



Stochastic Systems

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Importance Sampling for Rainbow Option Pricing

Leila Setayeshgar, Hui Wang

To cite this article:

Leila Setayeshgar, Hui Wang (2026) Importance Sampling for Rainbow Option Pricing. *Stochastic Systems* 16(1):90-107. <https://doi.org/10.1287/stsy.2025.0110>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Stochastic Systems*. Copyright © 2026 The Author(s). <https://doi.org/10.1287/stsy.2025.0110>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2026 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Importance Sampling for Rainbow Option Pricing

Leila Setayeshgar,^a Hui Wang^{b,*}

^aDepartment of Mathematics, The College of Charleston, Charleston, South Carolina 29424; ^bDivision of Applied Mathematics, Brown University, Providence, Rhode Island 02912

*Corresponding author

Contact: setayeshgarl@cofc.edu (LS); hui_wang@brown.edu,  <https://orcid.org/0009-0008-9902-5908> (HW)

Received: April 12, 2025

Revised: August 17, 2025

Accepted: December 15, 2025

Published Online in Articles in Advance:
February 10, 2026

Abstract. This paper studies the applications of state-dependent importance sampling in pricing exotic rainbow options. We demonstrate that for many rainbow options, efficient dynamic importance sampling schemes can be constructed from subsolutions to appropriate partial differential equations. We also introduce an alternative large deviations scaling, which leads to universal importance sampling for rainbow options.

<https://doi.org/10.1287/stsy.2025.0110>



Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Stochastic Systems*. Copyright © 2026 The Author(s). <https://doi.org/10.1287/stsy.2025.0110>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright: © 2026 The Author(s)

Keywords: high-dimensional rainbow options • importance sampling • option pricing

1. Introduction

Importance sampling is a general variance reduction technique in Monte Carlo simulation, and it has numerous applications in computational finance; see, for example, Glasserman et al. (1999, 2008), Glasserman (2004), Guasoni and Robertson (2008), and Wang (2012). In importance sampling, samples are simulated from alternative probability distributions, and an unbiased estimator is formed by multiplying each sample with an appropriate likelihood ratio. The literature on importance sampling is vast. Various methodologies and performance analyses can be found in this very partial list: Sigmund (1976), Heidelberger (1995), Sadowsky (1996), Dupuis and Wang (2004), Blanchet et al. (2007), Rubinstein and Kroese (2007), Bucklew (2010), and the references therein. Although most of the work on importance sampling is empirical, the goal of this paper is to build efficient importance sampling algorithms for the estimation of the price of high-dimensional rainbow options with rigorously provable asymptotically optimal schemes in the context of rare event simulation.

Almost all of the applications of importance sampling to computational finance are concerned with state-independent changes of measure. Within our context, state independence means that the change of measure can only depend on time. In other words, the alternative sampling distribution, which is usually obtained from solving an appropriate variational problem, remains fixed at any given time, regardless of the state of the system (Glasserman et al. 1999, Glasserman 2004, Guasoni and Robertson 2008). For such importance sampling schemes to attain efficiency, certain minimax conditions need to be imposed (see, e.g., Glasserman et al. 1999, condition 2.14 and Guasoni and Robertson 2008, condition 3.7). However, such minimax conditions will not hold in general, especially in the setting of rainbow options because many payoff functions are not convex. In Wang and Zhou (2015), a modified crossentropy scheme to select alternative sampling distributions from a class of mixtures for rainbow option simulation has been introduced. However, the scheme lacks theoretical performance analysis, and it requires pilot samples that may incur nontrivial computational overhead in higher dimensions.

In a series of papers, Dupuis and Wang (2004, 2007, 2009) have established that the change of measure needs to be state dependent (i.e., the change of measure can depend on both time and system state) in order to achieve efficiency in general. Such schemes, named *dynamic importance sampling* (dynamic IS) schemes, are built from suitable subsolutions to relevant Hamilton–Jacobi–Bellman (HJB) equations, and they have found much success in the setting of queueing networks. The goal of this paper is to systematically construct efficient state-dependent importance sampling schemes for rainbow options. We will demonstrate that the framework of dynamic importance sampling through subsolutions can be adapted to the simulation of rainbow options as well. However, this approach has limited success because it requires the option payoff function to satisfy certain conditions (Section 4.3). When applicable, the resulting dynamic importance sampling schemes are explicitly calculable and

are not difficult to implement. Moreover, rigorous performance analysis can be carried out to establish the efficiency or asymptotic optimality of such schemes.

The difficulty of applying this dynamic importance sampling to general rainbow options lies in the construction of appropriate subsolutions to the associated HJB equations. This is largely because of the complicated and high-dimensional nature of a rainbow option's payoff function. In order to circumvent this difficulty, we observe that all of these HJB equations are derived under the classical large deviation embedding, which has played an important role in the design of importance sampling schemes (Glasserman et al. 1999, Dupuis and Wang 2007, Guasoni and Robertson 2008). In this paper, we introduce a novel alternative large deviations embedding, which will reduce the complexity of the problem considerably. In short, it becomes sufficient to only investigate the estimation of probabilities instead of general expected values. Consequently, the construction of the important sampling schemes becomes much easier. The resulting asymptotically optimal (under this alternative embedding) importance sampling schemes are still state dependent, but the dependency is much simpler.

The paper is organized as follows. Section 2 establishes the model for the financial market. In Section 3, we give an overview of importance sampling and asymptotic optimality as well as a literature review. Section 4 considers the traditional large deviations embedding and establishes the connection between subsolutions and asymptotic optimality. We also discuss the construction of subsolutions when the terminal conditions take a certain form. The alternative large deviations embedding and the resulting importance sampling scheme are introduced in Section 5. Examples of numerical simulations are given in Section 6. We conclude the paper with further remarks and heuristic discussions in Sections 7 and 8.

2. Market Model and Rainbow Options

Because we are only interested in option pricing, we assume that the probability space (Ω, \mathcal{F}) is equipped with the risk-neutral probability measure P . Let r be the constant risk-free interest rate. Consider a financial market with d risky assets whose prices at time t are denoted by a column vector $S(t) := [S_1(t), \dots, S_d(t)]'$. We will assume that the financial market is complete and that stock prices are geometric Brownian motions. More precisely, let $B := \{B(t) = [B_1(t), \dots, B_d(t)]' : t \geq 0\}$ be a d -dimensional Brownian motion with strictly positive definite covariance matrix $\Sigma := [\Sigma_{ij}]$ of size $d \times d$ and $\Sigma_{ii} = 1$ for every $i = 1, 2, \dots, d$. Let $\sigma := [\sigma_1, \dots, \sigma_d]'$ be a column vector of strictly positive constants that denotes the volatilities of stock prices. For $i = 1, 2, \dots, d$, the stock price S_i is modeled by the geometric Brownian motion

$$S_i(t) = S_i(0) \exp \left\{ \left(r - \frac{1}{2} \sigma_i^2 \right) t + \sigma_i B_i(t) \right\}.$$

We are interested in estimating the price of a rainbow option with expiration date T and payoff $h(S(T)) := h(S_1(T), \dots, S_d(T))$, where $h : \mathbb{R}_+^d \rightarrow [0, \infty)$ is a nonnegative function defined on the open positive orthant

$$\mathbb{R}_+^d := \{x := [x_1, \dots, x_d]' \in \mathbb{R}^d : x_i > 0, i = 1, \dots, d\}.$$

The option price is simply the expected value $v := E[e^{-rT}h(S(T))]$. Prices of most rainbow options do not admit explicit formulae. The goal of our paper is to construct efficient importance sampling schemes for simulating such option prices.

For future analysis, it is convenient to introduce the log stock price process $X(t) := [X_1(t), \dots, X_d(t)]'$, where for each $i = 1, 2, \dots, d$,

$$X_i(t) := \log S_i(t) = \log S_i(0) + \left(r - \frac{1}{2} \sigma_i^2 \right) t + \sigma_i B_i(t).$$

In particular, letting Γ be a matrix such that $\Gamma\Gamma' := \Sigma$ (see Remark 1), we can rewrite $X(T) = A + C\Gamma Z$, where

$$A := [A_1, \dots, A_d]', \quad A_i := \log S_i(0) + \left(r - \frac{1}{2} \sigma_i^2 \right) T, \quad (1)$$

C is a $d \times d$ diagonal matrix with $C_{ii} = \sigma_i$ for $i = 1, \dots, d$, and $Z := [Z_1, \dots, Z_d]'$ is a d -dimensional normal random vector with distribution $N(0, TI_d)$. Therefore, the discounted payoff can be expressed as

$$e^{-rT}h(S(T)) = e^{-rT}h(\exp\{A + C\Gamma Z\}) := G(Z), \quad (2)$$

where we have adopted the notation that $\exp\{x\} := [\exp\{x_1\}, \dots, \exp\{x_d\}]'$ for any vector $x := [x_1, \dots, x_d]' \in \mathbb{R}^d$.

Finally, we would like to impose some mild technical conditions on the payoff function h , which are satisfied by nearly all rainbow options. Throughout the paper, we will assume that h is of linear growth; that is, there exist

some nonnegative constants a and b such that

$$0 \leq h(s) \leq a + b\|s\| \tag{3}$$

for every $s \in \mathbb{R}_+^d$. Under this assumption, the option price v is finite. We will assume that the set $D := \{h > 0\}$ is either open or closed and that there is some continuous nonnegative function $\bar{h} : \mathbb{R}_+^d \rightarrow [0, \infty)$ such that $\bar{h} = h$ on $\{h > 0\}$ or equivalently, $h = \bar{h}1_{\{h>0\}}$. The choice of \bar{h} is not essential, and we can choose \bar{h} so that it also satisfies the growth Condition (3); indeed, one can otherwise redefine \bar{h} to be the minimum of itself with $a + b\|s\|$.

