single* run of simulation.

- SLRPA estimators work for both continuous and jump processes, are easy to implement, and have comparatively low variance.

- SLRPA estimators are unbiased, assuming some mild regularity conditions.

It is worth pointing out that Glasserman (Chapter 7 in [44]) proposed an "LR-PW" estimator to estimate the second derivative for European call options, which is obtained by first applying the LR method and then the PW method; however, SLRPA is not a simple sequential application of the LR and IPA methods and hence is essentially different from the "LR-PW" method.

Finally, we propose a new general framework called Model-base Evolutionary Optimization for solving global optimization problems. The main idea of our method is to formulate the global optimization problem as an evolutionary game and to use dynamics in evolutionary game theory to study the evolution of the candidate solutions. The process of searching for the optimal solution is carried out through the procedure of reaching the equilibrium set of an evolutionary game. Specifically, we establish a connection be-

tween evolutionary game theory and the global optimization problem by partitioning the solution region of the global optimization problem into subsets and letting different players play strategies in different subsets. Differential dynamics such as replicator dynamics are used to govern the evolution of the candidate solutions for the optimization problem. Furthermore, we introduce probabilistic models to generate candidate solutions and formulate the global optimization problem as an evolutionary game with continuous strategy spaces, based on which, a Model-based Evolutionary Optimization (MEO) algorithm is developed. Moreover, to better capture the multimodal property of global optimization problems, we propose to use a population of models to generate candidate solutions and a new Population Model-based Evolutionary Optimization (PMEO) algorithm is proposed, in which evolutionary game theory is used to study the evolution of these models. The way we formulate global optimization problems as evolutionary games provides a new insight into the mechanism for generating new candidate solutions and the mechanism of model updating for model-based global optimization algorithms. For example, one special case of the MEO algorithm gives a new explanation for the CE method. This evolutionary game setting for global optimization problems makes it possible to study the convergence property of model-based algorithms by using analytical tools in the evolutionary game theory literature, and it also provides new possibilities to develop new algorithms, for example, the PMEO algorithm we developed.

1.4 Dissertation Outline

Chapter 2 provides the necessary background and literature on simulation optimization. It starts with the formulation of stochastic optimization and stochastic control problems. Then simulation optimization techniques such as stochastic approximation and

gradient estimation are reviewed. Finally, basic concepts of evolutionary game theory are introduced.

Chapter 3 presents our research work on a simulation optimization algorithm for solving a dynamic pricing and inventory control problem in revenue management. Section 3.1 gives the introduction and motivation. Section 3.2 describes the general model and the problem formulation, and gives an analytical solution for a special model. Section 3.3 describes the simulation algorithm for a revenue management problem over a finite horizon, and presents unbiased gradient estimators to be used in the stochastic approximation algorithm. Section 3.4 gives the problem formulation for jointly optimizing the prices and the initial inventory level, and proves the convergence of the stochastic approximation algorithm. Section 3.5 gives a simulation algorithm for the revenue management problem over an infinite horizon. Section 3.6 illustrates the proposed algorithms with numerical examples. Conclusions are given in Section 3.7.

Chapter 4 presents a new stochastic derivative estimation technique for performance functions containing discontinuities. Sections 4.1 and 4.2 give a literature review and background introduction of the IPA and the LR methods. A detailed description of the new derivative estimator is given in Section 4.3. In Section 4.4, unbiasedness of the given estimators is proved and some examples are given to illustrate the application of SLRIPA in Section 4.5. Numerical results are shown in Section 4.6 followed by conclusions in Section 4.7.

Chapter 5 presents a new simulation optimization framework for solving global optimization problems by establishing a connection between global optimization and evolutionary game theory. Section 5.1 gives a brief literature review and motivation. Section 5.2 establishes a connection between evolutionary game theory and global optimization.

Section 5.3 formulates the optimization problem as an evolutionary game with a continuous strategy space, and establishes a connection between a particular equilibrium set of the replicator dynamics and global optimal solutions of the optimization problem. Based on this connection, a model-based evolutionary optimization algorithm is presented. Section 5.4 presents a population model-based evolutionary optimization algorithm by using a mixture distribution as the probabilistic model which includes a population of individual models. Section 5.5 illustrates the performance of the algorithm on some benchmark problems. Conclusions are given in Section 5.6.

Chapter 6 concludes the dissertation and outlines some future research.

Chapter 2

Preliminaries

2.1 Simulation Optimization

In the literature of simulation, simulation optimization often considers the following stochastic optimization problem:

$$\sup_{\theta \in \Theta} E[H(\theta)], \tag{2.1}$$

where Θ is the domain of the parameter θ and $H(\cdot)$ is a random variable depending on θ. This is a static stochastic optimization problem with a parameterized decision variable θ. As opposed to the static nature of the stochastic optimization problem (2.1), researchers in the control community generally consider *optimal control* problems that have a dynamic nature. In the continuous-time case, let X_t be a Markov process and $X_0 = x$, and consider the following objective function

$$J_\pi(x) = E\Big[\int_0^T e^{-rt} R(t, X_t, u_t) dt + \Phi(X_T) | X_0 = x \Big],$$

where T is the time horizon of interest and r the discount rate; $X(t)$ can be viewed as the state of the system at epoch t, which generally satisfies a stochastic differential equation; $R(t, x, u)$ is a reward function depending on state x and control u at epoch t, and $\Phi(\cdot)$ is the terminal reward; π is a control policy and $u_t = \pi(t, X_t)$ is the control at epoch t.

Then the optimal control problem can be formulated as

$$\sup_{\pi} J_{\pi}(x). \tag{2.2}$$

The corresponding performance measure in discrete-time has the form

$$J_{\mu}(x) = E\Big[\sum_{k=0}^{T-1} \gamma^k G(x_k, u_k) + \Phi(T, x_T)|x_0 = x\Big],$$

where x_k is the state at epoch k of a controlled Markov process $\{x_k\}$ and its transition probability at epoch k depends on both the state x_{k-1} at epoch $k-1$ and the action $u_k = \mu_k(x_{k-1})$, where μ_k is an element of the control policy $\mu = \{\mu_0, \mu_1, \ldots, \mu_{T-1}\}$; γ is a discounting factor, $G(\cdot, \cdot)$ is a reward function, and $\Phi(\cdot, \cdot)$ is the terminal reward. The optimal control problem of this Markov decision process can be formulated as

$$\sup_{\mu} J_{\mu}(x). \tag{2.3}$$

If the policy π in (2.2) or μ in (2.3) is a stationary parameterized policy, i.e., $\pi(t, x) = f_{\theta}(x)$ for all t and all admissible x or $\mu = \{h_{\theta}, h_{\theta}, \ldots, h_{\theta}\}$, where f_{θ} and h_{θ} are two functions parameterized by θ, the optimization problem (2.2) or (2.3) is now in the form of (2.1). Generally, the stochastic optimization problem (2.1) can be viewed as a special case of (2.2) or (2.3).

When there is finite number of choices in the parameter set Θ of (2.1), techniques in the literature of ranking and selection can be applied to solve the problem ([30]). When parameters in Θ take continuous values, typically there are three classes of algorithms for solving the static optimization problem (2.1): sample average approximation (SAA)

([64]), gradient-based algorithms, and random search. The basic idea of SAA is to approximate the expectation using the corresponding sample average function and then solve the resulting sample average optimization problem. Then existing linear programming and nonlinear optimization techniques can be applied to solve the resulting sample average optimization problem; this is one of the major advantages of SAA. Gradient-based algorithms will be discussed in the following section.

2.1.1   Gradient-based Algorithms

2.1.1.1   Stochastic Approximation

Stochastic approximation (SA) is one of the most important and popular techniques for simulation optimization. The basic underlying assumption to apply SA is that the stochastic optimization problem (2.1) can be solved by finding a zero of the gradient, i.e. by solving

$$g(\theta) = 0,$$

where $g(\theta) = \frac{\partial E[H(\theta)]}{\partial \theta}$. For functions that are not convex, this may lead only to local optimality. Let $\hat{g}(\theta)$ be an estimator of the gradient $g(\theta)$; the SA algorithm generates iterates by the formula

$$\theta_{k+1} = \Pi_\Theta(\theta_k + a_k \hat{g}(\theta_k)), \tag{2.4}$$

where θ_k is the value of the parameter at the beginning of iteration k, $\{a_k\}$ is a sequence of positive step sizes, and Π_Θ is a projection onto the parameter set Θ. To ensure the convergence of the SA algorithm, the sequence $\{a_k\}$ needs to be selected appropriately,

generally satisfying: $a_k > 0, a_k \to 0, \sum_{k=1}^{\infty} a_k = \infty$. The rate of change of the appropriately chosen sequence $\{a_k\}$ will slow down as k becomes larger; this achieves some sort of averaging effect of the noisy observations. The insight of this averaging effect obtained by the properly chosen updating sequence and the associated proofs have led to an enormous literature on general recursive stochastic algorithms.

To simplify the presentation of some convergence results, we denote the bias of the gradient estimator at iteration k by β_k: $\beta_k = g(\theta_k) - \hat{g}(\theta_k)$, where $g(\theta_k)$ is the true gradient

$$g(\theta_k) = \frac{\partial E[H(\theta)]}{\partial \theta}\bigg|_{\theta=\theta_k}.$$

Define the cumulative step size as $s_n = \sum_{i=1}^{n-1} a_i$ and define a function $\rho(s) = \max\{n : s_n \leq s\}$. To prove the convergence of the algorithm 2.4, we make the following assumptions:

Assumption 2.1 *The sequence of step sizes $\{a_k\}$ satisfies: $a_k > 0, a_k \to 0, \sum_{k=1}^{\infty} a_k = \infty$.*

Assumption 2.2 *Assume that the domain of the parameter set Θ is defined by $\{l_i(\theta) \leq 0,\ i = 1, \ldots, s\}$, where $\{l_i(\cdot)\}$ are continuously differentiable functions. At each θ that is on the boundary of Θ, the gradients of the active constraints are linearly independent.*

Assumption 2.3 *For each $\epsilon > 0$ and $s > 0$, $\lim_{n\to\infty} P(\sup_{n \leq l \leq \rho(s_k+s)} \| \sum_{i=n}^{l} a_i \beta_i \| > \epsilon) = 0$.*

Assumption 2.4 *The expectation of the value function $E[H(\theta)]$ is continuously differentiable.*

Assumption 2.5 *$\{a_n \beta_n\}$ is a bounded sequence tending to zero w.p.1.*

Theorem 6.3.1 of [67] proves the convergence of the stochastic approximation algorithm; this theorem is restated below.

Theorem 2.1 *For the optimization problem (2.1), let $\Theta^\star$ be the set of Kuhn-Tucker points. Let Assumptions 2.1, 2.2, 2.3, 2.4, and 2.5 hold for the stochastic approximation algorithm (2.4). If $\Theta^\star$ is connected, the sequence of points $\{\theta_n\}$ converges to a point in $\Theta^\star$ in probability as n goes to infinity.*

2.1.1.2 Gradient Estimation

A good gradient estimator with small bias or no bias and low variance is critical to the performance of the stochastic approximation algorithm. Gradient estimation has been studied extensively in the literature of gradient estimation, and in the following we will give a brief introduction to the existing gradient estimation techniques.

For ease of explanation, assume that the performance function $H(\theta)$ in the stochastic optimization problem (2.1) is of the form $H(X(\omega;\theta))$, where $X(\omega;\theta)$ is a random variable defined on the probability space $(\Omega_\omega, \mathcal{F}_\omega, \mathbb{P}_\omega)$, and $\theta \in \Theta \subset \mathbb{R}^n$. For example, in a queueing system, θ could be the mean arrival time and X the interarrival time. $J(\theta)$ can be written as

$$J(\theta) = \int_\Omega H(X(\omega;\theta))d\mathbb{P}_\omega(\omega). \tag{2.5}$$

Suppose we are interested in the sensitivities of $J(\theta)$ with respect to the parameters θ. We introduce three major gradient estimation methods that have been widely applied in the literature: finite differences (FD), infinitesimal perturbation analysis (IPA), and the likelihood ratio method (LR).

A. Finite Differences

The finite differences method is a "brute force" or "naive" method, which simply estimates the objective function at two different values of the parameter and then takes the difference of the estimates divided by the difference of values of the parameter to obtain a gradient estimator. The estimator could be extremely noisy if the two values of the parameter are close to each other since the output is stochastic; hence there is a trade-off between bias and variance when choosing values of the parameter. The ith component of the one-sided forward difference gradient estimator is given by

$$\frac{H(X(\omega; \theta + \Delta\theta_i e_i)) - H(X(\omega; \theta))}{\Delta\theta_i},$$

where $\Delta\theta_i$ is a perturbation of the ith element of θ, and e_i denotes the unit vector in the ith direction.

The ith component of the two-sided symmetric difference gradient estimator is given by:

$$\frac{H(X(\omega; \theta + \Delta\theta_i e_i)) - H(X(\omega; \theta - \Delta\theta_i e_i))}{2\Delta\theta_i}.$$

For each realization of the sample ω, one gradient estimate is obtained.

B. Infinitesimal Perturbation Analysis

Different from FD, IPA examines the properties of sample paths and derives unbiased gradient estimators by directly operating on the sample paths. To obtain a gradient estimator for $J(\theta)$ with respect to θ, under appropriate conditions, IPA directly differentiates

the performance function H to obtain an unbiased gradient estimator

$$\nabla_\theta J(\theta) = E_{\mathbb{P}_\omega}[\nabla_\theta H(X(\omega; \theta))], ,$$
$$= E_{\mathbb{P}_\omega}\left[\frac{dH(X)}{dX}\left(\frac{\partial X(\omega; \theta)}{\partial \theta_1}, \ldots, \frac{\partial X(\omega; \theta)}{\partial \theta_n}\right)\right], \qquad (2.6)$$

where $\nabla_\theta = (\frac{\partial}{\partial \theta_1}, \cdots, \frac{\partial}{\partial \theta_n})$; $\frac{dH(X)}{dX}$ can be understood as the derivative of H given a random variate X, the derivation of which requires some knowledge about the structure of H; $\frac{\partial X(\omega; \theta)}{\partial \theta_i}$ is the derivative of a random variable with respect to the parameter θ_i ([100, 32]). Assume $\forall \theta \in \Theta$, $X(\omega; \theta)$ is differentiable w.p.1. The derivative of the random variable X for each ω is then defined as

$$\frac{\partial X(\omega; \theta)}{\partial \theta_i} = \lim_{\Delta\theta_i \to 0} \frac{X(\omega; \theta + \Delta\theta_i e_i) - X(\omega; \theta)}{\Delta\theta_i}.$$

For example, if X is exponentially distributed with a parameter λ, i.e., $X \sim Exp(\lambda)$, X can be generated from a uniformly distributed random variable $\omega \sim U[0, 1]$ as $X = -\lambda \ln \omega$. Then an IPA estimator of X with respect to λ is given by $\frac{X}{\lambda}$.

If the distribution function $F(\cdot; \theta)$ of $X(\omega; \theta)$ is known, we have

$$\frac{\partial X(\omega; \theta)}{\partial \theta_i} = -\frac{\partial F(X; \theta)/\partial \theta_i}{\partial F(X; \theta)/\partial X},$$

the derivation of which can be found in [100]. Since ω does not depend on θ, we only need to generate one single sample path ω to get an estimate of the sensitivity for each parameter of interest. IPA explores the structure of sample paths, and the resulting gradient estimator generally has a lower variance compared with FD and LR.

The interchangeability of differentiation and integration is needed to derive (2.6),

which generally requires H to be uniformly integrable. One sufficient condition given by Glasserman ([43]) is that H is Lipschitz continuous with respect to the parameter of interest. When there are discontinuities in sample paths of X, IPA generally cannot be applied directly. Deriving efficient gradient estimators for payoff functions with discontinuities is a challenge.

C. Likelihood Ratios (LR)

In contrast to IPA, LR differentiates probability measures instead of the performance functions to construct derivative estimators and hence does not require continuity of performance functions along sample paths. To derive an LR derivative estimator for $H(X(\omega;\theta))$ with respect to θ, we assume that $\mathbb{P}_X$ is the probability measure induced by the random variable $X(\omega;\theta)$, i.e., $\mathbb{P}_X(B;\theta) = \mathbb{P}_\omega\{\omega : X(\omega;\theta) \in B\}$ for any $B \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra. By the change of variable formula [96], we have

$$\int_\Omega H(X(\omega;\theta))d\mathbb{P}_\omega(\omega) = \int H(x)d\mathbb{P}_X(x;\theta).$$

Differentiating both sides of the above equation gives

$$\nabla_\theta J(\theta) = E_{\hat{\mathbb{Q}}}\Big[H(X)\Big(\nabla_\theta \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big)\frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big]$$

$$= E_{\mathbb{Q}}\Big[H(X)\Big(\nabla_\theta \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big)\frac{d\mathbb{P}_X}{d\mathbb{Q}}(X)\Big], \tag{2.7}$$

where $\hat{\mathbb{Q}}$ is the Lebesgue measure if X is a continuous random variable and is the counting measure if X is a discrete random variable (i.e., $d\mathbb{P}_X/d\hat{\mathbb{Q}}$ corresponds to the probability density and mass functions, respectively); $\mathbb{Q}$ is a probability measure such that $\mathbb{P}_X$ is absolutely continuous with respect to $\mathbb{Q}$. $H(X)\Big(\nabla_\theta \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big)\frac{d\mathbb{P}_X}{d\mathbb{Q}}(X)$ is called the LR

derivative estimator. The LR method does not require the performance function to be continuous, and the interchangeability of differentiation and integration is often not a problem [44]. However, it is not always possible to move parameters of interest into probability density functions; morover, note that the support of the distribution of X should not depend on θ_1 when deriving (2.7). Compared with IPA, LR estimators tend to have larger variances, especially when the input process involves an oft-repeated (e.g. i.i.d) random variable whose common distribution depends on the parameter of interest.

Although IPA is constructed by differentiating the performance function, whereas LR involves differentiation of the underlying probability measure, it is possible to unify them in a single framework as in [72].

2.2   Global Optimization

We consider the following optimization problem:

$$y^\star \in \arg\max_{y \in \mathcal{Y}} H(y), \tag{2.8}$$

where the solution space $\mathcal{Y} \subset \mathbb{R}^n$ is a nonempty set. The objective function $H : \mathcal{Y} \to \mathbb{R}$ is a deterministic function bounded from above, i.e., $\exists \mathscr{U} \in \mathbb{R}$ such that $H(y) \leq \mathscr{U} \; \forall y \in \mathcal{Y}$. $y^\star$ is a global optimal solution if $H(y^\star) \geq H(y) \; \forall y \neq y^\star, \; y \in \mathcal{Y}$.

In this section, we review the model-based cross-entropy method that can be applied to solve (2.8) and evolutionary game theory.

2.2.1 Cross-Entropy (CE) Algorithms

CE was first designed to estimate the probability of rare events and then it was realized that CE can also be used to solve global optimization problems. Let $\{f(\cdot; \theta)\}$ be a family of probability density functions defined on $\mathcal{Y}$ parameterized by a parameter (vector) θ. The CE algorithm initially tries to estimate the following probability

$$l = P_\theta(H(Y) \geq \gamma), \tag{2.9}$$

where γ is a real number. Suppose $H(Y)$ has a maximum $H^\star$. When γ is close to $H^\star$, $\{H(Y) \geq \gamma\}$ is a rare event. The optimal importance sampling distribution for this rare event estimation problem is given by

$$g^\star(y) = \frac{1\{H(Y) \geq \gamma\}f(y; \theta)}{l}, \tag{2.10}$$

where $1\{\cdot\}$ is an indicator function. It is easy to see that the optimal importance sampling distribution requires the probability density (mass) concentrated on the area $\{y : H(y) \geq \gamma\}$. Since l in (2.10) is unknown, $g^\star$ does not have an analytical form. In order to solve this rare event simulation problem, CE finds the optimal $f(\cdot; \theta^\star)$ in the parameterized family $\{f(\cdot; \theta)\}$ to approximate $g^\star$. To solve the global optimization problem (2.8), CE solves a sequence of rare event estimation problem and adaptively updates $\{\gamma_k\}$, so that the resulting p.d.f/p.m.f with parameter θ_k assigns most of the probability or mass to $\{y : H(y) \geq \gamma_k\}$ and thus candidate solutions close to the global optimal solutions can be sampled with a high probability. The main CE algorithm for optimization is summarized as follows.

Algorithm 2.2.1 (Main CE algorithm for optimization) Choose $\rho \in (0,1]$ to be the fraction of samples chosen for parameter updating and let N be the number of samples generated at each iteration.

1. Choose an initial parameter $\theta_0 \in \Theta$ and set $k = 0$.

2. Generate samples $Y_1, \ldots, Y_N$ from $f(\cdot; \theta_k)$ and compute the $1 - \rho$ quantile of the performance γ_k by ordering $\{H(Y_i)_{i=1}^N\}$ from smallest to largest.

3. Update the parameter θ_{k+1} by solving the optimization problem

$$\theta_{k+1} := \arg\max_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} 1\{H(Y_i) \geq \gamma_k\} \ln f(Y_i; \theta).$$

4. If for some $k > d$, say $d = 5$,

$$\gamma_k = \gamma_{k-1} = \ldots = \gamma_{k-d},$$

then stop; otherwise, set $k = k + 1$ and reiterate from step 2.

2.2.2   Evolutionary Games

In this section, we give a brief introduction to game theory and evolutionary game theory. Consider a simple two-player game. A is the payoff matrix for player I and B is the payoff matrix for player II. Player I has the pure strategy set $S_1 = \{1, \ldots, n\}$, and $S_2 = \{1, \ldots, n\}$ is the pure strategy set for player II. If player I plays the pure strategy i and player II plays j, player I receives a_{ij}, the (i,j) element of A, and player II receives b_{ij}, the (i,j) element of B. The mixed strategy of player I is a probability vector $x = (x_1, \ldots, x_n)^T$, where x_i is the probability of choosing strategy $i \in S_1$. Similarly, B

is the payoff matrix for player II with mixed strategy $y = (y_1, \ldots, y_n)^T$. If the game is symmetric, we have $B^T = A$. The expected payoff for player I and player II will be $x^T A y$ and $x^T B y$, respectively.

Evolutionary game theory studies the game in a population setting. Assume there is a population of agents which are programmed to play n different pure strategies in the set $\{1, \ldots, n\}$ and let x_i be the percentage of agents playing pure strategies i for $i \in \{1, \ldots, n\}$ in the population. We assume that x_i is a differentiable function of time t. If individuals meet randomly and then engage in a symmetric game with a payoff matrix A, then $(Ax)_i$ is the expected payoff for an individual playing strategy i and $x^T A x$ is the payoff of an agent that is randomly selected from the population, which is also the population average payoff. If the payoff of the individual playing i is greater than the population average payoff, the number of agents playing i will increase. Assume that the per capita rate of growth, i.e. the logarithmic derivative $(\ln x_i) := \dot{x}_i / x_i$, is given by the difference between the payoff for type i and the average payoff in the population. This yields the replicator equation ([107])

$$\dot{x}_i = x_i((Ax)_i - x^T A x) \qquad \forall i = 1, \ldots, n.$$

Replicator dynamics is a selection process, according to which, more successful strategies will spread in the population.

2.2.3   Other Dynamics

Besides replicator dynamics, there are other dynamics ([55]), and we introduce three as follows.

2.2.3.1 Imitation Dynamics

The imitation dynamics is given by $\dot{x}_i = x_i \sum_j [\phi_{ij}(x) - \phi_{ji}(x)] x_j$, where ϕ_{ij} is the rate at which an agent playing the strategy j adopts the strategy i. One plausible assumption is that this rate depends only on the payoffs achieved by the two agents, i.e., $\phi_{ij}(x) = \phi(f_i, f_j)$, where f_i, f_j are payoff functions and $\phi(u, v)$ defines the imitation rule (the same for all players). The simplest rule is to imitate the better, i.e.,

$$\phi(u, v) = \begin{cases} 0 & \text{if } u \leq v, \\ 1 & \text{if } u > v. \end{cases}$$

In this case, the percentage of agents playing a strategy increases if and only if its payoff is larger than the median of the payoffs of all the strategies.

2.2.3.2 Best Response and Logit Dynamics

Unlike imitation dynamics, best response dynamics requires more than imitating the better agent, and chooses the best reply to the current mean population strategy x:

$$\dot{x} \in BR(x) - x,$$

where $BR(x)$ is the set of best replies to x ([55]). The smoothed version of the best response dynamics-logit dynamics is given as follows:

$$\dot{x}_i = \frac{e^{f_i/\epsilon}}{\sum_j e^{f_j/\epsilon}} - x_i$$

with $\epsilon > 0$. As $\epsilon \to 0$, this converges to the best reply dynamics.

2.2.3.3 The Brown-von Neumann-Nash Dynamics

The Brown-von Neumann-Nash dynamics (BNN), is defined by

$$\dot{x}_i = k_i(x) - x_i \sum_{j=1}^{M} k_j(x),$$

where $k_i(x) = \max(0, f_i - \sum_{j=1}^{M} x_j f_j)$ denotes the positive part of the excess payoff for the strategy i. The discrete-time version of the above dynamics is given by

$$x_i(t+1) = \frac{x_i(t) + k_i(x(t))}{1 + \sum_{j=1}^{M} k_j(x(t))}, \quad \forall i = 1, \ldots, M.$$

Chapter 3

Simulation-based Algorithm for Dynamic Pricing and Inventory

Control

3.1 Introduction and Motivation

In this chapter we consider a joint optimization of dynamic pricing and inventory control problem and focus on developing a simulation-based algorithm to solve this problem. As mentioned in Chapter 1, in the literature of revenue management, there are two different types of demand models: discrete-time models and continuous-time models. For discrete-time models, dynamic pricing research works include [27, 24, 21]. Related works using continuous Brownian demands include [85, 7, 19, 20], and [5].

All these works using continuous Brownian demand models focus on inventory control and try to characterize the optimal ordering policy, whereas works addressing the problem of *jointly* optimizing pricing and ordering policies under Brownian demands are relatively rare. [86] analyzed a dynamic pricing problem with demand uncertainty modeled by a Brownian motion without incorporating ordering decisions. For a price-sensitive demand modeled by continuous diffusion processes, the dynamic pricing problem can be formulated as a stochastic control problem. For a class of pricing policies that allow a limited number of price changes, for example the markdown pricing policy, the problem of jointly optimizing the prices and the initial inventory level can be formulated as a continuous stochastic control problem with the initial ordering as a decision variable.

The aforementioned stochastic control problem generally does not admit an analyti-

cal solution. [22] considered a special continuous demand model, where the arrival rate and the uncertainty term in the demand only depend on the price. In their formulation, the dynamic pricing problem is transformed into a deterministic optimization problem and is solved numerically. But their method cannot be generalized to solve dynamic pricing problems with general demand models. There are two major types of numerical methods for stochastic control problems ([83]). Purely deterministic approximation methods discretize the HJB equation by finite differences or a finite element method, and thus obtain an approximate value function at points on the space-time grid. Probabilistic methods include the Markov chain approximation method presented by [70]. Markov chain approximation and finite differences cannot be applied directly to the dynamic pricing problem with a limited number of price changes because of the constraint on the pricing policy. Moreover, it is desirable to understand how the uncertainties in the demand and the changing of other parameters, such as parameters in the arrival rate and holding cost, would affect the pricing policy and the profit. This sensitivity information cannot be obtained directly using Markov chain approximation.

In the literature of supply chain management, simulation-based algorithms have been applied to solve inventory control problems ([31, 34, 45, 66, 60]) and network revenue management problems ([9, 104]). In simulation-based algorithms, the key is to obtain efficient gradient estimations. [9] proposed an innovative simulation-based optimization method for computing protection levels in a virtual nesting control scheme, in which a finite difference estimator is derived, because a discrete model for capacity and demand is used. [104] analyzed a continuous version of the problem, which enabled the derivation of a gradient estimator using infinitesimal perturbation analysis (IPA). However, all these methods can only be applied to discrete-time models.

Introduced by [52], IPA has been widely used in sensitivity analysis for discrete-event systems. IPA enables the sensitivity of a performance function to be estimated while observing a *single* sample path of a system and hence offers significant computational savings compared to "brute force" finite difference methods. IPA generally requires the performance function to be continuous, which makes IPA not applicable in many cases. By using conditional expectation, smoothed perturbation analysis (SPA) ([49, 37]) can sometimes "smooth" out the discontinuities in sample paths. Rather than differentiating a performance function as IPA does, the likelihood ratio (LR) method constructs derivative estimators from the derivatives of the probability measure associated with a simulation model. The method was proposed by [48], [87], and [91] to study discrete-event systems and has also been used in financial applications ([44]).

Motivated by the fact that the combined dynamic pricing and inventory control problem generally does not have an analytical solution, and numerical methods such as Markov Chain approximation and finite differences cannot be used to solve the problem effectively, we develop a simulation-based algorithm for solving this stochastic control problem in revenue management.

3.2   General Model

We consider a problem in which a vendor wants to sell a single item with an initial inventory level $X_0 = S$, and no inventory is ordered during the selling process. The price of the item can be adjusted dynamically over time. The selling process will continue until the inventory is depleted for an infinite horizon problem; for a finite horizon problem, the selling process stops before T, where T is a positive constant. We assume that the demand

depends on the price, and the cumulative demand up to time t follows a diffusion process:

$$D_t = \int_0^t \lambda(p_s)ds + \int_0^t \sigma(D_s, p_s)dB_s, \qquad t \geq 0, \tag{3.1}$$

where p_s is the price at time s, $\lambda(p)$ is the demand at price p; $\sigma(\cdot, \cdot)$ is a measure of the volatility, and $\{B_s\}$ is a standard Brownian motion. The differential form of the demand is given by: $dD_t = \lambda(p_t)dt + \sigma(D_t, p_t)dB_t$. In our model, D_t might be negative because of the Brownian motion, which can be interpreted as return of the product. Here we assume $\sigma(\cdot, \cdot)$ has a general form to emphasize wider applicability. Let X_t denote the inventory level at time t, given by $X_t = X_0 - D_t$, where X_0 is the initial inventory level. X_t satisfies the stochastic differential equation

$$dX_t = -dD_t = -\lambda(p_t)dt - \sigma(X_0 - X_t, p_t)dB_t. \tag{3.2}$$

Assume that the holding cost per unit time is $c(x)$ for an inventory level x. The epoch when all the inventory is sold is a stopping time, defined by $\tau = \inf\{t : X_t = 0\}$. We adjust the prices over time to sell the inventory over the horizon of interest $[0, T]$, with the goal of maximizing the expected total profit, which is given by $E\left[\int_0^{\tau \wedge T} p_s dD_s - c(X_s)ds\right]$, where $\tau \wedge T = \min\{\tau, T\}$, $p_s dD_s$ is the revenue obtained when dD_s inventory is sold, $c(X_s)ds$ is the holding cost associated with the inventory X_s for the time period ds, and T is a fixed constant for a finite horizon problem and ∞ for an infinite horizon problem. Here, we begin by considering a Markov control policy $\pi : [0, T] \times \mathbb{R}^+ \to \mathcal{P}$ such that the price at time t is given by $p_t = \pi(t, X_t)$, where $\mathcal{P} \subset \mathbb{R}^+$ is the domain of the price.

3.2.1 Dynamic Pricing Over a Fixed Time

With a Markov control policy π, the total expected profit starting from time t with inventory x is

$$
\begin{aligned}
J_\pi(t,x) =& E\Big[\int_t^{\tau \wedge T} -p_s dX_s - c(X_s)ds \Big| X_t = x \Big] \\
=& E\Big[\int_t^{\tau \wedge T} p_s[\lambda(p_s)ds + \sigma(X_t - X_s, p_s)dB_s] - c(X_s)ds \Big| X_t = x \Big] \\
=& E\Big[\int_t^{\tau \wedge T} [p_s \lambda(p_s) - c(X_s)]ds \Big| X_t = x \Big].
\end{aligned}
$$

The third equality above follows since $E[\int_t^T \sigma(X_t - X_s, p_s)dB_s | X_t = x] = 0$. The dynamic pricing problem can be formulated as

$$
\sup_\pi J_\pi(0, S). \tag{OPT1}
$$

Denote $V(t,x) = \sup_\pi J_\pi(t,x)$. The Hamilton-Jacobi-Bellman (HJB) equation for the optimization problem OPT1 is given by

$$
\sup_{p_s}[\mathcal{L}V(s, x_s) + p_s \lambda(p_s) - c(x_s)] = 0, \tag{3.3}
$$

with boundary conditions $V(T,x) = 0 \, \forall \, x \in [0, S]$, and $V(t,0) = 0 \, \forall t$, where

$$
\mathcal{L} \equiv \frac{\partial}{\partial t} + (-\lambda(p_s))\frac{\partial}{\partial x} + \frac{1}{2}[\sigma(X_0 - x_s, p_s)]^2 \frac{\partial^2}{\partial x^2}
$$

.

34

3.2.2 Dynamic Pricing Over Infinite Horizon

Consider the dynamic pricing problem with a selling period over an infinite horizon.
Given a pricing policy π, the total expected profit is

$$J_\pi(S) = E\Big[\int_0^\tau p_s(-dX_s) - c(X_s)ds|X_0 = S\Big]$$
$$= E\Big[\int_0^\tau [p_s\lambda(p_s) - c(X_s)]ds|X_0 = S\Big].$$

Then the dynamic pricing problem can be formulated as

$$\sup_\pi J_\pi(S). \tag{OPT2}$$

The value function V is defined as $V(x) = \sup_\pi J_\pi(x)$. For the optimization problem
OPT2, the HJB equation is

$$\sup_{p\in\mathcal{P}}[\mathcal{L}V(x) + p\lambda(p) - c(x)] = 0, \tag{3.4}$$

with the boundary condition $V(0) = 0$, where $\mathcal{L} \equiv (-\lambda(p))\frac{\partial}{\partial x} + \frac{1}{2}[\sigma(X_0 - x, p)]^2\frac{\partial^2}{\partial x^2}$.

The HJB equations (3.3) and (3.4) are difficult to solve, and there is no closed-form
solution in general. We give an analytical solution for a special case of the infinite horizon
problem in the following, in which $\lambda(p) = a - bp$, $\sigma(y, p) = \vartheta p\sqrt{y}$, and the cost function
$c(x) = rx^2$.

Theorem 3.1 *For the demand model with $\sigma(y, p) = \vartheta p\sqrt{y}$, arrival rate $\lambda(p) = a - bp$,*
and holding cost $c(x) = rx^2$, where a, b, ϑ, r are some positive constants, if the following

condition holds,

$$b + \frac{1}{4b} - \vartheta^2 > 0,$$

the optimal pricing policy for the problem (OPT2) is $p^\star = \frac{2b\alpha_2 x + b\alpha_1}{2\vartheta^2 \alpha_2 x + 2b - 2\vartheta^2 X_0}$ *and the optimal value function is* $V(x) = \alpha_2 x^2 + \alpha_1 x$, *where*

$$\alpha_1 = \frac{4ab\alpha_2 - 2\vartheta^4 \alpha_2 X_0}{4b^2}, \quad \alpha_2 = \sqrt{\frac{r}{b + \frac{1}{4b} - \vartheta^2}}.$$

Proof. We assume that the optimal function takes the form $V(x) = \alpha_2 x^2 + \alpha_1 x$; then we plug this value function into the HJB equation (3.4) and take the derivative with respect to p, which is set to zero to get the optimal p^*. Substituting this p^* back into (3.4) and letting all the coefficients of x and x^2 be zero gives the result. It is easy to check that the second derivative condition also holds.

3.3 Pricing with Fixed Number of Price Changes Over Finite Horizon

As mentioned in Section 3.2.2, the HJB equations are difficult to solve, and there is no analytical solution for general stochastic control problems except for some problems with special structure like the linear stochastic quadratic control problem. Numerical methods such as Markov chain approximation and finite difference may give us very good approximate results, but in the resulting pricing policy, prices could change frequently with the changing of the inventory level, and this may not be practical. Henceforth, we consider a class of pricing policies that only allow a limited number of price changes, which includes the well-known markdown pricing policy.

3.3.1 Formulation of the Optimization Problem

Consider selling the inventory over a finite time interval $[0, T]$. Let N be the maximal number of times that the price can be changed; let $S = S_0 > S_1 > \cdots > S_{N-1} > S_N = 0$ be fixed inventory levels, at which the price changes. Define the first time when the inventory drops to S_k as $\tau_k = \inf\{t \geq 0 : X_t = S_k\}$. As shown in Figure 3.1, before the inventory level falls to S_1 at τ_1, the price p_1 is charged; the price p_2 is then charged until the inventory drops to S_2 at τ_2; finally, the price p_N is charged until the inventory drops to $S_N = 0$. If all the inventory cannot be sold before T, the selling process stops at T. This type of policy includes the markdown pricing policy that is widely applied in high-tech and perishable products retailing ([101]). Let $\pi_f = \{p_k, k = 1, \cdots, N\}$ denote this non-Markov pricing policy; we try to find $\{p_k\}$ to maximize the total expected profit, given by

$$V_{\pi_f}(S) = E\Big[\sum_{k=1}^{N} \int_{\tau_{k-1}}^{\tau_k \wedge T} -p_k dX_s - c(X_s)ds | X_0 = S\Big], \qquad (3.5)$$

where we define $\tau_0 = 0$. The dynamic pricing problem can be formulated as

$$\sup_{\pi_f} V_{\pi_f}(S). \qquad (3.6)$$

To simplify the notation, we denote V_{π_f} by V henceforth and define the incremental value function by

$$\Delta V_k = E\Big[\int_{\tau_{k-1}}^{\tau_k \wedge T} -p_k dX_s - c(X_s)ds | X_{\tau_{k-1}} = S_{k-1}\Big]. \qquad (3.7)$$

There is no analytical solution for the above optimization problem (3.6) in general. We propose a simulation-based method to solve this problem. We first discretize the stochastic differential equation (3.2) and then present an algorithm to simulate the inventory level and estimate the stopping times $\{\tau_k\}$, from which we define discretized incremental value functions to estimate the incremental value function $\{\Delta V_k\}$. Second, we derive gradient estimators for the expectation of the discretized value function with respect to the prices $\{p_k\}$. Finally, we use stochastic approximation to find the optimal prices $\{p_k^\star\}$.
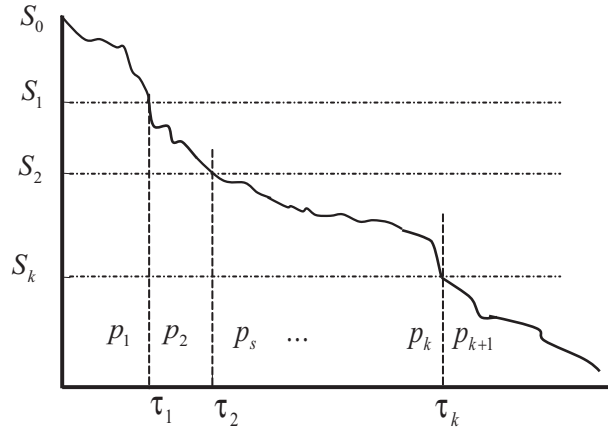
Figure 3.1: Pricing policy

3.3.2 Estimate the Value Function

In this section, we first use Euler discretization to discretize the stochastic differential equation (3.2), describe how to simulate the inventory level, and define the discretized value function. Then we analyze the properties of a stopping time and show how to estimate the conditional expectation of the stopping time in the discretized value function. Finally, we give a simulation algorithm to estimate the discretized value function.

3.3.2.1 The Inventory Level

To simulate the evolution of the inventory level given a pricing policy π_f, we discretize time with a fixed time step h. Set $\tau_0 = 0$ and denote the time when price p_k is first charged for $k \geq 1$ by $t_0^k = \tau_{k-1}$, where we use the superscript k in t_0^k to indicate that price p_k is active. Define $t_i^k = \tau_{k-1} + ih$ as the discretized time, with the corresponding inventory level denoted by X_i^k, which follows the Euler discretization of the stochastic differential equation (3.2):

$$X_{i+1}^k = X_i^k - \lambda(p_k)(t_{i+1}^k - t_i^k) - \sigma(X_0 - X_i^k, p_k)\sqrt{t_{i+1}^k - t_i^k}\, Z_{i+1}^k, \tag{3.8}$$

for $\tau_{k-1} \leq t_i^k < \tau_k$, where $\{Z_i^k\}$ are independent and identically distributed standard normal random variables. Denote a realization of X_i^k by x_i^k. Sample paths of the inventory level are generated as follows:

Start from $x_0^1 = S_0$ and generate x_1^1 using equation (3.8) with z_1^1 generated from a standard normal distribution. Then we want to decide if the inventory crosses S_1, i.e., if a stopping time occurs in $[t_0^1, t_1^1]$. Consider two different possibilities $x_1^1 > S_1$ and $x_1^1 \leq S_1$.

If $x_1^1 > S_1$, by the properties of Brownian motion, the inventory level crosses S_1 in $[t_0^1, t_1^1]$ with a probability $1 - \eta_1^1$, where η_1^1 defined later is a positive number depending on x_0^1, x_1^1 and S_1. To decide if a stopping time occurs in $[t_0^1, t_1^1]$, generate a random number $U_1^1 \sim U[0,1]$. If $U_1^1 < \eta_1^1$, the inventory level does not cross S_1, and then we continue to generate sample points of the inventory using equation (3.8) and check if a stopping time occurs. If $U_1^1 \geq \eta_1^1$, a stopping time denoted by τ_1^1 occurs in $[t_0^1, t_1^1]$. Then starting from τ_1^1, price p_2 is charged and more sample points are generated starting from τ_1^1 with a fixed time step h similarly.

If $x_1^1 \le S_1$, a stopping time denoted by τ_1^1 must have occurred in $[t_0^1, t_1^1]$. Then starting from τ_1^1, price p_2 is charged and more sample points are generated starting from τ_1^1 using equation (3.8) with a fixed time step h similarly.

A more detailed description of the sample path generation algorithm is presented below:

Step 0: Initialization. Set $i = 0, k = 1, x_0^1 = S_0$, and $t_0^1 = 0$.

Step 1: Set $t_{i+1}^k = t_i^k + h$. If $t_{i+1}^k + h < T$, generate x_{i+1}^k at t_{i+1}^k using equation (3.8) with z_{i+1}^k generated from a standard normal distribution; otherwise generate x_{i+1}^k by

$$x_{i+1}^k = x_i^k - \lambda(p_k)(T - t_i^k) - \sigma(S - x_i^k, p_k)\sqrt{T - t_i^k}\, z_{i+1}^k. \qquad (3.9)$$

Given x_i^k at t_i^k and x_{i+1}^k at t_{i+1}^k, we want to decide if the inventory crosses S_k in $[t_i^k, t_{i+1}^k]$, i.e., if a stopping time occurs in $[t_i^k, t_{i+1}^k]$. Consider two different possibilities $x_{i+1}^k > S_k$ and $x_{i+1}^k \le S_k$.

Step 2–Case 1: $x_{i+1}^k > S_k$. Define the minimum of x_t between time t_i^k and t_{i+1}^k given x_i^k and x_{i+1}^k by $\ell_{i+1}^k = \inf_{t_i^k < t \le t_{i+1}^k} x_t$, and define $\ell_0^k = S_{k-1}$ in particular. By [61], we have $P\left(\ell_{i+1}^k > S_k \big| x_i^k, x_{i+1}^k\right) = \eta_{i+1}^k(S_k, x_i^k, x_{i+1}^k)$, where

$$\eta_{i+1}^k(S_k, x_i^k, x_{i+1}^k) = 1 - \exp\left(-\frac{2(x_i^k - S_k)(x_{i+1}^k - S_k)}{(t_i - t_{i-1})[\sigma(X_0 - x_i^k, p_k)]^2}\right), \qquad x_i^k > S_k,\ x_{i+1}^k > S_k.$$

To decide if a stopping time occurs in $[t_i^k, t_{i+1}^k]$, generate $U_{i+1}^k \sim U[0, 1]$. If $U_{i+1}^k < \eta_{i+1}^k$, the inventory does not cross S_k; set $i = i + 1$, and return to Step 1. If $U_{i+1}^k \ge \eta_{i+1}^k$, the

inventory may cross S_k in $[t_i^k, t_{i+1}^k]$. Define

$$\tau_{i+1}^k = \min\{T, \inf\{t : x_t = 0, \text{ given } x_i^k \text{ and } x_{i+1}^k, \text{ and } \ell_{i+1}^k < 0\}\}. \qquad (3.10)$$

Set $\tau_k = \tau_{i+1}^k$, where the index $i+1$ is used to indicate that the stopping time occurs in $[t_i^k, t_{i+1}^k]$. Charge price p_{k+1} starting from τ_k. Define $t_0^{k+1} = \tau_k$, $x_0^{k+1} = S_k$, set $k = k+1, i = 0$, and return to Step 1.

Case 2: $x_{i+1}^k \leq S_k$. The inventory must have crossed S_k in $[t_i^k, t_{i+1}^k]$ and the first time that the inventory crossed S_k is defined by

$$\tau_{i+1}^k = \min\{T, \inf\{t : x_t = 0, \text{ given } x_i^k \text{ and } x_{i+1}^k \leq S_k\}\}.$$

Set $\tau_k = \tau_{i+1}^k$ and charge price p_{k+1} starting from τ_k. Define $t_0^{k+1} = \tau_k$, $x_0^{k+1} = S_k$, set $k = k+1, i = 0$, and return to Step 1.

To simplify the problem, define $v_k = \min\{i : x_i^k > S_k, x_{i+1}^k \leq S_k\}$, which indicates the first time interval in which the inventory level must have crossed S_k. Similarly, define $n_e = \min\{k : t_i^k < T \leq t_{i+1}^k \text{ for some } i\}$ and $v_e = \min\{i : t_i^{n_e} < T \leq t_{i+1}^{n_e}\}$. It is easy to see that when the selling stops between t_i^k and t_{i+1}^k, n_e is the index k of the active price p_k and v_e is the index i of time t_i^k, which indicate the time interval in which the selling stops and the index of the corresponding active price, respectively.

3.3.2.2 Discretized Value Function

Given a pricing policy π_f, the discretized incremental value function $\Delta \hat{V}_k$ corresponding to (3.7) computed along a sample path is given by

$$
\Delta \hat{V}_k = - p_k \Delta S_k - \sum_{m=0}^{v_k} \left\{ \left[\sum_{i=0}^{m-1} c(x_i^k) h + c(x_m^k) \left(E[\tau_{m+1}^k | x_m^k, x_{m+1}^k] - t_m^k \right) \right] \right.
$$
$$
\left. \times \prod_{j=0}^{m} I_{\{U_j^k < \eta_j^k\}} I_{\{U_{m+1}^k \geq \eta_{m+1}^k\}} \right\}, \tag{3.11}
$$

where $\Delta S_k = S_k - S_{k-1}$ is the change of the inventory level when p_k is active for $k < n_e$, and $\Delta S_{n_e} = x_{v_e+1}^{n_e} \prod_{j=0}^{v_e} I_{\{U_j^{n_e} < \eta_m^{n_e}\}} I_{\{U_{v_e+1}^{n_e} < \eta_{v_e+1}^{n_e}\}} I_{\{x_{v_e+1}^{n_e} \geq 0\}} - S_{n_e-1}$ is the change of the inventory level when p_{n_e} is active; for simplicity, define $\eta_0^k = 1$ and let $\sum_{i=0}^{m-1} c(x_i^k) h = 0$ for $m = 0$. The first part of the right hand side of $\Delta \hat{V}_k$ is the revenue obtained by selling $-\Delta S_k$ amount of inventory, and the second part is the holding cost; the product of indicator functions indicates the time interval in which the price changes.

