



**HAL**  
open science

# Multilevel Monte Carlo methods and statistical inference for financial models

Ahmed Kebaier

► **To cite this version:**

Ahmed Kebaier. Multilevel Monte Carlo methods and statistical inference for financial models. Probability [math.PR]. Université Paris 13, 2017. tel-01684717v2

**HAL Id: tel-01684717**

**<https://theses.hal.science/tel-01684717v2>**

Submitted on 23 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LAGA - CNRS UMR 7539  
UNIVERSITÉ PARIS 13

Mémoire d'habilitation à diriger des recherches

par

Ahmed KEBAIER

Méthodes Multilevel Monte Carlo et Statistiques  
des Processus en Finance

Soutenu le 1<sup>er</sup> Décembre 2017

Jury :

*Rapporteurs :*

M. Antoine LEJAY,	Directeur de recherche, INRIA.
M. Gilles PAGÈS,	Professeur à l'Université Pierre et Marie Curie.
M. Mark PODOLSKIJ,	Professeur à l'Université d'Aarhus, Danemark.

*Examineurs :*

Mme. Mireille BOSSY,	Directrice de recherche, INRIA.
M. Jean-Stéphane DHERSIN,	Professeur à l'Université Paris 13.
M. Yueyun HU,	Professeur à l'Université Paris 13.
M. Benjamin JOURDAIN,	Professeur à l'École des Ponts Paritech.
M. Damien LAMBERTON,	Professeur à l'Université Paris-Est.
M. Fabien PANLOUP,	Professeur à l'Université d'Angers.



# Remerciements

Il serait difficile de trouver les mots justes pour exprimer ma profonde reconnaissance à Vlad Bally, Faouzi Chaâbane & Damien Lambertson pour la passion qu'ils m'ont transmise pour mes thématiques de recherche, les divers domaines qu'ils m'ont invité à découvrir et les techniques qu'ils m'ont enseignées. Je suis très honoré par la présence de Damien Lambertson dans mon jury.

Je mesure mon privilège de pouvoir effectuer ma recherche dans deux cadres agréables et stimulants le LAGA et le CERMICS. A ce titre, je remercie chaleureusement l'ensemble de mes collègues. Plus particulièrement, je suis très reconnaissant envers Jean-Stéphane Dhersin & Yueyun Hu pour leurs encouragements continus, leur présence au jury me fait grand plaisir. Aussi, je ne saurais assez exprimer ma gratitude envers Benjamin Jourdain qui m'a offert les meilleures conditions au CERMICS pour mener à bien ces travaux, j'ai de la chance de pouvoir partager avec lui de nombreuses visions mathématiques et discussions inspirantes. Je suis très honoré par sa présence au jury. Mes pensées vont également à Aurélien Alfonsi et Mohamed Ben Alaya avec qui travailler m'a été et m'est toujours très agréable. J'en profite pour remercier tous mes collaborateurs pour les discussions enrichissantes que nous avons pu avoir.

Je tiens à exprimer ma gratitude à Gilles Pagès qui a accepté de présenter mon habilitation à diriger des recherches. Je le remercie pour l'intérêt continu qu'il porte à mes travaux. Je remercie chaleureusement Antoine Lejay & Mark Podolskij de m'avoir consacré du temps en écrivant un rapport sur mes travaux. Leur présence au jury me fait grand plaisir. Je suis également enchanté que Mireille Bossy et Fabien Panloup aient accepté d'en faire partie.

Je souhaite ensuite exprimer ma profonde reconnaissance à la chaire "Risques Financiers" et particulièrement à Nicole El Karoui, Emmanuel Gobet & Nizar Touzi pour leurs encouragements et pour l'intérêt qu'ils portent à mes travaux. Je remercie aussi toute l'équipe du projet INRIA "Mathrisk" et notamment Agnès Sulem et Antonino Zanette. Suite à l'invitation d'Arturo Kohatsu-Higa, j'ai passé un séjour remarquable à l'université de Ritsumeikan. Je souhaite lui exprimer toute ma gratitude pour son hospitalité et toutes les discussions mathématiques enrichissantes et fructueuses que nous avons pu partager. Je ne saurais oublier de remercier Francesco Russo qui a accompagné mes débuts au sein de l'équipe de probabilités & statistiques de l'université Paris 13.

Enfin, je souhaite remercier mes parents Mohamed & Najoua, mon épouse Nada, ma fille Dina & mon frère Yasser pour leur soutien moral et affectif inébranlable. Un grand merci du fond du cœur à ma grande famille pour l'amour sincère qu'elle me porte.



# Contents

<b>I</b>	<b>Multilevel Monte Carlo approximations</b>	<b>11</b>
<b>1</b>	<b>Asymptotic and non-asymptotic analysis of the Multilevel Monte Carlo approximation algorithms</b>	<b>13</b>
1.1	Introduction and motivation . . . . .	13
1.2	A two-level Monte Carlo approximation: the Statistical Romberg method	15
1.3	The Statistical Romberg method for density estimation . . . . .	16
1.4	Asymptotic analysis of the MLMC method . . . . .	19
1.5	Non-asymptotic analysis of the MLMC method . . . . .	24
1.5.1	MLMC parameters optimization revisited . . . . .	25
1.5.2	Concentration bounds . . . . .	26
1.6	Perspectives . . . . .	27
<b>2</b>	<b>Coupling Importance Sampling and Multilevel Monte Carlo</b>	<b>29</b>
2.1	Introduction . . . . .	29
2.2	Importance Sampling and Statistical Romberg method: a stochastic algorithm approach for Brownian diffusions . . . . .	30
2.2.1	The constrained stochastic algorithm . . . . .	31
2.2.2	The unconstrained stochastic algorithm . . . . .	32
2.2.3	A central limit theorem for the adaptive SR algorithm . . . . .	33
2.3	Importance Sampling and MLMC method: a stochastic algorithm approach for Brownian diffusions . . . . .	33
2.4	Importance Sampling and MLMC method: a sample average approach optimization for Brownian diffusions . . . . .	36
2.5	Importance Sampling and Statistical Romberg method for pure-jump Lévy processes . . . . .	39
2.6	Perspectives . . . . .	43
<b>II</b>	<b>Statistics for financial models: estimation from continuous and discrete time observations</b>	<b>45</b>
<b>3</b>	<b>Statistical properties for continuously observed financial models with and without jumps</b>	<b>47</b>
3.1	Introduction and motivation . . . . .	47
3.2	Maximum likelihood estimation for CIR processes . . . . .	48
3.3	Maximum likelihood estimation for Wishart processes . . . . .	51

3.4	Statistical inference for jump-type square root models . . . . .	55
3.4.1	Maximum likelihood estimation for stochastic volatility with jumps model (SVJ) . . . . .	55
3.4.2	Maximum likelihood estimation for the jump-type CIR model . . .	57
3.5	Perspectives . . . . .	60
<b>4</b>	<b>Statistical properties for a discretely observed CIR model</b>	<b>61</b>
4.1	Parameter estimation in a high frequency setting . . . . .	61
4.2	Local asymptotic properties of CIR process discretely observed at high frequency . . . . .	63
4.3	Perspectives . . . . .	65
<b>III</b>	<b>Weighted limit theorems for quasi-left continuous-time vector martingales</b>	<b>67</b>
<b>5</b>	<b>Generalized logarithmic limit theorems for quasi-left continuous mar- tingales</b>	<b>69</b>
5.1	Introduction and motivation . . . . .	69
5.2	Martingale with regular normalization . . . . .	70
5.3	Martingale with explosive normalization . . . . .	71
5.4	Application example: estimation of the variance of processes with station- ary independent increments . . . . .	72
5.5	Perspectives . . . . .	72

# Introduction

Dans ce document, nous présentons une synthèse de nos travaux de recherche en probabilité théorique et numérique ainsi qu'en statistiques des processus en finance. Nous avons regroupé ces travaux selon trois thèmes.

La première partie, dédiée à l'analyse des méthodes Multilevel Monte Carlo et de leurs applications en finance, comporte deux chapitres.

Le premier chapitre est consacré à des études asymptotiques et non-asymptotiques des méthodes Multilevel Monte Carlo. La première section présente une introduction et une motivation au problème de calcul d'espérance de fonctions ou de fonctionnelles d'une diffusion. Dans la deuxième section, nous analysons la méthode de Romberg Statistique ou two-level Monte Carlo pour des schémas de discrétisation de type Euler. Dans la troisième section, nous nous intéressons au comportement asymptotique de la loi de la dérivée de Malliavin de l'erreur de discrétisation et son application à l'analyse asymptotique de la méthode de Romberg Statistique pour le problème d'estimation de densités d'équations différentielles stochastiques. Dans la quatrième section, nous établissons un théorème central limite pour la méthode Multilevel Monte Carlo. Dans la dernière section, nous présentons des inégalités de concentrations Gaussiennes établies pour la méthode Multilevel Monte Carlo grâce à des techniques de calcul de Malliavin.

Le deuxième chapitre traite du couplage des méthodes Multilevel Monte Carlo avec les techniques de type importance sampling. Après une brève section introductive, nous nous focalisons dans une deuxième section sur l'étude d'une version adaptative de la méthode de Romberg Statistique combinée avec des versions tronquées et non-tronquées d'algorithmes stochastiques permettant de réduire la variance du problème, dans le cadre de diffusions Browniennes discrétisées. Dans la troisième section, nous étudions l'extension de ces résultats aux méthodes Multilevel Monte Carlo. La quatrième section traite de l'utilisation d'algorithmes déterministes de type Newton pour l'approximation du paramètre optimal de la méthode Multilevel Monte Carlo avec importance sampling. Dans la dernière section, nous nous intéressons à un nouveau concept d'utilisation de la méthode de Romberg Statistique et par défaut des méthodes Multilevel Monte Carlo dans le cadre où la diffusion Brownienne est remplacée par un processus de Lévy de sauts purs donc sans schéma de discrétisation. Nous présentons une analyse de cette nouvelle procédure ainsi que de son implémentation avec des techniques de type importance sampling. Des illustrations numériques de l'efficacité des algorithmes présentés dans cette partie sont fournies à travers des exemples concrets issus de la finance quantitative.



La deuxième partie est consacrée au problème d'inférence statistique des modèles appliqués en mathématiques financières. Elle comporte deux autres chapitres, décrits ci-dessous.

Le troisième chapitre traite du cadre des observations continues. Après une brève section introductive, nous nous intéressons dans une deuxième section au problème d'estimation des paramètres de drift dans le processus de Cox-Ingersoll-Ross. La troisième section est consacrée à l'étude du problème d'estimation paramétrique dans les processus matriciels de type Wishart. La dernière section développe le problème d'inférence statistique pour des modèles à volatilité stochastique en présence de sauts, notamment pour les modèles de Heston et de Cox-Ingersoll-Ross. Pour chacune de ces études nous examinons le comportement en temps long des estimateurs de type maximum de vraisemblance dans les régimes érgodique et non-érgodique en explicitant à chaque fois les différentes fonctions génératrices de moments des fonctionnelles liées aux processus étudiés.

Le quatrième chapitre est consacré au cadre des observations discrètes. Dans une première section, nous fournissons des conditions suffisantes portant sur la fréquence des observations du processus de Cox-Ingersoll-Ross pour garantir la validité des théorèmes limites obtenus dans le Chapitre 3 pour le cadre des observations continues. Dans la dernière section, nous nous basons sur des techniques de calcul de Malliavin récemment développées dans le cadre du processus de Cox-Ingersoll-Ross pour démontrer des propriétés asymptotiques locales de type LAN, LAMN et LAQ associées à l'estimation des paramètres de drift dans ce processus couvrant les régimes érgodique et non-érgodique.

La dernière partie, composée d'un seul et dernier chapitre, étend divers résultats théoriques de type théorème limite centrale presque sûre à des martingales quasi-continues à gauche. Nous développons ces résultats pour des normalisations régulières, explosives et mixtes. Nous illustrons à travers un exemple jouet l'application de ces résultats théoriques en pratique.

# List of articles

Published or accepted articles:

- [1] Kebaier, A. (2005) Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, **15**, 2681–2705.
- [2] Chaâbane, F. and Kebaier, A. (2008) Théorèmes limites avec poids pour les martingales vectorielles à temps continu. *ESAIM. Probability and Statistics*, **12**, 464–491.
- [3] Kebaier, A. and Kohatsu-Higa, A. (2008) An optimal control variance reduction method for density estimation. *Stochastic Processes and their Applications*, **118**, 2143–2180.
- [4] Fathallah, H. and Kebaier, A. (2012) Weighted limit theorems for continuous-time vector martingales with explosive and mixed growth. *Stochastic Analysis and Applications*, **30**, 238–257.
- [5] Ben Alaya, M. and Kebaier, A. (2012) Parameter estimation for the square-root diffusions: ergodic and nonergodic cases. *Stochastic Models*, **28**, 609–634.
- [6] Ben Alaya, M. and Kebaier, A. (2013) Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stochastic Analysis and Applications*, **31**, 552–573.
- [7] Ben Alaya, M. and Kebaier, A. (2014) Multilevel Monte Carlo for Asian options and limit theorems. *Monte Carlo Methods and Applications*, **20**, 181–194.
- [8] Ben Alaya, M. and Kebaier, A. (2015) Central limit theorem for the multilevel Monte Carlo Euler method. *The Annals of Applied Probability*, **25**, 211–234.
- [9] Ben Alaya, M., Hajji, K., and Kebaier, A. (2015) Importance sampling and statistical Romberg method. *Bernoulli*, **21**, 1947–1983.
- [10] Ben Alaya, M., Hajji, K., and Kebaier, A. (2016) Importance sampling and statistical Romberg method for Lévy processes. *Stochastic Processes and their Applications*, **126**, 1901–1931.
- [11] Alfonsi, A., Kebaier, A., and Rey, C. (2016) Maximum likelihood estimation for Wishart processes. *Stochastic Processes and their Applications*, **126**, 3243–3282.

- [12] Kebaier, A. and Lelong, J. (2017) Coupling importance sampling and multilevel monte carlo using sample average approximation. To appear in *Methodology and Computing in Applied Probability*.
- [13] Barczy, M., Ben Alaya, M., Kebaier, A., and Pap, G. (2017) Asymptotic properties of maximum likelihood estimator for the growth rate for a jump-type CIR process based on continuous time observations. To appear in *Stochastic Processes and their Applications*.

Submitted articles:

- [14] Barczy, M., Ben Alaya, M., Kebaier, A., and Pap, G. (2015) Asymptotic behavior of maximum likelihood estimators for a jump-type Heston model.  
<https://arxiv.org/pdf/1509.08869.pdf>.
- [15] Ben Alaya, M., Hajji, K., and Kebaier, A. (2016) Improved adaptive Multilevel Monte Carlo and applications to finance.  
<https://arxiv.org/pdf/1603.02959.pdf>.
- [16] Ben Alaya, M., Kebaier, A., and Tran, N. K. (2017) Local asymptotic properties for Cox-Ingersoll-Ross process with discrete observations.  
<https://arxiv.org/pdf/1708.07070.pdf>.
- [17] Jourdain, B., and Kebaier, A. (2017) Non-asymptotic error bounds for The Multi-level Monte Carlo Euler method applied to SDEs with constant diffusion coefficient.  
<https://arxiv.org/pdf/1708.07064.pdf>.

**Part I**

**Multilevel Monte Carlo  
approximations**



# Chapter 1

## Asymptotic and non-asymptotic analysis of the Multilevel Monte Carlo approximation algorithms

### 1.1 Introduction and motivation

The popularity of the Monte Carlo methods is due to their undisputed convenience in high dimensional settings and reliability, although they are also known to converge slowly. In particular in finance, one of the most problematic tasks faced by practitioners nowadays, is computing in the shortest possible time, risk indicators or similar quantities by Monte Carlo methods using huge samples of data. Worthy of mentioning is the future legislation: The Fundamental Review of the trading book (FRTB) which is compulsory for the efficiency of the calculation of quantities with nested expectations [Bas13]. Therefore, the need of finding new techniques to accelerate these methods becomes an important issue. The Multilevel Monte Carlo method, intensively developed during the last decade, is a new class of approximation methods that fulfill this need and has become a hot topic in numerical probability.

The content of this chapter follows from the articles [Keb05], [KKH08], [BAK15], [BAK14] and [JK17].

In this chapter, we assume that the underlying setting  $(X_t)_{0 \leq t \leq T}$  is solution to a stochastic differential equation

$$dX_t = b(X_t)dt + \sum_{j=1}^q \sigma_j(X_t)dW_t^j, \quad X_0 = x \in \mathbb{R}^d, \quad (1.1)$$

where  $W = (W^1, \dots, W^q)$  is a  $q$ -dimensional Brownian motion on some given filtered probability space  $\mathcal{B} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  and  $(\mathcal{F}_t)_{t \geq 0}$  is the standard Brownian filtration. The functions  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma_j : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $1 \leq j \leq q$ , satisfy usual global Lipschitz condition  $(\mathcal{H}_{b,\sigma}^{\text{GL}})$  ensuring strong existence and uniqueness of the solution of (1.1). In practice, we consider the Euler continuous approximation  $X^n$  of the process  $X$ , with time step  $\delta = T/n$ , given by

$$dX_t^n = b(X_{\eta_n(t)})dt + \sum_{j=1}^q \sigma_j(X_{\eta_n(t)})dW_t, \quad \eta_n(t) = [t/\delta]\delta \quad (1.2)$$

and we approximate  $\mathbb{E}\psi(X_t, 0 \leq t \leq T)$  by  $\mathbb{E}\psi(X_t^n, 0 \leq t \leq T)$  for a given payoff function  $\psi$ . The weak error is firstly studied by Talay and Tubaro in [TT90] and now it is well known that if  $\psi, b$  and  $(\sigma_j)_{1 \leq j \leq q}$  are in  $\mathcal{C}_P^4$ , they are four times differentiable and together with their derivatives at most polynomially growing, then we have (see Theorem 14.5.1 in Kloeden and Platen in [KP95])

$$\varepsilon_n := \mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T) = O(1/n).$$

The same result was extended in Bally and Talay in [BT96a] for a measurable function  $\psi$  but with a non degeneracy condition of Hörmander type on the diffusion. Also, it is worth noticing, the recent results of Kohatsu-Higa, Lejay and Yasuda [KHL17] on the rate of weak convergence of the Euler scheme for stochastic differential equations with non-regular drift.

In the context of possibly degenerate diffusions, if besides  $(\mathcal{H}_{b,\sigma}^{\text{GL}})$  we assume that  $b$  and  $(\sigma_j)_{1 \leq j \leq q}$  are  $\mathcal{C}^1$  and  $\psi$  satisfies condition

$$(\mathcal{H}_\psi) \quad \mathbb{P}(X_T \notin \mathcal{D}_\psi) = 0, \text{ where } \mathcal{D}_\psi := \{x \in \mathbb{R}^d \mid \psi \text{ is differentiable at } x\}$$

then, we proved in [Keb05] that  $\lim_{n \rightarrow \infty} \sqrt{n} \varepsilon_n = 0$ . Conversely, under the same assumptions, we show that the rate of convergence can be  $1/n^\alpha$ , for any  $\alpha \in (1/2, 1]$ . So, it is worth introducing assumption

$$(\mathcal{H}_{\varepsilon_n}) \quad \text{for } \alpha \in [1/2, 1] \quad n^\alpha \varepsilon_n \xrightarrow[n \rightarrow \infty]{} C_\psi(T, \alpha), \quad C_\psi(T, \alpha) \in \mathbb{R}.$$

Furthermore, as originally noticed in [TT90], it is possible to reduce the above error using a second Euler scheme with time step  $2n$ , since clearly the error  $\mathbb{E}[2\psi(X_T^{2n}) - \psi(X_T^n)] - \mathbb{E}\psi(X_T)$  is of order  $1/n^2$ . This idea is known in the literature as the Richardson-Romberg extrapolation. Later on, a multi-step version of the Richardson-Romberg extrapolation was introduced by Pagès [Pag07] with a higher order expansion and developed in the context of path dependent payoffs. Of course, it is also possible to reduce the weak error using other adequate discretization schemes instead of the Euler approximation (see e.g. the Ninomyia Victoir scheme proposed in [NV08]).

Also, we recall that under condition  $(\mathcal{H}_{b,\sigma}^{\text{GL}})$  we have the almost sure convergence of  $X^n$  towards  $X$  together with the following property (see e.g. Bouleau and Lépingle [BL94])

$$(\mathcal{P}) \quad \forall p \geq 1, \quad \sup_{0 \leq t \leq T} |X_t|, \quad \sup_{0 \leq t \leq T} |X_t^n| \in L^p \quad \text{and} \quad \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t - X_t^n|^p \right] \leq \frac{K_p(T)}{n^{p/2}},$$

where  $K_p(T)$  is a positive constant depending only on  $b, \sigma, T, p$  and  $q$ . This  $L^p$ -error or strong error can also be reduced using other approximation schemes like the Milstein approximation (see Section 10.3 in [KP95] and also Milstein's original article [Mil74]).

A standard Monte Carlo method approximates  $\mathbb{E}\psi(X_T)$  by  $\frac{1}{N} \sum_{i=1}^N \psi(X_{T,i}^n)$ . To find the optimal way in choosing the sample size, we prove a central limit theorem letting both  $n$  and  $N$  tend to infinity. More precisely, we prove in [Keb05] that under assumptions  $(\mathcal{H}_\psi)$  and  $(\mathcal{H}_{\varepsilon_n})$  we have

$$n^\alpha \left( \frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} \psi(X_{T,i}^n) - \mathbb{E}\psi(X_T) \right) \Longrightarrow \sqrt{\text{Var}(\psi(X_T))} G + C_\psi(T, \alpha), \text{ as } n \rightarrow \infty, \quad (1.3)$$

with  $G$  a standard normal. Hence, the optimal time complexity needed to achieve a precision of order  $1/n^\alpha$ ,  $\alpha \in [1/2, 1]$  is

$$C_{\text{MC}} \propto n^{2\alpha+1}. \quad (1.4)$$

Here the unit of calculation is one simulation of a random variable.

## 1.2 A two-level Monte Carlo approximation: the Statistical Romberg method

The statistical Romberg method (SR) introduced in [Keb05] is actually a two-level Monte Carlo method which can be presented as a kind of precursor of the popular Multilevel Monte Carlo introduced by [Gil08]. The SR method which is essentially inspired by the Richardson-Romberg detailed above, aims at reducing the statistical error in the Monte Carlo approximation. The main idea behind this approach is to use many sample paths with a coarse time discretization step and few additional sample paths with a fine time discretization step. In more details, we consider two Euler schemes with time steps  $T/n$  and  $T/n^\beta$ ,  $\beta \in (0, 1)$  and approximates  $\mathbb{E} \psi(X_T)$  by

$$V_n := \frac{1}{n^{\gamma_1}} \sum_{i=1}^{n^{\gamma_1}} \psi(\hat{X}_{T,i}^{n^\beta}) + \frac{1}{n^{\gamma_2}} \sum_{i=1}^{n^{\gamma_2}} \psi(X_{T,i}^n) - \psi(X_{T,i}^{n^\beta}),$$

where  $\gamma_1, \gamma_2$  are positive parameters,  $\hat{X}_T^{n^\beta}$  is a second Euler scheme with time step  $T/n^\beta$  and such that the Brownian paths used for  $X_T^n$  and  $X_T^{n^\beta}$  has to be independent of the Brownian paths used to simulate  $\hat{X}_T^{n^\beta}$ . In order to find the optimal parameters of the SR method namely  $\beta, \gamma_1$  and  $\gamma_2$ , it is possible to study the root mean squared error (RMSE) of our estimator and then optimize the parameters for a given precision  $\varepsilon > 0$  taking advantage of our knowledge of the convergence rate of the weak and strong errors. In [Keb05], we choose an other approach based on proving a central limit theorem (CLT) on our estimator. The advantage of this approach is that it gives a confidence interval for the SR estimator which is crucial for Monte Carlo methods and also a precise description of the choice of the optimal parameters as well. This choice of the parameters does not depend on any constant thereby avoiding any pre-computation step in the algorithm. To prove such a CLT on the SR method, we need to know the asymptotic behavior of the distribution of the Euler scheme's error. This convergence result was proven by [JP98] as an improvement on the result given by [KP91]:

$$\sqrt{n}U^n =: \sqrt{n}(X^n - X) \xrightarrow{\text{stably}} U, \text{ as } n \rightarrow \infty, \quad (1.5)$$

with  $U$  a  $d$ -dimensional process satisfying

$$dU_t = \dot{b}(X_t)U_t dt + \sum_{j=1}^q \dot{\sigma}_j(X_t)U_t dW_t^j - \sqrt{\frac{T}{2}} \sum_{j,\ell=1}^q \dot{\sigma}_j(X_t)\sigma_\ell(X_t) d\tilde{W}_t^{\ell j}, \quad (1.6)$$

where  $\tilde{W}$  is a  $q^2$ -dimensional standard Brownian motion, defined on the extension  $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \geq 0}, \tilde{P})$  of the original space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  independent of  $W$ , and  $\dot{b}$  (respectively  $(\dot{\sigma}_j)_{1 \leq j \leq q}$ ) is the Jacobian matrix of  $b$  (respectively  $(\sigma_j)_{1 \leq j \leq q}$ ).



Under assumptions  $(\mathcal{H}_\psi)$  and  $(\mathcal{H}_{\varepsilon_n})$ , we prove that for  $\gamma_1 = 2\alpha$  and  $\gamma_2 = 2\alpha - \beta$ , we have

$$n^\alpha(V_n - \mathbb{E}\psi(X_T)) \Longrightarrow \sigma_2 \tilde{G} + C_\psi(T, \alpha), \text{ as } n \rightarrow \infty, \quad (1.7)$$

with  $\sigma_2^2 = \text{Var}(\psi(X_T)) + \tilde{\text{Var}}(\nabla\psi(X_T) \cdot U_T)$  and  $\tilde{G}$  a standard normal. Thus, the optimal time complexity needed to achieve a precision of order  $1/n^\alpha$  is reached for  $\beta = 1/2$  and is given by

$$C_{\text{SR}} \propto n^{2\alpha+1/2}. \quad (1.8)$$

Thus, in view of (1.4), this clearly means that in this case the SR method outperforms the standard Monte Carlo method.

### 1.3 The Statistical Romberg method for density estimation

The problem of density estimation is a wide topic in applied probability and statistics. A commonly spread approach to deal with such problem is to use a kernel density approach. The aim is to estimate the density of a non-degenerate diffusion process  $X$  that satisfies the Hörmander condition ( see Section 2.3.2 of [Nua95] ) using the Euler scheme  $X^n$  given in (1.2). Under such assumption, the process  $X$  starting at point  $x_0$ , admits a smooth density  $p_T(x_0, x)$  ( see [KS85] ) and in order to simplify the notation, we denote  $p_T(x_0, x) := p(x)$ . Let us recall that in this framework, [BT96b] proved that

$$p(x) = p_n(x) + \frac{C}{n} + o(1/n),$$

where  $p_n(x)$  is a regularized density of the Euler scheme  $X^n$ . For the sake of simplicity, we first introduce the problem for  $d = 1$ . Then, we consider an integrable continuous function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\int \phi(x)dx = 1$  and define the kernels functions  $\phi_{h,x}(y) := \frac{1}{h}\phi\left(\frac{y-x}{h}\right)$ ,  $h > 0$  and  $x \in \mathbb{R}$ . Note that  $\phi_{h,x} \rightarrow \delta_x$  as  $h \rightarrow 0$ , in a weak sense, according to the assumptions on the function  $\phi$ . Thus, the idea is to approximate the density  $p(x) = \mathbb{E}\delta_x(X_T)$  by  $\mathbb{E}\phi_{h,x}(X_T^n)$ , where  $h = n^{-\alpha}$ ,  $\alpha > 0$  and then, use a Monte Carlo method to compute this last quantity. At this stage, two main problems arise when using this approach:

- **Weak error:** the first point concerns the evaluation problem of the weak error

$$\varepsilon_n := \mathbb{E}\phi_{h,x}(X_T^n) - p(x).$$

In their work [KHP02], Kohatsu-Higa and Pettersson proved that  $|\varepsilon_n| \leq C/n$  if  $\alpha \geq 1$  and no expansion of the error is provided.

- **Variance explosion:** the second point concerns the problem of the explosion of the variance of the r.v.  $\phi_{h,x}(X_T^n)$ . To deal with this problem, Kohatsu-Higa and Pettersson propose the use of the classic Malliavin integration by parts formula

$$\mathbb{E}\phi_{h,x}(X_T^n) = \mathbb{E}(\psi_{h,x}(X_T^n)H_n),$$

where  $\psi_{h,x}$  is the primitive function of  $\phi_{h,x}$  and  $H_n$  is the weight given by the Malliavin calculus. Then, they used a localization method in order to reduce the variance of the method. However, the disadvantage of this approach is that the computation time of their algorithm is higher than that of classical Monte Carlo methods with kernel density functions.

In [KKH08], we propose an alternative approach using the kernel estimation method through the use of the SR algorithm as a control method for variance reduction. It turns out that the kernels, as proposed before, in general do not lead to variance reduction. To properly implement the SR method, we consider a subclass of kernel functions known as super kernels of order  $s$ , where  $s > 2(d+1)$ . In more details, we consider a bounded function  $\phi \in C_b^\infty(\mathbb{R}; \mathbb{R})$  with bounded derivatives. The function  $\phi$  is a super kernel with order  $s > 2$  if

$$\int_{\mathbb{R}} \phi(x) dx = 1, \quad \int_{\mathbb{R}} x^i \phi(x) dx = 0 \quad \forall i = 1, \dots, s-1.$$

Moreover, we assume that  $\phi$  satisfies:

$$\int_{\mathbb{R}} |x|^{s+1} |\phi(x)| dx < \infty, \quad \text{and} \quad \int_{\mathbb{R}} |\phi'(x)|^2 dx < \infty, \quad \int_{\mathbb{R}} |\phi(x)|^l dx \text{ for } l = 1, 2, 3.$$

To build super kernels on  $\mathbb{R}^d$ , we take a product of  $d$  unidimensional super kernels. So, for  $\phi_i : \mathbb{R} \rightarrow \mathbb{R}$  with  $i = 1, \dots, d$ , we consider  $\phi(u_1, \dots, u_d) = \phi_1(u_1) \times \dots \times \phi_d(u_d)$  and  $\phi_{h,x}(y) = \frac{1}{h^d} \phi\left(\frac{y-x}{h}\right) = \prod_{i=1}^d \phi_{i,h,x}(y_i)$ . We say that  $\phi$  is a super kernel of order  $s$  if the functions  $\phi_i$ ,  $i = 1, \dots, d$  are unidimensional super kernels of order  $s$ .