Remark 1. When Σ is strictly positive definite, there is a particularly convenient choice of Γ based on Cholesky factorization (Glasserman 2004). In this case, Γ is lower triangular, and its components are explicitly available.

Remark 2. Even though our analysis is largely based on market models where the stock prices are geometric Brownian motions, this restriction is not essential. The idea is applicable to more general jump diffusion models. We will discuss this extension with more details in Section 7.

3. Overview of Importance Sampling

Importance sampling is a general variance reduction technique for Monte Carlo simulation. It is particularly effective in the context of rare event simulation. The idea is to generate samples from a different probability distribution and multiply each outcome with an appropriate likelihood ratio so that the resulting estimator is unbiased.

Hereon, we will equip the sample space (Ω, \mathcal{F}) with probability measures other than the original risk-neutral probability measure P . The expected value with respect to P will simply be denoted by $E[\cdot]$, whereas the expected value with respect to some other probability measure, say Q , will be denoted by $E^Q[\cdot]$. If those alternative probability measures are indexed by some parameter $\varepsilon > 0$, say $Q = P_\varepsilon$, then the expected value with respect to those probabilities will be denoted by $E^\varepsilon[\cdot]$ for simplicity when no confusion is incurred.

Suppose one is interested in estimating the expected value $v := E[X]$ of some random variable X . Let Q be an alternative probability measure such that $P \ll Q$. Then,

$$v = E^Q \left[X \frac{dP}{dQ} \right].$$

In other words, importance sampling produces an unbiased estimator by generating samples from an alternative probability measure Q and multiplying each sample of X by the likelihood ratio dP/dQ . The key in the design of importance sampling algorithms is the choice of the alternative sampling distribution Q . When appropriately designed, importance sampling can be very efficient, reducing the variance of the estimator by orders of magnitude. It is particularly useful in the estimation of probabilities or expected values related to rare events.

A commonly used criterion for selecting the alternative sampling distribution is the so-called ‘‘asymptotic optimality’’ (Sigmund 1976, Heidelberger 1995). Its definition requires that the present estimation problem be embedded into a sequence of estimation problems, where certain types of large deviations convergence results hold. To fix ideas, suppose that we wish to estimate the expected values of nonnegative random variables $\{X_\varepsilon\}$ indexed by $\varepsilon > 0$ (the original problem corresponds to some ε ; i.e., $X = X_\varepsilon$ for some ε). Denote $v_\varepsilon := E[X_\varepsilon]$. We assume that a large deviation type of asymptotics exists; that is, there exists a constant $\gamma \in \mathbb{R}$ such that

$$\gamma := \lim_{\varepsilon \rightarrow 0} \varepsilon \log v_\varepsilon.$$

Let P_ε be a new probability measure such that $P \ll P_\varepsilon$. The importance sampling estimator denoted by \hat{v}_ε is given by

$$\hat{v}_\varepsilon = X_\varepsilon \frac{dP}{dP_\varepsilon},$$

where dP/dP_ε denotes the Radon–Nikodým derivative or the likelihood ratio. Because the estimator is unbiased under P_ε , minimizing its variance is equivalent to minimizing its second moment. However, by the Jensen inequality and the unbiasedness of \hat{v}_ε , we have the lower bound

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log E^\varepsilon [\hat{v}_\varepsilon^2] \geq \liminf_{\varepsilon \rightarrow 0} 2\varepsilon \log E^\varepsilon [\hat{v}_\varepsilon] = \liminf_{\varepsilon \rightarrow 0} 2\varepsilon \log v_\varepsilon = 2\gamma.$$

Definition 1 (Asymptotic Optimality). An importance sampling scheme or change of measure $\{P_\varepsilon\}$ is said to be *asymptotically optimal* if this lower bound is achieved: that is, if

$$\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E^\varepsilon[\hat{\sigma}_\varepsilon^2] \leq 2\gamma.$$

Remark 3. To put everything into the framework of asymptotic optimality, we need to embed the original estimation problem into a sequence of estimation problems. Such embedding is not unique, which is a useful observation that we will exploit later on to deal with certain rainbow options.

4. Efficient Importance Sampling

It has been established in Glasserman and Wang (1997) and Dupuis and Wang (2004, 2007) that state-independent changes of measure will not lead to efficient importance sampling schemes in general, whereas asymptotically optimal state-dependent schemes can be constructed from classical subsolutions to related Hamilton–Jacobi–Bellman equations. For the financial market model under consideration, the stock prices are driven by Brownian motion B . Under an alternative sampling distribution that we consider, the Brownian motion B will become a Brownian motion with drift. Therefore, we expect that asymptotically optimal changes of measure correspond to state-dependent drifts in general.

4.1. The Classical Large Deviation Embedding

Recall the definition of G in (2). The goal is to estimate the option price $v := E[G(Z)]$, where $Z := [Z_1, \dots, Z_d]'$ is a d -dimensional normal random vector with distribution $N(0, T\mathbb{I}_d)$. Following the setup in Glasserman et al. (1999), define $F(z) := \log G(z)$ for any $z \in \mathbb{R}^d$. Note that F takes values in $\mathbb{R} \cup \{-\infty\}$ because G can be zero. Now, fix any $\varepsilon > 0$. Define

$$v_\varepsilon := E[e^{F(\sqrt{\varepsilon}Z)/\varepsilon}]. \quad (4)$$

The option price of interest corresponds to $\varepsilon = 1$. Under very mild conditions, one can establish the large deviations or exponential growth/decay rate of v_ε . The following result is a slight generalization of the Varadhan integral lemma, and its proof can be found in Glasserman et al. (1999, lemma 2.2).

Lemma 1. Suppose that $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ is continuous and satisfies the growth condition $f(z) \leq c_1 + c_2 \|z\|^2$ for some $c_2 < 1/2$ and for all $z \in D$, where $D \subseteq \mathbb{R}^d$ is either open or closed. Let Z be a d -dimensional normal random vector with distribution $N(0, T\mathbb{I}_d)$. Then,

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log E[e^{f(\sqrt{\varepsilon}Z)/\varepsilon} 1_D(\sqrt{\varepsilon}Z)] = \sup_{z \in D} [f(z) - \|z\|^2/2T].$$

Under our assumptions, $h = \bar{h}1_{\{h>0\}}$ for some nonnegative, continuous function $\bar{h} : \mathbb{R}_+^d \rightarrow [0, \infty)$ that satisfies the linear growth Condition (3). Now, define \bar{G} according to (2) similarly with h replaced by \bar{h} . Finally let $\bar{F} := \log \bar{G}$. It is not difficult to check that \bar{F} is continuous and satisfies the growth condition of Lemma 1; indeed, $\bar{F}(z) \leq c_1 + c_2 \|z\|^2$ for some c_1, c_2 and every $z \in \mathbb{R}^d$. Because h and \bar{h} agree on the set $\{h > 0\}$, it follows that $\bar{G} = G$ and thus, $\bar{F} = F$ on the set $D := \{F > -\infty\}$. Furthermore, because $\{h > 0\}$ is either open or closed and $z \mapsto \exp\{A + Cz\}$ is a continuous bijection from \mathbb{R}^d to \mathbb{R}_+^d , D is either open or closed. Thus, we may apply Lemma 1 to obtain that

$$\begin{aligned} \gamma &:= \lim_{\varepsilon \rightarrow 0} \varepsilon \log v_\varepsilon = \lim_{\varepsilon \rightarrow 0} \varepsilon \log E[e^{F(\sqrt{\varepsilon}Z)/\varepsilon}] \\ &= \lim_{\varepsilon \rightarrow 0} \varepsilon \log E[e^{F(\sqrt{\varepsilon}Z)/\varepsilon} 1_D(\sqrt{\varepsilon}Z)] \\ &= \lim_{\varepsilon \rightarrow 0} \varepsilon \log E[e^{\bar{F}(\sqrt{\varepsilon}Z)/\varepsilon} 1_D(\sqrt{\varepsilon}Z)] \\ &= \sup_{z \in D} [\bar{F}(z) - \|z\|^2/2T] \\ &= \sup_{z \in D} [F(z) - \|z\|^2/2T] \\ &= \sup_{z \in \mathbb{R}^d} [F(z) - \|z\|^2/2T]. \end{aligned} \quad (5)$$

Finally, we observe that the supremum is always attained at some $z^* \in \bar{D}$ when F is continuous on \bar{D} because of the linear growth condition of F .

