

*Discussion Paper Series 2008-02*

*Center for the Study of Finance and Insurance, Osaka University*

**Optimal importance sampling parameter search for  
Lévy processes via stochastic approximation**

**Reiichiro Kawai**

January 31, 2008

# 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 has to be 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, Robbins-Monro 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 in 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 minimizing the variance of iid Monte Carlo summands. In the Gaussian framework, it is well known by Girsanov theorem that the measure change is indexed by a single parameter, that

---

First version: February 9, 2007. This version: October 29, 2007.

Running head: Optimal importance sampling for Lévy processes

Email address: reiiichiro\_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.

is, the drift parameter, and several attempts have been made to find the optimum. In financial applications, for example, Glasserman, Heidelberger, and Shahabuddin [6] proposes an optimization setting 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 achieve the root of the gradient, with respect to the measure change parameter, of the Monte Carlo variance.

The aim of the present work is to apply the idea of [1, 14] to a class of Lévy processes with no Brownian motion, or equivalently after discretization, a class of infinitely divisible distributions without Gaussian component. In general, the measure change for Lévy processes involves every single jump, which forms the sample paths of Lévy processes. (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 ourselves to the simplest measure change, often called the Esscher transform, which needs not take jumps of Lévy processes into account but has only to look at the terminal marginals. The Esscher transform is nothing but the well known exponential tilting of distributions and is thus indexed by a single parameter, just as in the Gaussian case. Indeed, the Gaussian measure change is also an exponential tilting, which happens to be equivalent to the drift change due to the structure of the Gaussian distribution. A crucial difference is however that depending on the structure of the underlying Lévy measure, the exponential-tilting parameter for the measure change might have to be in a bounded 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, and 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  $\mathcal{B}(\mathbb{R}_0^d)$  is the Borel  $\sigma$ -field of  $\mathbb{R}_0^d$ .  $(\Omega, \mathcal{F}, \mathbb{P})$  is our underlying probability space.  $\text{Leb}(\cdot)$  denotes the Lebesgue measure, while  $\mathbb{P}|_{\mathcal{F}_t}$  is the restriction of a probability measure  $\mathbb{P}$  to the  $\sigma$ -field  $\mathcal{F}_t$ . Denote by  $\nabla$  the gradient, and by  $\text{Hess}[\cdot]$  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. 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

distributions 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  $\nu$  is a Lévy measure on  $\mathbb{R}_0^d$ , that is,  $\int_{\mathbb{R}_0^d} (\|z\|^2 \wedge 1) \nu(dz) < +\infty$ . When the above holds, then we say that the Lévy process  $\{X_t : t \geq 0\}$  is generated by the triplet  $(\gamma, A, \nu)$ . *In this paper, we restrict ourselves to pure-jump Lévy processes, that is, we set  $A \equiv 0$  throughout. Moreover, we also assume that all components of the marginal  $X_1$  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  $\nu$ . 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 a path-independent structure, which is built by exponential-tilting. The class is often called the Esscher transform.

Let  $\{X_t : t \geq 0\}$  be a Lévy process in  $\mathbb{R}^d$  generated by  $(\gamma, 0, \nu)$ , and let  $(\mathcal{F}_t)_{t \geq 0}$  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, and assume that  $\text{Leb}(\Lambda_1) > 0$  throughout. It is known that the set  $\Lambda_1$  contains the origin and is convex. For  $\lambda \in \Lambda_1$ , we denote by  $\varphi$  the cumulant generating function of the marginal distribution 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 convenience in notation, we also denote by  $\varphi_t(\lambda) := \ln \mathbb{E}_{\mathbb{P}}[e^{\langle \lambda, X_t \rangle}]$ ,  $t > 0$ , 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 distributions of Lévy processes. Note that  $\varphi(\lambda)$  is continuous 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  $t \in (0, +\infty)$ ,

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

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

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

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

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

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, the density function  $p_\lambda$  of  $X_t$  under  $\mathbb{Q}_\lambda$  is given by

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

## 2.2 Importance sampling variance reduction

Suppose we want to evaluate

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

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

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

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

**Lemma 2.2.** *The set  $\Lambda_2$  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

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

The convexity of  $\Lambda_1$  proves the assertion. □

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

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

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  $\text{Leb}(\Lambda_3) > 0$ .