Along a sample path, the discretized value function corresponding to (3.5) can be computed by $\hat{V} = \sum_{k=1}^{n_e} \Delta \hat{V}_k$. The dynamic pricing problem over a finite horizon (3.6) can be transformed to:

$$
\sup_{\pi_f} E\left[\hat{V}(S) \right], \tag{3.12}
$$

where $\pi_f = \{p_k, k = 1, \cdots, N\}$ is the pricing policy, and the expectation is taken under the policy π_f. Stochastic approximation will be used to solve this optimization problem later. In the following sections, we first analyze how to compute the conditional expectation of a stopping time in (3.11) that is needed to estimate $E[\hat{V}(S)]$ and give a simulation algorithm to estimate $E[\hat{V}(S)]$. Then we derive sample path-based gradient estimators for $E[\hat{V}(S)]$

with respect to pricing parameters $\{p_i, i = 1, \cdots, N\}$.

3.3.2.3   Estimating the Expectation of Stopping Times

To estimate the expectation of $\hat{V}(S)$ by using Monte Carlo simulation requires estimating the expectation of the stopping time τ_{i+1}^k given two end points. Given x_i^k and x_{i+1}^k at time t_i^k and t_{i+1}^k, denote the conditional probability density of τ_{i+1}^k by $f_k(t) = P(\tau_{i+1}^k \in [t, t + dt] | x_i^k, x_{i+1}^k)$, which is given by ([78]):

$$
\begin{aligned}
f_k(t) = {} & \frac{(x_i^k - S_k)}{\sqrt{2\pi(t - t_i^k)^3}[\sigma(X_0 - x_i^k, p_k)]^2} \sqrt{\frac{h}{t_{i+1}^k - t}} \\
& \times \exp\left\{ -\frac{[(x_i^k - x_{i+1}^k)(t - t_i^k) - (x_i^k - S_k)h]^2}{2h(t - t_i^k)(t_{v_k+1}^k - t)[\sigma(X_0 - x_i^k, p_k)]^2} \right\}.
\end{aligned}
\tag{3.13}
$$

Generating from this density can be accomplished by generating inverse Gaussian random variables as we now describe in the two possible cases.

Case 1. $X_i^k > S_k$ **and** $X_{i+1}^k < S_k$**:**

The probability density function of an inverse Gaussian random variable Y is given as

$$
g_k(y; \lambda_k, \theta_k) = \sqrt{\frac{\lambda_k}{2\pi y^3}} \exp\left(-\frac{\lambda_k(y - \theta_k)^2}{2\theta_k^2 y} \right),
\tag{3.14}
$$

where θ_k is the mean and λ_k is the shape parameter. Notice that $f_k(t)$ is the probability density function of a stopping time of a Brownian motion conditioned on two given end points, whereas the inverse Gaussian distribution describes the distribution of the time it takes for a Brownian motion to reach a certain level. It is easy to show that if a random variable Y has the density $g_k(y; \lambda_k, \theta_k)$ with $\theta_k = -\frac{x_i^k - S_k}{x_{i+1}^k - S_k}$ and $\lambda_k = \frac{(x_i^k - S_k)^2}{h[\sigma(X_0 - x_i^k, p_k)]^2}$, then

$t_i^k + \frac{hY}{1+Y}$ has the probability density function (3.13). Then we have

$$E[\tau_{i+1}^k | x_i^k, x_{i+1}^k] = \int_{t_i^k}^{t_{i+1}^k} t \frac{f_k(t)}{d} t = \int_0^\infty y g_k(y; \lambda_k, \theta_k) dy = E\left[t_i^k + \frac{hY}{1+Y}\right], \qquad (3.15)$$

where Y is an inverse Gaussian random variable that has the density $g(y; \lambda_k, \theta_k)$ with $\theta_k = -\frac{x_i^k - S_k}{x_{i+1}^k - S_k}$ and $\lambda_k = \frac{(x_i^k - S_k)^2}{h[\sigma(X_0 - x_i^k, p_k)]^2}$.

Case 2. $X_i^k > S_k$ **and** $X_{i+1}^k > S_k$**:**

In this case, by the definition of τ_{i+1}^k in (3.10), τ_{i+1}^k is well defined only when $\ell_{i+1}^k < S_k$, i.e., the minimum of the inventory level is less than S_k between time t_i^k and t_{i+1}^k. Conditioned on $\ell_{i+1}^k < S_k$, we have

$$E[\tau_{i+1}^k | x_i^k, x_{i+1}^k] = \int_{t_i^k}^{t_{i+1}^k} t \frac{f_k(t)}{P(\ell_{i+1}^k < S_k)} dt = \int_0^\infty y \frac{g_k(y; \lambda_k, \theta_k)}{P(\ell_{i+1}^k < S_k)} dy = E\left[t_i^k + \frac{h\bar{Y}}{1+\bar{Y}}\right],$$

$$(3.16)$$

where $\bar{Y}$ is an inverse Gaussian random variable that has the density $g(y; \lambda_k, -\theta_k)$. Therefore we can obtain unbiased estimates of the expectation of τ_{i+1}^k by generating inverse Gaussian random variates. Inverse Gaussian random variates can be generated by using the algorithm in [79].

3.3.2.4 Simulation Algorithm

We use the following algorithm to simulate the inventory level and to generate inverse Gaussian random variates that are used to estimate the expectation of the stopping time. By following this algorithm, an unbiased estimate of $E[\hat{V}(S)]$ can be obtained. When the discretized step size goes to 0, we have $E[\lim_{h \to 0} \hat{V}(S)] = E[V(S)]$.

Algorithm 1. _(Simulation Algorithm Over Finite Horizon)_

Initialization: For a given pricing policy π_f, choose a time step $h > 0$. Let $k = 1, t_0^1 = 0$, $I_0 = 1$, and $x_0^1 = S_0 = S$. Let $\tau_0 = 0$.

- step 1: Set $i = 0$, $t_0^k = \tau_{k-1}$ for $k > 1$, $I_k = 1$, and $x_0^k = S_{k-1}$. Go to step 2.

- step 2: Let the price be p_k according to the pricing policy π_f and generate $z_{i+1}^k \sim N(0,1)$. Let $t_{i+1}^k = t_i^k + h$.

  – When $t_{i+1}^k + h < T$, compute x_{i+1}^k by (3.8).

    * If $x_{i+1}^k \geq S_k$, generate $U_i^k \sim U(0,1)$, and compute $\eta_{i+1}^k(S_k, x_i^k, x_{i+1}^k)$. If $I_k = 1$, let $I_k = I_k I_{\{U_i^k < \eta_{i+1}^k\}}$, $\theta_k = \frac{x_i^k - S_k}{x_{i+1}^k - S_k}$, $\lambda_k = \frac{(x_i^k - S_k)^2}{h[\sigma(S - x_i^k, p_k)]^2}$, $i = i + 1$, then go to step 3;

    * If $x_i^k > S_k > x_{i+1}^k$, let $\upsilon_k = i$, $\theta_k = -\frac{x_i^k - S_k}{x_{i+1}^k - S_k}$, and $\lambda_k = \frac{(x_i^k - S_k)^2}{h[\sigma(S - x_i^k, p_k)]^2}$, then go to step 3.

  – If $t_i^k + h < T \leq t_{i+1}^k + h$, let $\upsilon_e = i, n_e = k$, and compute x_{i+1}^k by equation (3.9).

    * If $x_{\upsilon_e+1}^k > 0$, generate $U_i^k \sim U(0,1)$, and compute $\eta_{i+1}^k(0, x_{\upsilon_e}^k, x_{\upsilon_e+1}^k)$. Let $\theta_k = \frac{x_i^k}{x_{i+1}^k}$ and $\lambda_k = \frac{(x_i^k)^2}{h[\sigma(S - x_i^k, p_k)]^2}$.

    * If $x_{\upsilon_e+1}^k \leq 0$, let $\theta_k = -\frac{x_i^k}{x_{i+1}^k}$ and $\lambda_k = \frac{(x_i^k)^2}{h[\sigma(S - x_i^k, p_k)]^2}$.

    * Generate a sample y from (3.14). Let $\tau_{i+1}^k = t_i^k + \frac{hy}{1+y}$ and stop.

- step 3: Generate a sample y from (3.14). Let $\tau_{i+1}^k = t_i^k + \frac{hy}{1+y}$.

  – When $I_k = 1$, if $X_{i+1}^k > S_k$, the price does not change and go to step 2; if $X_{i+1}^k \leq S_k$, the price changes, let $\tau_k = \tau_{i+1}^k$, $k = k + 1$, and go to step 1.

  – When $I_k = 0$, the price changes, let $\tau_k = \tau_{i+1}^k$, and go to step 1.

45

3.3.3   Gradient Estimation

In this section, we derive gradient estimators for $E[\hat{V}(S)]$ with respect to the pricing parameters $\{p_k\}$, which involve the gradient estimators of the inventory level, and the stopping times. To circumvent the difficulty of differentiating indicator functions in $\hat{V}(S)$, smoothed perturbation analysis ([37]) is used. Before deriving the gradient estimators, we make the following assumptions.

Assumption 3.1 $\lambda(\cdot)$ *is differentiable, and $\sigma(\cdot,\cdot)$ is differentiable in both its arguments; $\lambda(\cdot)$ is Lipschitz continuous and $\sigma(\cdot,\cdot)$ is Lipschitz continuous in each argument, i.e., there exist positive constants K_λ, K_{g1}, and K_{g2} such that*

$$\|\lambda(p') - \lambda(p)\| \leq K_\lambda \|p' - p\|,$$

$$\|\sigma(X',p) - \sigma(X,p)\| \leq K_{g_1}\|X' - X\|,$$

$$\|\sigma(X,p') - \sigma(X,p)\| \leq K_{g_2}\|p' - p\|.$$

Assumption 3.2 *The prices $\{p_k\}$ are bounded from above and below, i.e. $p_k \in [\underline{p}, \bar{p}]\,\forall k = 1, \cdots, N$, where $\underline{p}$ and $\bar{p}$ are positive constants, and $\underline{p} < \bar{p}$; the arrival rate $\lambda(p)$ lies in the interval $[\underline{\lambda}, \bar{\lambda}]$ and $\sigma(\cdot,\cdot)$ belongs to $[\underline{\sigma}, \bar{\sigma}]$, where $\underline{\lambda}, \bar{\lambda}, \underline{\sigma},$ and $\bar{\sigma}$ are positive constants and $\underline{\lambda} < \bar{\lambda}, \underline{\sigma} < \bar{\sigma}$. The cost function $c(x)$ is a polynomial function of x.*

Assumption 3.1 is very common and is satisfied by many practical models. Assumption 3.2 puts some constraints on the holding cost. One of the special cases for the holding cost is the commonly used linear cost, which satisfies Assumption 3.2. We write $X_i^k(p_k)$ to show the dependence of X_i^k on price p_k.

Lemma 3.1 *If Assumptions 3.1 and 3.2 hold, there exist random variables $\{\Lambda_i^k, \bar{\Lambda}_i^k\}$ with*

finite nth moments for all positive integers n, such that X_i^k satisfies

$$\|X_i^k(p_k') - X_i^k(p_k)\| \le \Lambda_i^k \|p_k' - p_k\| \quad w.p.1 \quad \forall\, k \in \{1, \cdots, N\}, \tag{3.17}$$

$$\|X_i^k(p_k)\| \le \bar{\Lambda}_i^k \ w.p.1. \quad \forall\, k \in \{1, \cdots, N\} \tag{3.18}$$

for all $p_k, p_k' \in [\underline{p}, \bar{p}]$. There exist random variables $\{\Phi_i^k\}$ and $\{\bar{\Phi}_i^k\}$, whose nth moments are finite for all positive integers n, such that $\|c(X_i^k)\| \le \Phi_i^k$ w.p.1; and $\|c((X_i^k)' - c(X_i^k))\| \le \bar{\Phi}_i^k \|(X_i^k)' - X_i^k\|$ w.p.1.

Proof. The proof is by induction. We only prove the result for the states $\{X_i^1\}$ before the stopping time τ_1; other cases can be proved similarly. Let $\tau_0 = t_0$. Since $X_{\tau_1} = S_1$ is a scalar, the theorem is proved for X_{τ_1}. Similarly $X_0(t_0) = S$ is a constant, and the claim is trivial at t_0. The state $X_1(p_1)$ at t_1 is given by $X_1^1(p_1) = X_0 + \lambda(p_1)(t_1 - t_0) + \sigma(X_0 - X_0, p_1)\sqrt{t_1 - t_0}Z_1$. By Assumption 3.1, we have

$$\|X_1^1(p_1') - X_1^1(p_1)\|$$

$$\le \|\lambda(p_1')(t_1 - t_0) - \lambda(p_1)(t_1 - t_0)\| + \|\sigma(X_0 - X_0), p_1')\sqrt{t_1 - t_0}Z_1$$

$$- \sigma(X_0 - X_0, p_1)\sqrt{t_1 - t_0}Z_1\|$$

$$\le K_\lambda h\|p_1' - p_1\| + \sqrt{h}\|Z_1\|K_{g_2}\|p_1' - p_1\| = \Lambda_1^1\|p' - p\|, \quad w.p.1,$$

where $\Lambda_1^1 = hK_\lambda + \sqrt{h}K_{g_2}\|Z_1\|$ with $E[\Lambda_1^1] \le \mathcal{M}_1^1$ and $E[(\Lambda_1^1)^2] \le \mathcal{V}_1^1$, where $\mathcal{M}_1^1 = hK_\lambda + \sqrt{h}K_{g-2}$ and $\mathcal{V}_1^1 = (hK_\lambda)^2 + 2h\sqrt{h}K_\lambda K_{g_2} + h(K_{g_2})^2$. It is also easy to show there exists a $\mathcal{N}_1^1$ such that $E[(\Lambda_1^1)^4] \le \mathcal{N}_1^1$. For other parameters p_k, $k \in \{2, \cdots, N\}$, the claim

is trivial since X_1^1 is independent of p_k. We also have

$$\|X_1^1(p_1)\| \leq \|X_0^1\| + \|\lambda(p_1)(t_1 - t_0)\| + \|\sigma(X_0 - X_0), p_1)\sqrt{t_1 - t_0}Z_1\|$$

$$\leq S + \bar{\lambda}h + \bar{\sigma}\sqrt{h}\|Z_1\| = \bar{\Lambda}_1^1, \quad w.p.1,$$

where $\bar{\Lambda}_1^1 = S + \bar{\lambda}h + \bar{\sigma}\sqrt{h}\|Z_1\|$ with $E[\bar{\Lambda}_1^1] \leq \overline{\mathcal{M}}_1^1$ and $E[(\bar{\Lambda}_1^1)^2] \leq \overline{\mathcal{V}}_1^1$, where $\overline{\mathcal{M}}_1^1 = S + \bar{\lambda}h + \bar{\sigma}\sqrt{h}$ and $\overline{\mathcal{V}}_1^1 = (S+\bar{\lambda})^2 + 2(S+\bar{\lambda})\bar{\sigma}\sqrt{h} + \bar{\sigma}^2h$, since $E[\|Z_1\|] < 1$ and $E[\|Z_1\|^2] = 1$.

Now assume that the claim (3.18) is true for X_j^1. Then for X_{j+1}^1, we have

$$X_{j+1}^1 = X_j^1 + \lambda(p_1) + \sigma(X_0 - X_j^1, p_1)\sqrt{t_{j+1} - t_j}Z_{j+1}$$

$$\leq \|X_j^1\| + \bar{\sigma}\sqrt{h}\|Z_{j+1}\| \leq \bar{\Lambda}_j^1 + \bar{\sigma}\sqrt{h}\|Z_{j+1}\| = \bar{\Lambda}_{j+1}^1, \quad w.p.1.$$

Therefore (3.18) is true for X_{j+1}^1. Let $X_j^1(p_1')$ denote the value of X_j^1 when p_1 is perturbed to be p_1'. By Assumptions 3.1 and 3.2, we have

$$\|X_{j+1}^1(p_1') - X_{j+1}^1(p_1)\|$$

$$\leq \|X_j^1(p_1') - X_j^1(p_1)\| + (t_{j+1} - t_j)\|\lambda(p_1') - \lambda(p_1)\| +$$

$$\|\sqrt{t_{j+1} - t_j}Z_{j+1}\|\|\sigma(X_0 - X_j^1, p_1') - \sigma(X_0 - X_j^1, p_1)\|$$

$$\leq \Lambda_j^1\|p_1' - p_1\| + hK_\lambda\|p_1' - p_1\| + \sqrt{h}\|Z_{j+1}\|\left\{K_{g_1}\|X_j^1(p_1') - X_j^1(p_1)\| + K_{g_2}\|p_1' - p_1\|\right\}$$

$$\leq \Lambda_{j+1}^1\|p_1' - p_1\|, \quad w.p.1,$$

where $\Lambda_{j+1}^1 = \Lambda_j^1 + hK_\lambda + \sqrt{h}\|Z_{j+1}\|K_{g_1}\Lambda_j^1 + K_{g_2}\|$. Assume that there exist finite constants $\mathcal{M}_j^1$, $\mathcal{V}_j^1$, and $\mathcal{N}_j^1$ such that $E[\Lambda_j^1] < \mathcal{M}_j^1$, $E[(\Lambda_j^1)^2] < \mathcal{V}_j^1$, and and $E[(\Lambda_j^1)^4] < \mathcal{N}_j^1$. Moreover, notice that Z_{j+1} and Λ_j^1 are independent. It is easy to show that there exist

finite constants $\mathcal{M}_{j+1}^1$, $\mathcal{V}_{j+1}^1$, and $\mathcal{N}_{j+1}^1$ such that $E[\Lambda_{j+1}^1] \leq \mathcal{M}_{j+1}^1$, $E[(\Lambda_{j+1}^1)^2] \leq \mathcal{V}_{j+1}^1$, and $E[(\Lambda_{j+1}^1)^4] < \mathcal{N}_{j+1}^1$. The claims (3.17) and (3.18) for X_i^k follow from a similar proof, which is omitted. Since $\tau < T$, it at most takes $\lceil T/h \rceil$ steps to get to τ_1. Therefore the theorem holds for all K.

Similarly, we can show that all the nth moment of Λ_i^k is finite, where n is any positive integer. Since the nth moment of X_i^k is upper bounded by the nth moment of Λ_i^k, and $c(\cdot)$ is a polynomial function by Assumption 3.2, there exists a random variable Φ_{i+1}^k which has finite nth moment, such that $c(X_i^k) < \Phi_{i+1}^k$ w.p.1; there exists a random variable $\bar{\Phi}_{i+1}^k$ with finite nth moment, such that $\|c((X_i^k)') - c(X_i^k))\| \leq \bar{\Phi}_{i+1}^k \|(X_i^k)' - X_i^k\|$.

∎

3.3.3.1 Gradient Estimation for the Inventory

From equation (3.8), we can see that because $\lambda(p_k)$ and $\sigma(X_0 - x_i^k, p_k)$ are differentiable in p_k, then x_i^k is differentiable in p_k. For $x_i^k > S_k$, $x_{i+1}^k > S_k$, $t_{i+1}^k \leq T - h$, and $i \neq 0$ or $k = 1$ and $i = 0$, the gradient of x_i^k with respect to $p_k, k \in \{1, \cdots, N\}$ is given by

$$
\begin{aligned}
\frac{\partial x_{i+1}^k}{\partial p_k} = {} & \frac{\partial x_i^k}{\partial p_k} - \frac{\partial \lambda(p_k)}{\partial p_k}(t_{i+1}^k - t_i^k) \\
& - \left(\frac{\partial \sigma(X_0 - x_i^k, p_k)}{\partial p_k} - \frac{\partial \sigma(X_0 - x_i^k, p_k)}{\partial x_i^k} \frac{\partial x_i^k}{\partial p_k} \right) \sqrt{t_{i+1}^k - t_i^k} z_{i+1}^k.
\end{aligned}
\tag{3.19}
$$

The boundary conditions are

$$
\frac{\partial X_0}{\partial p_k} = 0, \quad \frac{\partial x_{\tau_k}^k}{\partial p_k} = 0.
\tag{3.20}
$$

The above boundary conditions hold because X_0 is a constant, and $x_{\tau_k}^k = S_k$.

When $x_i^k > S_k$ and $t_{i+1}^k > T-h > t_i^k$, from Algorithm 1, we know that $t_i^k = \tau_{k-1}+ih$. Equation (3.9) can be rewritten as

$$x_{i+1}^k = x_i^k - \lambda(p_k)\Big(T - \tau_{k-1} - ih\Big) - \sigma(X_0 - x_i^k, p_k)\sqrt{T - \tau_{k-1} - ih}z_{i+1}^k.$$

Differentiating both sides of the above equation with respect to p_l for $l \leq k$ yields

$$\frac{\partial x_{i+1}^k}{\partial p_l} = \lambda(p_k)\frac{\partial \tau_l}{\partial p_l} + \frac{\sigma(X_0 - x_i^k, p_k)}{2\sqrt{T - \tau_{k-1} - ih}}\frac{\partial \tau_l}{\partial p_l}z_{i+1}^k \qquad \forall\, l < k, \tag{3.21}$$

$$\frac{\partial x_{i+1}^k}{\partial p_k} = \frac{\partial x_i^k}{\partial p_k} - \frac{\partial \lambda(p_k)}{\partial p_k}\Big(T - \tau_{k-1} - ih\Big) - \Big(\frac{\partial \sigma(X_0 - x_i^k, p_k)}{\partial p_k}$$
$$- \frac{\partial \sigma(X_0 - x_i^k, p_k)}{\partial x_i^k}\frac{\partial x_i^k}{\partial p_k}\Big)\sqrt{T - \tau_{k-1} - ih}z_{i+1}^k. \tag{3.22}$$

3.3.3.2 Likelihood Ratio Method for Gradient Estimation of Stopping Times

Since the density of τ_{i+1}^k given two end points is known, we can use the likelihood ratio method ([48, 91]) to obtain a gradient estimator. The likelihood ratio gradient estimator is derived by assuming the interchangeability of the derivative and the integral:

$$\frac{\partial E[\tau_{i+1}^k | x_i^k, x_{i+1}^k]}{\partial p_k} = \int_{t_i}^{t_{i+1}} \tau \frac{\partial \ln f_k(\tau)}{\partial p_k} f_k(\tau)d\tau. \tag{3.23}$$

The above equality holds because $f_k(t_i) = 0$ and $f_k(t_{i+1}) = 0$. By using the inverse Gaussian distribution, we have

$$\frac{\partial E[\tau_{i+1}^k | x_i^k, x_{i+1}^k]}{\partial p_k} = \frac{\partial}{\partial p_k} \int_0^\infty \left(t_i^k + \frac{hy}{1+y} \right) g_k(y; \lambda_k, \theta_k) dy$$

$$= \int_0^\infty \frac{hy}{1+y} \frac{\partial \ln g_k(y; \lambda_k, \theta_k)}{\partial p_k} g_k(y; \lambda_k, \theta_k) dy.$$

We use

$$\mathscr{R}_k(\tau_{i+1}^k; x_i^k, x_{i+1}^k) = \frac{hy}{1+y} \frac{\partial \ln g_k(y; \lambda_k, \theta_k)}{\partial p_k} \tag{3.24}$$

to denote the likelihood ratio gradient estimator of τ_{i+1}^k with respect to p_k.

3.3.3.3 Sample Path-Based Gradient Estimator for Stopping Times

Although the likelihood ratio gradient estimator is easy to obtain, generally it has a larger variance compared with sample path-based gradient estimators. We derive a sample path-based gradient estimator by constructing a Brownian bridge, which starts from x_i^k, hits S_k at τ_{i+1}^k, and ends at x_{i+1}^k .

Let X_t^k be the inventory level at time t, where $t_i^k \leq t < t_{i+1}^k$; X_t^k is governed by the following equation

$$X_t^k = x_i^k - \lambda(p_k)(t - t_i^k) - \sigma(X_0 - x_i^k, p_k)W(t - t_i^k),$$

where $W(t - t_i^k)$ is a Brownian motion at time $t - t_i^k$. Given two end points x_i^k and x_{i+1}^k, we construct a Brownian bridge $\hat{X}_t^k$ starting from x_i^k and ending at x_{i+1}^k, which is defined

by ([62, 96]):

$$\hat{X}_t^k = x_i^k + \frac{x_{i+1}^k - x_i^k}{h}(t - t_i^k) - \sigma(X_0 - x_i^k, p_k)(t_{i+1}^k - t)\tilde{W}\left(\frac{t - t_i^k}{t_{i+1}^k - t}\right), \qquad (3.25)$$

where $\tilde{W}$ is a Brownian motion independent of W.

For the stochastic process $\hat{X}_t^k$, define the first time when $\hat{X}_t^k$ hits S_k by

$$\hat{\tau}_{i+1}^k = \inf\{t : \hat{X}_t^k = S_k\}. \qquad (3.26)$$

We derive an unbiased gradient estimator for $\frac{\partial E[\tau_{i+1}^k|(x_i^k, x_{i+1}^k)]}{\partial p_k}$ considering two different cases.

Case 1. $x_i^k > S_k$ *and* $x_{i+1}^k < S_k$**:**

In this case, $\hat{\tau}_{i+1}^k$ is well defined as it always exists, and we have the following lemma.

Lemma 3.2 *Given two points x_i^k and x_{i+1}^k, the conditional distribution of the random variable $\hat{\tau}_{i+1}^k$ defined by (3.26) is the same as the conditional distribution of the random variable τ_{i+1}^k. Hence we have*

$$E[\tau_{i+1}^k|x_i^k, x_{i+1}^k] = E[\hat{\tau}_{i+1}^k|x_i^k, x_{i+1}^k].$$

Proof. By the properties of a Brownian bridge, we know that the finite dimensional distributions of a Brownian bridge are the same as the finite dimensional distributions of the corresponding conditional Brownian motion. Since τ_{i+1}^k is the stopping time defined by X_t^k, whereas $\hat{\tau}_{i+1}^k$ is defined by $\hat{X}_t^k$ in (3.26), τ_{i+1}^k has the same distribution as $\hat{\tau}_{i+1}^k$

given x_i^k and x_{i+1}^k. Therefore we conclude that

$$E[\tau_{i+1}^k|(x_i^k, x_{i+1}^k)] = E[\hat{\tau}_{i+1}^k|(x_i^k, x_{i+1}^k)].$$

■

By Lemma 3.2, a gradient estimator for $E[\tau_{i+1}^k|(x_i^k, x_{i+1}^k)]$ with respect to p_k can be obtained by deriving a gradient estimator for $E[\hat{\tau}_{i+1}^k|(x_i^k, x_{i+1}^k)]$ with respect to p_k. By equation (3.25), there exists a realization $\tilde{w}_k$ of $\tilde{W}$ such that at time $\hat{\tau}_{i+1}^k$ we have

$$S_k = x_i^k + \frac{x_{i+1}^k - x_i^k}{h}(\hat{\tau}_{i+1}^k - t_i^k) - \sigma(X_0 - x_i^k, p_k)(t_{i+1}^k - \hat{\tau}_{i+1}^k)\tilde{w}_k. \qquad (3.27)$$

Given $\hat{\tau}_{i+1}^k$, $\tilde{w}_k$ can be computed by

$$\tilde{w}_k = \frac{-S_k + \frac{x_i^k}{h}(t_{i+1}^k - \hat{\tau}_{i+1}^k) + \frac{x_{i+1}^k}{h}(\hat{\tau}_{i+1}^k - t_i^k)}{\sigma(X_0 - x_i^k, p_k)(t_{i+1}^k - \hat{\tau}_{i+1}^k)}. \qquad (3.28)$$

An unbiased gradient estimator for $\frac{\partial E[\tau_{i+1}^k|(x_i^k, x_{i+1}^k)]}{\partial p_k}$ is given by the following lemma.

Lemma 3.3 *If Assumptions 3.1 and 3.2 hold, there exists a random variable Ψ_i^k with finite nth moment for all positive integers n, such that*

$$\mathscr{D}_k(\tau_{i+1}^k; x_i^k, x_{i+1}^k) \leq \frac{(t_{i+1}^k - \tau_{i+1}^k)\Psi_i^k}{S_k - x_{i+1}^k} \ w.p.1.,$$

where $\mathscr{D}_k(\tau_{i+1}^k; x_i^k, x_{i+1}^k)$ is given by

$$\mathscr{D}_k(\tau_{i+1}^k; x_i^k, x_{i+1}^k) = \frac{\frac{\partial x_i^k}{\partial p_k}\frac{t_{i+1}^k - \tau_{i+1}^k}{h} + \frac{\partial x_{i+1}^k}{\partial p_k}\frac{\tau_{i+1}^k - t_i^k}{h} - \frac{\partial \sigma(X_0 - x_i^k, p_k)}{\partial p_k}(t_{i+1}^k - \tau_{i+1}^k)\tilde{w}_k}{\frac{S_k - x_{i+1}^k}{t_{i+1} - \tau_{i+1}^k}}, \qquad (3.29)$$

53

in which $\tilde{w}_k$ is given by (3.28); $\mathscr{D}_k(\cdot;\cdot,\cdot)$ is an unbiased gradient estimator for $\frac{\partial E[\tau_{i+1}^k|(x_i^k,x_{i+1}^k)]}{\partial p_k}$.

Proof. Given x_i^k, x_{i+1}^k, the inventory levels at t_i^k and t_{i+1}^k with price p_k, and a realization $\tilde{w}_k$ of $\tilde{W}\left(\frac{\hat{\tau}_{i+1}^k - t_i^k}{t_{i+1}^k - \hat{\tau}_{i+1}^k}\right)$, we rewrite equation (3.27) below for convenience:

$$S_k = x_i^k + \frac{x_{i+1}^k - x_i^k}{h}(\hat{\tau}_{i+1}^k - t_i^k) - \sigma(X_0 - x_i^k, p_k)(t_{i+1}^k - \hat{\tau}_{i+1}^k)\tilde{W}\left(\frac{\hat{\tau}_{i+1}^k - t_i^k}{t_{i+1}^k - \hat{\tau}_{i+1}^k}\right). \quad (3.30)$$

Now we do a perturbation analysis. Perturb price p_k to $p_k + \Delta\theta$, where $\Delta\theta$ is small scalar. Denote $(x_i^k)', (x_{i+1}^k)'$, and $(\hat{\tau}_{i+1}^k)'$ the corresponding values of x_i^k, x_{i+1}^k, and $\hat{\tau}_{i+1}^k$ at price $p_k + \Delta\theta$. $(\hat{\tau}_{i+1}^k)'$ satisfies the following equation:

$$\begin{aligned}
S_k = (x_i^k)' + \frac{(x_{i+1}^k)' - (x_i^k)'}{h}((\hat{\tau}_{i+1}^k)' - t_i^k) \\
- \sigma(X_0 - (x_i^k)', p_k + \Delta\theta)(t_{i+1}^k - (\hat{\tau}_{i+1}^k)')\tilde{W}\left(\frac{(\hat{\tau}_{i+1}^k)' - t_i^k}{t_{i+1}^k - (\hat{\tau}_{i+1}^k)'}\right),
\end{aligned} \quad (3.31)$$

For simplicity we denote $\tilde{W}\left(\frac{\hat{\tau}_{i+1}^k - t_i^k}{t_{i+1}^k - \hat{\tau}_{i+1}^k}\right)$ by $\tilde{W}$ and $\tilde{W}\left(\frac{(\hat{\tau}_{i+1}^k)' - t_i^k}{t_{i+1}^k - (\hat{\tau}_{i+1}^k)'}\right)$ by $\tilde{W}'$. Let $\mathcal{F}_t$ be the filtration generated by $\left\{\tilde{W}\left(\frac{t - t_i^k}{t_{i+1}^k - t}\right)\right\}$. By properties of Brownian motion, $\tilde{W}\left(\frac{t - t_i^k}{t_{i+1}^k - t}\right)$ is a martingale. Without loss of generality, assume that $(\hat{\tau}_{i+1}^k)' \geq \hat{\tau}_{i+1}^k$. By the definition (3.26), $(\hat{\tau}_{i+1}^k)'$ is a stopping time. Therefore by optional stopping theory ([12])

$$E[\tilde{W}' - \tilde{W}|\tilde{W}] = 0. \quad (3.32)$$

Subtract equation (3.30) on each side of equation (3.31), take expectation on each side of the resulting equation conditioning on $\tilde{W}$, then divide each side by $\Delta\theta$, and take

the limit as $\Delta\theta$ goes to zero. We have

$$
\begin{aligned}
0 &= \frac{\partial x_i^k}{\partial p_k} + \left(\frac{\partial x_{i+1}^k}{\partial p_k} - \frac{\partial x_i^k}{\partial_k}\right)\frac{\hat{\tau}_{i+1}^k - t_i^k}{h} + \frac{x_{i+1}^k - x_i^k}{h}\frac{\partial E\left[\hat{\tau}_{i+1}^k|\tilde{W}\right]}{\partial p_k} \\
&\quad - \frac{\partial\sigma(X_0 - x_i^k, p_k)}{\partial p_k}(t_{i+1}^k - \hat{\tau}_{i+1}^k)\tilde{W} - \sigma(X_0 - x_i^k, p_k)t_{i+1}^k\lim_{\Delta\theta\to 0}\frac{1}{\Delta\theta}E[\tilde{W}' - \tilde{W}|\tilde{W}] \\
&\quad + \sigma(X_0 - x_i^k, p_k)\hat{\tau}_{i+1}^k\lim_{\Delta\theta\to 0}\frac{1}{\Delta\theta}E\left[(\hat{\tau}_{i+1}^k)'\tilde{W}' - \hat{\tau}_{i+1}^k\tilde{W}' + \hat{\tau}_{i+1}^k\tilde{W}' - \hat{\tau}_{i+1}^k\tilde{W}|\tilde{W}\right] \\
&= \frac{\partial x_i^k}{\partial p_k} + \left(\frac{\partial x_{i+1}^k}{\partial p_k} - \frac{\partial x_i^k}{\partial_k}\right)\frac{\hat{\tau}_{i+1}^k - t_i^k}{h} + \frac{x_{i+1}^k - x_i^k}{h}\frac{\partial E\left[\hat{\tau}_{i+1}^k|\tilde{W}\right]}{\partial p_k} \\
&\quad - \frac{\partial\sigma(X_0 - x_i^k, p_k)}{\partial p_k}(t_{i+1}^k - \hat{\tau}_{i+1}^k)\tilde{W} + \sigma(X_0 - x_i^k, p_k)\frac{\partial E[\hat{\tau}_{i+1}^k|\tilde{W}]}{\partial p_k}\tilde{W}, \quad\quad (3.33)
\end{aligned}
$$

where the second equality is obtained by (3.32). After some algebraic operations, we have

$$
\frac{\partial E[\hat{\tau}_{i+1}^k|\tilde{W}]}{\partial p_k} = \frac{\frac{\partial x_i^k}{\partial p_k}\frac{t_{i+1}^k - \hat{\tau}_{i+1}^k}{h} + \frac{\partial x_{i+1}^k}{\partial p_k}\frac{\hat{\tau}_{i+1}^k - t_i^k}{h} - \frac{\partial\sigma(X_0 - x_i^k, p_k)}{\partial p_k}(t_{i+1}^k - \hat{\tau}_{i+1}^k)\tilde{w}_k}{\frac{S_k - x_{i+1}^k}{t_{i+1} - \hat{\tau}_{i+1}^k}}.
$$

Denote $\frac{\partial E[\hat{\tau}_{i+1}^k|\tilde{W}]}{\partial p_k}$ by $\mathscr{D}_k(\hat{\tau}_{i+1}^k; x_i^k, x_{i+1}^k)$, which satisfies

$$
\begin{aligned}
\|\mathscr{D}_k(\hat{\tau}_{i+1}^k; x_i^k, x_{i+1}^k)\| &\leq \left\|\frac{\frac{\partial x_i^k}{\partial p_k}\frac{t_{i+1}^k - \hat{\tau}_{i+1}^k}{h} + \frac{\partial x_{i+1}^k}{\partial p_k}\frac{\hat{\tau}_{i+1}^k - t_i^k}{h} - \frac{\partial\sigma}{\partial p_k}(t_{i+1}^k - \hat{\tau}_{i+1}^k)\tilde{w}_k}{\frac{S_k - x_{i+1}^k}{t_{i+1} - \hat{\tau}_{i+1}^k}}\right\| \\
&\leq \frac{t_{i+1}^k - \hat{\tau}_{i+1}^k}{S_k - x_{i+1}^k}\left(\left\|\frac{\partial x_i^k}{\partial p_k}\right\| + \left\|\frac{\partial x_{i+1}^k}{\partial p_k}\right\| + \left\|\frac{\partial\sigma}{\partial p_k}\tilde{z}_k\right\|\right) \\
&\leq \frac{t_{i+1}^k - \hat{\tau}_{i+1}^k}{S_k - x_{i+1}^k}\left(\Lambda_i^k + \Lambda_{i+1}^k + (K_{g_1}\Lambda_i^k + K_{g_2})\frac{1}{\underline{\sigma}}(\overline{\Lambda}_i^k + \overline{\Lambda}_{i+1}^k + S_k)\right) \\
&= \frac{(t_{i+1}^k - \hat{\tau}_{i+1}^k)\Psi_i^k}{S_k - x_{i+1}^k}\ w.p.1, \quad\quad (3.34)
\end{aligned}
$$

where $\Psi_i^k = \left(\Lambda_i^k + \Lambda_{i+1}^k + (K_{g_1}\Lambda_i^k + K_{g_2})\frac{1}{\underline{\sigma}}(\overline{\Lambda}_i^k + \overline{\Lambda}_{i+1}^k + S_k)\right)$. By Assumptions 3.1, 3.2, and Lemma 3.1, Λ_{i+1}^k, Ψ_{i+1}^k has finite nth moment for any positive finite integer n. Given x_i^k and x_{i+1}^k, $\frac{(t_{i+1}^k - \hat{\tau}_{i+1}^k)\Psi_i^k}{S_k - x_{i+1}^k}$ is finite $w.p.1$, and has a finite expectation. By the dominated convergence theorem, $\mathscr{D}_k(\hat{\tau}_{i+1}^k; x_i^k, x_{i+1}^k)$ is unbiased. Since τ_{i+1}^k and $\hat{\tau}_{i+1}^k$ have the same

distribution, $\mathscr{D}_k(\tau_{i+1}^k; x_i^k, x_{i+1}^k)$ is an unbiased gradient estimator for $\frac{\partial E[\tau_{i+1}^k|(x_i^k, x_{i+1}^k)]}{\partial p_k}$, which has similar properties as $\mathscr{D}_k(\hat{\tau}_{i+1}^k; x_i^k, x_{i+1}^k)$.■

Case 2. $x_i^k > S_k$ *and* $x_{i+1}^k > S_k$:

By the definition of τ_{i+1}^k in (3.10), τ_{i+1}^k is well defined only when $\ell_{i+1}^k < S_k$, i.e., the minimum of the inventory level is less than S_k between time t_i^k and t_{i+1}^k. In this case, we know that the expectation of τ_{i+1}^k is given by (3.16) conditioned on $\ell_{i+1}^k < S_k$. It is not difficult to see that equation (3.16) is the same as equation (3.15) when the two given end points in (3.16) are x_i^k and $\tilde{x}_{i+1}^k$ where $\tilde{x}_{i+1}^k = 2S_k - x_{i+1}^k$. Therefore the the gradient estimation in this case is transformed into Case 1 with two given points x_i^k and $\tilde{x}_{i+1}^k$ at time t_i^k and t_{i+1}^k, respectively. As in Case 1, a pathwise gradient estimator for $E[\tau_{i+1}^k|(x_i^k, x_{i+1}^k)]$ given by (3.16) with respect to p_k is $\mathscr{D}_k(\tau_{i+1}^k; x_i^k, \tilde{x}_{i+1}^k)$.

3.3.3.4   Gradient Estimator for the Value Function

Now we show how to estimate the gradient of $E[\hat{V}(S)]$ with respect to the pricing policy parameters. Note that $\hat{V} = \sum_{k=1}^{n_e} \Delta\hat{V}_k$. Since there are indicator functions in $\Delta\hat{V}_k$ defined by (3.11), which are discontinuous, unbiased pathwise gradient estimators for $E[\hat{V}(S)]$ cannot be obtained directly. In the following, we give an unbiased pathwise gradient estimator for $E[\hat{V}(S)]$ using smoothed perturbation analysis (SPA).

Examining the incremental value function $\Delta\hat{V}_k$, it is easy to see that $\Delta\hat{V}_k$ is continuous at $X_j^k = S_k$ for $j = 1, \cdots, v_k + 1$. Hence the difficulty of differentiating $\Delta\hat{V}_k$ comes from the indicator functions $I_{\{U_i^k < \eta_i^k\}}$ and $I_{\{U_i^k \geq \eta_i^k\}}$ for $i = 1, \cdots, v_k + 1$, which make $\Delta\hat{V}_k$ discontinuous. From equation (3.11), we can see that given p_k and $\{X_n^k, \forall n \leq m\}$, the values of the indicator functions change at $U_i^{k\star} = \eta_i^k$ for $i = 1, \cdots, v_k + 1$, where $U_i^{k\star}$ is

one sample value of $U_i^k \sim U(0,1)$. We write $\eta_i^k(p_k)$ to explicitly show the dependence of η_i^k on the price p_k; similarly we write $\Delta \hat{V}_k(p_k)$ to show the dependence of $\Delta \hat{V}_k$ on the price p_k.

Now we perturb p_k to $p_k + \Delta \theta$ and $p_k - \Delta \theta$ for a small $\Delta \theta$. The values of η_i^k at the new parameters are $\eta_i^k(p_k + \Delta \theta)$ and $\eta_i^k(p_k - \Delta \theta)$, respectively. Without loss of generality, assume $U_i^{k\star} \leq \eta_i^k(p_k + \Delta \theta)$, and $U_i^{k\star} \geq \eta_i^k(p_k - \Delta \theta)$. Define

$$
\mathcal{B}_i^k(U_i^{k\star}, p_k, \Delta \theta) = \left\{ \ U_i^k : \begin{array}{l} U_i^{k\star} \leq U_i^k \leq \ \eta_i^k(p_k + \Delta \theta); \\[1mm] \text{and } U_i^{k\star} \geq U_i^k \geq \ \eta_i^k(p_k - \Delta \theta). \end{array} \right. \qquad \text{for } i = 1, \cdots, v_k.
$$

$\mathcal{B}_i^k(U_i^{k\star}, p_k, \Delta \theta)$ contains U_i^k such that $U_i^k - \eta_i^k$ with the price $p_k + \Delta \theta$ has a different sign than it does with the price $p_k - \Delta \theta$. Similarly, define $\mathcal{B}_e^k(U_e^{k\star}, p_k, \Delta \theta) = \{U_e^k : U_e^{k\star} \leq U_e^k \leq \ \eta_{v_e+1}^{n_e}(p_k + \Delta \theta), \text{or } U_e^{k\star} \geq U_e^k \geq \ \eta_{v_e+1}^{n_e}(p_k - \Delta \theta)\}$, where $U_e^k \sim U(0,1)$ for $k < n_e$. Let $\overline{\mathcal{B}_i^k}(U_i^{k\star}, p_k, \Delta \theta)$ be the complement of $\mathcal{B}_i^k(U_i^{k\star}, p_k, \Delta \theta)$. For simplicity, we use $\mathcal{B}_i^k$ to represent $\mathcal{B}_i^k(U_i^{k\star}, p_k, \Delta \theta)$, and define $\mathcal{B} = \mathcal{B}_e^k \bigcup_{i=1}^{v_k} \mathcal{B}_i^k$. By the property of the iterated conditional expectation, for $k < n_e$,

$$
\begin{aligned}
E[\Delta \hat{V}_k | \overline{\mathcal{B}}] = \ &- p_k \Delta S_k - E\Big[\sum_{m=0}^{v_k} \Big\{ \Big[\sum_{i=0}^{m-1} c(X_i^k) h + c(X_m^k)(E[\tau_{m+1}^k | (X_m^k, X_{m+1}^k)] \\
& - t_m^k) \Big] \prod_{j=0}^{m} I_{\{U_j^k < \eta_j^k\}} I_{\{U_{m+1}^k \geq \eta_{m+1}^k\}} \Big\} | \overline{\mathcal{B}} \Big].
\end{aligned}
$$

Restricted to the set $\overline{\mathcal{B}}$, $\Delta \hat{V}_k$ will not change abruptly if we perturb p_k, since $\Delta \hat{V}_k$ is continuous in p_k. Directly differentiating the part inside the right hand side expectation

in the above equation yields

$$\mathscr{D}_p^k \Delta \hat{V}_k = -\Delta S_k - p_k \frac{\Delta S_k}{\partial p_k} - \sum_{m=0}^{v_k} \left\{ \left[\sum_{i=0}^{m-1} \frac{\partial c(X_i^k)}{\partial p_k} h + \frac{\partial c(X_m^k)}{\partial p_k} (E[\tau_{m+1}^k | (X_m^k, X_{m+1}^k)] - t_m^k) \right. \right.$$

$$\left. \left. + c(X_m^k) E[\mathcal{G}_k(\tau_{m+1}^k; X_m^k, X_{m+1}^k)] \right] \prod_{j=0}^{m} I_{\{U_j^k < \eta_j^k\}} I_{\{U_{m+1}^k \geq \eta_{m+1}^k\}} \right\}, \tag{3.35}$$

where $\mathscr{D}_p^k \Delta \hat{V}_k$ is the derivative of $\Delta \hat{V}_k$ with respect to p_k conditioned on $\overline{\mathcal{B}}$; $\mathcal{G}_k$ is the gradient estimators of the stopping time with respect to p_k. When $X_m^k > S_k$ and $X_{m+1}^k > S_k$, we define $\mathcal{G}_k$ by: $\mathcal{G}_k(\tau_{m+1}^k; X_m^k, X_{m+1}^k) = \mathscr{D}_k(\tau_{m+1}^k; X_m^k, 2S_k - X_{m+1}^k)$; when $X_m^k > S_k$ and $X_{m+1}^k \leq S_k$, $\mathcal{G}_k(\tau_{m+1}^k; X_m^k, X_{m+1}^k) = \mathscr{D}_k(\tau_{m+1}^k; X_m^k, X_{m+1}^k)$ where $\mathscr{D}_k$ is defined by (3.29).

By equations (3.21) and (3.22), $X_{v_e+1}^{n_e}$ depends on p_l for $l \leq n_e$ since it depends on all the previous stopping times when the price changes, whereas $\{X_i^k\}$ do not depend on p_l for $l < k, k \neq n_e$ by (3.19). The derivative of $\Delta \hat{V}_{n_e}$ with respect to p_l for $l < n_e$ is given by

$$\mathscr{D}_p^l \Delta \hat{V}_{n_e} = -p_{n_e} \frac{\partial \Delta S_{n_e}}{p_l} - c(X_{v_e}^{n_e}) E\left[\mathcal{G}_l(\tau_{v_e+1}^{n_e}; X_{v_e}^{n_e}, X_{v_e+1}^{n_e})\right]. \tag{3.36}$$

Since $\{X_i^k\}$ do not depend on p_l for $l < k$ and $k \neq n_e$, $\Delta \hat{V}_k$ does not depends on p_l for $l < k$ and $k \neq n_e$. Hence the gradient estimator of $E[\hat{V}|\overline{\mathcal{B}}]$ with respect to p_k is given by

$$\mathscr{D}_p^k \hat{V} | \overline{\mathcal{B}} = \mathscr{D}_p^k \Delta \hat{V}_k | \overline{\mathcal{B}} + \mathscr{D}_p^k \Delta \hat{V}_{n_e} | \overline{\mathcal{B}}, \qquad \forall k \neq n_e \tag{3.37}$$

$$\mathscr{D}_p^{n_e} \hat{V} | \overline{\mathcal{B}} = \mathscr{D}_p^{n_e} \Delta \hat{V}_{n_e} | \overline{\mathcal{B}}, \tag{3.38}$$

where we use the sign $|\overline{\mathcal{B}}$ to mean that the gradient estimators apply when sample paths are restricted to $\overline{\mathcal{B}}$.