4.2. HJB Equation and Subsolution

Even though Dupuis and Wang (2004, 2007) are mostly concerned with discrete time dynamics, the technique therein can be naturally adapted to the diffusion setting. We should omit these repetitive heuristic details of deriving the HJB equation and directly state the definition of a subsolution. To this end, define $W := \Gamma^{-1}B$. Then, W is a d -dimensional standard Brownian motion. Define the scaled state process indexed by ε :

$$W_\varepsilon(t) := \sqrt{\varepsilon}W(t) = \sqrt{\varepsilon}\Gamma^{-1}B(t), \quad t \geq 0.$$

Then, the quantity of interest can be written as

$$v_\varepsilon = E[e^{F(\sqrt{\varepsilon}Z)/\varepsilon}] = E[e^{F(W_\varepsilon(T))/\varepsilon}].$$

We say a continuously differentiable function $V : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a *classical subsolution with bounded gradient* if ∇V is uniformly bounded and if V satisfies the partial differential inequality (see Remark 5)

$$\frac{\partial V}{\partial t} + \inf_{u \in \mathbb{R}^d} \left[\frac{1}{2} \|u\|^2 + \langle \nabla V, u \rangle \right] = \frac{\partial V}{\partial t} - \frac{1}{2} \|\nabla V\|^2 \geq 0 \quad (6)$$

for any $\forall(t, x) \in [0, T] \times \mathbb{R}^d$ with terminal condition

$$V(T, x) \leq -F(x), \quad \forall x \in \mathbb{R}^d. \quad (7)$$

Note that the term $\|u\|^2/2$ is the large deviation rate function for the standard d -dimensional normal distribution $N(0, I_d)$ and is closely associated with the large deviations of d -dimensional standard Brownian motion $\{W^\varepsilon\}$ (Dembo and Zeitouni 1998, theorem 5.2.3). The change of measure corresponding to a classical subsolution V is given by Dupuis and Wang (2007):

$$u^*(t, x) := -\nabla V(t, x). \quad (8)$$

More precisely, the alternative sampling distribution is defined by the probability measure P_ε , where

$$\frac{dP_\varepsilon}{dP} := \exp \left\{ \frac{1}{\sqrt{\varepsilon}} \int_0^T u^*(t, W_\varepsilon(t)) dW(t) - \frac{1}{2\varepsilon} \int_0^T \|u^*(t, W_\varepsilon(t))\|^2 dt \right\}. \quad (9)$$

Because u^* is assumed to be bounded, the right-hand side indeed defines a probability measure. By the Girsanov theorem (Karatzas and Shreve 1991), the process \bar{W} defined by

$$d\bar{W}(t) := dW(t) - \frac{1}{\sqrt{\varepsilon}} u^*(t, W_\varepsilon(t)) dt$$

is a standard Brownian motion under P_ε . The corresponding importance sampling estimate for v_ε is simply

$$\hat{v}_\varepsilon = e^{F(W_\varepsilon(T))/\varepsilon} \frac{dP}{dP_\varepsilon}.$$

The performance of this estimate is characterized by the following theorem, which is essentially a continuous-time version of Dupuis and Wang (2007, theorem 8.1). This is the key result in the subsolution approach to importance sampling.

Theorem 1. *Let $V : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a classical subsolution to the HJB Equations (6) and (7) with bounded gradient. If V is also twice continuously differentiable with respect to x with uniformly bounded Hessian matrix, then*

$$\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E^\varepsilon[\hat{v}_\varepsilon^2] \leq -2V(0, 0).$$

Proof. Let $\nabla^2 V := [\nabla_{ij}^2 V]$ denote the $d \times d$ Hessian matrix of V and C be a constant such that $|\nabla^2 V_{ij}(t, x)| \leq C$ for every $i, j = 1, \dots, d$ and $(t, x) \in [0, T] \times \mathbb{R}^d$. Define the process $M := \{M_t : t \in [0, T]\}$ with

$$M_t := \exp \left\{ -\frac{2}{\varepsilon} V(t, W_\varepsilon(t)) - \frac{1}{\sqrt{\varepsilon}} \int_0^t u^*(s, W_\varepsilon(s)) dW(s) + \frac{1}{2\varepsilon} \int_0^t \|u^*(s, W_\varepsilon(s))\|^2 ds - Cdt \right\}. \quad (10)$$

Because of the boundedness of all first-order derivatives of V , V is of linear growth in both t and x . Therefore, it is not difficult to see that $E[M_t^2] \leq K$ for some constant K and all $t \in [0, T]$. Applying the Itô formula and using (8), we have

$$\begin{aligned} \frac{dM_t}{M_t} &= \frac{2}{\varepsilon} \left[-\frac{\partial V}{\partial t} + \|\nabla V\|^2 + \frac{1}{2} \|u^*\|^2 + \langle \nabla V, u^* \rangle \right] dt - (Cd + \text{tr} \nabla^2 V) dt - \frac{1}{\sqrt{\varepsilon}} [2\nabla V + u^*] dW(t) \\ &= \frac{2}{\varepsilon} \left[-\frac{\partial V}{\partial t} + \frac{1}{2} \|\nabla V\|^2 \right] dt - (Cd + \text{tr} \nabla^2 V) dt - \frac{1}{\sqrt{\varepsilon}} [2\nabla V + u^*] dW(t), \end{aligned}$$

with the understanding that all of the coefficients on the right-hand side are taking values at $(t, W_\varepsilon(t))$. It follows that M is indeed a local supermartingale because the coefficients of dt are negative by the subsolution property of V and that $\|\nabla_{ii} V\|_\infty \leq C$ for all $i = 1, \dots, d$. Because M is nonnegative, it is thus a true supermartingale. In particular, $E[M_0] \geq E[M_T]$. Combined with the terminal Condition (7), it leads to

$$\begin{aligned} e^{-2V(0,0)/\varepsilon} &\geq e^{-CdT} E \left[e^{-2V(T, W_\varepsilon(T))/\varepsilon} \frac{dP}{dP_\varepsilon} \right] \\ &\geq e^{-CdT} E \left[e^{2F(W_\varepsilon(T))/\varepsilon} \frac{dP}{dP_\varepsilon} \right] \\ &= e^{-CdT} E^\varepsilon \left[e^{2F(W_\varepsilon(T))/\varepsilon} \left(\frac{dP}{dP_\varepsilon} \right)^2 \right] = e^{-CdT} E^\varepsilon [\hat{\sigma}_\varepsilon^2]. \end{aligned}$$

Now, taking logarithm on both sides, multiplying with ε , and then, sending ε to zero, we complete the proof. \square

Remark 4. Theorem 1 asserts that if $V(0,0) = -\gamma$, then the importance sampling estimate is asymptotically optimal. In many practical applications, $V(0,0)$ can be chosen to be arbitrarily close to $-\gamma$ for a classical subsolution V .

Remark 5. The equation should be viewed as an Isaacs equation for interpretation. The interesting point of this Isaacs equation is that it is equivalent to the HJB equation by a scaling of two, which is the intrinsic reason why state-dependent importance sampling schemes can attain asymptotic optimality. In this paper, we simply use the form of the equivalent HJB equation. Details of the Isaacs equation may be found in Dupuis and Wang (2004, 2007).

4.3. Construction of Classical Subolutions

Because of Theorem 1, building an efficient importance sampling scheme amounts to constructing a classical subsolution V of (6) and (7) with $V(0,0)$ as close to γ as possible. In many applications, such subsolutions can be obtained as a mollification of piecewise affine weak subsolutions (Dupuis and Wang 2007).

Throughout this section, we assume that F is continuous on \bar{D} , where $D := \{F > -\infty\}$. Recall the definition of γ in (5), and denote by $z^* \in \bar{D}$ the maximizing point. We start with the simplest case where F is concave. Define $V : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$V(t, x) := -\frac{1}{T} \langle z^*, x \rangle - \frac{1}{2T^2} (T-t) \|z^*\|^2 - \left[F(z^*) - \frac{1}{T} \|z^*\|^2 \right]. \quad (11)$$

We claim that V is a classical subsolution. Indeed, it satisfies (6) with equality. As for the terminal Inequality (7), we observe that $z^*/T \in \partial F$ because z^* attains the supremum of (5). Therefore, by concavity,

$$F(x) - F(z^*) \leq \frac{1}{T} \langle z^*, x - z^* \rangle,$$

which is exactly (7). Finally, $V(0,0) = -F(z^*) + \|z^*\|^2/2T = -\gamma$, and thus, the corresponding state-independent importance sampling scheme is asymptotically optimal.

A more common situation is that F is the maximum of a concave function. For many options, the corresponding F can be written in this form, such as multistrike call options. Assume that $F := F_1 \vee \dots \vee F_m$, where F_i is a concave function for each $i = 1, \dots, m$. Define for each i

$$V_i(t, x) := -\frac{1}{T} \langle z^i, x \rangle - \frac{1}{2T^2} (T-t) \|z^i\|^2 - \left[F_i(z^i) - \frac{1}{T} \|z^i\|^2 \right], \quad (12)$$

where $z^i \in \mathbb{R}^d$ is the maximizer for (5) with F_i in place of F . Note that each V_i satisfies the HJB Equation (6) with equality and $V_i(T, x) \leq -F_i(x)$. Define $V := V_1 \wedge \dots \wedge V_m$. Then, V satisfies (6) wherever it is differentiable with $V(T, x) \leq -F(x)$. Moreover,

$$\begin{aligned} -V(0, 0) &= \max_{i=1, \dots, m} \left[F_i(z^i) - \frac{1}{2T} \|z^i\|^2 \right] = \max_{i=1, \dots, m} \max_{z \in \mathbb{R}^d} \left[F_i(z) - \frac{1}{2T} \|z\|^2 \right] \\ &= \max_{z \in \mathbb{R}^d} \max_{i=1, \dots, m} \left[F_i(z) - \frac{1}{2T} \|z\|^2 \right] = \max_{z \in \mathbb{R}^d} \left[F(z) - \frac{1}{2T} \|z\|^2 \right] = \gamma. \end{aligned}$$

