Optimal importance sampling parameter search for Lévy processes via stochastic approximation

REIICHIRO KAWAI

Abstract

The author proposes stochastic approximation methods of finding the optimal measure change by the exponential tilting for Lévy processes in Monte Carlo importance sampling variance reduction. In accordance with the structure of the underlying Lévy measure, either a constrained or unconstrained algorithm of the stochastic approximation is chosen. For both cases, the almost sure convergence to a unique stationary point is proved. Numerical examples are presented to illustrate the effectiveness of our method.

Keywords: Esscher transform, Girsanov theorem, Monte Carlo simulation, infinitely divisible distribution, stochastic approximation algorithm, variance reduction.
2000 Mathematics Subject Classification: 65C05, 62L20, 60E07, 60G51.

1 Introduction

The importance sampling method is aimed at reducing the variance of iid Monte Carlo summands by appropriately transforming the underlying probability measure, from which interested random variables or stochastic processes are generated, so as to put more weight on important events and less on undesirable ones. Due to its practical effectiveness, it has long been thought of as one of the most important variance reduction methods in the Monte Carlo simulation and has been intensively studied with a view towards a wide range of applications, such as mathematical finance, queueing theory, sequential analysis, to mention just a few. For its principle with some numerical examples, see, for instance, Section 4.6 of Glasserman [7].

In the importance sampling "variance" reduction, the optimal measure change means nothing but the one attaining the minimal variance of iid Monte Carlo summands. In the Gaussian framework, the Girsanov measure change is often indexed by a single parameter, that is, the drift

Published in *SIAM Journal on Numerical Analysis* (2008) **47**(1) 293-307.
Running head: Optimal importance sampling for Lévy processes
Email address: reiichiro_kawai@ybb.ne.jp. Postal address: Center for the Study of Finance and Insurance, Osaka University, Toyonaka, 560-8531, Japan. TEL:+81-6-6850-6279. FAX:+81-6-6850-6092.

parameter, and several attempts have been made to find its optimum. In financial applications, for example, Glasserman, Heidelberger, and Shahabuddin [6] proposes an optimization procedure to find a nearly optimal measure change in pricing financial derivatives, while Su and Fu [14] and Arouna [1] apply the stochastic approximation so as to search for the root of the gradient of the Monte Carlo variance with respect to the measure change parameter.

The aim of the present work is to apply the idea of [1, 14] to Lévy processes without the Brownian motion, or equivalently after discretization, infinitely divisible laws without Gaussian component. In general, the measure change for Lévy processes involves every single jump, which forms the sample paths. (See Section 33 of Sato [12] for details. For an importance sampling method with such intricate measure changes, see Kawai [8].) In this paper, we however restrict our attention to the simplest measure change, often called the Esscher transform, which has only to look at the terminal marginals. The Esscher transform is nothing but the well known exponential tilting of laws and is thus indexed by a single (multidimensional) parameter. As we will investigate later, a crucial difficulty in the case of Lévy processes without Gaussian component is that depending on the structure of the underlying Lévy measure, the exponential tilting parameter might have to stay in a suitable compact set, while the drift parameter of the Gaussian distribution may be arbitrarily taken.

The rest of the paper is organized as follows. Section 2 recalls the Esscher transform and the principle of the importance sampling variance reduction, and constructs the basis of our analysis. In Section 3, the almost sure convergence of the stochastic approximation is proved separately for the constrained and unconstrained algorithms, depending on the structure of the underlying Lévy measure. Section 4 illustrates the effectiveness of our method via numerical examples for both constrained and unconstrained stochastic approximation algorithms. Finally, Section 5 concludes.

2 Preliminaries

Let us begin with some notations which will be used throughout the text. $\mathbb{N}$ is the collection of all positive numbers, with $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. $\mathbb{R}^d$ is the d-dimensional Euclidean space with the norm $\|\cdot\|$ and the inner product $\langle \cdot, \cdot \rangle$, $\mathbb{R}_0^d := \mathbb{R}^d \setminus \{0\}$ and $\mathscr{B}(\mathbb{R}_0^d)$ is the Borel σ-field of $\mathbb{R}_0^d$. $(\Omega, \mathscr{F}, \mathbb{P})$ is our underlying probability space. $\mathrm{Leb}(\cdot)$ denotes the Lebesgue measure, while $\mathbb{P}|_{\mathscr{F}_t}$ is the restriction of a probability measure $\mathbb{P}$ to the σ-field $\mathscr{F}_t$. Denote by ∇ the gradient, and by $\mathrm{Hess}[\cdot]$ the Hessian matrix. The interval $(0, -1]$ is understood to be $[-1, 0)$. The expression $f(x) \sim g(x)$ means $f(x)/g(x)$ tends to 1. The identity in law is denoted by $\stackrel{\mathscr{L}}{=}$. We say that a stochastic process $\{X_t : t \geq 0\}$ in $\mathbb{R}^d$ is a Lévy process if it has independent and stationary increments, if it is continuous in probability, and if $X_0 = 0$, *a.s.* By the Lévy-Khintchine representation theorem, the characteristic function of its marginal law is uniquely given by

$$\mathbb{E}\left[e^{i\langle y,X_t\rangle}\right]=\exp\left[t\left(i\langle y,\gamma\rangle-\frac{1}{2}\langle y,Ay\rangle+\int_{\mathbb{R}_0^d}\left(e^{i\langle y,z\rangle}-1-i\langle y,z\rangle\mathbb{1}_{(0,1]}(\|z\|)\right)\nu(dz)\right)\right],$$

where $\gamma\in\mathbb{R}^d$, A is a symmetric nonnegative-definite $d\times d$ matrix, and ν is a Lévy measure on $\mathbb{R}_0^d$, that is, $\int_{\mathbb{R}_0^d}(\|z\|^2\wedge 1)\nu(dz)<+\infty$. If the above holds, then we say that the Lévy process $\{X_t:t\geq 0\}$ is generated by the triplet (γ,A,ν). *In this paper, we restrict our attention to pure-jump Lévy processes, that is, we set $A\equiv 0$ throughout. Moreover, we also assume that all components are non-degenerate.* A function $f:\mathbb{R}^d\mapsto[0,\infty)$ is said to be submultiplicative if there exists a positive constant a such that $f(x+y)\leq af(x)f(y)$ for $x,y\in\mathbb{R}^d$. Letting $c\in\mathbb{R}$, $\gamma\in\mathbb{R}^d$, and $b>0$, if $f(x)$ is submultiplicative on $\mathbb{R}^d$, then $f(cx+\gamma)^b$ is submultiplicative, and the functions $\|x\|\vee 1$, $e^{\langle c,x\rangle}$ are submultiplicative, and a product of two submultiplicative functions is submultiplicative. We recall an important moment property of Lévy processes, which will be used often in what follows.

Theorem 2.1. (Sato [12], Theorem 25.3) *Let f be a submultiplicative, locally bounded, measurable function on $\mathbb{R}^d$, and let $\{X_t:t\geq 0\}$ be a Lévy process in $\mathbb{R}^d$ with Lévy measure ν. Then, $\mathbb{E}[f(X_t)]$ is finite for every $t>0$ if and only if $\int_{\|z\|>1}f(z)\nu(dz)<+\infty$.*