To make our asymptotic study on the use of the SR method for kernel density estimation possible, we have to prove an expansion of the weak error  $\varepsilon_n$ . However, in this setting, the Hörmander condition is not enough to guarantee that the Malliavin covariance matrix, associated to the Euler scheme  $X^n$ , is invertible (this would be clearly true under an ellipticity condition). To deal with this problem, we regularize the Euler scheme using  $X_n + Z_{n,\theta}$  instead of  $X_n$ , where  $Z_{n,\theta}$  denotes an independent vanishing random variable defined through the relation  $Z_{n,\theta} := \frac{\bar{W}_T}{n^{\frac{1}{2}+\theta}} \theta \geq 0$ , where  $\bar{W} := (W^{q+1}, \dots, W^{q+d})$  is a  $d$ -dimensional Brownian motion independent of  $W$ . We prove the following result.

**Theorem 1.3.1.** *Let  $h = n^{-\alpha}$ ,  $\alpha \geq 1/s$ . Then, there exists a constant  $C_{\phi,p(x)}^s > 0$  depending on the super kernel  $\phi$ , the density  $p(x)$  and the parameter  $s$  such that*

$$\mathbb{E}\left[\phi_{h,x}(X_T^n + Z_{n,\theta})\right] - p(x) = \frac{C_{\phi,p(x)}^s}{n} + o\left(\frac{1}{n}\right). \quad (1.9)$$

**A kernel density Monte Carlo estimator.** Once such expansion (1.9) is obtained, we are able to have a precise description of the optimal parameters needed for the implementation of a standard Monte Carlo approach. To this aim, we first set  $h = n^{-\alpha}$  the window size of the super kernel with order  $s$  with  $\alpha \geq 1/s$  and consider the estimator

$$S^n := \frac{1}{n^\gamma} \sum_{i=1}^{n^\gamma} \phi_{n^{-\alpha},x}(X_{T,i}^n + Z_{n,\theta}^i), \quad \gamma > 0,$$

where  $(X_{T,i}^n)_{1 \leq i \leq n^\gamma}$  and  $(Z_{n,\theta}^i)_{1 \leq i \leq n^\gamma}$  are independent copies of  $X_T^n$  and  $Z_{n,\theta}$ . If we choose  $\gamma = 2 + \alpha d$ , then we prove that

$$n(S^n - p(x)) \Longrightarrow \sigma G + C_{\phi,p(x)}^s, \quad \text{as } n \rightarrow \infty,$$

with  $\sigma^2 = \phi_2 p(x)$ ,  $\phi_2 = \int_{\mathbb{R}^d} |\phi(u)|^2 du$  and  $G$  a standard Gaussian random variable. Hence, in order to approximate the density  $p(x)$  through a kernel density Monte Carlo

method with a precision of order  $1/n$ , the optimal asymptotic choice of the parameters are  $h = n^\alpha$  and  $N = n^{2+\alpha d}$ , with  $\alpha \geq 1/s$ , where  $s$  denotes the order of the super kernel used for the estimation. This leads to the time complexity  $C_{\text{MC}} \propto n^{3+\alpha d}$ . Therefore, the optimal complexity of this algorithm is given by

$$C_{\text{MC}}^* \propto n^{3+\frac{d}{s}}. \quad (1.10)$$

Consequently, we conclude that the larger the order  $s$  of the super kernel is, the smaller the complexity is. Besides, it is worth noticing that it is always possible to construct super kernels of infinite order using inverse Fourier transform.

**A kernel density SR estimator.** The SR method involves two Euler schemes with time steps  $T/n$  and  $T/n^\beta$ , with  $\beta \in (0, 1)$  and approximates  $p(x)$  by

$$\begin{aligned} V_n := & \frac{1}{n^{\gamma_1}} \sum_{i=1}^{n^{\gamma_1}} \phi_{n^{-\alpha}, x}(\hat{X}_{T,i}^{n^\beta} + \hat{Z}_{n^\beta, \theta}^i) \\ & + \frac{1}{n^{\gamma_2}} \sum_{i=1}^{n^{\gamma_2}} \phi_{n^{-\alpha}, x}(X_{T,i}^n + Z_{n, \theta}^i) - \phi_{n^{-\alpha}, x}(X_{T,i}^{n^\beta} + Z_{n^\beta}^i), \text{ with } \gamma_1, \gamma_2 > 0. \end{aligned} \quad (1.11)$$

To properly analyze the SR method, we need to prove a central limit theorem that provides a precise description of the parameters choice. In order to prove such a result, we extend the limit theorem of [JP98] given in (1.5), by proving a new limit theorem on the asymptotic behavior of the joint distribution of the Euler scheme error and its Malliavin derivative.

**Theorem 1.3.1.** *Let  $(H_t^i)_{0 \leq t \leq T}$  be a continuous sequence of  $\mathbb{R}$ -valued processes (possibly non-adapted). Then  $(\sqrt{n}U_T^n, \sqrt{n} \int_0^T H_s^i D_s^i U_T^n ds) \xrightarrow{\text{stably}} (U_T, \int_0^T H_s^i D_s^i U_T ds)$ , as  $n \rightarrow \infty$ , where  $D^i U_T$  is the Malliavin derivative of  $U$  with respect to  $W^i$  and solution of*

$$\begin{aligned} D_s^i U_t = & \dot{\sigma}_i(X_s) U_s + \int_s^t \dot{b}(X_v) D_s^i U_v dv + \sum_{j=1}^q \int_s^t \dot{\sigma}_j(X_v) D_s^i U_v dW_v^j + \int_s^t D_s^i \dot{b}(X_v) U_v dv \\ & + \sum_{j=1}^q \int_s^t D_s^i \dot{\sigma}_j(X_v) U_v dW_v^j + \sqrt{\frac{T}{2}} \sum_{j, \ell=1}^q \int_s^t D_s^i (\dot{\sigma}_j(X_v) \sigma_\ell(X_v)) d\tilde{W}_v^{\ell j}, \end{aligned}$$

where  $\tilde{W}$  is the  $q \times q$ -dimensional Brownian motion of the limit process of (1.6).

The above limit theorem is the keystone used in proving a central limit theorem on the SR estimator given in (1.11). In order to characterize the limit variance, we need to introduce the following decomposition of the first derivatives of the kernel function  $\phi$

$$\frac{\partial \phi}{\partial x_i} := \phi_i^1(x) - \phi_i^2(x), \text{ where } \phi_i^1(x) := \left( \frac{\partial \phi}{\partial x_i} \right)_+ \text{ and } \phi_i^2(x) := \left( \frac{\partial \phi}{\partial x_i} \right)_-.$$

We also introduce the constant  $C_{ii'}^{jj'} = \int_{\mathbb{R}^d} \phi_i^j(x) \phi_{i'}^{j'}(x) dx, j, j' \in \{1, 2\}, i, i' \in \{1, \dots, d\}$ . We prove in [KKH08] the following central limit theorem.

**Theorem 1.3.2.** *Let*

$$\tilde{\sigma}^2 := \sum_{j,j'=1}^2 \sum_{i,i'=1}^d (-1)^{j+j'} C_{ii'}^{jj'} \left\{ \mathbb{E} \left[ \delta_x(X_T) U_T^i U_T^{i'} \right] + T \delta_{ii'} p(x) \mathbf{1}_{\theta=0} \right\},$$

where  $\delta_x(\cdot)$  stands for the Dirac delta function,  $\delta_{ii'}$  is the Kroeneker delta function and  $\theta$  is the parameter of the regularizing terms in (1.11). Assume that  $\gamma_1 = 2 + \alpha d$ ,  $\gamma_2 = (d+2)\alpha + 2 - \beta$  and  $1/s \leq \alpha \leq \beta/(d+2)$  with  $0 < \beta < 2/3$ . Then,

$$n(V_n - p(x)) \Longrightarrow \tilde{\sigma}G + C_{\phi, p(x)}^s, \text{ as } n \rightarrow \infty,$$

where  $G$  is a standard Gaussian random variable.

Hence, in order to approach the density  $p(x)$  with a precision of order  $1/n$ , the parameters needed to properly implement the SR algorithm are  $h = n^{-\alpha}$ ,  $\gamma_1 = 2 + \alpha d$ ,  $\gamma_2 = n(d+2)\alpha + 2 - \beta$  with  $\beta/(d+2) > \alpha \geq 1/s$ . Therefore, the time complexity needed for this algorithm is  $C_{\text{SR}} \propto n^{\beta+\alpha d+2} + n^{(d+2)\alpha-\beta+3}$ , where  $\beta/(d+2) > \alpha \geq 1/s$ . Hence, for  $\beta = \frac{1}{2} + \alpha$ , the SR method reaches an optimal time complexity given by

$$C_{\text{SR}}^* \propto n^{\frac{5}{2}+(d+1)\alpha},$$

which is clearly better than the optimal complexity of the Monte Carlo method given by (1.10). Therefore, the Romberg control variate method reduces the time complexity by a factor of order  $n^{1/2-\alpha}$ . Furthermore, taking into account that  $\beta/(d+2) > \alpha \geq 1/s$ , we see that if one uses super kernels of order  $s > 2(d+1)$ , we obtain a theoretical asymptotic optimal parameters choice for the method.

## 1.4 Asymptotic analysis of the MLMC method

Introduced by Giles [Gil08], the Multilevel Monte Carlo method (MLMC) extends the statistical Romberg (SR) method of Kebaier [Keb05] from a two-level to  $L$ -level Monte Carlo method with  $L \geq 2$ . It uses a quite similar approach as the Heinrich's multi-level method for parametric integration [Hei01]. For more references, we refer to the web page [https://people.maths.ox.ac.uk/gilesm/mlmc\\_community.html](https://people.maths.ox.ac.uk/gilesm/mlmc_community.html) dedicated to the MLMC research community. In more details, for a diffusion process solution to (1.1), the MLMC method uses information from a sequence of computations with decreasing step sizes and approximates the quantity  $\mathbb{E}\psi(X_T)$  by

$$Q_n = \frac{1}{N_0} \sum_{k=1}^{N_0} \psi(X_{T,k}^1) + \sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} (\psi(X_{T,k}^{\ell, m^\ell}) - \psi(X_{T,k}^{\ell, m^{\ell-1}})), \quad m \in \mathbb{N} \setminus \{0, 1\},$$

where the fine discretization step is equal to  $T/n$ , thereby  $L = \frac{\log n}{\log m}$ . For  $\ell \in \{1, \dots, L\}$ , the processes  $(X_{t,k}^{\ell, m^\ell}, X_{t,k}^{\ell, m^{\ell-1}})_{0 \leq t \leq T}$ ,  $k \in \{1, \dots, N_\ell\}$ , are independent copies of the process  $(X_t^{\ell, m^\ell}, X_t^{\ell, m^{\ell-1}})_{0 \leq t \leq T}$  whose components denote the Euler discretization schemes given by (1.2), with time steps  $m^{-\ell}T$  and  $m^{-(\ell-1)}T$ . However, for fixed  $\ell$ , the simulation of  $(X_t^{\ell, m^\ell})_{0 \leq t \leq T}$  and  $(X_t^{\ell, m^{\ell-1}})_{0 \leq t \leq T}$  has to be based on the same Brownian path like

it is the case for the SR method. Concerning the first empirical mean, the processes  $(X_{t,k}^1)_{0 \leq t \leq T}$ ,  $k \in \{1, \dots, N_0\}$  are independent copies of  $(X_t^1)_{0 \leq t \leq T}$  which denotes the Euler scheme with time step  $T$ . Here, it is important to point out that all these  $L + 1$  Monte Carlo estimators have to be based on different independent samples. Due to the above independence assumption on the paths, the variance of the multilevel estimator is given by

$$\text{Var}(Q_n) := N_0^{-1} \text{Var}(\psi(X_T^1)) + \sum_{\ell=1}^L N_\ell^{-1} \sigma_\ell^2, \text{ where } \sigma_\ell^2 = \text{Var}(\psi(X_T^{\ell, m^\ell}) - \psi(X_T^{\ell, m^{\ell-1}})).$$

According to property  $(\mathcal{P})$  (resp.  $(\mathcal{H}_{\varepsilon_n})$ ), the rate of convergence of the strong error (resp. weak error) of the Euler scheme is  $\beta = 1$  (resp.  $\alpha \in [1/2, 1]$ ). Then, if the function  $\psi$  is for instance Lipschitz continuous, it is easy to check that  $\text{Var}(Q_n) \leq c \sum_{\ell=0}^L N_\ell^{-1} m^{-\ell}$  for some positive non-explicit constant  $c$ . As already mentioned in Section 1.2, it is possible to study the root mean squared error (RMSE) of the above estimator and then optimize the parameters for a given precision  $\varepsilon > 0$  taking advantage of our knowledge of the convergence rates of the weak and strong errors. This is precisely the approach adopted in [Gil08] whereby for a desired RMSE that should be of the same order as the bias, say of order  $1/n^\alpha$ , the optimal choice is given by

$$N_\ell = 2cn^{2\alpha} \left( \frac{\log n}{\log m} + 1 \right) \frac{T}{m^\ell} \quad \text{for } \ell \in \{0, \dots, L\} \text{ and } L = \frac{\log n}{\log m}. \quad (1.12)$$

Hence, for a precision  $\varepsilon_n = 1/n^\alpha$ , this optimal choice leads to a time complexity for the MLMC method proportional to  $\varepsilon_n^{-2} (\log \varepsilon_n)^2$ , which is clearly better than the complexities performed by the SR and MC methods (see (1.8) and (1.4)). Interesting numerical tests, comparing the performance of the three methods in several specific examples were processed in Korn, Korn and Kroisandt [KKK10].

Before presenting our main contributions, let us first mention the main advances around the MLMC methods topic. Recently, Lemaire and Pagès [LP17] introduced a Multilevel Richardson-Romberg (MLRR) estimator combining the Multistep Richardson-Romberg method introduced in [Pag07] and the MLMC method. In more details, when the strong error of the discretization scheme is of order  $\beta = 1$  and the weak error  $\mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T)$  can be expanded at any order in  $n$ , then the global time complexity of the MLRR method is proportional to  $\varepsilon_n^{-2} \log(1/\varepsilon_n)$  instead of  $\varepsilon_n^{-2} (\log(1/\varepsilon_n))^2$  with the standard MLMC method. When the strong error of the discretization scheme is of order  $n^{-\frac{\beta}{2}}$ ,  $\beta < 1$ , the gain of MLRR over MLMC becomes even more pronounced. If other discretization schemes with higher order of strong convergence rate  $\beta > 1$  (as the Milstein scheme for instance [KP95]) are used instead of the Euler scheme, then for the same precision  $\varepsilon_n$ , the optimal time complexity of the MLMC method reaches the coveted minimal order  $\varepsilon_n^{-2}$ , so that it behaves exactly like an unbiased MC estimator. However, the practical implementation of these schemes with higher order of strong convergence rate is not always possible. This is the case of the Milstein's scheme whose implementation requires the simulation of iterated Itô integrals better known as Lévy areas for which there is no known efficient method especially in a multidimensional setting [GL94, RW01, Wik01]. To overcome this difficulty, Giles and Szpruch [GS14] introduced a new MLMC estimator with the Milstein scheme but with cancelling the Lévy areas.

Then an antithetic multilevel correction is embedded in the new estimator to take into account of the neglected Lévy areas. Thus, under additional regularity assumptions on the payoff function and the diffusion process coefficients, Giles and Szpruch [GS14] show that the antithetic multilevel Monte Carlo estimator reaches the optimal time complexity  $\varepsilon_n^{-2}$ . Also, Debrabant and Rössler [DR15] suggest to use a scheme with high order of weak convergence on the finest grid at the finest level  $L$  of the MLMC method. This approach reduces the constant in the computational complexity by decreasing the number of discretization levels. Inspired by [DR15], Al Gerbi, Jourdain and Clément [AGJC16] improved the MLMC estimator of Giles and Szpruch [GS14] suggesting to combine the Ninomiya-Victoir scheme, known for having a weak convergence rate of order  $\alpha = 2$ , on the finest grid at the finest level  $L$  with Giles-Szpruch antithetic estimator. A full study with new results comparing the performance of all these methods through advanced and challenging practical examples can be found in [Gio17].

Most of the above mentioned works used the Giles's approach based on finding the optimal parameters that minimize the computational complexity for a given precision. In [BAK15], we choose an alternative approach proving a central limit theorem on the MLMC method with an Euler discretization scheme. We get the following result.

**Theorem 1.4.1.** *Assume that  $b$  and  $\sigma$  are  $\mathcal{C}^1$  functions satisfying the global Lipschitz condition  $(\mathcal{H}_{b,\sigma}^{\text{GL}})$ . Let the payoff function  $\psi$  be a real-valued function satisfying condition  $(\mathcal{H}_\psi)$  and*

$$|\psi(x) - \psi(y)| \leq C(1 + |x|^p + |y|^p)|x - y| \quad \text{for some } C, p > 0. \quad (1.13)$$

Assume that condition  $(\mathcal{H}_{\varepsilon_n})$  is satisfied for some  $\alpha \in [1/2, 1]$  and constant  $C_\psi(T, \alpha)$ . Moreover, let us consider a real sequence  $(a_\ell)_{\ell \geq 1}$  of positive weights satisfying condition

$$(\mathcal{W}) \quad \lim_{n \rightarrow \infty} \sum_{\ell=1}^L a_\ell = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{\left(\sum_{\ell=1}^L a_\ell\right)^{p/2}} \sum_{\ell=1}^L a_\ell^{p/2} = 0, \quad \text{for } p > 2.$$

For the choice of  $N_\ell$ ,  $\ell \in \{0, 1, \dots, L\}$  given by

$$N_\ell = \frac{n^{2\alpha}(m-1)}{m^\ell a_\ell} \sum_{\ell=1}^L a_\ell, \quad \ell \in \{0, \dots, L\} \quad \text{and} \quad L = \frac{\log n}{\log m}, \quad (1.14)$$

we have  $n^\alpha(Q_n - \mathbb{E}(\psi(X_T))) \Longrightarrow \mathcal{N}(C_\psi(T, \alpha), \sigma^2)$ , as  $n \rightarrow \infty$ , with  $\sigma^2 = \tilde{\text{Var}}(\nabla\psi(X_T) \cdot U_T)$  and  $\mathcal{N}(C_\psi(T, \alpha), \sigma^2)$  denotes a normal distribution and  $U$  is the limit process solution to (1.6).

In order to be able to show this result, we first prove a stable law convergence theorem, for the Euler scheme error between two consecutive levels  $m^{\ell-1}$  and  $m^\ell$ , of the type obtained in Jacod and Protter [JP98]. Indeed, we prove the following functional limit theorem

$$\sqrt{\frac{mn}{(m-1)}}(X^{mn} - X^n) \xrightarrow{\text{stably}} U, \quad \text{as } n \rightarrow \infty. \quad (1.15)$$

In fact, their result, namely

$$\sqrt{n}(X^n - X) \xrightarrow{\text{stably}} U, \quad \text{as } n \rightarrow \infty, \quad (1.16)$$

is not sufficient to prove our Theorem 1.4.1, since the MLMC method involves the error process  $X^{mn} - X^n$  rather than  $X^n - X$ . However, we can recover (1.16) if we informally allow  $m = \infty$  in (1.15)

This generic form (1.14) for  $N_\ell$  allows us a straightforward use of the Toeplitz lemma, which is a crucial tool used in the proof of our central limit theorem. From a complexity analysis point of view, we can interpret Theorem 1.4.1 as follows. For a precision of order  $1/n^\alpha$ , the computational effort necessary to properly implement the MLMC method is given by the sequence of sample sizes specified by relation (1.14). The associated time complexity is given by  $C_{\text{MLMC}} \propto \frac{n^{2\alpha(m-1)T}}{a_0} \sum_{\ell=1}^L a_\ell + n^{2\alpha \frac{(m^2-1)T}{m}} \sum_{\ell=1}^L \frac{1}{a_\ell} \sum_{\ell=1}^L a_\ell$ . The minimum of the second term of this complexity is reached for the choice of weights  $a_\ell^* = 1$ ,  $\ell \in \{1, \dots, L\}$  and the optimal complexity is given by

$$C_{\text{MLMC}}^* \propto n^{2\alpha} (\log n)^2.$$

It turns out that for a given precision  $\varepsilon_n = 1/n^\alpha$ , the complexity is given by  $C_{\text{MLMC}} = O(\varepsilon_n^{-2} (\log \varepsilon_n)^2)$ . Note that this optimal choice  $a_\ell^* = 1$ ,  $\ell \in \{1, \dots, L\}$ , with taking  $a_0 = 1$  corresponds to the sample sizes given by

$$N_\ell = \frac{(m-1)}{m^\ell \log m} n^{2\alpha} \log n, \quad \ell \in \{0, \dots, L\}.$$

Hence, our optimal choice is consistent with that proposed by Giles [Gil08]. Nevertheless, unlike the parameters obtained in [Gil08] for the same setting (see relation (1.12)), our optimal choice of the sample sizes  $N_\ell, \ell \in \{1, \dots, L\}$  does not depend on any given constant, since our approach is based on proving a central limit theorem and not on getting upper bounds for the variance. It is worth noticing that an other advantage of the CLT approach is to construct a more accurate confidence interval. In fact, for a given root mean square error RMSE, the radius of the 90%-confidence interval constructed by the CLT is  $1.64 \times \text{RMSE}$ . However, without this latter result one can only use Chebyshev's inequality which yields a radius equal to  $3.16 \times \text{RMSE}$ . Furthermore, we compare the performance of the numerically optimized MLMC algorithm of Giles available on <https://people.maths.ox.ac.uk/gilesm/mlmc/> and the MLMC method equipped with our optimal parameters. Here, we emphasize that our algorithm is easier to implement, since it does not require any pre-computation step. The Figure 1.1 shows that both methods are perfectly online.

Then, in [BAK14], we investigate the application of this method to the pricing of Asian option call

$$e^{-rT} \mathbb{E} (I_T - K)_+, \quad \text{where } I_T := \frac{1}{T} \int_0^T S_u du,$$

under the Black-Scholes model  $S_t = S_0 \exp((r - \frac{\sigma^2}{2})t + \sigma W_t)$ , where  $S_0 > 0$ . We use the Riemann scheme to approximate the integral in the payoff process  $I_T^n = \frac{1}{n} \sum_{k=0}^{n-1} S_{k\delta}$ . It was shown in Lapeyre and Temam [LT01] that this approximation is of second order and the associated weak error  $\varepsilon_n$  is of order  $n^{-1}$ . At first, we study the asymptotic behavior of the error distribution and prove that, as  $n \rightarrow \infty$ , we have

$$\frac{mn}{\sqrt{m^2-1}} (I^{mn} - I^n) \xrightarrow{\text{stably}} \xi, \quad \text{where } \xi_t := \sqrt{\frac{m-1}{m+1}} \frac{S_t - S_0}{2} + \frac{1}{2\sqrt{3}} \int_0^t \sigma S_u dB_u,$$

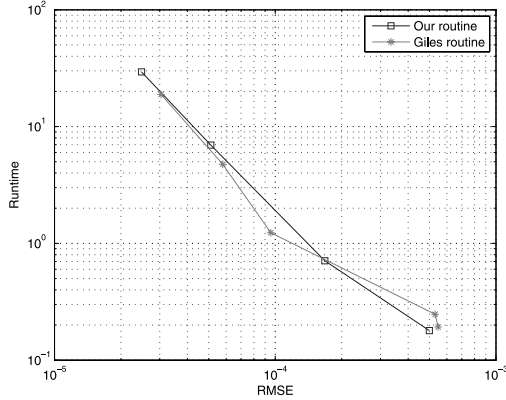


Figure 1.1: Comparison of both routines.

with  $B$  a standard Brownian motion on an extension  $\hat{\mathcal{B}}$  of  $\mathcal{B}$ , which is independent of  $W$ . Taking advantage of this study, we establish thanks to a new choice of the sample sizes  $(N_\ell)_{0 \leq \ell \leq L}$  a Lindeberg Feller central limit theorem type for the MLMC algorithm:

$$Q_n^I = f(s_0) + \sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \left( f(I_{T,k}^{\ell, m^\ell}) - f(I_{T,k}^{\ell, m^{\ell-1}}) \right), \text{ where } f(x) = e^{-rT}(x - K)_+.$$

Actually, we choose

$$N_\ell = \frac{n^2(m^2 - 1)}{m^{2\ell} a_\ell} \sum_{\ell=1}^L a_\ell, \quad \ell \in \{0, \dots, L\} \quad \text{and} \quad L = \frac{\log n}{\log m}, \quad (1.17)$$

with positive weights  $(a_\ell)_{\ell \geq 1}$  satisfying condition  $(\mathcal{W})$  and we prove that for an explicit constant  $C_f^I$ , we have

$$n(Q_n^I - \mathbb{E}(f(I_T))) \Rightarrow \mathcal{N}\left(C_f^I, \tilde{\text{Var}}(f'(I_T)\xi_T)\right).$$

We proved a similar result for the trapezoidal scheme. In other words, to achieve a precision of order  $1/n$ , when pricing Asian option calls, the MLMC method needs the sample sizes  $(N_\ell)_{0 \leq \ell \leq L}$  given by relation (1.17). Consequently, the computational complexity corresponds to  $C_{\text{MLMC}} \propto \frac{(m+1)^2(m-1)}{m} n^2 \sum_{\ell=1}^L \frac{1}{m^\ell a_\ell} \sum_{\ell=1}^L a_\ell$ . The minimum of this complexity is reached for the choice of weights  $a_\ell^* = m^{-\ell/2}$  which yields

$$N_\ell^* = \frac{m^2 - 1}{m^{3\ell/2}(\sqrt{m} - 1)} n^2 \left(1 - \frac{1}{\sqrt{n}}\right) \quad \text{and} \quad C_{\text{MLMC}}^{a_\ell^*} \propto n^2.$$

Even if this choice is consistent with the complexity analysis of Theorem 3.1 in [Gil08], when taking  $\beta = 2$ , we draw attention that it does not satisfy our technical condition  $(\mathcal{W})$  which is essential to check the classical Lyapunov type condition. However, numerical tests seem to confirm the validity of the central limit theorem for this choice (see Figure 1.4). Nevertheless, we provide a sequence  $(a_\ell)_{1 \leq \ell \leq L}$  satisfying assumption  $(\mathcal{W})$  and



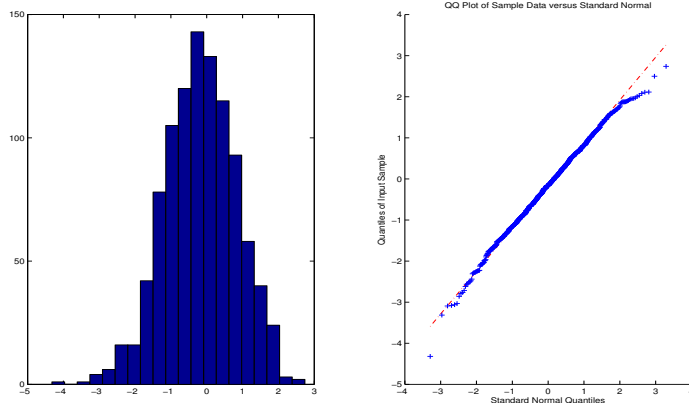


Figure 1.2: Numerical tests for the optimal choice  $a_\ell^* = m^{-\ell/2}$ .

letting the computational complexity quite close to the optimal  $n^2$  effort. Indeed, for  $\bar{a}_\ell = 1/(\ell \log \ell)$ , we get  $\bar{N}_\ell = \frac{(m^2 - 1)\ell \log \ell}{m^{2\ell}} n^2 \sum_{\ell=1}^L \frac{1}{\ell \log \ell}$  and  $C_{\text{MLMC}}^{\bar{a}_\ell} \propto n^2 \log \log \log n$ . It is worth noticing that the above complexity analysis remains valid for any discretization scheme having a strong convergence rate  $\beta = 2$ .

## 1.5 Non-asymptotic analysis of the MLMC method

In [JK17], we are interested in deriving non asymptotic error estimations for the MLMC estimator

$$\hat{Q} = \frac{1}{N_0} \sum_{k=0}^{N_0} f(X_{T,k}^1) + \sum_{\ell=0}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \left( f(X_{T,k}^{m^\ell}) - f(X_{T,k}^{m^{\ell-1}}) \right). \quad (1.18)$$

For  $\ell \in \{0, \dots, L\}$ , the processes  $((X_{t,k}^{m^\ell})_{0 \leq t \leq T})_k$  denote independent copies of the Euler discretization scheme with time step  $m^{-\ell}T$  of the  $\mathbb{R}^d$ -valued solution to the stochastic differential equation with additive noise

$$dX_t = b(X_t)dt + dW_t, \quad X_0 = x_0 \in \mathbb{R}^d, \quad (1.19)$$

driven by the  $d$ -dimensional Brownian motion  $W = (W^1, \dots, W^d)$  and with Lipschitz drift function  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . When  $d = 1$ , this additive noise setting is not restrictive. Indeed any stochastic differential equation  $dY_t = \sigma(Y_t)dW_t + \eta(Y_t)dt$  with multiplicative noise given by some function  $\sigma : \mathbb{R} \rightarrow \mathbb{R}_+^*$  such that  $\frac{1}{\sigma}$  is locally integrable can be reduced to (1.19) by the Lamperti transformation: for  $\varphi(y) = \int_{y_0}^y \frac{dz}{\sigma(z)}$ ,  $X_t = \varphi(Y_t)$  solves (1.19) with  $b(x) = \left( \frac{\eta}{\sigma} - \frac{\sigma'}{2} \right) (\varphi^{-1}(x))$ . Let us recall that, since the diffusion coefficient is constant, the Euler scheme coincides with the Milstein scheme and if  $b$  belongs to  $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d)$  with bounded derivatives, then the strong error estimation improves to  $\mathbb{E}[\sup_{0 \leq t \leq T} |X_t - X_t^n|^p] \leq \frac{K_p(T)}{n^p}$ , with  $K_p(T) < \infty$  (see for instance Milstein [Mil95]).