However, we cannot apply Theorem 1 directly because V is not differentiable everywhere. To do so, one can mollify the (weak) subsolution V to obtain a classical subsolution. To fix ideas, let δ be a small positive number, and define

$$V^\delta(t, x) := -\delta \log \left(\sum_{i=1}^m e^{-V_i(t, x)/\delta} \right). \quad (13)$$

Then, V^δ is a continuously differentiable function, and

$$\nabla V^\delta(t, x) = \sum_{i=1}^m \rho_i^\delta(t, x) \nabla V_i(t, x), \quad \frac{\partial V^\delta}{\partial t}(t, x) = \sum_{i=1}^m \rho_i^\delta(t, x) \frac{\partial V_i}{\partial t}(t, x) \quad (14)$$

with

$$\rho_i^\delta(t, x) := \frac{e^{-V_i(t, x)/\delta}}{\sum_{j=1}^m e^{-V_j(t, x)/\delta}}. \quad (15)$$

It is easy to verify because of the convexity of the mapping $x \mapsto \|x\|^2$ for $x \in \mathbb{R}^d$ that V^δ satisfies the subsolution property (6). Furthermore, by the definitions of V^δ and V ,

$$V^\delta(T, x) \leq -\delta \log(e^{-V(T, x)/\delta}) = V(T, x) \leq -F(x)$$

and

$$V^\delta(0, 0) \geq -\delta \log(me^{-V(0, 0)/\delta}) = V(0, 0) - \delta \log m = -(\gamma + \delta \log m).$$

In other words, V^δ is a classical subsolution, where $V(0, 0)$ is only $2\delta \log m$ away from asymptotic optimality (see Remark 6). In practice, δ is chosen to be small to guarantee good numerical performance.

Remark 6. To achieve optimality, one can let δ depend on ε , and a result analogous to Theorem 1 can be obtained. More precisely, if $\delta \rightarrow 0$ and $\delta/\varepsilon \rightarrow \infty$, then the corresponding importance sampling scheme is asymptotically optimal. The proof is verbatim to that of Theorem 1 except that the trace of V^δ is now bounded by C/δ for some constant C , which is the reason behind condition $\delta/\varepsilon \rightarrow \infty$.

5. An Alternative Embedding and Importance Sampling

In this section, we investigate an alternative large deviations embedding and the resulting importance sampling scheme.

5.1. Motivation and Overview

The classical large deviation embedding introduced by (4) and the subsequent subsolution approach yield a general framework for constructing asymptotically optimal importance sampling schemes. We have shown that if the payoff of a rainbow option can be expressed as the maximum of concave functions, the construction of subsolutions can be quite straightforward. However, many rainbow options (even simple ones, such as basket call) do not belong to this category. For those options, it is unclear, particularly in high dimensions, how to find systematic ways to construct subsolutions that lead to asymptotically optimal importance sampling schemes.

This prompts us to introduce a large deviation embedding other than (4). A distinct advantage of this alternative embedding is that the resulting importance sampling scheme is very easy to construct because it does not depend on the specific form of the payoff function. Rather, the construction will *only* rely on the distance from the origin to the region where the payoff function is strictly positive. For this reason, such importance sampling schemes are said to be “universal” (Dupuis and Wang 2004). They are still state-dependent importance sampling schemes, but the dependence is minimal; see (20).

5.2. An Alternative Large Deviation Embedding

To describe this new approach, we recall the definition of G in (2) and that of $D := \{G > 0\}$. Consider the following large deviations scaling. For any $\varepsilon > 0$, define

$$v_\varepsilon := E[G(\sqrt{\varepsilon}Z)], \quad (16)$$

where Z is a d -dimensional normal random vector with distribution $N(0, TI_d)$. The quantity of interest corresponds to $\varepsilon = 1$. Here, we have abused the notation as v_ε is also used in the classical embedding. But, there should be no ambiguity given the context. We also define the set (which actually corresponds to a binary call)

$$A_\varepsilon := \{z \in \mathbb{R}^d : G(\sqrt{\varepsilon}z) > 0\} = D/\sqrt{\varepsilon}$$

for every $\varepsilon > 0$. The heuristic is that a good importance sampling change of measure for estimating $P(A_\varepsilon)$ should also be good for estimating v_ε . Even though A_ε may not be convex itself or the union of convex sets, one can design universal importance sampling (universal IS) schemes for it (Dupuis and Wang 2007). We adopt the following mild assumptions throughout this section.

Assumption 1. D is nonempty, and $\bar{D}^\circ = \bar{D}$, where D° denotes the interior of D .

Assumption 2. G is continuous on \bar{D} .

Assumption 3. For some positive constants a and b , $|G(z)| \leq ae^{b\|z\|}$ for all $z \in \mathbb{R}^d$.

Assumption 4. \bar{D} does not contain the origin.

We would like to remark that Assumption 4 is imposed only to avoid the trivial case where importance sampling is not really necessary. It will not affect the theoretical results in any way.

Lemma 2. Let Z be a d -dimensional normal random vector with distribution $N(0, TI_d)$. Then, we have

$$\gamma := \lim_{\varepsilon \rightarrow 0} \varepsilon \log v_\varepsilon = \lim_{\varepsilon \rightarrow 0} \varepsilon \log P(Z \in A_\varepsilon) = -\frac{1}{2T} \inf_{z \in \bar{D}} \|z\|^2, \quad (17)$$

where the infimum is achieved at some $z^* \in \partial D$.

Proof. Note that the last equality in (17) follows from classic large deviations and Assumption 1. It is trivial that the infimum is attained at some $z^* \in \partial D$ because D is nonempty.

To analyze v_ε , we fix an arbitrarily small $\delta > 0$ and define the set $D_\delta := \{z \in \mathbb{R}^d : G(z) > \delta\}$. Note that D_δ° is nonempty for δ small enough because D° is nonempty. It follows that $v_\varepsilon \geq \delta P(\sqrt{\varepsilon}Z \in D_\delta)$, and classical large deviations theory yields

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log v_\varepsilon \geq \liminf_{\varepsilon \rightarrow 0} [\varepsilon \log \delta + \varepsilon \log P(\sqrt{\varepsilon}Z \in D_\delta)] \geq -\frac{1}{2T} \inf_{z \in D_\delta^\circ} \|z\|^2.$$

Now, letting $\delta \downarrow 0$, it is straightforward to show by Assumption 1 that $D_\delta^\circ \uparrow D^\circ$, and thus,

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log v_\varepsilon \geq -\frac{1}{2T} \inf_{z \in D^\circ} \|z\|^2 = -\frac{1}{2T} \inf_{z \in \bar{D}} \|z\|^2.$$

As for the other direction, take an arbitrarily large constant $M > 0$, and observe that

$$G(\sqrt{\varepsilon}Z) \leq G(\sqrt{\varepsilon}Z)1_{\{G(\sqrt{\varepsilon}Z) > M\}} + M1_{A_\varepsilon}(Z). \quad (18)$$

It follows that

$$\begin{aligned} \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log v_\varepsilon &\leq \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log \{E[G(\sqrt{\varepsilon}Z)1_{\{G(\sqrt{\varepsilon}Z) > M\}}] + MP(Z \in A_\varepsilon)\} \\ &= \max \left\{ \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[G(\sqrt{\varepsilon}Z)1_{\{G(\sqrt{\varepsilon}Z) > M\}}], \right. \\ &\quad \left. \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log [MP(Z \in A_\varepsilon)] \right\}. \end{aligned}$$

Because the second term on the right-hand side equals $\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log P(Z \in A_\varepsilon)$, it remains to show that

$$\lim_{M \rightarrow \infty} \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[G(\sqrt{\varepsilon}Z)1_{\{G(\sqrt{\varepsilon}Z) > M\}}] = -\infty.$$

Because of the growth condition of G , it is sufficient to show that

$$\lim_{M \rightarrow \infty} \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[ae^{b\sqrt{\varepsilon}\|Z\|} 1_{\{ae^{b\sqrt{\varepsilon}\|Z\|} > M\}}] = -\infty.$$

Observe that by the Hölder inequality,

$$E[ae^{b\sqrt{\varepsilon}\|Z\|} \mathbf{1}_{\{ae^{b\sqrt{\varepsilon}\|Z\|} > M\}}] \leq \{E[a^2 e^{2b\sqrt{\varepsilon}\|Z\|}]\} P(ae^{b\sqrt{\varepsilon}\|Z\|} > M)^{1/2}.$$

By the dominated convergence theorem, $E[a^2 e^{2b\sqrt{\varepsilon}\|Z\|}] \rightarrow a^2 e^{2b}$ as $\varepsilon \rightarrow 0$. Furthermore, classical large deviations theory implies that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \varepsilon \log P(ae^{b\sqrt{\varepsilon}\|Z\|} \geq M) &= -\frac{1}{2T} \inf\{\|z\|^2 : ae^{b\|z\|} \geq M\} \\ &= -\frac{1}{2Tb^2} \left(\log \frac{M}{a}\right)^2 \rightarrow -\infty \end{aligned}$$

as $M \rightarrow \infty$. This completes the proof. \square

5.3. Asymptotic Optimality

Below is the main result of the paper. It establishes the connection between the estimation of v_ε and that of $\{P(Z \in A_\varepsilon)\}$, and it confirms the heuristic that a good importance sampling scheme for the latter is also good for the former. In particular, if an importance sampling scheme is asymptotically optimal for estimating $\{P(Z \in A_\varepsilon)\}$, then it is also asymptotically optimal for estimating $\{v_\varepsilon\}$ under some mild conditions.