2.1 Esscher transform

Among the density transformations of Lévy processes, there is a simple class ending up with looking only at the marginals, which is built via the exponential tilting. The class is often called the Esscher transform in mathematical finance and actuarial science. Let $\{X_t:t\geq 0\}$ be a Lévy process in $\mathbb{R}^d$ generated by $(\gamma,0,\nu)$, and let $(\mathscr{F}_t)_{t\geq 0}$ be the natural filtration of $\{X_t:t\geq 0\}$. Define

$$\Lambda_1:=\left\{\lambda\in\mathbb{R}^d:\mathbb{E}_{\mathbb{P}}\left[e^{\langle\lambda,X_1\rangle}\right]<+\infty\right\}=\left\{\lambda\in\mathbb{R}^d:\int_{\|z\|>1}e^{\langle\lambda,z\rangle}\nu(dz)<+\infty\right\},$$

where the second equality holds by Theorem 2.1. We impose the condition $\mathrm{Leb}(\Lambda_1)>0$ throughout. Clearly, the set Λ_1 contains the origin and is convex. For $\lambda\in\Lambda_1$, we denote by φ the cumulant generating function of the marginal law at unit time of $\{X_t:t\geq 0\}$ under the probability measure $\mathbb{P}$, that is, $\varphi(\lambda):=\ln\mathbb{E}_{\mathbb{P}}[e^{\langle\lambda,X_1\rangle}]$. For ease in notation, we also write $\varphi_t(\lambda):=\ln\mathbb{E}_{\mathbb{P}}[e^{\langle\lambda,X_t\rangle}]$, $t>0$, in view of

$$\varphi_t(\lambda)=\ln\mathbb{E}_{\mathbb{P}}\left[e^{\langle\lambda,X_t\rangle}\right]=t\ln\mathbb{E}_{\mathbb{P}}\left[e^{\langle\lambda,X_1\rangle}\right]=t\varphi(\lambda),$$

where the second equality holds by the infinite divisibility of the marginal laws of Lévy processes. Note that $\varphi(\lambda)$ is continuous and $\nabla\varphi(\lambda)$ is well defined in $\lambda\in\Lambda_1$. Under the probability measure $\mathbb{Q}_\lambda$, where $\lambda\in\Lambda_1$ and which is defined via the Radon-Nikodym derivative, for every $t\in(0,+\infty)$,

$$\left.\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}}\right|_{\mathscr{F}_t}=\frac{e^{\langle\lambda,X_t\rangle}}{\mathbb{E}_{\mathbb{P}}\left[e^{\langle\lambda,X_t\rangle}\right]}=e^{\langle\lambda,X_t\rangle-\varphi_t(\lambda)},\quad\mathbb{P}\text{-}a.s.,$$

the stochastic process $\{X_t : t \geq 0\}$ is again a Lévy process generated by $(\gamma_\lambda, 0, \nu_\lambda)$, where $\gamma_\lambda = \gamma + \int_{\|z\|\leq 1} z(\nu_\lambda - \nu)(dz)$, and

$$\nu_\lambda(dz) = e^{\langle \lambda, z\rangle} \nu(dz). \tag{2.1}$$

Then, the probability measures $\mathbb{P}|_{\mathscr{F}_t}$ and $\mathbb{Q}_\lambda|_{\mathscr{F}_t}$ are mutually absolutely continuous for every $t \in (0,+\infty)$. We also have $\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\langle \lambda, X_1\rangle}] < +\infty$, and

$$\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda}\Big|_{\mathscr{F}_t} = \left(\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}}\Big|_{\mathscr{F}_t}\right)^{-1} = e^{-\langle \lambda, X_t\rangle + \varphi_t(\lambda)}, \quad \mathbb{Q}_\lambda\text{-}a.s.$$

For $t > 0$, let p be a probability density function on $\mathbb{R}^d$ of the random vector X_t under $\mathbb{P}$, provided that it is well defined. Then, a density function p_λ of X_t under $\mathbb{Q}_\lambda$ is given by

$$p_\lambda(x) = e^{\langle \lambda, x\rangle - \varphi_t(\lambda)} p(x), \quad x \in \mathbb{R}^d. \tag{2.2}$$

2.2 Importance sampling variance reduction

Suppose we are interested in evaluating

$$C := \mathbb{E}_{\mathbb{P}}[F(X)]$$

by Monte Carlo simulation, where $F(X) := F(\{X_t : t \in [0,T]\}) \in L^2(\Omega, \mathscr{F}_T, \mathbb{P})$, and assume $\mathbb{P}(F(X) \neq 0) > 0$. In view of the equality

$$\mathbb{E}_{\mathbb{P}}[F(X)] = \mathbb{E}_{\mathbb{Q}_\lambda}\left[\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda}\Big|_{\mathscr{F}_T} F(X)\right] = \mathbb{E}_{\mathbb{Q}_\lambda}\left[\left(\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}}\Big|_{\mathscr{F}_T}\right)^{-1} F(X)\right] = \mathbb{E}_{\mathbb{Q}_\lambda}\left[e^{-\langle \lambda, X_T\rangle + \varphi_T(\lambda)} F(X)\right],$$

define a set

$$\Lambda_2 := \Lambda_1 \cap \left\{\lambda \in \mathbb{R}^d : \mathbb{E}_{\mathbb{P}}\left[e^{-\langle \lambda, X_T\rangle} F(X)^2\right] < +\infty\right\},$$

and suppose that $\mathrm{Leb}(\Lambda_2) > 0$. Let us now give a lemma, whose proof will be often adapted in what follows.

Lemma 2.2. *The set Λ_2 is convex.*

Proof. For any $\lambda_1, \lambda_2 \in \Lambda_2$, and for any $m \in (0,1)$ and $n = 1-m$, the Hölder inequality gives

$$\mathbb{E}_{\mathbb{P}}\left[e^{-\langle m\lambda_1 + n\lambda_2, X_T\rangle} F(X)^2\right] \leq \mathbb{E}_{\mathbb{P}}\left[e^{-\langle \lambda_1, X_T\rangle} F(X)^2\right]^m \mathbb{E}_{\mathbb{P}}\left[e^{-\langle \lambda_1, X_T\rangle} F(X)^2\right]^n < +\infty.$$

The claim then follows from the convexity of Λ_1. $\square$

For $\lambda \in \Lambda_2$, the variance under the probability measure $\mathbb{Q}_\lambda$ is given by

$$\begin{aligned}
V(\lambda) &:= \mathbb{E}_{\mathbb{Q}_\lambda}\left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda}\Big|_{\mathscr{F}_T}\right)^2 F(X)^2\right] - C^2 \\
&= \mathbb{E}_{\mathbb{P}}\left[\left(\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}}\Big|_{\mathscr{F}_T}\right)^{-1} F(X)^2\right] - C^2 \\
&= \mathbb{E}_{\mathbb{P}}\left[e^{-\langle\lambda, X_T\rangle + \varphi_T(\lambda)} F(X)^2\right] - C^2.
\end{aligned}$$

Define also a set

$$\Lambda_3 := \Lambda_2 \cap \left\{\lambda \in \mathbb{R}^d : \mathbb{E}_{\mathbb{P}}\left[\|X_T\|^2 e^{-\langle\lambda, X_T\rangle} F(X)^2\right] < +\infty\right\},$$

and assume that $\mathrm{Leb}(\Lambda_3) > 0$.

Proposition 2.3. *The set Λ_3 is convex and $V(\lambda)$ is strictly convex in $\lambda \in \Lambda_3$.*

Proof. The convexity of Λ_3 can be proved in a similar manner to the proof of Lemma 2.2.