We prove Gaussian type concentration inequalities for  $\hat{Q} - \mathbb{E}[f(X_T)]$ . Recently, Frikha and Menozzi [FM12] derived concentration inequalities for  $f(X_T^n) - \mathbb{E}[f(X_T^n)]$

which appears in the classical Monte Carlo method. However, estimating the moment generating function of the differences  $f(X_T^{mn}) - f(X_T^n) - \mathbb{E}[f(X_T^{mn}) - f(X_T^n)]$  which, for  $n \in \{1, m, \dots, m^{L-1}\}$ , appear in  $\hat{Q} - \mathbb{E}[f(X_T)]$  is a much more delicate task and adapting their approach seems to be problematic. Nevertheless, the boundedness of the Malliavin derivatives  $DX_T^n$  and  $DX_T^{mn}$  in the additive noise setting permits to follow the approach of Houdré and Privault [HP02] based on the Clark-Ocone formula and this is one reason why we focus on their approach. Another reason is that for stochastic differential equations with multiplicative noise, more sophisticated schemes, like the Milstein scheme in the commutative case or the Giles and Szpruch [GS14] scheme in the general case, are necessary to improve to two the order one of convergence of the variance of  $(f(X_T^{m^\ell}) - f(X_T^{m^{\ell-1}}))^2$  and recover the unbiased Monte Carlo complexity.

### 1.5.1 MLMC parameters optimization revisited

In light of this non-asymptotical analysis, we first refresh the mean square error analysis carried out by Giles [Gil08] for discretization schemes with strong convergence rate  $\beta > 1$ . We prove under

#### Assumption (R1)

The function  $b \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d)$  and there exist finite constants  $[\dot{b}], a_{\Delta b} > 0$  such that

$$\begin{aligned} \forall x \in \mathbb{R}^d, \quad \|\nabla b(x)\| &\leq [\dot{b}]_\infty, \\ \forall x \in \mathbb{R}^d, \quad |\Delta b(x)| &\leq 2a_{\Delta b}(1 + |x - x_0|), \end{aligned} \quad (1.20)$$

that for some explicit positive constants  $K_{1,\infty}$  and  $K_{1,m}$  we have

$$\mathbb{E} \left[ \left( \hat{Q} - \mathbb{E}[f(X_T)] \right)^2 \right] \leq [\dot{f}]_\infty^2 \left( \frac{K_{1,\infty} T^2}{m^{2L}} + \frac{T}{N_0} + \sum_{\ell=1}^L \frac{K_{1,m}(m-1)T^2}{N_\ell m^{2\ell-1}} \right).$$

Therefore, to achieve a root mean square error  $\varepsilon > 0$ , one should choose  $[\dot{f}]_\infty T \sqrt{K_{1,\infty}}/m^L < \varepsilon$ . For such a choice, one should then choose  $(N_\ell)_{0 \leq \ell \leq L}$  such that

$$\sum_{\ell=1}^L \frac{m+1}{N_\ell m^{2\ell+1}} + \frac{C^2}{N_0} \leq \frac{m+1}{K_{1,m} m^2 (m-1)} \left( \frac{\varepsilon^2}{[\dot{f}]_\infty^2 T^2} - \frac{K_{1,\infty}}{m^{2L}} \right), \quad (1.21)$$

where  $C = \frac{1}{m} \sqrt{\frac{m+1}{K_{1,m}(m-1)}}$  minimizing the computation cost which is equal to  $N_0 + \sum_{\ell=1}^L N_\ell (m+1) m^{\ell-1}$ . This constrained minimization problem leads to  $N_0 = N \frac{C}{C + \sum_{\ell=1}^L m^{-3\ell/2}}$  and  $N_\ell = N \frac{m^{-3\ell/2}}{C + \sum_{\ell=1}^L m^{-3\ell/2}}$ , where the total number  $N$  of simulations is chosen in order to achieve equality in (1.21) :

$$N = \left( C + \sum_{\ell=1}^L m^{-3\ell/2} \right) \left( C + \frac{m+1}{m} \times \frac{1 - m^{-L/2}}{\sqrt{m} - 1} \right) \frac{K_{1,m} m^2 (m-1) [\dot{f}]_\infty^2 T^2}{(m+1)(\varepsilon^2 - [\dot{f}]_\infty^2 T^2 K_{1,\infty} m^{-2L})}.$$

Then, the computation cost is given by  $\text{Cost}(m, m^{-L})$ , where

$$\text{Cost}(m, x) = \left( C + \frac{m+1}{m} \times \frac{1 - \sqrt{x}}{\sqrt{m} - 1} \right)^2 \frac{K_{1,m} m^2 (m-1) [\dot{f}]_\infty^2 T^2}{(m+1)(\varepsilon^2 - [\dot{f}]_\infty^2 T^2 K_{1,\infty} x^2)}.$$

After some optimization steps, we end up with optimal  $L^\varepsilon$ ,  $N^\varepsilon$  and  $N_\ell^\varepsilon$  solving this problem. It turns out that as expected from [Gil08],  $\text{Cost}(m, m^{-L^\varepsilon}) = \mathcal{O}(\varepsilon^{-2})$  as  $\varepsilon \rightarrow 0$  for fixed  $m$  and  $N^\varepsilon = \mathcal{O}(\varepsilon^{-2})$ . However, the bias term  $[f]_\infty T \sqrt{K_{1,\infty}} m^{-L^\varepsilon}$  behaves as  $\mathcal{O}(\varepsilon^{4/3})$ . To the best of our knowledge, such a 4/3 order of convergence of the bias does not appear in the existing multilevel Monte Carlo methods literature.

### 1.5.2 Concentration bounds

Let us first introduce the following condition.

#### Assumption (R2)

The function  $b \in \mathcal{C}^3(\mathbb{R}^d, \mathbb{R}^d)$  and satisfies assumption (R1). Moreover, there exist finite constants  $[\ddot{b}]_\infty, a_{\nabla\Delta b} > 0$  such that

$$\begin{aligned} \forall j \in \{1, \dots, d\}, \forall x \in \mathbb{R}^d, \left\| \frac{\partial \nabla b}{\partial x_j}(x) \right\| &\leq [\ddot{b}]_\infty \\ \forall x \in \mathbb{R}^d, \|\nabla \Delta b(x)\| &\leq 2a_{\nabla\Delta b}(1 + |x - x_0|). \end{aligned} \quad (1.22)$$

Thanks to a clever decomposition of the error difference between the crude Euler scheme with  $n$  steps and the finer one with  $mn$  steps, we provide estimates of the moment generating functions of this squared difference and its Malliavin derivative. These two results lead us to prove the following concentration inequality on the MLMC algorithm.

**Theorem 1.5.1.** *Let assumption (R2) hold and  $f \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$  be a Lipschitz continuous function with constant  $[f]_\infty$  and such that  $\nabla f$  is also Lipschitz with constant  $[f]_{lip}$ . Then, there exist explicit positive constants  $\mathcal{C}_1$  and  $\mathcal{C}_2$  such that for all  $0 \leq \alpha \leq 2[f]_\infty^2 \left( \frac{T}{2N_0} + \sum_{\ell=1}^L \frac{C_1(m-1)T^2}{N_\ell m^{2\ell-1}} \right) \mathcal{C}_2 \min_{1 \leq \ell \leq L} m^\ell N_\ell$ , the MLMC estimator (1.18) satisfies*

$$\mathbb{P} \left( |\hat{Q} - \mathbb{E}f(X_T)| \geq \alpha \right) \leq 2 \exp \left( - \frac{\left( \alpha - \frac{[f]_\infty T \sqrt{K_{1,\infty}}}{m^L} \right)^2}{2[f]_\infty^2 \left( \frac{T}{N_0} + \sum_{\ell=1}^L \frac{2C_1(m-1)T^2}{N_\ell m^{2\ell-1}} \right)} \right).$$

Notice that the factor  $[f]_\infty^2 \left( \frac{T}{N_0} + \sum_{\ell=1}^L \frac{2C_1(m-1)T^2}{N_\ell m^{2\ell-1}} \right)$  in the denominator is closely related to the non-asymptotic upper-bound  $[f]_\infty^2 \left( \frac{T}{N_0} + \sum_{\ell=1}^L \frac{K_{1,m}(m-1)T^2}{N_\ell m^{2\ell-1}} \right)$  of the variance of  $\hat{Q}$  derived in Section 1.5.1. Following the discussion and notations of this latter subsection, for  $\varepsilon$  small enough, we consider  $\hat{Q}_\varepsilon$  the MLMC estimator (1.18) with the optimal parameters  $L^\varepsilon$ ,  $N^\varepsilon$  and  $N_\ell^\varepsilon$  leading to the a bias  $\frac{[f]_\infty T \sqrt{K_{1,\infty}}}{m^{L^\varepsilon}} = \mathcal{O}(\varepsilon^{4/3})$  and optimal  $\text{Cost}(m, m^{-L^\varepsilon}) = \mathcal{O}(\varepsilon^{-2})$  as  $\varepsilon \rightarrow 0$ . We rewrite the above concentration inequality as follows: there exist explicit positive constants  $c_1, c_2$  and  $c_3$  such that for  $\varepsilon$  small enough and  $0 \leq \alpha \leq c_1 \varepsilon^{2/3}$ , we have

$$\mathbb{P} \left( |\hat{Q}_\varepsilon - \mathbb{E}f(X_T)| \geq \alpha \right) \leq 2 \exp \left( - \frac{(\alpha - c_2 \varepsilon^{4/3})^2}{c_3 \varepsilon^2} \right) \leq 2 \exp \left( \frac{2c_1 c_2}{c_3} \right) \exp \left( - \frac{\alpha^2}{c_3 \varepsilon^2} \right).$$

Then, we provide an alternative choice of the parameters that does not increase neither the root mean square error  $\varepsilon$  nor the order in  $\varepsilon$  of the computational cost of the MLMC estimator and for which the upper bound on the deviation parameter  $\alpha$  is larger than  $c_1\varepsilon^{2/3}$ . Namely, for  $\beta > 1$ , we set  $N_\ell^{\varepsilon,\beta} = N_\ell^\varepsilon \times \frac{m^{\frac{\ell-1}{2}}}{m^{\frac{\ell-1}{2}} \wedge \ell^\beta}$  for  $\ell \in \{1, \dots, L^\varepsilon\}$  and

$$\hat{Q}_{\varepsilon,\beta} = \frac{1}{N_0^\varepsilon} \sum_{k=0}^{N_0^\varepsilon} f(X_{T,k}^1) + \sum_{\ell=1}^{L^\varepsilon} \frac{1}{N_\ell^{\varepsilon,\beta}} \sum_{k=1}^{N_\ell^{\varepsilon,\beta}} \left( f(X_{T,k}^{m^\ell}) - f(X_{T,k}^{m^{\ell-1}}) \right).$$

Since for each  $\ell$ ,  $N_\ell^{\varepsilon,\beta} \geq N_\ell^\varepsilon$ , the root mean square error of  $\hat{Q}_{\varepsilon,\beta}$  is not greater than the one of  $\hat{Q}_\varepsilon$  and therefore than  $\varepsilon$ . Moreover, as  $\varepsilon \rightarrow 0$ , the computational cost of  $\hat{Q}_{\varepsilon,\beta}$  is  $\mathcal{O}(\varepsilon^{-2})$ . Further, for  $\varepsilon$  small enough  $\min_{1 \leq \ell \leq L^\varepsilon} m^\ell N_\ell^{\varepsilon,\beta} = \frac{N^\varepsilon m^{-1/2}}{(L^\varepsilon)^\beta (C + \sum_{\ell=1}^{L^\varepsilon} m^{-3\ell/2})} = \mathcal{O}(\varepsilon^{-2} \ln^\beta(1/\varepsilon))$ . Hence, for this new choice we get the following concentration inequality : for  $\beta > 1$ , there exist explicit positive constants  $c_4, c_5$  and  $c_6$  such that for  $\varepsilon$  small enough and  $0 \leq \alpha \leq c_4 \ln^\beta(1/\varepsilon)$ , we have

$$\mathbb{P} \left( |\hat{Q}_{\varepsilon,\beta} - \mathbb{E}f(X_T)| \geq \alpha \right) \leq 2 \exp \left( - \frac{(\alpha - c_5 \varepsilon^{4/3})^2}{c_6 \varepsilon^2} \right).$$

## 1.6 Perspectives

**A Multilevel Monte Carlo algorithm for jump diffusions.** The aim of this current project is to study the use of MLMC methods with stochastic differential equations driven by a Lévy process instead of a Brownian motion. These processes are an excellent tool for modeling the underlying price in mathematical finance, see for instance Cont and Tankov [CT04]. When the Lévy process has a Brownian component, the rate of convergence of the MLMC method was already studied by Dereich and Li [DL16]. Their result does not cover the case of a stochastic differential equation driven by a pure-jump process. This topic is a part of the PhD project of Trâm Ngô that we currently co-supervise with Mohamed Ben Alaya.

**Optimized Multilevel Monte Carlo methods for jump diffusion processes.** With Arturo Kohatsu-Higa, we plan to provide an alternative scheme to the antithetic Multilevel Monte Carlo method introduced by Giles and Szpruch in [GS14]. Our approach is based on stochastic flow techniques. The advantage of this new approach is that it can be carried out for jump diffusion processes allowing the MLMC method in this latter case to behave exactly like an unbiased MC estimator.

**Concentration inequalities for MLMC methods for SDEs with multiplicative noise.** A natural question is to extend the concentration bounds obtained for the MLMC in the setting of additive noise SDEs (see Section 1.5) to the multidimensional multiplicative noise setting. This question is quite challenging since our technical approach based on the Clark-Ocone formula cannot be used due to the unboundedness of the Malliavin derivative of the Euler discretization scheme. Also, using the approach of Frikha and Menozzi [FM12] based on tensorization type arguments seems to be very

problematic even in the additive noise case. With Benjamin Jourdain, we then hope to develop new techniques allowing us to study the moment-Laplace generating function of  $f(X_T^{mn}) - f(X_T^n) - \mathbb{E}[f(X_T^{mn}) - f(X_T^n)]$  in the general multiplicative noise setting.

## Chapter 2

# Coupling Importance Sampling and Multilevel Monte Carlo

### 2.1 Introduction

In the implementation of Monte Carlo methods, one of the main concerns of practitioners is to reduce the associated variance to get more accurate estimations. Among the known variance reduction techniques we focus on the popular importance sampling method. In more details, if one has to use a Monte Carlo method to approximate  $\mathbb{E}\psi(X_T)$ , where  $(X_t)_{0 \leq t \leq T}$  is solution to (1.1), as usually the case, a Girsanov transform is applied to produce a shift parameter  $\theta$ , so that

$$\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T^\theta, W_T), \quad \text{where } g(\theta, X_T^\theta, W_T) = \psi(X_T^\theta)e^{-\theta \cdot W_T - \frac{1}{2}|\theta|^2 T}. \quad (2.1)$$

Here,  $g : \mathbb{R}^q \times \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}$ , and  $(X_t^\theta)_{0 \leq t \leq T}$  is solution to

$$dX_t^\theta = \left( b(X_t^\theta) + \sum_{j=1}^q \theta_j \sigma_j(X_t^\theta) \right) dt + \sum_{j=1}^q \sigma_j(X_t^\theta) dW_t^j. \quad (2.2)$$

Note that the case  $\theta = 0$  corresponds to the stochastic differential equation satisfied by  $(X_t)_{0 \leq t \leq T}$ . Then, it seems natural to implement a Monte Carlo procedure using the optimal  $\theta^*$  solution to the problem

$$\theta^* = \arg \min_{\theta \in \mathbb{R}^q} \mathbb{E}g^2(\theta, X_T, W_T),$$

since the quantity  $\mathbb{E}g^2(\theta, X_T, W_T)$  denotes the main term of the limit variance in the central limit theorem associated to the Monte Carlo method. But what about the effective computation of  $\theta^*$ ? Two strategies are possible:

- ▲ The first one consists on using stochastic algorithms of Robbins-Monro type to construct recursively a sequence of random variables  $(\theta_i)_{i \in \mathbb{N}}$  that approximates accurately  $\theta^*$ . Nevertheless, making this routine converge requires a quite restrictive non explosion condition (NEC) (see e.g. [BMP90, Duf96, KY03]), assuming that  $\mathbb{E}[\nabla_\theta g^2(\theta, X_T, W_T)]$  has a sub-quadratic growth in  $1 + |\theta|^2$ , which is clearly not satisfied in our setting. To avoid this restrictive condition, two improved versions of this routine are proposed in the literature:

- The first one is based on a truncation procedure introduced by Chen et al. [CZ86, CLG88] and investigated later by several authors (see, e.g. Andrieu, Moulines and Priouret in [AMP05] and Lelong in [Lel08]). The use of this technique in the context of importance sampling is initially proposed by Arouna in [Aro04] and investigated afterward by Lapeyre and Lelong in [LL11].
- The second alternative is more recent and introduced by Lemaire and Pagès in [LP10]. Indeed, they introduce an unconstrained procedure by using extensively the regularity of the involved density and they prove the convergence of this algorithm. In what follows, these two methods will be called respectively constrained and unconstrained algorithms.

In a different setting, namely that of pure-jump Lévy processes, Kawai [Kaw09] uses a constrained version of the Robbins-Monro algorithm to approximate the optimal shift  $\theta$  reducing the variance of the Monte Carlo method.

- ▲ The second strategy is proposed by Jourdain and Lelong [JL09], in a Brownian diffusion setting and consists on using deterministic optimization techniques for the empirical Monte Carlo estimator to approximate the optimal shift parameter.

The content of this chapter follows from :

- the results contained in [BAHK15] and [BHK16], in which we extend the use of the constrained and unconstrained versions of the Robbins-Monro algorithms for the Multilevel Monte Carlo (MLMC) methods in the setting of Brownian diffusions,
- the results of [BAHK16], in which we introduce a novel concept allowing the use of MLMC methods in the setting of pure-jump Lévy processes where no discretization schemes are involved. In this context, we also provide adaptive MLMC algorithms integrating efficient importance sampling procedure,
- the results of [KL17] that extend the works of [JL09] to the MLMC setting.

## 2.2 Importance Sampling and Statistical Romberg method: a stochastic algorithm approach for Brownian diffusions

In [BAHK15], we consider the problem of approximating  $\mathbb{E}g(\theta, X_T^\theta, W_T)$  introduced in (2.1) using an adaptive version of the Statistical Romberg (SR) method, which automatically computes the optimal parameter shift reducing the corresponding variance. Regarding the latter, in view of (1.7) the variance optimizer is defined by

$$\theta^* = \operatorname{argmin}_{\theta \in \mathbb{R}^q} v(\theta) \quad \text{with} \quad v(\theta) := \tilde{\mathbb{E}} \left( \left[ \psi(X_T^\theta)^2 + (\nabla \psi(X_T^\theta) \cdot U_T^\theta)^2 \right] e^{-2\theta \cdot W_T - |\theta|^2 T} \right), \quad (2.3)$$

where  $U^\theta$  is the weak limit process of the error  $\sqrt{n}(X^{n,\theta} - X^\theta)$  defined on the extension  $\tilde{\mathcal{B}}$  of the original probability space  $\mathcal{B}$  and solution to

$$dU_t^\theta = \left( \dot{b}(X_t^\theta) + \sum_{j=1}^q \theta_j \dot{\sigma}_j(X_t^\theta) \right) U_t^\theta dt + \sum_{j=1}^q \dot{\sigma}_j(X_t^\theta) U_t^\theta dW_t^j - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^q \dot{\sigma}_j(X_t^\theta) \sigma_\ell(X_t^\theta) d\tilde{W}_t^{\ell j}. \quad (2.4)$$

As  $v(\theta)$  is not explicit, we use the Euler scheme to approximate  $(X^\theta, U^\theta)$  by  $(X^{n,\theta}, U^{n,\theta})$  and we choose the associated

$$\theta_n^* := \operatorname{argmin}_{\theta \in \mathbb{R}^q} v_n(\theta) \quad \text{with} \quad v_n(\theta) := \tilde{\mathbb{E}} \left( \left[ \psi(X_T^{n,\theta})^2 + (\nabla \psi(X_T^{n,\theta}) \cdot U_T^{n,\theta})^2 \right] e^{-2\theta \cdot W_T - |\theta|^2 T} \right). \quad (2.5)$$

We prove a first result on the convergence of the variance optimizer of the discretized problem  $\theta_n^*$  toward  $\theta^*$ .

**Theorem 2.2.1.** *Suppose  $\sigma$  and  $b$  are in  $\mathcal{C}_b^{1,1*}$  and let the payoff function  $\psi$  satisfying  $\mathbb{P}(\psi(X_T) \neq 0) > 0$ ,  $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$  together with condition (1.13). Then,  $\theta \mapsto v(\theta)$  and  $\theta \mapsto v_n(\theta)$  are  $\mathcal{C}^2$  strictly convex functions and there exist unique  $\theta^*$  and  $\theta_n^*$  such that  $\min_{\theta \in \mathbb{R}^q} v(\theta) = v(\theta^*)$  and  $\min_{\theta \in \mathbb{R}^q} v_n(\theta) = v_n(\theta_n^*)$ . Moreover, we have*

$$\theta_n^* \longrightarrow \theta^* \quad \text{as} \quad n \rightarrow \infty. \quad (2.6)$$

The above regularity assumptions on the coefficients diffusions are essential to control the strong convergence rate of the error  $U^\theta - U^{n,\theta}$ . Now, following the classical Robbins-Monro approach, we construct recursively a sequence of random variables  $(\theta_i^n)_{i \in \mathbb{N}}$  in  $\mathbb{R}^q$  by adding independent innovations as follows

$$\theta_{i+1}^n = \theta_i^n - \gamma_{i+1} H(\theta_i^n, X_{T,i+1}^n, U_{T,i+1}^n, W_{T,i+1}), \quad i \geq 0, \quad \theta_0^n \in \mathbb{R}^q, \quad (2.7)$$

where  $(\theta, x, u, w) \mapsto H(\theta, x, u, w) := (\theta T - w) [\psi(x)^2 + (\nabla \psi(x) \cdot u)^2] e^{-\theta \cdot w + \frac{1}{2} |\theta|^2 T}$  and  $(\gamma_i)_{i \geq 1}$  is a decreasing sequence of positive real numbers satisfying  $\sum_{i=1}^{\infty} \gamma_i = \infty$  and  $\sum_{i=1}^{\infty} \gamma_i^2 < \infty$ . Since the (NEC) condition is clearly not satisfied in this context, we then use two different versions of the Robbins-Monro procedure.

### 2.2.1 The constrained stochastic algorithm

To avoid explosion in (2.7), the constrained algorithm as introduced by Chen et al. [CZ86, CLG88] uses a repeated random truncation each time the algorithm leaves a slowly growing compact set sequence to wait for stabilization. In [BAHK15], we develop this routine to approximate the variance optimize  $\theta^*$  of the SR method. In more details, let  $(\mathcal{K}_i)_{i \in \mathbb{N}}$  denote an increasing sequence of compact sets satisfying  $\bigcup_{i=0}^{\infty} \mathcal{K}_i = \mathbb{R}^d$  and  $\mathcal{K}_i \subsetneq \overset{\circ}{\mathcal{K}}_{i+1}, \forall i \in \mathbb{N}$ . For  $\theta_0^n \in \mathcal{K}_0$ ,  $\alpha_0^n = 0$ , we define the sequence  $(\theta_i^n, \alpha_i^n)_{i \in \mathbb{N}}$  recursively by

$$\begin{cases} \text{if} & \theta_i^n - \gamma_{i+1} H(\theta_i^n, X_{T,i+1}^n, U_{T,i+1}^n, W_{T,i+1}) \in \mathcal{K}_{\alpha_i^n}, \text{ then} \\ & \theta_{i+1}^n = \theta_i^n - \gamma_{i+1} H(\theta_i^n, X_{T,i+1}^n, U_{T,i+1}^n, W_{T,i+1}), \text{ and } \alpha_{i+1}^n = \alpha_i^n \\ \text{else} & \theta_{i+1}^n = \theta_0^n \text{ and } \alpha_{i+1}^n = \alpha_i^n + 1. \end{cases} \quad (2.8)$$

Actually, as we can see, if the  $(i+1)^{th}$  iteration of the Robbins-Monro recursion is in the compact set  $\mathcal{K}_{\alpha_i^n}$ , then the algorithm will behave like a regular one. However, if the  $(i+1)^{th}$  iteration is outside the compact set  $\mathcal{K}_{\alpha_i^n}$ , it is started over with a new larger compact set  $\mathcal{K}_{\alpha_i^n+1}$ , for the next shot. Under assumptions of Theorem 2.2.1, we easily prove that the above randomly truncated recursion satisfies

$$\lim_{i, n \rightarrow \infty} \theta_i^n = \lim_{i \rightarrow \infty} \left( \lim_{n \rightarrow \infty} \theta_i^n \right) = \lim_{n \rightarrow \infty} \left( \lim_{i \rightarrow \infty} \theta_i^n \right) = \theta^*, \quad \tilde{\mathbb{P}}\text{-a.s.} \quad (2.9)$$

---

\*For an integer  $k \geq 1$  and  $\delta \in [0, 1]$ , we denote by  $\mathcal{C}_b^{k, \delta}$  the set of functions  $g$  in  $\mathcal{C}^k$  with  $k^{th}$  order partial derivatives globally  $\delta$ -Hölder and all partial derivatives up to  $k^{th}$  order bounded. In case  $\delta = 0$  we simply use the usual notation  $\mathcal{C}_b^k$ .



### 2.2.2 The unconstrained stochastic algorithm

In practice, it is known that tuning the above random truncation is not always an easy task. Lemaire and Pagès [LP10] proposed an alternative stochastic algorithm without resorting to random truncation techniques. In [BAHK15], we follow their approach to develop an unconstrained stochastic algorithm reducing the variance of the SR method. To do so, we first apply Girsanov's theorem with a shift parameter  $-\theta$  to get a new representation of  $\nabla v_n(\theta) = \tilde{\mathbb{E}} \left[ (2\theta T - W_T) \left[ \psi(X_T^{n,(-\theta)})^2 + (\nabla \psi(X_T^{n,(-\theta)}) \cdot U_T^{n,(-\theta)})^2 \right] e^{|\theta|^2 T} \right]$ . The aim now is to find the unique zero of the above function. Then, we approximate the optimal shift parameter  $\theta^*$  by the following recursion

$$\theta_{i+1}^n = \theta_i^n - \gamma_{i+1} \tilde{H}_\eta(\theta_i^n, X_{T,i+1}^{n,(-\theta_i^n)}, U_{T,i+1}^{n,(-\theta_i^n)}, W_{T,i+1}), \quad \theta_0 \in \mathbb{R}^q, \quad (2.10)$$

where for a given  $\eta > 0$ ,

$$\tilde{H}_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T) = e^{-\eta|\theta|^2 T} (2\theta T - W_T) \left[ \psi(X_T^{n,(-\theta)})^2 + (\nabla \psi(X_T^{n,(-\theta)}) \cdot U_T^{n,(-\theta)})^2 \right].$$

The main idea here is that the coefficient  $e^{-\eta|\theta|^2 T}$  makes the above algorithm satisfy the (NEC) condition without changing its limit.

On the one hand, under assumptions of Theorem 2.2.1, we prove that the routine (2.10) satisfies:

(•) For all  $n \in \mathbb{N}$ ,  $\theta_i^n \xrightarrow{i \rightarrow \infty} \theta_n^*$   $\tilde{\mathbb{P}}$ -a.s.

(••) The sequence  $(\theta_i)_{i \geq 0}$ , obtained when replacing in routine (2.10)  $(X_{T,i}^n, U_{T,i}^n)_{i \geq 1}$  by their limit  $(X_{T,i}, U_{T,i})_{i \geq 1}$ , satisfies  $\theta_i \xrightarrow{i \rightarrow \infty} \theta^*$ ,  $\tilde{\mathbb{P}}$ -a.s.

Hence, combining (•) with (2.6) yields  $\lim_{n \rightarrow \infty} (\lim_{i \rightarrow \infty} \theta_i^n) = \theta^*$ ,  $\tilde{\mathbb{P}}$ -a.s.

On the other hand, studying the limit of  $\theta_i^n$  as  $n$  tends to infinity looks more complicated to achieve, since for a fixed  $i \geq 0$  the stochastic term  $\theta_i^n$  also appears in the drift part of the pair  $(X_{T,i+1}^{n,(-\theta_i^n)}, U_{T,i+1}^{n,(-\theta_i^n)})$ . To overcome this technical difficulty we strengthen our assumptions on the triplet  $(b, \sigma, \psi)$  and prove that the processes given by  $(X_t^{(-\theta)})_{t \in [0, T]}$  and  $(U_t^{(-\theta)})_{t \in [0, T]}$  have modifications of  $\mathcal{C}^1$  with respect to the parameter  $\theta$  and their partial derivatives are  $L^p$ -bounded for all  $p \geq 1$ . This allows us to prove the following technical result.