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

*Proof.* The convexity of  $\Lambda_3$  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

$$\mathbb{E}_{\mathbb{P}} \left[ \|X_T\| e^{-\langle \lambda, X_T \rangle} F(X)^2 \right] \leq \mathbb{E}_{\mathbb{P}} \left[ e^{-\langle \lambda, X_T \rangle} F(X)^2 \right]^{1/2} \mathbb{E}_{\mathbb{P}} \left[ \|X_T\|^2 e^{-\langle \lambda, X_T \rangle} F(X)^2 \right]^{1/2} < +\infty,$$

and thus with the help of the dominated convergence theorem, we can 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

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

Then, for  $y \in \mathbb{R}^d$ ,

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

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

**Remark 2.4.** The definition of the sets  $\Lambda_2$  and  $\Lambda_3$  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  $\Lambda'_2$  and  $\Lambda'_3$  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  $\Lambda_2$ , while the definition of  $\Lambda_3$  is verified with  $k = 2$ .

Meanwhile, as soon as  $F(X)$  reduces to  $f(X_T)$  with  $f$  being submultiplicative, the set  $\Lambda_3$  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 simplicity 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

$$\lambda_{n+1} = \Pi_H [\lambda_n - \epsilon_n Y_{n+1}], \quad (3.1)$$

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

$$\sum_{n \in \mathbb{N}_0} \epsilon_n = +\infty, \quad \sum_{n \in \mathbb{N}_0} \epsilon_n^2 < +\infty. \quad (3.2)$$

Moreover, define the filtration  $(\mathcal{G}_n)_{n \in \mathbb{N}_0}$  by

$$\mathcal{G}_n := \sigma(\{\lambda_k\}_{k \leq n}, \{X_k\}_{k \leq n}).$$

In what follows, “the constrained algorithms” means the algorithms where the constraint set  $H$  in (3.1) is a compact set in  $\mathbb{R}^d$  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 sequence  $\{\lambda_n\}_{n \in \mathbb{N}_0}$  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 as long as the iterates remain in  $\Lambda_4$ , the algorithm converges  $\mathbb{P}$ -*a.s.* to a unique stationary point. Hence, if  $\Lambda_4 = \mathbb{R}^d$ , then the algorithm is clearly unconstrained. It is however difficult to check whether or not that is the case, since the set  $\Lambda_4$  depends on the operator  $F$ . Meanwhile, to be unconstrained, 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  $\nu$  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 for the constrained algorithms. Their gradient-based structure simplifies the argument.

**Theorem 3.2.** Assume that  $\text{Leb}(\Lambda_4) \in (0, +\infty)$ , 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  $\lambda^*$ . 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 Hölder inequality,

$$\mathbb{E}_{\mathbb{P}} \left[ \|X_T\|^2 e^{-\langle \lambda, X_T \rangle} F(X)^2 \right] \leq \mathbb{E}_{\mathbb{P}} \left[ \|X_T\|^2 \right]^{1/2} \mathbb{E}_{\mathbb{P}} \left[ \|X_T\|^2 e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right]^{1/2} < +\infty.$$

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

$$\mathbb{E}_{\mathbb{P}} \left[ \|X_T\| e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] \leq \mathbb{E}_{\mathbb{P}} \left[ e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right]^{1/2} \mathbb{E}_{\mathbb{P}} \left[ \|X_T\|^2 e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right]^{1/2} < +\infty.$$

Now, note that

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

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  $\Lambda_4$  is convex, by Theorem 2.1 (pp.127) of Kushner and Yin [11], the sequence  $\{\lambda_n\}_{n \in \mathbb{N}_0}$  converges  $\mathbb{P}$ -a.s. to a unique stationary point in  $H$ . The last assertion follows from 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  $\Lambda_4$ .  $\square$

**Remark 3.4.** We may give some modifications of the set  $\Lambda_4$  so that it looks more intuitive, as in Remark 3.3. If  $F(X) = f(X_T)$  with  $f$  being submultiplicative, then  $\Lambda_4$  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  $\Lambda_4$ .  $\square$

## 3.2 Unconstrained algorithms

The following result asserts the existence of a single point  $\lambda^*$  such that  $\nabla V(\lambda^*) = 0$ . Recall again that in the constrained algorithm case, the limiting stationary point does not necessarily attain the root of the gradient.