We now prove that the gradient estimators (3.37) and (3.38) are unbiased.

Lemma 3.4 *If Assumptions 3.1 and 3.2 hold, there exist positive constants $\{\mathcal{G}_D^k\}$ and random variables $\{\Upsilon_D^k\}$ with $E[\Upsilon_D^k] \le \mathcal{G}_D^k$, such that*

$$\mathscr{D}_p^k \hat{V} | \overline{\mathcal{B}} \le \Upsilon_D^k \quad w.p.1, \qquad for\ k \le n_e,$$

where $\mathscr{D}_p^k \hat{V} | \overline{\mathcal{B}}$ is given by equations (3.37) and (3.38). The following result also holds:

$$\frac{\partial E\left[\hat{V}|\overline{\mathcal{B}}\right]}{\partial p_k} = E\left[\mathscr{D}_p^k \hat{V}|\overline{\mathcal{B}}\right], \qquad for\ k \le n_e.$$

Proof. First consider $\mathscr{D}_p^k \Delta \hat{V}_k$. From Assumptions 3.1 and 3.2 and Lemma 3.1

$$\left\| \sum_{i=1}^m \frac{\partial c(X_{i-1}^k)}{\partial p_k} h + \frac{\partial c(X_m^k)}{\partial p_k} (E[\tau_{m+1}^k | (X_m^k, X_{m+1}^k)] - t_m^k) \right\| \le \sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k h. \qquad (3.39)$$

Assume when $X_m^k > S_k, X_{m+1}^k < S_k$, and by Assumption 3.2 and Lemma 3.3,

$$\left\| c(X_m^k) E[\mathscr{G}_k(\tau_{m+1}^k; X_m^k, X_{m+1}^k)] I_{\{U_m^k < \eta_m^k\}} \right\| \le \Phi_m^k \frac{(t_{m+1}^k - \tau_{m+1}^k)\Psi_m^k}{S_k - X_{m+1}^k} I_{\{U_m^k < \eta_m^k\}}. \qquad (3.40)$$

From the density function (3.13), we have

$$\begin{aligned}
\chi &= E\left[\frac{(t_{m+1}^k - \tau_{m+1}^k)}{S_k - x_{m+1}^k} \Big| x_m^k, x_{m+1}^k \right] \\
&= \frac{1}{S_k - x_{m+1}^k} \int_{t_m^k}^{t_{m+1}^k} \frac{1}{\sqrt{2\pi\sigma^2}} (t - t_m^k)^{-3/2} (t_{m+1}^k - t)^{1/2} \\
&\quad \times \exp\left(-\frac{\left[(x_m^k - S_k)(t_{m+1}^k - t) - (S_k - x_{m+1}^k)(t - t_m^k)\right]^2}{2h(t - t_m^k)(t_{m+1}^k - t)\sigma^2} \right) dt
\end{aligned}$$

After some algebraic operations, we have

$$
\begin{aligned}
\chi &= \frac{1}{S_k - x_{m+1}^k} \int_0^h \exp\left(\frac{(x_m^k - x_{m+1}^k)^2}{2h\sigma^2}\right) u^{-3/2} \exp\left(-\frac{(x_m^k - S_k)^2}{2u\sigma^2}\right) \\
&\quad \times (h-u)^{1/2} \exp\left(-\frac{(S_k - x_{m+1}^k)^2}{2(h-u)\sigma^2}\right) du \qquad\qquad \text{let } u = t - t_m^k \\
&\leq \frac{1}{S_k - x_{m+1}^k} \exp\left(\frac{(x_m^k - x_{m+1}^k)^2}{2h\sigma^2}\right) \left(\int_0^h u^{-3} \exp\left(-\frac{(x_m^k - S_k)^2}{u\sigma^2}\right) du\right)^{1/2} \\
&\quad \times \left(\int_0^h (h-u) \exp\left(-\frac{(S_k - x_{m+1}^k)^2}{(h-u)\sigma^2}\right)\right)^{1/2} du. \qquad \text{by Hödler's inequality}
\end{aligned}
$$

(3.41)

Furthermore,

$$
\int_0^h u^{-3} \exp\left(-\frac{(x_m^k - S_k)^2}{u\sigma^2}\right) du = \int_{1/h}^\infty s \exp\left(-\frac{(x_m^k - S_k)^2 s}{\sigma^2}\right) ds
$$
$$
\leq \frac{\sigma^2}{(x_m^k - S_k)^2 h} + \frac{\sigma^4}{(x_m^k - S_k)^4}. \tag{3.42}
$$

Similarly, we have

$$
\begin{aligned}
&\int_0^h (h-u) \exp\left(-\frac{(S_k - x_{m+1}^k)^2}{(h-u)\sigma^2}\right) \\
&\leq \frac{(S_k - x_{m+1}^k)^2 h}{2\sigma^2}\left(1 - \exp\left(-\frac{-(S_k - x_{m+1}^k)^2}{\sigma^2 h}\right)\right) + \frac{(S_k - x_{m+1}^k)^4}{\sigma^4}.
\end{aligned}
$$

(3.43)

Applying (3.41), (3.42), and (3.43) to (3.40) gives

$$
\left\| c(X_m^k) E[\mathscr{G}_k(\tau_{m+1}^k; X_m^k, X_{m+1}^k)] I_{\{U_m^k < \eta_m^k\}} \right\| \leq M^k, \tag{3.44}
$$

where

$$M^k = \frac{1}{(X_m^k - S_k)^2} I_{\{U_m^k < \eta_m^k\}} \Phi_m^k \Psi_m^k \exp\left(\frac{(X_m^k - X_{m+1}^k)^2}{2h\sigma^2}\right)$$

$$\times \sqrt{\frac{\sigma^2(X_m^k - S_k)^2}{h} + \sigma^4} \sqrt{\frac{h}{2\sigma^2} + \frac{(S_k - X_{m+1}^k)^2}{\sigma^4}}.$$

To show that M^k has a finite expectation, we first show that $E\left[\frac{1}{(X_m^k - S_k)^2} I_{\{U_m^k < \eta_m^k\}}\right]$

has a finite expectation. Applying Assumption 3.2 and Lemma 3.1 yields

$$E\left[\frac{1}{(X_m^k - S_k)^2} I_{\{U_m^k < \eta_m^k\}}\right] = E\left[\frac{1}{(X_m^k - S_k)^2}\left(1 - \exp\left(-\frac{2(X_{m-1}^k - S_k)(X_m^k - S_k)}{h\sigma^2}\right)\right)\right]$$

$$\leq b_m^k, \tag{3.45}$$

where b_m^k is a positive constant. (3.45) holds since Ψ_m^k, X_{m-1}^k and X_m^j all have finite nth

moment for any positive integer n, and the exponential term goes to 1 exponentially when

$X_m^k - S_k$ is close to zero. It is easy to show that the above holds even when $X_m^k > S_k$ and

$X_{m+1}^k > S_k$.

Note that Φ_m^k and $\Phi_m^k \Psi_m^k$ in M^k are polynomial functions of X_i^k and they have

finite moments. By (3.45) and Cauchy-Schwarz inequality, it can be easily shown that

there exists a positive constant c_m^k, such that $E[M^k] \leq c^k$.

Combining the fact that $\upsilon_k h < T$, Assumption 3.2, and Lemma 3.1, we have

$$E[\sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k h] \leq \sum_{m=0}^{\lceil T/h \rceil} \sqrt{E[(\bar{\Phi}_m^k)^2]} \sqrt{E[(\Lambda_m^k)^2]} \leq d^k, \tag{3.46}$$

61

where d^k is a positive constant. Combining (3.39), (3.45), (3.44), and (3.46) yields

$$\|\mathscr{D}_p^k \Delta \hat{V}_k\| < \Upsilon_1^k, \tag{3.47}$$

where $\Upsilon_1^k = S_{k-1} - S_k + \sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k h + M^k$. From (3.39), (3.45), and (3.46), we know that $E[\Upsilon_1^k] \leq \mathcal{G}_1^k$, where $\mathcal{G}_1^k = (S_{k-1} - S_k) + c^k + d^k$. Similarly, we can prove that there exists a random variable Υ_2^k with $E[\Upsilon_2^k] \leq \mathcal{G}_2^k$, such that $\mathscr{D}_p^k \Delta \hat{V}_{n_e} \leq \Upsilon_2^k$. Together with (3.47), we have $\mathscr{D}_p^k \hat{V} \leq \Upsilon_D^k$, where $\Upsilon_D^k = \Upsilon_1^k + \Upsilon_2^k$, and $E[\Upsilon_D^k] \leq \mathcal{G}_D^k$, $\mathcal{G}_D^k = \mathcal{G}_1^k + \mathcal{G}_2^k$. By the dominated convergence theorem, for $k < n_e$, the first part of the theorem is proved. Similarly we prove the theorem for the gradient estimator $\mathscr{D}_p^{n_e} \hat{V}$ and the details are omitted. ∎

In the following, we derive a gradient estimator for $E[\hat{V}]$ with respect to p_k for $k = 1, \cdots, n_e$. We circumvent the difficulty of differentiating indicator function in $\hat{V}$ by conditioning on the set $\bar{\mathcal{B}}$.

Theorem 3.2 *If Assumptions 3.1 and 3.2 hold, we have*

$$\frac{\partial E[\hat{V}]}{\partial p_k} = E[\mathscr{C}_p^k \hat{V}] + E[\mathscr{D}_p^k \Delta \hat{V}_k + \mathscr{D}_p^k \Delta \hat{V}_{n_e}], \qquad for \ k < n_e; \tag{3.48}$$

$$\frac{\partial E[\hat{V}]}{\partial p_{n_e}} = E[\mathscr{C}_p^{n_e} \hat{V}] + E[\mathscr{D}_p^{n_e} \Delta \hat{V}_{n_e}], \tag{3.49}$$

where $\mathscr{C}_p^k \hat{V} = \mathscr{C}_e^k + \sum_{m=1}^{\upsilon_k} \mathscr{C}_{p,m}^k$, *in which* $\mathscr{C}_{p,m}^{n_e}$ *and* $\mathscr{C}_{p,m}^k$ *are given by*

$$\mathscr{C}_e^k = \frac{\partial \eta_{\upsilon_e+1}^{n_e}}{\partial p_k} \left(- \tau_{\upsilon_e+1}^{n_e} + T + x_{\upsilon_e+1}^{n_e} - S_{n_e} \right) \prod_{n=0}^{\upsilon_e} I_{\{U_n^{n_e} < \eta_n^{n_e}\}} I_{\{x_{\upsilon_e+1}^{n_e} \geq 0\}} \qquad for \ k < n_e,$$

$$\mathscr{C}_e^{n_e} = 0;$$

and

$$\mathscr{C}_{p,m}^k = \frac{\partial \eta_m^k}{\partial p_k} E\Big\{ \Big[c(X_{m-1}^k)(\tau_m^k - t_{m-1}^k) + \sum_{n=k+1}^{n_e} \Delta \hat{V}_n | t_0^{k+1} = \tau_m^k \Big] \prod_{n=0}^{m-1} I_{\{U_n^k < \eta_n^k\}}$$

$$- \sum_{j=m}^{v_k-1} \Big[\sum_{i=m}^{j} c(X_{i-1})h + c(X_j)(\tau_{j+1}^k - t_j^k) + \sum_{n=k+1}^{n_e} \Delta \hat{V}_n | t_0^{k+1} = \tau_{j+1}^k \Big]$$

$$\times \prod_{n=0, n \neq m}^{j} I_{\{U_n^k < \eta_n^k\}} I_{\{U_{n+1}^k \geq \eta_{n+1}^k\}} \Big| U_m^k = \eta_m^k \Big\},$$

(3.50)

$\mathscr{D}_p^k \Delta \hat{V}_k$ and $\mathscr{D}_p^k \Delta \hat{V}_{n_e}$ are defined by (3.35) and (3.36), respectively.

Proof. Assume $k < n_e$ and let $U_i^{k\star} = \eta_i^k(p_k)$. Without loss of generality, assume that

$U_i^{k\star} \leq U_i^k \leq \eta_i^k(p_k + \Delta\theta)$ and $U_i^{k\star} \geq U_i^k \geq \eta_i^k(p_k - \Delta\theta)$. Then

$$\frac{\partial}{\partial p_k} E[\hat{V}(p_k)] = \lim_{\Delta\theta \to 0} \frac{E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta)]}{2\Delta\theta} = E\Big[\sum_{m=1}^{v_k} \Xi_m^k + \Xi_e^k \Big] + \Pi + o(\Delta\theta),$$

where

$$\Pi = \lim_{\Delta\theta \to 0} \frac{E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta) | \overline{\mathcal{B}}] P(\overline{\mathcal{B}})}{2\Delta\theta},$$

$$\Xi_m^k = \frac{E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta) | \mathcal{B}_m^k] P(\mathcal{B}_m^k)}{2\Delta\theta},$$

$$\Xi_e^k = \frac{E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta) | \mathcal{B}_{v_e+1}^{n_e}] P(\mathcal{B}_{v_e+1}^{n_e})}{2\Delta\theta},$$

and $o(\Delta\theta)$ denotes the higher order terms in $\Delta\theta$.

For $1 \leq m \leq \upsilon_k$, Ξ_m is given by

$$
\begin{aligned}
\Xi_m^k &= \lim_{\Delta\theta \to 0} \frac{E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta)|\mathcal{B}_m^k]P(\mathcal{B}_m^k)}{2\Delta\theta} \\
&= \lim_{\Delta\theta \to 0} E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta)|\mathcal{B}_m^k] \lim_{\Delta\theta \to 0} \frac{P(\mathcal{B}_m^k)}{2\Delta\theta}.
\end{aligned}
$$

Since $\{U_i^k\}$ are random variables uniformly distributed on $[0, 1]$, we have

$$
\lim_{\Delta\theta \to 0} \frac{P(\mathcal{B}_m^k)}{2\Delta\theta} = \lim_{\Delta\theta \to 0} \frac{\int_{\eta_m^k(p_k - \Delta\theta)}^{\eta_m^k(p_k + \Delta\theta)} dt}{2\Delta\theta} = \lim_{\Delta\theta \to 0} \frac{\frac{\partial \eta_m^k(p_k)}{\partial p_k} \times 2\Delta\theta + o(\Delta\theta)}{2\Delta\theta} = \frac{\partial \eta_m^k(p_k)}{\partial p_k}.
$$

The second equality is obtained by using the mean value theorem. Furthermore, it is also easy to show that $\Xi_m^k = E[\mathscr{C}_p^k \hat{V}]$. Similarly we can also show that $\Xi_e^k = E[\mathscr{C}_e^k \hat{V}]$.

By Lemma 3.4 and the fact that $\lim_{\Delta\theta \to 0} P(\overline{\mathcal{B}}) = 1$, we have

$$
\Pi = \lim_{\Delta\theta \to 0} \frac{E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta)|\overline{\mathcal{B}}]P(\overline{\mathcal{B}})}{2\Delta\theta} = E[\mathscr{D}_p^k \Delta\hat{V}_k + \mathscr{D}_p^k \Delta\hat{V}_{n_e}].
$$

Now consider the case that $U_i^{k\star} - \eta_i^k(p_k + \Delta\theta)$ and $U_i^{k\star} - \eta_i^k(p_k - \Delta\theta)$ have the same sign; the indicator function $I_{\{U_i^k < \eta_i^k\}}$ is always 0 or 1 in this case for $i = 1, \cdots, \upsilon_k$, and hence there are no points of discontinuity in $\hat{V}$. Therefore by Lemma 3.4, we have

$$
\frac{\partial}{\partial p_k} E[\hat{V}(p_k)] = \lim_{\Delta\theta \to 0} \frac{E[\hat{V}(p_k + \Delta\theta) - \hat{V}(p_k - \Delta\theta)]}{2\Delta\theta} = E[\mathscr{D}_p^k \Delta\hat{V}_k + \mathscr{D}_p^k \Delta\hat{V}_{n_e}] \qquad k < n_e.
$$

Note that if $U_i^{k\star} - \eta_i^k(p_k + \Delta\theta)$ and $U_i^{k\star} - \eta_i^k(p_k - \Delta\theta)$ have the same sign for small enough $\Delta\theta$, we have $\frac{\partial \eta_i^k(p_k)}{\partial p_k}\big|_{\eta_i^k = U_i^{k\star}} = 0$. Hence (3.48) still holds. (3.49) can also be proved similarly. ∎

By Theorem 3.2, we can use the following gradient estimators to estimate the gra-

dient of $E[\hat{V}]$ with respect to prices $\{p_k, k = 1, \cdots, n_e\}$ along a sample path generated by Algorithm 1:

$$\mathscr{G}_p^k \hat{V} = \mathscr{C}_p^k \hat{V} + \mathscr{D}_p^k \Delta \hat{V}_k + \mathscr{D}_p^k \Delta \hat{V}_{n_e}, \qquad \text{for } k < n_e;$$

$$\mathscr{G}_p^{n_e} \hat{V} = \mathscr{C}_p^{n_e} \hat{V} + \mathscr{D}_p^{n_e} \Delta \hat{V}_{n_e}.$$

Remark 3.1 *The above gradient estimator contains two parts. For example, for $k < n_e$, the first part is $\mathscr{D}_p^k \Delta \hat{V}_k + \mathscr{D}_p^k \Delta \hat{V}_{n_e}$, which is derived from the continuous parts of the sample path. The other part is $\mathscr{S}_p^k \hat{V}$, which is derived from the discontinuous parts of the sample path by using the smoothed perturbation analysis technique.*

3.3.4   Stochastic Approximation Algorithm

With the gradient estimators given above, we seek to use stochastic approximation (SA) to solve the dynamic pricing problem (3.12). The SA algorithm goes back to the pioneering paper by [88]. For a recent exposition, please see [69]. The classical SA algorithm solves the following optimization problem by mimicking the gradient ascend method: $\max_\theta E[J(\theta)]$. Let $g(\theta) = \frac{\partial E[J(\theta)]}{\partial \theta}$ denote the gradient of $E[J(\theta)]$ with respect to the parameter θ. Let $\hat{g}(\theta)$ be an estimate of the gradient $g(\theta)$; SA generates iterates by the formula $\theta^{(n+1)} = \Pi_\Theta(\theta^{(n)} + a^{(n)}\hat{g}(\theta^{(n)}))$, where $\theta^{(n)}$ is the value of the parameter at the beginning of the iteration n, $a^{(n)}$ is a sequence of positive step sizes, and Π_Θ is a projection onto the parameter set Θ.

In our setting, to maximize $E\left[\sum_{k=1}^N \Delta \hat{V}_k\right]$ over the admissible pricing policy π_f, we assume that we have an initial pricing policy $\pi_f^{(0)} = \{p_k^{(0)}, k = 1, \cdots, N\}$, and a sequence

of step sizes $\{a^{(n)}\}$, which have the following properties:

$$a^{(n)} > 0, \qquad \lim_{n \longrightarrow \infty} a^{(n)} = 0, \qquad \sum_{n=1}^{\infty} a^{(n)} = +\infty, \qquad \sum_{n=1}^{\infty} [a^{(n)}]^2 < +\infty. \qquad (3.51)$$

Algorithm 2: Stochastic Approximation Algorithm

- Step 1. Given an initial feasible set of prices $\{p_k^{(0)}, k = 1, \cdots, N\}$

- Step 2. For $n = 1$ to I_n do :

  – Generate a sample path by Algorithm 1 and compute a gradient estimate using Theorem 3.2 :$\mathscr{G}_p^k \hat{V}(p_k^{(n-1)})$ for $k = 1, \cdots, n_e$. Pick a new step size $a^{(n)}$.

  – Update the pricing policy using $p_k^{(n)} := \Pi_{\mathcal{P}}(p_k^{(n-1)} + a^{(n)}\mathscr{G}_p^k \hat{V}(p_k^{(n-1)}))$, for $k = 1, \cdots, n_e$.

- Step 3. Return $\{p_k^{(I_n)}, k = 1, \cdots, n_e\}$, and stop.

3.3.4.1 Convergence of the Stochastic Approximation Algorithm

We show that Algorithm 2 has fairly robust local convergence properties. In Algorithm 2, the gradient estimator $\mathscr{G}_p^k \hat{V}(p_k^{(n)})$ is in fact a noisy estimate of the gradient of $E[\hat{V}(p_k^{(n)})]$. Let the bias (error) of the gradient estimate at iteration n be $\beta_k^{(n)}$:

$\beta_k^{(n)} = \mathscr{G}_p^k \hat{V}(p_k^{(n)}) - \frac{\partial E[\hat{V}]}{\partial p_k}$ for $k = 1, \cdots, n_e$. From Theorem 3.2, we know that

$$E[\beta_k^{(n)}|p_k^{(0)}, p_k^{(1)}, \cdots, p_k^{(n-1)}] = 0.$$

To simplify the notation, define the cumulative step size as $s_n = \sum_{i=1}^{n-1} a^{(i)}$ and define a function $\rho(s) = \max\{n : s_n \leq s\}$. To prove the convergence of Algorithm 2, we make the following assumptions.

Assumption 3.3 $\{a^{(n)}\}$ *is a sequence of positive real numbers such that* $a^{(n)} > 0$, *and* $\lim_{n \to \infty} a^{(n)} = 0$.

Assumption 3.4 *The constraint set* $\mathcal{P}$ *is closed and bounded, and is defined by* $\mathcal{P} = \{p : q_j(p) \leq 0, j = 1, \cdots, s\}$. $\{q_j(p)\}$ *are continuously differentiable. At each* $p \in \partial \mathcal{P}$, *the gradients of the active constraints are linearly independent.*

Assumption 3.5 *For each* $\epsilon > 0$ *and* $s > 0$, $\lim_{n \to \infty} P(\sup_{n \leq l \leq \rho(s_k+s)} \| \sum_{i=n}^{l} a^{(i)} \beta_k^{(i)} \| > \epsilon) = 0$.

Assumption 3.6 *The expectation of the value function* $E[\hat{V}]$ *is continuously differentiable.*

Assumption 3.7 $a^{(n)} E[\|\beta_k^{(n)}\|^2] \to 0$ *as* $n \to 0$.

To validate Assumptions 3.5, 3.6, and 3.7, we first prove the following two lemmas.

Lemma 3.5 *If Assumptions 3.1 and 3.2 hold,* $E[\hat{V}]$ *is continuously differentiable in* p_k *for* $k \leq N$.

Proof. The indicator functions in the gradient estimator $\mathscr{G}_p^k \hat{V}$ make it discontinuous at $U_m^k = \eta_m^k$ for $m = 1, \cdots, \upsilon_k - 1$. However, U_m^k is a continuous random variable; hence the gradient $\mathscr{G}_p^k \hat{V}$ is continuous w.p.1. Therefore for any sequence of $\{p_k^i\}$, $\lim_{p_k^i \to p_k} \mathscr{G}_p^k \hat{V}(p_k^i) = \mathscr{G}_p^k \hat{V}(p_k)$, w.p.1. Applying Assumption 3.2 and Lemma 3.1 yields

$$
\begin{aligned}
\frac{\partial \eta_{m+1}^k}{\partial p_k} &= \left(1 - \exp\left(-\frac{2(x_m^k - S_k)(x_{m+1}^k - S_k)}{h\sigma^2}\right)\right)\frac{2}{h\sigma^2} \\
&\quad \times \left(\frac{\partial x_m^k}{\partial p_k}(x_{m+1}^k - S_k) + \frac{\partial x_{m+1}^k}{\partial p_k}(x_m^k - S_k)\right) \\
&\leq \frac{2}{h\sigma^2}(\Lambda_m^k \bar{\Lambda}_{m+1}^k + \Lambda_{m+1}^k \bar{\Lambda}_m^k).
\end{aligned}
$$

Combining the above inequality, Assumption 3.2, and Lemma 3.1, we have

$$\|\mathscr{C}_p^k \hat{V}\| \leq \sum_{m=0}^{v_k} \frac{2}{h\sigma^2}(\Lambda_m^k \bar{\Lambda}_{m+1}^k + \Lambda_{m+1}^k \bar{\Lambda}_m^k)(\Phi_m^k h + 2(\bar{p}S + \sum_{i=1}^{N} \sum_{j=0}^{v_i} \Phi_j^i h))$$
$$+ \frac{2}{h\sigma^2}(\Lambda_{v_e}^{n_e} \bar{\Lambda}_{v_e+1}^{n_e} + \Lambda_{v_e+1}^{n_e} \bar{\Lambda}_{v_e}^{n_e})T.$$

Since $\{\Lambda_m^k\}$, $\{\bar{\Lambda}_m^k\}$, and $\{\Phi_m^k\}$ all have finite nth moment for any finite positive integer n, we have $E[\|\mathscr{C}_p^k \hat{V}\|] < \infty$. Together with Lemma 3.4, we have $E[\|\mathscr{G}_p^k \hat{V}\|] < \infty$. By the dominated convergence theorem

$$\lim_{p_k^i \to p_k} \frac{\partial E[\hat{V}(p)]}{\partial p}\bigg|_{p=p_k^i} = E[\mathscr{G}_p^k \hat{V}(p_k)]. \tag{3.52}$$

Applying (3.52) and Theorem 3.2 yields

$$\lim_{p_k^i \to p_k} \frac{\partial E[\hat{V}(p)]}{\partial p}\bigg|_{p=p_k^i} = \frac{\partial E[\hat{V}(p)]}{\partial p}\bigg|_{p=p_k}.$$

∎

The next lemma shows that the gradient estimators $\{\mathscr{G}_p^k \hat{V}\}$ given above have finite variances.

Lemma 3.6 *If Assumptions 3.1 and 3.2 hold, there exist positive constants $\{\mathcal{U}_k\}$, such that $E[(\mathscr{G}_p^k \hat{V})^2] < \mathcal{U}_k$.*

Proof. First we prove the theorem for $k < n_e$. The result for $k = n_e$ can be proved in a similar way.

By the Cauchy-Schwarz inequality,

$$E[(\mathscr{G}_p\hat{V})^2] = E[(\mathscr{C}_p^k\hat{V})^2 + 2\mathscr{C}_p^k\hat{V}_k(\mathscr{D}_p^k\Delta\hat{V}_k + \mathscr{D}_p^k\Delta\hat{V}_{n_e}) + (\mathscr{D}_p^k\Delta\hat{V}_k + \mathscr{D}_p^k\Delta\hat{V}_{n_e})^2]$$

$$\leq E[(\mathscr{C}_p^k\hat{V})^2] + E[(\mathscr{D}_p^k\Delta\hat{V}_k + \mathscr{D}_p^k\Delta\hat{V}_{n_e})^2]$$

$$+ 2\sqrt{E[(\mathscr{C}_p^k\hat{V})^2]}\sqrt{E[(\mathscr{D}_p^k\Delta\hat{V}_k + \mathscr{D}_p^k\Delta\hat{V}_{n_e})^2]}.$$

Therefore, to prove the theorem, we only need to prove that $E[(\mathscr{C}_p^k\hat{V})^2]$ and $E[(\mathscr{D}_p^k\Delta\hat{V}_k + \mathscr{D}_p^k\Delta\hat{V}_{n_e})^2]$ are upper bounded by constants.

By Assumption 3.2 and Lemma 3.1, we have

$$\frac{\partial\eta_m^k}{\partial p_k}$$

$$= \left(1 - \exp\left(-\frac{2(x_m^k - S_k)(x_{m+1}^k - S_k)}{h\sigma^2}\right)\right)\frac{2}{h\sigma^2}\left(\frac{\partial x_m^k}{\partial p_k}(x_{m+1}^k - S_k) + \frac{\partial x_{m+1}^k}{\partial p_k}(x_m^k - S_k)\right)$$

$$\leq \frac{2}{h\sigma^2}(\Lambda_m^k\bar{\Lambda}_{m+1}^k + \Lambda_{m+1}^k\bar{\Lambda}_m^k). \tag{3.53}$$

From Assumption 3.2 and Lemma 3.1, we know Λ_m^k and $\bar{\Lambda}_m^k$ are random variables with finite nth moments, for any positive integer n. By using Cauchy-Schwarz inequality, we can show that there exist constants $\{C_m^k\}$ such that

$$E[(\Phi_m^k)^4(\Lambda_m^k\bar{\Lambda}_{m+1}^k + \Lambda_{m+1}^k\bar{\Lambda}_m^k)^4] \leq C_m^k \qquad \forall\, m < T/h. \tag{3.54}$$

Combining Assumptions 3.1 and 3.2, Lemma 3.1, inequalities (3.53), (3.54), (3.58) and

the fact that $\upsilon_{n_e} h \leq T$, by Cauchy-Schwarz inequality we have

$$
\begin{aligned}
E[(\mathscr{C}_p^k \hat{V})^2] &\leq E\Big[\Big(\sum_{m=0}^{\upsilon_k} \frac{2}{h\sigma^2}(\Lambda_m^k \bar{\Lambda}_{m+1}^k + \Lambda_{m+1}^k \bar{\Lambda}_m^k)\Phi_m^k T\Big)^2\Big] \\
&\leq \frac{4T^4}{\sigma^4 h^4} \sum_{i=0,j=0,i>j}^{\lceil T/h \rceil} \Big\{ E\Big[\sqrt{(\Phi_i^k)^4(\Lambda_i^k \bar{\Lambda}_{i+1}^k + \Lambda_{i+1}^k \bar{\Lambda}_i^k)^4}\Big] \\
&\quad \times E\Big[\sqrt{(\Phi_j^k)^4(\Lambda_j^k \bar{\Lambda}_{j+1}^k + \Lambda_{j+1}^k \bar{\Lambda}_j^k)^4}\Big] \Big\} \\
&\leq U^k, \tag{3.55}
\end{aligned}
$$

where $U^k = \frac{4T^4}{\sigma^4 h^4} \sum_{i=0,j=0,i>j}^{\lceil T/h \rceil} C_i^k C_j^k$.

By Assumptions 3.1 and 3.2, and Lemmas 3.1 and 3.3, $E[(\mathscr{D}_p^k \Delta \hat{V}_k)^2]$ is bounded by

$$
\begin{aligned}
&E[(\mathscr{D}_p^k \Delta \hat{V}_k)^2] \\
&\leq E\Big[\Big(S + \sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k h + \sum_{m=0}^{\upsilon_k}\big(\Phi_m^k E[\mathcal{G}_k(\tau_{m+1}^k, X_m^k, X_{m+1}^k)] \prod_{j=0}^{m} I_{\{U_j^k < \eta_j^k\}}\big)\Big)^2\Big] \\
&= E\Big[S^2 + \big(\sum_{m=0}^{\upsilon_k}(\bar{\Phi}_m^k \Lambda_m^k h)^2 + \sum_{m=0}^{\upsilon_k}(\Phi_m^k)^2 E[(\mathcal{G}_k)^2(\tau_{m+1}^k, X_m^k, X_{m+1}^k)]I_{\{U_m^k < \eta_m^k\}} \\
&\quad + 2S\Big(\sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k h + \sum_{m=0}^{\upsilon_k} \Phi_m^k E[\mathcal{G}_k(\tau_{m+1}^k, X_m^k, X_{m+1}^k)]I_{\{U_m^k < \eta_m^k\}}\Big) \\
&\quad + 2\sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k \big(\Phi E[\mathcal{G}_k | X_{\upsilon_k}^k, X_{\upsilon_k+1}^k]I_{\{U_{\upsilon_k}^k < \eta_{\upsilon_k}^k\}}\Big]. \tag{3.56}
\end{aligned}
$$

Since $\{\Phi_m^k\}, \{\bar{\Phi}_m^k\}, \{\Lambda_m^k\}$ have finite nth moment for any positive integer n, and $\upsilon_k < T/h$, similar to the proof of (3.55), there exist positive constants η_1 and η_2, such that

$$
E[(\sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k h)^2] < \eta_1, \quad E[\sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k h] < \eta_2. \tag{3.57}
$$

As in the proof of (3.44) and (3.45), we can show that

$$E\Big[\sum_{m=0}^{\upsilon_k} \Phi_m^k E[\mathcal{G}_k | X_{\upsilon_k}^k, X_{\upsilon_k+1}^k] I_{\{U_{\upsilon_k}^k < \eta_{\upsilon_k}^k\}}\Big] < \mu_1,$$

$$E\Big[\sum_{m=0}^{\upsilon_k} \bar{\Phi}_m^k \Lambda_m^k \Phi_m^k E[\mathcal{G}_k | X_{\upsilon_k}^k, X_{\upsilon_k+1}^k]\Big] < \mu_2, \tag{3.58}$$

where μ_1 and μ_2, are positive constants. By Assumptions 3.1 and 3.2, and Lemmas 3.1 and 3.3, as in the proof of (3.45), we have

$$E\Big[\sum_{m=0}^{\upsilon_k} (\Phi_m^k)^2 E[(\mathcal{G}_k)_k^2(\tau_{m+1}^k, X_m^k, X_{m+1}^k)] I_{\{U_m^k < \eta_m^k\}}\Big] \leq \mu_3. \tag{3.59}$$

Substituting (3.57), (3.58), and (3.59) into (3.56) yields

$$E[(\mathscr{D}_p^k \Delta \hat{V}_k)^2] \leq S^2 + \eta_1 + \mu_3 + 2S(\eta_2 + \mu_1) + 2\mu_2 = \beta_1^k, \tag{3.60}$$

where β_k is a positive constant. Similarly, we can prove that there exist positive constants β_2^k, β_3^k such that

$$E[(\mathscr{D}_p^k \Delta \hat{V}_{n_e})^2] \leq \beta_2^k, \quad E[\mathscr{D}_p^k \Delta \hat{V}_k \mathscr{D}_p^k \Delta \hat{V}_{n_e}] \leq \beta_3^k. \tag{3.61}$$

Combining (3.55), (3.60), and (3.61) yields

$$E[(\mathscr{G}_p^k \hat{V})^2] \leq U^k + \beta_1^k + \beta_2^k + 2\beta_3^k + 2\sqrt{U^k}\sqrt{\beta_1^k + \beta_2^k + 2\beta_3^k} = \mathcal{U}_k.$$

∎

The following theorem shows the convergence property of Algorithm 2.

71

Theorem 3.3 *For the optimization problem (3.12), let $\mathcal{P}^\star$ be the set of Kuhn-Tucker points. Then Assumptions 3.3, 3.4, 3.5, 3.6, and 3.7 hold for the Stochastic Approximation Algorithm 2. If $\mathcal{P}^\star$ is connected, the sequence of points $\{p_k^{(n)}\}$ converges to a point in $\mathcal{P}^\star$ in probability as n goes to infinity.*

Proof. We prove the theorem by applying Theorem 6.3.1 of [67]. First we verify that Assumptions 3.3, 3.4, 3.5, 3.6, and 3.7 hold for our algorithm. Assumption 3.3 holds since we choose the step-size according to (3.51). Assumption 3.4 holds because the constrained set is $\{\underline{p} \leq p_k \leq \bar{p}\}$. Note that $\|\sum_{i=n}^{l} a^{(i)} \beta_k^{(i)}\|, l \geq n$ is a submartingale since $\sum_{i=n}^{l} a^{(i)} \beta_k^{(i)}, l \geq n$ is a martingale and the Euclidean norm is convex. Hence

$$P\Big(\sup_{n \leq l \leq \rho(s_k+s)} \|\sum_{i=n}^{l} a^{(i)} \beta_k^{(i)}\| > \epsilon\Big) = P\Big(\sup_{n \leq l \leq \rho(s_k+s)} \big(\sum_{i=n}^{l} a^{(i)} \beta_k^{(i)}\big)^2 > \epsilon^2\Big)$$

$$\leq \frac{E\Big[\big(\sum_{i=k}^{\rho(s_k+s)} a^{(i)} \beta_k^{(i)}\big)^2\Big]}{\epsilon^2}$$

$$= \frac{\sum_{i=n}^{\rho(s_k+s)} [a^{(i)}]^2 E[[\beta_k^{(i)}]^2]}{\epsilon^2}$$

$$\leq \frac{\sum_{i=n}^{\infty} [a^{(i)}]^2 E[[\beta_k^{(i)}]^2]}{\epsilon^2}.$$

Assumption 3.6 holds because of Lemma 3.5. Assumption 3.7 is valid by the choice of $a^{(i)}$ and Lemma 3.6. By Theorem 6.3.1 of [67], we complete the proof. ∎

3.4 Joint Optimization of Dynamic Pricing and Initial Order

Algorithm 1 and Algorithm 2 address the dynamic pricing problem with a fixed amount of inventory. Now we extend the dynamic pricing problem and allow to choose the initial order S freely. A similar pricing policy as in Section 3.3.1 is adopted, in which the price changes at $S_k = \frac{N-k}{N} S$ for $i = 1, \cdots, N$. The inventory levels at which the price

72

can be changed are proportional to the initial inventory level with a fixed ratio. Given an initial order S and a pricing policy π_f, we can use Algorithm 1 to generate sample paths, along which discretized incremental value functions $\{\Delta \hat{V}_k\}$ can be calculated by (3.11). The joint dynamic pricing and initial order optimization problem can be formulated as

$$\sup_{\pi_f, S} E\left[\hat{V}(S)\Big| X_0 = S\right]. \tag{3.62}$$

Similar to the dynamic pricing problem (3.12), we use SA to solve (3.62). The gradient estimators $\{\mathscr{G}_p^k \hat{V}\}$ still work here. In the following, we derive gradient estimators for the expectation of the discretized value function with respect to the initial order S.

3.4.1 Derivative estimator with respect to the initial order

The gradient of x_i^k with respect to S can be easily obtained. For $x_i^k > S_k$, $x_{i+1}^k > S_k$ and $t_{i+1}^k \le T - h$,

$$\frac{\partial x_{i+1}^k}{\partial S} = \frac{\partial x_i^k}{\partial S} - \left(\frac{\partial \sigma(X_0 - x_i^k, p_k)}{\partial x_i^k}\frac{\partial x_i^k}{\partial S}\right)\sqrt{t_{i+1}^k - t_i^k}z_{i+1}^k.$$

The boundary conditions are $\frac{\partial X_0}{\partial S} = 1$, $\frac{\partial x_{\tau_k}^k}{\partial S} = \frac{N-k}{N}$.

When $x_i^k > S_k$, and $t_{i+1}^k > T - h > t_i^k$, we have

$$\frac{\partial x_{i+1}^k}{\partial S} = \frac{\partial x_i^k}{\partial S} + \lambda(p_k)\frac{\partial \tau_{k-1}}{\partial S} + \frac{\partial \sigma(X_0 - x_i^k, p_k))}{\partial x_i^k}\frac{\partial x_i^k}{\partial S}\sqrt{T - \tau_{k-1} - ih}z_{i+1}^k$$

$$+ \frac{\sigma(X_0 - x_i^k, p_k)}{2\sqrt{T - \tau_{k-1} - ih}}\frac{\partial \tau_k}{\partial S}z_{i+1}^k.$$

3.4.2 Sample Path-Based Gradient Estimators for Stopping Times

We derive a sample path-based gradient estimator for the expectation of τ_{i+1}^k with respect to S along a similar line as in Section 3.3.3.3. When $x_i^k > S_k$ and $x_{i+1}^k < S_k$, the gradient estimator is given by

$$\frac{\partial \tau_{i+1}^k}{\partial S} = \frac{-\frac{N-k}{N} + \frac{\partial x_i^k}{\partial S}\frac{t_{i+1}^k - \tau_{i+1}^k}{h} + \frac{\partial x_{i+1}^k}{\partial S}\frac{\tau_{i+1}^k - t_i^k}{h} - \frac{\partial \sigma(X_0 - x_i^k, p_k)}{\partial S}\sqrt{N(\tau_{i+1}^k)}Z_k}{-\frac{x_{i+1}^k}{h} + \frac{x_i^k}{h} + \sigma(X_0 - x_i^k, p_k)\frac{\frac{\partial N(\tau_{i+1}^k)}{\partial \tau_{i+1}^k}}{2\sqrt{N(\tau_{i+1}^k)}}Z_k}$$

$$= \mathscr{D}_S^\tau(\tau_{i+1}^k; x_i^k, x_{i+1}^k), \tag{3.63}$$

where Z_k is defined in (3.28). When $x_i^k > S_k$ and $x_{i+1}^k > S_k$, the gradient estimator is given by $\frac{\partial \tau_{i+1}^k}{\partial S} = \mathscr{D}_S^\tau(\hat{\tau}_k; x_i^k, \tilde{x}_{i+1}^k)$ where $\tilde{x}_{i+1}^k = 2S_k - x_{i+1}^k$.

3.4.3 Likelihood Ratio Gradient Estimators for Stopping Times

A likelihood ratio gradient estimator of the expectation of the stopping time τ_{i+1}^k with respect to S can be obtained in a similar way as we derived (3.23) and (3.24). The likelihood ratio gradient estimator is derived by assuming the interchangeability of the derivative and the integral:

$$\frac{\partial E[\tau_{i+1}^k | x_i^k, x_{i+1}^k]}{\partial S} = \int_{t_i}^{t_{i+1}} \tau \frac{\partial \ln f_k(\tau)}{\partial S} f_k(\tau) d\tau.$$

By using the inverse Gaussian distribution, we can rewrite the above equation as

$$\frac{\partial E[\tau_{i+1}^k | x_i^k, x_{i+1}^k]}{\partial S} = \int_0^\infty \frac{hy}{1+y} \frac{\partial \ln g_k(y; \lambda_k, \theta_k)}{\partial S} g_k(y; \lambda_k, \theta_k) dy.$$

We use $\mathscr{R}_S(\tau_{i+1}^k; x_i^k, x_{i+1}^k) = \frac{hy}{1+y} \frac{\partial \ln g_k(y; \lambda_k, \theta_k)}{\partial S}$ to denote the likelihood ratio gradient esti-

mator.

3.4.4 Gradient Estimator for Value Function

Similarly as in Section 3.3.3.4, the values of the indicator functions in $\Delta \hat{V}_k$ change

at $U_i^{k\star} = \eta_i^k(S)$. We use conditional expectation to smooth out the discontinuities. Now

we perturb S to $S + \Delta\theta$ and $S - \Delta\theta$ for a small $\Delta\theta$. The values of η_i^k at the new

parameters are $\eta_i^k(S + \Delta\theta)$ and $\eta_i^k(S - \Delta\theta)$, respectively. Without loss of generality,

assume $U_i^{k\star} \leq \eta_i^k(S + \Delta\theta)$, and $U_i^{k\star} \geq \eta_i^k(S - \Delta\theta)$. Define the following sets

$$\mathcal{A}_i^k(U_i^{k\star}, S, \Delta\theta) = \left\{ \; U_i^k : \begin{array}{l} U_i^{k\star} \leq U_i^k \leq \eta_i^k(S + \Delta\theta); \\[2mm] \text{and } U_i^{k\star} \geq U_i^k \geq \eta_i^k(S - \Delta\theta). \end{array} \right. \qquad \text{for } i = 1, \cdots, \upsilon_k.$$

$\mathcal{A}_i^k(U_i^{k\star}, S, \Delta\theta)$ contains U_i^k such that $U_i^k - \eta_i^k$ with the price $S + \Delta\theta$ has a different sign than

it does with the price $S - \Delta\theta$. Let $\overline{\mathcal{A}_i^k}(U_i^{k\star}, S, \Delta\theta)$ be the complement of $\mathcal{A}_i^k(U_i^{k\star}, S, \Delta\theta)$.

For simplicity, we use $\mathcal{A}_i^k$ to represent $\mathcal{A}_i^k(U_i^{k\star}, S, \Delta\theta)$, and define $\mathcal{A} = \bigcup_{i=1}^{\upsilon_k} \mathcal{A}_i^k$.

Restricted to the set $\overline{\mathcal{A}}$, the signs of the indicator functions in $\Delta \hat{V}_k$ will not change.

By direct differentiation, we can easily obtain the following gradient estimator for $E[\{\Delta \hat{V}_k\} | \overline{\mathcal{A}}]$:

$$\mathscr{D}_S \Delta \hat{V}_k = -p_k \frac{\partial \Delta S_k}{\partial S} - \sum_{m=0}^{\upsilon_k} \left\{ \left[\sum_{i=0}^{m-1} \left(\frac{\partial c(X_i^k)}{\partial S} h + \frac{\partial c(X_m^k)}{\partial S} (E[\tau_{m+1}^k | (X_m^k, X_{m+1}^k)] - t_m^k) \right. \right. \right.$$

$$\left. \left. \left. + c(X_m^k) E[\mathcal{G}_S(\tau_{m+1}^k; X_m^k, X_{m+1}^k)] \right) \right] \prod_{j=0}^{m} I_{\{U_j^k < \eta_j^k\}} I_{\{U_{m+1}^k \geq \eta_{m+1}^k\}} \right\}, \qquad (3.64)$$

where

$$\mathcal{G}_S(\tau_{m+1}^k; X_m^k, X_{m+1}^k) = \mathscr{D}_S^\tau(\tau_{m+1}^k; X_m^k, X_{m+1}^k) \quad \text{when } X_m^k > S_k, X_{m+1}^k \le S_k,$$

$$\mathcal{G}_S(\tau_{m+1}^k; X_m^k, X_{m+1}^k) = \mathscr{D}_S^\tau(\tau_{m+1}^k; X_m^k, 2S_k - X_{m+1}^k) \quad \text{when } X_m^k > S_k, X_{m+1}^k > S_k.$$

The following lemma states that the above gradient estimator is unbiased.

Lemma 3.7 *If Assumptions 3.1 and 3.2 hold, we have*

$$\frac{\partial E\big[\Delta \hat{V}_k | \overline{\mathcal{A}}\big]}{\partial S} = E\big[\mathscr{D}_S \Delta \hat{V}_k | \overline{\mathcal{A}}\big],$$

where $\mathscr{D}_S$ is defined by (3.64), i.e., the gradient estimator (3.64) is unbiased.

Proof. The proof is similar to the proof of Lemma 3.4. ∎

Theorem 3.4 *If Assumptions 3.1 and 3.2 hold, we have*

$$\frac{\partial E[\Delta \hat{V}_k]}{\partial S} = E[\mathscr{S}_S \Delta \hat{V}_k] + E\big[\mathscr{D}_S \Delta \hat{V}_k\big], \tag{3.65}$$

where $\mathscr{S}_S \Delta \hat{V}_k = \sum_{m=1}^{v_k} \mathscr{C}_{S,m}^k$, in which $\mathscr{C}_{S,m}^k$ is given by

$$\begin{aligned}
\mathscr{C}_{S,m}^k = \frac{\partial \eta_m^k}{\partial S} E\Big\{ &\big[c(X_{m-1}^k)(\tau_m^k - t_{m-1}^k) + \sum_{n=k+1}^{n_e} \hat{V}_n | t_0^{k+1} = \tau_m^k\big] \prod_{n=0}^{m-1} I_{\{U_n^k < \eta_n^k\}} \\
&- \sum_{j=m}^{v_k} \Big[\sum_{i=m}^{j} c(X_{i-1})h + c(X_j)(\tau_{j+1}^k - t_j^k) + \sum_{n=k+1}^{n_e} \hat{V}_n | t_0^{k+1} = \tau_{j+1}^k \Big] \\
&\times \prod_{n=0, n \ne m}^{j} I_{\{U_n^k < \eta_n^k\}} I_{\{U_{n+1}^k \ge \eta_{n+1}^k\}} \Big\};
\end{aligned}$$

$\mathscr{D}_S \Delta \hat{V}_k$ is defined by (3.64); τ_m^k is the first time that the inventory level hits S_k given X_{m-1}^k and X_m^k.