**Theorem 2.2.2.** *Let  $b$  and  $\sigma$  in  $\mathcal{C}_b^{2, \delta}$ ,  $\delta > 0$ . Assume that  $\psi$  is  $\mathcal{C}^2$  with polynomial growth as well as all its partial derivatives until order two and satisfies  $\mathbb{P}(\psi(X_T) \neq 0) > 0$  and  $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$ , for all  $n \geq 1$ . Then,  $\forall i \in \mathbb{N}$  and  $\forall p \geq 1$ , there exists  $C > 0$  depending only on  $i, p, b, \sigma$  and  $T$ , such that*

$$\forall n \in \mathbb{N}^*, \quad \tilde{\mathbb{E}} |\theta_{i+1}^n - \theta_{i+1}|^{2p} \leq \frac{C}{n^p}.$$

Consequently,  $\forall i \in \mathbb{N}$   $\theta_i^n \xrightarrow{n \rightarrow \infty} \theta_i$   $\tilde{\mathbb{P}}$ -a.s.

Thus, combining this last result with (••) we deduce that the sequence given by the unconstrained recursion (2.10) satisfies

$$\lim_{i,n \rightarrow \infty} \theta_i^n = \lim_{i \rightarrow \infty} (\lim_{n \rightarrow \infty} \theta_i^n) = \lim_{n \rightarrow \infty} (\lim_{i \rightarrow \infty} \theta_i^n) = \theta^*, \quad \tilde{\mathbb{P}}\text{-a.s.} \quad (2.11)$$

So, the same result as for the constrained algorithm (2.9) is satisfied, but under additional regularity assumptions.

### 2.2.3 A central limit theorem for the adaptive SR algorithm

The double limit results obtained above are crucial for the study of the asymptotic normality of the adaptive SR method. For  $\beta \in (0, 1)$ ,  $\gamma_1 > 1$ ,  $\gamma_2 > 1$  and  $g(\theta, x, y) = \psi(x)e^{-\theta \cdot y - \frac{1}{2}|\theta|^2 T}$  with  $x \in \mathbb{R}^d$  and  $y \in \mathbb{R}^q$ , this latter method is given by

$$\begin{aligned} V_n := & \frac{1}{n^{\gamma_1}} \sum_{i=1}^{n^{\gamma_1}} g(\hat{\theta}_i^{n^\beta}, \hat{X}_{T,i+1}^{n^\beta, \hat{\theta}_i^{n^\beta}}, \hat{W}_{T,i+1}) \\ & + \frac{1}{n^{\gamma_2}} \sum_{i=1}^{n^{\gamma_2}} \left( g(\theta_i^n, X_{T,i+1}^{n, \theta_i^n}, W_{T,i+1}) - g(\theta_i^n, X_{T,i+1}^{n^\beta, \theta_i^n}, W_{T,i+1}) \right), \end{aligned} \quad (2.12)$$

where  $(\theta_i^n)_{i \geq 0, n \geq 1}$  is any of the double indexed sequences (2.8) or (2.10) satisfying respectively (2.9) or (2.11) with the corresponding regularity assumptions given respectively in Theorem 2.2.1 or Theorem 2.2.2. Moreover, for a payoff function  $\psi$  satisfying  $(\mathcal{H}_\psi)$ ,  $(\mathcal{H}_{\varepsilon_n})$  with constant  $C_\psi(T, \alpha) \in \mathbb{R}$  and  $\alpha \in (1/2, 1]$  (see Section 1.1), if we choose  $N_1 = n^{2\alpha}$ ,  $N_2 = n^{2\alpha - \beta}$ , then

$$n^\alpha (V_n - \mathbb{E}\psi(X_T)) \implies \mathcal{N}(C_\psi(T, \alpha), \sigma^2), \quad \text{as } n \rightarrow \infty,$$

where  $\sigma^2 := \tilde{\mathbb{E}} \left[ \left( \psi(X_T)^2 + [\nabla \psi(X_T) \cdot U_T]^2 \right) e^{-\theta^* \cdot W_T - \frac{1}{2}|\theta^*|^2 T} \right] - [\mathbb{E}\psi(X_T)]^2$ . The same computational complexity analysis as in Section 1.2 holds with  $\beta^* = 1/2$ . Then, we provide in Section 5 of [BAHK15] numerical tests in which we show the gain obtained when using the SR method combined with importance sampling over a classical Monte Carlo method combined with Importance sampling to price European calls in the Heston model.

## 2.3 Importance Sampling and MLMC method: a stochastic algorithm approach for Brownian diffusions

In [BHK16], we focus on extending the results of the previous section to the setting of MLMC methods. Following the same notations, the variance optimizer in this case is, according to Theorem 1.4.1, defined by

$$\theta^* = \arg \min_{\theta \in \mathbb{R}^q} v(\theta) \quad \text{where } v(\theta) := \tilde{\text{Var}} \left( \nabla \psi(X_T^\theta) \cdot U_T \right). \quad (2.13)$$

At this stage, we come up with a new idea by changing the way of approximating  $\theta^*$ . More precisely, we suggest to approximate  $\theta^*$  by  $\theta_\ell^* := \arg \min_{\theta \in \mathbb{R}^q} v_\ell(\theta)$ , with

$$\begin{aligned} v_\ell(\theta) &:= \mathbb{E} \left[ \left( \sqrt{\frac{m^\ell}{m-1}} (\psi(X_T^{m^\ell, \theta}) - \psi(X_T^{m^{\ell-1}, \theta})) \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \sqrt{\frac{m^\ell}{m-1}} (\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})) \right)^2 e^{-\theta \cdot W_T + \frac{1}{2} |\theta|^2 T} \right], \text{ for } \ell \geq 1, \end{aligned} \quad (2.14)$$

$$v_0(\theta) := \mathbb{E} \left[ \psi(X_T^{m^0})^2 e^{-\theta \cdot W_T + \frac{1}{2} |\theta|^2 T} \right]. \quad (2.15)$$

Unlike (2.5), this new variance associated to the MLMC is only based on the discretization of the process  $X$  and does not involve  $\nabla \psi$ , so clearly it is more convenient in practice. Now, let us introduce the following assumption

$$(\mathcal{R}_{\psi, a}) : \begin{cases} \text{Conditions } (\mathcal{H}_\psi) \text{ and (1.13) are satisfied and } \mathbb{P}((\nabla \psi(X_T) \cdot U_T) \neq 0) > 0. \\ \text{There exists } a > 1 \text{ such that } \mathbb{E} [|\nabla \psi(X_T)|^{2a}] < \infty. \end{cases}$$

**Proposition 2.3.1.** *Let assumption  $(\mathcal{R}_{\psi, a})$  holds and the diffusion coefficients  $\sigma$  and  $b$  be in  $\mathcal{C}_b^1$ . Assume that  $\mathbb{P}((\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})) \neq 0) > 0$  for  $\ell \geq 1$  and  $\mathbb{P}((\psi(X_T^{m^0}) \neq 0) > 0$ . Then,  $\theta \mapsto v(\theta)$  and  $\theta \mapsto v_\ell(\theta)$  are  $\mathcal{C}^2$  strictly convex functions. Then, there exist unique  $\theta^*$  and  $\theta_\ell^*$  such that  $\min_{\theta \in \mathbb{R}^q} v(\theta) = v(\theta^*)$  and  $\min_{\theta \in \mathbb{R}^q} v_\ell(\theta) = v_\ell(\theta_\ell^*)$ . Moreover, we have*

$$\theta_\ell^* \longrightarrow \theta^*, \quad \text{as } \ell \rightarrow \infty.$$

From a technical point of view, the use of the constrained and unconstrained versions of the Robbins-Monro algorithms (see (2.8) and (2.10)) seems to be problematic in the study of the asymptotic normality of the MLMC estimator coupled with the importance sampling technique. To cope with this difficulty, we consider an alternative version of Robbins-Monro type algorithms, namely the stochastic algorithm with projection on a fixed compact set (see e.g. Laruelle et al. [LLP13]). In more details, we consider a compact convex subset  $K \subset \mathbb{R}^q$ , satisfying  $0 \in \overset{\circ}{K}$  (the interior of  $K$ ). For a deterministic  $\theta_0 \in K$ , we introduce the sequences  $(\theta_i)_{i \in \mathbb{N}}$  and  $(\theta_i^{m^\ell})_{i \in \mathbb{N}}$ ,  $\ell \in \mathbb{N}$ , defined recursively by

$$\begin{cases} \theta_{i+1} &= \Pi_K [\theta_i - \gamma_{i+1} H(\theta_i, X_{T, i+1}, U_{T, i+1}, W_{T, i+1})], \\ \theta_{i+1}^{m^\ell} &= \Pi_K \left[ \theta_i^{m^\ell} - \gamma_{i+1} H_\ell(\theta_i^{m^\ell}, X_{T, i+1}^{m^\ell}, X_{T, i+1}^{m^{\ell-1}}, W_{T, i+1}) \right], \quad \theta_0^{m^\ell} = \theta_0, \end{cases} \quad (2.16)$$

where  $\Pi_K$  is the Euclidean projection onto the constraint set  $K$ ,  $H$  and  $H_\ell$  are given respectively by the following expressions

$$\begin{aligned} H(\theta, X_T, U_T, W_T) &:= (\theta T - W_T) (\nabla \psi(X_T) \cdot U_T)^2 e^{-\theta \cdot W_T + \frac{1}{2} |\theta|^2 T} \\ H_\ell(\theta, X_T^{m^\ell}, X_T^{m^{\ell-1}}, W_T) &:= (\theta T - W_T) \left( \sqrt{\frac{m^\ell}{m-1}} (\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})) \right)^2 \\ &\quad \times e^{-\theta \cdot W_T + \frac{1}{2} |\theta|^2 T}, \text{ for } \ell \geq 1 \\ H_0(\theta, X_T^{m^0}, X_T^{m^{-1}}, W_T) &:= (\theta T - W_T) \psi(X_T^{m^0})^2 e^{-\theta \cdot W_T + \frac{1}{2} |\theta|^2 T}. \end{aligned}$$

Under assumptions of (2.3.1), we prove that  $\lim_{n \rightarrow \infty} (\lim_{i \rightarrow \infty} \theta_i^n) = \theta^*$   $\tilde{\mathbb{P}}$ - a.s. However, when reverting the indexes order we get a new type of convergence

$$\theta_i^{m^\ell} \xrightarrow{\text{stably}} \theta_i, \text{ as } \ell \rightarrow \infty \text{ and } \theta_i \xrightarrow{a.s.} \theta^* \text{ as } i \rightarrow \infty.$$

To the best of our knowledge such result does not exist in the literature around stochastic algorithms and it turns out to be crucial for the study of the adaptive MLMC algorithm. In this respect, this latter approximates our initial quantity of interest  $\mathbb{E}\psi(X_T) = \mathbb{E} \left[ \psi(X_T^\theta) e^{-\theta \cdot W_T - \frac{1}{2}|\theta|^2 T} \right]$  by

$$Q_n := \frac{1}{N_0} \sum_{i=1}^{N_0} g(\theta_{i-1}^{m^0}, X_{T,i}^{m^0}, \theta_{i-1}^{m^0}, \hat{W}_{T,i}) + \sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left( g(\theta_{i-1}^{\ell, m^\ell}, X_{T,i}^{\ell, m^\ell}, \theta_{i-1}^{\ell, m^\ell}, W_{T,i}^\ell) - g(\theta_{i-1}^{\ell, m^\ell}, X_{T,i}^{\ell, m^{\ell-1}}, \theta_{i-1}^{\ell, m^\ell}, W_{T,i}^\ell) \right), \quad (2.17)$$

where for all  $x \in \mathbb{R}^d$  and  $y \in \mathbb{R}^q$ ,  $g(\theta, x, y) = \psi(x) e^{-\theta \cdot y - \frac{1}{2}|\theta|^2 T}$ ,  $L = \frac{\log n}{\log m}$ . As it can be noticed, here one importance sampling parameter is carried out per each level so that we reduce the effective variance in each MLMC block. Then, we prove the following result.

**Theorem 2.3.1.** *Let assumptions of Proposition 2.3.1 hold. If moreover assumption  $(\mathcal{H}_{\varepsilon_n})$  is satisfied with constant  $C_\psi(T, \alpha) \in \mathbb{R}$  and  $\alpha \in (1/2, 1]$ , then for the choice of  $N_\ell, \ell \in \{0, 1, \dots, L\}$  given by (2.19), we have as  $n \rightarrow \infty$ ,*

$$n^\alpha (Q_n - \mathbb{E}\psi(X_T)) \implies \mathcal{N}(C_\psi(T, \alpha), \tilde{\sigma}^2), \text{ where } \tilde{\sigma}^2 = \tilde{\mathbb{E}} \left[ [\nabla \psi(X_T) \cdot U_T]^2 e^{-\theta^* \cdot W_T + \frac{1}{2}|\theta^*|^2 T} \right].$$

Due to its design, it is clear that the adaptive importance sampling Multilevel Monte Carlo approach (AIS MLMC) is more time consuming than the standard MLMC method. However, in practice we do not need to reach the optimal variance but just to be close enough to it. Based on this idea, we enforce the adaptive stochastic algorithm to stop after  $I \in \mathbb{N}$  iterations. Therefore, the time complexity of the stopped AIS MLMC method is given by  $C_{\text{AISMLMC}} = C \times I \times \sum_{\ell=0}^L m^\ell + C' \times \left( \sum_{\ell=0}^L N_\ell (m^\ell + m^{\ell-1}) \right)$  with  $C, C' > 0$ . For the same specific choice  $a_\ell^* = 1$ , the optimal complexity is then given by

$$C_{\text{AISMLMC}} \propto n^{2\alpha} (\log n)^2 \left( 1 + \frac{I}{n^\alpha (\log n)^2} \right).$$

We implement the AIS MLMC method with both routines the random truncation with increasing compact sets (Chen's projection) (2.8) and the stochastic algorithm with projection on a fixed compact set (2.16). According to our numerical simulations (see Figure 2.1), the performance of the AIS MLMC method seems to be quite similar if it is stopped after  $I = 15000$  or  $I = 1000$  iterations and for both types of stochastic algorithms. Hence, it is comforting to notice that our AIS MLMC method approximates efficiently the optimal parameter  $\theta^*$  reducing the total variance after just 1000 iterations. For our numerical tests, we consider the problem of pricing call European options under the Black & Scholes model.

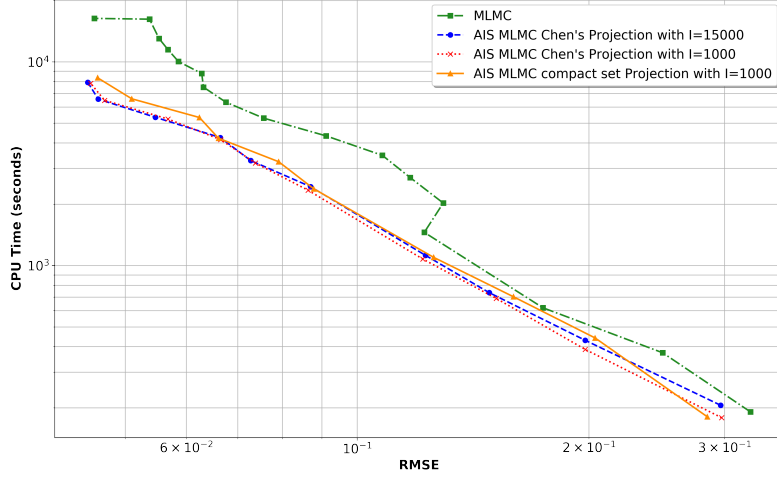


Figure 2.1: *CPU time vs. RMSE for an European call option under the B&S model.*

## 2.4 Importance Sampling and MLMC method: a sample average approach optimization for Brownian diffusions

In [KL17], we use a different approach to approximate the variance optimizer  $\theta^*$  (see (2.13))

$$\theta^* = \arg \min_{\theta \in \mathbb{R}^q} v(\theta), \quad \text{where } v(\theta) := \tilde{\text{Var}} \left( \nabla \psi(X_T^\theta) \cdot U_T \right).$$

Here, we apply a deterministic optimization to sample average estimators instead of stochastic routines. This idea was introduced in the setting of Monte Carlo methods by Jourdain and Lelong [JL09]. To do so, we introduce, for  $\lambda_1, \dots, \lambda_L \in \mathbb{R}^q$ , the MLMC estimator

$$\begin{aligned} Q_L(\lambda_0, \dots, \lambda_L) &= \frac{1}{N_0} \sum_{k=1}^{N_0} \psi(\tilde{X}_{T,0,k}^{m_0}(\lambda_0)) \mathcal{E}^-(\tilde{W}_{0,k}, \lambda_0) \\ &\quad + \sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \left( \psi(\tilde{X}_{T,\ell,k}^{m_\ell}(\lambda_\ell)) - \psi(\tilde{X}_{T,\ell,k}^{m_{\ell-1}}(\lambda_\ell)) \right) \mathcal{E}^-(\tilde{W}_{\ell,k}, \lambda_\ell), \end{aligned} \quad (2.18)$$

with  $\mathcal{E}^-(W, \theta) := \exp(-\theta \cdot W_T - \frac{1}{2}|\theta|^2 T)$ . For any fixed  $\ell \in \{1, \dots, L\}$ , the random variables  $(\tilde{W}_{\ell,k})_{1 \leq k \leq N_\ell}$  are independent and are distributed according to the Brownian law. We assume that for  $\ell, \ell' \in \{1, \dots, L\}$ , with  $\ell \neq \ell'$ , the blocks  $(\tilde{W}_{\ell,k})_{1 \leq k \leq N_\ell}$  and  $(\tilde{W}_{\ell',k})_{1 \leq k \leq N_{\ell'}}$  are independent. To actually minimize the empirical variance of  $Q_L$ , we consider the sample average approximation of size  $N'_\ell$ , given by

$$\begin{aligned} v_{0,N'_0}(\lambda_0) &:= \frac{1}{N'_0} \sum_{k=1}^{N'_0} \psi(X_{T,0,k}^{m_0})^2 \mathcal{E}^+(W_{0,k}, \lambda_0), \\ v_{\ell,N'_\ell}(\lambda_\ell) &:= \frac{1}{N'_\ell} \sum_{k=1}^{N'_\ell} \frac{m_\ell}{(m_\ell - 1)} \left| \psi(X_{T,\ell,k}^{m_\ell}) - \psi(X_{T,\ell,k}^{m_{\ell-1}}) \right|^2 \mathcal{E}^+(W_{\ell,k}, \lambda_\ell). \end{aligned}$$

We prove that  $v_{\ell, N'_\ell}$  has a unique minimum

$$\widehat{\lambda}_\ell = \arg \min_{\lambda \in \mathbb{R}^q} v_{\ell, N'_\ell}(\lambda).$$

and we have the following result.

**Theorem 2.4.1.** *Assume  $b$  and  $\sigma$  are  $\mathcal{C}_b^1$ . Let the payoff function  $\psi$  satisfy  $(\mathcal{H}_\psi)$  and (1.13) such that  $\nabla\psi$  has polynomial growth. Then, the sequence of random functions  $(v_{\ell, N'_\ell} : \lambda \in \mathbb{R}^q \rightarrow v_{\ell, N'_\ell}(\lambda))_\ell$  converges a.s. locally uniformly to the strongly convex function  $v : \mathbb{R}^q \rightarrow \mathbb{R}$  given by (2.13). Moreover,*

$$\widehat{\lambda}_\ell \xrightarrow{\mathbb{P}\text{-a.s.}} \theta^* \quad \text{as } \ell \rightarrow \infty.$$

In this setting, the minimization step is performed by applying the Newton–Raphson algorithm to  $\nabla v_{\ell, N'_\ell}$ . Here, we emphasize once again, that the samples  $(W_{\ell, k})_{\ell, k \geq 1}$  used to compute  $\widehat{\lambda}_L$  are independent of the variables  $(\widetilde{W}_{\ell, k})_{\ell, k}$  needed for the implementation of the MLMC method. From a practical point of view, the efficiency of the Newton–Raphson algorithm mainly depends on the convexity of the  $v_{\ell, N'_\ell}$  functions. However, for  $\mathcal{E}^+(W, \lambda) := e^{-\lambda \cdot W_T + \frac{1}{2}|\lambda|^2 T}$ , the smallest eigenvalue of the Hessian matrix  $\nabla^2 v_{\ell, N'_\ell}$  is essentially given by  $\frac{T}{N'_\ell} \sum_{k=1}^{N'_\ell} \frac{m^\ell}{(m-1)} |\psi(X_{T, \ell, k}^{m^\ell}) - \psi(X_{T, \ell, k}^{m^{\ell-1}})|^2 \mathcal{E}^+(W_{\ell, k}, \lambda)$ , which can become extremely small and then conflicts with the will to have the strongest possible convexity in order to speed up Newton–Raphson’s algorithm. We overcome this difficulty by noticing that  $\widehat{\lambda}_\ell$  is at the same time the root of  $\nabla u_{\ell, N'_\ell}$  with

$$u_{\ell, N'_\ell}(\lambda) = \frac{|\lambda|^2 T}{2} + \log \left( \frac{1}{N'_\ell} \sum_{k=1}^{N'_\ell} \frac{m^\ell}{(m-1)} |\psi(X_{T, \ell, k}^{m^\ell}) - \psi(X_{T, \ell, k}^{m^{\ell-1}})|^2 e^{-\lambda \cdot W_{T, \ell, k}} \right),$$

since it can be easily seen that the Hessian matrix of  $u_{\ell, N'_\ell}$  is lower bounded by  $TI_q$ , which is a good feature for our numerical approximation.

In [KL17], we prove a strong large law of numbers for the MLMC algorithm. Considering a sequence  $(a_\ell)_{\ell \in \mathbb{N}}$  of positive real numbers such that  $\lim_{L \rightarrow \infty} \sum_{\ell=1}^L a_\ell = \infty$ , we assume that the sample size  $N_\ell$  has the following form

$$N_{\ell, L}^\rho = \frac{\rho(L)}{m^\ell a_\ell} \sum_{k=1}^L a_k, \quad \ell \in \{0, \dots, L\} \quad (2.19)$$

for some increasing function  $\rho : \mathbb{N} \rightarrow \mathbb{R}$ . Then,

**Theorem 2.4.2.** *Assume that  $\sup_L \sup_\ell \frac{L^2 a_\ell}{\rho(L) \sum_{k=1}^L a_k} < +\infty$ . Then, under the assumptions of Theorem 2.4.1,  $Q_L(\widehat{\lambda}_0, \dots, \widehat{\lambda}_L) \xrightarrow{\text{a.s.}} \mathbb{E}[\psi(X_T)]$  when  $L \rightarrow +\infty$ .*

For the choice  $a_\ell = 1$  for all  $\ell$ , the condition on  $\rho$  reduces to  $\sup_L \frac{L}{\rho(L)} < +\infty$ . Further, if we assume the weights  $a_\ell$  satisfy assumption  $(\mathcal{W})$ , given in Theorem 1.4.1, we prove a central limit theorem.

**Theorem 2.4.3.** *Under assumptions of Theorem 2.4.2, we consider  $N_{\ell,L}^p$  as given in (2.19) with  $\rho(L) = m^{2\gamma L}(m-1)T$ . If condition  $(\mathcal{H}_{\varepsilon_n})$  is satisfied for some  $\alpha \in [1/2, 1]$  and constant  $C_\psi(T, \alpha)$ , then*

$$m^L(Q_L(\hat{\lambda}_0, \dots, \hat{\lambda}_L) - \mathbb{E}[\psi(X_T)]) \implies \mathcal{N}(C_\psi(T, \alpha), v(\theta^*)), \text{ when } L \rightarrow \infty.$$

The global cost of our algorithm writes as the sum of the cost of the computation of the  $(\hat{\lambda}_\ell)_\ell$  and of the standard multilevel estimator

$$C_{\text{ML IS}} = \sum_{\ell=0}^L N'_\ell(m^\ell + 3K_\ell) + \sum_{\ell=0}^L N_\ell m^\ell,$$

where  $K_\ell$  is the number of iterations of Newton–Raphson’s algorithm to approximate  $\hat{\lambda}_\ell$  and the factor 3 corresponds to the fact that building  $\nabla u_{\ell, N'_\ell}$  and  $\nabla^2 u_{\ell, N'_\ell}$  basically boils down to three Monte Carlo summations. In practice,  $K_\ell \leq 5$  as the problem is strongly convex. Because the same random variables are used at each iteration of the optimisation step, they must be stored, which makes the memory footprint of our algorithm proportional to  $N'_\ell$ . From a practical point of view, choosing  $N'_\ell = \frac{N_\ell m^\ell}{m^{\ell+15}} \wedge 500000$ , the total cost of our ML IS algorithm is roughly twice the cost of the standard multilevel estimator (ML). Numerical tests, on pricing basket options in a multidimensional local volatility model, confirm that the ML IS estimator outperforms the standard MLMC estimator. To conclude let us precise that this approach has three main advantages. First,

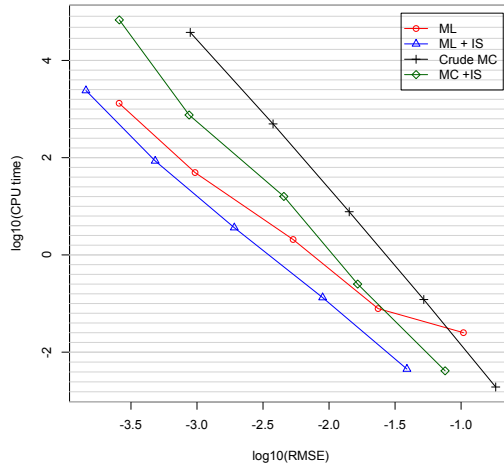


Figure 2.2: RMSE vs. CPU time for a basket option in a local volatility model.

the computations within the different levels remain independent. Second, the variance of each level  $\ell$  only depends on  $\lambda_\ell$ , which reduces the global minimization problem to several smaller minimization problems. Third, we actually minimize the real variance of the estimator and not its asymptotic value.