Since $\lambda \in \Lambda_3$, by the Hölder inequality, we have

$$\mathbb{E}_{\mathbb{P}}\left[\|X_T\| e^{-\langle\lambda, X_T\rangle} F(X)^2\right]^2 \leq \mathbb{E}_{\mathbb{P}}\left[e^{-\langle\lambda, X_T\rangle} F(X)^2\right] \mathbb{E}_{\mathbb{P}}\left[\|X_T\|^2 e^{-\langle\lambda, X_T\rangle} F(X)^2\right] < +\infty,$$

and thus with the help of the dominated convergence theorem, we obtain the gradient

$$\nabla V(\lambda) = \mathbb{E}_{\mathbb{P}}\left[(\nabla\varphi_T(\lambda) - X_T) e^{-\langle\lambda, X_T\rangle + \varphi_T(\lambda)} F(X)^2\right].$$

and also the Hessian

$$\mathrm{Hess}[V(\lambda)] = \mathbb{E}_{\mathbb{P}}\left[\left(\mathrm{Hess}[\varphi_T(\lambda)] + (\nabla\varphi_T(\lambda) - X_T)(\nabla\varphi_T(\lambda) - X_T)'\right) e^{-\langle\lambda, X_T\rangle + \varphi_T(\lambda)} F(X)^2\right].$$

Then, we have for $y \in \mathbb{R}^d_0$,

$$y'\mathrm{Hess}[V(\lambda)]y = \mathbb{E}_{\mathbb{P}}\left[\left(y'\mathrm{Hess}[\varphi_T(\lambda)]y + \langle y, \nabla\varphi_T(\lambda) - X_T\rangle^2\right) e^{-\langle\lambda, X_T\rangle + \varphi_T(\lambda)} F(X)^2\right] > 0,$$

since $\mathrm{Hess}[\varphi_T(\lambda)]$ reduces to the variance-covariance matrix of the random vector X_T under the probability measure $\mathbb{Q}_\lambda$, which is clearly positive definite. $\square$

Remark 2.4. The definition of the sets Λ_2 and Λ_3 is less intuitive and is of less practical use. We may instead give more intuitive definition in connection with the Lévy measure by giving up some part of its domain as

$$\Lambda_2' = \left\{\lambda \in \mathbb{R}^d : \int_{\|z\|>1} e^{-q\langle\lambda, z\rangle} \nu(dz) < +\infty, \mathbb{E}_{\mathbb{P}}\left[|F(X)|^{2p}\right] < +\infty, \ \frac{1}{p} + \frac{1}{q} = 1 \text{ for some } p > 1\right\},$$

and

$$\Lambda_3' = \left\{ \lambda \in \mathbb{R}^d : \int_{\|z\|>1} \|z\|^{2q} e^{-q\langle \lambda, z\rangle} \nu(dz) < +\infty, \mathbb{E}_{\mathbb{P}}\left[|F(X)|^{2p}\right] < +\infty,\ \frac{1}{p}+\frac{1}{q} = 1 \text{ for some } p > 1 \right\}.$$

It is easy to check that both Λ_2' and Λ_3' are convex, and that $\Lambda_2' \subseteq \Lambda_2$, $\Lambda_3' \subseteq \Lambda_3$, and $\Lambda_3' \subseteq \Lambda_2'$. They are derived as follows. By the Hölder inequality, with $1/p + 1/q = 1$ for some $p > 1$ and for $k = 0, 2$,

$$\mathbb{E}_{\mathbb{P}}\left[\|X_T\|^k e^{-\langle \lambda, X_T\rangle} F(X)^2\right] \leq \mathbb{E}_{\mathbb{P}}\left[|F(X)|^{2p}\right]^{1/p} \mathbb{E}_{\mathbb{P}}\left[\|X_T\|^{kq} e^{-q\langle \lambda, X_T\rangle}\right]^{1/q}.$$

By Theorem 2.1, the finiteness of the second expectation of the above right hand side for each $k = 0, 2$ is equivalent to $\int_{\|z\|>1} \|z\|^{kq} e^{-q\langle \lambda, z\rangle} \nu(dz) < +\infty$ for corresponding k. This, with $k = 0$, asserts the definition of Λ_2, while the definition of Λ_3 is verified with $k = 2$.

Meanwhile, as soon as $F(X)$ reduces to $f(X_T)$ with f being submultiplicative, the set Λ_3 is identical to

$$\left\{ \lambda \in \mathbb{R}^d : \int_{\|z\|>1} \left[e^{\langle \lambda, z\rangle} \vee \left(\|z\|^2 e^{-\langle \lambda, z\rangle} f(z)^2 \right) \right] \nu(dz) < +\infty \right\},$$

by Theorem 2.1. $\square$

3 Convergence of stochastic approximation algorithms

We begin with recalling the stochastic approximation algorithms. Let $\{X_{n,t} : t \in [0,T]\}_{n\in\mathbb{N}}$ be iid copies of the stochastic process $\{X_t : t \in [0,T]\}$. For ease in notation, we will write $X_n := X_{n,T}$ for $n \in \mathbb{N}$, and $F(X)_n := F(\{X_{n,t} : t \in [0,T]\})$. Let H be a connected set in $\mathbb{R}^d$ with $\{0\} \in H$, and define a sequence $\{Y_n\}_{n\in\mathbb{N}}$ of random vectors in $\mathbb{R}^d$ by

$$Y_{n+1} = (\nabla\varphi(\lambda_n) - X_{n+1})\, e^{-\langle \lambda_n, X_{n+1}\rangle + \varphi(\lambda_n)} F(X)_{n+1}^2,$$

where $\lambda_0 \in H$, $\{\lambda_n\}_{n\in\mathbb{N}}$ is a sequence of random vectors in $\mathbb{R}^d$ iteratively generated by

$$\lambda_{n+1} = \Pi_H\left[\lambda_n - \varepsilon_n Y_{n+1}\right], \tag{3.1}$$

where Π_H is the projection onto the constraint set H and where $\{\varepsilon_n\}_{n\in\mathbb{N}_0}$ is a sequence of positive constants satisfying

$$\textstyle\sum_{n\in\mathbb{N}_0} \varepsilon_n = +\infty, \quad \sum_{n\in\mathbb{N}_0} \varepsilon_n^2 < +\infty. \tag{3.2}$$

Moreover, define the filtration $(\mathscr{G}_n)_{n\in\mathbb{N}_0}$ by $\mathscr{G}_n := \sigma(\{\lambda_k\}_{k\leq n}, \{X_k\}_{k\leq n})$.

In what follows, the term "the constrained algorithms" means the algorithms where the constraint set H in (3.1) is not $\mathbb{R}^d$ and the sequence $\{\lambda_n\}_{n\in\mathbb{N}_0}$ is required to stay in the set, while by "the unconstrained algorithms," we mean the ones whose constraint set H is extended to $\mathbb{R}^d$, that is, the elements of $\{\lambda_n\}_{n\in\mathbb{N}_0}$ may be arbitrarily taken in $\mathbb{R}^d$.

Define a set

$$\Lambda_4 := \Lambda_1 \cap \left\{ \lambda \in \mathbb{R}^d : \mathbb{E}_{\mathbb{P}} \left[\|X_T\|^k e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] < +\infty,\ k = 0, 2 \right\}.$$

We will below see that the algorithm is unconstrained if $\Lambda_4 = \mathbb{R}^d$. It is however difficult to check whether or not that is the case, since the the operator F is involved. Meanwhile, to have an unconstrained algorithm, we need at least $\Lambda_4 \subseteq \Lambda_1 = \mathbb{R}^d$. In this sense, let us give a rough illustration of the situation in the following.

Lemma 3.1. *If the Lévy measure ν has a compact support, then $\Lambda_1 = \mathbb{R}^d$. If $\int_{\|z\|>1} e^{\|z\|^{1+\delta}} \nu(dz) < +\infty$ for some $\delta > 0$, then $\Lambda_1 = \mathbb{R}^d$.*

3.1 Constrained algorithms

The following proves the almost sure convergence of the constrained algorithms. Their gradient-based structure simplifies the argument.