Theorem 2. Assume that $\{P_\varepsilon\}$ is asymptotically optimal for the estimation of $\{P(Z \in A_\varepsilon)\}$. If for some $q > 1$,

$$\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[1_{A_\varepsilon}(Z)(dP/dP_\varepsilon)^q] < \infty,$$

then $\{P_\varepsilon\}$ is asymptotically optimal for the estimation of $\{v_\varepsilon\}$ as well.

Proof. Denote the likelihood $L_\varepsilon := dP/dP_\varepsilon$. The importance sampling estimator is just

$$\hat{v}_\varepsilon = G(\sqrt{\varepsilon}Z)L_\varepsilon,$$

where Z is sampled from P_ε . Because of Lemma 2, it is sufficient to show that

$$\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E^\varepsilon[\hat{v}_\varepsilon^2] \leq 2\gamma, \quad (19)$$

where γ is defined to be the right-hand side of (17). Plugging in the form of \hat{v}_ε , the left-hand side (LHS) of (19) equals

$$\text{LHS} = \lim_{\varepsilon \rightarrow 0} \varepsilon \log E[\hat{v}_\varepsilon^2 L_\varepsilon^{-1}] = \lim_{\varepsilon \rightarrow 0} \varepsilon \log E[G^2(\sqrt{\varepsilon}Z)L_\varepsilon].$$

Take any large number M . It follows from (18) and $(a+b)^2 \leq 2a^2 + 2b^2$ that

$$G^2(\sqrt{\varepsilon}Z) \leq 2G^2(\sqrt{\varepsilon}Z)\mathbf{1}_{\{G(\sqrt{\varepsilon}Z) > M\}} + 2M^2\mathbf{1}_{A_\varepsilon}(Z).$$

By assumption, $\{P_\varepsilon\}$ is asymptotically optimal for estimating $\{P(Z \in A_\varepsilon)\}$. It follows that

$$\begin{aligned} \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[2M^2\mathbf{1}_{A_\varepsilon}(Z)L_\varepsilon] &= \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[1_{A_\varepsilon}(Z)L_\varepsilon] \\ &= \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E^\varepsilon[1_{A_\varepsilon}(Z)L_\varepsilon^2] \end{aligned}$$

Now, for the other term, we can again use the Hölder inequality. Define α such that $1/\alpha + 1/q = 1$. Then,

$$\begin{aligned} &\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[2G^2(\sqrt{\varepsilon}Z)\mathbf{1}_{\{G(\sqrt{\varepsilon}Z) > M\}}L_\varepsilon] \\ &= \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[G^2(\sqrt{\varepsilon}Z)\mathbf{1}_{\{G(\sqrt{\varepsilon}Z) > M\}}\mathbf{1}_{A_\varepsilon}(Z)L_\varepsilon] \\ &\leq \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log \{E[G^{2\alpha}(\sqrt{\varepsilon}Z)\mathbf{1}_{\{G(\sqrt{\varepsilon}Z) > M\}}]\}^{1/\alpha} \{E[1_{A_\varepsilon}(Z)L_\varepsilon^q]\}^{1/q} \\ &\leq \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon/\alpha \log E[G^{2\alpha}(\sqrt{\varepsilon}Z)\mathbf{1}_{\{G(\sqrt{\varepsilon}Z) > M\}}] + \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon/q \log E[1_{A_\varepsilon}(Z)L_\varepsilon^q]. \end{aligned}$$

Because of the assumption, the second term on the right-hand side cannot take a value of ∞ . Thus, it remains to show that

$$\lim_{M \rightarrow \infty} \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[G^{2\alpha}(\sqrt{\varepsilon}Z)\mathbf{1}_{\{G(\sqrt{\varepsilon}Z) > M\}}] = -\infty.$$

This is almost verbatim to the second half of the proof to Lemma 2, and we omit the details. Therefore, Inequality (19) follows readily. This completes the proof. \square

5.4. Universal Importance Sampling

To apply Theorem 2, we need efficient importance sampling schemes for estimating $\{P(Z \in A_\varepsilon)\}$. Here, we investigate the universal schemes brought up in Dupuis and Wang (2007, section 10.7) and argue that it is indeed asymptotically optimal for the estimation of $\{P(Z \in A_\varepsilon)\}$ as well as $\{v_\varepsilon\}$.

Recall $D := \{G > 0\}$ and γ in (17). Fix any vector $\theta \in \mathbb{R}^d$ such that $\|\theta\| \leq 1$, and define $u^* : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ by

$$u^*(t, x) := \sqrt{\frac{-2\gamma}{T}} \left(\frac{x}{\|x\|} 1_{\{x \neq 0\}} + \theta 1_{\{x=0\}} \right) \quad (20)$$

The importance sampling change of measure $\{P_\varepsilon\}$ is again given by the Girsanov transform

$$\frac{dP_\varepsilon}{dP} := \exp \left\{ \frac{1}{\sqrt{\varepsilon}} \int_0^T u^*(t, W_\varepsilon(t)) dW(t) - \frac{1}{2\varepsilon} \int_0^T \|u^*(t, W_\varepsilon(t))\|^2 dt \right\}.$$

Theorem 3. $\{P_\varepsilon\}$ is asymptotically optimal for the estimation of both $\{P(Z \in A_\varepsilon)\}$ and $\{v_\varepsilon\}$.

Proof. We first show that $\{P_\varepsilon\}$ is asymptotically optimal for estimating $\{P(Z \in A_\varepsilon)\}$. This part is very similar to the proof of Theorem 1. Fix any $\delta > 0$, and define $V : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ by (see Dupuis and Wang 2007, section 10.7)

$$V(t, x) := -\sqrt{\frac{-2\gamma}{T}(\|x\|^2 + \delta)} - \gamma \left(1 + \frac{t}{T} \right).$$

It is not difficult to show that the gradient and Hessian of V are uniformly bounded by some constant, say C . Analogous to the proof of Theorem 1, we define the process $M := \{M_t : t \in [0, T]\}$ exactly as in (10). Because u^* is bounded, we easily have $E[M_t^2] \leq K$ for some constant K and all $t \in [0, T]$. Applying the Itô formula, we have

$$\begin{aligned} \frac{dM_t}{M_t} &= \frac{4\gamma}{\varepsilon T} \left(\sqrt{\frac{\|\sqrt{\varepsilon}W_\varepsilon(t)\|^2}{\|\sqrt{\varepsilon}W_\varepsilon(t)\|^2 + \delta}} - \frac{\|\sqrt{\varepsilon}W_\varepsilon(t)\|^2}{\|\sqrt{\varepsilon}W_\varepsilon(t)\|^2 + \delta} \right) 1_{\{W_\varepsilon(t) \neq 0\}} dt \\ &\quad + \frac{2\gamma}{\varepsilon T} (1 - \|\theta\|^2) 1_{\{W_\varepsilon(t)=0\}} dt - (Cd + \text{tr}\nabla^2 V) dt - \frac{1}{\sqrt{\varepsilon}} [2\nabla V + u^*] dW(t) \end{aligned}$$

with the understanding that all of the coefficients on the right-hand side are taking values at $(t, W_\varepsilon(t))$. Clearly, the coefficients of dt are negative (note that $\gamma \leq 0$). Define $F(x) := \log 1_D(x)$. We observe that for $x \in \bar{D}$, $\|x\|^2 \geq -2\gamma T$ by the definition of γ in (17). Thus, for all $x \in \bar{D}$,

$$V(T, x) = -\sqrt{\frac{-2\gamma}{T}(\|x\|^2 + \delta)} - 2\gamma \leq -\sqrt{\frac{-2\gamma}{T}}\|x\| - 2\gamma \leq 0.$$

In particular, we have $V(T, x) \leq -F(x)$ for all $x \in \mathbb{R}^d$. The rest of the proof is almost verbatim to that of Theorem 1, and we arrive at

$$\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E^\varepsilon [1_{A_\varepsilon}(Z)(dP/dP_\varepsilon)^2] \leq -2V(0, 0) = 2\gamma + 2\sqrt{\frac{-2\gamma\delta}{T}}.$$

Letting $\delta \rightarrow 0$, we conclude that $\{P_\varepsilon\}$ is asymptotically optimal for estimating $\{P(Z \in A_\varepsilon)\}$. To complete the proof, it is sufficient to show that for some $q > 1$,

$$\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[1_{A_\varepsilon}(Z)(dP/dP_\varepsilon)^q] \leq \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log E[(dP/dP_\varepsilon)^q] < \infty. \quad (21)$$

We then invoke Theorem 2. This is trivial because $\|u^*\|$ is uniformly bounded. Thus,

$$\begin{aligned} E[(dP/dP_\varepsilon)^q] &= E \left[\exp \left\{ \frac{-q}{\sqrt{\varepsilon}} \int_0^T u^*(t, W_\varepsilon(t)) dW(t) + \frac{q}{2\varepsilon} \int_0^T \|u^*(t, W_\varepsilon(t))\|^2 dt \right\} \right] \\ &\leq E \left[\exp \left\{ \frac{-q}{\sqrt{\varepsilon}} \int_0^T u^*(t, W_\varepsilon(t)) dW(t) - \frac{q^2}{2\varepsilon} \int_0^T \|u^*(t, W_\varepsilon(t))\|^2 dt \right\} \right] \\ &\quad \times \exp \left\{ \frac{(q^2 + q)T \|u^*\|_\infty^2}{2\varepsilon} \right\} = \exp \left\{ \frac{(q^2 + q)T \|u^*\|_\infty^2}{2\varepsilon} \right\}, \end{aligned}$$

and Inequality (21) follows readily. This completes the proof. \square

Remark 7. Both the dynamic importance sampling scheme outlined in Section 4 and the universal importance sampling schemes (20) are state dependent. However, the latter is much easier to construct because it only requires information on γ or equivalently, the distance between the origin and the region D where the payoff function is strictly positive. Moreover, it is more broadly applicable because it does not require the payoff function to take any specific form. However, the price that we pay for this simplicity and generality is the efficiency. Even though the universal schemes are asymptotically optimal under the alternative large deviation scaling, they will not be as efficient as those schemes in Section 4 when the latter is applicable. This is quite obvious as the universal schemes do not utilize the details of the payoff functions. It is also demonstrated by the empirical evidence in the subsequent numerical experiments.