## 2.5 Importance Sampling and Statistical Romberg method for pure-jump Lévy processes

In [BAHK16], we focus on finding a way to implement the SR method and implicitly the MLMC method, when only pure-jump Lévy processes  $(L_t)_{t \geq 0}$  are used to model the underlying price instead of  $(X_t)_{t \geq 0}$  the solution of the stochastic differential equation (1.1). In what follows, we assume that  $(L_t)_{t \geq 0}$  has a generating triplet  $(\gamma, 0, \nu)$  and a Lévy Kintchine representation

$$\mathbb{E}e^{iu \cdot L_t} = \exp \left\{ t \left( i\gamma \cdot u + \int_{\mathbb{R}^d} (e^{iu \cdot x} - 1 - iu \cdot x \mathbf{1}_{|x| \leq 1}) \nu(dx) \right) \right\}, \quad u \in \mathbb{R}^d,$$

where  $\gamma \in \mathbb{R}^d$  and  $\nu$  is a Lévy measure on  $\mathbb{R}^d \setminus \{0\}$  verifying  $\int_{\mathbb{R}^d \setminus \{0\}} (|x|^2 \wedge 1) \nu(dx) < \infty$ . At a first glance, it seems quite unlikely that such a procedure with pure-jump Lévy processes would work, since the design of MLMC methods requires the use of a discretization scheme or at least an inner iterative routine that can be recycled from the finest level to crudest one. However, it is known in the literature (see e.g. Asmussen and Rosiński [AR01]) that when the increments of the jump process cannot be simulated,  $L$  can be represented as a sum of a compound Poisson process and an almost sure limit of compensated compound Poisson process  $L_t = \lim_{\varepsilon \rightarrow 0} L_t^\varepsilon$  a.s. where for  $0 < \varepsilon < 1$

$$L_t^\varepsilon = \gamma t + \sum_{0 < s \leq t} \Delta L_s \mathbf{1}_{|\Delta L_s| > 1} + \left( \sum_{0 < s \leq t} \Delta L_s \mathbf{1}_{\varepsilon \leq |\Delta L_s| \leq 1} - t \int_{\varepsilon \leq |x| \leq 1} x \nu(dx) \right), \quad t \geq 0. \quad (2.20)$$

The error process  $R^\varepsilon := L - L^\varepsilon$  is also a Lévy process independent of  $L^\varepsilon$  with characteristic function  $\mathbb{E}e^{iu \cdot R_t^\varepsilon} = \exp \left\{ t \int_{|x| \leq \varepsilon} (e^{iu \cdot x} - 1 - iu \cdot x) \nu(dx) \right\}$ . This independence feature of the error process is the keystone on which we build the implementation of MLMC type methods, for this setting. Consequently,  $\mathbb{E}[R_t^\varepsilon] = 0$  and the variance-covariance matrix  $\mathbb{E}[R_t^\varepsilon (R_t^\varepsilon)^\top] = t \Sigma_\varepsilon$  where  $\Sigma_\varepsilon = \int_{|x| \leq \varepsilon} x x^\top \nu(dx)$ . The asymptotic behavior of the distribution of  $R^\varepsilon$  is firstly studied by Asmussen and Rosiński [AR01] in the one dimensional case and later extended to the multidimensional case by Cohen and Rosiński [CR07]. In what follows,  $W = (W_t)_{t \geq 0}$  is a standard Brownian motion in  $\mathbb{R}^d$  independent of  $(L_t)_{t \geq 0}$ . If  $\Sigma_\varepsilon$  is invertible for every  $\varepsilon \in (0, 1]$ , then they prove that

$$\Sigma_\varepsilon^{-1/2} R^\varepsilon \Rightarrow W, \quad \text{as } \varepsilon \rightarrow 0, \quad (2.21)$$

if and only if for each  $k > 0$

$$\lim_{\varepsilon \rightarrow 0} \int_{\langle \Sigma_\varepsilon^{-1} x, x \rangle > k} \langle \Sigma_\varepsilon^{-1} x, x \rangle \mathbf{1}_{|x| \leq \varepsilon} \nu(dx) = 0. \quad (2.22)$$

**Strong error.** If  $\nu$  is given in polar coordinates by  $\nu(dr, du) = \mu(dr|u) \lambda(du)$ ,  $r > 0$ ,  $u \in S^{d-1}$ , where  $\{\mu(\cdot|u) : u \in S^{d-1}\}$  is a measurable family of Lévy measures on  $(0, \infty)$  and  $\lambda$  is a finite measure on the unit sphere  $S^{d-1}$ , then

$$\Sigma_\varepsilon = \int_{S^{d-1}} \int_0^\varepsilon r^2 u u^\top \mu(dr|u) \lambda(du).$$



If we define  $\sigma^2(\varepsilon, u) := \int_0^\varepsilon r^2 \mu(dr|u)$  and  $\sigma^2(\varepsilon) := \int_{S^{d-1}} \sigma^2(\varepsilon, u) \lambda(du)$ , then

$$\mathbb{E}|L_t - L_t^\varepsilon|^2 = t \text{Tr}(\Sigma_\varepsilon) = t \sigma^2(\varepsilon). \quad (2.23)$$

Moreover, if the support of the measure  $\lambda$  is not contained in any proper linear subspace of  $\mathbb{R}^d$ , they proved that if

$$\lim_{\varepsilon \rightarrow 0} \frac{\sigma(\varepsilon, u)}{\varepsilon} = \infty, \lambda - a.e., \quad (2.24)$$

then  $\Sigma_\varepsilon$  is invertible and condition (2.22) holds. Furthermore, following the proof of Proposition 2.1 given in Dia [Dia13] for a one-dimensional setting, we prove that the  $d$ -dimensional error Lévy process  $R^\varepsilon$  satisfies

$$\mathbb{E}|R_t^\varepsilon|^q \leq K_{q,T} \sigma_0(\varepsilon)^q, \quad \text{where } K_{q,T} > 0 \text{ and } \sigma_0(\varepsilon) = \sigma(\varepsilon) \vee \varepsilon. \quad (\text{SE})$$

**Weak error.** Concerning the weak error, under some regularity conditions on function  $F$ , we can obtain an expansion of the weak error as in Proposition 2.2 and Remark 2.3 of [Dia13]. So, it is worth introducing the following assumption: there exist  $C_F \in \mathbb{R}$  and  $v_\varepsilon \searrow 0$  as  $\varepsilon \searrow 0$  such that

$$v_\varepsilon^{-1} (\mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)) \longrightarrow C_F, \quad \text{as } \varepsilon \searrow 0. \quad (\text{WE}_{v_\varepsilon})$$

**SR method with pure-jump processes.** The main idea of this new method is to consider two cut-off sizes  $\varepsilon$  and  $\varepsilon^\beta$ ,  $\beta \in (0, 1)$  and then approximate  $\mathbb{E}F(L_T)$  by  $Q_\varepsilon := \frac{1}{N_1} \sum_{i=1}^{N_1} F(\hat{L}_{T,i}^{\varepsilon^\beta}) + \frac{1}{N_2} \sum_{i=1}^{N_2} F(L_{T,i}^\varepsilon) - F(L_{T,i}^{\varepsilon^\beta})$ . The samples  $(L_{T,i}^\varepsilon)_{1 \leq i \leq N_2}$  and  $(L_{T,i}^{\varepsilon^\beta})_{1 \leq i \leq N_2}$  have to be independent of  $(\hat{L}_{T,i}^\varepsilon)_{1 \leq i \leq N_1}$ . Moreover, for  $1 \leq i \leq N_2$ , we first simulate  $L_{T,i}^{\varepsilon^\beta}$  and then recycle it to simulate  $L_{T,i}^\varepsilon$  as the sum of  $L_{T,i}^{\varepsilon^\beta}$  and an independent r.v.  $L_{T,i}^{\varepsilon, \varepsilon^\beta}$  with Lévy measure  $\nu_{\{|x| \leq \varepsilon\}}$  which is also simulatable as a compound Poisson process. Then, we prove two central limit theorems for the standard MC method and for  $Q_\varepsilon$  leading us to the following computational complexity analysis. In more details, the computational effort of the MC method needed to achieve a precision of order  $\sigma(\varepsilon)$  is random with expectation of order  $C_{\text{MC}} = \mathcal{K}(\varepsilon)N = \mathcal{K}(\varepsilon)\sigma^{-2}(\varepsilon)$ . In the same way, the computational effort of the SR method necessary to achieve a precision of order  $\sigma(\varepsilon)$  is random with expectation of order  $C_{\text{SR}} = \mathcal{K}(\varepsilon^\beta)N_1 + \mathcal{K}(\varepsilon)N_2 = (\mathcal{K}(\varepsilon^\beta) + \mathcal{K}(\varepsilon)\sigma^2(\varepsilon^\beta))\sigma^{-2}(\varepsilon)$ . Thanks to Karamata's theorem (see e.g. Bingham, Goldie and Teugels [BGT87]), we were able to compute the time complexity ratio given by

$$\frac{C_{\text{SR}}}{C_{\text{MC}}} = \frac{L(\varepsilon^\beta) + L(-\varepsilon^\beta)}{L(\varepsilon) + L(-\varepsilon^\beta)} \varepsilon^{Y(1-\beta)} + \frac{L(\varepsilon^\beta) + L(-\varepsilon^\beta)}{2 - Y} \varepsilon^{\beta(2-Y)}.$$

If  $L(\varepsilon)$  is constant in the neighborhood of zero, which is the case for the Carr, Geman, Madan and Yor (CGMY) model [CGMY02], then we easily get

$$\frac{C_{\text{SR}}}{C_{\text{MC}}} = O\left(\varepsilon^{Y(1-\beta)} + \varepsilon^{\beta(2-Y)}\right).$$

Optimizing the order of this last quantity yields  $\beta = Y/2$ , which leads us to a gain of a complexity of order  $\varepsilon^{Y(Y/2-1)}$  that asymptotically increases as soon as  $\varepsilon$  becomes small.

Since the efficiency of the Monte Carlo simulation considerably depends on the smallness of the variance in the estimation, we propose a stochastic approximation method to find the optimal measure change by Esscher transform for Lévy processes for MC and SR methods when coupled with the importance sampling technique. Kawai [Kaw09] applied this technique for MC algorithms. His approach is exploitable only when the Lévy process  $(L_t)_{0 \leq t \leq T}$  is simulatable without any approximation. Note also that in his study there is no results on the rate of convergence of the obtained algorithm. To do so, we define the family of  $\{\mathbb{P}_\theta, \theta \in \Theta_1\}$  with

$$\Theta_1 := \{\theta \in \mathbb{R}^d : \mathbb{E}[e^{\theta \cdot L_t}] < +\infty\} = \{\theta \in \mathbb{R}^d : \int_{|x|>1} e^{\theta \cdot x} \nu(dx) < \infty\}, \quad (2.25)$$

as all the equivalent probability measures with respect to  $\mathbb{P}$  such that

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \frac{e^{\theta \cdot L_t}}{\mathbb{E}[e^{\theta \cdot L_t}]} = e^{\theta \cdot L_t - t\kappa(\theta)},$$

where  $\kappa$  denotes the cumulant generating function given by  $\kappa(\theta) = \ln \mathbb{E}[e^{\theta \cdot L_1}]$ . If we introduce the Lévy process  $\{L_t^\theta; t \geq 0\}$ , with generating triplet  $(\gamma_\theta, 0, \nu_\theta)$  under  $\mathbb{P}$ , where  $\gamma_\theta = \gamma + \int_{|x| \leq 1} x(\nu_\theta - \nu)(dx)$  and  $\nu_\theta(dx) = e^{\theta \cdot x} \nu(dx)$  (see e.g. Cont and Tankov [CT04]), then the random variable  $L_T$  under  $\mathbb{P}_\theta$  has the same law as  $L_T^\theta$  under  $\mathbb{P}$  and we get

$$\mathbb{E}[F(L_T)] = \mathbb{E}\left[F(L_T^\theta) e^{-\theta \cdot L_T + T\kappa(\theta)}\right].$$

Furthermore, we use this representation twice in the SR algorithm with considering  $\theta_1$  and  $\theta_2$  in  $\mathbb{R}^d$  to approximate  $\mathbb{E}[F(L_T)]$  by

$$\frac{1}{N_1} \sum_{k=1}^{N_1} F(L_{T,k}^{\varepsilon, \theta_1}) e^{-\theta_1 \cdot L_{T,k}^{\varepsilon, \theta_1} + T\kappa(\theta_1)} + \frac{1}{N_2} \sum_{k=1}^{N_2} (F(L_{T,k}^{\varepsilon, \theta_2}) - F(L_{T,k}^{\varepsilon, \theta_2})) e^{-\theta_2 \cdot L_{T,k}^{\varepsilon, \theta_2} + T\kappa(\theta_2)}.$$

Then, we minimize separately the two quantities involved in the limit variance of the above estimator given by,

$$v_i(\theta) := \mathbb{E}\left[F_i(L_T) e^{-\theta \cdot L_T + T\kappa(\theta)}\right], \quad i \in \{1, 2\} \text{ with } F_1 \equiv F^2 \text{ and } F_2 \equiv \nabla F \cdot \Sigma \nabla F. \quad (2.26)$$

To ensure the existence of  $v_i(\theta)$ , we introduce a first set

$$\Theta_{i,2} := \Theta_1 \cap \left\{ \theta \in \mathbb{R}^d : \mathbb{E}\left[F_i(L_T) e^{-\theta \cdot L_T}\right] < +\infty \right\}$$

and to make sense for the first and second derivatives of  $v_i(\theta)$  we introduce a second set

$$\Theta_{i,3} := \Theta_{i,2} \cap \left\{ \theta \in \mathbb{R}^d : \mathbb{E}\left[|L_T|^2 F_i(L_T) e^{-\theta \cdot L_T}\right] < +\infty \right\}.$$

Under assumptions  $Leb(\Theta_{i,3}) > 0$ , the sets  $\Theta_{i,2}$  and  $\Theta_{i,3}$  are convex. For  $\varepsilon > 0$ , the same result holds for the approximated Lévy process  $(L_t^\varepsilon)_{t \geq 0}$ , by considering the associated sets  $\Theta_1^\varepsilon$ ,  $\Theta_{i,2}^\varepsilon$  and  $\Theta_{i,3}^\varepsilon$  and functions  $\kappa_\varepsilon$  and  $v_{i,\varepsilon}$ ,  $i \in \{1, 2\}$ . Introducing, for  $i \in \{1, 2\}$ ,

$$\theta_{i,\varepsilon}^* := \arg \min_{\theta \in \Theta_{i,3}^\varepsilon} v_{i,\varepsilon}(\theta) \quad \text{and} \quad \theta_i^* := \arg \min_{\theta \in \Theta_{i,3}} v_i(\theta), \quad (2.27)$$

we focus on the convergence of  $\theta_{i,\varepsilon}^*$  toward  $\theta_i^*$  as  $\varepsilon$  tends to zero. For this purpose, we define, for  $q > 1$ , the set  $\Theta_q := \left\{ \theta \in \mathbb{R}^d : \int_{|x|>1} |x|^{2q} e^{-q\theta \cdot x} \nu(dx) < +\infty \right\}$ .

**Theorem 2.5.1.** *Let  $i \in \{1, 2\}$ . Suppose that  $x \mapsto F_i(x)$  is continuous, that is for the case  $i = 1$  the function  $F$  is continuous and for  $i = 2$  the function  $F$  is of class  $\mathcal{C}^1$ . Moreover, assume  $\mathbb{P}(F_i(L_T) \neq 0) > 0$ ,  $\mathbb{P}(F_i(L_T^\varepsilon) \neq 0) > 0$  for all  $\varepsilon > 0$  and there exists  $a > 1$  such that  $\mathbb{E}[F_i^a(L_T)]$  and  $\sup_{\varepsilon > 0} \mathbb{E}[F_i^a(L_T^\varepsilon)]$  are finite. Then,  $\theta \mapsto v_i(\theta)$  and  $\theta \mapsto v_{i,\varepsilon}(\theta) = \mathbb{E}[F_i(L_T^\varepsilon)e^{-\theta L_T^\varepsilon + T\kappa_\varepsilon(\theta)}]$  are  $\mathcal{C}^2$  strictly convex functions on  $\Theta_{i,3}$  and  $\Theta_{i,3}^\varepsilon$  respectively. Moreover, let  $K$  be a compact set such that  $K \subset \mathring{\Theta}_q$  with  $q > \frac{a}{a-1}$  and assume that the sequence  $(\theta_{i,\varepsilon}^*)_{\varepsilon > 0} \in K$ . Then,*

$$\theta_{i,\varepsilon}^* \longrightarrow \theta_i^* \in K, \quad \text{as } \varepsilon \rightarrow 0.$$

Now we turn to the stochastic algorithms needed to approximate the above variance optimizers. Let  $K$  be a compact convex subset of  $\Theta_1 \subset \mathbb{R}^d$  with  $\{0\} \in K$ . For fixed  $i \in \{1, 2\}$  and  $\theta_{i,0} \in K$ , we construct recursively the sequences of  $\mathbb{R}^d$ -valued random variables  $(\theta_{i,n})_{n \in \mathbb{N}}$  and  $(\theta_{i,\varepsilon,n})_{n \in \mathbb{N}}$  defined by the system

$$\begin{cases} \theta_{i,n+1} &= \Pi_K [\theta_{i,n} - \gamma_{n+1} H_i(\theta_{i,n}, L_{T,n+1})] \\ \theta_{i,\varepsilon,n+1} &= \Pi_K [\theta_{i,\varepsilon,n} - \gamma_{n+1} H_i(\theta_{i,\varepsilon,n}, L_{T,n+1}^\varepsilon)] \end{cases}, \quad (2.28)$$

where  $\Pi_K$  is the Euclidean projection onto the constraint set  $K$ ,  $H_1$  and  $H_2$  are given by  $H_i(\theta, L_T) = (T\nabla\kappa(\theta) - L_T)F_i(L_T) \exp(-\theta \cdot L_T + T\kappa(\theta))$ . We get the following result

**Theorem 2.5.2.** *Suppose assumptions of Theorem 2.5.1 hold. Moreover, let  $K$  be a compact set such that  $K \subset \mathring{\Theta}_{2a/(a-1)}$  for some  $a > 1$ . For  $i \in \{1, 2\}$*

- *if the unique  $\theta_i^* = \arg \min_{\theta \in \Theta_{i,3}} v_i(\theta)$  satisfies  $\theta_i^* \in K$  then  $\theta_{i,n} \xrightarrow[n \rightarrow +\infty]{a.s.} \theta_i^*$ ,*
- *if the unique  $\theta_{i,\varepsilon}^* = \arg \min_{\theta \in \Theta_{i,3}^\varepsilon} v_{i,\varepsilon}(\theta)$  satisfies  $\theta_{i,\varepsilon}^* \in K$  then  $\theta_{i,\varepsilon,n} \xrightarrow[n \rightarrow +\infty]{a.s.} \theta_{i,\varepsilon}^*$ .*

Clearly, according to the above results we have the following convergence result on our stochastic approximation, namely the constrained algorithm given by routine (2.28) satisfies

$$\text{For } i \in \{1, 2\}, \quad \lim_{n \rightarrow \infty} \theta_{i,\varepsilon,n} = \lim_{\varepsilon \rightarrow 0} (\lim_{n \rightarrow \infty} \theta_{i,\varepsilon,n}) = \lim_{n \rightarrow \infty} (\lim_{\varepsilon \rightarrow 0} \theta_{i,\varepsilon,n}) = \theta_i^*, \quad \mathbb{P}\text{-a.s.} \quad (2.29)$$

This last property is the essential key tool needed to prove the following central limit theorem on the adaptive version of the SR method (ISSR) that approximates our initial quantity of interest  $\mathbb{E}F(L_T)$  by

$$\begin{aligned} Q_\varepsilon^{\text{ISSR}} &:= \frac{1}{N_1} \sum_{k=1}^{N_1} F(L_{T,k}^{\varepsilon^\beta, \theta_{1,k-1}^{\varepsilon^\beta}}) e^{-\theta_{1,k-1}^{\varepsilon^\beta} \cdot L_{T,k}^{\varepsilon^\beta, \theta_{1,k-1}^{\varepsilon^\beta}} + T\kappa_{\varepsilon^\beta}(\theta_{1,k-1}^{\varepsilon^\beta})} \\ &\quad + \frac{1}{N_2} \sum_{k=1}^{N_2} \left( F(L_{T,k}^{\varepsilon, \theta_{2,k-1}^{\varepsilon}}) - F(L_{T,k}^{\varepsilon^\beta, \theta_{2,k-1}^{\varepsilon^\beta}}) \right) e^{-\theta_{2,k-1}^{\varepsilon} \cdot L_{T,k}^{\varepsilon, \theta_{2,k-1}^{\varepsilon}} + T\kappa_\varepsilon(\theta_{2,k-1}^{\varepsilon})}. \end{aligned} \quad (2.30)$$

**Theorem 2.5.3.** *Let  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  be a  $\mathcal{C}^1$  function satisfying assumption  $(\text{WE}_{v_\varepsilon})$  and s.t.  $\sup_{0 < \varepsilon \leq 1} \mathbb{E}F^{2a}(L_T^\varepsilon)$  and  $\sup_{0 < \varepsilon \leq 1} \mathbb{E}|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))|^{2a}$  are finite, for  $a > 1$ . Suppose also that the following assumptions are satisfied.*

H1. Condition (2.22) holds and  $\exists \Sigma$  a definite positive matrix s.t.  $\lim_{\varepsilon \rightarrow 0} \sigma^{-2}(\varepsilon) \Sigma_\varepsilon = \Sigma$ .

H2. For  $0 < \beta < 1$ , we have  $\lim_{\varepsilon \rightarrow 0} \sigma(\varepsilon) \sigma^{-1}(\varepsilon^\beta) = 0$  and  $\lim_{\varepsilon \rightarrow 0} v_\varepsilon \sigma^{-1}(\varepsilon^\beta) = 0$ .

Moreover, assume that  $\text{Leb}(\Theta_q) > 0$  with  $q > a/(a-1)$  and, for  $i \in \{1, 2\}$ , there exists a double indexed family  $(\theta_{i,k}^\varepsilon)_{k \in \mathbb{N}, \varepsilon > 0}$  satisfying (2.29) and belonging to some compact subset  $K_i \subset \dot{\Theta}_q$ . If we choose  $N_1 = v_\varepsilon^{-2}$  and  $N_2 = v_\varepsilon^{-2} \sigma^2(\varepsilon^\beta)$ , then

$$v_\varepsilon^{-1} (Q_\varepsilon^{\text{ISSR}} - \mathbb{E}F(L_T)) \implies \mathcal{N}(C_F, \sigma^2), \quad \text{as } \varepsilon \rightarrow 0,$$

where  $\sigma^2 = T \mathbb{E} [(F^2(L_T) + \nabla F(L_T) \cdot \Sigma \nabla F(L_T)) e^{-\theta^* \cdot L_T + T \kappa(\theta^*)}]$ .

We turn to illustrate the efficiency of the ISSR method that reduces at the same time the variance and the computational effort associated to the effective computation of option prices when the underlying asset process follows a 2-dimensional exponential pure jump CGMY model. Let us recall that a one-dimensional CGMY process has a generating triplet  $(0, 0, \nu)$  where for  $C > 0, G > 0, M > 0$  and  $Y < 2$

$$\nu(dx) = C \frac{e^{-Mx}}{x^{1+Y}} \mathbf{1}_{x>0} dx + \frac{C e^{-G|x|}}{|x|^{1+Y}} \mathbf{1}_{x<0} dx. \quad (2.31)$$

We focus on the computation of a price of the form  $e^{-rT} \mathbb{E}F(S_T^1, S_T^2)$ , where  $F(x, y) = (x + y - \text{Strike})_+$  and the couple  $(S_t^1, S_t^2)_{0 \leq t \leq T}$  denotes the underlying asset process. Here, we choose  $(S_t^1, S_t^2) = (S_0 e^{rt+L_t^1}, S_0 e^{rt+L_t^2})$ , where  $(L_t^1)_{0 \leq t \leq T}$  and  $(L_t^2)_{0 \leq t \leq T}$  are two independent CGMY processes with generating triplets  $(\gamma_1, 0, \nu_1)$  and  $(\gamma_2, 0, \nu_2)$ . In this setting there is no available explicit formula. The superiority of the ISSR method over the classical MC approach in the setting of Lévy processes is highlighted by Figure 2.3.

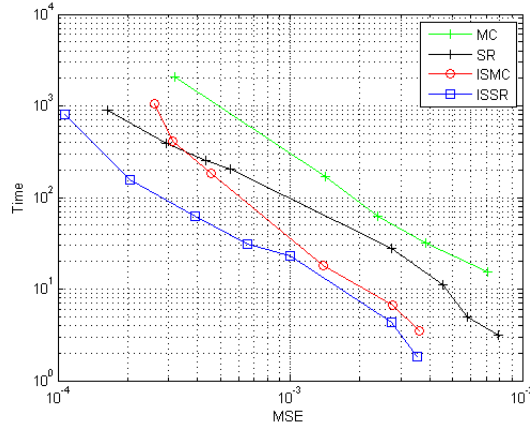


Figure 2.3: CPU time versus MSE in the two-dimensional setting.

## 2.6 Perspectives

In the PhD project of Trâm Ngô, that we co-supervise with Mohamed Ben Alaya, we also plan to study the antithetic Multilevel Monte Carlo in [GS14] when coupled with stochastic algorithms to approximate the variance optimizer of the corresponding importance

sampling weight. Proving a central limit theorem on the simple antithetic Multilevel Monte Carlo is a quite challenging problem on its own. We also hope to apply these techniques to the simulation of rare events. The analysis and simulations of rare events are crucial to prevent the worst by analyzing the cascade of events leading to the extremes. Using a new algorithm based on the combination between the Multilevel Monte Carlo method and the importance sampling techniques explained above will be of great interest in this setting, since the payoff function is less regular. It seems that the use of Malliavin calculus techniques will be of great help.

## Part II

# Statistics for financial models: estimation from continuous and discrete time observations



## Chapter 3

# Statistical properties for continuously observed financial models with and without jumps

### 3.1 Introduction and motivation

During the past decades, inference for diffusion models has become one of the core areas in statistical sciences. The basic statistical properties are well-summarized in Lipster and Shiriyayev [LS77] and Kutoyants [Kut04] with continuous observations and in Kessler and Sørensen [KLSr12] with discrete ones. The setting of continuously observed diffusions is rather theoretical, since the real data are discrete time observations. Nevertheless, if the process is observed at high frequency with a negligible discretization error then the statistical properties obtained for the continuously observed model remain valid for discrete time observations too. Most of the existing results dealing with this topic concern the case of ergodic diffusions with coefficients satisfying the Lipschitz and linearity growth conditions. In the literature, only few results can be found for non-ergodic diffusions or diffusions with non-regular coefficients such as the Cox Ingersoll Ross (CIR), Heston or Wishart processes. We also point out that only few results are available for parameter estimation for jump processes with non regular coefficients and in non-ergodic regimes.

To our knowledge, one of the first papers having studied the problem of drift parameter estimation in the one-dimensional CIR model is that of Fournié and Talay [FT91]. They have obtained a nice explicit formula of the maximum likelihood estimator (MLE) of the drift parameters  $\theta := (a, b)$  and have established its asymptotic normality only in the ergodic case, that is the subcritical case in terms of continuous branching processes. Afterward, Overbeck [Ove98] considers the CIR model in the context of a continuous branching process with immigration rate  $a > 0$  (CBI) and gives more detailed results by including singular subdomains. By establishing all three properties LAN (local asymptotic normality), LAMN (local asymptotic mixed normality) and LAQ (local asymptotic quadraticity), he obtains consistency and asymptotic normality for the MLE for different sub-models. Simultaneously, Overbeck and Ryden [OR97] examine the same problem from the discrete point of view. Note that, in practice, the diffusion parameter  $\sigma$  is usually assumed to be known and one can estimate it separately using the quadratic variation of the process  $X$ . For the jump-type CIR processes, Mai [Mai12] studied the



asymptotic normality of the MLE also in the subcritical case. Yet, this latter property of the MLE is not difficult to obtain in the subcritical case, since it is based on the classical martingale central limit theorem. Otherwise, for the critical and supercritical cases this argument is no more valid. The content of this chapter follows from :

- the results contained in [BAK12], [BAK13], in which we prove original limit theorems on the drift parameters MLE of the continuously observed CIR process including the critical and subcritical cases and provide sufficient conditions so that these limit theorems can be easily carried out for the discretely observed process.
- the results of [AKR16], in which we extend the above results to the setting of matrix Wishart processes providing asymptotic behavior and local asymptotic properties of the associated drift parameters, in the ergodic and several non-ergodic cases.
- the results of [BBKP15] and [BBKP17], in which we study the drift parameters MLE properties for jump-type Heston and CIR models.

## 3.2 Maximum likelihood estimation for CIR processes

The Cox-Ingersoll-Ross (CIR) process is widely used in mathematical finance to model the evolution of short-term interest rates. It is also used in the valuation of interest rate derivatives. It was introduced by Cox, Ingersoll and Ross [CIR85] as solution to the stochastic differential equation (SDE)

$$dX_t = (a - bX_t)dt + \sqrt{2\sigma|X_t|}dW_t, \quad (3.1)$$

where  $X_0 = x > 0$ ,  $a > 0$ ,  $b \in \mathbb{R}$ ,  $\sigma > 0$  and  $(W_t)_{t \geq 0}$  is a standard Brownian motion. This SDE has a unique non-negative strong solution  $(X_t)_{t \geq 0}$  (see Ikeda and Watanabe [IW81], p. 221). In the particular case  $b = 0$  and  $\sigma = 2$ , we recover the square of a  $a$ -dimensional Bessel process starting at  $x$ . For extensive studies on Bessel processes we refer to Revuz and Yor [RY99] and Pitman and Yor [PY81, PY82]. The behavior of the CIR process  $X$  mainly depends on the sign of  $b$ . Indeed, in the case  $b > 0$ , there exists a unique stationary distribution, say  $\pi$ , of  $X$  and the stationary CIR processes enjoy the ergodic property that is: for all  $h \in L^1(\pi)$ ,  $\frac{1}{t} \int_0^t h(X_s)ds$  converges almost surely to  $\int_{\mathbb{R}} h(x)\pi(dx)$ . In the case  $a \geq \sigma$ , the CIR process  $X$  stays strictly positive; for  $0 < a < \sigma$ , it hits 0 with probability  $p \in ]0, 1[$  if  $b < 0$  and almost surely if  $b \geq 0$ , the state 0 is instantaneously reflecting (see e.g. Göing-Jaeschke and Yor [GJY03] for more details).

Based on Lie symmetry methods, Craddock and Lenox [CL09] give an explicit form of the fundamental solution  $(t, x, y) \mapsto p(t, x, y)$  of the PDE  $u_t = \sigma x u_{xx} + au_x - (\frac{\mu}{x} + \lambda x)u$ ,  $\lambda > 0, \mu > 0$ . In [BAK12] and [BAK13], we take advantage of this result to compute explicitly the moment generating-Laplace transform of the quadruplet  $(\log X_t, X_t, \int_0^t X_s ds, \int_0^t \frac{ds}{X_s})$ , since we have that  $\mathbb{E}_x \left( X_t^\eta e^{-\rho X_t - \lambda \int_0^t X_s ds - \mu \int_0^t \frac{ds}{X_s}} \right) = \int_0^\infty y^\eta e^{-\rho y} p(t, x, y) dy$ . In more details, we prove

1. (**Case**  $b = 0$ ). For  $\rho \geq 0, \lambda \geq 0, \mu > 0$  and  $\eta \in ] -k - \frac{\nu}{2} - \frac{1}{2}, +\infty[$ , we have

$$\begin{aligned} \mathbb{E}_x \left( X_t^\eta e^{-\rho X_t - \lambda \int_0^t X_s ds - \mu \int_0^t \frac{ds}{X_s}} \right) &= \frac{\Gamma(\eta + k + \frac{\nu}{2} + \frac{1}{2})}{\Gamma(\nu + 1)} x^\eta \exp \left( -\frac{\sqrt{\sigma \lambda} x}{\sigma} \coth(\sqrt{\sigma \lambda} t) \right) \\ &\times \left( \frac{\sqrt{\sigma \lambda} x}{\sigma \sinh(\sqrt{\sigma \lambda} t)} \right)^{\frac{\nu}{2} + \frac{1}{2} - k - \eta} \left( (\sqrt{\sigma \rho} / \sqrt{\lambda}) \sinh(\sqrt{\sigma \lambda} t) + \cosh(\sqrt{\sigma \lambda} t) \right)^{-\frac{\nu}{2} - \frac{1}{2} - k - \eta} \\ &\times {}_1F_1 \left( \eta + k + \frac{\nu}{2} + \frac{1}{2}, \nu + 1, \frac{\sqrt{\sigma \lambda} x}{\sigma \sinh(\sqrt{\sigma \lambda} t) \left( (\sqrt{\sigma \rho} / \sqrt{\lambda}) \sinh(\sqrt{\sigma \lambda} t) + \cosh(\sqrt{\sigma \lambda} t) \right)} \right), \end{aligned} \quad (3.2)$$

where  $k = \frac{a}{2\sigma}$ ,  $\nu = \frac{1}{\sigma} \sqrt{(a - \sigma)^2 + 4\mu\sigma}$  and  ${}_1F_1$  is the confluent hypergeometric function defined by  ${}_1F_1(u, v, z) = \sum_{n=0}^{\infty} \frac{u_n z^n}{v_n n!}$ , with  $u_0 = v_0 = 1$ , and for  $n \geq 1$ ,  $u_n = \prod_{k=0}^{n-1} (u + k)$  and  $v_n = \prod_{k=0}^{n-1} (v + k)$ .