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

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

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

**Remark 3.6.** Most applications use Lévy processes with independent components and with each component possessing small jumps in both positive and negative directions. In such cases, their Lévy measures are supported on all the axes of  $\mathbb{R}^d$ , that is,

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

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

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

In Example 4.2 below, we discretize a Lévy process of 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  $\gamma_c \in \mathbb{R}^d$ ,

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

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

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

Therefore, we get

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

which explodes as  $\|\lambda\| \uparrow +\infty$ . This proves the assertion.  $\square$

The unconstrained algorithms often show a rough numerical behavior, and unfortunately, that is the case in our setting. 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, 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

$$\lambda_{n+1} = \Pi_{H_{\sigma(n)}}[\lambda_n - \epsilon_n Y_{n+1}], \quad (3.4)$$

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  $\lambda^*$  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  $\lambda^*$ . Moreover,  $\lim_{n \uparrow +\infty} \sigma(n) < +\infty$ ,  $\mathbb{P}$ -a.s.*

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

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

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  $(\mathcal{G}_n)_{n \in \mathbb{N}_0}$ , 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.,

$$\begin{aligned} & \sum_{n \in \mathbb{N}_0} \epsilon_n^2 \mathbb{E}_{\mathbb{P}} \left[ \|Y_{n+1}\|^2 \mathbb{1}(\|\lambda_n\| \leq m) \mid \mathcal{G}_n \right] \\ &= \sum_{n \in \mathbb{N}_0} \epsilon_n^2 \mathbb{E}_{\mathbb{P}} \left[ \|\nabla\varphi_T(\lambda_n) - X_T\|^2 e^{-2\langle\lambda_n, X_T\rangle + 2\varphi_T(\lambda_n)} F(X)^4 \mathbb{1}(\|\lambda_n\| \leq m) \mid \mathcal{G}_n \right] < +\infty. \end{aligned}$$

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

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

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  $e^{\langle\lambda, z\rangle} - 1 - \langle\lambda, z\rangle \sim \langle\lambda, z\rangle^2 \leq \|\lambda\|^2 \|z\|^2$  as  $\|z\| \downarrow 0$ . For  $C_3(m)$ , the Hölder inequality gives the assertion, that is, with  $1/p + 1/q = 1$  for some  $p > 1$ ,

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

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

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

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

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

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

Next, note that

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

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

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

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

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

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

## 4 Numerical examples

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

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

**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  $\nu$  on  $\mathbb{R}_0$  for each component is the Meixner type of Schoutens and Teugels [13], characterized by three parameters  $(a, b, d)$  in the form of

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

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

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

It is easy to prove that

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

Then, for  $\lambda \in \Lambda_1$ , we can obtain

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

Since the exponential tail decay of the Lévy measure dominates each polynomial moment, we get, for  $\lambda \in \Lambda_1$  and for each  $k \in \mathbb{N}$ ,  $\int_{|z|>1} |z|^k e^{\lambda z} \nu(dz) < +\infty$ , or equivalently  $\mathbb{E}_{\mathbb{P}}[\|X\|^k e^{\langle \lambda, X \rangle}] < +\infty$ . This verifies the finiteness of every polynomial moment and thus

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

For a numerical example, we will consider an Asian payoff

$$F(X) = \max \left[ 0, \frac{1}{5} \sum_{k=1}^5 S_0 e^{\sum_{i=1}^k X_i - k\varphi((1,0,0,0,0)')} - K \right].$$

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

$$\Lambda'_2 = \Lambda_1 \cap \left( \frac{b - \pi}{a} \left( \left[ 1 - \frac{2a}{\pi - b} \right] \wedge 1 \right), \frac{b + \pi}{a} \left( \left[ 1 - \frac{2a}{\pi - b} \right] \wedge 1 \right) \right)^5,$$

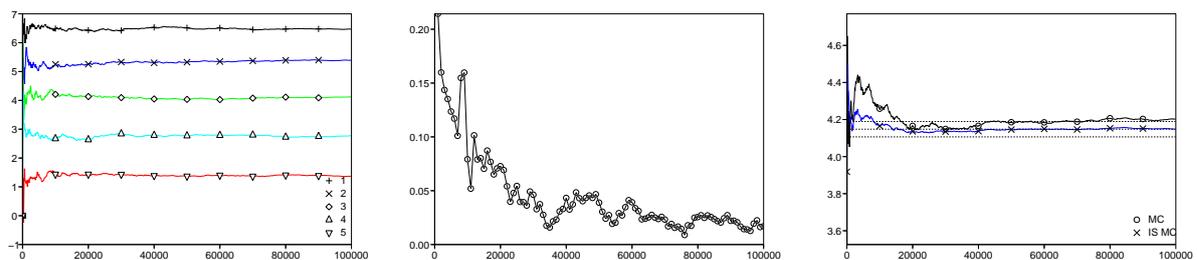
provided that  $\pi - b > 2a$ . In a similar manner, we can prove  $\Lambda'_3 = \Lambda'_2$  and

$$\Lambda'_4 = \Lambda_1 \cap \left( \frac{b - \pi}{2a} \left( \left[ 1 - \frac{4a}{\pi - b} \right] \wedge 1 \right), \frac{b + \pi}{2a} \left( \left[ 1 - \frac{4a}{\pi - b} \right] \wedge 1 \right) \right)^5,$$