The gradient estimator of $\hat{V}$ with respect to S is given by $\mathscr{G}_S\hat{V} = \sum_{k=1}^{n_e}\left(\mathscr{S}_S\Delta\hat{V}_k+\mathscr{D}_S\Delta\hat{V}_k\right)$.

3.5   Pricing Over Infinite Horizon

In this section, we consider selling all the inventory over an infinite horizon using a pricing policy π_f as in Section 3.3 . Similar to Algorithm 1, we use the following algorithm to simulate the discretized value function.

***Algorithm* 3. (*Simulation Algorithm Over Infinite Horizon*)**

Initialization: for a given pricing policy π_f, choose a time step $h > 0$. Let $k = 1, t_0^1 = 0$, $I_0 = 1$, $x_0^1 = S_0 = S$, and $\tau_0 = 0$.

- step 1: If $k > N$, stop; otherwise set $i = 0$, $t_0^k = \tau_{k-1}$ for $k > 1$, $I_k = 1$, and $x_0^k = S_{k-1}$. Go to step 2.

- step 2: Let the price be p_k according to the pricing policy π_f and generate $z_{i+1}^k \sim N(0,1)$. Compute x_{i+1}^k by (3.8).

  - When $x_{i+1}^k \geq S_k$, generate U_i^k uniformly on $[0,1]$, and compute η_{i+1}^k. If $I_k = 1$, let $I_k = I_k I_{\{U_i^k < \eta_{i+1}^k\}}$, $\theta_k = \frac{x_i^k - S_k}{x_{i+1}^k - S_k}$, and $\lambda_k = \frac{(x_i^k - S_k)^2}{h[\sigma(S-x_i^k,p_k)]^2}$, set $i = i+1$, then go to step 3.

  - When $x_i^k > S_k > x_{i+1}^k$, let $v_k = i$, $\theta_k = -\frac{x_i^k - S_k}{x_{i+1}^k - S_k}$, and $\lambda_k = \frac{(x_i^k - S_k)^2}{h[\sigma(S-x_i^k,p_k)]^2}$, then go to step 3.

- step 3: Generate a sample y_k from the inverse Gaussian distribution (3.14). Let $\tau_{i+1}^k = t_i^k + \frac{hy}{1+y}$.

  - When $I_k = 1$, if $X_{i+1}^k > S_k$, the price does not change and go to Step 2, and if $X_{i+1}^k \leq S_k$, let $\tau_k = \tau_{k+1}^k$, $k = k + 1$, and go to Step 1.

– When $I_k = 0$ and $X_{i+1}^k > S_k$, the price changes, let $\tau_k = \tau_{i+1}^k$, then go to Step 1.

With the above simulation scheme, the remaining formulation of the joint inventory control and dynamic pricing problem over an infinite horizon is very similar to that over a finite time horizon. Instead of solving the optimization problem (3.12), we solve the following optimization problem:

$$\sup_{\pi_f, S} E\Big[\sum_{i=1}^{N} \Delta \hat{V}_i \Big| X_0 = S\Big]. \tag{3.66}$$

Note that the arrival rate of the demand (3.1) at time t only depends on the price at t, and $\sigma(X_0 - X_t^k, p_k)$ depends on the cumulative demand and the price at t, but not explicitly on the time t. This property simplifies the problem of estimating the gradient of the expected discretized profit functions, and yields the following result.

Theorem 3.5 *Under Assumptions 3.1 and 3.2, we have*

$$\frac{\partial \sum_{i=1}^{N} E\Big[\Delta \hat{V}_i\Big]}{\partial p_k} = \frac{\partial E\Big[\Delta \hat{V}_k\Big]}{\partial p_k} = E\Big[\frac{\Delta \hat{V}_k}{\partial p_k}\Big].$$

Proof. From the boundary conditions (3.20), $x_{\tau_{k-1}}^{k-1}$ does not depend on p_{k-1}, and hence x_i^k does not depend on p_{k-1} by (3.8). In $\Delta \hat{V}_k$ defined by (3.11), $\tau_{m+1}^k - t_m^k$ only depends on x_m^k and x_{m+1}^k, which do not depend on p_{k-1}. It is easy to show iteratively that $\Delta \hat{V}_k$ does not depends on p_i for $i < k - 1$. Hence the theorem is proved. ∎

The gradient estimators given by (3.19) and the boundary condition (3.20) for $\{x_i^k\}$ are still valid. We can also use the gradient estimator (3.29) for $\{\tau_{i+1}^k\}$. However, for dynamic pricing problems over an infinite horizon, the stopping time τ_N could be infinity,

which is different from the dynamic pricing problem over a finite horizon T. We show that the stopping times $\{\tau_k\}$ now have the following property.

Lemma 3.8 *If Assumptions 3.1 and 3.2 hold, there exist positive constants $\mathcal{M}_1$ and $\mathcal{M}_2$, such that the first and second moments of the stopping time $\{\tau_k\}$ satisfy*

$$E[\tau_k] \leq \mathcal{M}_1, \qquad E[(\tau_k)^2] \leq \mathcal{M}_2. \qquad \forall k = 1, \cdots, N.$$

Proof. We prove this lemma by applying optional stopping theory. Without loss of generality, we start from the inventory level S_{k-1} at time τ_{k-1} with the selling price p_k. From Algorithm 3, we know that

$$X_1^k = S_{k-1} - \lambda(p_k)h - \sigma\sqrt{h}Z_1^k,$$

$$\ldots$$

$$X_{i+1}^k = X_i^k - \lambda(p_k)h - \sigma\sqrt{h}Z_{i+1}^k.$$

Summing up all the above equations gives

$$X_{i+1}^k = S_{k-1} - \lambda(p_k)h \times i - \sum_{l=0}^{i} \sigma\sqrt{h}Z_{l+1}^k.$$

Define a new stochastic process $M_{i+1} = \sum_{l=0}^{i} g(X_0 - X_l^k, p_k)\sqrt{h}Z_{l+1}^k = S_{k-1} - \lambda(p_k)(i+1)h - X_{i+1}^k$. Let $\mathcal{F}_i^k$ be the filtration generated by random variables $\{Z_m^k, m = 1, \cdots, i\}$. Note that M_i is a martingale with respect to the filtration $\mathcal{F}_i^k$. It is easy to show that $\upsilon_k + 1 = \inf\{i + 1 : x_i^k > S_k, x_{i+1}^k \leq S_k\}$ is a stopping time. By optional stopping theory,

$E[M_{v_k+1}] = E[M_0] = 0$, i.e,

$$E[M_{v_k+1}] = E\big[S_{k-1} - \lambda(p_k)(v_k + 1)h - X^k_{v_k+1}\big]$$

$$= S_{k-1} - \lambda(p_k)E[v_k + 1]h - E[X^k_{v_k+1}] = 0. \qquad (3.67)$$

Note that $X^k_{v_k+1}$ is no greater than S_k. Therefore $E[X^k_{v_k+1}] \leq S_k$, together with (3.67),

yields $\frac{S_{k-1}-S_k}{\lambda(p_k)} - h \leq E[v_k h]$. From (3.67), we know that

$$S_{k-1} - \lambda(p_k)E[v_k + 1]h = E[X^k_{v_k+1}] = E[X^k_v - \lambda(p_k)h - \sigma\sqrt{h}Z^k_{v_k+1}]$$

$$> S_k - \lambda(p_k)h - \sigma\sqrt{h}\int_{\frac{X^k_{v_k}-\lambda h - S_k}{\sigma\sqrt{h}}}^{\infty} \frac{1}{\sqrt{2\pi}} z\exp\big(\frac{-z^2}{2}\big)dz > S_k - \lambda h - \sigma\sqrt{h}.$$

The first inequality holds since $X^k_{v_k}$ is greater than S_k. Hence we have

$$E[v_k h] < \frac{S_{k-1} - S_k + \sigma\sqrt{h}}{\lambda(p_k)}. \qquad (3.68)$$

Combining (3.68) and the fact $v_k h \leq \tau_k - \tau_{k-1} \leq (v_k + 1)h$ yields

$$E[\tau_k] - E[\tau_{k-1}] \leq \frac{S_{k-1} - S_k + \sigma\sqrt{h}}{\lambda(p_k)} + h \leq \frac{S_{k-1} - S_k}{\underline{\lambda}} + h.$$

By iteratively applying the above inequality to the stopping times τ_i for $i = 1, \cdots, k$, and

then taking the sum, we have

$$E[\tau_N] \leq \mathcal{M}_1, \qquad (3.69)$$

where $\mathcal{M}_1 = \frac{S_0 - S_k + N\sigma\sqrt{h}}{\underline{\lambda}} + Nh$.

To derive a bound on the second moment of the stopping times, we define a stochastic

process $A_{i+1} = (\sum_{l=0}^{i} g(X_0 - X_l^k, p_k)\sqrt{h}Z_{l+1}^k)^2 - \sum_{l=0}^{i} g^2(X_0 - X_l^k, p_k)h$. Note that A_{i+1}

is also a martingale with respect to the filtration $\mathcal{F}_{i+1}^k$ and $v_k + 1$ is a stopping time. By

the optional stopping theory, $E[A_{v_k+1}] = E[A_0] = 0$, i.e.,

$$
\begin{aligned}
E[A_{v_k+1}] &= E[\big(S_{k-1} - \lambda(p_k)(v_k + 1)h - X_{v_k+1}^k\big)^2] - E[\sum_{l=0}^{v_k} g^2(X_0 - X_l^k, p_k)h] \\
&= (S_{k-1})^2 + \lambda^2(p_k)E[((v_k + 1)h)^2] - 2S_{k-1}\lambda(p_k)E[(v_k + 1)h] - 2S_{k-1}E[X_{v_k+1}^k] \\
&\quad + 2\lambda(p_k)E[(v_k + 1)hX_{v_k+1}^k] + E[(X_{v_k+1}^k)^2] - E\big[\sum_{l=0}^{v_k} g^2(X_0 - X_{l-1}^k, p_k)\big]h = 0.
\end{aligned}
$$

(3.70)

Note that $X_{v_k+1}^k$ is no greater than S_k. Combining the inequalities (3.68) and (3.70) yields

$$
\begin{aligned}
E[((v_k + 1)h)^2] &= \frac{1}{\lambda^2(p_k)}\big[-S_{k-1}^2 + 2S_{k-1}\lambda(p_k)E[(v_k + 1)h] + 2S_{k-1}E[X_{v_k+1}^k] \\
&\quad - 2\lambda(p_k)E[(v_k + 1)hX_{v_k+1}^k] - E[(X_{v_k+1}^k)^2] + E[\sum_{l=0}^{v_k} g^2(X_0 - X_{l-1}^k, p_k)h]\big] \\
&\leq b_1,
\end{aligned}
$$

where $b_1 = \frac{1}{\lambda^2(p_k)}\big[S_{k-1}^2 + 2S_{k-1}\lambda(p_k)h + \overline{\sigma}^2\big(\frac{S_{k-1}-S_k+\sigma\sqrt{h}}{\underline{\lambda}} + h\big)\big]$. Therefore we have

$$
E[(\tau_k - \tau_{k-1})^2] \leq E[(v_k h + h)^2] \leq b_1.
$$

(3.71)

Combining (3.69) and (3.71), we conclude that there exists a positive constant $\mathcal{M}_2$, such

that $E[(\tau_N)^2] \leq \mathcal{M}_2.$∎

With Lemma 3.8, by using similar techniques as in Section 3.3, we have

Theorem 3.6 *If Assumptions 3.1 and 3.2 hold, we have*

$$\frac{\partial E[\Delta \hat{V}_k]}{\partial p_k} = E[\mathscr{C}_{p,I}^k \Delta \hat{V}_k] + E\big[\mathscr{D}_p^k \Delta \hat{V}_k\big], \qquad \text{for all } k \leq N,$$

$$\frac{\partial E[\Delta \hat{V}_k]}{\partial S} = E[\mathscr{C}_{S,I}^k \Delta \hat{V}_k] + E\big[\mathscr{D}_S \Delta \hat{V}_k\big], \qquad \text{for all } k \leq N,$$

where $\mathscr{D}_p^k \Delta \hat{V}_k$ and $\mathscr{D}_S \Delta \hat{V}_k$ are given by (3.35) and (3.64), respectively; $\mathscr{C}_{p,I} \Delta \hat{V}_k$ and $\mathscr{C}_{S,I} \Delta \hat{V}_k$ are given by $\mathscr{C}_{p,I}^k \Delta \hat{V}_k = \sum_{m=1}^{v_k} \mathscr{C}_{p,I,m}^k$, $\mathscr{C}_{S,I} \Delta \hat{V}k = \sum_{m=1}^{v_k} \mathscr{C}_{S,I,m}^k$, in which $\mathscr{C}_{p,I,m}^k$ and $\mathscr{C}_{S,I,m}^k$ are defined by

$$\mathscr{C}_{p,I,m}^k = \frac{\partial \eta_m^k}{\partial p_k} E\Big[c(X_{m-1}^k)(\tau_m^k - t_m^k) \prod_{n=0}^{m-1} I_{\{U_n^k < \eta_n^k\}} - \sum_{j=m}^{v_k} \Big(\sum_{i=m}^{j} c(X_{i-1})h$$
$$+ c(X_j)(\tau_{j+1}^k - t_j^k)\Big) \prod_{n=0, n \neq m}^{j} I_{\{U_n^k < \eta_n^k\}} I_{\{U_{n+1}^k \geq \eta_{n+1}^k\}} \Big],$$

$$\mathscr{C}_{S,I,m}^k = \frac{\partial \eta_m^k}{\partial S} E\Big[c(X_{m-1}^k)(\tau_m^k - t_m^k) \prod_{n=0}^{m-1} I_{\{U_n^k < \eta_n^k\}} - \sum_{j=m}^{v_k} \Big(\sum_{i=m}^{j} c(X_{i-1})h$$
$$+ c(X_j)(\tau_{j+1}^k - t_j^k)\Big) \prod_{n=0, n \neq m}^{j} I_{\{U_n^k < \eta_n^k\}} I_{\{U_{n+1}^k \geq \eta_{n+1}^k\}} \Big].$$

The gradient estimators of $\hat{V}$ with respect to p_k and S are given by $\mathscr{G}_p^k \hat{V} = \sum_{k=1}^{N} \big(\mathscr{S}_{p,I}^k \hat{V}_k + \mathscr{D}_p^k \hat{V}_k \big)$, and $\mathscr{G}_S^k \hat{V} = \sum_{k=1}^{N} \big(\mathscr{S}_{S,I} \hat{V}_k + \mathscr{D}_S^k \hat{V}_k \big)$, respectively. With these gradient estimators, we can use SA to find the optimal initial inventory level and the optimal pricing policy. Convergence results for the SA algorithms for the infinite horizon problem can be established similarly, and all the details are omitted.

3.6 Numerical Experiments

In this section, we test the gradient estimators of the expected discretized profit function and the proposed stochastic approximation algorithm 2 for revenue management

problems both over an infinite horizon and over a finite horizon with different parameter settings, using a class of pricing policy π_f, in which the price can only be changed a limited number of times.

3.6.1 Infinite Horizon Problem

Consider the revenue management problem (3.66) in Section 3.5 over an infinite horizon. In order to show the effectiveness of the proposed algorithm, we assume that the arrival rate of the demand is $\lambda(p) = 30 - p$, the uncertainty parameter σ in the demand is a constant, and the holding cost per unit time is $c(x) = \alpha x$, where α is a positive scalar. For this model, analytical solutions of the optimal pricing policy, and the sensitivities of the expected discretized profit with respect to various parameters can be obtained by using the method in [22].

3.6.1.1 Sensitivity Analysis

We first conduct a simulation experiment to compute sensitivities of the expected discretized profit with respect to various parameters, including the pricing parameters, the initial inventory level, the uncertainty parameter σ, and the holding cost parameter α. In this specific experiment, we assume that the price can be changed twice, i.e., $N = 3$. Let S be the initial inventory level. A price p_1 is charged until the inventory level falls to $\frac{2S}{3}$, then a price p_2 is charged until the inventory level falls to $\frac{S}{3}$, and finally p_3 is used until all the items are sold.

To compute sensitivities, we first generate sample paths by using Algorithm 3; along each sample path, sensitivities of the expected discretized profit function with respect to various parameters can be obtained. Let the time step in Algorithm 3 be $h = 0.1$, the

Table 3.1: Infinite Time Sensitivity Estimation Results (Standard Error in Parentheses)

	$\sigma = 1$	$\sigma = 3$	$\sigma = 5$	$\sigma = 7$
$\partial \hat{V}/\partial p_1$				
LR	5.91(2.55)	4.02(1.27)	4.30(0.78)	5.03(1.24)
PW	5.47(0.18)	5.40(0.32)	5.08(0.40)	3.88(0.50)
ANA	5.52	5.26	4.72	3.92
$\partial \hat{V}/\partial p_2$				
LR	18.85(2.00)	16.03(0.51)	15.44(0.38)	14.57(0.32)
PW	16.70(0.09)	16.47(0.17)	16.04(0.22)	15.43(0.27)
ANA	16.63	16.367	15.83	15.03
$\partial \hat{V}/\partial p_3$				
LR	27.75(0.01)	27.47(0.03)	26.97(0.06)	26.12(0.09)
PW	27.73(0.01)	27.47(0.03)	26.99(0.06)	26.36(0.09)
ANA	27.74	27.48	26.94	26.14
$\partial \hat{V}/\partial S$				
LR	10.28(0.32)	9.85(0.14)	9.88(0.08)	9.83(0.13)
PW	10.05(0.02)	10.03(0.04)	9.98(0.04)	9.86(0.05)
ANA	10.00	9.96	9.88	9.76
$\partial \hat{V}/\partial \alpha$				
LR	-505.2(0.3)	-508.2(0.9)	-515.3(1.5)	-531.1(2.2)
PW	-505.3 (0.3)	-509.9(0.9)	-516.0(1.5)	-530.8(2.1)
ANA	-500.5	-504.5	-512.5	-524.5
$\partial \hat{V}/\partial \sigma$				
LR	0.27(1.66)	-1.75(0.90)	-4.46(0.48)	-5.99(0.48)
PW	-1.51(0.31)	-3.28(0.34)	-4.642(0.39)	-7.38(0.44)
ANA	-1.00	-3.00	-5.00	-7.00

holding cost parameter $\alpha = 1$, and the initial inventory level $S = 100$. We let the prices $p_1 = p_2 = p_3 = 20$. In the simulation experiment, we generate 4000 independent sample paths. Along each sample path, we implement two estimators for each parameter denoted by LR and PW. LR represents the LR estimator when LR for the expected stopping times is used to derive gradient estimators for the expected discretized profit; PW represents the pathwise estimator when the sample path-based estimator for the expected stopping times is used to derive gradient estimators for the expected discretized profit.

Table 3.1 lists the sensitivities of the expected discretized profit with respect to various parameters for different uncertainty levels in the demand model by using different methods. The numbers in parentheses are standard errors. The proposed pathwise algorithm generally performs better than the LR method for most of the parameters at different uncertainty levels, especially for the prices p_1, p_2, and S, which are used in SA to find the optimal prices and the optimal initial inventory level. The simulation results of the proposed method are very close to the analytical ones with a comparatively small standard error.

3.6.1.2 Joint Optimization of Prices and Initial Inventory Level

With the above gradient estimation results, we use SA to jointly optimize prices and the initial inventory level. The optimal pricing policy can be obtained numerically using the method in [22]. We conduct two experiments with the number of price changes $N = 2$ and $N = 3$. In the demand model, we set $\sigma = 1$.

When $N = 2$, the price is allowed to change once when the inventory hits $S/2$, where S is the initial inventory level. To implement the SA algorithm, we first generate 20 sample paths by using Algorithm 3, and compute gradient estimates along these sample paths for parameters of interest. Then we use Algorithm 2 to update these parameters.

In Algorithm 2, we set $a = 0.10/k$ when updating prices, where k is the iteration number, and $a = 15/k$ when updating S. Figures 3.2(a) and 3.2(b) show the prices and the initial inventory levels, respectively, at different iterations when we iteratively update the prices and the inventory level by using the SA algorithm. In Figure 3.2(a), p_1 is the price that is charged before the inventory level falls to $S/2$, and p_2 is the price after the inventory level first falls bellow $S/2$; $p_1^\star$ and $p_2^\star$ are the corresponding optimal

prices. In Figure 3.2(b), S is the initial inventory level, and $S^\star$ is the corresponding optimal value. V in Figure 3.4(a) is the average profit, and $V^\star$ is the expected discretized profit corresponding to the optimal policy. From the simulation results, we can see our simulation algorithm converges to the optimal values quickly.
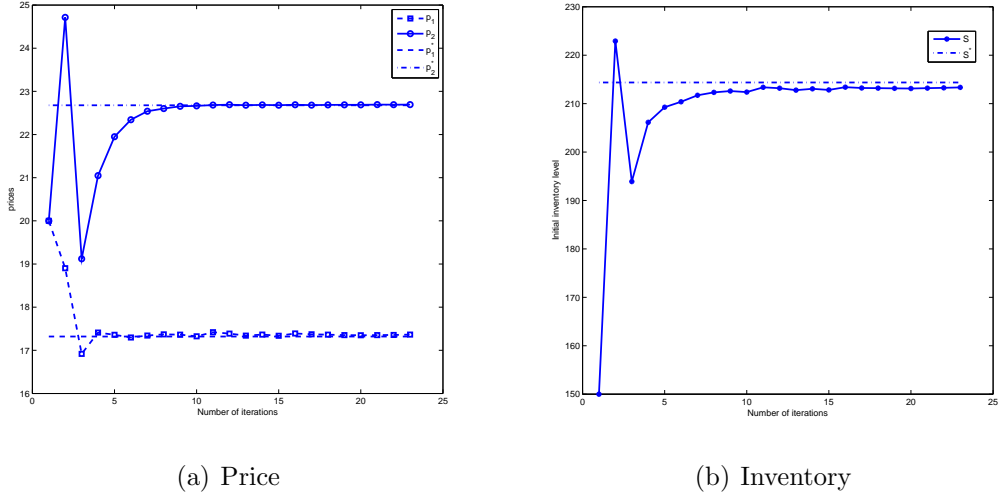

(a) Price

(b) Inventory

Figure 3.2: Prices and Inventory when $N = 2$
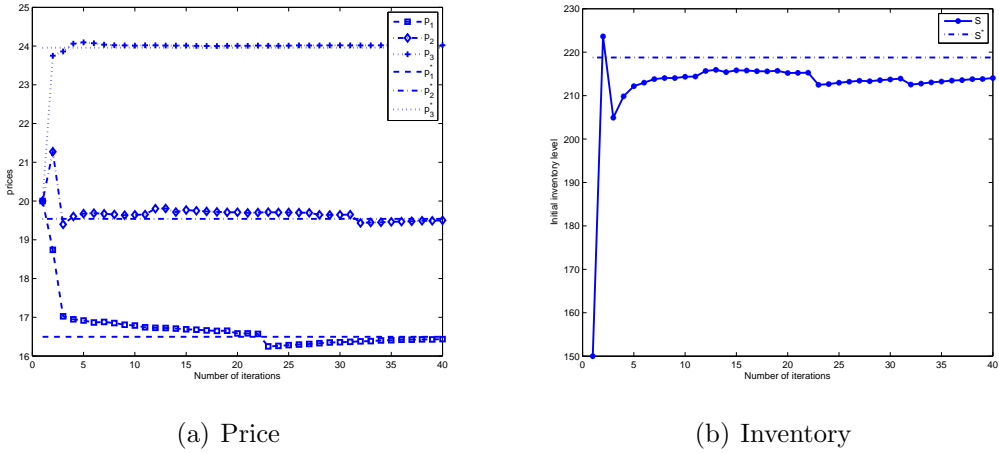

(a) Price

(b) Inventory

Figure 3.3: Prices and Inventory when $N = 3$

Figures 3.3(a), 3.3(b), and 3.4(b) show the convergence of the prices, the initial inventory level, and the average profit of the SA algorithm, respectively, when the price is allowed to change twice, i.e., $N = 3$. Again, we can see that the proposed algorithm

86

works well. When $N = 3$, Figure 3.5(a) shows how the optimal pricing policies change when we change the initial inventory level. The optimal prices decrease when the initial inventory level increases, since with the increase of the initial inventory level, lowering the prices will increase the demand, and hence reduce the holding cost. Figure 3.5(b) shows the optimal average profits for different initial inventory levels. We can see that at the beginning, an increase in the initial inventory level brings an increase in the profit.
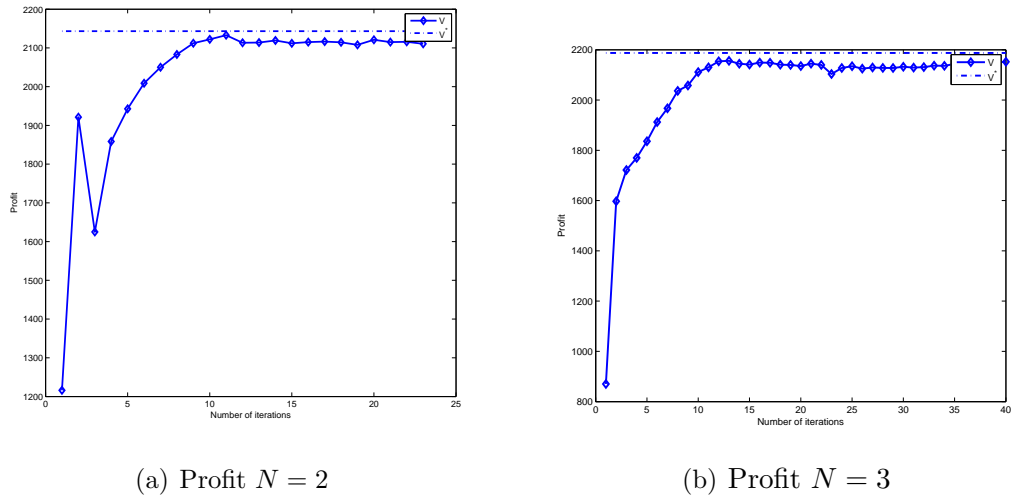

(a) Profit $N = 2$ (b) Profit $N = 3$

Figure 3.4: Profit

However up to a certain level, an increase in the initial inventory level will reduce the profit since the revenue increase due to the increase in the initial inventory level is less than the holding cost increase. This figure shows the benefit of optimizing the initial inventory level. The average profit improves significantly if we jointly optimize the prices and the initial inventory level. Figure 3.6(a) shows the simulation results of dynamic pricing with a fixed initial inventory level $S = 150$, when we vary the number of price changes allowed in the pricing policy. We can see that when the price changes are more than 4, the increase in the profit is not remarkable any more. In other words, we can say that pricing policies that allow the price to change continuously only provide limited

87

advantage over the pricing policies that allow a limited number of price changes.
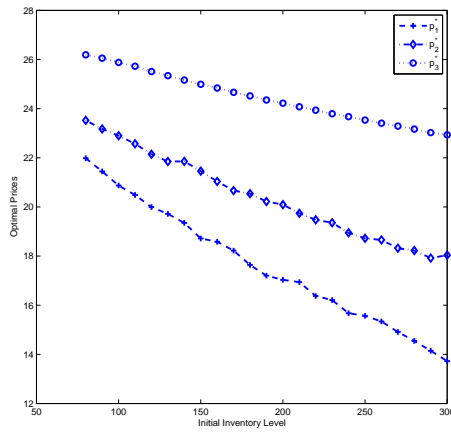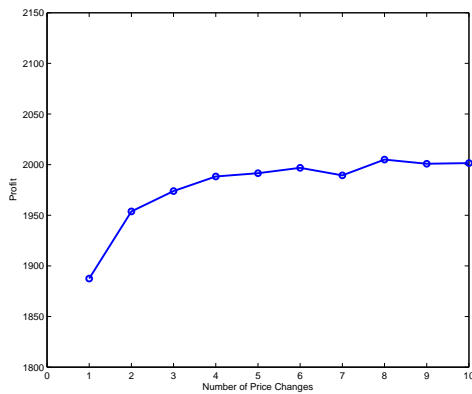

(a) Pricing Policy Changes
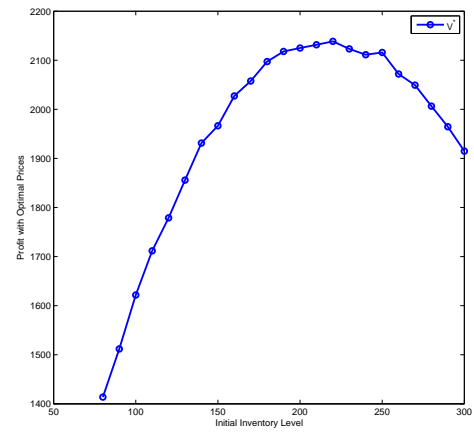
(b) Profit Changes

Figure 3.5: Pricing Policy and Profit Changes

(a) Effect of Number of Price Changes
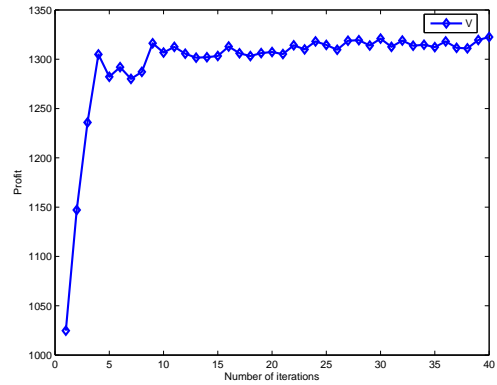
(b) Profit with $N = 2$ over Finite Time

Figure 3.6: Effect of Price Changes and Profit

3.6.2 Finite Horizon Problem

Now we consider the revenue management problem (3.12) over a finite horizon; we use the same demand model as the one over an infinite time horizon. We assume that the arrival rate of the demand is $\lambda(p) = 30 - p$, σ is a constant, and the holding cost per

unit time is $c(x) = \alpha x$. We adopt the pricing policy in which only a limited number of price changes can be made during the selling process. The selling process stops at $T = 8$. There is no analytical form for the sensitivities of the expected discretized profit function with respect to parameters of interest, such as the pricing parameter $p_i, i = 1, \cdots, N$, and the initial inventory level S. There is also no analytical solution for the optimal pricing policy or the optimal initial inventory level.

We assume that the price can only be changed once, i.e., $N = 2$. We generate sample paths using Algorithm 1; along each sample path, sensitivities of the expected discretized profit function with respect to various parameters can be obtained. The time step in Algorithm 1 is $h = 0.1$; the holding cost parameter $\alpha = 1$; the initial inventory level is $S = 100$. We let $p_1 = 20$ and $p_2 = 20$. In the simulation experiment, we generate 4000 independent sample paths. Along each sample path, we implement two estimators denoted by LR and PW. LR represents the LR estimator when LR for the expected stopping times is used to derive gradient estimators for the expected discretized profit; PW is the pathwise estimator when the sample path-based estimator for the expected stopping times is used to derive gradient estimators for the expected discretized profit.

Table 3.2 lists the gradient estimates of the expected discretized profit function with respect to the prices p_1, p_2, and the initial inventory level S. The pathwise estimators outperform the LR estimators with a much smaller standard error. The pathwise estimators also have relatively constant performances for different values of σ. Again we only need one simulation run to obtain the gradients of the expected discretized profit function with respect to all parameters of interest by using the proposed method.

Table 3.2: Finite Time Sensitivity Estimation Results (Standard Error in Parentheses)

	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$
$\partial \hat{V}/\partial p_1$				
LR	-59.3(17)	-51.3(12.7)	-59.9(11.1)	-63.1(9.1)
PW	-63.8(1.4)	-62.2(1.3)	-63.8(1.4)	-64.2(1.4)
$\partial \hat{V}/\partial p_2$				
LR	-34.27(0.05)	-34.21(0.08)	-34.3(0.1)	-34.4(0.1)
PW	-34.28(0.05)	-34.40(0.08)	-34.2(0.1)	-34.2(0.1)
$\partial \hat{V}/\partial S$				
LR	-6.0(1.7)	-5.0(1.3)	-5.9(1.1)	-5.6(0.9)
PW	-6.5(0.2)	-6.0(0.2)	-5.9(0.1)	-5.8(0.1)

3.6.2.1   Joint Optimization of Prices and Initial Inventory Level

We assume that the price can only be changed a limited number of times in the pricing policies. With the above gradient estimation results, we use SA to jointly optimize prices and the initial inventory level. In the SA algorithm, we set $a = 0.10/k$ when computing p_1 and p_2, where k is the iteration number, and $a = 15/k$ when computing S. We present simulation results for the case $N = 2$; the price is allowed to change one
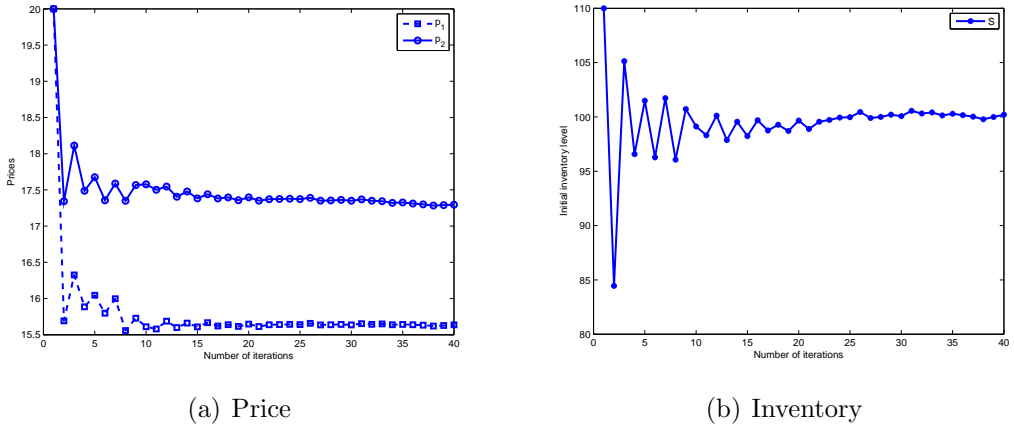

(a) Price

(b) Inventory

Figure 3.7: Prices and Inventory with $N = 2$ over Finite Time

time when the inventory hits $S/2$, where S is the initial inventory level. Figures 3.7(a), 3.7(b), and 3.6(b) show that the prices, the initial inventory level, and the average profit

converge quickly to the optimal values. Compared with the results of the corresponding infinite horizon revenue management problem in Figures 3.2(a), 3.2(b), and 3.4(a), we can see that the optimal initial inventory level changes sharply if we only allow selling the inventory over a finite horizon. The corresponding pricing policy and the final profit also change.

3.7   Conclusions

We considered dynamic pricing problems for inventory systems with price-sensitive demand, which follows a continuous-time, continuous-state stochastic process instead of the commonly used discrete-time stochastic process. We formulated the dynamic pricing problem as a stochastic control problem, and gave an analytical solution for a special demand model.

When only a finite number of price changes is allowed in the pricing policy, we proposed a simulation-based method for solving the pricing problem under a broad range of demand models. We gave a new simulation scheme to simulate the evolution of the inventory level. Based on the generated sample paths, we derived gradient estimators of the expected discretized profit function with respect to various parameters. Specifically, we gave a pathwise gradient estimator for stopping times by using a Brownian bridge. When we derived the gradient estimators for the expected discretized profit function, we circumvented the difficulty of differentiating a performance function with discontinuous sample paths by using smoothed perturbation analysis. We showed the unbiasedness of the resulting estimators. We also showed the convergence of the SA algorithm with the proposed gradient estimators. Simulation examples demonstrated that the proposed algorithm works well.

Chapter 4

A New Gradient Estimator for Discontinuous Payoff Functions

4.1 Introduction and Motivation

In this chapter, we consider performance functions of the following form:

$$L(X(\omega;\theta_1), Z;\theta_2)1\{X(\omega;\theta_1) > \theta_2\}, \tag{4.1}$$

which is a function of a real-valued vector of parameters $\theta_1 \in \Theta_1 \subset \mathbb{R}^d$, a real-valued parameter $\theta_2 \in \Theta_2 \subset \mathbb{R}$, a random variable $X(\omega;\theta_1)$, and a vector of other random variables Z. ω is the randomness in the random variable X and $1\{\cdot\}$ is the indicator function. For example, for a European option, θ_1 could be the initial stock price and X the stock price at time T, where T is the time when the European option expires. We are interested in obtaining an efficient gradient estimator for $\frac{\partial E[]L(X(\omega;\theta_1),Z;\theta_2)1\{X(\omega;\theta_1)>\theta_2\}}{\partial\theta}$, where θ could be an element of θ_1 or θ_2.

Payoff functions of the form (4.1) are found in many financial applications. For example, the European call option payoff has the form (4.1), but it still happens to be continuous, so IPA can be applied. However, the first derivative of the payoff will be discontinuous due to the indicator function, so that IPA would not be applicable, e.g., for estimating the *gamma*. The American call option pricing problem considered in [36] and [51] also has payoff functions of a similar form with indicator functions that make the payoff functions discontinuous; hence IPA does not work. By using conditional expectation,

smoothed perturbation analysis (SPA) ([49, 37]) can sometimes overcome the difficulty of differentiating an indicator function, but SPA requires choosing what to condition on, and an SPA estimator often requires more computation because of the estimation of conditional expectations. Although LR doesn't require the payoff function to be continuous, it is not applicable directly when the support of the random variable $X(\omega; \theta_1)$ involves the parameters θ_1 and/or θ_2.

4.1.1 Related Literature

In the last three decades, derivative estimation has been studied extensively in the simulation literature. Infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method are two of the main techniques ([53, 90]). Other techniques include smoothed perturbation analysis (SPA), which can be applied to performance functions containing discontinuities, and methods based on weak derivatives (WD) and Malliavin calculus ([29, 23]), and finite-difference-based and kernel estimation-based methods ([57, 56, 74]).

Introduced by [52], IPA has been widely used in sensitivity analysis for discrete-event systems (see applications of IPA in queueing systems and inventory control problems in ([100, 43, 53, 35, 46, 72, 3])), as well as for financial derivatives ([44, 17, 36]). IPA enables the sensitivity of a performance function to be estimated while observing a *single* sample path of a system and hence offers significant computational savings compared to "brute force" finite difference methods. However, IPA generally requires the performance function to be continuous with respect to parameters of interest, which makes IPA not applicable in many cases. For example, second-order derivatives cannot be obtained by IPA for European call options ([17]).

Rather than differentiating a performance function as IPA does, the likelihood ratio

(LR) method constructs derivative estimators from the derivatives of the probability measure associated with a simulation model. The method was proposed by [48], [87], and [91] to study discrete event systems and has also been used in financial applications ([17, 44]). It is also called the score function (SF) method ([90]). For further information on the LR method, see [91, 90, 32]. The LR method does not require continuity of performance functions and hence is more widely applicable than IPA. However, parameters of interest have to be in probability density functions in order to apply LR, whereas in many cases they appear naturally in performance functions. The push-out method may be able to move parameters of interest out of the performance function and push them into a probability density function ([90]); then the LR method can be applied. However the push-out method has only been demonstrated on some simple cases ([90]).

4.1.2   Background of Gradient Estimation

Let $H(X(\omega; \theta_1); \theta_2)$ be a performance function of a stochastic system, where $X(\omega; \theta_1)$ is a random variable defined on the probability space $(\Omega_\omega, \mathcal{F}_\omega, \mathbb{P}_\omega)$, and θ_1 and θ_2 are defined the same way as in (4.1). For example, in a queueing system, θ_1 could be the mean arrival time and X the interarrival time. Let $J(\theta_1, \theta_2) = E_{\mathbb{P}_\omega}[H(X(\omega; \theta_1); \theta_2)]$, where $E_{\mathbb{P}_\omega}[\cdot]$ denotes expectation with respect to the probability measure $\mathbb{P}_\omega$. $J(\theta_1, \theta_2)$ can be written as

$$J(\theta_1, \theta_2) = \int_\Omega H(X(\omega; \theta_1); \theta_2) d\mathbb{P}_\omega(\omega). \tag{4.2}$$

Random variables can be generated from $U(0, 1)$ random numbers by transforming them in an appropriate way. Hence $J(\theta_1, \theta_2)$ can always be written in the form of (4.2). Suppose

we are interested in the sensitivities of $J(\theta_1, \theta_2)$ with respect to the parameters θ_1 and θ_2. Under appropriate conditions, we have

$$\nabla_{\theta_1} J(\theta_1, \theta_2) = E_{\mathbb{P}_\omega}[\nabla_{\theta_1} H(X(\omega; \theta_1); \theta_2)],$$

$$\frac{\partial J(\theta_1, \theta_2)}{\partial \theta_2} = E_{\mathbb{P}_\omega}\left[\frac{\partial H(X(\omega; \theta_1); \theta_2)}{\partial \theta_2}\right], \qquad (4.3)$$

where $\nabla_{\theta_1} = (\frac{\partial}{\partial \theta_{11}}, \cdots, \frac{\partial}{\partial \theta_{1d}})$. Note that

$$\frac{\partial H(X(\omega; \theta_1); \theta_2)}{\partial \theta_{1i}} = \frac{\partial H(X(\omega; \theta_1); \theta_2)}{\partial X(\omega; \theta_1)} \frac{\partial X(\omega; \theta_1)}{\partial \theta_{1i}},$$

where $\frac{\partial X(\omega; \theta_1)}{\partial \theta_{1i}}$ is the derivative of a random variable ([100, 32]). $\nabla_{\theta_1} H(X(\omega; \theta_1); \theta_2)$ and $\frac{\partial H(X(\omega; \theta_1); \theta_2)}{\partial \theta_2}$ are the IPA derivative estimators. Since ω does not depend on θ_1 and θ_2, we only need to generate one single sample path ω to get an estimate of the sensitivity for each parameter of interest. IPA requires the interchange of orders for integration and differentiation, which usually requires the performance function $H(X(\omega; \theta_1); \theta_2)$ to be continuous.

In contrast to IPA, LR differentiates probability measures instead of the performance functions to construct derivative estimators and hence does not require continuity of performance functions along sample paths. To derive an LR derivative estimator for $H(X(\omega; \theta_1); \theta_2)$ with respect to θ_1, we assume that $\mathbb{P}_X$ is the probability measure induced by the random variable $X(\omega; \theta_1)$, i.e., $\mathbb{P}_X(B; \theta_1) = \mathbb{P}_\omega\{\omega : X(\omega; \theta_1) \in B\}$ for any $B \subset \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is a Borel set. By the change of variable formula ([96]), we have

$$\int_\Omega H(X(\omega; \theta_1); \theta_2) d\mathbb{P}_\omega(\omega) = \int H(x; \theta_2) d\mathbb{P}_X(x; \theta_1),$$

Differentiating both sides of the above equation gives

$$\nabla_{\theta_1} J(\theta_1, \theta_2) = E_{\hat{\mathbb{Q}}}\Big[H(X; \theta_2)\Big(\nabla_{\theta_1} \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big)\frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big]$$

$$= E_{\mathbb{Q}}\Big[H(X; \theta_2)\Big(\nabla_{\theta_1} \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big)\frac{d\mathbb{P}_X}{d\mathbb{Q}}(X)\Big], \tag{4.4}$$

where $\hat{\mathbb{Q}}$ is the Lebesgue measure if X is a continuous random variable and is the counting measure if X is a discrete random variable (i.e., $d\mathbb{P}_X/d\hat{\mathbb{Q}}$ corresponds to the probability density and mass functions, respectively); $\mathbb{Q}$ is a probability measure such that $\mathbb{P}_X$ is absolutely continuous with respect to $\mathbb{Q}$.

$$H(X; \theta_2)\Big(\nabla_{\theta_1} \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)\Big)\frac{d\mathbb{P}_X}{d\mathbb{Q}}(X)$$

is called the LR derivative estimator. The LR method does not require the performance function to be continuous, and the interchangeability of differentiation and integration is often not a problem ([44]). However, first, it is not always possible to move parameters of interest into probability density functions; second, note that the support of the distribution of X should not depend on θ_1 when deriving (4.3) and (4.4). For performance functions in the form of (4.1), LR is not applicable directly if the support of the distribution of X depends on the parameter of interest. Compared with IPA, LR estimators tend to have larger variance, especially when the input process involves an oft-repeated (e.g. i.i.d) random variable whose common distribution depends on the parameter of interest.

Although IPA is constructed by differentiating the performance function, whereas LR involves differentiation of the underlying probability measure, it is possible to unify them in a single framework as in [72].

4.2 SLRIPA Derivative Estimator

The expectation of the performance function (4.1) is denoted as

$$\tilde{J}(\theta_1, \theta_2) = E_{\mathbb{P}_\omega}[L(X(\omega; \theta_1), Z; \theta_2)1\{X(\omega; \theta_1) > \theta_2\}], \tag{4.5}$$

where Z is defined on a probability space $(\Omega_Z, \mathcal{F}_Z, \mathbb{P}_Z)$. Let $(\Omega, \mathcal{F})$ be the product space

of $(\Omega_\omega, \mathcal{F}_\omega)$ and $(\Omega_Z, \mathcal{F}_Z)$. Here Ω_Z is a sample space and $\mathcal{F}_Z$ is a σ-algebra defined

on Ω_Z; X and Z are independent and are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$,

where $d\mathbb{P}(\omega, z) = d\mathbb{P}_\omega(\omega) \times d\mathbb{P}_Z(z)$. Here the expectation is actually computed with

respect to the measure $\mathbb{P}$, but to emphasize the effect of $\mathbb{P}_\omega$, we use the measure $\mathbb{P}_\omega$

as a subscript of the expectation. We assume that the random variable $X(\omega; \theta_1)$ has

a probability density or mass function $f(x; \theta_1)$ with support $[a(\theta_1), b(\theta_1)]$, where $\theta_1 =$

$(\theta_{11}, \cdots, \theta_{1d})^T$ is a real-valued vector of continuous parameters, and $a(\theta_1)$ and $b(\theta_1)$ can

be $-\infty$ and $+\infty$, respectively. To simplify the notation, we sometimes use X to denote

$X(\omega; \theta_1)$. Note that if L equals 1, then $\tilde{J}(\theta_1, \theta_2)$ becomes the probability that X is greater

than θ_2. We are interested in estimating $\frac{\partial \tilde{J}(\theta_1, \theta_2)}{\partial \theta}$, where θ could be θ_{1i} for $i = 1, 2, \cdots, d$,

or θ_2. Since there is an indicator function in $\tilde{J}(\theta_1, \theta_2)$, the direct IPA estimator is biased.

4.2.1 Derivation of SLRIPA

Our method consists of two steps. First we move the parameter θ_2 out of the

indicator function through an appropriate change of random variables; the critical feature

of the change is that the support of the distribution of the new random variable no longer

depends on the parameter of interest. The resulting form allows us to obtain the unbiased

SLRIPA estimator by simultaneously applying LR and IPA.