2. (**Case**  $b \neq 0$ ). For  $\rho \geq 0, \lambda \geq 0$  and  $\mu > 0$ , we have

$$\begin{aligned} \mathbb{E}_x \left( e^{-\rho X_t - \lambda \int_0^t X_s ds - \mu \int_0^t \frac{ds}{X_s}} \right) &= \frac{\Gamma(k + \frac{\nu}{2} + \frac{1}{2})}{\Gamma(\nu + 1)} \exp \left( \frac{b}{2\sigma} (at + x) - \frac{Ax}{2\sigma} \coth(At/2) \right) \\ &\times \left( \frac{Ax}{2\sigma \sinh(At/2)} \right)^{\frac{\nu}{2} + \frac{1}{2} - k} \left( \frac{2\sigma\rho + b}{A} \sinh(At/2) + \cosh(At/2) \right)^{-\frac{\nu}{2} - \frac{1}{2} - k} \\ &\times {}_1F_1 \left( k + \frac{\nu}{2} + \frac{1}{2}, \nu + 1, \frac{A^2 x}{2\sigma \sinh(At/2) \left( (2\sigma\rho + b) \sinh(At/2) + A \cosh(At/2) \right)} \right), \end{aligned} \quad (3.3)$$

where  $k = \frac{a}{2\sigma}$ ,  $A = \sqrt{b^2 + 4\sigma\lambda}$  and  $\nu = \frac{1}{\sigma} \sqrt{(a - \sigma)^2 + 4\mu\sigma}$ .

These explicit moment generating-Laplace transforms are crucial for getting a full precise description of the asymptotic behavior of the quadruplet  $(\log X_T, X_T, \int_0^T X_s ds, \int_0^T \frac{ds}{X_s})$  that naturally appears in the error of the MLE  $\hat{\theta} := (\hat{a}, \hat{b})$  of the drift CIR parameters  $\theta = (a, b)$ , which is defined only when  $a \geq \sigma$  and given by

$$\hat{\theta}_T - \theta = \begin{cases} \hat{a}_T - a &= \frac{\left( \log X_T - \log x + (\sigma - a) \int_0^T \frac{ds}{X_s} \right) \int_0^T X_s ds - T (X_T - x - aT)}{\int_0^T \frac{ds}{X_s} \int_0^T X_s ds - T^2} \\ \hat{b}_T - b &= \frac{T \left( \log X_T - \log x + bT + \sigma \int_0^T \frac{ds}{X_s} \right) - \left( X_T - x + b \int_0^T X_s ds \right) \int_0^T \frac{ds}{X_s}}{\int_0^T \frac{ds}{X_s} \int_0^T X_s ds - T^2}. \end{cases}$$

**Subcritical case** ( $b > 0$ ). In this case, we recover easily the asymptotic normality of the MLE thanks to the classical central limit theorem for continuous-time martingales.

For  $a > \sigma$ ,  $\mathcal{L}_\theta \left\{ \sqrt{T}(\hat{\theta}_T - \theta) \right\} \implies \mathcal{N}_{\mathbb{R}^2} \left( 0, 2\sigma C^{-1} \right)$ , as  $T \rightarrow \infty$ , with  $C = \begin{pmatrix} \frac{b}{a-\sigma} & -1 \\ -1 & \frac{a}{b} \end{pmatrix}$ .

However, all the following limit theorems cannot be obtained using this same argument and are derived thanks to (3.2) and (3.3). In the subcritical case, we prove that

$$\text{For } a = \sigma, \quad \mathcal{L}_\theta \left\{ \text{diag}(T, \sqrt{T})(\hat{\theta}_T - \theta) \right\} \Longrightarrow \left( \frac{b}{\tau_2}, \sqrt{2b}G \right), \text{ as } T \rightarrow \infty,$$

where  $G$  is a standard normal random variable independent of  $\tau_2$  the hitting time associated with Brownian motion  $\tau_2 = \inf\{t > 0 : W_t = \frac{b}{\sqrt{2\sigma}}\}$ .

**Critical case** ( $b = 0$ ). We prove that

$$\text{For } a = \sigma, \quad \mathcal{L}_\theta \left\{ \text{diag}(\log T, T)(\hat{\theta}_T - \theta) \right\} \Longrightarrow \left( \frac{1}{\tau_1}, \frac{a - R_1}{I_1} \right), \text{ as } T \rightarrow \infty,$$

where  $(R_t)$  is the CIR process, starting from 0, solution to

$$dX_t =adt + \sqrt{2\sigma X_t}dW_t, \tag{3.4}$$

$I_t = \int_0^t R_s ds$ , and  $\tau_1$  is the hitting time associated with Brownian motion  $\tau_1 = \inf\{t > 0 : W_t = \frac{1}{\sqrt{2\sigma}}\}$ . The couple  $(R_1, I_1)$  and the random time  $\tau_1$  are independent.

$$\text{For } a > \sigma, \quad \mathcal{L}_\theta \left\{ \text{diag}(\sqrt{\log T}, T)(\hat{\theta}_T - \theta) \right\} \Longrightarrow \left( \sqrt{2\sigma(a - \sigma)}G, \frac{a - R_1}{I_1} \right), \text{ as } T \rightarrow \infty,$$

where  $(R_1, I_1)$  is defined in the previous case,  $G$  is a standard normal random variable independent of  $(R_1, I_1)$ .

**Supercritical case** ( $b < 0$ ). For  $a \geq \sigma$ : the MLE estimator  $\hat{\theta}_T$  is not consistent, since in this specific case the process  $\left(\int_0^t \frac{ds}{X_s}\right)_{t \geq 0}$  is an increasing process converging to a finite random variable without any normalization. Nevertheless, if we assume that the parameter  $a$  is known, then the MLE  $\hat{b}_T$  in this case simplifies to  $\hat{b}_T = (aT + x - X_T) / \int_0^T X_s ds$ . In [BAK12], we prove that

$$\mathcal{L}_b \left\{ e^{-bT/2}(\hat{b}_T - b) \right\} \Longrightarrow \frac{G}{R}, \text{ as } T \rightarrow \infty,$$

where  $(G, R)$  is a couple of random variables characterized with its joint moment generating-Laplace transform. For  $\lambda \in \mathbb{R}$  and  $\mu \geq 0$ ,

$$\mathbb{E} \left( e^{\lambda G - \mu R} \right) = \left( \frac{b}{\mu\sigma/b + b} \right)^{\frac{\mu}{\sigma}} \exp \left( x \frac{\sigma\lambda^2/b + \mu}{\mu\sigma/b + b} \right).$$

Therefore,  $G$  and  $R$  are correlated,  $G$  is normal and  $R$  has the same distribution as  $t_0 \tilde{R}_{t_0}$ ,  $t_0 = -1/b$ , where  $(\tilde{R}_t)_{t \geq 0}$  is the CIR process, starting from  $x$ , solution to (3.1).

### 3.3 Maximum likelihood estimation for Wishart processes

The problem of estimating the parameters of the underlying Wishart process may be of big interest for practical purposes and especially in finance, where the data are particularly consistent and plentiful. Even though these processes were originally considered by Bru [Bru87] to model some biological data they are currently widely used in financial models in order to describe the evolution of the dependence between assets. We point to Alfonsi [Alf15] for detailed results on the Wishart process and its practical use.

Let  $d \in \mathbb{N}^*$  denote the dimension,  $\mathcal{M}_d$  be the set of real  $d$ -square matrices,  $\mathcal{S}_d^+$  (resp.  $\mathcal{S}_d^{+,*}$ ) be the subset of positive semidefinite (resp. definite) matrices,  $\mathcal{S}_d$  (resp.  $\mathcal{A}_d$ ) the subset of symmetric (resp. antisymmetric) matrices. Wishart processes are defined by the following SDE

$$\begin{cases} dX_t = [\alpha a^\top a + bX_t + X_t b^\top] dt + \sqrt{X_t} dW_t a + a^\top dW_t^\top \sqrt{X_t}, & t > 0 \\ X_0 = x \in \mathcal{S}_d^+, \end{cases} \quad (3.5)$$

where  $\alpha \geq d - 1$ ,  $a \in \mathcal{M}_d$ ,  $b \in \mathcal{M}_d$  and  $(W_t)_{t \geq 0}$  denotes a  $d$ -square matrix made of independent Brownian motions. We recall that for  $x \in \mathcal{S}_d^+$ ,  $\sqrt{x}$  is the unique matrix in  $\mathcal{S}_d^+$  such that  $\sqrt{x}^2 = x$ . It is shown by Bru [Bru91] and Cuchiero et al. [CFMT11] in a more general affine setting that the SDE (3.5) has a unique strong solution when  $\alpha \geq d + 1$  and a unique weak solution when  $\alpha \geq d - 1$ . Besides, we have  $X_t \in \mathcal{S}_d^{+,*}$  for any  $t \geq 0$  when  $x \in \mathcal{S}_d^{+,*}$  and  $\alpha \geq d + 1$ . In this part, we will denote by  $WIS_d(x, \alpha, b, a)$  the law of  $(X_t, t \geq 0)$ . In dimension  $d = 1$ , Wishart processes are known as CIR processes in the literature. It is worth recalling that the law of  $X$  only depends on  $a$  through  $a^\top a$  since we have  $WIS_d(x, \alpha, b, a) = WIS_d(x, \alpha, b, \sqrt{a^\top a})$ . Therefore, the parameters to estimate are  $\alpha$ ,  $b$ , since the diffusion parameter  $a^\top a$  can easily be estimated using the explicit expression of the quadratic variation of (3.5). Since the process defined by  $Y_t = (a^\top)^{-1} X_t a^{-1}$  follows the law  $WIS_d((a^\top)^{-1} x a^{-1}, \alpha, (a^\top)^{-1} b a^\top, I_d)$ , it is sufficient to focus on the estimation of  $\theta = (b, \alpha)$  when  $a = I_d$ , which we consider now. To do so, let us denote by  $\mathbb{P}_\theta$  the original probability measure under which  $X$  satisfies

$$dX_t = \left[ \alpha I_d + bX_t + X_t b^\top \right] dt + \sqrt{X_t} dW_t + dW_t^\top \sqrt{X_t}, \text{ where } \alpha \geq d + 1 \text{ and } x \in \mathcal{S}_d^{+,*}. \quad (3.6)$$

We consider  $\alpha_0 \geq d + 1$  and set  $\theta_0 = (\alpha_0, 0)$ . We will assume, for the joint estimation of  $\alpha$  and  $b$ , that  $\alpha \geq d + 1$ . This restriction is needed to ensure that  $X_t \in \mathcal{S}_d^{+,*}$  for any  $t > 0$ . Thanks to this assumption, we know by Theorem 4.1 in Mayerhofer [May12] that

$$\frac{d\mathbb{P}_{\theta_0, T}}{d\mathbb{P}_{\theta, T}} = \exp \left( \int_0^T \text{Tr}[H_s dW_s] - \frac{1}{2} \int_0^T \text{Tr}[H_s H_s^\top] ds \right) \text{ with } H_t = \frac{\alpha_0 - \alpha}{2} (\sqrt{X_t})^{-1} - b \sqrt{X_t}$$

defines a probability measure under which  $\tilde{W}_t = W_t - \int_0^t H_s^\top ds$  is a  $d \times d$ -Brownian motion, where  $\mathbb{P}_{\theta, T}$  is the restriction of  $\mathbb{P}_\theta$  to the  $\sigma$ -algebra  $\sigma(W_s, s \in [0, T])$ .

Here, we draw attention to an unusual phenomena in statistical inference of stochastic differential equations that we faced in this study. Actually, unlike the one-dimensional setting (see Section 3.2), we prove in [AKR16] that the above Radon-Nikodym derivative might not coincide with the likelihood ratio needed to derive the MLE of  $\theta$ . In more details, according to Lipster and Shiryaev [LS01], the likelihood ratio is given by

$L_T^{\theta, \theta_0} = (\mathbb{E}[d\mathbb{P}_{\theta_0, T}/d\mathbb{P}_{\theta, T} | \mathcal{F}_T^X])^{-1}$ , where  $(\mathcal{F}_t^X)_{t \geq 0}$  denotes the filtration generated by the process  $X$ . We prove the following result.

**Proposition 3.3.1.** *For  $X \in \mathcal{S}_d^{+, *}$ , we have*

1. *The linear application  $\mathcal{L}_X : \mathcal{S}_d \rightarrow \mathcal{S}_d$  defined by  $\mathcal{L}_X(Y) = XY + YX$  is invertible and the likelihood of  $(X_t, t \in [0, T])$  is given by*

$$\begin{aligned} L_T^{\theta, \theta_0} &= \left( \frac{\det[X_T]}{\det[x]} \right)^{\frac{\alpha - \alpha_0}{4}} \times \exp\left( -\frac{\alpha - \alpha_0}{4} \left( \frac{\alpha + \alpha_0}{2} - 1 - d \right) \int_0^T \text{Tr}[X_s^{-1}] ds \right. \\ &\quad \left. + \frac{1}{2} \int_0^T \text{Tr} \left[ \mathcal{L}_{X_t}^{-1} \left( bX_t + X_t b^\top \right) dX_t \right] - \frac{\alpha T}{2} \text{Tr}[b] \right. \\ &\quad \left. - \frac{1}{4} \int_0^T \text{Tr} \left[ \mathcal{L}_{X_t}^{-1} \left( bX_t + X_t b^\top \right) \left( bX_t + X_t b^\top \right) \right] dt \right). \end{aligned} \quad (3.7)$$

2. *The Radon-Nikodym derivative  $\frac{d\mathbb{P}_{\theta_0, T}}{d\mathbb{P}_{\theta, T}} \in \mathcal{F}_T^X$  if, and only if  $b \in \mathcal{S}_d$  in which case the above likelihood ratio simplifies to*

$$\begin{aligned} L_T^{\theta, \theta_0} &= \left( \frac{\det[X_T]}{\det[x]} \right)^{\frac{\alpha - \alpha_0}{4}} \times \exp\left( \frac{\text{Tr}[bX_T] - \text{Tr}[bx]}{2} - \frac{1}{2} \int_0^T \text{Tr}[b^2 X_s] ds \right. \\ &\quad \left. - \frac{\alpha - \alpha_0}{4} \left( \frac{\alpha + \alpha_0}{2} - 1 - d \right) \int_0^T \text{Tr}[X_s^{-1}] ds - \frac{\alpha T}{2} \text{Tr}[b] \right). \end{aligned} \quad (3.8)$$

For brevity's sake, we only give the results for  $b \in \mathcal{S}_d$ . For the general case  $b \in \mathcal{M}_d$ , we proved the asymptotic normality of the global MLE of  $\theta = (\alpha, b)$  when  $\alpha > d + 1$  and the asymptotic normality of the MLE of  $b$  when the parameter  $\alpha \geq d - 1$  is supposed to be known. For these latter results, the rate of convergence is the standard  $\sqrt{T}$ . To get explicit formulas for the MLE error, we introduce for  $X \in \mathcal{S}_d$  and  $a \in \mathbb{R}$  the linear applications

$$\begin{aligned} \mathcal{L}_X : \mathcal{S}_d &\rightarrow \mathcal{S}_d & \text{and } \mathcal{L}_{X, a} : \mathcal{S}_d &\rightarrow \mathcal{S}_d \\ Y &\mapsto YX + XY & Y &\mapsto YX + XY - 2a\text{Tr}[Y]I_d \end{aligned} \quad (3.9)$$

together with the following shorthand notations

$$R_T := \int_0^T X_s ds, \quad Q_T := \left( \int_0^T \text{Tr}[X_s^{-1}] ds \right)^{-1}, \quad Z_T := \log \left( \frac{\det[X_T]}{\det[x]} \right). \quad (3.10)$$

Note that  $Q_T$  and  $Z_T$  are defined only for  $\alpha \geq d + 1$ , while  $R_T$  is defined for  $\alpha \geq d - 1$  and belongs almost surely to  $\mathcal{S}_d^{+, *}$ . Then, for  $M_t := \int_0^t \sqrt{X_s} dW_s + \int_0^t dW_s^\top \sqrt{X_s}$  and  $N_t := \int_0^t \text{Tr}[(\sqrt{X_s})^{-1} dW_s]$ , we provide an explicit expression for the MLE error given by

$$\begin{cases} \sqrt{T}(\hat{\alpha}_T - \alpha) &= 2TQ_T \frac{N_T}{\sqrt{T}} - 2TQ_T \text{Tr} \left[ \mathcal{L}_{\frac{R_T}{T}, TQ_T}^{-1} \left( \frac{M_T}{\sqrt{T}} - 2TQ_T I_d \frac{N_T}{\sqrt{T}} \right) \right] \\ \sqrt{T}(\hat{b}_T - b) &= \mathcal{L}_{\frac{R_T}{T}, TQ_T}^{-1} \left( \frac{M_T}{\sqrt{T}} - 2TQ_T I_d \frac{N_T}{\sqrt{T}} \right). \end{cases}$$

We also provide an explicit Laplace transform crucial to the study of the asymptotic behavior of the MLE of  $\theta = (b, \alpha)$ , in both ergodic and non-ergodic cases.

**Theorem 3.3.1.** *Let  $\alpha \geq d-1$ ,  $x \in \mathcal{S}_d^+$ ,  $b \in \mathcal{S}_d$  and  $X \sim WIS_d(x, \alpha, b, I_d)$ . Let  $v, w \in \mathcal{S}_d$  be such that*

$$\exists m \in \mathcal{S}_d, \frac{v}{2} - mb - bm - 2m^2 \in \mathcal{S}_d^+ \text{ and } \frac{w}{2} + m \in \mathcal{S}_d^+. \quad (3.11)$$

*Then, we have for  $t \geq 0$*

$$\mathbb{E} \left[ \exp \left( -\frac{1}{2} \text{Tr} [wX_t + vR_t] \right) \right] = \frac{\exp \left( -\frac{\alpha}{2} \text{Tr} [b]t \right)}{\det[V_{v,w}(t)]^{\frac{\alpha}{2}}} \exp \left( -\frac{1}{2} \text{Tr} [(V'_{v,w}(t)V_{v,w}(t)^{-1} + b)x] \right),$$

*with  $V_{v,w}(t) = \left( \sum_{k=0}^{\infty} t^{2k+1} \frac{\tilde{v}^k}{(2k+1)!} \right) \tilde{w} + \sum_{k=0}^{\infty} t^{2k} \frac{\tilde{v}^k}{(2k)!}$ ,  $\tilde{v} = v + b^2$ , and  $\tilde{w} = w - b$ . If besides  $\tilde{v} = v + b^2 \in \mathcal{S}_d^{+,*}$ , we have  $V_{v,w}(t) = (\sqrt{\tilde{v}})^{-1} \sinh(\sqrt{\tilde{v}}t)\tilde{w} + \cosh(\sqrt{\tilde{v}}t)$  and then  $V'_{v,w}(t) = \cosh(\sqrt{\tilde{v}}t)\tilde{w} + \sinh(\sqrt{\tilde{v}}t)\sqrt{\tilde{v}}$ .*

This last result extends the recent findings of Gnoatto and Grasselli [GG14] and is clearly of independent interest.

**Subcritical case:**  $(-b \in \mathcal{S}_d^{+,*})$ .

When  $-b \in \mathcal{S}_d^{+,*}$ , the Wishart process  $X_t$  converges in law as  $t \rightarrow +\infty$  to the stationary law  $X_\infty \sim WIS_d(0, \alpha, 0, \sqrt{2q_\infty}; 1/2)$  with  $q_\infty = \int_0^\infty e^{2sb} ds$  for any starting point  $x \in \mathcal{S}_d^+$ . Therefore, this is the unique stationary law which is thus extremal and that it is then ergodic, see e.g. Pagès [Pag01], Annex A. Thus, we have

$$\frac{R_T}{T} \xrightarrow{a.s.} \bar{R}_\infty := \mathbb{E}_\theta(X_\infty), \quad \text{as } T \rightarrow +\infty. \quad (3.12)$$

Besides, when  $\alpha \geq d+1$ ,  $\bar{Q}_\infty = \frac{1}{\mathbb{E}_\theta(\text{Tr}[X_\infty^{-1}])}$  is finite and again, the ergodic Birkhoff's theorem gives

$$TQ_T \xrightarrow{a.s.} \bar{Q}_\infty = \frac{1}{\mathbb{E}_\theta(\text{Tr}[X_\infty^{-1}])}, \quad \text{as } T \rightarrow +\infty. \quad (3.13)$$

- For the case  $\alpha > d+1$ , we prove thanks to (3.12), (3.13) and the central limit theorem for martingales the asymptotic normality of the normalized MLE error  $\left( \sqrt{T}(\hat{b}_T - b, \hat{\alpha}_T - \alpha) \right)$  and characterize the limit distribution given by a centered Gaussian vector  $(\mathbf{G}, H)$  taking values in  $\mathcal{S}_d \times \mathbb{R}$ .
- However, for the case  $\alpha = d+1$ , the rate of convergence of the MLE of  $\alpha$  is even better as stated by the following theorem.

**Theorem 3.3.2.** *Assume  $-b \in \mathcal{S}_d^{+,*}$  and  $\alpha = d+1$ . Then, under  $\mathbb{P}_\theta$ , the couple  $\left( \sqrt{T}(\hat{b}_T - b), T(\hat{\alpha}_T - \alpha) \right)$  converges in law when  $T \rightarrow +\infty$  to  $\left( \mathbf{G}, -2\tau_{-\text{Tr}[b]}^{-1} \text{Tr}[b] \right)$ , where  $\tau_a = \inf\{t \geq 0, B_t = a\}$  with  $(B_t)_{t \geq 0}$  a given one-dimensional standard Brownian motion and  $\mathbf{G}$  is a Gaussian vector independent of  $B$  such that  $\mathbb{E}_\theta[\exp(\text{Tr}[c\mathbf{G}])] = \exp\left(\text{Tr}[c\mathcal{L}_{\bar{R}_\infty}^{-1}(c)]\right)$ ,  $c \in \mathcal{S}_d$ .*

**Critical case:** ( $b = 0$ ).

In this case, we combine together Proposition 3.3.1 and several subtle properties of the Wishart process to prove the following result.

**Theorem 3.3.3.** *Assume  $b = 0$ .*

- If  $\alpha > d + 1$ , then,  $(T(\hat{b}_T - b), \sqrt{\log(T)}(\hat{\alpha}_T - \alpha))$  converges in law under  $\mathbb{P}_\theta$ , when  $T \rightarrow +\infty$  to

$$\left( \mathcal{L}_{R_1^0}^{-1}(X_1^0 - \alpha I_d), 2\sqrt{\frac{\alpha - (d+1)}{d}}G \right),$$

where  $X_t^0 = \alpha t I_d + \int_0^t \sqrt{X_s^0} dW_s + dW_s^\top \sqrt{X_s^0}$  is a Wishart process with the same parameters but starting from 0,  $R_t^0 = \int_0^t X_s^0 ds$  and  $G \sim \mathcal{N}(0, 1)$  is an independent standard Normal variable.

- If  $\alpha = d + 1$ , then,  $(T(\hat{b}_T - b), \log(T)(\hat{\alpha}_T - \alpha))$  converges in law under  $\mathbb{P}_\theta$ , when  $T \rightarrow +\infty$  to

$$\left( \mathcal{L}_{R_1^0}^{-1}(X_1^0 - \alpha I_d), \frac{4}{d\tau_1} \right),$$

where  $X_t^0 = \alpha t I_d + \int_0^t \sqrt{X_s^0} dW_s + dW_s^\top \sqrt{X_s^0}$  is a Wishart process with the same parameters but starting from 0,  $R_t^0 = \int_0^t X_s^0 ds$  and  $\tau_1 = \inf\{t \geq 0, B_t = 1\}$ , where  $B$  is a standard Brownian motion independent from  $W$ .

**Supercritical case: a specific choice** ( $b = b_0 I_d, b_0 > 0$ ).

Similarly, as for the one-dimensional case, for the supercritical case we treat only the special case of estimating  $b$  while  $\alpha$  is supposed to be known. In this case, the MLE of  $b$  is given by  $\hat{b}_T = \mathcal{L}_{R_T}^{-1}(X_T - x - \alpha T I_d)$ .

Obviously, the case  $b = b_0 I_d$  is very particular. One would like to consider more general non-ergodic cases or ideally to be able to state a general convergence result of  $\hat{b}_T$  towards  $b$  for any  $b \in \mathcal{S}_d$ . Such result seems to be difficult to obtain. The reason why we can handle the ergodic case and the non-ergodic case  $b = b_0 I_d$  with  $b_0 \geq 0$  is that the convergence of all the matrix terms occurs at the same speed, namely  $1/\sqrt{T}$  for the ergodic case,  $1/T$  for  $b = 0$  and  $e^{-b_0 T}$  when  $b_0 > 0$ . In the other cases, there is no such a simple scalar rescaling. Heuristically, there may be different speeds of convergence that are difficult to disentangle because of the different matrix products. So we prove the following result.

**Theorem 3.3.4.** *Assume that  $\alpha \geq d - 1$ . Then, as  $T \rightarrow +\infty$ ,  $\exp(b_0 T)(\hat{b}_T - b)$  converges in law under  $\mathbb{P}_\theta$  to  $\mathcal{L}_X^{-1}\left(\sqrt{X}\tilde{G} + \tilde{G}\sqrt{X}\right)$  where  $X \sim WIS_d\left(\frac{x}{2b_0}, \alpha, 0, I_d; \frac{1}{4b_0^2}\right)$  and  $\tilde{G}$  is an independent  $d$ -square matrix whose elements are independent standard Normal variables.*

Though limited to some non-ergodic cases, we however recover and extend results obtained by [BAK12] and [BAK13] for the one-dimensional CIR process. We also prove local asymptotic properties (LAN and LAQ) of the derived MLE in the different cases.

### 3.4 Statistical inference for jump-type square root models

As pointed out in Section 3.1, there are only few studies examining MLE estimators for jump diffusions with non-regular coefficients and almost none dealing with this problem in non-ergodic regimes. We address this question in [BBKP15] and [BBKP17].

#### 3.4.1 Maximum likelihood estimation for stochastic volatility with jumps model (SVJ)

In [BBKP15], for  $\theta, \kappa, \sigma \in (0, \infty)$ ,  $\mu \in \mathbb{R}$  and  $\varrho \in (-1, 1)$ , we consider the jump-type Heston model given by

$$\begin{cases} dY_t = \kappa(\theta - Y_t) dt + \sigma\sqrt{Y_t} dW_t, Y_0 \geq 0 \\ dS_t = \mu S_t dt + S_t\sqrt{Y_t}(\varrho dW_t + \sqrt{1 - \varrho^2} dB_t) + S_{t-} dL_t, S_0 > 0, \end{cases} \quad (3.14)$$

where  $(L_t)_{t \in [0, \infty)}$  is a purely non-Gaussian Lévy process of not necessarily bounded variation, independent of  $(W_t, B_t)_{t \in [0, \infty)}$  with Lévy–Khintchine representation

$$\mathbb{E}(e^{iuL_1}) = \exp \left\{ i\gamma u + \int_{-1}^{\infty} (e^{iuz} - 1 - iuz\mathbb{1}_{[-1,1]}(z)) m(dz) \right\}, \quad u \in \mathbb{R}, \quad (3.15)$$

where  $\gamma \in \mathbb{R}$  and  $m$  is a Lévy measure concentrating on  $(-1, \infty)$  with  $m(\{0\}) = 0$ . We first prove that there is a (pathwise) unique strong solution  $(Y_t, S_t)_{t \in [0, \infty)}$  of the SDE (3.14) such that  $\mathbb{P}(Y_t \in [0, \infty)$  and  $S_t \in (0, \infty)$  for all  $t \in [0, \infty)) = 1$  and

$$S_t = S_0 \exp \left\{ \int_0^t \left( \mu - \frac{1}{2} Y_u \right) du + \int_0^t \sqrt{Y_u} (\varrho dW_u + \sqrt{1 - \varrho^2} dB_u) + L_t \right\} \times \prod_{u \in [0, t]} (1 + \Delta L_u) e^{-\Delta L_u}, \quad (3.16)$$

where for  $u \in [0, \infty)$ ,  $\Delta L_u := L_u - L_{u-}$ , and the above (possibly) infinite product is absolutely convergent. The above model (3.14) is quite popular in finance with the special choice of the Lévy process  $L$  as a compound Poisson process

$$N_t := \sum_{i=1}^{\pi_t} (e^{J_i} - 1), \quad t \in [0, \infty), \quad (3.17)$$

where  $(\pi_t)_{t \in [0, \infty)}$  is a Poisson process with intensity 1,  $(J_i)_{i \in \mathbb{N}}$  is a sequence of independent identically distributed random variables having no atom at zero and being independent of  $\pi$  as well. One can interpret  $J$  as the jump size of the log returns of the asset prices. When the distribution of  $J$  is chosen as a normal distribution, the above process is widely known in quantitative finance as the Bates model [Bat96]. Moreover, Bakshi et al. [BCC97] used this model for studying (European style) S&P500 options, e.g., they derived a practically implementable closed-form pricing formula and Broadie and Kaya [BK06] gave an exact simulation algorithm for this model.

Further, we prove that for all  $t \in [0, T]$ ,  $L_t$  and  $Y_t$  are measurable functions of  $(S_t)_{t \in [0, T]}$ . Then, as the continuous martingale part  $(\log S)_t^{\text{cont}}$  of  $\log S$  is  $(\log S)_t^{\text{cont}} =$



$\int_0^t \sqrt{Y_u} (\varrho dW_u + \sqrt{1 - \varrho^2} dB_u)$ ,  $t \in [0, \infty)$  and since for all  $T > 0$ ,

$$\begin{bmatrix} \sigma^2 & \varrho\sigma \\ \varrho\sigma & 1 \end{bmatrix} = \frac{1}{\int_0^T Y_s ds} \begin{bmatrix} \langle Y \rangle_T & \langle Y, (\log S)^{\text{cont}} \rangle_T \\ \langle Y, (\log S)^{\text{cont}} \rangle_T & \langle (\log S)^{\text{cont}} \rangle_T \end{bmatrix} =: \widehat{\Sigma}_T,$$

we deduce that  $\widehat{\Sigma}_T$  is a statistic, that is we can assume that parameters  $\sigma, \rho$  are known.