Theorem 3.2. *Assume that* $\mathrm{Leb}(\Lambda_4) \in (0, +\infty)$*, and $\lambda_0 \in \Lambda_4$. Let H be a compact set such that $H \subseteq \Lambda_4$. Then, there exists $\lambda^* \in H$ such that the sequence $\{\lambda_n\}_{n \in \mathbb{N}_0}$ in (3.1) converges $\mathbb{P}$-a.s. to λ^*. Moreover, $V(\lambda^*) \leq V(0)$.*

Proof. First, note that $\mathbb{E}_{\mathbb{P}}[e^{-\langle \lambda, X_T \rangle} F(X)^2] < +\infty$ since $\mathbb{E}_{\mathbb{P}}[e^{-2\langle \lambda, X_T \rangle} F(X)^4] < +\infty$, and that by the Cauchy-Schwartz inequality,

$$\mathbb{E}_{\mathbb{P}} \left[\|X_T\|^2 e^{-\langle \lambda, X_T \rangle} F(X)^2 \right]^2 \leq \mathbb{E}_{\mathbb{P}} \left[\|X_T\|^2 \right] \mathbb{E}_{\mathbb{P}} \left[\|X_T\|^2 e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] < +\infty.$$

Hence, $\Lambda_4 \subseteq \Lambda_3$. The convexity of Λ_4 can be proved in a similar manner to the proof of Lemma 2.2. Moreover, we have

$$\mathbb{E}_{\mathbb{P}} \left[\|X_T\| e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right]^2 \leq \mathbb{E}_{\mathbb{P}} \left[e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] \mathbb{E}_{\mathbb{P}} \left[\|X_T\|^2 e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] < +\infty.$$

Now, since

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}} \left[\|Y_n\|^2 \right] \leq \sup_{\lambda \in H} \mathbb{E}_{\mathbb{P}} \left[\|\nabla \varphi_T(\lambda) - X_T\|^2 e^{-2\langle \lambda, X_T \rangle + 2\varphi_T(\lambda)} F(X)^4 \right],$$

and since $\|\nabla \varphi(\lambda)\| < +\infty$ and $\varphi(\lambda) < +\infty$, for $\lambda \in H$, the expectation of the above right hand side is finite if and only if $\mathbb{E}_{\mathbb{P}}[\|X_T\|^k e^{-2\langle \lambda, X_T \rangle} F(X)^4] < +\infty$ for each $k = 0, 1, 2$. This proves $\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}}[\|Y_n\|^2] < +\infty$. Since Λ_4 is convex, it follows from Theorem 2.1 (pp.127) of Kushner and Yin [11] that the sequence $\{\lambda_n\}_{n \in \mathbb{N}_0}$ converges $\mathbb{P}$-*a.s.* to a unique stationary point in H. The last claim holds by the strict convexity of V on H. $\square$

Remark 3.3. It is not clear whether or not there exists $\lambda \in H$ such that $\nabla V(\lambda) = 0$, and thus the above stationary point $\lambda^* \in H$ does not necessarily attain $\nabla V(\lambda^*) = 0$. If, however, there happens to exist $\lambda \in H$ such that $\nabla V(\lambda) = 0$, then $\nabla V(\lambda^*) = 0$ is guaranteed by the strict convexity of V on Λ_4. $\square$

Remark 3.4. We may give some modifications of the set Λ_4 so that it looks more intuitive, as in Remark 3.3. If $F(X) = f(X_T)$ with f being submultiplicative, then Λ_4 can be rewritten as

$$\Lambda_4 = \left\{ \lambda \in \mathbb{R}^d : \int_{\|z\|>1} \left[e^{\langle \lambda, z\rangle} \vee \|z\|^2 e^{-2\langle \lambda, z\rangle} f(z)^4 \right] \nu(dz) < +\infty \right\}.$$

Otherwise, by the Hölder inequality,

$$\Lambda_4' = \left\{ \lambda \in \mathbb{R}^d : \int_{\|z\|>1} \|z\|^{2q} e^{-2q\langle \lambda, z\rangle} \nu(dz) < +\infty, \mathbb{E}_{\mathbb{P}} \left[|F(X)|^{4p} \right] < +\infty, \ \frac{1}{p} + \frac{1}{q} = 1 \text{ for some } p > 1 \right\},$$

which is a convex subset of Λ_4. $\square$

3.2 Unconstrained algorithms

We begin with the main result.

Proposition 3.5. *If $\Lambda_4 = \mathbb{R}^d$ and if there exists $c > 0$ such that*

$$M := \inf_{\|y\|=1} \int_{\|z\|\leq c} \langle y, z\rangle^2 \mathbb{1}_{[0,+\infty)}(\langle y, z\rangle) \nu(dz) > 0, \tag{3.3}$$

then there exists a unique $\lambda^* \in \mathbb{R}^d$ such that $\nabla V(\lambda^*) = 0$.

Remark 3.6. In most applications, Lévy processes are chosen to have independent components, each of which possesses small jumps in both positive and negative directions. Then, their Lévy measures are supported on all the axes of $\mathbb{R}^d$, that is,

$$\nu(dz_1, \ldots, dz_d) = \sum_{k=1}^d \delta_0(dz_1) \cdots \delta_0(dz_{k-1}) \nu_k(dz_k) \delta_0(dz_{k+1}) \cdots \delta_0(dz_d),$$

for some Lévy measures $\{\nu_k\}_{k=1,\ldots,d}$ on $\mathbb{R}_0$. We then get

$$M = \inf_{\|y\|=1} \sum_{k=1}^d y_k^2 \int_{z_k \in (0, \mathrm{sgn}(y_k) c]} z_k^2 \nu_k(dz_k) > 0.$$

In Example 4.2 below, we discretize the sample paths of Lévy processes with both positive and negative jumps on a finite time horizon into a few independent increments, and thus the condition (3.3) holds true. $\square$

Proof. By Proposition 2.3 with $\Lambda_3 \supseteq \Lambda_4 = \mathbb{R}^d$, it suffices to show that $\lim_{\|\lambda\|\uparrow+\infty} V(\lambda) = +\infty$. First, note that with a suitable $\gamma_c \in \mathbb{R}^d$,

$$\varphi(\lambda) - \langle \lambda, \gamma_c \rangle = \int_{\|z\|>c} \left(e^{\langle \lambda, z \rangle} - 1 \right) \nu(dz) + \int_{\|z\|\leq c} \left(e^{\langle \lambda, z \rangle} - 1 - \langle \lambda, z \rangle \right) \nu(dz).$$

The first component of the right hand side above is bounded from below by $-\nu(\{z \in \mathbb{R}_0^d : \|z\| > c\})$ since $\Lambda_1 = \mathbb{R}^d$. For the second component, since $e^x - 1 - x \geq 0$, $x \in \mathbb{R}$, we have

$$\begin{aligned}
\int_{\|z\|\leq c} \left(e^{\langle \lambda, z \rangle} - 1 - \langle \lambda, z \rangle \right) \nu(dz) &\geq \inf_{\|y\|=1} \int_{\|z\|\leq c} \left(e^{\|\lambda\|\langle y, z \rangle} - 1 - \|\lambda\| \langle y, z \rangle \right) \nu(dz) \\
&\geq \inf_{\|y\|=1} \int_{\|z\|\leq c} \left(e^{\|\lambda\|\langle y, z \rangle} - 1 - \|\lambda\| \langle y, z \rangle \right) \mathbb{1}_{[0,+\infty)}(\langle y, z \rangle) \nu(dz) \\
&\geq M \|\lambda\|^2.
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2 \right] &= e^{\varphi_T(\lambda) - T\langle \lambda, \gamma_c \rangle} \mathbb{E}_{\mathbb{P}}\left[e^{-\langle \lambda, X_T - T\gamma_c \rangle} F(X)^2 \right] \\
&\geq e^{\varphi_T(\lambda) - T\langle \lambda, \gamma_c \rangle} \mathbb{E}_{\mathbb{P}}\left[e^{-\|\lambda\|\|X_T - T\gamma_c\|} F(X)^2 \mathbb{1}\left(\|X_T - T\gamma_c\| \leq M\|\lambda\|/2 \right) \right] \\
&\geq e^{T(M\|\lambda\|^2/2 - \nu(\{z \in \mathbb{R}_0^d : \|z\|>c\}))} \mathbb{E}_{\mathbb{P}}\left[F(X)^2 \mathbb{1}(\|X_T - T\gamma_c\| \leq M\|\lambda\|/2) \right],
\end{aligned}$$

which explodes as $\|\lambda\| \uparrow +\infty$. This proves the claim. □

The unconstrained algorithms often show a rough numerical behavior. This phenomenon is mainly due to the extremely fast grow of $\mathbb{E}_{\mathbb{P}}[\|\nabla\varphi_T(\lambda) - X_T\|^2 e^{-2\langle \lambda, X_T \rangle + 2\varphi_T(\lambda)} F(X)^4]$ with respect to $\|\lambda\|$. Alternatively, Chen, Guo and Gao [4] proposes a projection procedure. In essence, by forcing the iterates to stay in an increasing sequence of compact sets, the procedure avoids the explosion of the algorithm during the early stage. Meanwhile, we adapt the results of Chen and Zhu [3] and Delyon [5]. Let $\{H_n\}_{n\in\mathbb{N}_0}$ be an increasing sequence of compact sets such that $\cup_{n\in\mathbb{N}_0} H_n = \mathbb{R}^d$, and modify the algorithm (3.1) as

$$\lambda_{n+1} = \Pi_{H_{\sigma(n)}} \left[\lambda_n - \varepsilon_n Y_{n+1} \right], \tag{3.4}$$

where $\sigma(n)$ counts the number of projections up to the n-th step.