Assume that we can define a random variable $Y = \Gamma(X; \theta_1, \theta_2)$, where Γ is a function of X with parameters θ_1 and θ_2, such that Y satisfies two conditions. First, the support of the random variable Y does not depend on θ_1 and θ_2, and Γ is invertible as a function of X. The specific form of Γ will be given for several different cases, in which the support of the distribution of X is in different forms, in the following subsections. Second, there exists a function $\psi(Y)$, such that ψ does not depend on θ_1 and θ_2 explicitly, and $X(\omega; \theta_1) > \theta_2$ is equivalent to $\psi(Y) > 0$. Let $\mathbb{P}_Y$ be the probability measure induced by the random variable Y. We have

$$
\begin{aligned}
\frac{\partial \tilde{J}(\theta_1, \theta_2)}{\partial \theta} =& \frac{\partial}{\partial \theta} E_{\mathbb{P}_\omega}[L(X(\omega; \theta_1), Z; \theta_2)1\{X(\omega; \theta_1) > \theta_2\}] \\
=& \frac{\partial}{\partial \theta} E_{\mathbb{P}_Y}[L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)1\{\psi(Y) > 0\}] \\
=& \frac{\partial}{\partial \theta} E_{\hat{\mathbb{Q}}}\Big[L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)1\{\psi(Y) > 0\}\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y)\Big], \qquad (4.6)
\end{aligned}
$$

where $\hat{\mathbb{Q}}$ is the Lebesgue measure if Y is a continuous random variable and is the counting measure if Y is a discrete random variable. Similar as in (4.5), we use $\mathbb{P}_Y$ as the subscripts of the expectation operator to emphasize the effect of $\mathbb{P}_Y$ when computing the above expectation.

Note that (4.6) has two terms related to the parameter θ_1 and/or the parameter θ_2: $L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)$ and the Radon-Nikodym derivative $\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y)$. The indicator function $1\{X(\omega; \theta_1) > \theta_2\}$ becomes $1\{\psi(Y) > 0\}$, which no longer has any dependence on the parameters. Hence, given Y, a perturbation in the parameter θ will affect the function $L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)$ and the Radon-Nikodym derivative $\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y)$, but not the indicator function $1\{\psi(Y) > 0\}$. In other words, the effect of the parameter θ has been moved out of the indicator function and is transferred into the Radon-Nikodym derivative.

Thus, we can apply IPA to $L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)$ and use LR for $\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)$. Let $\tilde{\mathbb{Q}}$ be a probability measure, such that $\mathbb{P}_Y$ is absolutely continuous with respect to $\tilde{\mathbb{Q}}$. Then assuming differentiation and expectation can be interchanged in (4.6), we have

$$
\begin{aligned}
\frac{\partial \tilde{J}(\theta_1, \theta_2)}{\partial \theta} =& E_{\tilde{\mathbb{Q}}}\Big[\frac{\partial L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)}{\partial \theta} 1\{\psi(Y) > 0\}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\Big] \\
& + E_{\tilde{\mathbb{Q}}}\Big[L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)1\{\psi(Y) > 0\}\frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\Big]. \quad (4.7)
\end{aligned}
$$

Mild regularity conditions to be given later, allow the interchange of expectation and differentiation in (4.6), whereas the interchange of expectation and differentiation would not be valid in directly differentiating (4.1). Here we call

$$
\begin{aligned}
& \frac{\partial L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)}{\partial \theta} 1\{\psi(Y) > 0\}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) \\
& + L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)1\{\psi(Y) > 0\}\frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) \quad (4.8)
\end{aligned}
$$

the SLRIPA derivative estimator. In particular, if $\tilde{\mathbb{Q}}$ is the probability measure induced by Y, then the SLRIPA derivative estimator is

$$
\begin{aligned}
& \frac{\partial L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)}{\partial \theta} 1\{\psi(Y) > 0\} \\
& + L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)1\{\psi(Y) > 0\}\frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta}. \quad (4.9)
\end{aligned}
$$

Remark 4.1 *It is critical to choose the functions Γ and ψ such that the support of Y and the indicator function do not depend on the parameters of interest. For payoff functions of the form (4.5), we give specific SLRIPA estimators for several different cases later.*

Remark 4.2 *Other than writing the SLRIPA estimator in terms of the newly defined*

random variable Y, it might be desirable to represent the SLRIPA derivative estimator in terms of the original random variable X, when X has specific physical meanings. One way to accomplish this is letting $\tilde{\mathbb{Q}}$ be the probability measure induced by X, and then we can write the SLRIPA estimator in terms of the original random variable X as follows:

$$
\frac{\partial L(\Gamma^{-1}(Y; \theta_1, \theta_2), Z; \theta_2)}{\partial \theta}\bigg|_{Y=\Gamma(X;\theta_1,\theta_2)} 1\{X > \theta_2\}
$$
$$
+ L(X, Z; \theta_2) 1\{X > \theta_2\} \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta}\bigg|_{Y=\Gamma(X;\theta_1,\theta_2)}. \tag{4.10}
$$

More generally, the choice of $\tilde{\mathbb{Q}}$ might be based on the ease and/or efficiency in carrying out the simulations depending on the specific problem. One natural approach is to design $\tilde{\mathbb{Q}}$ to reduce the variance of the SLRIPA derivative estimator by viewing $\tilde{\mathbb{Q}}$ as a probability measure for importance sampling; formulating and solving this problem is an interesting topic for future research.

4.2.1.1 $\quad b(\theta_1) = +\infty$ and $a(\theta_1)$ is finite

We assume that $\theta_2 > a(\theta_1)$; otherwise, the indicator function $1\{X > \theta_2\}$ would always be 1, and this degenerate case will be considered later. Define a new random variable by $Y = \frac{X-a(\theta_1)}{\theta_2-a(\theta_1)}$, and note the distribution of the new random variable Y has support $[0, +\infty]$, which does not depend on the parameters θ_1 and θ_2. Furthermore $X - \theta_2 > 0 \iff Y - 1 > 0$. Hence we can define $\Gamma(X; \theta_1, \theta_2) = \frac{X-a(\theta_1)}{\theta_2-a(\theta_1)}$ and $\psi(Y) = Y - 1$.

Then $X = \Gamma^{-1}(Y; \theta_1, \theta_2) = (\theta_2 - a(\theta_1))Y + a(\theta_1)$. The SLRIPA estimator (4.8) becomes

$$\frac{\partial L((\theta_2 - a(\theta_1))Y + a(\theta_1), Z; \theta_2)}{\partial \theta} 1\{\psi(Y) > 0\} \frac{d\mathbb{P}_Y}{d\widetilde{\mathbb{Q}}}(Y)$$

$$+ L((\theta_2 - a(\theta_1))Y + a(\theta_1), Z; \theta_2) 1\{\psi(Y) > 0\} \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\widetilde{\mathbb{Q}}}(Y)}{\partial \theta} \frac{d\mathbb{P}_Y}{d\widetilde{\mathbb{Q}}}(Y). \qquad (4.11)$$

4.2.1.2 $b(\theta_1)$ is finite and $a(\theta_1)$ is finite.

We do a similar analysis as in the above case and define a random variable:

$$Y = \Gamma(X; \theta_1, \theta_2) = \frac{(X - a(\theta_1))(b(\theta_1) - \theta_2)}{(b(\theta_1) - X)(\theta_2 - a(\theta_1))}.$$

Again the distribution of the new random variable Y has support $[0, \infty)$, which does not depend on the parameters θ_1 and θ_2. We also assume that $a(\theta_1) < \theta_2 < b(\theta_1)$; otherwise, the indicator function is always the constant 1 or 0, and this degenerate case will be considered later. Notice the relation $X - \theta_2 > 0 \iff \psi(Y) = Y - 1 > 0$. The SLRIPA estimator (4.8) is given by

$$\frac{\partial}{\partial \theta} L\Big(\frac{b(\theta_1)(\theta_2 - a(\theta_2))Y + a(\theta_1)(b(\theta_1) - \theta_2)}{(\theta_2 - a(\theta_1))Y + (b(\theta_1) - \theta_2)}, Z; \theta_2\Big) 1\{\psi(Y) > 0\} \frac{d\mathbb{P}_Y}{d\widetilde{\mathbb{Q}}}(Y)$$

$$+ L\Big(\frac{b(\theta_1)(\theta_2 - a(\theta_2))Y + a(\theta_1)(b(\theta_1) - \theta_2)}{(\theta_2 - a(\theta_1))Y + (b(\theta_1) - \theta_2)}, Z; \theta_2\Big) 1\{\psi(Y) > 0\} \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\widetilde{\mathbb{Q}}}(Y)}{\partial \theta} \frac{d\mathbb{P}_Y}{d\widetilde{\mathbb{Q}}}(Y).$$

$$(4.12)$$

4.2.1.3   Other cases

The case where $a(\theta_1) = -\infty$ and $b(\theta_1)$ is finite can be done similarly as in the case when $a(\theta_1)$ is finite and $b(\theta_1) = +\infty$. When $a(\theta_1) = -\infty$ and $b(\theta_1) = +\infty$, we can define the new random variable as X/θ_2 or $X - \theta_2$. This case is trivial and we omit the details.

When the function L does not depend on θ_2 after the change of random variables, the SLRIPA estimator for the sensitivity with respect to the parameter θ_2 will be the same as the estimator obtained by the push-out method ([90]). In this case, the push-out method can be viewed as a special case of the SLRIPA method.

Remark 4.3 *For payoff functions of more general forms, for example, in which X is the sum of random variables $X = \sum_{i=1}^{m} X_i$ or the product of random variables $X = \prod_{k=1}^{m} X_i$, SLRIPA can still be applied. In simple cases in which the support of X_i does not depend on any parameter, we can define $Y_i = X_i/\theta_2$ for the sum case and $Y_1 = X_1/\theta_2$ for the product case. For problems in which the support of X_i depends on parameters of interest, we can define Y_i according to the two different cases in Sections 4.2.1.1 and 4.2.1.2. Practically, for many financial applications, defining $Y = X/\theta_2$ will work.*

4.2.2   LR and IPA as Special Cases of SLRIPA

From the description in Section 4.1.2, we can see that LR and IPA are closely related to each other and they are connected by a change of random variable; this relationship has been well known in the literature ([72]). We show how IPA and LR can be viewed as special cases of SLRIPA. Here we assume that the indicator function in the performance function (4.1) is always 1, and we are interested in estimating $\frac{\partial \hat{J}(\theta_1, \theta_2)}{\partial \theta}$, where

$$\hat{J}(\theta_1, \theta_2) = E[L(X(\omega; \theta_1), Z; \theta_2)]. \tag{4.13}$$

First we derive the SLRIPA derivative estimators for $\frac{\partial E[\hat{J}(\theta_1, \theta_2)]}{\partial \theta}$ before we show that LR and IPA are special cases of SLRIPA.

Case 1. $b(\theta_1) = +\infty$ and $a(\theta_1)$ is finite

Since there is no indicator function, we only need to define a new random variable whose support does not depend on either parameter. Define $Y = X - a(\theta_1)$. Along the same line as the analysis in Section 4.2, the SLRIPA estimator is

$$\frac{\partial L(Y + a(\theta_1), Z; \theta_2)}{\partial \theta} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) + L(Y + a(\theta_1), Z; \theta_2) \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y). \qquad (4.14)$$

Case 2. $b(\theta_1)$ is finite and $a(\theta_1)$ is finite

Now we define $Y = \frac{X - a(\theta_1)}{b(\theta_1) - X}$. Similarly, the SLRIPA estimator is

$$\frac{\partial}{\partial \theta} L\Big(\frac{b(\theta_1)Y + a(\theta_1)}{Y + 1}, Z; \theta_2\Big) \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) + L\Big(\frac{b(\theta_1)Y + a(\theta_1)}{Y + 1}, Z; \theta_2\Big) \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y).$$

$$(4.15)$$

The case that $a(\theta_1) = -\infty$ and $b(\theta_1)$ is finite can be done similarly as the case when $a(\theta_1)$ is finite and $b(\theta_1) = +\infty$. For the case that $a = -\infty$ and $b = +\infty$, the LR or IPA method can be applied directly.

4.2.2.1   LR and IPA as Special Cases of SLRIPA

Now we show that LR and IPA are just special cases of SLRIPA when a and b do not depend on θ_1.

First we consider the estimator in (4.14). Note that $P(Y \leq y) = P(X - a \leq y)$. Hence we have $d\mathbb{P}_Y(Y) = d\mathbb{P}_X(Y + a)$. We can always choose a probability measure $\mathbb{Q}$ such that $d\tilde{\mathbb{Q}}(Y) = d\mathbb{Q}(Y + a)$. For example, if there is a probability density function g_Y such that $d\tilde{\mathbb{Q}}(Y) = g_Y(Y)dY$, then $d\mathbb{Q}(Y)$ can be defined as $d\mathbb{Q}(Y) = g_Y(Y - a)dY$.

Therefore we have $\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) = \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(Y+a)$. If $\theta = \theta_{1i}$ for $i = 1, 2, \cdots, d$, we have

$$
\begin{aligned}
\frac{\partial \hat{J}(\theta_1, \theta_2)}{\partial \theta_{1i}} &= E_{\tilde{\mathbb{Q}}}\Big[L(Y+a, Z; \theta_2) \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y)}{\partial \theta_{1i}} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) \Big] \\
&= E_{\mathbb{Q}}\Big[L(Y+a, Z; \theta_2) \frac{\partial \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(Y+a)}{\partial \theta_{1i}} \frac{d\mathbb{P}_X}{d\mathbb{Q}}(Y+a) \Big] \\
&= E_{\mathbb{Q}}\Big[L(X, Z; \theta_2) \frac{\partial \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)}{\partial \theta_{1i}} \frac{d\mathbb{P}_X}{d\mathbb{Q}}(X) \Big].
\end{aligned}
$$

Note that $L(X, Z; \theta_2) \frac{\partial \ln \frac{d\mathbb{P}_X}{d\hat{\mathbb{Q}}}(X)}{\partial \theta_{1i}} \frac{d\mathbb{P}_X}{d\mathbb{Q}}(X)$ is an LR estimator. Hence we show that the LR estimator is equivalent to an SLRIPA estimator.

If $\theta = \theta_2$, we have

$$
\begin{aligned}
\frac{\partial \hat{J}(\theta_1, \theta_2)}{\partial \theta_2} &= E_{\tilde{\mathbb{Q}}}\Big[\frac{\partial L(Y+a, Z; \theta_2)}{\partial \theta_2} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) \Big] \\
&= E_{\mathbb{Q}}\Big[\frac{\partial L(Y+a, Z; \theta_2)}{\partial \theta_2} \frac{d\mathbb{P}_X}{d\mathbb{Q}}(Y+a) \Big] \\
&= E_{\mathbb{Q}}\Big[\frac{\partial L(X, Z; \theta_2)}{\partial \theta_2} \frac{d\mathbb{P}_X}{d\mathbb{Q}}(X) \Big].
\end{aligned}
$$

Note that $\frac{\partial L(X, Z; \theta_2)}{\partial \theta_2} \frac{d\mathbb{P}_X}{d\mathbb{Q}}(X)$ is an IPA estimator, which is equivalent to the SLRIPA estimator we obtained. Similarly, we can also show (4.15) is equivalent to an LR estimator or an IPA estimator. From the above analysis, we can see the SLRIPA technique includes LR and IPA as special cases and can be applied to more general classes of functions.

4.2.3   Comparison between SLRIPA and SPA Derivative Estimators

SPA is another technique that can deal with performance functions with discontinuities in parameters of interest. It is critical to choose what to condition on in order to get a good SPA derivative estimator. Under appropriate conditions, an SPA estimator

can be obtained as in [105]:

$$\frac{\partial \tilde{J}(\theta_1, \theta_2)}{\partial \theta} = E_{\mathbb{P}_X}\left[L(X, Z; \theta_2)\frac{\partial(-X + b(\theta_1))}{\partial \theta}\Big| X = b(\theta_1)\right]f(b(\theta_1); \theta_1)$$

$$+ E_{\mathbb{P}_X}\left[L(X, Z; \theta_2)\frac{\partial(X - \theta_2)}{\partial \theta}\Big| X = \theta_2\right]f(\theta_2; \theta_1)$$

$$+ E_{\mathbb{P}_X}\left[\frac{\partial L(X, Z; \theta_2)}{\partial \theta}1\{X > \theta_2\}\right].$$

Hence an SPA estimator is given by

$$E_{\mathbb{P}_X}\left[L(X, Z; \theta_2)\frac{\partial(-X + b(\theta_1))}{\partial \theta}\Big| X = b(\theta_1)\right]f(b(\theta_1); \theta_1)$$

$$+ E_{\mathbb{P}_X}\left[L(X, Z; \theta_2)\frac{\partial(X - \theta_2)}{\partial \theta}\Big| X = \theta_2\right]f(\theta_2; \theta_1) + \frac{\partial L(X, Z; \theta_2)}{\partial \theta}1\{X > \theta_2\}.$$

Note that there are conditional expectation terms in estimators obtained by SPA, which generally involves more computation, whereas only a *single* run of simulation is needed to obtain derivative estimates with respect to various parameters of interest with the SLRIPA estimator.

Remark 4.4 *[56] proposed an approach to obtain derivatives of a probability function. [74] generalized the results of [56] to functions of a more general form containing an indicator function and obtained a derivative estimator that generally has two parts, including a conditional expectation part, estimated using the kernel method. The conditional expectation part can be intuitively viewed as the derivative of the indicator function in the method of [74]. In SLRIPA, the effects of parameters of interest in the indicator function have been transferred into two different parts through a change of variables. One part of the effect is transferred into the function L, from which the IPA part of SLRIPA is obtained; the other part is in the Radon-Nikodym derivative, from which the LR part is obtained.*

Thus, the additional part in SLRIPA estimators involves no conditional expectation terms.

4.3   Unbiasedness of the Estimator

In this section, we give conditions such that the SLRIPA estimators derived in Section 4.2 are unbiased. Specifically, we show that the interchange of integral and derivative is valid under these conditions. Recall the random variable Y is defined as $Y = \Gamma(X; \theta_1, \theta_2)$, where Γ is a function of X with parameters $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$, where Θ_1 and Θ_2 are assumed to be open subsets of $\mathbb{R}^d$ and $\mathbb{R}$, respectively. Therefore the Radon-Nikodym derivative $\frac{d\mathbb{P}_Y}{d\mathbb{Q}}$ is a function of the parameters θ_1 and θ_2. To prove the unbiasedness of the estimators, we need the following assumptions:

A1. $\frac{\partial L(\cdot, Z; \theta_2)}{\partial X}$ and $\frac{\partial L(X, Z; \cdot)}{\partial \theta_2}$ exist almost everywhere.

A2. $\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}$ is differentiable almost everywhere with respect to θ_{1i} for $i = 1, \cdots, d$ and θ_2.

A3. The support of the distribution of X is given by $[a(\theta_1), \infty]$, where $a(\theta_1)$ is differentiable with respect to θ_1.

(a), For every θ_{1i}, $i = 1, \cdots, d$, which is a component of the vector $\theta_1 \in \Theta_1$, there exists an $\epsilon > 0$, such that if $B(\epsilon, \theta_{1i}) = \{\tilde{\theta}_{1i} : |\tilde{\theta}_{1i} - \theta_{1i}| < \epsilon, (\theta_{11}, \cdots, \tilde{\theta}_{1i}, \cdots, \theta_{1d}) \in \Theta_1\}$, then

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_{1i} \in B(\epsilon, \theta_{1i})} \left|\frac{\partial L((\theta_2 - a)Y + a, Z; \theta_2)}{\partial \tilde{\theta}_{1i}} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right] < \infty,$$

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_{1i} \in B(\epsilon, \theta_{1i})} \left|L((\theta_2 - a)Y + a, Z; \theta_2)\frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}}{\partial \tilde{\theta}_{1i}}(Y)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right] < \infty.$$

(b), For every $\theta_2 \in \Theta_2$, there exists an $\epsilon > 0$, such that if $B(\epsilon, \theta_2) = \{\tilde{\theta}_2 \in \Theta_2 : |\tilde{\theta}_2 - \theta_2| <$

$\epsilon\}$, then

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_2\in B(\epsilon,\theta_2)}\left|\frac{\partial L((\theta_2-a)Y+a,Z;\theta_2)}{\partial\tilde{\theta}_2}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right]<\infty,$$

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_2\in B(\epsilon,\theta_2)}\left|L((\theta_2-a)Y+a,Z;\theta_2)\frac{\partial\ln\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}}{\partial\tilde{\theta}_2}(Y)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right]<\infty.$$

A4. The support of the distribution of X is given by $[a(\theta_1),b(\theta_1)]$, where $a(\theta_1)$ and $b(\theta_1)$ are differentiable with respect to θ_1.

(a), For every θ_{1i}, $i=1,\cdots,d$, which is a component of the vector $\theta_1\in\Theta_1$, there exists an $\epsilon>0$, such that if $B(\epsilon,\theta_{1i})=\{\tilde{\theta}_{1i}:|\tilde{\theta}_{1i}-\theta_{1i}|<\epsilon,(\theta_{11},\cdots,\tilde{\theta}_{1i},\cdots,\theta_{1d})\in\Theta_1\}$, then

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_{1i}\in B(\epsilon,\theta_{1i})}\left|\frac{\partial L\left(\frac{b(\theta_2-a)Y+a(b-\theta_2)}{(\theta_2-a)Y+(b-\theta_2)},Z;\theta_2\right)}{\partial\tilde{\theta}_{1i}}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right]<\infty,$$

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_{1i}\in B(\epsilon,\theta_{1i})}\left|L\left(\frac{b(\theta_2-a)Y+a(b-\theta_2)}{(\theta_2-a)Y+(b-\theta_2)},Z;\theta_2\right)\frac{\partial\ln\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}}{\partial\tilde{\theta}_{1i}}(Y)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right]<\infty.$$

(b), For every $\theta_2\in\Theta_2$, there exists an $\epsilon>0$, such that if $B(\epsilon,\theta_2)=\{\tilde{\theta}_2\in\Theta_2:|\tilde{\theta}_2-\theta_2|<\epsilon\}$, then

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_2\in B(\epsilon,\theta_2)}\left|\frac{\partial L\left(\frac{b(\theta_2-a)Y+a(b-\theta_2)}{(\theta_2-a)Y+(b-\theta_2)},Z;\theta_2\right)}{\partial\tilde{\theta}_2}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right]<\infty,$$

$$E_{\tilde{\mathbb{Q}}}\left[\sup_{\tilde{\theta}_2\in B(\epsilon,\theta_2)}\left|L\left(\frac{b(\theta_2-a)Y+a(b-\theta_2)}{(\theta_2-a)Y+(b-\theta_2)},Z;\theta_2\right)\frac{\partial\ln\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}}{\partial\tilde{\theta}_2}(Y)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\right|\right]<\infty.$$

Theorem 4.1 *Under conditions A1, A2, A3(a), the SLRIPA stochastic derivative estimator given by (4.11) with $\theta=\theta_{1i}$ is unbiased for $\frac{\partial\tilde{J}(\theta_1,\theta_2)}{\partial\theta_{1i}}$, where θ_{1i} is a component of $\theta_1\in\Theta_1$ for $i=1,\cdots,d$. Under conditions A1, A2, A3(b), the SLRIPA stochastic derivative estimator given by (4.11) with $\theta=\theta_2\in\Theta_2$ is unbiased for $\frac{\partial\tilde{J}(\theta_1,\theta_2)}{\partial\theta_2}$.*

Proof. Assume that we are interested in the gradient estimator for $\frac{\partial\tilde{J}(\theta_1,\theta_2)}{\partial\theta_2}$. To indicate

that the probability measure $\mathbb{P}_Y$ depends on the parameter θ_2, we rewrite it as $\mathbb{P}_Y(\theta_2)$ and write the Radon-Nikodym derivative $\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}$ as $\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)$ in the following part of the proof.

For simplicity, denote

$$\varphi(\theta_2) = E_{\hat{\mathbb{Q}}}\Big[L((\theta_2-a)Y+a,Z;\theta_2)1\{Y>1\}\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)\Big]$$

$$M(Y,Z;\theta_2) = L((\theta_2-a)Y+a,Z;\theta_2)1\{Y>1\}.$$

Assume δ is sufficiently small, such that $\theta_2+\delta \in B(\epsilon,\theta_2)$. We have

$$\frac{\varphi(\theta_2+\delta)-\varphi(\theta_2)}{\delta} = E_{\hat{\mathbb{Q}}}\Big[\frac{1}{\delta}\Big(M(Y,Z;\theta_2+\delta)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2+\delta)$$

$$- M(Y,Z;\theta_2)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)\Big)\Big].$$

By A1, A2, and the mean value theorem, there exist $\eta_1 \in [\theta_2-\delta,\theta_2+\delta]$ and $\eta_2 \in [\theta_2-\delta,\theta_2+\delta]$, such that

$$E_{\hat{\mathbb{Q}}}\Big[\frac{1}{\delta}\Big(M(Y,Z;\theta_2+\delta)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2+\delta) - M(Y,z;\theta_2)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)\Big)\Big]$$

$$=E_{\hat{\mathbb{Q}}}\Big[\frac{1}{\delta}\Big(M(Y,Z;\theta_2+\delta)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2+\delta) - M(Y,Z;\theta_2)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2+\delta)\Big)$$

$$+ \frac{1}{\delta}\Big(M(Y,Z;\theta_2)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2+\delta) - M(Y,Z;\theta_2)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)\Big)\Big]$$

$$=E_{\hat{\mathbb{Q}}}\Big[\frac{\partial}{\partial\tilde{\theta}_2}M(Y,Z;\tilde{\theta}_2)\Big|_{\tilde{\theta}_2=\eta_1(Y,Z)}\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2+\delta) \qquad (4.16)$$

$$+ M(Y,Z;\theta_2)\frac{\partial}{\partial\tilde{\theta}_2}\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\tilde{\theta}_2)\Big|_{\tilde{\theta}_2=\eta_2(Y,Z)}\Big]$$

$$=E_{\tilde{\mathbb{Q}}}\Big[\frac{\partial}{\partial\tilde{\theta}_2}M(Y,Z;\tilde{\theta}_2)\Big|_{\tilde{\theta}_2=\eta_1(Y,Z)}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y,\theta_2+\delta)$$

$$+ M(Y,Z;\theta_2)\frac{\partial}{\partial\tilde{\theta}_2}\ln\Big(\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\tilde{\theta}_2)\Big)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y,\tilde{\theta}_2)\Big|_{\tilde{\theta}_2=\eta_2(Y,Z)}\Big] \qquad (4.17)$$

Therefore, by A3 and (4.16), we have

$$\lim_{\delta \to 0} E_{\hat{\mathbb{Q}}}\Big[\frac{1}{\delta}\Big(M(Y,Z;\theta_2+\delta)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2+\delta) - M(y,z;\theta_2)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)\Big)\Big]$$

$$\leq \lim_{\delta \to 0} E_{\tilde{\mathbb{Q}}}\Big[\sup_{\tilde{\theta}_2 \in B(\epsilon,\theta_2)} \Big|\frac{\partial}{\partial \tilde{\theta}_2} M(Y,Z;\tilde{\theta}_2)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y,\theta_2+\delta)\Big|_{\tilde{\theta}_2=\eta_1(Y,Z)}\Big]$$

$$+ \lim_{\delta \to 0} E_{\tilde{\mathbb{Q}}}\Big[\sup_{\tilde{\theta}_2 \in B(\epsilon,\theta_2)} \Big|M(Y,Z;\theta_2)\frac{\partial}{\partial \tilde{\theta}_2}\ln\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\tilde{\theta}_2)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y,\tilde{\theta}_2)\Big|_{\tilde{\theta}_2=\eta_2(Y,Z)}\Big] < \infty.$$

By the dominated convergence theorem, we have

$$\frac{\partial}{\partial \theta_2} E_{\hat{\mathbb{Q}}}\Big[M(Y,Z;\theta_2)\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)\Big] = E_{\tilde{\mathbb{Q}}}\Big[\Big(\frac{\partial}{\partial \theta_2}M(Y,Z;\theta_2)$$

$$+ M(Y,Z;\theta_2)\frac{\partial}{\partial \theta_2}\ln\frac{d\mathbb{P}_Y}{d\hat{\mathbb{Q}}}(Y,\theta_2)\Big)\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y,\theta_2)\Big].$$

Hence, the estimator (4.11) is unbiased. The theorem can be proved analogously when $\theta = \theta_{1i}$ for $i = 1, \cdots, d$.

Theorem 4.2 *Under conditions A1, A2, A4(a), the SLRIPA stochastic derivative estimator given by (4.12) with $\theta = \theta_{1i}$ for $\frac{\partial \tilde{J}(\theta_1,\theta_2)}{\partial \theta_{1i}}$ is unbiased, where θ_{1i} is a component of $\theta_1 \in \Theta_1$ for $i = 1, \cdots, d$. Under conditions A1, A2, A4(b), the SLRIPA stochastic derivative estimator given by (4.12) with $\theta = \theta_2 \in \Theta_2$ for $\frac{\partial \tilde{J}(\theta_1,\theta_2)}{\partial \theta_2}$ is unbiased.*

Proof. This theorem follows from an analogous argument as in the proof of Theorem 4.1.

4.4 Application of SLRIPA Derivative Estimator

In this section, we give several examples to illustrate the application of the proposed SLRIPA derivative estimator.

4.4.1 Derivative Estimation for European Call Options

4.4.1.1 Black-Scholes Model

We consider a European call option with the underlying security price $\{S_t\}$ satisfying the Black-Scholes model $dS_t = S_t(rdt + \sigma dB_t)$, where r is the riskless rate, σ is the volatility parameter, and B_t is a standard Brownian motion. The payoff function of a European call option is given by $J_E(S_T) = e^{-rT}(S_T - K)1\{S_T \geq K\}$, where T is the expiration date of the option and K is the strike price. The lognormal random variable S_T can be represented as $S_T = S_0 \exp[(r - \sigma^2/2)T + \sigma\sqrt{T}Z]$, where Z is a standard normal random variable and S_0 is the initial stock price. Suppose that we are interested in estimating *delta*, which is the derivative of $E[J_E(S_T)]$ with respect to S_0. Since the payoff function $J_E(S_T)$ is continuous, we have

$$delta = \frac{\partial E[J_E(S_T)]}{\partial S_0} = E\Big[e^{-rT}\frac{\partial S_T}{\partial S_0}1\{S_T \geq K\}\Big] = E\Big[e^{-rT}\frac{S_T}{S_0}1\{S_T \geq K\}\Big].$$

$e^{-rT}\frac{S_T}{S_0}1\{S_T \geq K\}$ is the IPA estimator. Similarly, the IPA estimator for $\frac{\partial E[J_E]}{\partial K}$ is given by $-e^{-rT}1\{S_T \geq K\}$. However, IPA cannot be used to estimate the second derivative *gamma*, which is defined as $gamma = \frac{\partial^2 E[J_E(S_T)]}{\partial S_0^2} = \frac{\partial}{\partial S_0}E\Big[e^{-rT}\frac{S_T}{S_0}1\{S_T \geq K\}\Big]$, since $e^{-rT}\frac{S_T}{S_0}1\{S_T \geq K\}$ is not a continuous function of S_T. The LR method can be used to obtain a derivative estimator for *gamma* ([17]), but it is no longer applicable to estimate

$\frac{\partial E[J_E]}{\partial K}$. SLRIPA can be applied to estimate first derivatives with respect to all the parameters of interest and to estimate the second derivative *gamma*. In this simple example, both IPA and SLRIPA are applicable for first derivatives, and IPA is preferred since it generally has lower variance.

To obtain SLRIPA estimators, we define $Y = S_T/K$; the probability density function of Y is given by $f_{E,Y}(y) = \frac{dP(S_T/K \le y)}{dy} = \frac{1}{\sigma\sqrt{T}y}n(d)$, where $n(\cdot)$ is the probability density of a standard normal random variable, and $d = \frac{\ln(Ky/S_0) - (r - \sigma^2)T}{\sigma\sqrt{T}}$. The SLRIPA derivative estimator (4.8) for $\frac{\partial E[J_E]}{\partial \theta}$ (θ could be S_0, r, K, or σ) is given by

$$e^{-rT}\frac{\partial(KY - K)}{\partial \theta}1\{Y > 1\}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) + e^{-rT}(KY - K)1\{Y > 1\}\frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y),$$

where $d\mathbb{P}_Y = f_{E,Y}(Y)dY$, and $\tilde{\mathbb{Q}}$ is a probability measure such that $\mathbb{P}_Y$ is absolutely continuous with respect to $\tilde{\mathbb{Q}}$. The second derivative *gamma* can be obtained by the SLRIPA technique:

$$
\begin{aligned}
gamma &= \frac{d^2 E[J_E(S_T)]}{dS_0^2} = \frac{d}{dS_0}E_{\tilde{\mathbb{Q}}}\Big[e^{-rT}(KY - K)1\{Y > 1\}\frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial S_0}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\Big] \\
&= E_{\tilde{\mathbb{Q}}}\Big[e^{-rT}(KY - K)1\{Y > 1\}\frac{\partial^2(\ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y))}{\partial S_0}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\Big] \\
&\quad + E_{\tilde{\mathbb{Q}}}\Big[e^{-rT}(KY - K)1\{Y > 1\}\Big(\frac{\partial(\ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y))}{\partial S_0}\Big)^2\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)\Big].
\end{aligned}
$$

Since the support of the distribution of S_T is the interval $(0, \infty)$, which does not depend on S_0, the SLRIPA derivative estimator with respect to the parameter S_0 turns out to be the same as the LR derivative estimator. This is true for other parameters as well, except for K, where LR cannot be applied directly.

4.4.1.2   Ornstein-Uhlenbeck Model

We consider a European call option with the underlying security price $\{S_t\}$ satisfying the Ornstein-Uhlenbeck model: $dS_t = \rho(\mu - S_t)dt + \sigma dB_t$, where ρ is a positive constant, σ the volatility parameter, and B_t a standard Brownian motion. The payoff function of a European call option is given by $J_E(S_T) = e^{-rT}(S_T - K)1\{S_T \geq K\}$. S_T can be written exactly as $S_T = S_0 e^{-\rho T} + \mu(1 - e^{-\rho T}) + \sigma\sqrt{\frac{1-e^{-2\rho T}}{2\rho}}Z$, where Z is a standard normal random variable and S_0 is the initial stock price.

Suppose that we are interested in estimating $\frac{\partial E[J_E(S_T)]}{\partial K}$. The LR method cannot be applied, since K is in an indicator function. Now we derive an SLRIPA estimator. We define $Y = S_T/K$; the probability density function of Y is given by $f_{E,Y}(y) = \frac{dP(S_T/K \leq y)}{dy} = \frac{K}{m_2}n(d)$, where $m_1 = S_0 e^{-\rho T} + \mu(1 - e^{-\rho T})$, $m_2 = \sigma\sqrt{\frac{1-e^{-2\rho T}}{2\rho}}$, and $d = \frac{Ky-m_1}{m_2}$. The SLRIPA derivative estimator (4.8) for $\frac{\partial E[J_E(S_T)]}{\partial K}$ is given by

$$e^{-rT}(Y-1)1\{Y>1\}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) + e^{-rT}(KY - K)1\{Y > 1\}\frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial K}\frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y),$$

where $d\mathbb{P}_Y = f_{E,Y}(Y)dY$ and $\tilde{\mathbb{Q}}$ is a probability measure such that $\mathbb{P}_Y$ is absolutely continuous with respect to $\tilde{\mathbb{Q}}$. If $\tilde{\mathbb{Q}}$ is the probability measure induced by S_T, the SLRIPA estimator for $\frac{\partial E[J_E(S_T)]}{\partial K}$ in terms of S_T is given as

$$\mathcal{D}_K = e^{-rT}(S_T/K - 1)1\{S_T > K\} + e^{-rT}(S_T - K)1\{S_T > K\}\frac{(m_2)^2 - S_T(S_T - m_1)}{(m_2)^2 K}.$$

4.4.2   Derivative Estimation for Barrier Option

We consider a discretely monitored up-and-out European barrier option, which is worthless if the underlying security price exceeds a barrier H. The discounted payoff

function is given by $J_B(S_m) = e^{-rT}(S_m - K)1\{S_m \geq K\}1\{\max\{S_1, \cdots, S_m\} \leq H\}$, where S_i denotes the discretely monitored price S_{t_i} for $i = 1, \cdots, m$ and $\{0 = t_0 < t_1 < \cdots < t_m = T\}$. $J_B(S_m)$ can be rewritten as $J_B(S_m) = e^{-rT}(S_m - K)1\{S_m \geq K\}\prod_{i=1}^m 1\{S_i \leq H\}$. Assume S_t follows a geometric Brownian motion, i.e., $S_t = S_0 e^{(r-\sigma^2/2)t+\sigma B_t}$.

We are interested in estimating $\frac{\partial E[J_B(S_m)]}{\partial \theta}$, where θ could be S_0, r, σ, and H, which correspond to *delta, rho*, and *vega*, respectively, for the first three parameters. Now we derive SLRIPA derivative estimators. We know that $S_{i+1} = S_i e^{(r-\sigma^2/2)(t_{i+1}-t_i)+\sigma\sqrt{t_{i+1}-t_i}Z_{i+1}}$, where $\{Z_{i+1}\}$ are independent and identically distributed standard normal random variables. Define $Y = S_1/H$, and $X_{i+1} = e^{(r-\sigma^2/2)(t_{i+1}-t_i)+\sigma\sqrt{t_{i+1}-t_i}Z_{i+1}}$ for $i = 1, \cdots, m-1$. It is easy to see that $S_1 = YH$ and $S_i = S_1 \prod_{k=2}^i X_k$ for $i > 1$; $S_i \leq H$ is equivalent to $Y \leq 1$ for $i = 1$, and $Y \prod_{k=2}^i X_k \leq 1$ for $i > 1$.

Let f_Y denote the probability density function of Y, and let f_i denote the probability density function of X_i for $i = 2, \cdots, m$. By the results of Section 4.2, we write the SLRIPA estimator in terms of Y and $\{X_2, \cdots, X_m\}$ as (see the details of the derivation in the appendix):

$$
\begin{aligned}
\mathcal{D}_\theta =& \frac{\partial e^{-rT}\left(HY\prod_{i=2}^m X_i - K\right)}{\partial \theta}1\{HY\prod_{k=2}^m X_k \geq K\}1\{Y \leq 1\}\Big\{\prod_{i=2}^m 1\{Y\prod_{k=2}^i X_k \leq 1\}\Big\} \\
&+ e^{-rT}\left(HY\prod_{k=2}^m X_k - K\right)1\{HY\prod_{k=2}^m X_k \geq K\}1\{Y \leq 1\}\prod_{i=2}^m 1\{Y\prod_{k=2}^i X_k \leq 1\} \\
&\times \left(\frac{\partial \ln f_Y}{\partial \theta} + \sum_{i=1}^m \frac{\partial \ln f_i}{\partial \theta}\right).
\end{aligned}
$$

In particular, the SLRIPA derivative estimator for *delta* can be written in terms of the

original random variables $S_1, \cdots, S_m$ as

$$\mathcal{D}_{S_0} = e^{-rT}(S_m - K)1\{S_m \geq K\} \prod_{i=1}^{m} 1\{S_i \leq H\} \left(\frac{\ln S_1 - \ln S_0 - (r - \sigma^2/2)t_1}{S_0 \sigma^2 t_1} \right).$$

The SLRIPA derivative estimator for $\frac{\partial E[J_B(S_m)]}{\partial H}$ is

$$\mathcal{D}_H = e^{-rT}\frac{S_m}{H}1\{S_m \geq K\} \prod_{i=1}^{m} 1\{S_i \leq H\} + e^{-rT}(S_m - K)1\{S_m \geq K\}$$

$$\times \prod_{i=1}^{m} 1\{S_i \leq H\} \left(-\frac{\ln S_1 - \ln S_0 - (r - \sigma^2/2)t_1}{H \sigma^2 t_1} \right).$$

Compared with the payoff function of the European call option, there are multiple indicator functions in the payoff function of the barrier option. By using the SLRIPA method, we circumvent the difficulty of differentiating all the indicator functions. Although the SLRIPA derivative estimators with respect to S_0, r, σ are the same as the LR estimators, SLRIPA is applicable for $\frac{\partial E[J_B(S_m)]}{\partial H}$, whereas LR is not. Again we only need to generate one sample path to obtain a derivative estimate for all the parameters of interest.

4.4.3 Sensitivity Analysis for American Call Option

In this section, we will apply SLRIPA to an American call option pricing problem.

4.4.3.1 Price Model

Assume that the underlying stock of the American call option distributes a known cash dividend of amount D_j at time $t_j = \sum_{i=1}^{j} \tau_j$ $(\tau_j > 0)$, $j = 1, \ldots, \eta(T)$, where $\eta(T)$ is the number of dividends distributed during the lifetime of the call option, τ_1 denotes the time until the first ex-dividend date, τ_i, $i = 2, \ldots, \eta(T) - 1$ denote the time between

subsequent ex-dividends, and $\tau_{\eta(T)}$ denotes the time from the last ex-dividend date to the expiration date. We denote the stock price at time t as S_t.

Following the standard models (e.g., [98]), assume that after the ex-dividend, the stock price drops by the amount of the dividend, i.e., $S_{t_j^+} = S_{t_j^-} - D_j$. We also denote $\tau_{\eta(T)+1} = T - \sum_{i=1}^{\eta(T)} \tau_i$, $\tau_0 = 0$, $t_{\eta(T)+1} = T$. The dividend amounts are assumed to be deterministic and known. Although an American call option can be exercised at any time before the expiration date T, it is well known that the option should only be exercised right before an ex-dividend date or at the expiration date. Therefore, we assume that on every ex-dividend date τ_j, there is a corresponding threshold stock price $s_j (\geq K)$ that does not depend on the other parameters. and the option is exercised if $S_{t_j^-} > s_j$, where K is the strike price of the American option. The payoff function of the American call option can be written as $J_T = e^{-rT} \hat{J}_T$, where

$$
\hat{J}_T = \Big(\sum_{i=1}^{\eta(T)} \Big[\prod_{j=1}^{i-1} 1\{S_{t_j^-} \leq s_j\} 1\{S_{t_i^-} > s_i\}(S_{t_i^-} - K)^+ e^{r(T-t_i)} \Big]
$$
$$
+ \prod_{j=1}^{\eta(T)} 1\{S_{t_j^-} \leq s_j\}(S_T - K)^+ \Big).
$$

We are interested in estimating the sensitivity of the option price $E[J_T]$ with respect to the parameter θ, which could be $r, s_i, \ i = 1, \cdots, \eta(T)$, or other parameters of interest, i.e., our goal is to estimate: $\frac{\partial E[J_T]}{\partial \theta} = e^{-rT} \Big[\frac{\partial E[\hat{J}_T]}{\partial \theta} - E[\hat{J}_T] \frac{\partial (rT)}{\partial \theta} \Big]$. The option price is equal to the expected payoff $E[J_T]$ under the optimal (payoff maximizing) set of threshold values $\{s_j^\star\}$.

Assume that aside from the discrete jumps at the ex-dividend dates, the stock price changes continuously, i.e., according to a function $h(Z; S, t, r, \sigma)$, where S is the current stock price, σ is a parameter, and Z is a random vector that does not depend on

other parameters. $h(Z; S, t, r, \sigma)$ gives the stock price after duration t from the present given current stock price S. Define the continuous part of the stock price as $\tilde{S}_t$: $\tilde{S}_t = h(Z; \tilde{S}_0, t, r, \sigma)$. The stock price with dividends can be recovered from $\tilde{S}_t$ by discounting the dividends over the correct period of time as follows:

$$S_t = \tilde{S}_t + \sum_{i=j+1}^{\eta(T)} D_i e^{-r(t_i - t)}, \quad for \ t_j \le t < t_{j+1}, \ j = 0, 1, \ldots, \eta(T). \qquad (4.18)$$

Just prior to ex-dividend dates where an early exercise decision is made, the stock price is given by $S_{t_j^-} = \tilde{S}_{t_j} + \sum_{i=j}^{\eta(T)} D_i \exp\left(-r \sum_{k=j+1}^{i} \tau_k \right)$, $j = 1, \ldots, \eta(T)$. To simplify the problem, we first consider the case when $\eta(T) = 1$. We will drop the subscript on the dividend, i.e., $D_1 \triangleq D$ and $S_{t_1^+} = S_{t_1^-} - D$. Let s be the threshold parameter. The payoff function without discount is:

$$\hat{J}_T = 1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+ e^{r(T - t_1)} + 1\{S_{t_1^-} \le s\}(S_T - K)^+. \qquad (4.19)$$

4.4.3.2 SLRIPA Gradient Estimator

Assume the underlying security price of the American call option is governed by a geometric Brownian motion, i.e., the continuous part of the stock price is given by $h(Z; \tilde{S}_0, t, r, \sigma) = \tilde{S}_0 e^{(r - \sigma^2/2)t + \sigma\sqrt{t}Z}$, and $h^{-1}(y, \tilde{S}_0, t, r, \sigma) = \frac{\ln y - \ln \tilde{S}_0 - (r - \sigma^2/2)t}{\sigma\sqrt{t}}$, where $\tilde{S}_0 = S_0 - De^{-r\tau_1}$ and Z is a standard normal random variable. We have $\tilde{S}_{t_1^-} = \tilde{S}_0 e^{(r - \sigma^2/2)\tau_1 + \sigma\sqrt{\tau_1}Z_1}$, $S_{t_1^-} = \tilde{S}_{t_1^-} + D$, $S_T = \tilde{S}_T = (S_{t_1^-} - D)e^{(r - \sigma^2/2)\tau_2 + \sigma\sqrt{\tau_2}Z_2}$, where Z_1 and Z_2 are two independent standard normal random variables. The payoff function (4.19) falls into the type of payoff functions given by (4.1) to which SLRIPA can be applied.

Denote $\hat{J}_T^1 = 1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+ e^{r(T - t_1)}$ and $\hat{J}_T^2 = 1\{S_{t_1^-} \le s\}(S_T - K)^+$.