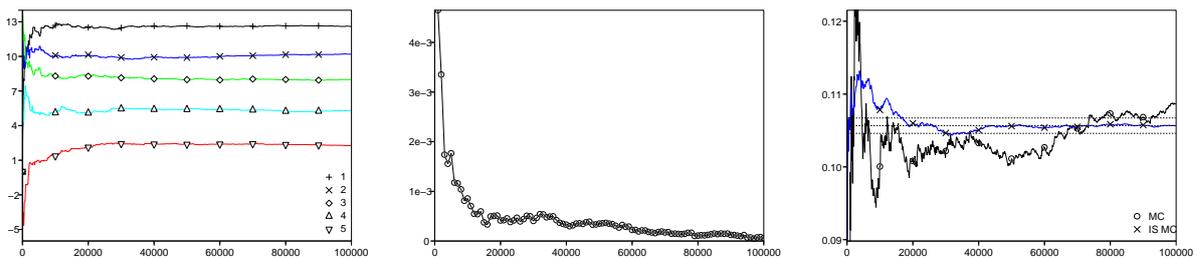
provided that  $\pi - b > 4a$ .

We set the parameters for the Meixner distribution by  $(a, b, d) = (0.1, 0.0, 1.0)$ , 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 summands with the full help of the closed-form density function (4.1). With those runs, we perform the constrained algorithm (3.1) with  $\epsilon_n = \alpha/(n + 1)$  and  $\lambda_0 = \{0\}$ . For the ATM

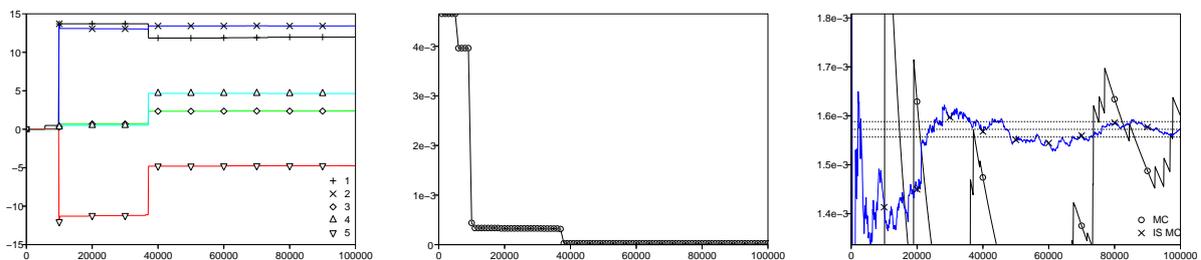
case ( $K = 100$ ), a OTM case ( $K = 125$ ), and a deep OTM case ( $K = 150$ ), we draw in Figure 1 a sequence  $\{\lambda_n\}_{n \in \mathbb{N}_0}$  of iteratively generated exponential-tilting parameters in  $\mathbb{R}^5$  (left), and a sequence  $\{\|\nabla V(\lambda_n)\|\}_{n \in \mathbb{N}_0}$  of the absolute gradient levels (right), 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  $\lambda^*$  such that  $\nabla V(\lambda^*) = 0$ .) Finally, each figure on the right presents 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  $\lambda_N$  is the exponential-tilting parameter obtained after  $1e+5$  of the stochastic approximation iterations, while the three vertical lines indicate  $\tilde{C} := \mathbb{E}_{\mathbb{Q}_{\lambda_N}}[(d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X)]$ ,  $0.99\tilde{C}$  and  $1.01\tilde{C}$ .



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



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



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

Figure 1:  $\{\lambda_n\}_{n \in \mathbb{N}_0}$  (left) and  $\{\|\nabla V(\lambda_n)\|\}_{n \in \mathbb{N}_0}$  (middle), 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).