Theorem 3.7. *Assume that $\Lambda_4 = \mathbb{R}^d$ and that there exists a unique λ^* such that $\nabla V(\lambda^*) = 0$. Then, the sequence $\{\lambda_n\}_{n\in\mathbb{N}_0}$ in (3.4) converges $\mathbb{P}$-a.s. to λ^*. Moreover, $\lim_{n\uparrow+\infty} \sigma(n) < +\infty$, $\mathbb{P}$-a.s.*

Proof. Let $m \in \mathbb{N}$ and define for $n \in \mathbb{N}_0$,

$$M_n := \sum_{k=0}^{n} \varepsilon_k \left(Y_{k+1} - \mathbb{E}_{\mathbb{P}}\left[Y_{k+1} \,|\, \mathscr{G}_k \right] \right) \mathbb{1}(\|\lambda_k\| < m).$$

By Proposition 3.5 and the results in [3, 5], we are only to show that for each $m \in \mathbb{N}$, $\{M_n\}_{n\in\mathbb{N}_0}$ converges $\mathbb{P}$-*a.s.* Since the sequence $\{M_n\}_{n\in\mathbb{N}_0}$ is a martingale with respect to the filtration $(\mathscr{G}_n)_{n\in\mathbb{N}_0}$, it suffices to show that $\{M_n\}_{n\in\mathbb{N}_0}$ is a L^2-martingale. To this end, for each $m \in \mathbb{N}$, we will show that, $\mathbb{P}$-*a.s.*,

$$\sum_{n\in\mathbb{N}_0} \varepsilon_n^2 \mathbb{E}_{\mathbb{P}}\left[\|Y_{n+1}\|^2 \mathbb{1}(\|\lambda_n\| \le m) \,\middle|\, \mathscr{G}_n\right]$$
$$= \sum_{n\in\mathbb{N}_0} \varepsilon_n^2 \mathbb{E}_{\mathbb{P}}\left[\|\nabla\varphi_T(\lambda_n) - X_T\|^2 e^{-2\langle\lambda_n, X_T\rangle + 2\varphi_T(\lambda_n)} F(X)^4 \mathbb{1}(\|\lambda_n\| \le m) \,\middle|\, \mathscr{G}_n\right] < +\infty.$$

We begin with proving that for each $m \in \mathbb{N}$, the following four quantities are well defined:

$$C_1(m) := \sup_{\|\lambda\|\le m} \left|\int_{\|z\|>1} \left(e^{\langle\lambda,z\rangle} - 1\right) \nu(dz)\right|,$$
$$C_2(m) := \sup_{\|\lambda\|\le m} \left|\int_{\|z\|\le 1} \left(e^{\langle\lambda,z\rangle} - 1 - \langle\lambda,z\rangle\right) \nu(dz)\right|,$$
$$C_3(m) := \sup_{\|\lambda\|\le m} \int_{\|z\|>1} \|z\| e^{\langle\lambda,z\rangle} \nu(dz),$$
$$C_4(m) := \sup_{\|\lambda\|\le m} \int_{\|z\|\le 1} \|z\| \left|e^{\langle\lambda,z\rangle} - 1\right| \nu(dz).$$

Clearly, $C_1(m)$ is finite since $\Lambda_1 = \mathbb{R}^d$ and $\nu(\{z \in \mathbb{R}_0^d : \|z\| > 1\}) < +\infty$, while the finiteness of $C_2(m)$ follows from $e^{\langle\lambda,z\rangle} - 1 - \langle\lambda,z\rangle \sim \langle\lambda,z\rangle^2 \le \|\lambda\|^2\|z\|^2$ as $\|z\| \downarrow 0$. For $C_3(m)$, the Hölder inequality gives the assertion, that is, with $1/p + 1/q = 1$ for some $p > 1$,

$$\int_{\|z\|>1} \|z\| e^{\langle\lambda,z\rangle} \nu(dz) \le \left[\int_{\|z\|>1} \|z\|^p \nu(dz)\right]^{1/p} \left[\int_{\|z\|>1} e^{\langle q\lambda,z\rangle} \nu(dz)\right]^{1/q} < +\infty,$$

again with the help of $\Lambda_1 = \mathbb{R}^d$. Finally, the finiteness of $C_4(m)$ holds by $\|z\||e^{\langle\lambda,z\rangle} - 1| \sim \|z\||\langle\lambda,z\rangle| \le \|\lambda\|\|z\|^2$ as $\|z\| \downarrow 0$.

Let us now proceed to the main part of the proof. First, as previously, note that

$$\varphi(\lambda) - \langle\lambda,\gamma\rangle = \int_{\|z\|>1} \left(e^{\langle\lambda,z\rangle} - 1\right) \nu(dz) + \int_{\|z\|\le 1} \left(e^{\langle\lambda,z\rangle} - 1 - \langle\lambda,z\rangle\right) \nu(dz).$$

Both the first and the second integrals of the right hand side above are well defined due to the finiteness of $C_1(m)$ and $C_2(m)$, respectively. Hence, we get

$$|\varphi(\lambda) - \langle\lambda,\gamma\rangle| \le C_1(m) + C_2(m) =: C_5(m).$$

Next, note that

$$\nabla(\varphi(\lambda) - \langle\lambda,\gamma\rangle) = \int_{\|z\|>1} z e^{\langle\lambda,z\rangle} \nu(dz) + \int_{\|z\|\le 1} z\left(e^{\langle\lambda,z\rangle} - 1\right) \nu(dz),$$

where the passages to the gradient operator are verified by the finiteness of $C_3(m)$ and $C_4(m)$, and thus

$$\|\nabla(\varphi(\lambda) - \langle\lambda,\gamma\rangle)\| \le C_3(m) + C_4(m) =: C_6(m).$$

In total, we get for each $m \in \mathbb{N}$,

$$\mathbb{E}_{\mathbb{P}}\left[\|\nabla\varphi_T(\lambda)-X_T\|^2 e^{-2\langle\lambda,X_T\rangle+2\varphi_T(\lambda)}F(X)^4\mathbb{1}(\|\lambda\|\le m)\,\middle|\,\mathscr{G}_n\right]$$
$$\le \mathbb{E}_{\mathbb{P}}\left[(\|X_T-\gamma T\|+C_6(m)T)^2 e^{-2\langle\lambda,X_T-\gamma T\rangle+2C_5(m)T}F(X)^4\mathbb{1}(\|\lambda\|\le m)\,\middle|\,\mathscr{G}_n\right],$$

which is bounded $\mathbb{P}$-*a.s.*, since $\Lambda_4=\mathbb{R}^d$. The proof is complete. □

4 Numerical illustrations

In this section, we give two numerical examples, corresponding to the constrained algorithm and the unconstrained one. We will evaluate the efficiency of the importance sampling variance reduction by the ratio of variances (vratio), defined by

$$(\text{vratio}) := \frac{\mathrm{Var}_{\mathbb{P}}(F(X))}{\mathrm{Var}_{\mathbb{Q}_{\lambda_N}}((d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X))}.$$