### Existence and uniqueness of MLE

Let us denote by  $\psi := (\theta, \kappa, \mu) \in (0, \infty)^2 \times \mathbb{R} =: \Psi$  the parameter to be estimated. At first, we rewrite the Heston model (3.14) in the form

$$\begin{bmatrix} dY_t \\ dS_t \end{bmatrix} = A(Y_t, S_t) H(\psi) dt + \Gamma(Y_t, S_t) \begin{bmatrix} dW_t \\ dB_t \end{bmatrix} + \begin{bmatrix} 0 \\ S_{t-} dL_t \end{bmatrix}, \quad t \in [0, \infty), \quad (3.18)$$

where the functions  $A : [0, \infty) \times (0, \infty) \rightarrow \mathbb{R}^{2 \times 3}$ ,  $\Gamma : [0, \infty) \times (0, \infty) \rightarrow \mathbb{R}^{2 \times 2}$  and  $H : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  are defined by

$$A(y, s) := \begin{bmatrix} 1 & -y & 0 \\ 0 & 0 & s \end{bmatrix}, \quad \Gamma(y, s) := \sqrt{y} \begin{bmatrix} \sigma & 0 \\ \varrho\sigma & \sqrt{1 - \varrho^2} s \end{bmatrix}, \quad H(x_1, x_2, x_3) := \begin{bmatrix} x_1 x_2 \\ x_2 \\ x_3 \end{bmatrix}$$

for  $(y, s) \in [0, \infty) \times (0, \infty)$  and  $(x_1, x_2, x_3) \in \mathbb{R}^3$ . Let us introduce the function  $\Sigma : [0, \infty) \times (0, \infty) \rightarrow \mathbb{R}^{2 \times 2}$  given by

$$\Sigma(y, s) := \Gamma(y, s) \Gamma(y, s)^\top = y \begin{bmatrix} \sigma^2 & \varrho\sigma s \\ \varrho\sigma s & s^2 \end{bmatrix}, \quad (y, s) \in [0, \infty) \times (0, \infty)$$

and the processes

$$\mathbf{G}_t := \int_0^t A(Y_u, S_u)^\top \Sigma(Y_u, S_u)^{-1} A(Y_u, S_u) du, \quad t \in [0, \infty),$$

and

$$\mathbf{f}_t := \int_0^t A(Y_{u-}, S_{u-})^\top \Sigma(Y_{u-}, S_{u-})^{-1} \begin{bmatrix} dY_u \\ dS_u - S_{u-} dL_u \end{bmatrix}, \quad t \in [0, \infty),$$

provided that  $\mathbb{P}(Y_t \in (0, \infty) \text{ for all } t \in [0, \infty)) = 1$ , which holds if  $\theta\kappa \in [\frac{\sigma^2}{2}, \infty)$ . By Jacod and Shiryaev [JS03, Theorem III.5.34], we prove that for  $\tilde{\psi} := (\tilde{\theta}, \tilde{\kappa}, \tilde{\mu}) \in \Psi$  with  $\theta\kappa, \tilde{\theta}\tilde{\kappa} \in [\frac{\sigma^2}{2}, \infty)$ , the probability measures  $\mathbb{P}_{\psi, T}$  and  $\mathbb{P}_{\tilde{\psi}, T}$  are absolutely continuous with respect to each other, and, under  $\mathbb{P}$ ,

$$\log \frac{d\mathbb{P}_{\psi, T}}{d\mathbb{P}_{\tilde{\psi}, T}}(\tilde{Y}, \tilde{S}) = (H(\psi) - H(\tilde{\psi}))^\top \tilde{\mathbf{f}}_T - \frac{1}{2} (H(\psi) - H(\tilde{\psi}))^\top \tilde{\mathbf{G}}_T (H(\psi) + H(\tilde{\psi})), \quad (3.19)$$

where  $\tilde{Y}, \tilde{S}, \tilde{\mathbf{G}}$  and  $\tilde{\mathbf{f}}$  are the processes corresponding to the parameter  $\tilde{\psi}$ .

## Asymptotic behavior of MLE

Taking advantage of the explicit expression of the above likelihood ratio (3.19), we prove that for all  $T \in (0, \infty)$ , there exists a unique MLE  $\hat{\boldsymbol{\psi}}_T = (\hat{\theta}_T, \hat{\kappa}_T, \hat{\mu}_T)^\top$  of  $\boldsymbol{\psi} = (\theta, \kappa, \mu)^\top$  on the set  $\mathbb{R}^3$  based on the observations  $(Y_t, S_t)_{t \in [0, T]}$  taking the form

$$\hat{\boldsymbol{\psi}}_T = \begin{bmatrix} \hat{\theta}_T \\ \hat{\kappa}_T \\ \hat{\mu}_T \end{bmatrix} = \begin{bmatrix} (\mathbf{G}_T^{-1} \mathbf{f}_T)_1 \\ (\mathbf{G}_T^{-1} \mathbf{f}_T)_2 \\ (\mathbf{G}_T^{-1} \mathbf{f}_T)_3 \end{bmatrix}, \quad (3.20)$$

provided that  $\mathbf{G}_T$  is strictly positive definite and  $(\mathbf{G}_T^{-1} \mathbf{f}_T)_2 \neq 0$ , which hold almost surely. Moreover, when  $\theta\kappa > \sigma^2/2$ , we prove strong consistency and asymptotic normality of  $\hat{\boldsymbol{\psi}}_T$ . However, when  $\theta\kappa = \sigma^2/2$ , we only get weak consistency as a consequence of the following result:

$$\begin{bmatrix} T^{1/2}(\hat{\theta}_T - \theta) \\ T^{1/2}(\hat{\kappa}_T - \kappa) \\ T(\hat{\mu}_T - \mu) \end{bmatrix} \xrightarrow{\mathcal{D}} \begin{bmatrix} -\frac{\sigma^2 \sqrt{1-\rho^2}}{\sqrt{2\kappa^3}} Z_1 \\ \sqrt{2(1-\rho^2)}\kappa Z_1 \\ \frac{\rho\sigma}{\kappa\tau} + \frac{\sigma\sqrt{1-\rho^2}}{\kappa\sqrt{\tau}} Z_2 \end{bmatrix} \quad \text{as } T \rightarrow \infty, \quad (3.21)$$

where  $\tau := \inf\{t \in [0, \infty) : \mathcal{W}_t = 1\}$  with a standard Wiener process  $(\mathcal{W}_t)_{t \in [0, \infty)}$ , and  $Z_1$  and  $Z_2$  are independent standard normally distributed random variables, independent from  $\tau$ .

### 3.4.2 Maximum likelihood estimation for the jump-type CIR model

In [BBKP17], we consider a jump-type CIR process driven by a standard Wiener process and a subordinator

$$dY_t = (a - bY_t) dt + \sigma\sqrt{Y_t} dW_t + dJ_t, \quad t \in [0, \infty), Y_0 \geq 0, \quad (3.22)$$

where  $a, \sigma \in [0, \infty)$ ,  $b \in \mathbb{R}$ ,  $(W_t)_{t \in [0, \infty)}$  is a standard Wiener process and  $(J_t)_{t \in [0, \infty)}$  is a subordinator with zero drift and a Lévy measure  $m$  concentrating on  $(0, \infty)$  such that

$$\int_0^\infty z m(dz) \in [0, \infty), \quad (3.23)$$

that is,  $\mathbb{E}(e^{uJ_t}) = \exp\{t \int_0^\infty (e^{uz} - 1)m(dz)\}$  for any  $t \in [0, \infty)$  and for any complex number  $u$  with  $\Re(u) \in (-\infty, 0]$ . We suppose that  $(W_t)_{t \in [0, \infty)}$  and  $(J_t)_{t \in [0, \infty)}$  are independent. Note that the subordinator  $J$  has sample paths of bounded variation on every compact time interval almost surely, see, e.g., Sato [Sat13, Theorem 21.9].

Under these assumptions we first prove that there is a (pathwise) unique strong solution of the SDE (3.22) such that  $\mathbb{P}(Y_t \in [0, \infty) \text{ for all } t \in [0, \infty)) = 1$ . If, in addition,  $Y_0 > 0$  or  $a > 0$ , then  $\mathbb{P}(\int_0^t Y_s ds > 0) = 1$  for all  $t > 0$ . Moreover, we prove that  $(Y_t)_{t \in [0, \infty)}$  is a special continuous state and continuous time branching process with immigration (CBI process) having branching mechanism  $R(u) = \frac{\sigma^2}{2}u^2 - bu$  and immigration mechanism  $F(u) = au + \int_0^\infty (e^{uz} - 1)m(dz)$ , for  $u \in \mathbb{C}$  with  $\Re(u) \leq 0$ .

The jump-type CIR process in (3.22) includes the so-called basic affine jump-diffusion (BAJD) as a special case, in which the drift takes the form  $\kappa(\theta - Y_t)$  with some  $\kappa \in (0, \infty)$

and  $\theta \in [0, \infty)$ , and the Lévy process  $(J_t)_{t \in [0, \infty)}$  is a compound Poisson process with exponentially distributed jump sizes, namely,  $m(dz) = c\lambda e^{-\lambda z} \mathbb{1}_{(0, \infty)}(z) dz$  with some constants  $c \in [0, \infty)$  and  $\lambda \in (0, \infty)$ . For describing the dynamics of default intensity, the BAJD was introduced by Duffie and Gârleanu [DN01]. Filipović [Fil01] and Keller-Ressel and Steiner [KRS08] used the BAJD as a short-rate model. We point out that, the volatility parameter  $\sigma$  can be estimated thanks to

$$\frac{1}{\frac{1}{n} \sum_{i=1}^{\lfloor nT \rfloor} Y_{\frac{i-1}{n}}} \left( \sum_{i=1}^{\lfloor nT \rfloor} (Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}})^2 - \sum_{u \in [0, T]} (\Delta Y_u)^2 \right) \xrightarrow{\mathbb{P}} \sigma^2 \quad \text{as } n \rightarrow \infty. \quad (3.24)$$

So, we can suppose  $\sigma$  to be known and focus on the drift parameter estimation of the jump-type CIR process (3.22) in critical and supercritical cases ( $b = 0$  and  $b \in (-\infty, 0)$ , respectively), which have not been addressed in previous research. We also study the subcritical case ( $b \in (0, \infty)$ ) and we get results extending those of Mai [Mai12, Theorem 4.3.1] in several aspects: we do not suppose the ergodicity of the process  $Y$  and we make explicit the expectation of the unique stationary distribution of  $Y$ .

To derive the MLE for the parameter  $b$  based on the observations  $(Y_t)_{t \in [0, T]}$ , we provide an explicit expression of the Radon–Nikodym derivative. For  $b, \tilde{b} \in \mathbb{R}$ , the probability measures  $\mathbb{P}_{b, T}$  and  $\mathbb{P}_{\tilde{b}, T}$  are absolutely continuous with respect to each other and

$$\log \left( \frac{d\mathbb{P}_{b, T}}{d\mathbb{P}_{\tilde{b}, T}}(\tilde{Y}) \right) = -\frac{b - \tilde{b}}{\sigma^2} (\tilde{Y}_T - y_0 - aT - J_T) - \frac{b^2 - \tilde{b}^2}{2\sigma^2} \int_0^T \tilde{Y}_s ds, \quad (3.25)$$

where  $\tilde{Y}$  is the process corresponding to the parameter  $\tilde{b}$ . This last expression allows us to get an explicit expression of the MLE of  $b$  given by

$$\hat{b}_T = -\frac{Y_T - y_0 - aT - J_T}{\int_0^T Y_s ds}, \quad (3.26)$$

provided that  $\int_0^T Y_s ds > 0$  which holds almost surely under the above assumptions. Here, let us point out that using the SDE (3.22), we have  $\Delta J_t = \Delta Y_t$ ,  $t \in \mathbb{R}_+$  and then, by the property  $J_t = \sum_{s \in [0, t]} \Delta J_s$ , which is valid under condition (3.23), we obtain  $J_t = \sum_{s \in [0, t]} \Delta Y_s$ , for all  $t \in \mathbb{R}_+$ . Consequently, for all  $t \in [0, T]$ ,  $J_t$  is a measurable function of  $(Y_u)_{u \in [0, T]}$ , so that the right hand side of (3.26) is a statistic as desired. Besides, in the subcritical case  $b > 0$ , we have the existence of a unique stationary distribution and the exponential ergodicity for the process  $(Y_t)_{t \geq 0}$ , see e.g. Pinsky [Pin72].

**Theorem 3.4.1.** *Let  $a \geq 0$ ,  $b > 0$ ,  $\sigma > 0$ , and let  $m$  be a Lévy measure on  $(0, \infty)$  satisfying (3.23). Let  $(Y_t)_{t \geq 0}$  be the unique strong solution of the SDE (3.22) satisfying  $\mathbb{P}(Y_0 \geq 0) = 1$  and  $\mathbb{E}(Y_0) < \infty$ .*

(i) *Then  $(Y_t)_{t \geq 0}$  converges in law to its unique stationary distribution  $\pi$  given by*

$$\int_0^\infty e^{uy} \pi(dy) = \exp \left\{ \int_u^0 \frac{F(v)}{R(v)} dv \right\} = \exp \left\{ \int_u^0 \frac{av + \int_0^\infty (e^{vz} - 1)m(dz)}{\frac{\sigma^2}{2}v^2 - bv} dv \right\}, u < 0.$$

*Moreover,  $\pi$  has a finite expectation given by  $\int_0^\infty y \pi(dy) = \left( a + \int_0^\infty z m(dz) \right) \frac{1}{b} > 0$ .*

- (ii) If, in addition,  $a > 0$  and the extra moment condition  $\int_0^1 z \log\left(\frac{1}{z}\right) m(dz) < \infty$  holds, then the process  $(Y_t)_{t \geq 0}$  is exponentially ergodic, namely, there exist constants  $\beta \in (0, 1)$  and  $C > 0$  such that

$$\|\mathbb{P}_{Y_t|Y_0=y} - \pi\|_{\text{TV}} \leq C(y+1)\beta^t, \quad t \geq 0, \quad y \geq 0,$$

where  $\|\mu\|_{\text{TV}}$  denotes the total-variation norm of a signed measure  $\mu$  on  $(0, \infty)$  defined by  $\|\mu\|_{\text{TV}} := \sup_{A \in \mathcal{B}(\mathbb{R}_+)} |\mu(A)|$ . Moreover, for all Borel measurable functions  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  with  $\int_0^\infty |f(y)| \pi(dy) < \infty$ , we have

$$\frac{1}{T} \int_0^T f(Y_s) ds \xrightarrow{\text{a.s.}} \int_0^\infty f(y) \pi(dy) \quad \text{as } T \rightarrow \infty. \quad (3.27)$$

Here, we draw attention that it might be possible to use (3.27) together with the central limit theorem for martingales to easily derive the asymptotic normality of the MLE (3.26) in the subcritical case, but at an extra cost in terms of the additional moment condition we need on  $m$ . To avoid such restrictive condition, we prove the following result.

**Theorem 3.4.2.** *Let  $a \geq 0$ ,  $b \in \mathbb{R}$ ,  $\sigma > 0$ , and let  $m$  be a Lévy measure on  $(0, \infty)$  satisfying (3.23). Let  $(Y_t)_{t \geq 0}$  be the unique strong solution of the SDE (3.22) satisfying  $\mathbb{P}(Y_0 = y_0) = 1$  with some  $y_0 \geq 0$ . For all  $u, v \leq 0$ , if we denote  $\gamma_v := \sqrt{b^2 - 2\sigma^2 v}$ , then*

$$\mathbb{E} \left[ \exp \left\{ uY_t + v \int_0^t Y_s ds \right\} \right] = \exp \left\{ \psi_{u,v}(t)y_0 + \int_0^t \left( a\psi_{u,v}(s) + \int_0^\infty (e^{z\psi_{u,v}(s)} - 1)m(dz) \right) ds \right\}$$

for  $t \geq 0$ , where the function  $\psi_{u,v} : \mathbb{R}_+ \rightarrow \mathbb{R}_-$  takes the form

$$\psi_{u,v}(t) = \begin{cases} \frac{u\gamma_v \cosh\left(\frac{\gamma_v t}{2}\right) + (-ub + 2v) \sinh\left(\frac{\gamma_v t}{2}\right)}{\gamma_v \cosh\left(\frac{\gamma_v t}{2}\right) + (-\sigma^2 u + b) \sinh\left(\frac{\gamma_v t}{2}\right)} & \text{if } v < 0 \text{ or } b \neq 0 \text{ (i.e., if } \gamma_v > 0), \\ \frac{u}{1 - \frac{\sigma^2 u}{2}t} & \text{if } v = 0 \text{ and } b = 0 \text{ (i.e., if } \gamma_v = 0). \end{cases}$$

Our above joint Laplace transform is in accordance with the corresponding one obtained in Keller-Ressel [KR08, Theorem 4.10] in case of a regular affine process and with the one in Jiao et al. [JMS16, Proposition 4.3] in case of a general CBI process. However, our contribution here is to provide a new proof for it based on the fact that the couple  $(Y_t, \int_0^t Y_s ds)_{t \in [0, \infty)}$  still defines a 2-dimensional CBI process which follows also from Keller-Ressel [KR08, Theorem 4.10]. This observation allows us to write the joint Laplace transform in an affine form as for the continuous CIR process and make the solutions of the corresponding Riccati-type differential equations explicit. That is, Theorem 3.4.2 allows not only to avoid the above extra moment condition on the Lévy measure  $m$  in the subcritical case but also make the study of the critical and supercritical cases possible.

### Subcritical case ( $b > 0$ ).

We prove that if  $a > 0$  and the Lévy measure  $m$  satisfies (3.23), then

$$\sqrt{T}(\hat{b}_T - b) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \frac{\sigma^2 b}{a + \int_0^\infty z m(dz)} \right) = \mathcal{N} \left( 0, \frac{\sigma^2}{\int_0^\infty y \pi(dy)} \right) \quad \text{as } T \rightarrow \infty,$$

where  $\pi$  denotes the unique stationary distribution of  $(Y_t)_{t \geq 0}$  (see (i) of Theorem 3.4.1).

**Critical case** ( $b = 0$ ).

Let  $a \geq 0$ ,  $(Y_t)_{t \geq 0}$  be the unique strong solution of the SDE (3.22) such that  $\mathbb{P}(Y_0 = y_0) = 1$  with some  $y_0 \geq 0$  and the Lévy measure  $m$  satisfying (3.23). Suppose that  $a > 0$  or  $a = 0$ ,  $y_0 > 0$ ,  $\int_0^\infty z m(dz) > 0$ . Then, we prove that

$$T(\widehat{b}_T - b) = T\widehat{b}_T \xrightarrow{\mathcal{D}} \frac{a + \int_0^\infty z m(dz) - \mathcal{Y}_1}{\int_0^1 \mathcal{Y}_s ds} \quad \text{as } T \rightarrow \infty,$$

where  $(\mathcal{Y}_t)_{t \geq 0}$  is the unique strong solution of a critical (diffusion type) CIR model

$$d\mathcal{Y}_t = \left( a + \int_0^\infty z m(dz) \right) dt + \sigma \sqrt{\mathcal{Y}_t} d\mathcal{W}_t, \quad t \in \mathbb{R}_+,$$

with initial condition  $\mathcal{Y}_0 = 0$  and  $(\mathcal{W}_t)_{t \geq 0}$  is a standard Wiener process.

**Supercritical case** ( $b < 0$ ).

Let  $a \geq 0$ ,  $(Y_t)_{t \geq 0}$  be the unique strong solution of the SDE (3.22) such that  $\mathbb{P}(Y_0 = y_0) = 1$  with some  $y_0 \geq 0$  and the Lévy measure  $m$  satisfying (3.23). Suppose that  $a > 0$  or  $a = 0$ ,  $y_0 > 0$ ,  $\int_0^\infty z m(dz) > 0$ . Then, we prove that

$$e^{-bT/2}(\widehat{b}_T - b) \xrightarrow{\mathcal{D}} \sigma Z \left( -\frac{V}{b} \right)^{-1/2} \quad \text{as } T \rightarrow \infty,$$

where  $V$  is a positive random variable having an explicit Laplace transform

$$\mathbb{E}(e^{uV}) = \left( 1 + \frac{\sigma^2 u}{2b} \right)^{-\frac{2a}{\sigma^2}} \exp \left\{ \frac{uy_0}{1 + \frac{\sigma^2 u}{2b}} + \int_0^\infty \left( \int_0^\infty \left( \exp \left\{ \frac{zue^{by}}{1 + \frac{\sigma^2 u}{2b} e^{by}} \right\} - 1 \right) m(dz) \right) dy \right\}$$

and  $Z$  is a standard normally distributed random variable, independent of  $V$ .

## 3.5 Perspectives

**Large deviations for Wishart processes.** With Aurélien Alfonsi we plan to establish large deviation principles for the maximum likelihood estimators presented in Section 3.3. Based on a new strategy introduced by Bercu and Richou [BR15], Du Roy De Chaumaray [De 14] established large deviation results for the CIR process taking advantage of the explicit moment generating-Laplace transform established in [BAK13] (see Section 3.2). We then hope to use a similar strategy taking advantage of the explicit Laplace transform we established for the Wishart process (see Theorem 3.3.1).

**Parameter estimation in rough volatility models** In their recent paper Gatheral, Jaisson and Rosenbaum [GJR14] proved that fractional Brownian motion with short memory have a better aptitude for reproducing stylized effects of both historical and implied volatilities. So they adopt the fractional Heston model originally introduced by Comte, Coutin and Renault [CCR12] with a small Hurst parameter. We aim at studying the problem of parameter estimation for these challenging models.

## Chapter 4

# Statistical properties for a discretely observed CIR model

During the last forty years or so continuous-time stochastic differential equations have been widely applied in mathematical finance for describing stock prices, interest rates, volatility evolution. However, only discrete-time observations are available in practice which makes the problem of parameter estimation more and more challenging in this context. We refer the reader to the book of Kessler and Sørensen [KLSr12] for an intensive survey on this topic. As pointed out in Section 3.1, only few results can be found in the literature for non-ergodic diffusions or diffusions with non-regular coefficients. The content of this chapter follows from the articles [BAK13] and [BKK17].

### 4.1 Parameter estimation in a high frequency setting

In [BAK13], we consider rather a discrete sample  $(X_{t_k})_{0 \leq k \leq n}$  of the CIR diffusion at deterministic and equidistant instants  $(t_k = k\Delta_n)_{0 \leq k \leq n}$ . Our aim is to study the problem of estimating  $\theta = (a, b)$  from discrete observations, under high frequency conditions, namely  $\Delta_n \rightarrow 0$ , and infinite horizon,  $n\Delta_n \rightarrow \infty$ . A common way to do that is to consider a discretization of the logarithm likelihood (see e.g. [KLSr12]). In our case this method yields the contrast

$$\frac{1}{2\sigma} \sum_{k=0}^{n-1} \frac{a - bX_{t_k}}{X_{t_k}} (X_{t_{k+1}} - X_{t_k}) - \frac{1}{4\sigma} \sum_{k=0}^{n-1} \Delta_n \frac{(a - bX_{t_k})^2}{X_{t_k}}.$$

Our approach is slightly different since we discretize the continuous time MLE of  $\theta = (a, b)$ , obtained in Section 3.2, instead of considering the maximum argument of the above contrast. Doing so, we take advantage of our limit theorems obtained in the continuous time observations context. The discretized version of the MLE is given by

$$\left\{ \begin{array}{l} \hat{a}_{t_n}^{\Delta_n} = \frac{\left( \log X_{t_n} - \log x + \sigma \sum_{k=0}^{n-1} \frac{\Delta_n}{X_{t_k}} \right) \sum_{k=0}^{n-1} \Delta_n X_{t_k} - t_n (X_{t_n} - x)}{\sum_{k=0}^{n-1} \frac{\Delta_n}{X_{t_k}} \sum_{k=0}^{n-1} \Delta_n X_{t_k} - t_n^2} \\ \hat{b}_{t_n}^{\Delta_n} = \frac{t_n \left( \log X_{t_n} - \log x + \sigma \sum_{k=0}^{n-1} \frac{\Delta_n}{X_{t_k}} \right) - (X_{t_n} - x) \sum_{k=0}^{n-1} \frac{\Delta_n}{X_{t_k}}}{\sum_{k=0}^{n-1} \frac{\Delta_n}{X_{t_k}} \sum_{k=0}^{n-1} \Delta_n X_{t_k} - t_n^2}. \end{array} \right. \quad (4.1)$$

As mentioned in Section 3.2, the relevant study cases are the subcritical and critical ( $b = 0$ ) since for the remaining supercritical case the MLE estimator of  $\theta = (a, b)$  is not even consistent. In this setting we prove the following results

**Subcritical case** ( $b > 0$ ).

**Theorem 4.1.1.** *For all  $\eta \in ] -\frac{a}{\sigma}, +\infty[$ , we have*

$$\mathbb{E}_x(X_t^\eta) \sim \left(\frac{\sigma^2}{2b}\right)^\eta \frac{\Gamma(\frac{a}{\sigma} + \eta)}{\Gamma(\frac{a}{\sigma})}, \text{ as } t \rightarrow +\infty$$

and for all  $0 < t - s < 1$  there exist  $C_1 > 0$  and  $C_2 > 0$  such that:

1. For all  $q \geq 1$ ,  $\mathbb{E}_x |X_t - X_s|^q \leq C_1(t - s)^{\frac{q}{2}}$ .
2. For all  $a > 2\sigma$ ,  $\mathbb{E}_x \left| \frac{1}{X_t} - \frac{1}{X_s} \right| \leq C_2(t - s)^{\frac{1}{2}}$ .

Moreover, for  $a > 2\sigma$ , if  $n\Delta_n^2 \rightarrow 0$ , then we have

$$\mathcal{L}_\theta \left\{ \sqrt{t_n}(\hat{\theta}_{t_n}^{\Delta_n} - \theta) \right\} \xrightarrow[n \rightarrow \infty]{} \mathcal{N}_{\mathbb{R}^2} (0, 2\sigma\Gamma^{-1}), \text{ with } \Gamma = \begin{pmatrix} \frac{b}{a-\sigma} & -1 \\ -1 & \frac{a}{b} \end{pmatrix}.$$

**Critical case** ( $b = 0$ ).

**Theorem 4.1.2.** *For all  $\eta \in ] -\frac{a}{\sigma}, +\infty[$ , we have*

$$\mathbb{E}_x(X_t^\eta) \sim \sigma^\eta \frac{\Gamma(\frac{a}{\sigma} + \eta)}{\Gamma(\frac{a}{\sigma})} t^\eta, \text{ as } t \rightarrow +\infty$$

and for all  $0 < t - s < 1$  there exist  $C_3 > 0$  and  $C_4 > 0$  such that:

1. For all  $q \geq 2$ ,  $\mathbb{E}_x |X_t - X_s|^q \leq C_3(t - s)^{\frac{q}{2}} \sup_{s \leq u \leq t} \mathbb{E}_x(X_u^{\frac{q}{2}})$ .
2. For all  $1 \leq q < 2$ ,  $\mathbb{E}_x |X_t - X_s|^q \leq C_4(at + x)^{\frac{q}{2}}(t - s)^{\frac{q}{2}}$ .
3. For all  $a > 2\sigma$ , there exists  $q \geq 2$  and  $2 < p < \frac{a}{\sigma}$ , such that

$$\mathbb{E}_x \left| \frac{1}{X_t} - \frac{1}{X_s} \right| \leq C(t - s)^{\frac{1}{2}} \sup_{s \leq u \leq t} \left( \mathbb{E}_x(X_u^{\frac{q}{2}}) \right)^{\frac{1}{q}} \|X_t^{-1}\|_p \|X_s^{-1}\|_p.$$

Moreover, for  $a > 2\sigma$ , if  $n\Delta_n^2 \rightarrow 0$  and  $\frac{n\Delta_n^{\frac{3}{2}}}{\log(n\Delta_n)} \rightarrow 0$ , then we have

$$\mathcal{L}_\theta \left\{ \text{diag}(\sqrt{\log t_n}, t_n)(\hat{\theta}_{t_n}^{\Delta_n} - \theta) \right\} \xrightarrow[n \rightarrow \infty]{} \left( \sqrt{2\sigma(a - \sigma)}G, \frac{a - R_1}{I_1} \right),$$

where  $(R_t)_{t \geq 0}$  is the CIR process, starting from 0, solution to (3.4),  $I_t = \int_0^t R_s ds$ , and  $G$  is a standard normal random variable independent of  $(R_1, I_1)$ .

Here, we draw attention that the condition  $n^2\Delta_n \rightarrow 0$  is consistent with those of papers in the literature dealing with the same problem for ergodic diffusions with regular coefficients (see e.g. Yoshida [Yos92] and the references there). However, for  $b = 0$ , the condition  $\max(n^2\Delta_n, \frac{n\Delta_n^{\frac{3}{2}}}{\log(n\Delta_n)}) \rightarrow 0$  seems to be quite distinctive since it concerns a non-ergodic case.

## 4.2 Local asymptotic properties of CIR process discretely observed at high frequency

A fundamental concept in asymptotic theory of statistics is the local asymptotic normality (LAN) property introduced by Le Cam [LC60] and then extended by Jeganathan [Jeg82] to the local asymptotic mixed normality (LAMN) property. The local asymptotic quadraticity (LAQ) property was introduced by [LCY90]. Initiated by Gobet [Gob01], Malliavin calculus techniques have recently been applied in order to analyze the log-likelihood ratio of the discrete observation of continuous diffusion processes. Concretely, Gobet [Gob01, Gob02] obtained the LAMN and LAN properties respectively for multi-dimensional elliptic diffusions and ergodic diffusions on the basis of discrete observations at high frequency. As far as we know, the validity of the local asymptotic properties for CIR process on the basis of discrete observations has never been addressed in the literature. One main difficulty comes from the fact that its diffusion coefficient is the square root function.