Note that $D \le S_{t_1^-} < \infty$. Let $\theta_1 = (r, \sigma, S_0, D, K)$, $\theta_2 = s$, and $a(\theta_1) = D, b = \infty$. From the analysis in Section 4.2.1.1, we define $Y = \Gamma(S_{t_1^-}; \theta_1, \theta_2) = \frac{S_{t_1^-} - D}{s - D} = \frac{\tilde{S}_{t_1^-}}{s - D}$. Then $S_{t_1^-} = \Gamma^{-1}(Y; \theta_1, \theta_2) = Y(s - D) + D$, and $S_{t_1^-} > s$ is equivalent to $\psi(Y) = Y - 1 > 0$. Since $s > K > D$, $s - D$ is always positive, the probability density function of the new random variable Y is given by

$$f_Y(Y \le y) = \frac{\partial P(Y \le y)}{\partial y} = \frac{1}{y\sigma\sqrt{\tau_1}} n(h^{-1}(y(s - D); \tilde{S}_0, \tau_1, r, \sigma)).$$

The payoff function for $E[\hat{J}_T^1]$ is $L(S_{t_1^-}; \theta_1, \theta_2) = (S_{t_1^-} - K)^+$, and the SLRIPA estimator (4.8) for $\frac{\partial E[\hat{J}_T^1]}{\partial \theta}$, where θ could be $S_0, r, \sigma, s,$ or K, is given by

$$\begin{aligned}
\mathcal{D}_1 =& 1\{Y > 1\} \frac{\partial}{\partial \theta}[(Y(s - D) + D - K)^+ e^{r(T - t_1)}] \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) \\
& + 1\{Y > 1\} e^{r(T - t_1)}(Y(s - D) + D - K)^+ \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y).
\end{aligned}$$

For the derivative of $E[\hat{J}_T^2]$ with respect to θ, note that $S_{t_1^-} \le s$ is equivalent to $\psi(Y) = -Y + 1 \ge 0$, and the payoff function is $L(S_{t_1^-}, Z_2; \theta_1, \theta_2) = [h(Z_2; S_{t_1^-} - D, \tau_2, r, \sigma) - K]^+$. The SLRIPA estimator (4.8) for $\frac{\partial E[\hat{J}_T^2]}{\partial \theta}$ is given by

$$\begin{aligned}
\mathcal{D}_2 =& 1\{Y \le 1\} \frac{\partial}{\partial \theta}[h(Z_2; Y(s - D), \tau_2, r, \sigma) - K]^+ \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y) \\
& + 1\{Y \le 1\}[h(Z_2; Y(s - D), \tau_2, r, \sigma) - K]^+ \frac{\partial \ln \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y)}{\partial \theta} \frac{d\mathbb{P}_Y}{d\tilde{\mathbb{Q}}}(Y).
\end{aligned}$$

Since $\frac{\partial E[\hat{J}_T]}{\partial \theta} = \frac{\partial E[\hat{J}_T^1]}{\partial \theta} + \frac{\partial E[\hat{J}_T^2]}{\partial \theta}$, the derivative estimator for $E[J_T]$ with respect to θ is $\mathcal{D}_\theta = e^{-rT}(\mathcal{D}_1 + \mathcal{D}_2)$. In particular, if $\tilde{\mathbb{Q}}$ is the probability measure induced by $S_{t_1^-}$, the

117

SLRIPA estimator for $\frac{\partial E[J_T]}{\partial s}$ can be written in terms of $S_{t_1^-}$ and S_T as

$$\mathcal{D}_s = e^{-rt_1}1\{S_{t_1^-} > s\}1\{S_{t_1^-} > K\}\frac{S_{t_1^-} - D}{s - D}$$

$$+ e^{-rt_1}1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+\frac{1}{\sigma^2\tau_1(s - D)}M$$

$$+ e^{-rT}1\{S_{t_1^-} \le s\}\frac{S_T}{s - D} + e^{-rT}1\{S_{t_1^-} \le s\}(S_T - K)^+\frac{1}{\sigma^2\tau_1(s - D)}M, \quad (4.20)$$

where $M = \left(\ln(S_{t_1^-} - D) - \ln\tilde{S}_0 - (r - \sigma^2/2)\tau_1\right)$.

The SLRIPA estimators for $E[J_T]$ with respect to parameters K, S_0, σ, r, and D written in terms of original random variables are given as follows. To simplify the notation, we define $M = \left(\ln(S_{t_1^-} - D) - \ln\tilde{S}_0 - (r - \sigma^2/2)\tau_1\right)$.

$1 : \theta = K$

$$e^{-rt_1}1\{S_{t_1^-} > s\}1\{S_{t_1^-} > K\} - e^{-rT}1\{S_{t_1^-} \le s\}1\{S_T > K\}.$$

$2 : \theta = S_0$

$$e^{-rt_1}1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+\frac{1}{\sigma^2\tilde{S}_0\tau_1}M + e^{-rT}1\{S_{t_1^-} \le s\}(S_T - K)^+\frac{1}{\sigma^2\tilde{S}_0\tau_1}M.$$

$3 : \theta = \sigma$

$$e^{-rt_1}1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+\left[\frac{-1}{\sigma} + \right.$$

$$\left.\frac{1}{\sigma^3\tau_1}\left(\left(\ln(S_{t_1^-} - D) - \ln\tilde{S}_0 - r\tau_1\right)^2 - \frac{\sigma^4(\tau_1)^2}{4}\right)\right]$$

$$+ e^{-rT}1\{S_{t_1^-} \le s\}1\{S_T > K\}S_T(-\sigma\tau_2 + \sqrt{\tau_2}Z_2)$$

$$+ e^{-rT}1\{S_{t_1^-} \le s\}(S_T - K)^+\left[\frac{-1}{\sigma}\right.$$

$$\left.+ \frac{1}{\sigma^3\tau_1}\left(\left(\ln(S_{t_1^-} - D) - \ln\tilde{S}_0 - r\tau_1\right)^2 - \frac{\sigma^4(\tau_1)^2}{4}\right)\right].$$

$$(4.21)$$

$4 : \theta = r$

$$e^{-rt_1} 1\{S_{t_1^-} > s\}(T - t_1)(S_{t_1^-} - K)^+ + e^{-rt_1} 1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+ \frac{M}{\sigma^2}$$

$$+ e^{-rT} 1\{S_{t_1^-} \leq s\}\tau_2(S_T - K)^+ + e^{-rT} 1\{S_{t_1^-} \leq s\}(S_T - K)^+ \frac{M}{\sigma^2}.$$

$5 : \theta = D$

$$e^{-rt_1} 1\{S_{t_1^-} > s\} 1\{S_{t_1^-} > K\}\Big(-\frac{S_{t_1^-} - D}{s - D} + 1\Big)$$

$$+ e^{-rt_1} 1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+ \frac{M}{\sigma^2 \tau_1(s - D)}$$

$$+ e^{-rT} 1\{S_{t_1^-} \leq s\} 1\{S_T > K\}\frac{S_T}{-s + D}$$

$$+ e^{-rT} 1\{S_{t_1^-} \leq s\}(S_T - K)^+ \frac{M}{\sigma^2 \tau_1(s - D)}.$$

Remark 4.5 *Note that the indicator functions in $\hat{J}_T$ have a parameter s, so the direct IPA estimator with respect to the parameter s is biased. Using the smoothing property of conditional expectation, [36] derived an unbiased smoothed perturbation analysis (SPA) estimator, but it requires additional simulations.*

Remark 4.6 *For American options with multiple ex-dividend dates, SLRIPA estimators are given in the online e-companion. Each estimate of the derivative with respect to all the parameters including s_i, $i = 1, \ldots, \eta(T)$ can be obtained by generating only one sample path, which is not the case for estimators given in [36] and [51].*

Remark 4.7 *In the American call option pricing example, if Euler discretization with time step Δt is used to simulate the stock price, then the number of terms in the sum of likelihood ratios in SLRIPA increases as $\Delta t \to 0$; thus the variance of SLRIPA increases linearly with the number of time steps just as in the LR method. Similarly, the number of terms in the sum of likelihood ratios increases linearly with the number of ex-dividend dates. However for the threshold parameters, s_i only appears in one of the marginal probability*

densities, and hence the variance of the derivative estimator with respect to s_i does not increase as the number of ex-dividend dates increases.

4.4.3.3   Unbiasedness Result

Based on Theorem 4.1, unbiasedness of all the estimators can be verified. The following theorem establishes the result for the threshold parameter in the single dividend case, and proofs for the other examples and parameters follow analogous arguments.

Theorem 4.3 *The estimator given by (4.20) is unbiased for $\frac{\partial E[J_T]}{\partial s}$.*

Proof.   The proof consists in finding a bound required by the dominated convergence theorem to justify the interchange of integration and differentiation. Note that $s \geq K$ and we have

$$E[S_{t_1^-}] = E[\tilde{S}_0 e^{(r-\sigma^2/2)\tau_1 + \sigma\sqrt{\tau_1}Z_1} + D] < \infty,$$

$$E[S_T] = E[\tilde{S}_0 e^{(r-\sigma^2/2)\tau_1 + \sigma\sqrt{\tau_1}Z_1} e^{(r-\sigma^2/2)(T-t_1) + \sigma\sqrt{T-t_1}Z_2}] < \infty.$$

Therefore

$$E\left[e^{-rt_1} \sup_s \left\{ 1\{S_{t_1^-} > s\} 1\{S_{t_1^-} - K > 0\} \frac{S_{t_1^-} - D}{s - D} \right\} \right] \leq E\left[e^{-rt_1} \frac{S_{t_1^-} - D}{K - D}\right] < \infty.$$

Similarly, we also have

$$E\left[e^{-rt_1} \sup_s \left\{ 1\{S_{t_1^-} > s\}(S_{t_1^-} - K)^+ \frac{1}{\sigma^2\tau_1(s-D)} \right.\right.$$
$$\left.\left. \times \left(\ln\left(\frac{(S_{t_1^-} - D)}{\tilde{S}_0}\right) - (r - \sigma^2/2)\tau_1 \right) \right\} \right]$$
$$\leq E\left[e^{-rt_1}(S_{t_1^-} - K)^+ \frac{1}{\sigma^2\tau_1(K-D)} \left(\ln\left(\frac{S_{t_1^-}}{\tilde{S}_0}\right) - (r - \sigma^2/2)\tau_1 \right) \right] < \infty,$$

and

$$E\Big[e^{-rT}\sup_s\Big\{1\{S_{t_1^-}\le s\}\frac{S_T}{s-D}\Big\}\Big]\le E\Big[e^{-rT}\frac{S_T}{K-D}\Big]<\infty.$$

We also have

$$E\Big[e^{-rT}\sup_s\Big\{1\{S_{t_1^-}\le s\}(S_T-K)^+\frac{1}{\sigma^2\tau_1(s-D)}$$
$$\times\Big(\ln\Big(\frac{(S_{t_1^-}-D)}{\tilde S_0}\Big)-(r-\sigma^2/2)\tau_1\Big)\Big\}\Big]$$
$$\le E\Big[e^{-rT}(S_T-K)^+\frac{1}{\sigma^2\tau_1(K-D)}\Big(\ln\Big(\frac{S_{t_1^-}}{\tilde S_0}\Big)-(r-\sigma^2/2)\tau_1\Big)\Big]<\infty.$$

Therefore condition A3(b) in Section 4.3 holds. Conditions A1 and A2 hold by the property

of the payoff function J_T. The proof is completed by Theorem 4.1. ∎

4.4.3.4 The American Call Option with a Different Stock Price Model

Every m days, the continuous part of the stock price is given as: $\tilde S_t = \tilde S_0 e^{rt}X$,

where $X\sim U(1-\sigma\sqrt{t},1+\sigma\sqrt{t})$ and $\sigma\sqrt{t}<1$. Let $h_i(Z_i;S,t,r,\sigma)=Se^{rt}(1+\sigma\sqrt{t}Z_i)$,

where $\{Z_i\}$ are iid random variables distributed as $U(-1,1)$. Define h_i as $h_i(S;t,r,\sigma)=$

$Se^{rt}\Big[\prod_{j=1}^i(1+\sigma\sqrt{m}Z_j)\Big](1+\sigma\sqrt{t-im}Z_{i+1})$. for $t\in[im,(i+1)m]$. For simplicity, we

will assume $\tau_1=m$, $\tau_2<m$, $T=\tau_1+\tau_2$. Then

$$h_1^{-1}(y;\tilde S_0,\tau_1,r,\sigma)=\frac{1}{\sigma\sqrt{\tau_1}}\Big(\frac{y}{\tilde S_0 e^{r\tau_1}}-1\Big);$$

$$\tilde S_0=S_0-De^{-r\tau_1};$$

$$\tilde S_{t_1^-}=h_0(\tilde S_0;\tau_1,r,\sigma)=\tilde S_0 e^{r\tau_1}(1+\sigma\sqrt{\tau_1}Z_1);$$

121

$$S_{t_1^-} = \tilde{S}_{t_1^-} + D = \tilde{S}_0 e^{r\tau_1}(1 + \sigma\sqrt{\tau_1}Z_1) + D;$$

$$\tilde{S}_T = (S_{t_1^-} - D)e^{r\tau_2}(1 + \sigma\sqrt{T - \tau_1}Z_2).$$

Let $\theta_1 = (r, \sigma, S_0, D, K)$ and $\theta_2 = s$. Note that $S_{t_1^-} \in [a, b]$, where $a = \tilde{S}_0 e^{r\tau_1}(1 - \sigma\sqrt{\tau_1}) + D$ and $b = \tilde{S}_0 e^{r\tau_1}(1 + \sigma\sqrt{\tau_1}) + D$, hence we can apply the SLRIPA estimator in Section 4.2.1.2. Define a new random variable

$$Y = \Gamma(S_{t_1^-}; \theta_1, \theta_2) = \frac{(S_{t_1^-} - a)(b - s)}{(b - S_{t_1^-})(s - a)}, \tag{4.22}$$

then

$$S_{t_1^-} = \Gamma^{-1}(Y; \theta_1, \theta_2) = \frac{bY(s - a) + a(b - s)}{Y(s - a) + (b - s)}.$$

Hence

$$P(Y \leq y) = P\left(\frac{(S_{t_1^-} - a)(b - s)}{(b - S_{t_1^-})(s - a)} \leq y\right) = P\left(S_{t_1^-} \leq \Gamma^{-1}(y; \theta_1, \theta_2)\right)$$

$$= P\left(Z_1 \leq h_1^{-1}(\Gamma^{-1}(y; \theta_1, \theta_2) - D; \tilde{S}_0, \tau_1, r, \sigma)\right)$$

$$= \int_{-1}^{h_1^{-1}(\Gamma^{-1}(y; \theta_1, \theta_2) - D; \tilde{S}_0, \tau_1, r, \sigma)} f_1(z_1) dz_1,$$

where $f_1(z_1) = 1/2$ for $z_1 \in (-1, 1)$, so the probability density function of Y is

$$f_Y(y) = \frac{\partial P(Y \leq y)}{\partial y} = \frac{1}{2}\frac{\partial h_1^{-1}(\Gamma^{-1}(y; \theta_1, \theta_2) - D; \tilde{S}_0, \tau_1, r, \sigma)}{\partial y}$$

$$= \frac{1}{2\sigma\sqrt{\tau_1}\tilde{S}_0 e^{r\tau_1}}\frac{(b - a)(b - s)(s - a)}{(y(s - a) + (b - s))^2}.$$

The payoff function for $E[\hat{J}_T^1]$ is $L(S_{t_1^-}; \theta_1, \theta_2) = (S_{t_1^-} - K)^+$, and by the analysis in

Section 4.2, the SLRIPA estimator (4.9) for $\frac{\partial E[\hat{J}_T^1]}{\partial \theta}$ is given by:

$$1\{Y > 1\}1\{\Gamma^{-1}(Y;\theta_1,\theta_2) > K\}\frac{\partial(\Gamma^{-1}(Y;\theta_1,\theta_2) - K)}{\partial \theta}e^{r(T-\tau_1)}$$

$$+ 1\{Y > 1\}e^{r(T-\tau_1)}(\Gamma^{-1}(Y;\theta_1,\theta_2) - K)^+\frac{\partial \ln f_Y(Y)}{\partial \theta}.$$

Consider estimating the derivative of $E[\hat{J}_T^2]$ with respect to θ. Note that $S_{t_1^-} \leq s$ is equivalent to $\psi(Y) = -Y + 1 \geq 0$, and the payoff function is $L(S_{t_1^-}, Z_2; \theta_1, \theta_2) = [h(Z_2; S_{t_1^-} - D, \tau_2, r, \sigma) - K]^+$. By the analysis in Section 4.2, then the SLRIPA estimator (4.9) for $\frac{\partial E[\hat{J}_T^2]}{\partial \theta}$ is given by

$$1\{Y \leq 1\}\frac{\partial(h_2(Z_2; \Gamma^{-1}(Y;\theta_1,\theta_2) - D, \tau_2, r, \sigma) - K)^+}{\partial \theta}$$

$$+ 1\{Y \leq 1\}(h_2(Z_2; \Gamma^{-1}(Y;\theta_1,\theta_2) - D, \tau_2, r, \sigma) - K)^+\frac{\partial \ln f_Y(Y)}{\partial \theta}.$$

So the full SLRIPA estimator for $E[J_T]$ with respect to θ is given by

$$e^{-r\tau_1}1\{Y > 1\}1\{\Gamma^{-1}(Y;\theta_1,\theta_2) > K\}\frac{\partial(\Gamma^{-1}(Y;\theta_1,\theta_2) - K)}{\partial \theta}$$

$$+ e^{-r\tau_1}1\{Y > 1\}(\Gamma^{-1}(Y;\theta_1,\theta_2) - K)^+\frac{\partial \ln f_Y(Y)}{\partial \theta}$$

$$+ e^{-rT}1\{Y \leq 1\}\frac{\partial(h_2(Z_2; \Gamma^{-1}(Y;\theta_1,\theta_2) - D, \tau_2, r, \sigma) - K)^+}{\partial \theta}$$

$$+ e^{-rT}1\{Y \leq 1\}(h_2(Z_2; \Gamma^{-1}(Y;\theta_1,\theta_2) - D, \tau_2, r, \sigma) - K)^+\frac{\partial \ln f_Y(Y)}{\partial \theta}$$

$$- \left[e^{-r\tau_1}1\{Y > 1\}(\Gamma^{-1}(Y;\theta_1,\theta_2) - K)^+\right.$$

$$+ e^{-rT}1\{Y \leq 1\}(h_2(Z_2; \Gamma^{-1}(Y;\theta_1,\theta_2) - D, \tau_2, r, \sigma) - K)^+]\frac{\partial(rT)}{\partial \theta}. \qquad (4.23)$$

4.4.3.5 The American Call Option with Multiple Ex-Dividend Dates

We now extend the results in Section 4.4.3 to the cases where there are multiple ex-dividend dates. To simplify the problem, we denote $d_j = \sum_{i=j}^{\eta(T)} D_i \exp\Big(-r\sum_{k=j+1}^{i}\tau_k\Big)$, $j = 1,\ldots,\eta(T)$. Since the support of the distribution of the random variables $S_{t_i^-}$, $i = 1, \cdots, \eta(T)$ depends on the parameters of interest and the indicator function also contains parameters, we do the following change of random variables to move the parameters out of the indicator functions and make the support independent of the parameters of interests. Define $Y_1 = g_1(S_{t_1^-}) = \frac{S_{t_1^-} - d_1}{s_1 - d_1}, \ldots Y_{\eta(T)} = g_{\eta(T)}(S_{t_{\eta(T)}^-}) = \frac{S_{t_{\eta(T)}^-} - d_{\eta(T)}}{s_{\eta(T)} - d_{\eta(T)}}, Y_{\eta(T)+1} = g_{\eta(T)+1}(S_T) = S_T$. Therefore we have

$$
\begin{aligned}
\frac{\partial E[\hat{J}_T]}{\partial \theta} =& \frac{\partial}{\partial \theta} E\Big[\sum_{i=1}^{\eta(T)} \Big(\prod_{j=1}^{i-1} 1\{S_{t_j^-} \leq s_j\} 1\{S_{t_i^-} > s_i\}(S_{t_i^-} - K)^+ e^{r(T-t_i)}\Big) \\
&+ \prod_{j=1}^{\eta(T)} 1\{S_{t_j^-} \leq s_j\}(S_T - K)^+ \Big] \\
=& \frac{\partial}{\partial \theta} \Big\{ \sum_{i=1}^{\eta(T)} \Big(\int_0^\infty \prod_{j=1}^{i-1} 1\{y_j \leq 1\} 1\{y_i > 1\}(y_i(s_i - d_i) + d_i)^+ e^{r(T-t_i)} \\
&\times f_1(y_1) f_{2;1}(y_2; y_1) \cdots f_{i;i-1}(y_i; y_{i-1}) dy_1 \cdots dy_i \Big) \\
&+ \int_0^\infty \prod_{j=1}^{\eta(T)} 1\{y_j \leq 1\}(y_{\eta(T)+1} - K)^+ f_1(y_1) f_{2;1}(y_2; y_1) \cdots \\
&\times\ f_{\eta(T)+1;\eta(T)}(y_{\eta(T)+1}; y_{\eta(T)}) \\
&\times dy_1 \cdots dy_{\eta(T)+1}\Big\},
\end{aligned}
$$

where f_1 is the probability density function of the random variable Y_1, and $f_{i;i-1}$ is the probability density function of the random variable Y_i given Y_{i-1}, which can be easily obtained. Assuming interchangeability of integral and derivative, we have $\frac{\partial E[\hat{J}_T]}{\partial \theta} =$

$\frac{\partial E[\hat{J}_T^1]}{\partial \theta} + \frac{\partial E[\hat{J}_T^2]}{\partial \theta}$, where

$$\frac{\partial E[\hat{J}_T^1]}{\partial \theta} = \sum_{i=1}^{\eta(T)} \Big(\int_0^\infty \prod_{j=1}^{i-1} 1\{y_j \le 1\} 1\{y_i > 1\} \frac{\partial (y_i(s_i - d_i) + d_i)^+ e^{r(T-t_i)}}{\partial \theta}$$

$$\times f_1(y_1) \cdots f_{i;i-1}(y_i; y_{i-1}) dy_1 \cdots dy_i \Big)$$

$$+ \sum_{i=1}^{\eta(T)} \Big(\int_0^\infty \prod_{j=1}^{i-1} 1\{y_j \le 1\} 1\{y_i > 1\} (y_i(s_i - d_i) + d_i)^+ e^{r(T-t_i)}$$

$$\times \Big(\sum_{k=1}^{i} \frac{\partial \ln f_{i;i-1}(y_i; y_{i-1})}{\partial \theta} \Big) f_1(y_1) f_{2;1}(y_2; y_1) \cdots$$

$$\times f_{i;i-1}(y_i; y_{i-1}) dy_1 \cdots dy_i \Big),$$

$$\frac{\partial E[\hat{J}_T^2]}{\partial \theta} = \int_0^\infty \prod_{j=1}^{\eta(T)} 1\{y_j \le 1\} \frac{\partial (y_{\eta(T)+1} - K)^+}{\partial \theta}$$

$$\times f_1(y_1) f_{2;1}(y_2; y_1) \cdots f_{\eta(T)+1;\eta(T)}(y_{\eta(T)+1}; y_{\eta(T)}) dy_1 \cdots dy_{\eta(T)+1}$$

$$+ \int_0^\infty \prod_{j=1}^{\eta(T)} 1\{y_j \le 1\} (y_{\eta(T)+1} - K)^+ \Big(\sum_{i=1}^{\eta(T)+1} \frac{\partial \ln f_{i;i-1}(y_i; y_{i-1})}{\partial \theta} \Big)$$

$$\times f_1(y_1) f_{2;1}(y_2; y_1) \cdots f_{\eta(T)+1;\eta(T)}(y_{\eta(T)+1}; y_{\eta(T)}) dy_1 \cdots dy_{\eta(T)+1}.$$

The SLRIPA estimator for $\frac{\partial E[J_T]}{\partial \theta}$ is given by

$$\sum_{i=1}^{\eta(T)} \Big(e^{-rt_i} \prod_{j=1}^{i-1} 1\{Y_j \le 1\} 1\{Y_i > 1\} \frac{\partial (Y_i(s_i - d_i) + d_i)^+}{\partial \theta} \Big)$$

$$+ \sum_{i=1}^{\eta(T)} \Big(e^{-rt_i} \prod_{j=1}^{i-1} 1\{Y_j \le 1\} 1\{Y_i > 1\} (Y_i(s_i - d_i) + d_i)^+ \Big(\sum_{k=1}^{i} \frac{\partial \ln f_{i;i-1}(Y_i; Y_{i-1})}{\partial \theta} \Big) \Big)$$

$$+ e^{-rT} \prod_{j=1}^{\eta(T)} 1\{Y_j \le 1\} \frac{\partial (Y_{\eta(T)+1} - K)^+}{\partial \theta}$$

$$+ e^{-rT} \prod_{j=1}^{\eta(T)} 1\{Y_j \le 1\} (Y_{\eta(T)+1} - K)^+ \Big(\sum_{i=1}^{\eta(T)+1} \frac{\partial \ln f_{i;i-1}(Y_i; Y_{i-1})}{\partial \theta} \Big) - \Phi \qquad (4.24)$$

where

$$\Phi = -\Big(\sum_{i=1}^{\eta(T)} e^{-rt_i} \prod_{j=1}^{i-1} 1\{Y_j \leq 1\} 1\{Y_i > 1\}(Y_i(s_i - d_i) - K)^+$$

$$+ e^{-rT} \prod_{j=1}^{\eta(T)} 1\{Y_j \leq 1\}(Y_{\eta(T)+1} - K)^+ \Big) \frac{\partial(rT)}{\partial \theta}.$$

Remark 4.8 *Although the estimator (4.24) is written in terms of $\{Y_i\}$, it can be rewritten in terms of $\{S_{t_i^-}\}$ by applying $Y_i = g_i(S_{t_i^-})$.*

4.4.4 Discussion

As shown by examples in this section, SLRIPA is generally applicable for applications that have payoff functions containing indicator functions and is especially useful in the setting where the payoff function is discontinuous, so IPA cannot be applied and other methods usually require additional simulation. In practical applications, for payoff functions not exactly of the form $L(X(\omega; \theta_1), z; \theta_2)1\{X(\omega; \theta_1) > \theta_2\}$, SLRIPA may still be applicable, for example, for problems with multiple indicator functions as in the barrier option and American call option examples. The idea is to make the payoff function continuous with respect to parameters of interest through changes of variables and make the support of new random variables independent of any parameter of interest as well. In order to apply SLRIPA, some technical conditions are required. For instance, the part of the payoff function not containing indicator functions should satisfy some continuity conditions, and the probability density function of the random variable inside the indicator function can be obtained. These technical requirements are generally not a problem for financial applications. However for problems with random variables that have more complicated forms and cannot be written in simple functions of random variables with

known probability density functions, it might be more difficult to apply SLRIPA. One of the key advantages of SLRIPA is that it is easy to implement and only requires a single run of simulation to obtain one estimate, i.e., no resimulation is needed.

4.5   Simulation Results

In this section, we give some numerical results when applying the SLRIPA estimator to sensitivity analysis of barrier options and American option pricing problems.

4.5.1   European Barrier Option

Consider the up-and-out European barrier option in Section 4.4.2, where the stock price $S_t = S_0 e^{(r-\sigma^2/2)t+\sigma B_t}$. We let $\sigma = 0.1, r = 0.05, T = 1, H = 110, S_0 = K = 100$, with m the number of discretized points, and $h = T/m$ the time step. For the SLRIPA derivative estimators $\mathcal{D}_\theta$ given in Section 4.4.2, we give numerical results for $\mathcal{D}_H$, as this SLRIPA estimator is new, and there is no LR or IPA derivative estimator for the price of this barrier option with respect to H. Using the approach in [105], we can also derive the following SPA estimator:

$$\sum_{p=1}^{m} \left\{ e^{-rT} \prod_{i=1}^{p} 1\{S_i \leq H\} E\left[(S_m - K)1\{S_m \geq K\} \prod_{i=p}^{m} 1\{S_i \leq H\} | S_p = H \right] \frac{1}{H\sigma\sqrt{h}} n(\phi) \right\},$$

where $\phi = \frac{\ln H - \ln S_{p-1} - (r-\sigma^2/2)h}{\sigma\sqrt{h}}$.

We compare the performance of the proposed SLRIPA derivative estimator with that of the above SPA estimator and a finite-difference estimator. For the finite difference estimator with respect to H, we perturb H to $H + h$ and divide the difference of the payoff function at H and $H + h$ by h, where h is a small positive number; common

Table 4.1:   Sensitivity of Barrier Option: $\partial J_T/\partial H$ (Standard Error in Parentheses)

| | $r = 0.05, \sigma = 0.10$ | | |
	$m = 10$	$m = 20$	$m = 30$
FD	0.329(0.084)	0.310(0.077)	0.271(0.071)
SPA	0.286(0.011)	0.262(0.013)	0.254(0.015)
SLRIPA	0.276(0.018)	0.251(0.024)	0.266(0.028)

random numbers are used. In our simulation setting, we set $h = 0.2$ to compute the finite difference estimator. For the SPA estimator, we generate 10 sample paths to estimate each conditional expectation term in the SPA estimators separately. We performed $2,000$ independent replications for each estimator, and the results are summarized in Table 4.1 with the standard error in the parentheses. From the simulation results, we can see that *SLRIPA has comparable performance with SPA, with a slightly larger standard error. However, SLRIPA only needs to generate a single sample path to obtain one estimate, whereas SPA needs to estimate m conditional expectations, thus requiring a substantially higher amount of computation.* It is not clear whether SPA performs better than the SLRIPA method in terms of variance if the variances introduced by the conditional expectations in SPA are considered, but there are many ways in which this could be done, so this is an interesting topic for future research. In any case, our estimator requires less computation, as no additional simulation is required.

4.5.2   American Call Option

4.5.2.1   Sensitivity Analysis

We consider the sensitivity estimation of an American option with a single ex-dividend date and compare the results of the SLRIPA estimators, finite difference esti-

mators, and the estimators given in [36]. Two different values for D, K, σ are tested, and parameter values are $S_0 = 40$; $r = 0.05$; $D = 0.5$ $(D = 1.0)$; $K = 40$ $(K = 45)$, $\sigma = 0.10$ $(\sigma = 0.30)$; $t_1 = 4/12$; $T = 6/12$.

For the finite difference estimator with respect to θ, we perturb θ to $\theta + h$ and divide the difference of the payoff function at θ and $\theta + h$ by h, where h is a small positive number; common random numbers are used. The SPA estimators in [36] require additional simulations for the conditional expectation term for various parameters especially when there are multiple ex-dividend dates, and we do a separate estimation to estimate the conditional expectation term. In our simulation setting, we set $h = 0.1$ to compute finite difference estimators with respect to parameters S_0, s, and $h = 0.01$ with respect to σ, r, D and K. We generated 10 sample paths to estimate conditional expectation terms in the SPA estimators separately. We performed $2,000$ independent replications for each estimator. From the simulation results in Table 4.2, we can see that SLRIPA generally performs better than the finite difference method. *Compared with SPA, SLRIPA has comparative performance in terms of variance, but with much less computation. Again, a single sample path is needed to obtain one estimate for all the parameters of interest.*

4.5.2.2 Optimal Threshold Policy

With an estimate of the derivative with respect to the early exercise threshold parameters, we can solve the American option pricing problem as an optimization problem by applying stochastic approximation (SA) ([68]). The classical SA algorithm solves the following optimization problem by mimicking the gradient ascent method: $\max_\theta E[J_T(\theta)]$. In our settings, J_T is the return of the American option as a function of θ, which is the early exercise threshold parameter s for the single ex-dividend case and a vector when

Table 4.2: Sensitivity of American Option (Standard Error in Parentheses)

| | $r = 0.05, \sigma = 0.30$ | | |
	$D = 0.5, s = K = 40,$	$D = 0.5, s = K = 45,$	$D = 1.0, s = K = 40$
$\partial J_T/\partial K$			
FD	-0.584(0.011)	-0.342(0.010)	-0.581(0.011)
SPA	-0.584(0.011)	-0.342(0.010)	-0.581(0.011)
SLRIPA	-0.584(0.011)	-0.342(0.010)	-0.581(0.011)
$\partial J_T/\partial S_0$			
FD	0.489(0.093)	0.288(0.079)	0.562(0.054)
SPA	0.557(0.013)	0.316(0.014)	0.564(0.013)
SLRIPA	0.559(0.031)	0.320(0.021)	0.553(0.033)
$\partial J_T/\partial s$			
FD	0.169(0.070)	0.048(0.034)	0.089(0.048)
SPA	0.113(0.004)	0.086(0.003)	0.102(0.004)
SLRIPA	0.111(0.031)	0.094(0.016)	0.112(0.029)
$\partial J_T/\partial \sigma$			
FD	9.544(0.373)	9.058(0.872)	9.703(0.446)
SPA	10.031(0.331)	8.960(0.363)	9.793(0.326)
SLRIPA	9.972(1.334)	8.730(0.910)	10.592(1.679)
$\partial J_T/\partial r$			
FD	5.528(1.078)	4.205(0.819)	6.920(0.451)
SPA	6.836(0.167)	4.395(0.191)	7.000(0.169)
SLRIPA	6.839(0.374)	4.422(0.259)	6.836(0.399)
$\partial J_T/\partial D$			
FD	-0.156(0.007)	-0.170(0.008)	-0.161(0.007)
SPA	-0.158(0.007)	-0.159(0.008)	-0.162(0.007)
SLRIPA	-0.158(0.007)	-0.159(0.007)	-0.162(0.007)

there are multiple ex-dividend points.

The optimal parameter θ^* determines the optimal exercise threshold policy that defines the value of the American option. Let $\hat{g}(\theta)$ be an estimator of the gradient $\frac{\partial E[J_T(\theta)]}{\partial \theta}$; the SA algorithm generates iterates by the formula: $\theta_{k+1} = \Pi_\Theta(\theta_k + a_k \hat{g}(\theta_k))$, where θ_k is the value of the parameter at the beginning of iteration k, a_k is a sequence of positive step sizes, and Π_Θ is a projection onto the parameter set Θ. There are various conditions on the sequence of step sizes that ensure the convergence of the algorithm ([68]). In our

simulation experiments, we adopt the harmonic series $a_n = a/n$, which is also used in [36].

Table 4.3: Optimal Threshold and Corresponding Expected Payoff (Standard Error Based on 10 Runs)

| | $r = 0.05, \sigma = 0.10$ | | | |
	$D = 0.5$ $K = 40$	$D = 0.5$ $K = 45$	$D = 1.0$ $K = 40$	$D = 1.0$ $K = 45$
$s^\star$	41.69	47.19	40.62	45.84
$\tilde{s}^\star$	41.77(0.03)	45.92(0.01)	40.61(0.02)	45.59(0.01)
$J_T^\star$	1.411	0.089	1.323	0.063
$\tilde{J}_T^\star$	1.410	0.087	1.323	0.062
	$r = 0.05, \sigma = 0.30$			
	$D = 0.5$ $K = 40$	$D = 0.5$ $K = 45$	$D = 1.0$ $K = 40$	$D = 1.0$ $K = 45$
$s^\star$	47.97	55.08	44.48	50.56
$\tilde{s}^\star$	47.48(0.14)	52.01(0.08)	44.85(0.08)	50.41(0.06)
$J_T^\star$	3.586	1.756	3.414	1.618
$\tilde{J}_T^\star$	3.586	1.753	3.413	1.618

Given the derivative estimator with respect to the threshold parameter s, we use the SA algorithm to compute the optimal threshold policy and obtain estimates of the American option price. In the SA algorithm, we set $a = 20$ for the case of parameter values $r = 0.05, \sigma = 0.1$, and $a = 60$ for $r = 0.05, \sigma = 0.3$. The algorithm is started with $s = K$ and is stopped if the difference between two consecutive iterative values of s is less than $\epsilon = 5 \times 10^{-5}$. 2000 sample paths are generated to estimate the gradient for each iteration of the SA algorithm. We performed 10 independent replications. Once we have the threshold parameter s, the expected payoff can be computed analytically. In Table 4.3, $s^\star$ is the optimal threshold policy computed by Roll-Geske-Whaley formula [98] and $J_T^\star$ is the corresponding option price, $\tilde{s}^\star$ is the threshold policy obtained by SA, and $\tilde{J}_T^\star$ is the corresponding expected payoff. From Table 4.3, we can see that $\tilde{s}^\star$ is very close to $s^\star$, and the expected payoff with the threshold parameter obtained by SA agrees with the true

option price within a penny. Figure 1 shows a typical sample path of the rapidly converging stochastic approximation algorithm with parameters $r = 0.05, \sigma = 0.10, K = 40, D = 1.0,$ and $S_0 = 40$.
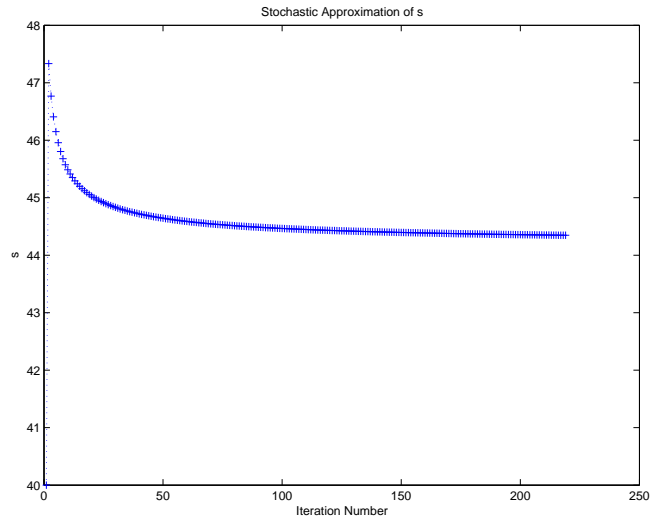

Figure 4.1: Convergence of SA with $r = 0.05, \sigma = 0.10, K = 40, D = 1.0,$ and $S_0 = 40$

4.6   Conclusions

We have presented a new stochastic derivative estimation approach for discontinuous payoff functions of a special form, for which IPA and LR may give biased estimates. Although unbiased estimators for these types of functions have been derived using other techniques such as SPA, the resulting estimators require additional simulations. By combining a change of variables with both IPA and LR, we show how to derive unbiased estimators that can be computed using only a single simulation replication. Simulation experiments illustrate the effectiveness of the proposed method.

SLRIPA uses changes of variables to smooth out discontinuities, whereas SPA uses conditional expectations. We note that generally SLRIPA has two parts: an LR part

and an IPA part, while SPA is composed of an IPA part and a conditional expectation part. An interesting future research topic is to explore the connection between SPA and SLRIPA. Since LR works in a general Markov chain setting, extending SLRIPA to this general setting and studying the properties of the resulting SLRIPA estimators is another natural future research direction.

Chapter 5

Model-based Evolutionary Optimization

5.1 Introduction and Motivation

In this chapter, we focus on developing a novel simulation optimization framework based on evolutionary game theory for solving the following global optimization problem

$$y^{\star} \in \arg\max_{y \in \mathcal{Y}} H(y),$$

where the solution space $\mathcal{Y} \subset \mathbb{R}^n$ is a nonempty set.

As in Chapter 1 and Chapter 2, according to the criteria in [110], we classify global optimization algorithms as *instance-based*, which include simulated annealing ([63]), genetic algorithms ([97]), tabu search ([47]), and nested partitions ([93]), and *model-based* methods, which includes estimation of distribution algorithms (EDAs) ([80]), the cross-entropy (CE) method ([15, 75]), model reference adaptive search (MRAS) ([59]), and particle filtering-based method ([109]).

Since the emerging of genetic algorithms, many instance-based algorithms have been developed and well studied. On the other hand, the age of model-based algorithms is relatively young and the convergence behavior and performance of model-based algorithms are not well understood. Further exploration of the properties of model-based algorithms and developing new computational efficient algorithms are the principal purpose of the research presented in this Chapter.

EDAs, CE, MRAS, and algorithms based on particle filtering are different from each other in the way of interpreting and updating the probabilistic models. In these *model-based* global optimization algorithms, candidate solutions are generated from probabilistic models, which are updated by these candidate solutions in such a way that better solutions will have a higher chance to be sampled at the next iteration. Similarly, in evolutionary games, the dynamics are such that better strategies will spread in the population. This similarity motivates us to connect global optimization problems with evolutionary games. The dynamics that are used to study the evolution of strategies in evolutionary games provides us a powerful tool to investigate the model updating in model-based algorithms.

Game theory studies the strategic interaction of players using different strategies; it has been applied in many areas such as economics, engineering, and biology ([38, 95]). Recently, [71] and [41] have applied game theory to solve discrete optimization problems, where they model the optimization problem as a potential game. Fictitious play and joint fictitious play are adopted to obtain the Nash equilibrium, and two sampled version of fictitious and joint fictitious play are also proposed in [71] and [41]. For a potential game, although fictitious play has been proven to converge to a mixed strategy Nash equilibrium, the mixed strategy equilibrium might not be a feasible solution for the optimization problem. The algorithms in [41] and [71] only work for discrete optimization problems with a finite solution space, and moreover the Nash equilibrium obtained by fictitious play might only be a locally optimal solution.

Evolutionary game theory applies game theory to study the evolution of the number of players playing different strategies in a population setting. After being introduced by the biologist [77], evolutionary game theory has become popular in biology and has attracted increasing interest from researchers in other areas. As opposed to static games,

135

evolutionary games involve a population of players and study the evolution of players by using replicator dynamics, which sets the growth rate of the proportion of players using a certain strategy equal to the difference between the average payoff of that strategy and the average payoff of the entire population. Replicator dynamics can also be used as a learning algorithm to study the behavior of multiple agents ([103]).

The main idea of our method is to formulate the global optimization problem as an evolutionary game and to use dynamics from evolutionary game theory to study the evolution of the candidate solutions. Searching for the optimal solution is carried out through the dynamics of reaching equilibrium points in evolutionary games. Specifically, we establish a connection between evolutionary game theory and optimization by partitioning the solution region of a global optimization problem and letting different players play strategies in different subsets. Differential dynamics such as replicator dynamics are used to govern the evolution of the candidate solutions for the optimization problem. Furthermore, we introduce probabilistic models to generate candidate solutions and formulate the global optimization problem as an evolutionary game with continuous strategy spaces. We show that there is a strong connection between a particular equilibrium set of the replicator dynamics and the global optimal solutions. By using Lyapunov theory, we also show that the particular equilibrium set is asymptotically stable under mild conditions. Based on the connection between the equilibrium points and global optimal solutions, a Model-based Evolutionary Optimization (MEO) algorithm is developed. Moreover, to better capture the multimodal property of global optimization problems, we propose to use a population of models to generate candidate solutions and a new Population Model-based Evolutionary Optimization (PMEO) algorithm is proposed, in which evolutionary game theory is used to study the evolution of those individual models, and models with

best performances will survive eventually. Motivated by the idea of mutation strategies in evolutionary games, we introduce mutation strategies into the framework of PMEO to further improve the algorithm's ability for escaping trap of low quality solutions. In preliminary numerical experiments, PMEO is able to find good solutions.

The way we formulate global optimization problems as evolutionary games provides new insights into the mechanism for generating new candidate solutions and the mechanism of model updating for model-based global optimization algorithms. For example, one special case of the MEO algorithm leads to a new interpretation of the CE method. This evolutionary game setting for global optimization problems makes it possible to study the convergence property of model-based algorithms by using analytical tools in the evolutionary game theory literature and it also provides possibilities for developing new algorithms, such as the PMEO algorithm developed in this Chapter.

5.2 Connecting Optimization and Evolutionary Game Theory

Consider the following optimization problem:

$$y^\star \in \arg\max_{y \in \mathcal{Y}} H(y), \tag{5.1}$$

where the solution space $\mathcal{Y} \subseteq \mathbb{R}^n$ is a nonempty set. The objective function $H : \mathcal{Y} \to \mathbb{R}$ is a deterministic function that is continuous almost everywhere. Assume the set of global optimal solutions is nonempty and finite. $y^\star$ is a global optimal solution if $H(y^\star) \geq H(y) \ \forall y \neq y^\star, \ y \in \mathcal{Y}$.

Assume that the solution space $\mathcal{Y}$ can be partitioned into M disjoint subsets $\mathcal{G}_1, ..., \mathcal{G}_M$. At each iteration, generate $N_1, \ldots, N_M$ candidate solutions using some random sampling

algorithm in the corresponding subsets $\mathcal{G}_1, ..., \mathcal{G}_M$, and the total number of solutions generated is a fixed number: $\sum_{i=1}^{M} N_i = N$. Our goal is to sample more around an optimal solution $y^\star$, and thus increase the chance of finding an optimal solution. From an evolutionary game theory perspective, we view the N samples (candidate solutions) as agents, who are programmed to play M different pure strategies $\{1, \ldots, M\}$. By playing the pure strategy i, we mean sampling a candidate solution in the subset $\mathcal{G}_i$. Hence there are N_i agents playing the pure strategy i. An agent playing a pure strategy i will receive a payoff f_i given by

$$f_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \varphi(H(y_{ij})), \quad \forall i = 1, \ldots, M, \tag{5.2}$$

where $\varphi : \mathbb{R} \to \mathbb{R}^+$ is a strictly increasing function and $\{y_{ij}, j = 1, \ldots, N_i\}$ are the candidate solutions generated from the subset $\mathcal{G}_i$. Note that the payoff f_i only depends on the actions of agents playing strategy i. Define $x_i = \frac{N_i}{N}$ as the percentage of agents playing the pure strategy i for all $i \in \{1, \ldots, M\}$ and we have $\sum_{i=1}^{M} x_i = 1$. According to the analysis in Section 2.2.2, the evolution of the number of agents playing different strategies in this evolutionary game is governed by the replicator dynamics

$$\dot{x}_i = x_i(f_i - \sum_{j=1}^{M} x_j f_j), \quad \forall i = 1, \ldots, M. \tag{5.3}$$

From (5.3), it is easy to see that if the payoff of strategy i is greater than the average payoff, i.e., $f_i > \sum_{j=1}^{M} x_j f_j$, the proportion of agents playing i will increase. From the viewpoint of simulation-based optimization, more samples will be drawn from the more promising area - the subset $\mathcal{G}_i$. It is easy to check equation (5.3) preserves $\sum_{i=1}^{M} x_i = 1$.

Note that from equation (5.2), we can see f_i only depends on the actions of agents playing strategy i, which is independent of $x_j, j \neq i$. In other words, the payoff received

by agents playing i is irrelevant to the actions of agents playing other different strategies. In this evolutionary game setting, agents can be viewed as competing against nature [106]. The function f_i is random; it is a Monte Carlo estimate of the mean of $\varphi(H(\cdot))$ in set $\mathcal{G}$.

Note that the replicator dynamics (5.3) is a differential equation. Since our optimization algorithm is simulation-based, we need a discretized version of replicator dynamics. In matrix games introduced in Section 2.2.2, the discrete replicator dynamics is given by

$$x_i(t+1) = x_i(t) \frac{(Ax(t))_i + c}{x(t)^T Ax(t) + c},$$

where c is some constant to make sure that the denominator is not zero. The discretized replicator dynamics corresponding to (5.3) is

$$x_i(t+1) = x_i(t) \frac{f_i^t + c}{\sum_{j=1}^{M} x_j(t) f_j^t + c}, \quad \forall i = 1, \ldots, M, \tag{5.4}$$

where f_j^t is the payoff received by an agent playing a strategy j at time t. We can see from (5.4) that the percentage $\{x_i(t+1)\}$ of agents playing each strategy changes at each iteration by a factor proportional to the average payoff of the corresponding strategy. The percentage of agents playing a strategy increases only if its payoff is greater than the average payoff, and the amount of increase depends on the difference of the payoff of the particular strategy and the average payoff.

The discrete replicator dynamics (5.4) governs the evolution of percentages of samples in different regions. By incorporating the ideas of the nested partitions method ([93]), promising regions can further be partitioned to perform a finer search, and non-promising regions can be combined. Then replicator dynamics can be applied again to concentrate most of the samples on the most promising region. How to partition the solution region

139

and to generate candidate solutions efficiently in the resulting subsets is critical to solving the global optimization problem and requires further investigation. We will not pursue the idea of incorporating nested partitions method further in this Chapter. Instead, we consider the problem from the perspective of evolutionary games with continuous strategies spaces.

5.3   Model-based Evolutionary Optimization

Now we consider an evolutionary game with a continuous strategy space. Let $\mathcal{B}$ be a Borel σ-algebra on $\mathcal{Y}$, the strategy space of the game; for each t, let $\mathbb{P}_t$ be a probability measure defined on $(\mathcal{Y}, \mathcal{B})$. Let Δ denote the set of all strategies (probability measures) on $\mathcal{Y}$. Let $\mathcal{M}^e(\mathcal{Y}, \mathcal{B})$ denote the linear span of Δ, the space of all finite and signed measures.