Example 4.1. (Constrained algorithm) Let $X := (X^1,\dots,X^5)'$ be an infinitely divisible random vector with independent and identically distributed components under the probability measure $\mathbb{P}$, where the common Lévy measure ν on $\mathbb{R}_0$ for each component is of the Meixner type of Schoutens and Teugels [13]. It is characterized by three parameters (a,b,d) in the form

$$\nu(dz) = d\frac{\exp(bz/a)}{z\sinh(\pi z/a)}dz, \quad z\in\mathbb{R}_0,$$

where $a>0$, $b\in(-\pi,\pi)$, and $d>0$, while the probability density function p of X_1 is given in closed form by

$$p(x) = \frac{(2\cos(b/2))^{2d}}{2a\pi\Gamma(2d)}e^{bx/a}\left|\Gamma\left(d+\frac{ix}{a}\right)\right|^2. \tag{4.1}$$

We can derive that

$$\Lambda_1 = \left\{\lambda\in\mathbb{R} : \int_{|z|>1} e^{\lambda z}\nu(dz) < +\infty\right\}^5 = \left(\frac{-\pi-b}{a},\frac{\pi-b}{a}\right)^5,$$

and that for $\lambda\in\Lambda_1$,

$$\varphi(\lambda) = \textstyle\sum_{k=1}^5 2d\left[\ln\left(\cos\frac{b}{2}\right)-\ln\left(\cos\frac{b+a\lambda_k}{2}\right)\right],$$

and

$$\nabla\varphi(\lambda) = \left(2ad\tan\frac{b+a\lambda_1}{2},\cdots,2ad\tan\frac{b+a\lambda_5}{2}\right)'.$$

Consider an Asian payoff

$$F(X) = \max\left[0,\frac{1}{5}\textstyle\sum_{k=1}^5 S_0e^{\sum_{l=1}^k X_l-k\varphi((1,0,0,0,0)')}-K\right].$$

For the condition $\mathbb{E}_{\mathbb{P}}[|F(X)|^{2p}] < +\infty$, it is sufficient to have $\int_{|z|>1} e^{2pz}\nu(dz) < +\infty$. With $p > 1$, we get $p \in (1,+\infty) \cap ((-\pi - b)/(2a), (\pi - b)/(2a))$, provided that $\pi - b > 2a$. Next, the condition $\int_{|z|>1} |z|^{2q} e^{-q\lambda z}\nu(dz) < +\infty$ yields $b/a - q\lambda \in (-\pi/a, \pi/a)$, and in view of the interval of q, we get

$$\Lambda_2' = \Lambda_1 \cap \left(\frac{b-\pi}{a}\left(\left[1 - \frac{2a}{\pi - b}\right] \wedge 1\right), \frac{b+\pi}{a}\left(\left[1 - \frac{2a}{\pi - b}\right] \wedge 1\right)\right)^5,$$

provided that $\pi - b > 2a$. In a similar manner, we can prove $\Lambda_3' = \Lambda_2'$ and

$$\Lambda_4' = \Lambda_1 \cap \left(\frac{b-\pi}{2a}\left(\left[1 - \frac{4a}{\pi - b}\right] \wedge 1\right), \frac{b+\pi}{2a}\left(\left[1 - \frac{4a}{\pi - b}\right] \wedge 1\right)\right)^5,$$

provided that $\pi - b > 4a$.

We set the parameters of the Meixner distribution $(a,b,d) = (0.1, 0.0, 1.0)$, and thus an effective domain is approximately $\Lambda_4' = (-13.707963, 13.707963)^5$. The constraint set H must be compact, so it is safe to set $H = [-13.70796, 13.70796]^5 \subset \Lambda_4'$. We generate $N =$1e+5 Monte Carlo runs with the full help of the closed form density function (4.1). With those runs, we perform the constrained algorithm (3.1) with $\varepsilon_n = \alpha/(n+1)$ and $\lambda_0 = \{0\}$. We examine three cases; the ATM case ($K = 100$), a OTM case ($K = 125$), and a deep OTM case ($K = 150$). The left figures in Figure 1 draw a sequence $\{\|\nabla V(\lambda_n)\|\}_{n\in\mathbb{N}_0}$ of the absolute gradient levels, which is "desired" to achieve $\lim_{n\uparrow+\infty} \|\nabla V(\lambda_n)\| = 0$, $\mathbb{P}$-*a.s.* (As pointed out in Remark 2.4, it is not clear whether or not the constraint set H contains λ^* such that $\nabla V(\lambda^*) = 0$.) The figures on the right present the convergence of the Monte Carlo estimate $\mathbb{E}_{\mathbb{P}}[F(X)]$ (MC) and that of the importance sampling Monte Carlo estimate $\mathbb{E}_{\mathbb{Q}_{\lambda_N}}[(d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X)]$ (IS MC), of which λ_N is the exponential tilting parameter obtained after $N =$1e+5 of the stochastic approximation iterations, while the three vertical lines indicate $\widetilde{C} := \mathbb{E}_{\mathbb{Q}_{\lambda_N}}[(d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X)]$, $0.99\widetilde{C}$ and $1.01\widetilde{C}$.

The absolute gradient level tends to decrease as desired, and the resulting importance sampling succeeds in reducing the Monte Carlo variance. The absolute gradient level seems to have already converged to zero, while we have observed that a component of $\{\lambda_n\}_{n\in\mathbb{N}_0}$ seems to stay at the upper boundary (=+13.70796) in the ATM (K=100) and in the OTM (K=125). Those are delicate issues in the constrained algorithms. □

Example 4.2. (Unconstrained algorithm) Let $X := (X_1, \ldots, X_5)'$ be an infinitely divisible random vector in $\mathbb{R}^5$ with independent and identically distributed components, whose common Lévy measure ν on $\mathbb{R}_0$ of each component is given in the form of the standard Gaussian density function,

$$\nu(dz) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

Evidently, for each $\lambda \in \mathbb{R}$, $\int_{|z|>1} e^{\lambda z}\nu(dz) < +\infty$. Letting $\lambda := (\lambda_1, \ldots, \lambda_5)'$, with the help of the independence of the components, we get $\Lambda_1 = \mathbb{R}^5$,

$$\varphi(\lambda) = \sum_{k=1}^5 \left(e^{\frac{1}{2}\lambda_k^2} - 1\right),$$

and

$$\nabla\varphi(\lambda) = \left(\lambda_1 e^{\frac{1}{2}\lambda_1^2}, \ldots, \lambda_5 e^{\frac{1}{2}\lambda_5^2}\right)'.$$

Due to the compound Poisson structure, the random vector under the probability measure $\mathbb{P}$ can be generated via

$$X \stackrel{\mathscr{L}}{=} \left(\sum_{n=1}^{N_1} W_{1,n}, \ldots, \sum_{n=1}^{N_5} W_{5,n}\right)',$$

where $\{N_n\}_{n\le 5}$ is a sequence of iid Poisson random variables with unit parameter and $\{W_{k,n}\}_{k\le 5, n\in\mathbb{N}}$ is an iid standard Gaussian random array. In view of (2.1), the Lévy measure under the probability measure $\mathbb{Q}_\lambda$ is given by

$$\nu_\lambda(dz) = e^{\lambda z}\nu(dz) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}\lambda^2} e^{-\frac{1}{2}(z-\lambda)^2} dz,$$

which is just like a drift shift of the Gaussian density by λ (up to the constant $e^{\frac{1}{2}\lambda^2}$). Hence, the random vector under the new probability measure $\mathbb{Q}_\lambda$ can be generated via the identity

$$\begin{aligned} X &\stackrel{\mathscr{L}}{=} \left(\sum_{n=1}^{N_1}(W_{1,n}+\lambda_1), \ldots, \sum_{n=1}^{N_5}(W_{5,n}+\lambda_5)\right)' \\ &= \left(\sum_{n=1}^{N_1} W_{1,n} + \lambda_1 N_1, \ldots, \sum_{n=1}^{N_5} W_{5,n} + \lambda_5 N_5\right)', \end{aligned} \tag{4.2}$$

where $\{N_n\}_{n\le 5}$ is now a sequence of iid Poisson random variables with parameter $e^{\frac{1}{2}\lambda^2}(\ge 1)$ and where $\{W_{k,n}\}_{k\le 5, n\in\mathbb{N}}$ remains to be an iid standard Gaussian random array. For any $\lambda \in \mathbb{R}_0^d$, the componentwise variance tends to increase by factor $e^{\frac{1}{2}\lambda_k^2}$, since $\mathbb{E}_{\mathbb{Q}_\lambda}[\sum_{n=1}^{N_k} W_{k,n}] = \mathbb{E}_{\mathbb{Q}_\lambda}[N_k]\mathbb{E}_{\mathbb{Q}_\lambda}[W_{k,1}]$, while the drift shift λ_k is further accelerated by factor $e^{\frac{1}{2}\lambda_k^2}$ on average.