The absolute gradient level tends to decrease as expected, and the resulting importance sampling succeeds in reducing the Monte Carlo variance. Although the sequence  $\{\lambda_n\}_{n \in \mathbb{N}_0}$  seems to have converged to a point in the interior of  $H$ , the absolute gradient level has not converged to zero. On contrary, the absolute gradient level seems to have already converged to zero, while 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 and unsolved matters in the constrained algorithms.

**Example 4.2.** (Unconstrained algorithm) Let  $X := (X_1, \dots, X_5)'$  be an infinitely divisible random vector in  $\mathbb{R}^5$  with independent and identically distributed components, whose common Lévy measure  $\nu$  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  $\lambda := (\lambda_1, \dots, \lambda_5)'$ , 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

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

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

$$X \leftarrow \left( \sum_{n=1}^{N_1} W_{1,n}, \dots, \sum_{n=1}^{N_5} W_{5,n} \right)',$$

where  $\{N_n\}_{n \leq 5}$  is a sequence of iid Poisson random variables with unit parameter and  $\{W_{k,n}\}_{k \leq 5, n \in \mathbb{N}}$  is an iid standard Gaussian random array. Now, 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  $\lambda$  (up to the constant  $e^{\frac{1}{2}\lambda^2}$ ). Then, the random vector under the new probability measure  $\mathbb{Q}_\lambda$  can be generated as

$$\begin{aligned} X &\leftarrow \left( \sum_{n=1}^{N_1} (W_{1,n} + \lambda_1), \dots, \sum_{n=1}^{N_5} (W_{5,n} + \lambda_5) \right)' \\ &= \left( \sum_{n=1}^{N_1} W_{1,n} + \lambda_1 N_1, \dots, \sum_{n=1}^{N_5} W_{5,n} + \lambda_5 N_5 \right)', \end{aligned} \quad (4.2)$$

where  $\{N_n\}_{n \leq 5}$  is now a sequence of iid Poisson random variables with parameter  $e^{\frac{1}{2}\lambda^2} (\geq 1)$  and where  $\{W_{k,n}\}_{k \leq 5, n \in \mathbb{N}}$  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  $\lambda_k$  is further accelerated by factor  $e^{\frac{1}{2}\lambda_k^2}$  on average.

In the above setting, we 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  $S_n = 100 \exp[\sum_{k=1}^n X_k - n\varphi((1, 0, 0, 0, 0)')]$ ,  $n \leq 5$ . Since  $|F(X)| \leq 1$ ,  $\mathbb{P}$ -*a.s.*, we get  $\Lambda_4 = \mathbb{R}^5$ . We generate  $N = 1e+5$  Monte Carlo summands and perform the unconstrained algorithm (3.4) with  $\epsilon_n := \alpha/(n+1)$ ,  $H_n := \{\lambda \in \mathbb{R}^d : \|\lambda\| \leq 10 \ln(100(n+1))\}$ , and  $\lambda_0 := \{0\}$ . In Figure 2, for  $K = 5, 20$ , and  $40$ , we draw a sequence  $\{\lambda_n\}_{n \in \mathbb{N}_0}$  of iteratively generated exponential-tilting parameters in  $\mathbb{R}^5$  (left), and a sequence  $\{\|\nabla V(\lambda_n)\|\}_{n \in \mathbb{N}_0}$  of the absolute gradient levels (right), which is “supposed” to achieve  $\lim_{n \uparrow +\infty} \|\nabla V(\lambda_n)\| = 0$ ,  $\mathbb{P}$ -*a.s.* Each figure on the right presents 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  $\lambda_N$  is the exponential-tilting parameter obtained after  $1e+5$  of the stochastic approximation iterations, while the three vertical lines indicate  $\tilde{C} := \mathbb{E}_{\mathbb{Q}_{\lambda_N}}[(d\mathbb{P}/d\mathbb{Q}_{\lambda_N})F(X)]$ ,  $0.98\tilde{C}$  and  $1.02\tilde{C}$ .