Every single point $y \in \mathcal{Y}$ can be viewed as a pure strategy. The fraction of agents playing the pure strategy y at time t is $\mathbb{P}_t(dy)$. An agent playing the pure strategy y obtains a fitness $\varphi(H(y))$. The fraction of agents adopting different strategies in the continuous game is described by the probability measure $\mathbb{P}_t$ defined on the strategy space $\mathcal{Y}$, so the average payoff of the whole population is given by

$$E_{\mathbb{P}_t}[\varphi(H(Y))] = \int_{\mathcal{Y}} \varphi(H(y))\mathbb{P}_t(dy).$$

In evolutionary game theory ([82]), the evolution of this probability measure is governed by some dynamics such as replicator dynamics. Let $\mathscr{A}$ be a measurable set in $\mathcal{Y}$. The replicator dynamics with a continuous strategy space is given by

$$\dot{\mathbb{P}}_t(\mathscr{A}) = \int_{\mathscr{A}} (\varphi(H(y)) - E_{\mathbb{P}_t}[\varphi(H(Y))])\mathbb{P}_t(dy). \tag{5.5}$$

From (5.5), we can see that if $\varphi(H(y))$ outperforms $E_{\mathbb{P}_t}[\varphi(H(Y))]$, the probability measure around y will increase. In this evolutionary game setting, the payoff $\varphi(H(y))$ depends only on y, and not on the other strategies. Similar to the analysis in Section 5.2, in this game, agents are competing against nature. Equation (5.5) is inspired by extending (5.3) to a continuous setting.

Since the probability measure doesn't have a specific structure, it would be very difficult to use (5.5) directly. If there exists a probability density function p_t, such that $\mathbb{P}_t(dy) = p_t \mu(dy)$, where $\mu(\cdot)$ is the Lebesgue measure defined on $(\mathcal{Y}, \mathcal{B})$, then (5.5) becomes

$$\dot{p}_t(y) = (\varphi(H(y)) - E_{\mathbb{P}_t}[\varphi(H(Y))])p_t(y), \tag{5.6}$$

which governs the evolution of the probability density function on the continuous strategy space. When $p_t(y)$ is used as our model to generate candidate solutions for the global optimization problem (5.1), the differential equation (5.6) can be used to update the model $p_t(y)$, with the final goal of making the probability density function $p_t(y)$ concentrated on a small set containing the global optimal solutions. Then the global optimization problem can be easily solved by sampling from the obtained probability density function.

Remark 5.1 *What is interesting about equation (5.6) is that it is a nonlinear ordinary differential equation, indexed by y, which is different from the linear partial (Fokker-Planck) differential equation that governs the probability density evolution of a Markov process. This nonlinearity also makes the convergence analysis that will be given below more difficult. The random process it describes is known as a nonlinear Markov process [65].*

5.3.1 Analysis of the Replicator Dynamics with a Continuous Strategy Space

In this section, we study the properties of the equilibrium points of (5.5) and their connection with the global optimal solutions for the optimization problem by exploring the tools of equilibrium analysis in the literature of game theory and stability analysis in dynamic systems.

5.3.1.1 Convergence Analysis of Equilibrium Set

We make the following assumption to conduct the convergence analysis.

Assumption 5.1 *There exist constants $\mathscr{L}$ and $\mathscr{M}$ such that $\mathscr{L} \leq \varphi(H(y)) \leq \mathscr{M}$ for all $y \in \mathcal{Y}$.*

Assumption 5.1 is needed to ensure that the average fitness function $E_{\mathbb{P}_t}[\varphi(H(Y))]$ is well defined. Functions that are lower bounded by $-\infty$ can be truncated by a constant $\mathscr{L}$, which does not change the solution of the optimization problem since maximization problems are considered here.

We first show that the replicator dynamics (5.5) is well defined.

Theorem 5.1 *If Assumption 5.1 holds, for each $\mathbb{P}_0$ defined on $(\mathcal{Y}, \mathcal{B})$, the ordinary differential equation (5.5) has a unique solution for $t \in [0, \infty]$.*

Proof: Our proof follows the method proposed by [81] and [54]. Let $\| \cdot \|$ denote the variational norm on $(\mathcal{M}^e, \mathcal{B})$, which is defined as

$$\|\mathbb{P}\| = \sup_g \Big| \int_{\mathcal{Y}} g(y)\mathbb{P}(dy) \Big|,$$

where the sup is taken over all measurable functions $g : \mathcal{Y} \to \mathbb{R}$ and $\sup_{y \in \mathcal{Y}} |g(y)| \leq 1$. For simplicity, let

$$J(\mathbb{P}_t) = \int_{\mathscr{A}} (\varphi(H(y)) - E_{\mathbb{P}_t}[\varphi(H(Y))])\mathbb{P}_t(dy).$$

By Assumption 5.1 that $\mathscr{L} \leq \varphi(H(y)) \leq \mathscr{M}, \forall y \in \mathcal{Y}$, we have

$$\|J(\mathbb{P})\| \leq 2 \max(|\mathscr{L}|, |\mathscr{M}|).$$

Hence $J(\mathbb{P})$ is bounded $\forall \mathbb{P} \in \Delta$. We now show that J is Lipschitz continuous. Let $\mathbb{P}$ and $\mathbb{Q}$ be two different probability measures in Δ; then

$$\|J(\mathbb{P}) - J(\mathbb{Q})\|$$

$$= \| \int_{\mathscr{A}} \varphi(H(y))\mathbb{P}(dy) - E_{\mathbb{P}}[\varphi(H(Y))]\mathbb{P} - (\int_{\mathscr{A}} \varphi(H(y))\mathbb{Q}(dy) - E_{\mathbb{Q}}[\varphi(H(Y))\mathbb{Q}])\|$$

$$\leq \|E_{\mathbb{P}}[\varphi(H(Y))]\mathbb{P} - E_{\mathbb{P}}[\varphi(H(Y))]\mathbb{Q} + E_{\mathbb{P}}[\varphi(H(Y))]\mathbb{Q} - E_{\mathbb{Q}}[\varphi(H(Y))]\mathbb{Q}\|$$

$$+ \int_{\mathscr{A}} |\varphi(H(y))|\|\mathbb{P} - \mathbb{Q}\|(dy)$$

$$\leq |E_{\mathbb{P}}[\varphi(H(Y))]|\|\mathbb{P} - \mathbb{Q}\| + \int_{\mathcal{Y}} |\varphi(H(y))|\|\mathbb{P} - \mathbb{Q}\|(dy) + \int_{\mathscr{A}} |\varphi(H(y))|\|\mathbb{P} - \mathbb{Q}\|(dy)$$

$$\leq 3 \max(|\mathscr{L}|, |\mathcal{M}|)\|\mathbb{P} - \mathbb{Q}\|.$$

Therefore, J is Lipschitz continuous on the set of probability measures with variational norm. By Corollary 3.9 of [108], the ordinary differential equation $\dot{\mathbb{P}}_t(\mathscr{A}) = J(\mathbb{P}_t)$ with an initial measure $\mathbb{P}_0 \in \Delta$ has a unique solution $\mathbb{P}_t$. $\blacksquare$

Assume that $\mathbb{P}^\star$ is an equilibrium distribution for the replicator dynamics (5.5), and thus we have $J(\mathbb{P}^\star) = 0$. It is easy to see that $\mathbb{P}^\star = \delta_{y_i^\star}$ for $i = 1, \ldots, m$ are equilibrium

points of (5.5), and we might guess there is a strong connection between the equilibrium points of (5.5) and the optimal solutions of the global optimization problem (5.1). To further study the properties of the equilibrium points of the replicator dynamics (5.5), the Prokhorov metric is used to measure the distance between different strategies (probability measures):

$$\rho(\mathbb{P}, \mathbb{Q}) := \inf\{\epsilon > 0 : \mathbb{Q}(\mathscr{A}) \leq \mathbb{P}(\mathscr{A}^\epsilon) + \epsilon \ \text{and} \ \mathbb{P}(\mathscr{A}) \leq \mathbb{Q}(\mathscr{A}^\epsilon) + \epsilon, \quad \forall \mathscr{A} \in \mathcal{B}\},$$

where $\mathscr{A}^\epsilon := \{x : \exists \tilde{y} \in \mathscr{A}, d(\tilde{y}, x) < \epsilon\}$, in which d is a metric defined on $\mathcal{Y}$. Then the convergence of $\rho(\mathbb{Q}_n, \mathbb{Q}) \to 0$ is equivalent to the weak convergence of $\mathbb{Q}_n$ to $\mathbb{Q}$ ([2]).

The following definition specifies the dynamic stability concepts we will be using for the infinite dimensional system (5.5).

Definition 5.1 *Let $\mathcal{E}$ be a set in Δ. For a point $\mathbb{P} \in \Delta$, define the distance between $\mathbb{P}$ and $\mathcal{E}$ as $\rho(\mathbb{P}, \mathcal{E}) = \inf\{\rho(\mathbb{P}, \mathbb{Q}), \forall \mathbb{Q} \in \mathcal{E}\}$. $\mathcal{E}$ is called Lyapunov stable if for all $\epsilon > 0$, there exists $\eta > 0$ such that $\rho(\mathbb{P}_0, \mathcal{E}) < \eta \implies \rho(\mathbb{P}_t, \mathcal{E}) < \epsilon$ for all $t > 0$.*

Definition 5.2 *Let $\mathcal{E}$ be a set in Δ. $\mathcal{E}$ is called asymptotically stable if $\mathcal{E}$ is Lyapunov stable and there exists $\eta > 0$ such that $\rho(\mathbb{P}_0, \mathcal{E}) < \eta \implies \rho(\mathbb{P}_t, \mathcal{E}) \to 0$ as $t \to \infty$.*

The following theorem shows that the overall fitness of the strategy (probability measure) $\mathbb{P}_t$ is monotonically increasing over time.

Theorem 5.2 *Let $\mathbb{P}_t$ be a solution of the replicator dynamics (5.5). If Assumption 5.1 holds, the average payoff of the entire population $E_{\mathbb{P}_t}[\varphi(H(Y))]$ is monotonically increasing with time t. If Assumption 5.1 holds and $\mathbb{P}_t$ is not an equilibrium point of (5.5), then $E_{\mathbb{P}_t}[\varphi(H(Y))]$ is strictly increasing with time t.*

144

Proof: Differentiate the average payoff:

$$\frac{d}{dt}E_{\mathbb{P}_t}[\varphi(H(Y))] = \int_{\mathcal{Y}} \varphi(H(y))\frac{d\mathbb{P}_t}{dt}(dy)$$

$$= \int_{\mathcal{Y}} \varphi(H(y))(\varphi(H(y)) - E_{\mathbb{P}_t}[\varphi(H(Y))])\mathbb{P}_t(dy)$$

$$= \left(E_{\mathbb{P}_t}[\varphi^2(H(Y))] - (E_{\mathbb{P}_t}[\varphi(H(Y))])^2\right) \geq 0.$$

The first equality above holds by the dominated convergence theorem, since $\varphi(H(\cdot))$ is upper and lower bounded by finite constants, and hence the derivative of $\mathbb{P}_t$ is bounded, using (5.5); the second equality above is obtained by applying (5.5); the last inequality holds since $E_{\mathbb{P}_t}[\varphi^2(H(Y))] - (E_{\mathbb{P}_t}[\varphi(H(Y))])^2$ is just the variance of $\varphi(H(\cdot))$ under the measure $\mathbb{P}_t$. Therefore the first claim is proved.

The second claim is proved by contradiction. Assume for some t that $\frac{d}{dt}E_{\mathbb{P}_t}[\varphi(H(Y))] = 0$. It is easy to see that when $\frac{d}{dt}E_{\mathbb{P}_t}[\varphi(H(Y))] = 0$, we have $\varphi(H(y)) = C$, where $C = E_{\mathbb{P}_t}[\varphi(H(Y))]$ is a constant, which is an equilibrium point of (5.5). This contradicts the fact that $\mathbb{P}_t$ is not an equilibrium point. Therefore, we must have $\frac{d}{dt}E_{\mathbb{P}_t}[\varphi(H(Y))] > 0$ when $\mathbb{P}_t$ is not an equilibrium point, and the theorem is proved. ∎

Before presenting the main theorem of the convergence analysis, we give the definition of a particular strategy set.

Definition 5.3 $\Delta_0 \subset \Delta$ *is the set containing all $\mathbb{P}_0$ for which there exists a $y_k^\star$ such that $\mathbb{P}_0(\tilde{\mathcal{A}}) > 0$ for any set $\tilde{\mathcal{A}} \in \mathcal{B}$ that contains $y_k^\star$ and has a positive Lebesgue measure $\mu(\tilde{\mathcal{A}}) > 0$. Let $\mathcal{C} = \{\mathbb{P}^\star : \mathbb{P}^\star = \lim_{t\to\infty} \mathbb{P}_t$ starting from some $\mathbb{P}_0 \in \Delta_0\}$.*

We also need the following assumptions.

Assumption 5.2 *There is a finite number of global optimal solutions $\{y_1^\star, \ldots, y_m^\star\}$ for*

the optimization problem (5.1), where m is a positive integer.

Assumption 5.3 $\varphi(\cdot)$ *is continuous and strictly increasing.*

Theorem 5.3 *If Assumptions 5.1, 5.2, and 5.3 hold, then the following hold:*

C1. *for any $\mathbb{P}^\star \in \mathcal{C}$, there exist $\alpha_i \geq 0$, for $i = 1, \ldots, m$ with $\sum_{i=1}^m \alpha_i = 1$ such that*

$\mathbb{P}^\star = \sum_{i=1}^m \alpha_i \delta_{y_i^\star};$

C2. *the set $\mathcal{C}$ can be represented as $\mathcal{C} = \{\mathbb{P}^\star : \mathbb{P}^\star = \sum_{i=1}^m \alpha_i \delta_{y_i^\star}, \text{ for some } \sum_{i=1}^m \alpha_i = 1, \alpha_i \geq 0, \forall i = 1, \ldots, m\}$; in addition, it is asymptotically stable.*

Proof: Let H^* denote the global optimal value of $\varphi(H(\cdot))$, i.e., $H^* = \max_{y \in \mathcal{Y}} \varphi(H(y))$ and we prove the claim C1 by contradiction. Assume there exists a $\tilde{\mathbb{P}}$, an equilibrium point in $\mathcal{C}$ that is not of the form $\tilde{\mathbb{P}} = \sum_{i=1}^m \alpha_i \delta_{y_i^\star}$. Combining with Assumption 5.2, we see that $\tilde{\mathbb{P}}$ does not put all the measure to global optimal solutions, and hence there exist positive constants ϵ_1 and ϵ_2, and a measurable set $\mathcal{A}_1 \in \mathcal{B}$ such that $\tilde{\mathbb{P}}(\mathcal{A}_1) > \epsilon_1$ and $\varphi(H(y)) < H^* - \epsilon_2$ for almost all $y \in \mathcal{A}_1$. Let $\epsilon = \epsilon_1 \epsilon_2$; then we have

$$E_{\tilde{\mathbb{P}}}[\varphi(H(Y))] \leq H^*(1 - \epsilon_1) + \epsilon_1(H^* - \epsilon_2) = H^* - \epsilon. \tag{5.7}$$

By the definition of $\mathcal{C}$ in Definition 5.3 and the fact $\tilde{\mathbb{P}} \in \mathcal{C}$, there exist a $\mathbb{P}_0 \in \Delta_0$ and a trajectory $\mathbb{P}_t$ starting from $\mathbb{P}_0$ such that $\mathbb{P}_t \to \tilde{\mathbb{P}}$ as $t \to \infty$. We have

$$E_{\mathbb{P}_t}[\varphi(H(Y))] \leq E_{\tilde{\mathbb{P}}}[\varphi(H(Y))] \leq H^* - \epsilon, \tag{5.8}$$

since $E_{\mathbb{P}_t}[\varphi(H(Y))]$ is monotonically increasing over time by Theorem 2, and by inequality (5.7).

Since $\mathbb{P}_0 \in \Delta_0$ and by Definition 5.3, there exist a point $y_k^\star \in \tilde{\mathcal{A}}$ and a measurable set $\tilde{\mathcal{A}}$ with $\mu(\tilde{\mathcal{A}}) > 0$ containing $y_k^\star$ such that $\mathbb{P}_0(\tilde{\mathcal{A}}) > 0$. Then by the continuity of $\varphi(\cdot)$ from Assumption 5.3 and the continuity of $H(\cdot)$ at $y_k^\star$, there exists a measurable set $\mathcal{S} \in \tilde{\mathcal{A}}$ with $\mu(\mathcal{S}) > 0$ such that $\mathbb{P}_0(S) > 0$, $y_k^\star \in \mathcal{S}$, and $\varphi(H(y)) > H^* - \epsilon/2$ for all $y \in \mathcal{S}$. Note that

$$\dot{\mathbb{P}}_t(\mathcal{S}) = \int_{\mathcal{S}} (\varphi(H(y)) - E_{\mathbb{P}_t}[\varphi(H(Y))]) \mathbb{P}_t(dy)$$

$$= \int_{\mathcal{S}} \varphi(H(y)) \mathbb{P}_t(dy) - E_{\mathbb{P}_t}[\varphi(H(Y))] \mathbb{P}_t(\mathcal{S})$$

$$\geq (H^* - \epsilon/2 - (H^* - \epsilon)) \mathbb{P}_t(\mathcal{S}) = \frac{\epsilon}{2} \mathbb{P}_t(\mathcal{S}), \tag{5.9}$$

where the third inequality holds by (5.8). Therefore we have

$$\frac{\dot{\mathbb{P}}_t}{\mathbb{P}_t}(\mathcal{S}) \geq \frac{\epsilon}{2} \qquad \forall t \geq 0.$$

By integrating both sides of the above inequality from 0 to t, we obtain $\mathbb{P}_t(\mathcal{S}) \geq e^{\frac{\epsilon}{2}t} \mathbb{P}_0(\mathcal{S})$ for all $t \geq 0$. Since $\mathbb{P}_0(\mathcal{S})$ is a positive constant, $\mathbb{P}_t(\mathcal{S})$ goes to infinity as t goes to infinity, which contradicts the fact that $\mathbb{P}^\star = \lim_{t \to \infty} \mathbb{P}_t$ is a probability measure in $\mathcal{C}$; this concludes the first part of the proof.

Now we prove the claim C2. First we prove that the set $\mathcal{C}$ is equal to the set $\mathcal{E} = \{\mathbb{P}^\star : \mathbb{P}^\star = \sum_{i=1}^m \alpha_i \delta_{y_i^\star}, \text{ for some } \sum_{i=1}^m \alpha_i = 1, \alpha_i \geq 0, \forall i = 1, \ldots, m\}$. By the claim C1, $\mathcal{C} \subseteq \mathcal{E}$. It is also easy to check that $\mathcal{E} \subseteq \Delta_0$, and then by the definition of $\mathcal{C}$, $\mathcal{E} \subseteq \mathcal{C}$. Hence $\mathcal{C} = \mathcal{E}$.

Now we prove $\mathcal{C}$ is asymptotically stable by using a generalized Lyapunov theory on the metric space $(\mathcal{M}^e, \rho)$. We first show that the set $\mathcal{C}$ is compact. To prove $\mathcal{C}$ is

compact in a metric space, we show that any sequence $\{\mathbb{P}_n^\star = \sum_{i=1}^m \alpha_i^n \delta_{y_i^\star}\}$ in $\mathcal{C}$ has a convergent subsequence that converges to a point in $\mathcal{C}$. Note that $(\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n) \in \mathbb{R}^m$ lives in a simplex for each n. Since a simplex in the Euclidean space is closed and hence is compact, there exists a subsequence $\{(\alpha_1^{n_1}, \alpha_2^{n_1}, \ldots, \alpha_m^{n_1})\}$ that converges to a point $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ in the simplex. Corresponding to this subsequence, there exists a subsequence of probability measures $\{\mathbb{P}_{n_1}^\star = \sum_{i=1}^m \alpha_i^{n_1} \delta_{y_i^\star}\}$ converging to a point $\mathbb{P}^\star = \sum_{i=1}^m \alpha_i \delta_{y_i^\star} \in \mathcal{C}$. Therefore $\mathcal{C}$ is compact.

Define a Lyapunov function $V(\mathbb{P}) = E_{\mathbb{P}^\star}[\varphi(H(Y))] - E_{\mathbb{P}}[\varphi(H(Y))]$, where $\mathbb{P}^\star$ is a point in $\mathcal{C}$; notice that $V(\mathbb{P}_t)$ is positive for all $\mathbb{P}_t \in \Delta \backslash \mathcal{C}$ and $V(\mathbb{P}_t) = 0$ for $\mathbb{P}_t \in \mathcal{C}$. From the proof of Theorem 5.2, we know that $\dot{V}(\mathbb{P}_t) < 0$ for all $t > 0$ if $\mathbb{P}_t$ is not in $\mathcal{C}$. From a generalization of Lyapunov's theorem (see Chapter V of [11]), the compact set $\mathcal{C}$ is asymptotically stable. ∎

Remark 5.2 *Chapter V of [11] presented a generalized Lyapunov's theorem on a general metric space. In the proof of Theorem 5.3, we applied this generalization of Lyapunov's theorem on the metric space $(\mathcal{M}^e, \rho)$.*

Remark 5.3 *The above analysis in this section is done under the assumption that $H(\cdot)$ is continuous almost everywhere. Similar results can be obtained when $H(\cdot)$ is not continuous almost everywhere under some more general assumptions by extending the above analysis.*

Remark 5.4 *Theorems 5.2 and 5.3 build a connection between the optimal solutions of the global optimization problem (5.1) and a particular equilibrium set of the corresponding evolutionary game. By Theorem 5.3, global optimal solutions can be approached by following a trajectory of the replicator dynamics (5.5) starting from a point in Δ_0.*

Remark 5.5 *The construction of the initial probability measure $\mathbb{P}_0$ is critical for designing algorithms to obtain equilibrium points. Assumption 5.2 ensures that a small set around an optimal solution can be sampled with a positive probability. One possible choice of $\mathbb{P}_0$ is a probability measure with a continuous probability density function and support on all of $\mathcal{Y}$, for example, the probability density function of a Gaussian distribution.*

5.3.2 General Model-based Evolutionary Optimization Algorithm

In this section, we give a Model-based Evolutionary Optimization Algorithm based on the connection between the evolutionary game and the global optimization problem. Assume that there is a probability density function p_t associated with the probability measure $\mathbb{P}_t$.

There are many dynamics in evolutionary game theory that can be used to govern the evolution of the fraction of agents playing different strategies, as shown in Section 2.2.3.1 for evolutionary games with finite strategies. Similarly, these dynamics can also be extended to games with a continuous strategies space to govern the evolution of the probabilistic model $p_t(y)$. For example, for replicator dynamics, we have equation (5.6), which is rewritten here:

$$\dot{p}_t(y) = (\varphi(H(y)) - E_{\mathbb{P}_t}[\varphi(H(Y))])p_t(y).$$

To describe these dynamics in a unified form, we use the following compact representation:

$$\dot{p}_t(y) = D(\varphi(H(y)), E_{p_t}[\varphi(H(Y))], p_t(y)), \tag{5.10}$$

where D is simply a function with three arguments and $\dot{p}_t(y)$ is a function of these three

149

quantities. The corresponding discrete-time version is

$$p_{k+1}(y) = D_d(\varphi(H(y)), E_{p_k}[\varphi(H(Y))], p_k(y)).$$

5.3.2.1 Model-based Evolutionary Optimization

From the analysis in Section 5.3.1, we know that global optimal solutions can be obtained by generating samples from some equilibrium distributions of the replicator dynamics (5.5); these equilibrium distributions can be approached by following trajectories of (5.5) starting from $\mathbb{P}_0 \in \Delta_0$. Based on this connection, we give the following Model-based Evolutionary Optimization (MEO) algorithm.

Model-based Evolutionary Optimization Algorithm

0. Initialization. Choose $\rho \in (0,1]$ and an initial p_0 defined on $\mathcal{Y}$. Set $k = 0$ and $\gamma_0 = -\infty$.

1. Quantile calculation. Calculate the $(1 - \rho)$ quantile γ_k:

$$\gamma_k = \sup_l \{l : P_k(H(y) \geq l) \geq \rho\}.$$

If $\gamma_k < \gamma_{k-1}$ and $k > 1$, set $\gamma_k = \gamma_{k-1}$. Set $k = k + 1$ and go to step 2.

2. Update the probabilistic model:

$$p_k(y) = D_d(\varphi(H(y))I_{\{H(y) \geq \gamma_{k-1}\}}, E_{p_t}[\varphi(H(Y))I_{\{H(Y) \geq \gamma_{k-1}\}}], p_{k-1}(y)).$$

3. Stop if some stopping criterion is satisfied; otherwise go to step 1.

In the MEO algorithm, the parameter ρ specifies the proportion of samples that will be used to update the probabilistic model; ρ also defines a sequence of $1 - \rho$ quantiles $\{\gamma_k, k = 1, 2, \ldots\}$. These quantiles are used to obtain a sequence of nondecreasing thresholds that are used to select samples for model updating. Only those samples that perform better than these thresholds will be used for model updating in step 2. These quantiles help to concentrate the computational effort of the algorithm on promising candidate solutions, and they also help to rule out samples with bad performance. The overall idea of the algorithm is conceptually simple. We use the sequence of nondecreasing thresholds to select promising samples, which are then used to update the probabilistic model governed by evolutionary dynamics. Once the probabilistic model approaches the equilibrium distribution, optimal solutions or solutions very close to optimal solutions can be easily obtained by sampling from the resulting distribution.

5.3.2.2   Monte Carlo Version of MEO

In MEO, the structure of the density p_k is not specified, and it might be difficult to generate candidate solutions from a general density p_k. The choice of p_k is crucial to the performance of the MEO algorithm. Note that by Theorem 5.3, the equilibrium points obtained by starting from $\mathbb{P}_0 \in \Delta_0$ are of the form $\mathbb{P}^\star = \sum_{i=1}^{m} \alpha_i \delta_{y_i^\star}$, where $\sum_{i=1}^{m} \alpha_i = 1$ and $\alpha_i \geq 0$ for $i = 1, \ldots, m$, which suggests using a sum of Dirac functions to approximate p_t. Assume a group of candidate solutions $\{y_t^i\}_{i=1}^{N}$ is generated from p_t; then the probability density function p_t can be approximated by

$$\hat{p}_t(y) = \sum_{i=1}^{N} w_t^i \delta(y - y_t^i),$$

151

where δ denotes the Dirac function, and $\{w_t^i\}_{i=1}^N$ are weights satisfying $\sum_{i=1}^N w_t^i = 1$. If we use this approximation $\hat{p}_t$ as our probabilistic model, we can rewrite (5.10) as

$$\sum_{i=1}^N \frac{\partial w_t^i}{\partial t}\delta(y - y_t^i) = D\Big(\varphi(H(y)), \sum_{j=1}^N w_t^j \varphi(H(y_t^j)), \sum_{i=1}^N w_t^i \delta(y - y_t^i)\Big),$$

which is equivalent to

$$\frac{\partial w_t^i}{\partial t} = D\Big(\varphi(H(y_t^i)), \sum_{j=1}^N w_t^j \varphi(H(y_t^j)), w_t^i\Big), \quad \forall i = 1, \dots, N. \tag{5.11}$$

The discrete-time version of (5.11) is

$$w_{k+1}^i = D_d\Big(\varphi(H(y_k^i)), \sum_{j=1}^N w_k^j \varphi(H(y_k^j)), w_k^i\Big), \quad \forall i = 1, \dots, N.$$

In particular, for replicator dynamics, we have

$$w_{k+1}^i = \frac{\varphi(H(y_k^i))}{\sum_{j=1}^N w_k^j \varphi(H(y_k^j))} w_k^i, \quad \forall i = 1, \dots, N.$$

Although an updated density approximation $\hat{p}_{k+1}(y) = \sum_{i=1}^N w_{k+1}^i \delta(y - y_k^i)$ is obtained, it cannot be used directly to generate new candidate solutions. We construct a new continuous density to approximate $\hat{p}_{k+1}$, which is done by projecting $\hat{p}_{k+1}$ onto some parameterized family of distributions g_θ. The idea of projection onto a parameterized family has also been used in CE ([89]) and MRAS ([59]). Specifically, we try to minimize the Kullback-Leibler (KL) distance between the parameterized distribution g_θ and $\hat{p}_{k+1}$:

$$\theta_{k+1} = \arg\min_{\theta \in \Theta} \mathcal{D}_{KL}\big(\hat{p}_{k+1}\|g_\theta\big), \tag{5.12}$$

where Θ is the domain of θ, and the KL distance is defined as

$$\mathcal{D}_{KL}(\hat{p}_{k+1}, g_\theta) = \int_{y \in \mathcal{Y}} \ln \frac{\hat{p}_{k+1}(y)}{g_\theta(y)} \hat{p}_{k+1}(y) dy$$

$$= \int_{y \in \mathcal{Y}} \ln \hat{p}_{k+1}(y) \hat{p}_{k+1}(y) dy - \int_{y \in \mathcal{Y}} \ln g_\theta(y) \hat{p}_{k+1}(y) dy.$$

Since the first term does not depend on the parameter θ, the minimization problem (5.12) is equivalent to

$$\max_{\theta \in \Theta} \int_{y \in \mathcal{Y}} \ln g_\theta(y) \hat{p}_{k+1}(y) dy,$$

which can be rewritten as

$$\max_{\theta \in \Theta} \sum_{i=1}^{N} w_{k+1}^i \ln g_\theta(y_k^i).$$

Based on the above analysis, a Monte Carlo simulation version of the MEO algorithm is given as follows.

Simulated Model-based Evolutionary Optimization Algorithm (SMEO)

0. Initialization. Specify N as the total number of candidate solutions generated at each iteration. Choose $\rho \in (0, 1]$ and an initial g_{θ_0} defined on $\mathcal{Y}$. Set $k = 0$, $w_0^i = 1/N$ for $i = 1, \ldots, N$, and $\gamma_0 = -\infty$.

1. Quantile Calculation. Generate N candidate solutions $\{y_k^i\}_{i=1}^N$ from g_{θ_k}. Calculate the $1 - \rho$ quantile γ_k of $\{y_k^i\}_{i=1}^N$. If $\gamma_k < \gamma_{k-1}$ and $k > 1$, set $\gamma_k = \gamma_{k-1}$ and

$w_{k-1}^i = 1/N$ for $i = 1, \ldots, N$. Set $k = k + 1$ and go to step 2.

2. Updating the probabilistic model. The discrete approximation of the model is

   $\hat{p}_k(y) = \sum_{i=1}^N w_k^i \delta(y - y_{k-1}^i)$, where

   $$w_k^i = D_d\Big(\varphi(H(y_{k-1}^i))I_{\{H(y_{k-1}^i) \geq \gamma_{k-1}\}}, \sum_{j=1}^N w_{k-1}^j \varphi(H(y_{k-1}^j))I_{\{H(y_{k-1}^j) \geq \gamma_{k-1}\}}, w_{k-1}^i\Big),$$

   for $i = 1, \ldots, N$.

3. Density projection. Construct g_θ by projecting the density $\hat{p}_k = \sum_{i=1}^N w_k^i \delta(y - y_{k-1}^i)$

   onto g_θ:

   $$\theta_k = \arg\max_{\theta \in \Theta} \sum_{i=1}^N w_k^i \ln g_\theta(y_{k-1}^i).$$

4. Stop if some stopping criterion is satisfied; otherwise go to step 1.

Generally it is not easy to solve the optimization problem (5.12), which depends on the choice of g_θ. However for g_θ in an exponential family, analytical solutions exist ([15]).

If replicator dynamics is used in the Simulated Model-based Evolutionary Optimization Algorithm above, then in step 2, we have

$$w_k^i = \frac{\frac{1}{N}\varphi(H(y_{k-1}^i))I_{\{H(y_{k-1}^i) \geq \gamma_{k-1}\}}}{\sum_{j=1}^N \frac{1}{N}\varphi(H(y_{k-1}^j))I_{\{H(y_{k-1}^j) \geq \gamma_{k-1}\}}}, \quad \forall i = 1, \ldots, N.$$

It is easy to show that the SMEO algorithm gives the same updated probabilistic model g_θ as the extended CE algorithm in [15] when multivariate normal distributions with independent components are used for g_θ.

When the Brown-von Neumann-Nash dynamics is used, define $N_e = \sum_{j=1}^N I_{\{H(y_{k-1}^j) \geq \gamma_{k-1}\}}$,

and then we have

$$w_k^i = \frac{\frac{1}{N_e} + \max(0, \varphi(H(y_{k-1}^i))I_{\{H(y_{k-1}^i) \geq \gamma_{k-1}\}} - \sum_{j=1}^N \varphi(H(y_{k-1}^j))I_{\{H(y_{k-1}^i) \geq \gamma_{k-1}\}})}{1 + \sum_{l=1}^N \max(0, \varphi(H(y_{k-1}^l))I_{\{H(y_{k-1}^l) \geq \gamma_{k-1}\}} - \sum_{j=1}^N \varphi(H(y_{k-1}^j))I_{\{H(y_{k-1}^i) \geq \gamma_{k-1}\}})},$$

for $\forall i = 1, \ldots, N$.

Remark 5.6 *In CE, the parameterized density function is chosen at the beginning of the algorithm and then the parameter is adaptively updated by an adaptively updated group of elite solutions. In SMEO, the probabilistic model is estimated by some weighted Dirac functions and the evolution of the estimated density function is governed by replicator dynamics. Projecting the estimated density function onto a parameterized family of density functions is the final step in SMEO. The Dirac function only gives a coarse approximation of the density function. One direction to improve SMEO is to explore effective approximations of the density function based on generated samples.*

5.4 Population Model-based Evolutionary Optimization

In the SMEO algorithm given in Section 5.3.2, the probability density approximation $\hat{p}_k$ is inherently multimodal for global optimization problems with many local maxima. The projection of $\hat{p}_k$ onto a family of unimodal probability density functions g_θ cannot capture this multimodal property. Motivated by the work of [58], in which candidate solutions were generated from a group of models with the emphasis on optimization of budget allocation, we consider using a mixture distribution as our probabilistic model in the SMEO algorithm, and focus on studying the evolving behavior of the individual models in the mixture distribution. The global optimization problem is formulated as an evolutionary game along similar lines as in Section 5.3.

5.4.1 Population Model

Let $\Gamma = \{g_{\theta_t^1}, \ldots, g_{\theta_t^M}\}$ be a set of M parameterized probability density functions at time t on $\mathcal{Y}$. Assume that $\{g_{\theta_t^i}, i = 1, \ldots, M\}$ belong to a parameterized distribution family $\{g_\theta, \theta \in \Theta\}$, where Θ is the parameter space. Consider generating N samples from the following probability density function at t:

$$g_t(y) = \sum_{i=1}^M \alpha_t^i g_{\theta_t^i}(y), \tag{5.13}$$

where $\alpha_t^i \geq 0$ and $\sum_{i=1}^M \alpha_t^i = 1$. We can use stratified sampling to generate $N^i = \lfloor \alpha_t^i N \rfloor$ samples $\{y_n^i, n = 1, \ldots, N_i\}$ from $g_{\theta_t^i}$ for $i = 1, \ldots, M-1$, and $N^M = N - \sum_{i=1}^{M-1} N^i$ samples $\{y_n^M, n = 1, \ldots, N_m\}$ from $g_{\theta_t^M}$, respectively. With these samples, the probability density function $g_{\theta_t^i}$ can be approximated by

$$\hat{g}_{\theta_t^i}(y) = \sum_{n=1}^{N^i} w_{t,n}^i \delta(y - y_n^i), \quad \forall i = 1, \ldots, M, \tag{5.14}$$

where $\{w_{t,n}^i\}$ satisfy $w_{t,n}^i \geq 0, \sum_{n=1}^{N^i} w_{t,n}^i = 1$ can be obtained and updated from the SMEO algorithm. Thus, $g_t(y)$ can be approximated as

$$\hat{g}_t(y) = \sum_{i=1}^M \alpha_t^i \sum_{n=1}^{N^i} w_{t,n}^i \delta(y - y_n^i). \tag{5.15}$$

In the above equation, both α_t^i and $w_{t,n}^i$ are evolving with time, and they are also interacting with each other. To analyze the evolution of $\hat{g}_t$, assume that there are two time scales in (5.15), i.e., α_t^i is evolving on a fast time scale, and $w_{t,n}^i$ is changing on a slow time scale. $w_{t,n}^i$ can be viewed as a constant when α_t^i is changing. If replicator dynamics

is used in (5.10), then we have

$$\sum_{i=1}^{M} \frac{\partial \alpha_t^i}{\partial t} \sum_{n=1}^{N^i} w_{t,n}^i \delta(y - y_n^i) = \sum_{i=1}^{M} \alpha_t^i \sum_{n=1}^{N^i} w_{t,n}^i \delta(y - y_n^i) \Big[\varphi(H(y)) - \sum_{i=1}^{M} \alpha_t^i \sum_{n=1}^{N^i} w_{t,n}^i \varphi(H(y_n^i)) \Big],$$

which leads to

$$\frac{\partial \alpha_t^i}{\partial t} = \alpha_t^i \Big(\hat{\mathcal{I}}^i - \sum_{j=1}^{M} \alpha_t^j \hat{\mathcal{I}}^j \Big), \quad \forall i = 1, \ldots, M, \tag{5.16}$$

where $\hat{\mathcal{I}}^j = \sum_{n=1}^{N^j} w_{t,n}^j \varphi(H(y_n^j))$. It is easy to see from (5.16) that if $\hat{\mathcal{I}}^i$, the performance of the probability model i, is greater than the average performance $\sum_{j=1}^{M} \alpha_t^j \hat{\mathcal{I}}^j$, more samples will be generated from the model i. The evolution of the weights $\{\alpha_t^i\}$ can be viewed as an evolution of the balance between exploration and exploitation when searching for optimal solutions. The discrete-time version of (5.16) is

$$\alpha_{k+1}^i = \alpha_k^i \frac{\hat{\mathcal{I}}_k^i}{\sum_{j=1}^{M} \alpha_k^j \hat{\mathcal{I}}_k^j}. \tag{5.17}$$

When Brown-von Neumann-Nash Dynamics is used, we have

$$\alpha_{k+1}^i = \frac{\alpha_k^i + \max(0, \hat{\mathcal{I}}_k^i - \sum_{j=1}^{M} \alpha_k^j \hat{\mathcal{I}}_k^j)}{1 + \sum_{l=1}^{M} \max(0, \hat{\mathcal{I}}_k^l - \sum_{j=1}^{M} \alpha_k^j \hat{\mathcal{I}}_k^j)}. \tag{5.18}$$

Remark 5.7 *By the strong law of large numbers, when $N^i \to \infty$, $\hat{\mathcal{I}}^i \to E_{\theta^i}[\varphi(H(Y))]$, where the expectation is taken under the probability measure induced by g_{θ^i}.*

Remark 5.8 *The above analysis is done by viewing each $y \in \mathcal{Y}$ as a strategy, as in Section 5.3. Since samples are generated from a population of models, we can set up an evolutionary game from a different perspective. Assume there are $\{1, \ldots, M\}$ pure strategies in an*

evolutionary game corresponding to the probability density functions $\{g_{\theta^i}, i = 1, \ldots, M\}$.

The action of playing a pure strategy i can be viewed as generating samples from g_{θ^i}; α_t^i is

the percentage of agents that play i at each iteration. Assume that the payoff of playing the

pure strategy i is $\hat{\mathcal{I}}^i$. By replicator dynamics, the same conclusion that $\{\alpha_t^i, i = 1, \ldots, M\}$

is governed by (5.16) can be reached.

Remark 5.9 *A mean-field equation similar to (5.16) has been obtained in [106], in which*

evolutionary game theory is used to study strategies for pursuit. By deriving the mean-field

equation as an orthogonal projection (with respect to the Fisher-Rao-Shahshahani metric)

of a linear vector field in the positive orthant, [106] showed that

$$\alpha_t^i = \frac{\alpha_0^i e^{\hat{\mathcal{I}}^i t}}{\sum_j^M \alpha_0^j e^{\hat{\mathcal{I}}^j t}} \tag{5.19}$$

solves (5.16). The model with the highest average payoff dominates all the other models,

which agrees with the intention of generating more samples from promising models.

5.4.2   Updating Population of the Probability Models

From the replicator dynamics (5.16), the weights of more promising models will

increase, resulting in generating more samples from the promising models. Besides doing

exploration on a fast time scale, we also exploit individual models to further improve the

performance of the algorithm on a slow time scale, which can be carried out through

biasing the probability density functions $\{g_{\theta^i}, i = 1, \ldots, M\}$ towards promising areas by

using previously generated candidate solutions.

5.4.2.1 Updating of Probabilistic Model

Many model-based methods such as EDAs, CE, and MRAS can be incorporated into our framework to update the population of probabilistic models. Here we use the SMEO method (CE can be viewed as a special case of SMEO; see Section 5.3.2.2). Let g_θ be a parameterized family of distributions. Generate $N_i^k = \lfloor \alpha_k^i N \rfloor$ samples $\{y_{k,j}^i, j = 1, \ldots, N_k^i\}$ from $g_{\theta_k^i}$ for all $i = 1, \ldots, M-1$, and $N - \sum_{i=1}^{M-1} N_k^i$ samples from $g_{\theta_k^M}$, and calculate the performances $H(y_{k,j}^i)$ at iteration k. Then as in the SMEO algorithm, we project (5.14) onto an exponential family, and when replicator dynamics is used, the parameter of $g_{\theta_k^i}$ is updated by

$$
\theta_k^i = \arg\max_{\theta \in \Theta} \sum_{j=1}^{N_{k-1}^i} \frac{\frac{1}{N_{k-1}^i} \varphi(H(y_{k-1}^{i,j})) I_{\{H(y_{k-1}^{i,j}) \geq \gamma_{k-1}\}}}{\sum_{m=1}^{N_{k-1}^i} \frac{1}{N_{k-1}^i} \varphi(H(y_{k-1}^{i,m})) I_{\{H(y_{k-1}^{i,m}) \geq \gamma_{k-1}\}}} \ln g_\theta(y_{k-1}^{i,j}), \quad \forall i = 1, \ldots, M,
$$

(5.20)

where γ_{k-1} is a positive threshold parameter.

The exponential family contains a broad class of distributions, such as the Gaussian and binomial. A closed form solution of θ_k^i in (5.20) can be obtained if the exponential family is used to update the distributions. For continuous optimization problems, it is convenient to use multivariate Gaussian distributions with independent components. At iteration k, assume that the parameterized distribution has the following form:

$$
g_{\theta_k^i}(y) = \prod_{d=1}^{n} \frac{1}{\sqrt{2\pi(\sigma_k^{i,d})^2}} \exp\left(-\frac{(y_d - \mu_k^{i,d})^2}{2(\sigma_k^{i,d})^2}\right), \quad \forall i = 1, \ldots, M, \qquad (5.21)
$$

where n is the problem dimension, y_d is the dth element of y, and $\mu_k^{i,d}, \sigma_k^{i,d}$ are parameters.

Define the "elite" sets $\mathscr{L}_k^i = \{y_k^{i,j} : H(y_k^{i,j}) \geq \gamma\} \quad \forall i = 1, \ldots, M$, and solve the

optimization problem (5.20). We obtain

$$\tilde{\mu}_{k+1}^{i,d} = \frac{\sum_{y \in \mathscr{L}_k^i} \varphi(H(y)) y_d}{\sum_{y \in \mathscr{L}_k^i} \varphi(H(y))},$$

$$(\tilde{\sigma}_{k+1}^{i,d})^2 = \frac{\sum_{y \in \mathscr{L}_k^i} \varphi(H(y))(y_d - \mu_{k+1}^{i,d})^2}{\sum_{y \in \mathscr{L}_k^i} \varphi(H(y))},$$

for all $d = 1, \ldots, n$ and $i = 1, \ldots, M$. Let $\tilde{N}_k^i$ be the total number of elements in $\mathscr{L}_k^i$ for $i = 1, \ldots, M$. When Brown-von Neumann-Nash dynamics is used, similar to the case of replicator dynamics, we can show that

$$\tilde{\mu}_{k+1}^{i,d} = \frac{\sum_{y \in \mathscr{L}_k^i} \left[\frac{1}{\tilde{N}_k^i} + \max(0, \varphi(H(y)) - \frac{1}{\tilde{N}_k^i} \sum_{y \in \mathscr{L}_k^i} \varphi(H(y))) \right] y_d}{1 + \sum_{y \in \mathscr{L}_k^i} \left[\max(0, \varphi(H(y)) - \frac{1}{\tilde{N}_k^i} \sum_{y \in \mathscr{L}_k^i} \varphi(H(y))) \right]},$$

$$(\tilde{\sigma}_{k+1}^{i,d})^2 = \frac{\sum_{y \in \mathscr{L}_k^i} \left[\frac{1}{\tilde{N}_k^i} + \max(0, \varphi(H(y)) - \frac{1}{\tilde{N}_k^i} \sum_{y \in \mathscr{L}_k^i} \varphi(H(y))) \right] (y_d - \mu_{k+1}^{i,d})^2}{1 + \sum_{y \in \mathscr{L}_k^i} \left[\max(0, \varphi(H(y)) - \frac{1}{\tilde{N}_k^i} \sum_{y \in \mathscr{L}_k^i} \varphi(H(y))) \right]},$$

for all $d = 1, \ldots, n$ and $i = 1, \ldots, M$. Let ν be a mixing parameter, and update the parameters by

$$\mu_{k+1}^{i,d} = (1 - \nu)\tilde{\mu}_{k+1}^{i,d} + \nu\mu_k^{i,d},$$

$$(\sigma_{k+1}^{i,d})^2 = (1 - \nu)(\tilde{\sigma}_{k+1}^{i,d})^2 + \nu\big((\sigma_k^{i,d})^2 + (\mu_{k+1}^{i,d} - \mu_k^{i,d})^2\big), \tag{5.22}$$

for all $d = 1, \ldots, n$ and $i = 1, \ldots, M$. The mixing parameter, which is widely applied in model-based algorithms ([15, 59]), can prevent the premature convergence of the algorithm.

5.4.3 PMEO Algorithm

Our algorithm is a simulation-based optimization algorithm, and hence stochastic counterparts are used to estimate expectations of random variables. The approximate

performance $\hat{\mathcal{I}}_k^i(\gamma)$ is given by

$$\hat{\mathcal{I}}_k^i(\gamma) = \frac{1}{\sum_{j=1}^{N_k^i} I_{\{H(y_k^{i,j}) \geq \gamma\}}} \sum_{j=1}^{N_k^i} \varphi(H(y_k^{i,j})) I_{\{H(y_k^{i,j}) \geq \gamma\}}, \quad \forall i = 1, \ldots, M. \qquad (5.23)$$

Based on the above analysis, we give the following Population Model-based Evolutionary Optimization algorithm.

Population Model-based Evolutionary Optimization Algorithm

0. Initialization: Choose N as the number of total samples at each iteration. Specify the initial weights $\{\alpha_0^i, i = 1, \ldots, M\}$ and the probability density function $\{g_{\theta_0^i}, i = 1, \ldots, M\}$. Choose ρ as the quantile parameter. Set $k = 0$ and $\gamma_0 = -\infty$.