Consider a digital payoff

$$F(X) = \mathbb{1}(S_1 < 100-K, S_2 > 100+K, S_3 < 100-K, S_4 > 100+K, S_5 < 100-K),$$

for a suitable K and where $S_n = 100\exp[\sum_{k=1}^n X_k - n\varphi((1,0,0,0,0)')]$, $n \le 5$. Since $|F(X)| \le 1$, $\mathbb{P}$-*a.s.*, we get $\Lambda_4 = \mathbb{R}^5$. We generate $N =$1e+5 Monte Carlo runs and perform the unconstrained algorithm (3.4) with $\varepsilon_n := \alpha/(n+1)$, $H_n := \{\lambda \in \mathbb{R}^d : \|\lambda\| \le 10\ln(100(n+1))\}$, and $\lambda_0 := \{0\}$. We examine the three cases; $K = 5$, 20, and 40. We present the results in Figure 2, where the three vertical lines in the right figures here indicate $\widetilde{C} := \mathbb{E}_{\mathbb{Q}_{\lambda_N}}[(d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X)]$, $0.98\widetilde{C}$ and $1.02\widetilde{C}$.

In similar to the results in Example 4.1, the algorithm reduces the absolute gradient level, and the resulting importance sampling succeeds in reducing the Monte Carlo variance. Unlike in the constrained algorithm case, we know that there exists a unique optimum λ^* which makes the absolute gradient level zero. It seems that the zero is fairly attained. □

Remark 4.3. In the above numerical illustrations, we have chosen the Lévy measures of the Meixner type and of the Gaussian density. It is a clear reason of the choice that they are somewhat invariant with respect to the Esscher transform owing to the exponential component of their Lévy measure and thus remain very easy to generate in simulation even after the measure change. It should be mentioned here that from a computational point of view, Lévy measures without such an invariance property may not be a good candidate for simulation in our framework. However, this should not be a crucial drawback since most recent popular Lévy measures possess an exponential component, for instance, the Lévy measure of gamma processes. □

5 Concluding remarks

In this paper, we have developed stochastic approximation methods of finding the optimal measure change for Lévy processes in Monte Carlo importance sampling variance reduction. Our analysis is valid on the basis of the restriction to the exponential tilting measure change, that is, limiting the density to a function the terminal value X_T. Nevertheless, our method should be applicable to a variety of applications since its principle is not specific to the structure of Monte Carlo estimator itself. It may be of interest to extend to the intricate series representation setting of [8] by using characterizing parameters of the Lévy measure in the stochastic approximation procedure. Extensions to an optimal parameter search for the combined importance sampling and control variates variance reduction are studied using a two-time-scale version of the stochastic approximation algorithm in subsequent papers [9, 10].

Acknowledgments

The author is grateful to anonymous referees for a careful reading and various useful suggestions, and in particular for remarks which helped improve Proposition 3.5. Part of this work was done at Daiwa Securities SMBC Co.Ltd. The author would like to thank Tatsuya Toshida and the Financial Engineering team for their support and encouragement. This work is supported in part by JSPS Core-to-Core program. At last but not least, it is a great pleasure to express my appreciation to Arturo Kohatsu-Higa for stimulating discussions.

References

[1] Arouna, B. (2004) Robbins-Monro algorithms and variance reduction in finance, *Journal of Computational Finance*, **7**(2) 35-61.

[2] Arouna, B. (2004) Adaptive Monte Carlo method, a variance reduction technique, *Monte Carlo Methods and Applications*, **10**(1) 1-24.

[3] Chen, K.-F., Zhu, Y. (1986) *Stochastic Approximation Procedure with randomly varying truncations,* Scientia Sinica Series.

[4] Chen, H.-F., Guo, L., Gao, A.-J. (1987) Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds, *Stochastic Processes and their Applications*, **27**, 217-231.

[5] Delyon, B. (1996) General results on the convergence of stochastic algorithms, *IEEE Transactions on Automatic Control*, **41**(9) 1245-1255.

[6] Glasserman, P., Heidelberger, P., Shahabuddin, P. (1999) Asymptotic optimal importance sampling and stratification for pricing path-dependent options, *Mathematical Finance*, **9**(2) 117-152.

[7] Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York.

[8] Kawai, R. (2006) An importance sampling method based on the density transformation of Lévy processes, *Monte Carlo Methods and Applications,* **12**(2) 171-186.

[9] Kawai, R. (2007) Adaptive Monte Carlo variance reduction with two-time-scale stochastic approximation. *Monte Carlo Methods and Applications*, **13**(3) 197-217.

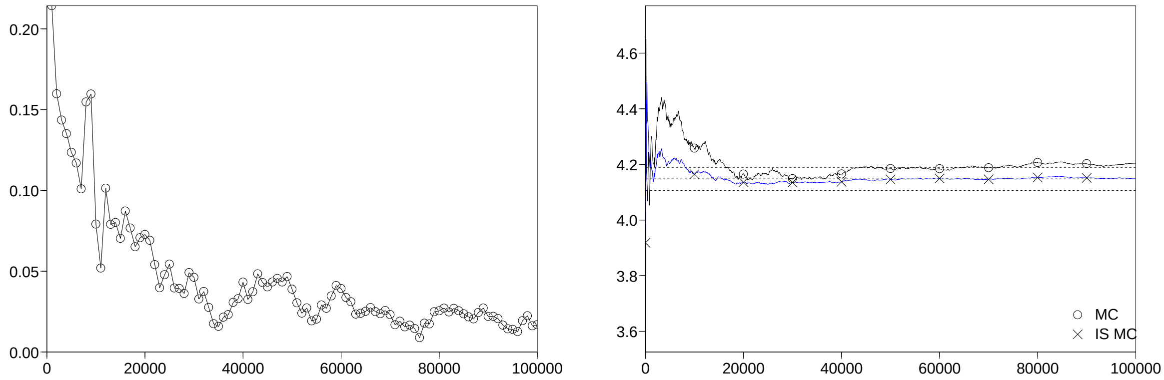
[10] Kawai, R. (2008) Adaptive Monte Carlo variance reduction for Lévy processes with two-time-scale stochastic approximation, *Methodology and Computing in Applied Probability*, **10**(2) 199-223.

[11] Kushner, H.J., Yin, G. (2003) *Stochastic Approximation and Recursive Algorithms and Applications* (2nd ed.), Springer-Verlag, New York.

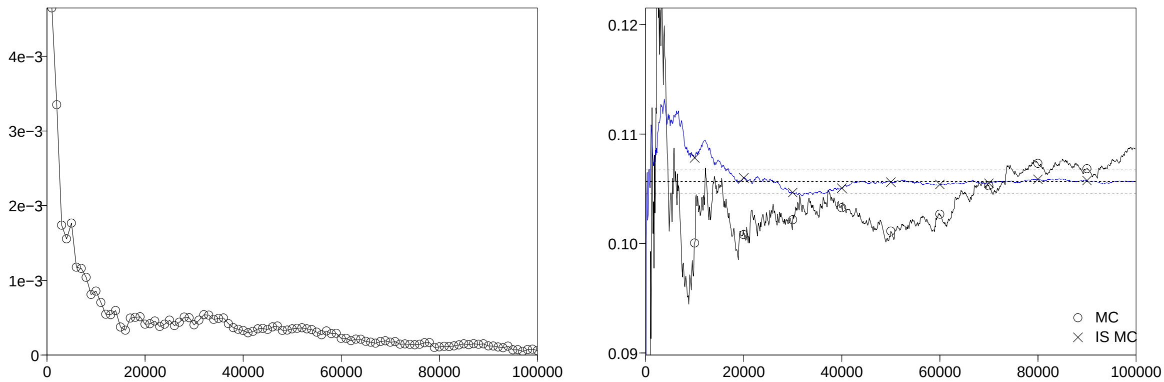
[12] Sato, K. (1999) *Lévy processes and infinitely divisible distributions,* Cambridge University Press, Cambridge.