6. Numerical Experiments

In this section, we will present numerical examples on the pricing of rainbow options. For some options, we will use the dynamic importance sampling scheme outlined in Section 4, and for all options, we will also perform the universal importance sampling outlined in Section 5. In Tables 1–7, which show the numerical results, “R.E.” denotes the relative error of the importance sampling estimate (standard error divided by the estimate), and “var ratio” denotes the ratio of the variance of the plain Monte Carlo estimate to that of the importance sampling estimate. In all of the simulations, the time interval $[0, T]$ is divided into $k = 50$ equal-length subintervals (see Remark 8). Each estimate is based on $n = 100,000$ samples.

All of the rainbow options that we consider satisfy the conditions in Section 2, including the linear growth Condition (3). Also, recall the definitions of vector A and square matrices C and Γ in (1) and thereafter. Note that $F(z) := \log G(z)$, where G is defined in (2). When we apply universal IS, the set of interest is $D := \{G > 0\}$.

Example 1. In this example, we consider a spread call option with $d = 2$ underlying assets and payoff

$$h(S) := [S_1(T) - S_2(T) - K]^+,$$

where $K > 0$ is a constant. In this case, for any $z \in \mathbb{R}^2$, the corresponding $F(z)$ and D are given by

$$\begin{aligned} F(z) &= \log(\exp\{x_1\} - \exp\{x_2\} - K)^+, \\ D &= \{z : \exp\{x_1\} - \exp\{x_2\} > K\}, \end{aligned}$$

where $x := [x_1, x_2]' = A + C\Gamma z$ for $z \in \mathbb{R}^2$ (Table 1).

1. *Dynamic IS.* One can show that F is a continuous concave function (we omit the technical details). Therefore, V in (11) is a classical subsolution with z^* as the unique maximizer:

$$z^* = \arg \max_{z \in \mathbb{R}^d} \left[F(z) - \frac{1}{2T} \|z\|^2 \right].$$

The corresponding change of measure is given by (8) and (9), where $u^* = -\nabla V(t, x) = z^*/T$. Note that z^* must satisfy $\nabla F(z^*) = z^*$.

2. *Universal IS.* The corresponding γ is given by Lemma 2, where the minimizing z must lie on $\partial D = \{z : \exp\{x_1\} - \exp\{x_2\} = K\}$.

We use the MATLAB function `fminunc` supplied with gradient (one can also use the iterative method in Glasserman et al. 1999) to solve for z^* . The equation of γ can be converted into an equation of a single variable and solved by the bisection method. We omit the technical details and simply present the simulation results. For future

Table 1. Spread Call Option with Two Underlying Assets

	$K = 20$		$K = 40$		$K = 60$	
	Dynamic	Universal	Dynamic	Universal	Dynamic	Universal
Estimate	1.2038	1.1828	0.0897	0.0903	0.0060	0.0060
R.E., %	0.24	0.70	0.33	1.25	0.39	1.83
Var ratio	18.4	2.4	138.1	9.6	1,699	76.1
z^*, \hat{z}	(1.4, -0.9)	(0.8, -0.6)	(2.4, -1.1)	(1.9, -1.0)	(3.2, -1.2)	(2.9, -1.1)

Notes. R.E. denotes the relative error of the importance sampling estimate (standard error divided by the estimate). Var ratio denotes the ratio of the variance of the plain Monte Carlo estimate to that of the importance sampling estimate.

reference, we also record z^* and the maximizing \hat{z} for Lemma 2. The parameters are set at $r = 0.05$, $T = 1$, and

$$S(0) = \begin{bmatrix} 35 \\ 30 \end{bmatrix}, \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}.$$

Example 2. In this example, we consider an outperformance digital call option with d underlying assets and payoff

$$h(S) = 1_{\{S_1(T) \vee \dots \vee S_d(T) \geq K\}},$$

where $K > 0$ is a constant (Table 2). In this case, for any $z \in \mathbb{R}^d$, the corresponding $F(z)$ and D are given by

$$F(z) = \log 1_{\{x_1 \vee \dots \vee x_d \geq \log K\}},$$

$$D = \{z : x_1 \vee \dots \vee x_d \geq \log K\},$$

where $x := [x_1, \dots, x_d]' = A + CFz$.

1. *Dynamic IS.* $F := F_1 \vee \dots \vee F_d$, where $F_i(z) := \log 1_{\{x_i \geq \log K\}}$ is a concave function for each i . Therefore, $V := V_1 \wedge \dots \wedge V_d$ in (11) is a (weak) subsolution, where for each i , V_i is given by (12) with z^i as the unique maximizer:

$$z^i = \arg \max_{z \in \mathbb{R}^d} \left[F_i(z) - \frac{1}{2T} \|z\|^2 \right].$$

Note that F_i takes values of either zero or $-\infty$. Thus, the maximizer z^i necessarily satisfies $F_i(z^i) = 0$. From this observation, it is very straightforward to show that

$$z^i = \frac{\log K - A_i}{\|\theta^i\|^2} \cdot \theta^i,$$

where θ^i denotes the i th column vector of the matrix $(CF)'$. A classical subsolution V^δ can be obtained by mollification (13) with derivatives (14) and (15). The corresponding change of measure is given by (8) and (9), where $u^* = -\nabla V^\delta(t, x)$.

2. *Universal IS.* In this case, the corresponding γ is given by Lemma 2, where obviously, the minimizing z must coincide with one of the z^i 's above, and it is easily identified as

$$\gamma = -\frac{1}{2T} \min_{i=1, \dots, d} \|z^i\|^2 = -\frac{1}{2T} \min_{i=1, \dots, d} \frac{(\log K - A_i)^2}{\|\theta^i\|^2}.$$

For simulation, we consider a digital call with $d = 3$ underlying assets. The mollification parameter δ is set to be 0.1 in dynamic IS, and the other parameters are $r = 0.05$, $T = 1$, and

$$S(0) = \begin{bmatrix} 40 \\ 35 \\ 40 \end{bmatrix}, \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.1 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 1 & -0.5 \\ 0.3 & -0.5 & 1 \end{bmatrix}.$$

Example 3. In this example, we consider a multistrike call option with d underlying assets and payoff

$$h(S) := \max\{S_1(T) - K_1, \dots, S_d(T) - K_d, 0\},$$

Table 2. Outperformance Digital Call Option with Three Underlying Assets

	K = 80		K = 100		K = 120	
	Dynamic	Universal	Dynamic	Universal	Dynamic	Universal
Estimate	3.41E-3	3.45E-3	2.41E-4	2.35E-4	2.05E-5	2.09E-5
R.E., %	0.64	2.97	0.65	3.40	0.68	5.44
Var ratio	71.2	3.2	1,012	38.3	9,256	138.9

Notes. R.E. denotes the relative error of the importance sampling estimate (standard error divided by the estimate). Var ratio denotes the ratio of the variance of the plain Monte Carlo estimate to that of the importance sampling estimate.

where $K > 0$ is a constant (Table 3). In this case, for any $z \in \mathbb{R}^d$, the corresponding $F(z)$ and D are given by

$$F(z) = \max\{\log[\exp\{x_1\} - K_1]^+, \dots, \log[\exp\{x_d\} - K_d]^+\},$$

$$D = \{z : \max\{x_1 - \log K_1, \dots, x_d - \log K_d\} > 0\},$$

where $x := [x_1, \dots, x_d]' = A + CTz$.

1. *Dynamic IS.* $F := F_1 \vee \dots \vee F_d$, where $F_i(z) := \log[\exp\{x_i\} - K_i]^+$ is a concave function for each i . Therefore, $V := V_1 \wedge \dots \wedge V_d$ of (11) is a (weak) subsolution, where for each i , V_i is given by (12) with z^i as the unique maximizer:

$$z^i = \arg \max_{z \in \mathbb{R}^d} \left[F_i(z) - \frac{1}{2T} \|z\|^2 \right].$$

It can be derived (we omit the technical details) that for each i , $z^i = x^i \theta^i$, where θ^i denotes the i th column vector of the matrix $(CT)'$ and x^i is the unique solution to the equation

$$1 + \frac{K_i}{e^{A_i + x^i \|\theta^i\|^2} - K_i} - \frac{x^i}{T} = 0, \quad x^i \geq \frac{\log K_i - A_i}{\|\theta^i\|^2}.$$

Each x^i can be obtained by the bisection method. A classical subsolution V^δ can be obtained by mollification (13) with derivatives (14) and (15). The corresponding change of measure is given by (8) and (9), where $u^* = -\nabla V^\delta(t, x)$.