1. Generate $N_k^i = \lfloor \alpha_k^i N \rfloor$ samples $\{y_k^{i,j}, j = 1, \ldots, N_k^i\}$ from $g_{\theta_k^i}$ for all $i = 1, \ldots, M-1$, and $N - \sum_{i=1}^{M-1} N_k^i$ samples from $g_{\theta_k^M}$; compute the performances $H(y_k^{i,j})$ over all i, j. Order the performances from smallest to largest, $H_{(1)} \leq \ldots \leq H_{(N)}$. Let γ_k be the $(1 - \rho)$ sample quantile of performances: $\gamma_k = H_{(\lceil (1-\rho)N \rceil)}$. When $k > 0$, if $\gamma_k \leq \gamma_{k-1}$, set $\gamma_k = \gamma_{k-1}$. Generate the "elite" sets $\mathscr{L}_k^i = \{y_k^{i,j} : H(y_k^{i,j}) \geq \gamma_k\}$ for all $i = 1, \ldots, M$. Compute $\hat{\mathcal{I}}_k^i$ by (5.23).

2. Let

$$\alpha_{k+1}^i = \alpha_k^i \frac{\hat{\mathcal{I}}_k^i}{\sum_{j=1}^M \alpha_j \hat{\mathcal{I}}_k^j}.$$

Update the parameter θ_{k+1}^i according to (5.22) for $i = 1, \ldots, M$.

3. If a stopping rule is met, then stop; otherwise set $k = k + 1$ and go to step 1.

If Brown-von Neumann-Nash dynamics is used, then (5.18) will be used in Step 2 of the above algorithm. Maximization problems are considered in the above algorithm, which can be easily adjusted for solving minimization problems.

161

5.4.3.1 Properties of Population Model Updating of PMEO

In this section, we show that the overall fitness of the probabilistic model g_k in PMEO is monotonically increasing.

Theorem 5.4 *Let $g_k(y) = \sum_{i=1}^{M} \alpha_k^i g_{\theta_k^i}(y)$ be the probabilistic model used at iteration k in the PMEO algorithm. For a given γ_{k+1}, we have*

$$E_{g_{k+1}}[\varphi(H(Y))I_{\{H(y) \geq \gamma_{k+1}\}}] \geq E_{g_k}[\varphi(H(Y)))I_{\{H(y) \geq \gamma_{k+1}\}}].$$

Proof: First we show that the fitness of individual models $\{g_{\theta_k^n}\}$, i.e., the individual model updated by SMEO, is increasing. In this part of the proof we use a similar technique as in the proof of Theorem 2 in [59]. From SMEO, assume that at iteration $k + 1$, we generate $\{y_{k+1}^i\}_{i=1}^N$ candidate solutions from $g_{\theta_k^n}$. Define

$$\tilde{p}_{k+2}(y) = \sum_{i+1}^{N} \frac{\frac{1}{N}\varphi(H(y_{k+1}^i))I_{\{H(y_{k+1}^i) \geq \gamma_{k+1}\}} \frac{g_{\theta_{k+1}^n}(y_{k+1}^i)}{g_{\theta_k^n}(y_{k+1}^i)}}{\sum_{j=1}^{N} \frac{1}{N}\varphi(H(y_{k+1}^j))I_{\{H(y_{k+1}^j) \geq \gamma_{k+1}\}}} \delta(y - y_{k+1}^i).$$

From the SMEO algorithm in Section 5.3.2.2, we have

$$\hat{p}_{k+1} = \sum_{i+1}^{N} \frac{\frac{1}{N}\varphi(H(y_{k+1}^i))I_{\{H(y_i^{k+1}) \geq \gamma_{k+1}\}}}{\sum_{j=1}^{N} \frac{1}{N}\varphi(H(y_{k+1}^j))I_{\{H(y_{k+1}^j) \geq \gamma_{k+1}\}}} \delta(y - y_{k+1}^i).$$

By equation (5.12), θ_{k+1}^n minimizes the K-L divergence $\mathcal{D}_{KL}(\hat{p}_{k+1}\|g_\theta)$, and hence we have

$$0 \leq \mathcal{D}_{KL}(\hat{p}_{k+1}\|g_{\theta_k^n}) - \mathcal{D}_{KL}(\hat{p}_{k+1}\|g_{\theta_{k+1}^n})$$

$$\leq \mathcal{D}_{KL}(\hat{p}_{k+1}\|g_{\theta_k^n}) - \mathcal{D}_{KL}(\hat{p}_{k+1}\|g_{\theta_{k+1}^n}) + \mathcal{D}_{KL}(\hat{p}_{k+1}\|\tilde{p}_{k+2})$$

$$= E_{\hat{p}_{k+1}}[\ln\hat{p}_{k+1}] - E_{\hat{p}_{k+1}}[\ln g_{\theta_k^n}] - E_{\hat{p}_{k+1}}[\ln\hat{p}_{k+1}] + E_{\hat{p}_{k+1}}[\ln g_{\theta_{k+1}^n}]$$

$$+ E_{\hat{p}_{k+1}}\left[\frac{g_{\theta_k^n}}{g_{\theta_{k+1}^n}}(Y)\right] + \ln\frac{\sum_{j=1}^N \varphi(H(y_{k+1}^j))I_{\{H(y_{k+1}^j)\geq\gamma_{k+1}\}}\frac{g_{\theta_{k+1}^n}}{g_{\theta_k^n}}(y_{k+1}^j)}{\sum_{i=1}^N \varphi(H(y_{k+1}^i))I_{\{H(y_{k+1}^i)\geq\gamma_{k+1}\}}}$$

$$= \ln\frac{\sum_{j=1}^N \varphi(H(y_{k+1}^j))I_{\{H(y_{k+1}^j)\geq\gamma_{k+1}\}}\frac{g_{\theta_{k+1}^n}}{g_{\theta_k^n}}(y_{k+1}^j)}{\sum_{i=1}^N \varphi(H(y_{k+1}^i))I_{\{H(y_{k+1}^i)\geq\gamma_{k+1}\}}},$$

where the second inequality holds since the K-L divergence is always nonnegative. Therefore we have

$$\sum_{j=1}^N \varphi(H(y_{k+1}^j))I_{\{H(y_{k+1}^j)\geq\gamma_{k+1}\}}\frac{g_{\theta_{k+1}^n}}{g_{\theta_k^n}}(y_{k+1}^j) \geq \sum_{j=1}^N \varphi(H(y_{k+1}^j))I_{\{H(y_{k+1}^j)\geq\gamma_{k+1}\}}.$$

We divide both sides of the above equation by N, and take limit on both sides as $N \to \infty$, and then by the strong law of large numbers,

$$E_{\theta_{k+1}^n}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}] \geq E_{\theta_k^n}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}]. \tag{5.24}$$

Now we prove that the fitness of g_k is monotonically increasing. By the definition of α_{k+1}^i,

$$E_{g_{k+1}}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}]$$

$$= \sum_{i=1}^M \alpha_{k+1}^i E_{\theta_{k+1}^i}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}]$$

$$= \sum_{i=1}^M \frac{\alpha_k^i E_{\theta_{k+1}^i}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}]}{\sum_{j=1}^M E_{\theta_k^j}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}]}E_{\theta_{k+1}^i}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}].$$

163

Therefore we have

$$
\begin{aligned}
& E_{g_{k+1}}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}] - E_{g_k}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}] \\
& = \sum_{i=1}^{M} \frac{\alpha_k^i (E_{\theta_{k+1}^i}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}])^2}{\sum_{j=1}^{M} E_{\theta_k^j}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}]} - \sum_{j=1}^{M} \alpha_k^j E_{\theta_k^j}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}] \\
& \geq \frac{\sum_{i=1}^{M} \alpha_k^i (E_{\theta_k^i}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}])^2 - (\sum_{i=1}^{M} \alpha_k^i E_{\theta_k^i}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}])^2}{\sum_{j=1}^{M} E_{\theta_k^j}[\varphi(H(Y))I_{\{H(y)\geq\gamma_{k+1}\}}]} \geq 0.
\end{aligned}
$$

The first equality holds by equation (5.24) and the last inequality holds since the variance of a random variable is always nonnegative. ∎

Remark 5.10 *[106] showed that the right hand side of (5.16) is the gradient of the average fitness $\sum_{j=1}^{M} \alpha_t^j \hat{\mathcal{I}}^j$ with respect to a Fisher-Rao-Shahshahani metric, and thus concluded that (5.16) is a gradient ascent equation, which gave another interesting explanation that the average fitness of the probabilistic model in PMEO is monotonically increasing as stated in Theorem 4.*

5.4.3.2 PMEO with Mutation Strategies

In the mixture distribution (5.13), weights of models that have lower than average performances will approach zero eventually by (5.19). This will reduce the diversity of the probabilistic models in the mixture that is used to generate candidate solutions. To maintain the diversity of the probabilistic models in the population, we propose to introduce a mutant probabilistic model into the model population. The parameter in the mutant probabilistic model is given by

$$
\tilde{\theta} = \theta_{i^\star} + \Delta\theta,
$$

164

where $i^\star = \{i : \mathcal{I}^i \geq \mathcal{I}^j, \forall j = 1, \ldots, M\}$, and $\Delta\theta$ is a small random perturbation. For the probabilistic model (5.21), we could let $\Delta\mu = -\lambda + \frac{1}{2}\lambda U$, where λ is a constant and U is an n dimensional random variable with each component uniformly distributed on $[0, 1]$, and $\Delta\sigma_d^2 = \upsilon, d = 1, \ldots, n$, where υ is a positive constant.

Population Model-based Evolutionary Optimization Algorithm with Mutation

0. Initialization: Specify N as the number of total samples at each iteration. Specify the weights $\{\alpha_0^i, i = 1, \ldots, M\}$ and the probability density functions $\{g_{\theta_0^i}, i = 1, \ldots, M\}$. Let ρ be the quantile parameter. Set $k = 0$ and $\gamma_0 = -\infty$. Specify a threshold weight $\tilde{\alpha}$ and a λ for a mutant strategy. Specify a positive integer N_d, and let $count = 0$.

1. Let $i_\star = \{i : \alpha_k^i \leq \alpha_k^j, j = 1, \ldots, M\}$ and $i^\star = \{i : \alpha_k^i \geq \alpha_k^j, j = 1, \ldots, M\}$.

   – If $\alpha_k^{i_\star} < \tilde{\alpha}$, let $\theta_k^{i_\star} = \theta_k^{i^\star} + \Delta\theta$.

   – If $count = N_d$, let $\theta_k^{i_\star} = \theta_k^{i^\star} + \Delta\theta$ and $count = 0$.

   – Generate $N_k^i = \lfloor \alpha_k^i N \rfloor$ samples $\{y_k^{i,j}, j = 1, \ldots, N_k^i\}$ from $g_{\theta_k^i}$ for all $i = 1, \ldots, M-1$, and $N - \sum_{i=1}^{M-1} N_k^i$ samples from $g_{\theta_k^M}$; compute the performances $H(y_k^{i,j})$. Order the performances from smallest to largest, $H_{(1)} \leq \ldots \leq H_{(N)}$. Let γ_k be the $(1 - \rho)$ sample quantile of performances: $\gamma_k = H_{(\lceil (1-\rho)N \rceil)}$. When $k > 0$, if $\gamma_k \leq \gamma_{k-1} + \epsilon$, set $\gamma_k = \gamma_{k-1}$ and $count = count + 1$; otherwise set $count = 0$. Define the "elite" sets $\mathscr{L}_k^i = \{y_k^{i,j} : H(y_k^{i,j}) \geq \gamma_k\}$ for all $i = 1, \ldots, M$. Compute $\hat{\mathcal{I}}_k^i$ by (5.23).

2. Let

$$\alpha_i^{k+1} = \alpha_i^k \frac{\hat{\mathcal{I}}_k^i}{\sum_{j=1}^M \alpha_j \hat{\mathcal{I}}_k^j}.$$

Update the parameter θ_{k+1}^i according to (5.22) for $i = 1, \ldots, M$.

3. If a stopping rule is met, then stop; otherwise set $k = k + 1$ and go to step 1.

In the above algorithm, $i_\star$ is the index of the model with the worst performance, and $i^\star$ is the index of the model with the best performance. *count* is used to count the number of iterations in which the improvement of the threshold γ_k is less than a positive constant ϵ. When the weight of the $i_\star$ model is less than the threshold $\tilde{\alpha}$, this model will be discarded and a mutation model based on the model with the best performance will be introduced. If the improvement of γ_k is less than ϵ in N_d consecutive iterations, a mutation strategy will also be introduced.

5.5   Numerical Examples

In this section, we illustrate the performance of the PMEO algorithm on various benchmark problems that are well known in global optimization and compare its performance with the CE method.

Since minimization problems are considered, whereas PMEO was in the form of solving maximization problems, the following modifications are made. When computing the $1 - \rho$ quantile γ_k, we first order the performance function from largest to smallest, and then take the $\lceil (1-\rho)N \rceil$ statistics. In the PMEO with mutation algorithm, $\varphi(\cdot)$ is a strictly increasing function, which is used to update weights of different models. Minimization problems require a strictly decreasing function, which we take as $\varphi(x) = r/x$, where $r = \alpha \gamma_k$ at iteration k and α is a positive constant.

166

We test two instantiations of PMEO, one with replicator dynamics denoted by PMEO-RP and the other with Brown-von Neumann-Nash dynamics denoted by PMEO-BNND. We test our algorithms on the following benchmark problems, which have been previously studied by [84], [97], and [59].

H1. Dejong's 5th function ($n = 2$).

$$H_1(y) = \left[0.002 + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^{2}(y_i - a_{j,i})^6}\right],$$

where $a_{j,1} = \{-32, -16, 0, 16, 32, -32, -16, 0, 16, 32, -32, -16, 0, 16, 32, -32, -16, 0, 16,$ $32, -32, -16, 0, 16, 32\}; a_{j,1} = \{-32, -32, -32, -32, -32, -16, -16, -16, -16, -16, 0, 0,$ $0, 0, 0, 16, 16, 16, 16, 16, 32, 32, 32, 32, 32, \}$; the global optimal solution is $y^* = (-32, -32)^T$, $H_1(y^*) \approx 0.998$.

H2. Rosenbrock function ($n = 20$).

$$H_2(y) = \sum_{i=1}^{n-1} 100(y_{i+1} - y_i^2)^2 + (y_i - 1)^2,$$

where the global optimal solution is $y^* = (1, \ldots, 1)^T, H_2(y^*) = 0$.

H3. Powell singular function ($n = 20$).

$$H_3(y) = \sum_{i=1}^{n-2} \left[(y_{i-1} + 10y_i)^2 + 5(y_{i+1} - y_{i+2})^2 + (y_i - 2y_{i+1})^4 + 10(y_{i-1} - y_{i+2})^4\right],$$

where the global optimal solution is $y^* = (0, \ldots, 0)^T, H_3(y^*) = 0$.

H4. Trigonometric function ($n = 20$).

$$H_4(y) = 1 + \sum_{i=1}^{n} 8\sin^2(7(y_i - 0.9)^2) + 6\sin^2(14(y_i - 0.9)^2) + (y_i - 0.9)^2,$$

where the global optimal solution is $y^* = (0.9, \ldots, 0.9)^T$, $H_4(y^*) = 1$.

H5. Griewank function ($n = 20$).

$$H_5(y) = 1 + \sum_{i=1}^{n} 8\sin^2(7(y_i - 0.9)^2) + 6\sin^2(14(y_i - 0.9)^2) + (y_i - 0.9)^2,$$

where the global optimal solution is $y^* = (0, \ldots, 0)^T$, $H_5(y^*) = 0$.

H6. Pintér's function ($n = 20$).

$$H_6(y) = \sum_{i=1}^{n} iy_i^2 + \sum_{i=1}^{n} 20i\sin^2(y_{i-1}\sin y_i - y_i + \sin y_{i+1}) +$$
$$\sum_{i=1}^{n} i\log_{10}(1 + i(y_{i-1}^2 - 2y_i + 3y_{i+1} - \cos y_i + 1)^2),$$

where the global optimal solution is $y^* = (0, \ldots, 0)^T$, $H_2(y^*) = 0$.

Function H_1 is low dimensional with only a few local minima, which are separated by plateaus and are relatively far apart. H_2 and H_3 are 20-dimensional badly scaled functions. H_4 is highly multimodal. Function H_5 is highly multimodal, and H_6 is both badly scaled and highly multimodal.

In PMEO-RP and PMEO-BNND, for the six test problems, the same set of parameters is used. Let $\rho = 0.03, \nu = 0.8, \epsilon = 0.02, N_d = 4$, and $\tilde{\alpha} = 0.1$. Let $\alpha = 2$, and use $M = 3$ different models in the mixture distribution. For the exponential family, multivariate normal distributions with independent components are used. Each component of the

mean vector of the four models is uniformly selected from $[-30, 20], [-20, 30], [-50, 50],$ and the covariance matrix is a diagonal matrix with each diagonal element equal to 500. $N = 1400$ samples are generated at each iteration. By sampling from different intervals to obtain the mean vector for different models, a certain level of exploration is gained. However at the same time, since a fixed number of samples are generated at each iteration, fewer samples are generated from each model when a group of models are used, which reduces exploitation of each model. The balance between exploration and exploitation is adjusted by the weight updating in the algorithm. To introduce the mutation strategy, we let $\Delta\theta = (-\frac{1}{2}\max(\theta_{i*}^k) + \max(\theta_{i*}^k)U) \times 10/k$, where U is an n-dimensional standard normal random variable. Since better candidate solutions are more likely to be found around the best model, we use $\max(\theta_{i*}^k)$ to define the range of our random perturbation.

For the CE method, a multivariate normal distribution with independent components is used as the probability model. Specify the quantile parameter $\rho = 0.03$, and the smoothing parameter $\nu = 0.7$. Each element of the initial mean vector of the probability density function is uniformly selected from $[-50, 50]$, and the covariance matrix is a diagonal matrix with each diagonal element equal to 500. $N = 1400$ samples are generated at each iteration. To prevent the variance matrix from converging quickly to zero, which makes the algorithm get trapped in regions with low quality solutions, we also tried the CE method with some other smoothing parameters, and the numerical results are reported for CE with $\nu = 0.2$, which seemed to give the best overall performance, and for CE with $\nu = 0.7$

20 independent replications of PMEO-RP, PMEO-BNND, CE with $\nu = 0.7$, and CE with $\nu = 0.2$ are performed for each problem, and the numerical results are reported in Figure 5.1(a) to Figure 5.3(b), which show the average of the current best solutions
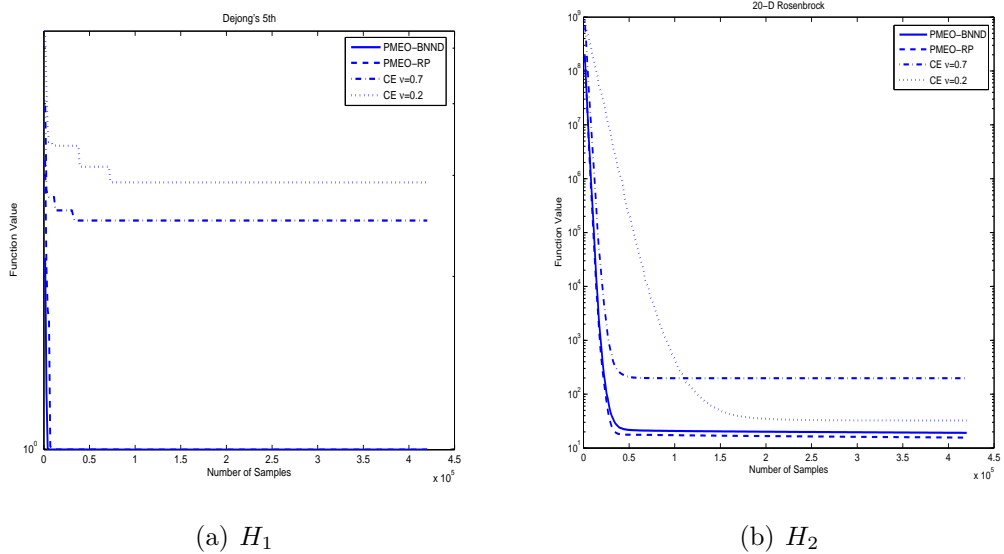
(a) H_1 (b) H_2

Figure 5.1: Dejong H_1 and Rosenbrock H_2

given all the samples generated. Our comparison is based on the same computational effort. Since evaluating the performance function accounts for most of the computational time for each algorithm, the total sample size is used to evaluate the computational effort. Figure 5.1(a) to Figure 5.3(b) show the comparison between the CE method, PMEO-RP, and PMEO-BNND. A dotted line represents CE with $\nu = 0.2$, a dash doted line represents CE with $\nu = 0.7$, a dashed line represents PMEO-RP, and a solid line represents the PMEO-BNND.

Function H_1 has only a few local minima, but the local minima are surrounded by plateaus. From Figure 5.1(a), we can see that CE with $\nu = 0.7$ and CE with $\mu = 0.2$ converge quickly to local minimal points, and fail to escape out of them. Both PMEO-RP and PMEO-BNND clearly outperform CE with both $\nu = 0.7$ and $\nu = 0.2$. Because of the multimodal property of PMEO, PMEO-RP and PMEO-BNND can quickly find a better solution and converge to the global optimal solution.

For the badly scaled functions H_2 and H_3, both PMEO-RP and PMEO-BNND
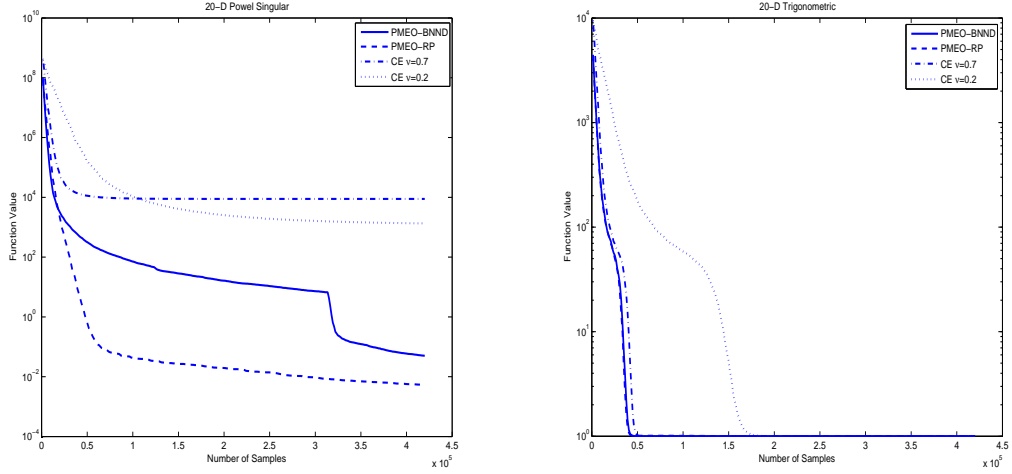
(a) H_3 (b) H_4

Figure 5.2: Powell Singular H_3 and Trigonometric H_4

perform better than CE with $\nu = 0.7$ and CE with $\nu = 0.2$. CE with $\nu = 0.7$ converges quickly to a local minimum, and never gets out of it. CE with $\mu = 0.2$ converges to slightly better solutions, although it converges slowly. PMEO-BNND quickly locates a local optimal solution, and has been able to get out it. PMEO-RP has better performance than PMEO-BNND, and particularly, in Figure 5.3(a), it can consistently get out of local minima and finds better solutions.

PMEO-RP, PMEO-BNND, and CE with $\nu = 0.7$ all work very well for H_4 and H_5, which have many local minima, and the all three algorithms converge quickly to the optimal solution. CE with $\nu = 0.2$ converges much slower than the other three algorithms since it spends more time doing exploration. For the badly scaled and highly multimodal function H_6, PMEO-RP and PMEO-BNDD are superior to both CE with $\nu = 0.7$ and CE with $\nu = 0.2$. In the first 70000 samples, PMEO-RP, PMEO-BNND, and CE with $\nu = 0.7$ have similar performances. Then CE with $\nu = 0.7$ gets trapped around low quality solutions. On the other hand, PMEO-RP and PMEO-BNND successfully escape low quality local solutions, and find better solutions.

171
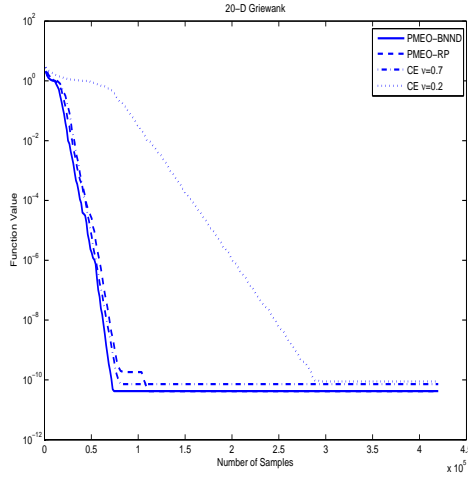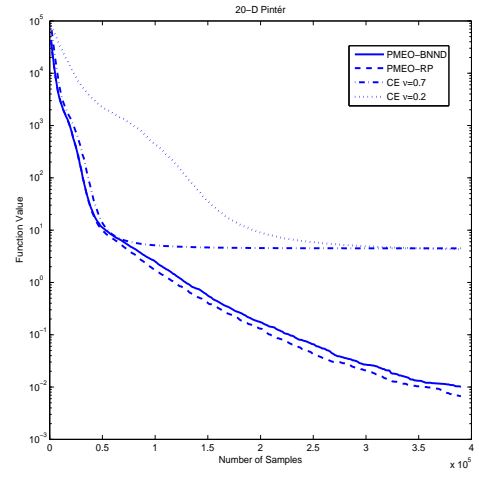
(a) H_5　　　　　　　　　　(b) H_6

Figure 5.3: Griewank H_5 and Pintér H_6

In summary, we can see that PMEO-RP and PMEO-BNND are able to find better solutions than CE with different smoothing parameters. PMEO-RP and PMEO-BNND are inherently based on multimodal probability models and have a better mechanism to escape the trap of local optimal solutions.

5.6   Conclusions

We have developed a new framework for solving global optimization problems by formulating the global optimization problem as an evolutionary game. By using Lyapunov theory, we showed that a particular equilibrium set of the replicator dynamics in the evolutionary game has a strong connection with the global optimal solutions, and this equilibrium set is asymptotically stable under mild conditions. Based on this connection, we proposed a model-based evolutionary optimization (MEO) algorithm, which includes the extended CE algorithm as an instantiation. We also proposed a population model-based evolutionary optimization algorithm, which better captures the multimodal

172

property of global optimization problems. Motivated by the ideas of mutation strategies in evolutionary games, we introduced mutation strategies by adding random perturbations to the best model. This enables the PMEO algorithm to escape local optimal solutions effectively.

Preliminary numerical tests for PMEO with mutation were reported on some widely used benchmark examples. Although simulation results show that PMEO with mutation might be able provide high quality solutions, there are still challenges when implementing the PMEO algorithm. Most of the parameters in PMEO are determined by trial and error, although in our preliminary numerical tests, we found a set of parameters that work well for all the test functions. How to effectively select parameters is an interesting topic for future research.

As a general framework for global optimization, PMEO holds the promise to be extended to solve other optimization problems, such as optimization problems with stochastic objective functions, and Markov decision processes with a large policy space.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This dissertation has proposed and developed new simulation-based optimization methods for solving stochastic control problems with parameterized policies and global optimization problems, and has also proposed a new stochastic gradient estimator for performance functions containing discontinuities that can be used for simulation optimization as well as sensitivity analysis. After a brief overview of simulation optimization and evolutionary game theory in Chapter 2, we presented three main results in Chapters 3, 4, and 5, which compose the main body of this dissertation.

In Chapter 3, we considered a dynamic pricing and inventory control problem arising in revenue management, in which the demand follows a continuous-time, continuous-state stochastic process instead of the commonly used discrete-time stochastic process. We developed a simulation-based algorithm for solving a class of stochastic control problems with parameterized policies, motivated by the fact that the combined dynamic pricing and inventory control problem generally does not have analytical solutions and numerical methods such as Markov Chain approximations and finite differences cannot be applied directly to solve the problem. When a continuous pricing policy is allowed, we modeled the dynamic pricing problem as a stochastic control problem and gave a theoretical solution for a special case. When only a finite number of price changes is allowed in the pricing policy, we proposed a simulation-based method for solving the pricing problem

for a broad range of demand models. We gave a new simulation scheme to simulate the evolution of the inventory level. Based on the generated sample paths, we derived gradient estimators for the expected discretized profit function with respect to various parameters. When we derived the gradient estimators for the expected discretized profit function, we circumvented the difficulty of differentiating a performance function with discontinuous sample paths by using smoothed perturbation analysis. We showed the unbiasedness of the resulting estimators. We also showed the convergence of the SA algorithm with the proposed gradient estimators. Simulation examples demonstrated that the proposed algorithm works well.

In Chapter 4, we considered a derivative estimation problem that contains discontinuous payoff functions, motivated by the fact that existing derivative estimation techniques are either not valid or not efficient. Inspired by IPA and LR, we derived a new computationally efficient derivative estimation technique called the Support independent unified Likelihood Ratio and Infinitesimal Perturbation Analysis (SLRPA), which applies an appropriate change of variables to circumvent the difficulty of differentiating indicator functions and gives an unbiased derivative estimator. One critical feature of SLRIPA is that it needs no additional simulations, i.e., sensitivities with respect to various parameters can be obtained by a *single* run of simulation. We applied SLRIPA to sensitivity analysis for European options and barrier options, and to American option pricing. Simulation results demonstrate the effectiveness of this new derivative estimation technique.

In Chapter 5, we developed a new framework for solving global optimization problems by formulating the global optimization problem as an evolutionary game. By using Lyapunov theory, we showed that a particular equilibrium set of the replicator dynamics in the evolutionary game has a strong connection with the global optimal solutions, and

this equilibrium set is asymptotically stable under mild conditions. Based on this connection, we proposed a model-based evolutionary optimization (MEO) algorithm, which includes the extended CE algorithm as an instantiation. We also proposed a population model-based evolutionary optimization algorithm, which better captures the multimodal property of global optimization problems. Motivated by the ideas of mutation strategies in evolutionary games, we introduced mutation strategies by adding random perturbations to the best model. This enables the PMEO algorithm to escape local optimal solutions effectively. Preliminary numerical tests for PMEO with mutation were reported on some widely used benchmark examples. The way we formulate global optimization problems as evolutionary games provides a new insight into the mechanism for generating new candidate solutions and the mechanism of model updating for model-based global optimization algorithms. For example, one special case of the MEO algorithm gives a new explanation for the CE method. This evolutionary game setting for global optimization problems makes it possible to study the convergence property of model-based algorithms by using analytical tools in the evolutionary game theory literature.

6.2 Future Work

Our research has initiated some new and promising ideas in the field of simulation optimization. There is still room to further refine the proposed methods and to explore new applications. I plan to continue my research along the following lines.

In Chapter 3, we introduced a new simulation-based algorithm to solve the joint inventory control and dynamic pricing problem in revenue management, which is a special case of the stochastic control and impulse control problems. In a general stochastic impulse control problem, times when impulse actions are applied are stopping times and the

176

magnitude of the impulse action is also a random variable. The optimal policies of many stochastic impulse control problems have a threshold structure, i.e., the impulse action is applied when the state of the system enters a certain region. If we can parameterized the optimal policy, it will be possible to apply the technique developed in Chapter 3 to solve these stochastic impulse control problems. Hence, the algorithms in Chapter 3 can be extended to solve more general stochastic control and impulse control problems that have parameterized policies.

The stochastic control problem formulated in Chapter 3, does not have jumps in the underlying dynamics. Stochastic control and impulse control problems with jump diffusion processes arise in many applications, such as finance and management (inventory control). The jump term in the underlying models poses challenges to simulation-based methodologies because of the discontinuity of the sample path. Extending the algorithm in Chapter 3 to problems with jump diffusion models is not straightforward and requires further investigation. How to develop an efficient simulation scheme and gradient estimation technique for these jump-diffusion models will be one of our future research directions.

For the global optimization algorithm MEO, there are several interesting directions for future research. Studying the global optimization problem from an evolutionary game perspective introduces dynamics such as replicator dynamics to study the evolution of the probabilistic models, which also motivates us to consider proving the convergence rate of model-based optimization algorithms by examining the properties of replicator dynamics and using analytical tools in the literature of evolutionary game theory. In one of the instantiations of the MEO algorithm given in Chapter 5, a sum of Dirac functions is used to approximate the probability density function, and it serves as the probabilistic model that is used to generate samples in the algorithm. This model provides a very coarse rep-

resentation of the probability density model, and other methods such as kernel estimation might provide better representations of the probabilistic model. How to construct effective and efficient probabilistic models for MEO is a promising direction to pursue. Another interesting research direction is to extend the MEO algorithm to solve Markov decision processes (MDP), which can be done by modeling MDPs as optimization problems over policy spaces.

Bibliography

[1] V. Aguirregabiria. The dynamics of markups and inventories in retailing firms. *Review of Economic Studies*, 66:275–308, 1999.

[2] Patrick B. *Convergence of Probability Measures.* John Wiley & Sons, Inc., New York, 1999.

[3] S. Bashyam and M. C. Fu. Optimization of (s, S) inventory systems with random lead times and a service level constraint. *Management Science*, 44(12):S243–S256, 1998.

[4] J.A. Bather. A continuous time inventory model. *Journal of Applied Probability*, 3:538–549, 1966.

[5] L. Benkherouf. On a stochastic inventory model with a generalized holding costs. *European Journal of Operations Research*, 182(2):730–737, 2007.

[6] A. Bensoussan, R.H. Liu, and S.P. Sethi. Optimality of an (s, S) policy with compound possion and diffusion demands: a quasi-variational inequalities approach. *SIAM Journal on Control and Optimization*, 44(5):1650–1675, 2005.

[7] A. Bensoussan and C.S. Tapiero. Impulsive control in management: Prospects and applications. *Journal of Optimization Theory and Applications*, 37:419–442, 1982.

[8] D. P. Bertsekas. *Dynamic Programming and Optimal Control, Vols. 1 & 2.* Athena Scientific, 1995.

[9] D. Bertsimas and S. de Boer. Simulation-based booking limits for airline revenue management. *Operations Research*, 53(1):90–106, 2005.

[10] D. Beyer. An inventory model with wiener demand process and positive lead time. *Optimization*, 181–193, 1994.

[11] N.P. Bhatia and G.P. Szegö. *Stability Theory of Dynamical Systems.* Springer Verlag, 1970.

[12] P. Billingsley. *Probability and measure.* Wiley-India, 2008.

[13] G. Bitran and R. Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.

[14] G.R. Bitran and S.V. Mondschein. Periodic pricing of seasonal products in retailing. *Management Science*, 43(1):64–79, 1997.

[15] P.T. De Boer, D.P. Kroese, S. Mannor, and R.Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operation Research*, 134:19–67, 2005.

[16] V.S. Borkar. *Stochastic approximation: a dynamical systems viewpoint*. Cambridge University Press, 2008.

[17] M. Broadie and P. Glasserman. Estimating security price derivatives using simulation. *Management Science*, 42(2):269–285, 1996.

[18] G.W. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13:374–376, 1951.

[19] S. Browne and P. Zipkin. Inventory models with continuous stochastic demands. *The Annals of Applied Probability*, 1:419–435, 1991.

[20] H. Chao. The eoq model with stochastic demand and discounting. *European Journal of Operations Research*, 59:434–443, 1992.

[21] F.Y. Chen, S. Ray, and Y. Song. Optimal pricing and inventory control policy in periodic review systems with fixed ordering cost and lost sales. *Naval Research Logistics*, 53(2):117–136, 2006.

[22] H. Chen, . Q. Wu, and . D. Yao. On the benefit of inventory-based dynamic pricing strategies. *Production and Operations Management*, 19(3):249–260, 2009.

[23] N. Chen and P. Glasserman. Malliavin Greeks without Malliavin calculus. *Stochastic Processes and their Applications*, 117:1689–1723, 2007.

[24] X. Chen and D. Simchi-Levi. Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: the finite horizon case. *Operations Research*, 52(6):887–896, 2004.

[25] G. Constantinides and S. Richard. Existence of optimal simple policies for discounted-cost inventory and cash management in continuous time. *Operations Research*, 26:620–636, 1978.

[26] V.F. Farias and B. Van Roy. Dynamic pricing with a prior on market response. *Operations Research*, 58(1):16, 2010.

[27] A. Federgruen and A. Heching. Combined pricing and inventory control under uncertainty. *Operations Research*, 47(3):454–475, 1999.

[28] Y. Feng and G. Gallego. Perishable asset revenue management with Markovian time dependent demand intensities. *Management Science*, 46(7):941–956, 2000.

[29] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. An application of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3:391–412, 1999.

[30] M. C. Fu. Optimization via simulation: A review. *Annals of Operations Research*, 53:199–248, 1994.

[31] M. C. Fu. Sample path derivatives for (s, S) inventory systems. *Operations Research*, 42(2):351–364, 1994.

[32] M. C. Fu. Gradient estimation. In S. G. Henderson and B. L. Nelson, editors, *Handbooks in Operations Research and Management Science: Simulation*, chapter 19, pages 575–616. Elsevier, 2006.

[33] M. C. Fu, F. W. Glover, and J. April. Simulation optimization: A review, new developments, and applications. In *Proceedings of the 2005 Winter Simulation Conference*, pages 83–95, Piscataway, NJ, 2005. IEEE.

[34] M. C. Fu and J. Q. Hu. Smoothed perturbation analysis derivative estimation for Markov chains. *Operations Research Letters*, 15(5):241–251, 1994.

[35] M. C. Fu and J. Q. Hu. (s, S) inventory systems with random lead times: Harris recurrence and its implications in sensitivity analysis. *Probability in the Engineering and Informational Sciences*, 8(3):355–376, 1994.

[36] M. C. Fu and J. Q. Hu. Sensitivity analysis for Monte Carlo simulation of option pricing. *Probability in the Engineering and Informational Sciences*, 9(3):417–446, 1995.

[37] M. C. Fu and J. Q. Hu. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Kluwer Academic Publishers, 1997.

[38] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. The MIT press, 1998.

[39] G. Gallego and G. Van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8), 1994.

[40] G. Gallego and G. Van Ryzin. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research*, 45(1):24–41, 1997.

[41] A. Garcia, S. Patek, and K. Sinha. A Decentralized Approach for Simulation-based Optimization. *Operations Research*, 55(4):1–16, 2007.

[42] A. Garcia, S.D. Patek, and K. Sinha. A decentralized approach to discrete optimization via simulation: Application to network flow. *Working Paper*, 2006.

[43] P. Glasserman. *Gradient Estimation via Perturbation Analysis*. Kluwer Academic Publishers, Boston, Massachusetts, 1991.

[44] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.

[45] P. Glasserman and S. Tayur. The stability of a capacitated, multi-echelon production-inventory system under a base-stock policy. *Operations Research*, 42:913–925, 1994.

[46] P. Glasserman and S. Tayur. Sensitivity analysis for base-stock levels in multi-echelon production-inventory systems. *Management Science*, 41:263–281, 1995.

[47] F. W. Glover. Tabu search: A tutorial. *Interfaces*, 20:74–94, 1990.

[48] P. W. Glynn. Likelihood ratio gradient estimation: An overview. In *Proceedings of the 1987 Winter Simulation Conference*, pages 366–375, Piscataway, NJ, 1987. IEEE.

[49] W. B. Gong and Y. C. Ho. Smoothed perturbation analysis of discrete-event dynamic systems. *IEEE Transactions on Automatic Control*, AC-32:858–867, 1987.

[50] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[51] B. Heidergott. Option pricing via Monte Carlo simulation - a weak derivative approach. *Probability in the Engineering and Informational Sciences*, 15:335–349, 2001.

[52] Y. C. Ho and X. R. Cao. Optimization and perturbation analysis of queueing networks. *Journal of Optimization Theory and Applications*, 40:559–582, 1983.

[53] Y. C. Ho and X. R. Cao. *Perturbation Analysis and Discrete Event Dynamic Systems.* Kluwer Academic, 1991.

[54] J. Hofbauer, J. Oechssler, and F. Riedel. Brown-von Neumann-Nash dynamics: the continuous strategy case. *Games and Economic Behavior*, 65(2):406–429, 2009.

[55] J. Hofbauer and K. Sigmund. Evolutionary game dynamics. *Bulletin (New Seires) of the American Mathematical Society*, 40(4):479–519, 2003.

[56] L.J. Hong. Estimating quantile sensitivities. *Operations research*, 57(1):118–130, 2009.

[57] L.J. Hong and G. Liu. Pathwise Estimation of Probability Sensitivities Through Terminating or Steady-State Simulations. *Operations Research*, 58(2):357–370, 2010.

[58] J. Hu, H.S. Chang, M.C. Fu, and S.I. Marcus. Dynamic sample budget allocation in model-based optimization. *Journal of Global Optimization*, pages 1–22, 2010.

[59] J. Hu, M. C. Fu, and S. I. Marcus. A model reference adaptive search method for global optimization. *Operations Research*, 55(3):549–568, 2007.

[60] R. Kapuscinski and S.R. Tayur. Optimal policies and simulation based optimization for capacitated production inventory systems. In S.R. Tayur, R. Ganeshan, and M.J. Magazine, editors, *Quantitative Models for Supply Chain Management*, pages 7–40. Kluwer Academic, Boston, MA, 1999.

[61] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag, 1991.

[62] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981.

[63] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[64] A. Kleywegt, A. Shapiro, and T. Homem-de-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12:479–502, 2001.

[65] V.N. Kolokoltsov. *Nonlinear Markov processes and Kinetic Equations*, volume 182. Cambridge Univ Pr, 2010.

[66] S. Kunnumkal and H. Topaloglu. Using Stochastic Approximation Methods to Compute Optimal Base-Stock Levels in Inventory Control Problems. *Operations Research*, 56(3):646–664, 2008.

[67] H. J Kushner and D.S Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, 1978.

[68] H. J Kushner and G. G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, 1997.

[69] H. J Kushner and G. G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, 2nd edition, 2003.

[70] H.J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*. Applications of Mathematics, 24, Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 2001.

[71] T.J. Lambert, M.A. Epelman, and R.L. Smith. A fictitious play approach to large-scale optimization. *Operations Research*, 53(3):477–489, 2005.

[72] P. L'Ecuyer. A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science*, 36:1364–1383, 1990.

[73] A.E.B. Lim and J.G. Shanthikumar. Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research*, 55(2):198–214, 2007.

[74] G. Liu and LJ. Hong. Kernel estimation of the Greeks for options with discontinuous payoffs. *Operations Research*, page forthcoming, 2010.

[75] S. Mannor, R.Y. Rubinstein, and Y. Gat. The cross-entropy method for fast policy search. *International Conference on Machine Learning*, pages 512–519, 2003.

[76] J. Maynard Smith. *Evolutionary and the Theory of Games*. Cambridge Univ, 1982.

[77] J. Maynard Smith and G.R. Price. The logic of animal conflict. *Nature*, 246(5427):15–18, 1973.

[78] A.K. Metwally and A.F. Atiya. Using brownian bridge for fast simulation of jump-diffusion processes and barrier options. *The Journal of Derivatives*, 10:43–54, 2002.

[79] J.R. Michael, W.R. Schucany, and R.W. Haas. Generating random variates using transformations with multiple roots. *American Statistician*, 30(2):88–90, 1976.

[80] H. Mühlenbein and G. Paaß. From recombination of genes to the estimation of distributions: I. binary parameters. In Ingo Rechenberg Hans-Michael Voigt, Werner Ebeling and Hans-Paul Schwefel, editors, *Parallel Problem Solving from Nature - PPSN IV*, pages 178–187. Springer, 1996.

[81] J. Oechssler and F. Riedel. Evolutionary dynamics on infinite strategy spaces. *Economic Theory*, 17(1):141–162, 2001.

[82] J. Oechssler and F. Riedel. On the dynamics foundation of evolutionary stability in continuous models. *Journal of Economic Theory*, 107:223–252, 2002.

[83] H. Pham. On some recent aspects of stochastic control and their applications. *Probability Surveys*, 2:506–549, 2005.

[84] J. D. Pintér. *Global Optimization in Action*. Kluwer, 1996.

[85] M.L. Puterman. A diffsion process model for a storage system. In M.A. Geisler, editor, *Logistics*, pages 143–159. North-Holland Publishing Co., American Elsevier Publishing Co., 1975.

[86] K. Raman and R. Chatterjee. Optimal monopolist pricing under demand uncertainty in dynamic markets. *Management Science*, 41(1):144–162, 1995.

[87] M.I. Reiman and A. Weiss. Sensitivity analysis for simulations via likelihood ratios. *Operations Research*, 37:830–844, 1989.

[88] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[89] R. Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operations Research*, 99:89–112, 1997.

[90] R. Y. Rubinstein and A. Shapiro. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method.* John Wiley & Sons, 1993.

[91] R.Y. Rubinstein. Sensitivity analysis of computer simulation models via the score efficient. *Operations Research*, 37:72–81, 1989.

[92] R.Y. Rubinstein and B. Melamed. *Modern Simulation and Modeling.* Wiley Series in Probability and Statistics, 1998.

[93] L. Shi and S. Ólafsson. Nested partitions method for global optimization. *Operations Research*, 48:390–407, 2000.

[94] L. Shi and S. Ólafsson. Nested partitions method for stochastic optimization. *Methodology and Computing in Applied Probability*, 2:271–291, 2000.

[95] Y. Shoham, R. Powers, and T. Grenager. Multi-agent reinforcement learning: a critical survey. In *AAAI Fall Symposium on Artificial Multi-Agent Learning*, 2004.

[96] S.E. Shreve. *Stochastic Calculus for Finance: Continuous-Time Models.* Birkhäuser, 2004.

[97] M. Srinivas and L. M. Patnaik. Genetic algorithms: A survey. *IEEE Computer*, 27:17–26, 1994.

[98] Hans R. Stoll and Robert E. Whaley. *Futures and Options: Theory and Applications.* Cincinnati, Ohio: South-Western Pub. Co., 1993.

[99] A. Sulem. A solvable one-dimensional model of a diffusion inventory system. *Mathematics of Operations Research*, 11:125–133, 1986.

[100] R. Suri and M. A. Zazanis. Perturbation analysis gives strongly consistent sensitivity estimates for the $M/G/1$ queue. *Management Science*, 34:39–64, 1988.

[101] K.T. Talluri and G. Van Ryzin. *The Theory and Practice of Revenue Management.* Springer Verlag, 2005.

[102] H. Topaloglu. A stochastic approximation method to compute bid prices in network revenue management problems. *INFORMS Journal on Computing*, 20(4):596, 2008.

[103] K. Tuyls and S. Parsons. What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence*, 171(7):406–416, 2007.

[104] G. van Ryzin and G. Vulcano. Simulation-based optimization of virtual nesting controls for network revenue management. *Operations Research*, 56(4):865–880, 2008.

[105] Y. Wang, M.C. Fu, and S.I. Marcus. Sensitivity analysis for barrier options. In *Proceedings of the 2009 Winter Simulation Conference*, pages 1272–1282, 2009.

[106] E. Wei, E.W. Justh, and P.S. Krishnaprasad. Pursuit and an evolutionary game. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 465:1539–1559, 2009.

[107] J.W. Weibull. *Evolutionary Game Theory*. The MIT Press, 1995.

[108] E. Zeidler. *Nonlinear Functional Analysis and Its Applications, Vol. I*. Springer-Verlag, New York, 1986.

[109] E. Zhou, M.C. Fu, and S.I. Marcus. A particle filtering framework for randomized optimization algorithms. In *Proceeds of the 2010 Winter Simulation Conference*, pages 647–654, 2008.

[110] M. Zlochin, M. Birattari, N. Meuleau, and M. Dorigo. Model-based search for combinatorial optimization: A critical survey. *Annals of Operations Research*, 131:373–395, 2004.