[13] Schoutens, W. and Teugels, J.L. (1998) Lévy processes, polynomials and martingales, *Communications in Statistics: Stochastic Models,* **14**(1-2) 335-349.

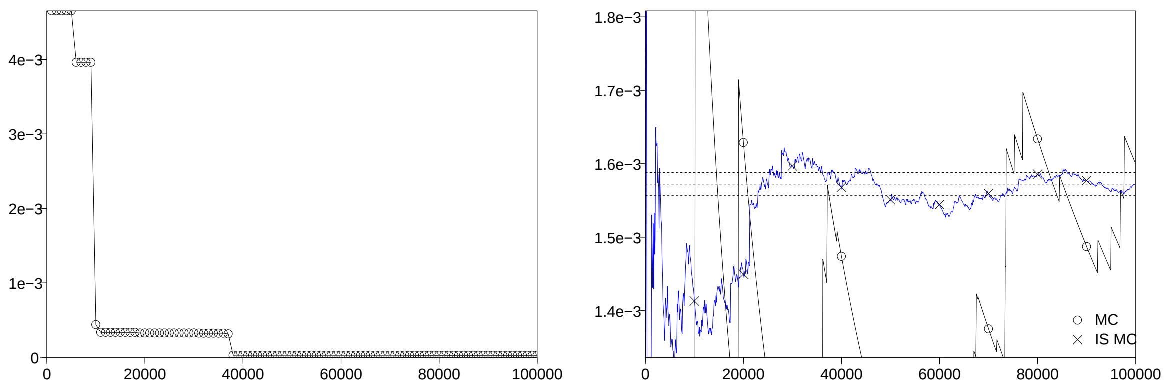
[14] Su, Y., Fu. M.-C. (2002) Optimal importance sampling in securities pricing, *Journal of Computational Finance,* **5**(4) 27-50.



The ATM (K=100) case with $\alpha = 5\text{e}{+}0$. The vratio is 5.063.

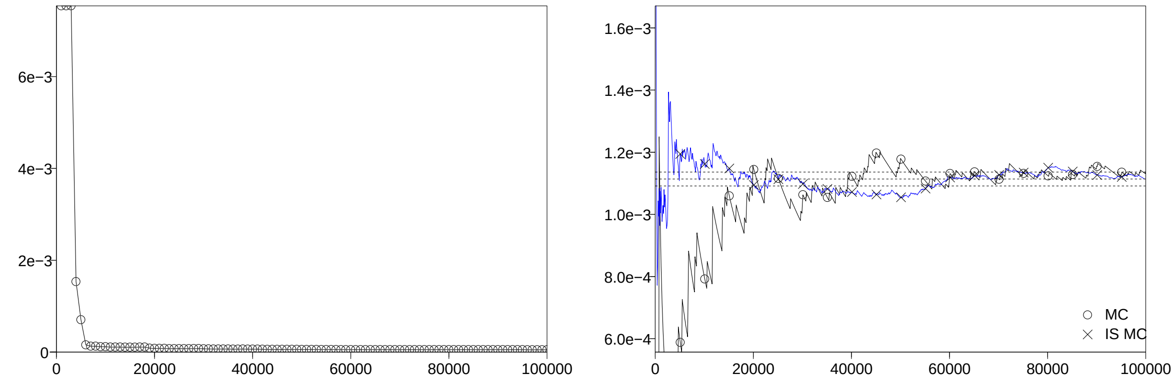


An OTM (K=125) case with $\alpha = 1\text{e}{+}3$. The vratio is 35.26.

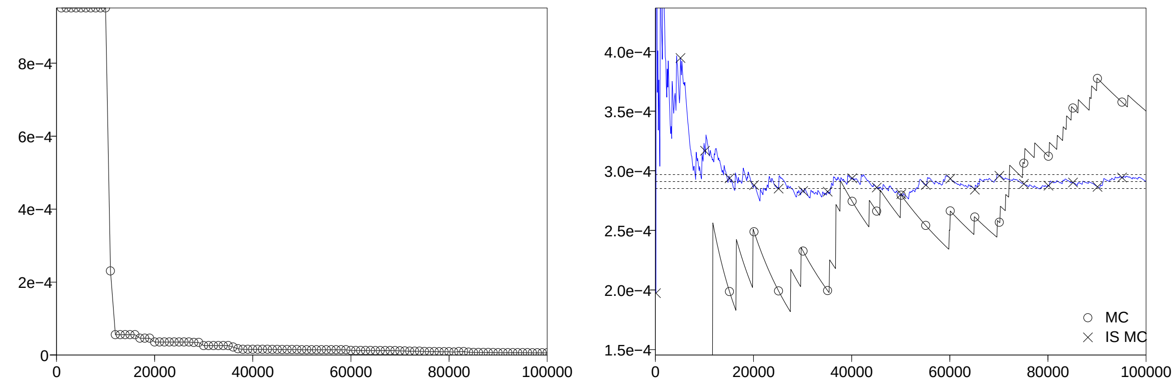


A deep OTM (K=150) case with $\alpha = 1\text{e}{+}4$. The vratio is 60.67.

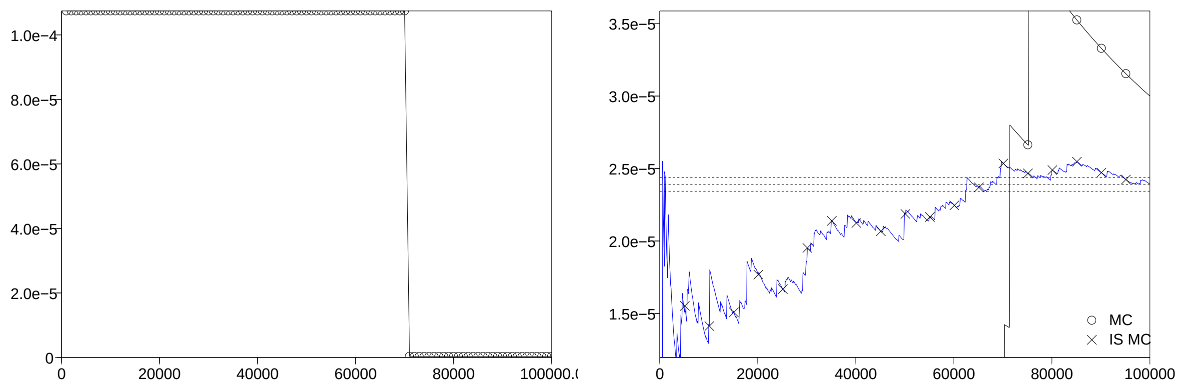
Figure 1: Results for the constrained algorithm; $\{\|\nabla V(\lambda_n)\|\}_{n\in\mathbb{N}_0}$ (left figures), while $\mathbb{E}_{\mathbb{P}}[F(X)]$ (MC) and $\mathbb{E}_{\mathbb{Q}_{\lambda_N}}[(d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X)]$ (IS MC) (right figures).



K=5 with $\alpha = 8\text{e}{+}2$. The vratio is 6.605.



K=20 with $\alpha = 1\text{e}{+}4$. The vratio is 11.75.



K=40 with $\alpha = 3\text{e}{+}4$. The vratio is 57.27.

Figure 2: Results for the unconstrained algorithm; $\{\|\nabla V(\lambda_n)\|\}_{n\in\mathbb{N}_0}$ (left figures), while $\mathbb{E}_{\mathbb{P}}[F(X)]$ (MC) and $\mathbb{E}_{\mathbb{Q}_{\lambda_N}}[(d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X)]$ (IS MC) (right figures).