2. *Universal IS.* The calculation of γ is almost verbatim to that of Example 2 with K replaced by K_i in each instance. Indeed,

$$\gamma = \min_{i=1, \dots, d} \frac{(\log K_i - A_i)^2}{\|\theta^i\|^2}.$$

The mollification parameter δ is set to be 0.1 in dynamic IS. We consider $d = 4$ underlying assets with parameters $r = 0.05$, $T = 1$, and

$$S(0) = \begin{bmatrix} 40 \\ 35 \\ 30 \\ 30 \end{bmatrix}, \quad \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.2 \\ 0.2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.2 & 0.3 & 0 \\ 0.2 & 1 & 0.4 & 0.2 \\ 0.3 & 0.4 & 1 & 0.3 \\ 0 & 0.2 & 0.3 & 1 \end{bmatrix},$$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ K_4 \end{bmatrix} = S(0) + \Delta \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

In order to investigate how the dimension of the problem affects the performance of these schemes, we also consider a group of $N = 12$ underlying assets and multistrike call options associated with the first d assets, with $d = 2, 4, 6, 8, 10, 12$. The initial stock price $S(0)$, the volatility vector $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_N]'$, and the covariance matrix Σ

Table 3. Multistrike Call Option with Four Underlying Assets

	$\Delta = 20$		$\Delta = 30$		$\Delta = 40$	
	Dynamic	Universal	Dynamic	Universal	Dynamic	Universal
Estimate	5.38E-2	5.45E-2	2.95E-3	3.01E-3	1.36E-4	1.38E-4
R.E., %	0.47	1.59	0.59	2.68	0.72	3.90
Var ratio	58.0	4.9	481.9	21.4	2,113	70.5

Notes. R.E. denotes the relative error of the importance sampling estimate (standard error divided by the estimate). Var ratio denotes the ratio of the variance of the plain Monte Carlo estimate to that of the importance sampling estimate.

are randomly generated. Below is a realization of these parameters:

$$S(0) = \begin{bmatrix} 18 \\ 12 \\ 22 \\ 24 \\ 18 \\ 18 \\ 12 \\ 21 \\ 30 \\ 25 \\ 15 \\ 21 \end{bmatrix}, \quad \sigma = \begin{bmatrix} 0.13 \\ 0.15 \\ 0.18 \\ 0.13 \\ 0.19 \\ 0.17 \\ 0.12 \\ 0.21 \\ 0.11 \\ 0.14 \\ 0.20 \\ 0.26 \end{bmatrix}, \quad \begin{matrix} K_1 \\ K_2 \\ K_3 \\ K_4 \\ K_5 \\ K_6 \\ K_7 \\ K_8 \\ K_9 \\ K_{10} \\ K_{11} \\ K_{12} \end{matrix} = S(0) + 30 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} 1.00 & 0.08 & 0.23 & -0.15 & -0.33 & 0.32 & 0.19 & -0.31 & 0.05 & 0.12 & 0.27 & -0.13 \\ 0.08 & 1.00 & 0.16 & 0.11 & 0.16 & 0.24 & 0.26 & 0.16 & 0.17 & -0.31 & 0.21 & -0.12 \\ 0.23 & 0.16 & 1.00 & -0.35 & -0.35 & -0.18 & 0.39 & 0.03 & -0.14 & -0.36 & -0.18 & 0.27 \\ -0.15 & 0.11 & -0.35 & 1.00 & 0.33 & -0.20 & -0.12 & -0.07 & 0.30 & 0.22 & -0.17 & -0.35 \\ -0.33 & 0.16 & -0.35 & 0.33 & 1.00 & -0.30 & -0.08 & 0.28 & 0.27 & -0.09 & 0.06 & -0.14 \\ 0.32 & 0.24 & -0.18 & -0.20 & -0.30 & 1.00 & -0.21 & -0.38 & 0.12 & -0.01 & 0.41 & -0.17 \\ 0.19 & 0.26 & 0.39 & -0.12 & -0.08 & -0.21 & 1.00 & 0.06 & -0.04 & 0.05 & -0.03 & -0.13 \\ -0.31 & 0.16 & 0.03 & -0.07 & 0.28 & -0.38 & 0.06 & 1.00 & 0.00 & -0.40 & -0.33 & -0.01 \\ 0.05 & 0.17 & -0.14 & 0.30 & 0.27 & 0.12 & -0.04 & 0.00 & 1.00 & 0.01 & -0.06 & 0.02 \\ 0.12 & -0.31 & -0.36 & 0.22 & -0.09 & -0.01 & 0.05 & -0.40 & 0.01 & 1.00 & -0.14 & -0.37 \\ 0.27 & 0.21 & -0.18 & -0.17 & 0.06 & 0.41 & -0.03 & -0.33 & -0.06 & -0.14 & 1.00 & -0.01 \\ -0.13 & -0.12 & 0.27 & -0.35 & -0.14 & -0.17 & -0.13 & -0.01 & 0.02 & -0.37 & -0.01 & 1.000 \end{bmatrix}.$$

The other parameters in the simulation remain the same except that the sample size is increased to $n = 400,000$ (Table 4).

The variance ratios to the plain Monte Carlo estimates are not presented because it is often the case that the plain Monte Carlo cannot produce meaningful estimates/variances for $d = 2, 4, 6, 8, 10$. For example, a typical plain Monte Carlo estimate for $d = 2$ is zero, and a typical estimate for $d = 8$ is $1.16E-5$ with a 40% relative error (in these cases, the true values of the expectation and variance are both vastly underestimated, a common occurrence in rare event simulation) (Glasserman and Wang 1997, Dupuis and Wang 2004). It is only for $d = 12$ that the plain Monte

Table 4. Multistrike Call Option with d Underlying Assets

	$d = 2$		$d = 4$		$d = 6$		$d = 8$		$d = 10$		$d = 12$	
	Dynamic	Universal	Dynamic	Universal	Dynamic	Universal	Dynamic	Universal	Dynamic	Universal	Dynamic	Universal
Estimate	2.01E-13	1.96E-13	4.10E-6	3.73E-6	4.24E-6	4.60E-6	5.52E-5	5.21E-5	5.51E-5	5.42E-5	1.5E-3	1.6E-3
R.E., %	0.31	2.21	0.24	3.44	0.35	6.38	0.29	7.70	0.28	9.95	0.28	8.07

Note. R.E. denotes the relative error of the importance sampling estimate (standard error divided by the estimate).

Carlo produces meaning results, in which case the variance ratios are 2,300 and 4, respectively. In all of these simulations, one can see that both the dynamic IS and the universal IS outperform the plain Monte Carlo simulation. The dynamic IS, when feasible, is more efficient than the universal IS.

Example 4. In this example, we show the advantage of universal IS by considering three different types of rainbow options. Let K , $\{c_i : i = 1, \dots, d\}$, and $\{K_i : i = 1, \dots, d\}$ be strictly positive constants:

1. basket option: $h(S) = [c_1 S_1(T) + \dots + c_d S_d(T) - K]^+$;
2. pyramid option: $h(S) = [|S_1(T) - K_1| + \dots + |S_d(T) - K_d| - K]^+$; and
3. Madonna option: $h(S) = [\sqrt{|S_1(T) - K_1|^2 + \dots + |S_d(T) - K_d|^2} - K]^+$.

For all of these options, the corresponding F cannot be expressed as the maximum of concave functions. Therefore, it is not clear how one can construct appropriate subsolutions and consequently, dynamic IS algorithms.

However, the universal IS algorithm is still easily applicable. For all simulations, we consider $d = 3$ underlying assets with parameters $r = 0.05$, $T = 1$, and

$$S(0) = \begin{bmatrix} 40 \\ 35 \\ 30 \end{bmatrix}, \quad \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 1 & 0.4 \\ 0.3 & 0.4 & 1 \end{bmatrix},$$

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.3 \\ 0.4 \end{bmatrix}, \quad \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 35 \\ 35 \\ 35 \end{bmatrix}.$$

We use the MATLAB function `fmincon` to solve the minimization problem in Lemma 2. There is no guarantee that the local minimum found by the function is the global minimum because of the lack of concavity (even though we have used random initial search conditions for `fmincon` to alleviate the problem to some degree). However, the simulation shows great improvement in performance, which is our ultimate goal. In Tables 5–7, “plain MC” stands for the plain Monte Carlo scheme.

6.1. Summary

The numerical results have shown that both dynamic IS and universal IS greatly improve the efficiency of the Monte Carlo simulation, especially when the options are deep out of the money. When applicable, the dynamic IS outperforms the universal IS, which is not surprising considering that the universal scheme only uses very limited knowledge of the model. On the other hand, universal IS (still provably asymptotically optimal in the alternative scaling) leads to great variance reduction compared with the plain Monte Carlo algorithm, and it is much more broadly applicable. Further, it also is much easier to construct for various rainbow options with complicated payoffs.