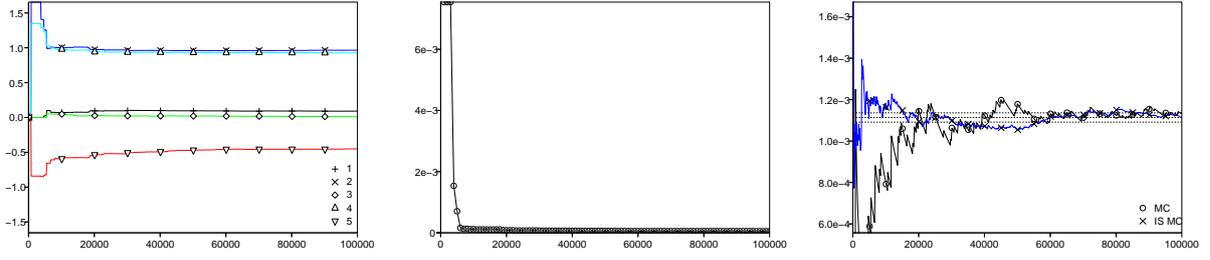
As the result in Example 4.1, the algorithm reduces the absolute gradient level as expected, and the resulting importance sampling succeeds in reducing the Monte Carlo variance. Unlike in the constrained algorithm case, there exists a unique optimum  $\lambda^*$  which makes the absolute gradient level zero, which seems to be in fact nearly attained.

## 5 Concluding remarks

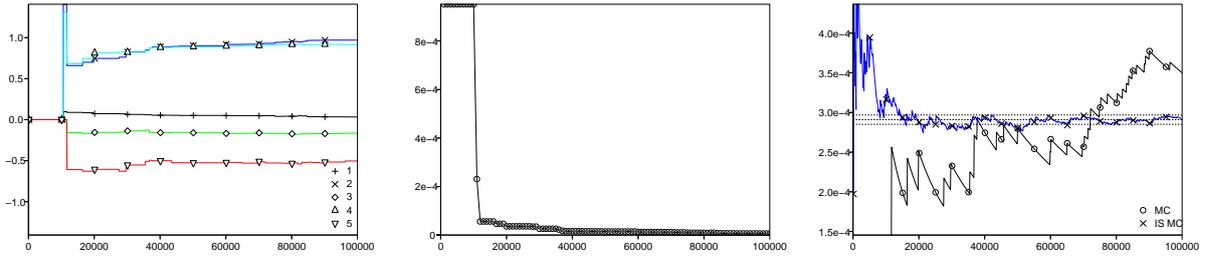
In this paper, we have studied stochastic approximation methods of finding the optimal measure change for Lévy processes in Monte Carlo importance sampling variance reduction. Our analysis is, on one hand, valid on the basis of the restriction to the exponential-tilting measure change, that is, limiting the density to a function the terminal value  $X_T$ . It would thus be very interesting to extend to the intricate series representation setting of [8] by using characterizing parameters of the Lévy measure in the stochastic approximation procedure, which will be studied elsewhere. On the other hand, our method should be applicable to a variety of applications in the sense that its principle is not specific to the structure of Monte Carlo estimator itself. Finally, the extension of our framework to the combined importance sampling and control variates is investigated using a two-time-scale version of the stochastic approximation algorithm in subsequent papers [9, 10].

## Acknowledgments

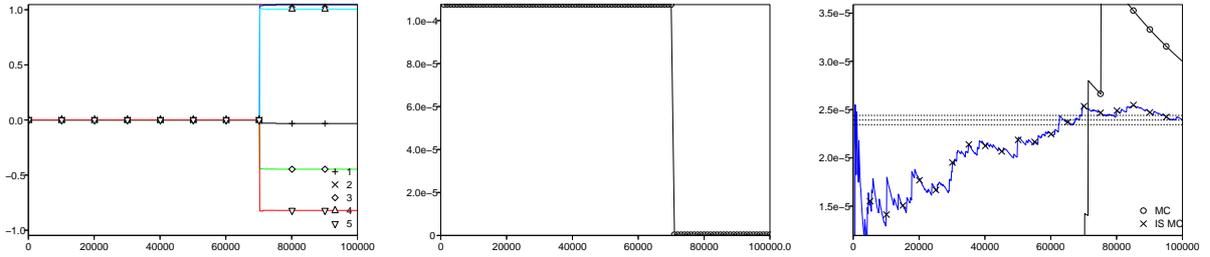
The author is grateful to anonymous referees for a careful reading and various suggestions, and in particular for remarks which helped improve Proposition 3.5. It is a pleasure to express many thanks to Arturo Kohatsu-Higa for stimulating discussions.



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



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



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

Figure 2:  $\{\lambda_n\}_{n \in \mathbb{N}_0}$  (left) and  $\{\|\nabla V(\lambda_n)\|\}_{n \in \mathbb{N}_0}$  (middle), 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).

## 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. to appear in *Methodology and Computing in Applied Probability*, **10**(1).
- [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.