Remark 8. The implementation of the dynamic importance sampling or the universal importance sampling requires the partition of the time interval. Even though we have picked a fairly large partition with $k = 50$ here, the results remain nearly identical even if k is much smaller: for example, $k = 10$. Note that in this paper, the plain Monte Carlo simulation does not require the partition of the time interval, whereas the state-dependent importance sampling does. Such computational overhead is insignificant compared with the reduction in variance, especially for the simulation of deep out-of-the-money options.

Table 5. Basket Call Option with Three Underlying Assets

	$K = 50$		$K = 55$		$K = 60$	
	Universal	Plain MC	Universal	Plain MC	Universal	Plain MC
Estimate	7.47E-3	7.09E-3	5.22E-4	4.73E-4	3.30E-5	1.20E-4
R.E., %	2.51	6.96	2.54	24.7	4.16	61.6
Var ratio	7.0	—	77.5	—	2,891	—

Notes. Plain MC stands for the plain Monte Carlo scheme. R.E. denotes the relative error of the importance sampling estimate (standard error divided by the estimate). Var ratio denotes the ratio of the variance of the plain Monte Carlo estimate to that of the importance sampling estimate.

Table 6. Pyramid Call Option with Three Underlying Assets

	K = 50		K = 60		K = 70	
	Universal	Plain MC	Universal	Plain MC	Universal	Plain MC
Estimate	2.66E-2	2.74E-2	4.67E-3	5.03E-3	8.22E-4	7.75E-4
R.E., %	1.61	6.72	2.00	14.5	2.38	30.1
Var ratio	18.3	—	60.6	—	142.1	—

Notes. Plain MC stands for the plain Monte Carlo scheme. R.E. denotes the relative error of the importance sampling estimate (standard error divided by the estimate). Var ratio denotes the ratio of the variance of the plain Monte Carlo estimate to that of the importance sampling estimate.

7. Further Remarks

This paper discusses two state-dependent importance sampling schemes: the “dynamic IS” based on the classical large deviations embedding and the “universal IS” based on an alternative large deviations embedding. Both of them greatly improve the efficiency of Monte Carlo simulation of rainbow options. Even though the theory is based on market models with lognormal stock prices and path-independent options, in principle one may be able to extend the algorithms to more general market models and path-dependent options. For the dynamic IS, the challenge lies in the overly complicated structure of the HJB equations, which is inevitable when one deviates from the simple lognormal stock price model and/or introduces path dependency, as well as the construction of appropriate subsolutions. On the other hand, the universal IS will reduce the complexity of the problem considerably by only dealing with probabilities. Even so, it still leads to a number of challenging yet interesting future research directions.

For example, some commonly used stock price models are jump diffusion models. The building blocks in those models, for instance, can take the form

$$W(t) + \sum_{i=1}^{N(t)} Y_i,$$

where W is a Brownian motion, N is a Poisson process, and $\{Y_1, Y_2, \dots\}$ determine jump sizes. It would be interesting to construct an explicit and asymptotically efficient algorithm similar to (20) by analyzing the large deviation rate functions associated with those building blocks. Another example is path-dependent options, such as barrier options and Asian options. Regardless of whether such options are discretely or continuously monitored, one has to deal with a very high-dimensional or infinite-dimensional system. The construction and analysis of universal schemes within this context are not immediately clear. The last example that we would like to mention is with respect to American options. To the best of our knowledge, there has been no work on importance sampling for deep-out-of-the-money American options with provable asymptotic optimality. It would be interesting to see if any rigorous analysis can be done under the alternative large deviations embedding.

8. Heuristic Connection Between Two Embeddings

Finally, we would like to discuss the connection between the two large deviations embedding heuristically. It is difficult to give any rigorous justification, and hence, we should only proceed heuristically. The first large deviation scaling (4) leads to the rate function given by (5), whereas the second scaling (16) leads to (17). Intuitively speaking, the most important point for importance sampling under each large deviations scaling is the optimizing point in (5) and (17). Our goal is to heuristically justify that these two optimizing points are indeed close to each other in practice. In the special case where G is an indicator function (i.e., binary options), these two points

Table 7. Madonna Call Option with Three Underlying Assets

	K = 40		K = 50		K = 60	
	Universal	Plain MC	Universal	Plain MC	Universal	Plain MC
Estimate	1.07E-2	0.93E-2	1.06E-3	1.52E-3	1.09E-4	1.86E-4
R.E., %	1.67	9.53	2.36	27.2	4.28	59.4
Var ratio	24.7	—	282.8	—	554.4	—

Notes. Plain MC stands for the plain Monte Carlo scheme. R.E. denotes the relative error of the importance sampling estimate (standard error divided by the estimate). Var ratio denotes the ratio of the variance of the plain Monte Carlo estimate to that of the importance sampling estimate.

are identical. More generally, let us assume that $G(z) = [f(z) - K]^+$ for some continuous function f (i.e., various call options). We also let $T = 1$ for notational convenience. In this case, (5) is just

$$\sup_{z \in \mathbb{R}^d} \left(\log[f(z) - K]^+ - \frac{1}{2} \|z\|^2 \right) = \sup_{z \in D} \left(\log[f(z) - K] - \frac{1}{2} \|z\|^2 \right), \quad (22)$$

where $D := \{f(z) > K\} = \{G(z) > 0\}$. Assume that f is continuously differentiable and $c_1 f(z) \leq \|\nabla f(z)\| \leq c_2 f(z)$ for some constants $c_1, c_2 > 0$ on the set \bar{D} . Given the form of the stock prices, this assumption is very natural. The maximizing $z^* \in \bar{D}$ should satisfy the equation

$$\frac{\nabla f(z^*)}{f(z^*) - K} = z^*,$$

and thus,

$$\frac{c_1 f(z^*)}{f(z^*) - K} \leq \|z^*\| \leq \frac{c_2 f(z^*)}{f(z^*) - K}.$$

Because we are interested in the case where K is large and $\|z\| \rightarrow \infty$ as $K \rightarrow \infty$ if $z \in \bar{D}$ (because of the continuity of f), we can assume that $\|z^*\| > c_2$, and thus, the preceding inequality amounts to

$$K \frac{\|z^*\|}{\|z^*\| - c_1} \leq f(z^*) \leq K \frac{\|z^*\|}{\|z^*\| - c_2}.$$

Thus, the maximization problem for the first large deviations scaling or Equation (22) can be taken over all z such that

$$z \in D' := D \cap \left\{ z : K \frac{\|z\|}{\|z\| - c_1} \leq f(z) \leq K \frac{\|z\|}{\|z\| - c_2} \right\}. \quad (23)$$

For all such z (note that $\|z\|$ is large when K is large), we have

$$\log[f(z) - K] - \frac{1}{2} \|z\|^2 = \log K + \left(-\frac{1}{2} \|z\|^2 + \phi(z) \right),$$

where $\phi(z) = o(\|z\|^2)$ because

$$\log \frac{c_1}{\|z\| - c_1} \leq \phi(z) \leq \log \frac{c_2}{\|z\| - c_2}.$$

Therefore, intuitively, the optimization problem is more or less equivalent to the following:

$$\sup_{z \in D'} -\frac{1}{2} \|z\|^2.$$

However, recalling the definition of D' in (23) and that for each $i = 1, 2$,

$$\frac{\|z\|}{\|z\| - c_i} \approx 1,$$

we conclude that D' is very close to ∂D . This justifies that the optimizer for (5) is very close to that of (17).

Acknowledgments

The authors express their appreciation to the anonymous referee for careful suggestions.

References

- Blanchet JH, Glynn P, Liu JC (2007) Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. *Queueing Systems* 57(2):99–113.
- Bucklew JA (2010) *Introduction to Rare Event Simulation* (Springer-Verlag, New York).
- Dembo A, Zeitouni O (1998) *Large Deviations Techniques and Applications* (Springer-Verlag, New York).
- Dupuis P, Wang H (2004) Importance sampling, large deviations, and differential games. *Stochastics Stochastic Rep.* 76(6):481–508.
- Dupuis P, Wang H (2007) Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math. Oper. Res.* 32(3):723–757.
- Dupuis P, Wang H (2009) Importance sampling for Jackson networks. *Queueing Systems* 62(1):113–157.

- Glasserman P (2004) *Monte Carlo Methods in Financial Engineering* (Springer-Verlag, New York).
- Glasserman P, Wang Y (1997) Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.* 7(3):731–746.
- Glasserman P, Heidelberger P, Shahabuddin P (1999) Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance* 9(2):117–152.
- Glasserman P, Kang W, Shahabuddin P (2008) Fast simulation of multifactor portfolio credit risk. *Oper. Res.* 56(5):1200–1217.
- Guasoni P, Robertson S (2008) Optimal importance sampling with explicit formulas in continuous time. *Finance Stochastics* 12(1):1–19.
- Heidelberger P (1995) Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Simulation* 5(1):43–85.
- Karatzas I, Shreve SE (1991) *Brownian Motion and Stochastic Calculus* (Springer-Verlag, New York).
- Rubinstein RY, Kroese DP (2007) *Simulation and the Monte Carlo Method* (John Wiley & Sons, Hoboken, NJ).
- Sadowsky JS (1996) On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Probab.* 6(2):399–422.
- Sigmund JS (1976) Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* 4(4):673–684.
- Wang H (2012) *Monte Carlo Simulation with Applications to Finance* (CRC Press, Boca Raton, FL).
- Wang H, Zhou X (2015) A cross-entropy scheme for mixtures. *ACM Trans. Model. Comput. Simulation* 25(1):6.