In [BKK17], we focus on deriving the LAN property in the subcritical case, the LAQ property in the critical case and the LAMN property in the supercritical case. To obtain these results, we use the Malliavin calculus techniques developed recently for CIR process by Alòs et al. [AE08] and Altmayer et al. [AN15] together with the  $L^p$ -norm estimation for positive and negative moments of the CIR process obtained in [BAK13] (see theorems 4.1.1 and 4.1.2). We consider a CIR process  $(X_t^{a,b})_{t \geq 0}$  given by

$$X_t^{a,b} = x_0 + \int_0^t (a - bX_s^{a,b}) ds + \int_0^t \sqrt{2\sigma X_s^{a,b}} dB_s, \quad (4.2)$$

where  $X_0^{a,b} = x_0 > 0$  is a given initial condition and we assume that  $a > \sigma > 0$ . Here,  $B = (B_t)_{t \geq 0}$  is a standard Brownian motion. For any  $t > s$ , the law of  $X_t^{a,b}$  conditioned on  $X_s^{a,b} = x$  admits a positive transition density  $p^{a,b}(t-s, x, y)$ , which is explicit and differentiable w.r.t.  $a$  and  $b$ . Then, we consider a discrete observation sample at deterministic and equidistant times  $t_k = k\Delta_n$ ,  $k \in \{0, \dots, n\}$  of the process  $X^{a,b}$ , which is denoted by  $X^{n,a,b} = (X_{t_0}^{a,b}, X_{t_1}^{a,b}, \dots, X_{t_n}^{a,b})$ , where  $\Delta_n \leq 1$  for all  $n \geq 1$ . We assume that the high-frequency and infinite horizon conditions hold ( $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  as  $n \rightarrow \infty$ ). We denote by  $\mathbb{P}_n^{a,b}$  the probability law of the random vector  $X^{n,a,b}$ . For a reference parameter  $(a_0, b_0) \in \Theta \times \Sigma$ , we consider a discrete observation of the process  $X^{a_0, b_0}$  given by  $X^{n, a_0, b_0} = (X_{t_0}^{a_0, b_0}, X_{t_1}^{a_0, b_0}, \dots, X_{t_n}^{a_0, b_0})$ . The aim here is to study the asymptotic behavior of the log-likelihood ratio given by  $\log \frac{d\mathbb{P}_n^{a_n, b_n}}{d\mathbb{P}_n^{a_0, b_0}}(X^{n, a_0, b_0})$  in ergodic and non-ergodic regimes. To do so, we use the Markov property and the mean value theorem to get

$$\begin{aligned} \log \frac{d\mathbb{P}_n^{a_n, b_n}}{d\mathbb{P}_n^{a_0, b_0}}(X^{n, a_0, b_0}) &= \sum_{k=0}^{n-1} (a_n - a_0) \int_0^1 \frac{\partial_a p^{a(\ell), b_0}}{p^{a(\ell), b_0}}(\Delta_n, X_{t_k}^{a_0, b_0}, X_{t_{k+1}}^{a_0, b_0}) d\ell \\ &\quad + \sum_{k=0}^{n-1} (b_n - b_0) \int_0^1 \frac{\partial_b p^{a_n, b(\ell)}}{p^{a_n, b(\ell)}}(\Delta_n, X_{t_k}^{a_0, b_0}, X_{t_{k+1}}^{a_0, b_0}) d\ell, \end{aligned}$$

where  $a(\ell) := a_0 + \ell(a_n - a_0)$  and  $b(\ell) := b_0 + \ell(b_n - b_0)$ . Then, combining the Malliavin calculus techniques together with the  $L^p$ -norm estimation for positive and negative



moments of the CIR process (see theorems 4.1.1 and 4.1.2), to prove the following results.

**Subcritical case** ( $b > 0$ ).

**Theorem 4.2.1.** *Assume that  $a/\sigma > 5 + 3\sqrt{2}$ . Then, the LAN property holds for the likelihood at  $(a_0, b_0)$  with rate of convergence  $(\sqrt{n\Delta_n}, \sqrt{n\Delta_n})$  and asymptotic Fisher information matrix  $I(a_0, b_0)$ . That is, for all  $z = (u, v)^\top \in \mathbb{R}^2$ , as  $n \rightarrow \infty$ ,*

$$\log \frac{dP_n^{a_n, b_n}}{dP_n^{a_0, b_0}} \left( X^{n, a_0, b_0} \right) \xrightarrow{\mathcal{L}(\hat{P}^{a_0, b_0})} z^\top \mathcal{N}(0, I(a_0, b_0)) - \frac{1}{2} z^\top I(a_0, b_0) z,$$

where  $a_n := a_0 + \frac{u}{\sqrt{n\Delta_n}}$ ,  $b_n := b_0 + \frac{v}{\sqrt{n\Delta_n}}$ , and  $\mathcal{N}(0, I(a_0, b_0))$  is a centered  $\mathbb{R}^2$ -valued

Gaussian vector with covariance matrix  $I(a_0, b_0) := \frac{1}{2\sigma} \begin{pmatrix} \frac{b_0}{a_0 - \sigma} & -1 \\ -1 & \frac{a_0}{b_0} \end{pmatrix}$ .

**Critical case** ( $b = 0$ ).

**Theorem 4.2.2.** *Assume that  $a/\sigma > 5 + 3\sqrt{2}$  and  $\frac{n\Delta_n^{\frac{3}{2}}}{\log(n\Delta_n)} \rightarrow 0$  as  $n \rightarrow \infty$ . Then, the LAQ property holds for the likelihood at  $(a_0, 0)$  with rates of convergence  $(\sqrt{\log(n\Delta_n)}, n\Delta_n)$  and random matrix  $I(a_0, 0)$ . That is, for all  $z = (u, v)^\top \in \mathbb{R}^2$ , as  $n \rightarrow \infty$ ,*

$$\log \frac{dP_n^{a_n, b_n}}{dP_n^{a_0, 0}} \left( X^{n, a_0, 0} \right) \xrightarrow{\mathcal{L}(\hat{P}^{a_0, 0})} z^\top U(a_0, 0) - \frac{1}{2} z^\top I(a_0, 0) z,$$

where  $a_n := a_0 + u/\sqrt{\log(n\Delta_n)}$ ,  $b_n := 0 + v/(n\Delta_n)$ , and  $U(a_0, 0)$  is a  $\mathbb{R}^2$ -valued random vector given by  $U(a_0, 0) := \left( \frac{1}{\sqrt{2\sigma(a_0 - \sigma)}} G, \frac{a_0 - R_1^{a_0, 0}}{2\sigma} \right)^\top$  with covariance matrix  $I(a_0, 0) := \frac{1}{2\sigma} \text{diag}\left(\frac{1}{a_0 - \sigma}, \int_0^1 R_s^{a_0, 0} ds\right)$ . Here,  $(R_t^{a_0, 0})_{t \geq 0}$  is the process starting at 0 and solution to (3.4) and  $G$  is an independent standard normal random variable.

**Supercritical case** ( $b < 0$ ).

**Theorem 4.2.3.** *Assume that  $a/\sigma > 5 + 3\sqrt{2}$  and  $n\Delta_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ . Then, for all  $z = (u, v)^\top \in \mathbb{R}^2$ , as  $n \rightarrow \infty$ ,*

$$\log \frac{dP_n^{a_n, b_n}}{dP_n^{a_0, b_0}} \left( X^{n, a_0, b_0} \right) \xrightarrow{\mathcal{L}(\hat{P}^{a_0, b_0})} z^\top U(a_0, b_0) - \frac{1}{2} z^\top I(a_0, b_0) z,$$

where  $a_n := a_0 + u$ ,  $b_n := b_0 + v/(e^{-b_0 \frac{n\Delta_n}{2}})$  and  $U(a_0, b_0)$  is a  $\mathbb{R}^2$ -valued random vector given by  $U(a_0, b_0) = \left( \frac{V}{2\sigma}, \frac{1}{\sqrt{2\sigma}} \left(-\frac{1}{b_0} R_{-\frac{1}{b_0}}^{a_0}\right)^{\frac{1}{2}} Z_1 \right)^\top$  with covariance matrix  $I(a_0, b_0) := \frac{1}{2\sigma} \text{diag}\left(\int_0^{-\frac{1}{b_0}} R_s^{a_0} ds, -\frac{1}{b_0} R_{-\frac{1}{b_0}}^{a_0}\right)$ . Here,  $V := \log R_{-\frac{1}{b_0}}^{a_0} - \log x_0 - (a_0 - \sigma) \int_0^{-\frac{1}{b_0}} R_s^{a_0} ds$ , and  $Z_1$  is a standard normal random variable independent of  $R_{-\frac{1}{b_0}}^{a_0}$ . Here,  $(R_t^{a_0})_{t \geq 0}$  is the process starting at  $x_0$  and solution to (3.4).

We draw attention that in the particular case, when  $a$  is known and  $b$  is unknown to be estimated, the LAMN property holds for the likelihood at  $b_0$  with rate of convergence  $e^{-b_0 \frac{n\Delta_n}{2}}$  and asymptotic random Fisher information  $I(b_0) := -\frac{1}{2\sigma b_0} R_{-\frac{1}{b_0}}^a$ . That is, for all  $v \in \mathbb{R}$ , as  $n \rightarrow \infty$ ,

$$\log \frac{dP_n^{a,b_n}}{dP_n^{a,b_0}} \left( X^{n,a,b_0} \right) \xrightarrow{\mathcal{L}(\widehat{P}^{a,b_0})} v \sqrt{I(b_0)} \mathcal{N}(0, 1) - \frac{v^2}{2} I(b_0),$$

where  $\mathcal{N}(0, 1)$  is a centered standard Gaussian random variable which is independent of  $I(b_0)$ . In this case, condition  $n\Delta_n^2 \rightarrow 0$  as  $n \rightarrow \infty$  is no longer needed.

It is worth noticing that thanks to these local asymptotic properties, we deduce by the Hájek-Le Cam convolution theorem and the minimax theorem [Jeg82, Proposition 2], the asymptotical efficiency of the estimator (4.1) derived in [BAK13] which asymptotically achieves the Cramér-Rao lower bound  $I(a_0, b_0)^{-1}$  for the estimation variance.

### 4.3 Perspectives

#### **Local asymptotic properties for discretely observed jump-type CIR processes.**

A natural question is whether one can prove local asymptotic properties similar to the ones of Section 4.2 for a CIR process with jumps. In the presence of jumps, several cases have been largely investigated, see e.g. Aït-Sahalia and Jacod [ASJ07], Kawai [Kaw13], Clément, Delattre and Gloter [CDG14], Clément and Gloter [CG15], Kohatsu-Higa, Nualart and Tran [KHNT14, KHNT17], and Tran [Tra17]. However, all these results deal with the stochastic differential equations whose coefficients are continuously differentiable and satisfy a global Lipschitz condition. The case where the coefficient functions of the model do not satisfy these standard assumptions, for instance the square root diffusion function in the jump-type CIR model still remains an open problem. This topic is currently under study with Mohamed Ben Alaya and Khue Tran, where we have obtained partial results on this question.

#### **Local asymptotic properties of square root type processes under the presence of microstructure noise.**

As confirmed by recent empirical studies in finance, it is well known that the market microstructure contaminates the high-frequency data and may cause serious damage to the approximation of the volatility. As explained in Zhang, Mykland and Aït-Sahalia [ZMAS05], in practice the realized volatility estimator as we give in (3.24) does not converge as the sampling frequency increases. Several approaches build volatility consistent estimators reducing the effect of this noise are available in the literature. We mention for example the works of Zhang, Mykland, and Aït-Sahalia [ZMAS05], Barndorff-Nielsen et al. [BNHLS11] and Podolskij and Vetter [PV09]. There are also works assuming both microstructure noise and non-synchronous observations as in Christensen, Kinnebrock, and Podolskij [CKP10] and Christensen, Podolskij, and Vetter [CPV13]. For this last setting, Ogihara [Ogi14] studied parametric inference for diffusion processes proving local asymptotic properties under restrictive assumption on the diffusion coefficient of the model. We aim at proving such local asymptotic properties for the square root type diffusion. This is a quite challenging question since such results does not seem to have been proved even for noisy and equidistant observations case.



## Part III

# Weighted limit theorems for quasi-left continuous-time vector martingales



## Chapter 5

# Generalized logarithmic limit theorems for quasi-left continuous martingales

### 5.1 Introduction and motivation

A classical property satisfied by a standard Brownian motion  $B$  is the logarithmic strong law

$$(LSL) \quad \forall f \in L^1(G), \quad (\log t)^{-1} \int_1^t f\left(\frac{B_s}{\sqrt{s}}\right) \frac{ds}{s} \xrightarrow[t \rightarrow \infty]{\text{a.s.}} \int f dG, \text{ where } G \sim \mathcal{N}(0, 1).$$

An immediate consequence of the LSL property is the celebrated almost-sure central limit theorem

$$(ASCLT), \quad (\log t)^{-1} \int_1^t \frac{ds}{s} \delta_{\frac{B_s}{\sqrt{s}}} \Longrightarrow G \text{ a.s. when } t \rightarrow \infty,$$

established in a functional form by Brosamler [Bro88].

The content of this chapter follows from the articles [CK08] and [FK12], where we focus on extending the above properties to the setting of vector quasi-left continuous martingales with generalized logarithmic weights. We provide a possible application of these theoretical results in the estimation of drift and variance of processes with stationary independent increments. In this setting, we consider a  $d$ -dimensional quasi-left continuous martingale  $M = (M_t, t \geq 0)$ , locally square integrable, defined on a filtered space of probability  $(\Omega, \mathcal{F}, (\mathcal{F})_{t \geq 0}, \mathbb{P})$  and we consider a deterministic  $d \times d$  non-singular matrix process  $V = (V_t)_{t \geq 0}$ . For  $u \in \mathbb{R}^d$ , we set

$$\phi_t(u) := \exp\left(-\frac{1}{2}u^\top \langle M^c \rangle_t u + \int_0^t \int_{\mathbb{R}^d} (\exp\{i\langle u, x \rangle\} - 1 - i\langle u, x \rangle) \nu^M(ds, dx)\right),$$

where  $\nu^M$  denotes the compensator of the random measure associated to the jumps of  $M$ . Our results are based on the generalized CLT for martingales stated below (see Touati [Tou91]).

**Theorem 5.1.1.** *Under the above notations, we define a probability  $\mathcal{Q}$  on the space  $C(\mathcal{X}, \mathbb{R}^d)$  of continuous functions from  $\mathcal{X}$  to  $\mathbb{R}^d$  (where  $\mathcal{X}$  indicates a vector space of finite dimension). If the couple  $(M, V)$  satisfies the following assumption*

$$(\mathcal{H}) \quad \phi_t((V_t^\top)^{-1}u) \xrightarrow[t \rightarrow \infty]{\text{a.s.}} \phi_\infty(\eta, u) \neq 0 \text{ a.s.},$$

where  $\eta$  denotes a r.v., possibly degenerated taking values in  $\mathcal{X}$  and

$$\phi_\infty(z, u) = \int_{\mathbb{R}^d} \exp\{i\langle u, \xi \rangle\} \pi(z, d\xi), \quad (z, u) \in \mathcal{X} \times \mathbb{R}^d,$$

denotes the Fourier transform of the one-dimensional conditionals laws  $(\pi(x, \cdot), x \in \mathcal{X})$  of  $\mathcal{Q}$ . Then,

$$(\text{GCLT}), \quad Z_t := V_t^{-1}M_t \xrightarrow{\text{stably}} Z_\infty := \Sigma(\eta), \text{ when } t \rightarrow \infty,$$

where  $(\Sigma(z), z \in \mathcal{X})$  is a  $\mathcal{Q}$  law process independent of the r.v.  $\eta$ .

In what follows,  $\|\cdot\|$  denotes the Frobenius matrix norm.

## 5.2 Martingale with regular normalization

We say that a given normalization  $V$  is *regular*, if it satisfies  $(\mathcal{C}) = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$  :

- $(\mathcal{C}_1)$   $t \mapsto V_t$  is of class  $\mathcal{C}^1$ ,
- $(\mathcal{C}_2)$  there exists  $s_0 \geq 0$  such that for every  $t \geq s \geq s_0$  we have  $V_s V_s^\top \leq V_t V_t^\top$  (in the sense of real positive-semidefinite symmetric matrices),
- $(\mathcal{C}_3)$  there exists a continuous function  $a = (a_t)$  taking values on  $\mathbb{R}_+$ , decreasing to 0, such that  $A_t = \int_0^t a_s ds \xrightarrow[t \rightarrow \infty]{} \infty$  and a matrix  $U_1$  such that

$$\begin{cases} a_t^{-1} V_t^{-1} \frac{dV_t}{dt} - U_1 = \Delta_{t,1}, & \text{with } \Delta_{t,1} \xrightarrow[t \rightarrow \infty]{} 0, \\ U_1 + U_1^\top = S_1, & \text{where } S_1 \text{ is a non-singular matrix.} \end{cases}$$

These conditions are simple to verify especially in the case of a scalar normalization.

In [CK08], we prove that under assumptions:  $(\mathcal{H})$ ,

$$(\mathcal{H}_1) \quad V_t^{-1} \langle M \rangle_t (V_t^\top)^{-1} \xrightarrow[t \rightarrow \infty]{\text{a.s.}} C$$

and if furthermore condition  $(\mathcal{C}_3)$  is obtained with  $\|\Delta_{t,1}\| = \mathcal{O}(A_t^{-\beta})$ , for  $\beta > 1$ , the generalized almost-sure central limit theorem holds.

$$(\text{GASCLT}) \quad \log(\det[V_t^2])^{-1} \int_0^t \delta_{Z_s} d(\log(\det[V_s^2])) \Longrightarrow \mu_\infty \text{ a.s., when } t \rightarrow \infty,$$

where  $\mu_\infty$  denotes the law of  $Z_\infty$ . If in addition, the couple  $(M, V)$  satisfies assumptions

$$(\mathcal{H}_2) \quad V_t^{-1} [M]_t (V_t^\top)^{-1} \xrightarrow[t \rightarrow \infty]{\text{a.s.}} C$$

and

$$(\mathcal{H}_3) \quad C = \int xx^\top d\mu_\infty(x),$$

then we prove a regular quadratic strong law

$$(\text{QSL}) \quad (\log(\det[V_t^2]))^{-1} \int_0^t V_s^{-1} M_s M_s^\top (V_s^\top)^{-1} d(\log(\det[V_s^2])) \xrightarrow[t \rightarrow \infty]{\text{a.s.}} C$$

and an associated central limit theorem.

### 5.3 Martingale with explosive normalization

We say that  $V$  is an *explosive* normalization, if it satisfies the following condition  $(\mathcal{C}')$  =  $\{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}'_3\}$  where

- $(\mathcal{C}'_3)$  there exists a matrix  $U_2$  such that

$$\begin{cases} V_t^{-1} \frac{dV_t}{dt} - U_2 = \Delta_{t,2}, & \text{with } \Delta_{t,2} \xrightarrow[t \rightarrow \infty]{} 0, \\ U_2 + U_2^\top = S_2, & \text{where } S_2 \text{ a non-singular matrix.} \end{cases}$$

Note that these conditions are fulfilled in the particular case where  $V_t$  is of scalar type that is:  $V_t = v_t I_d$  with  $v_t$  a scalar function given by  $v_t = c e^{bt}$  with  $c, b \in \mathbb{R}$ . We prove in [FK12], the following result.

**Theorem 5.3.1.** *Let  $M$  be a  $d$ -dimensional quasi-left continuous local martingale with  $M_0 = 0$ . let  $V$  be an explosive normalization of non-singular matrices satisfying condition  $(\mathcal{C}')$  with  $\|\Delta_{t,2}\| = \mathcal{O}(t^{-\beta})$ , as  $t \rightarrow \infty$  with  $\beta > 1$ . If the couple  $(M, V)$  satisfies assumptions:  $(\mathcal{H})$  and  $(\mathcal{H}_1)$  then*

$$(\text{GASCLT}) \quad \mu_t = t^{-1} \int_0^t \delta_{Z_s} ds \implies \mu_\infty \text{ a.s., when } t \rightarrow \infty,$$

where  $\mu_\infty$  is the limit law of  $Z_\infty$ . If in addition the couple  $(M, V)$  satisfies assumptions:  $(\mathcal{H}_2)$  and  $(\mathcal{H}_3)$ , then the quadratic strong law holds.

$$(\text{QSL}) \quad t^{-1} \int_0^t V_s^{-1} M_s M_s^\top (V_s^\top)^{-1} ds \xrightarrow[t \rightarrow \infty]{\text{a.s.}} C.$$

Finally, if the couple  $(M, V)$  satisfies assumptions:  $(\mathcal{H})$ ,  $(\mathcal{H}_1)$ ,  $(\mathcal{H}_3)$ ,

$$(\mathcal{H}''_2) \quad \exists p \in [1, 2], \text{ s.t. } \int_0^\infty \int_{\mathbb{R}^d} (1+s)^{-p/2} \|V_s^{-1} x\|^{2p} \nu^M(ds, dx) < \infty \text{ a.s.,}$$

and if furthermore condition  $(\mathcal{C}'_3)$  is obtained with  $\|\Delta_{t,2}\| = \mathcal{O}(t^{-\frac{3}{2}})$  as  $t \rightarrow \infty$ , then,

$$(\text{CLT}) \quad t^{-1/2} \int_0^t \{U_2 \tilde{D}_s + \tilde{D}_s U_2^\top\} ds \implies \nu_\infty, \text{ when } t \rightarrow \infty,$$

where  $\tilde{D}_s := V_s^{-1} (M_s M_s^\top - \langle M \rangle_s) (V_s^\top)^{-1}$  and  $\nu_\infty$  is the distribution of a random variable of the form  $\mathcal{C}^{1/2} G$  where  $G$  denotes a  $d$ -dimensional standard Gaussian vector and  $\mathcal{C}$  an explicit independent random matrix.



We also prove similar results when the normalization is mixed. We say that  $V = (V_t, t \geq 0) \in \mathbb{R}^{d \times d}$  is a *mixed* normalization, if it is of the form  $V = \text{diag}(V_1, V_2)$ , where  $V_1$  (resp.  $V_2$ ) denotes a deterministic family of non-singular matrices on  $\mathbb{R}^{d_1 \times d_1}$  (resp.  $\mathbb{R}^{d_2 \times d_2}$ ) satisfying condition  $(\mathcal{C})$  (resp. condition  $(\mathcal{C}')$ ). In this case, we deal with a  $d$ -dimensional quasi-left continuous local martingale  $M = (M_1, M_2)$  with  $M_0 = 0$  where  $M_1 \in \mathbb{R}^{d_1}$  and  $M_2 \in \mathbb{R}^{d_2}$  and  $Z_t := V_t^{-1}M_t$ .

## 5.4 Application example: estimation of the variance of processes with stationary independent increments

Let  $(S_t)_{t \geq 0}$  be a processes with stationary independent increments with Lévy measure  $\nu(dt, dx) = dt F(dx)$  with  $\int_{\mathbb{R}} |x|^{2p} F(dx) < \infty$ , for some  $p > 1$ , where  $F$  is a positive measure on  $\mathbb{R}$ . Let

$$m = \mathbb{E} S_1, \quad \sigma^2 = \mathbb{E} S_1^2 - m^2.$$

The QSL given in Subsection 5.2 gives a strongly consistent estimator of  $\sigma^2$ .

$$\hat{\sigma}_t^2 := (\log(1+t))^{-1} \int_0^t \frac{(S_r - mr)^2}{(1+r)^2} dr \xrightarrow[t \rightarrow \infty]{\text{a.s.}} \sigma^2.$$

If in addition, for some  $\rho > 1/2$ , there exists  $\mathcal{C} > 0$  such that

$$(1+t)^{-1} \sum_{r \leq t} (\Delta S_r)^2 - \int_{\mathbb{R}} |x|^2 F(dx) \leq \mathcal{C} [\log(1+t)]^{-\rho} \text{ a.s.},$$

then the functional central limit theorem corresponding to the QSL applies and we get

$$\sqrt{\log(1+t)}(\hat{\sigma}_t^2 - \sigma^2) \Longrightarrow \mathcal{N}(0, 4\sigma^4).$$

In [CK08], we improve the above speeds of convergence using weighting techniques.

## 5.5 Perspectives

Mâaouia and Touati [MT05] developed a global approach to build an asymptotic global confidence region for the covariance matrices of the reproduction distributions involved in a multitype branching process. Their approach is based on the central limit theorem corresponding to the QSL-type results proved by Chaâbane, Mâaouia and Touati [CM00], for the setting of discrete-time vector martingales. More recently, Barczy, Li and Pap [BLP15] introduced the notion of a multitype continuous-state branching process having  $d$ -types as a solution of a  $d$ -dimensional stochastic differential equation. Later on, Kyprianou and Palau [KP16] used a different approach for a representation allowing for up to countable infinity of types. We then aim at taking advantage of the central limit theorem developed for the setting of continuous time martingales (see Theorem 5.3.1) to build confidence region for the parameters involved in the multitype continuous-state branching process.

# Bibliography

- [AE08] E. Alòs and C.-O. Ewald. Malliavin differentiability of the Heston volatility and applications to option pricing. *Adv. in Appl. Probab.*, 40(1):144–162, 2008.
- [AGJC16] A. Al Gerbi, B. Jourdain, and E. Clément. Ninomiya-Victoir scheme: strong convergence, antithetic version and application to multilevel estimators. *Monte Carlo Methods Appl.*, 22(3):197–228, 2016.
- [AKR16] A. Alfonsi, A. Kebaier, and C. Rey. Maximum likelihood estimation for Wishart processes. *Stochastic Process. Appl.*, 126(11):3243–3282, 2016.
- [Alf15] A. Alfonsi. *Affine diffusions and related processes: simulation, theory and applications*, volume 6 of *Bocconi & Springer Series*. Springer, Cham; Bocconi University Press, Milan, 2015.
- [AMP05] C. Andrieu, E. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44(1):283–312, 2005.
- [AN15] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo quadrature of discontinuous payoffs in the generalized Heston model using Malliavin integration by parts. *SIAM J. Financial Math.*, 6(1):22–52, 2015.
- [AR01] S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.*, 38(2):482–493, 2001.
- [Aro04] B. Arouna. Adaptive Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1):1–24, 2004.
- [ASJ07] Y. Aït-Sahalia and J. Jacob. Volatility estimators for discretely sampled Lévy processes. *Ann. Statist.*, 35(1):355–392, 2007.
- [BAHK15] M. Ben Alaya, K. Hajji, and A. Kebaier. Importance sampling and statistical Romberg method. *Bernoulli*, 21(4):1947–1983, 2015.
- [BAHK16] M. Ben Alaya, K. Hajji, and A. Kebaier. Importance sampling and statistical Romberg method for Lévy processes. *Stochastic Process. Appl.*, 126(7):1901–1931, 2016.

- [BAK12] M. Ben Alaya and A. Kebaier. Parameter estimation for the square-root diffusions: ergodic and nonergodic cases. *Stoch. Models*, 28(4):609–634, 2012.
- [BAK13] M. Ben Alaya and A. Kebaier. Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stoch. Anal. Appl.*, 31(4):552–573, 2013.
- [BAK14] M. Ben Alaya and A. Kebaier. Multilevel Monte Carlo for Asian options and limit theorems. *Monte Carlo Methods Appl.*, 20(3):181–194, 2014.
- [BAK15] M. Ben Alaya and A. Kebaier. Central limit theorem for the multilevel Monte Carlo Euler method. *Ann. Appl. Probab.*, 25(1):211–234, 2015.
- [Bas13] Basel Committee On Banking Supervision. Fundamental review of the trading book: A revised market risk framework. Technical report, Bank of International Settlements, Basel, 2013.
- [Bat96] D. S. Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1):69–107, 1996.
- [BBKP15] M. Barczy, M. Ben Alaya, A. Kebaier, and G. Pap. Asymptotic behavior of maximum likelihood estimators for a jump-type Heston model. *ArXiv e-prints*, September 2015.
- [BBKP17] M. Barczy, M. Ben Alaya, A. Kebaier, and G. Pap. Asymptotic properties of maximum likelihood estimator for the growth rate for a jump-type {CIR} process based on continuous time observations. *Stochastic Processes and their Applications*, 2017.
- [BCC97] G. Bakshi, C. Cao, and Z. Chen. Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5):2003–2049, 1997.
- [BGT87] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.
- [BHK16] M. Ben Alaya, K. Hajji, and A. Kebaier. Improved adaptive Multilevel Monte Carlo and applications to finance. *ArXiv e-prints*, March 2016.
- [BK06] M. Broadie and O. Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Oper. Res.*, 54(2):217–231, 2006.
- [BKK17] M. Ben Alaya, A. Kebaier, and N. Khue Tran. Local asymptotic properties for Cox-Ingersoll-Ross process with discrete observations. *ArXiv e-prints*, August 2017.
- [BL94] N. Bouleau and D. Lépingle. *Numerical methods for stochastic processes*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1994. A Wiley-Interscience Publication.

- [BLP15] M. Barczy, Z. Li, and G. Pap. Stochastic differential equation with jumps for multi-type continuous state and continuous time branching processes with immigration. *ALEA Lat. Am. J. Probab. Math. Stat.*, 12(1):129–169, 2015.
- [BMP90] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.
- [BNHLS11] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics*, 162(2):149–169, 2011.
- [BR15] B. Bercu and A. Richou. Large deviations for the Ornstein-Uhlenbeck process with shift. *Adv. in Appl. Probab.*, 47(3):880–901, 2015.
- [Bro88] G. A. Brosamler. An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.*, 104(3):561–574, 1988.
- [Bru87] M. Bru. *Thèse 3<sup>ème</sup> cycle. Résistance d’Escherichie coli aux antibiotiques. Sensibilités des analyses en composantes principales aux perturbations Browniennes et simulation*. PhD thesis, Université Paris Nord, 1987.
- [Bru91] M. Bru. Wishart processes. *J. Theoret. Probab.*, 4(4):725–751, 1991.
- [BT96a] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1):43–60, 1996.
- [BT96b] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods Appl.*, 2(2):93–128, 1996.
- [CCR12] F. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2):337–378, May 2012.
- [CDG14] E. Clément, S. Delattre, and A. Gloter. Asymptotic lower bounds in estimating jumps. *Bernoulli*, 20(3):1059–1096, 2014.
- [CFMT11] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann. Affine processes on positive semidefinite matrices. *Ann. Appl. Probab.*, 21(2):397–463, 2011.
- [CG15] E. Clément and A. Gloter. Local asymptotic mixed normality property for discretely observed stochastic differential equations driven by stable Lévy processes. *Stochastic Process. Appl.*, 125(6):2316–2352, 2015.
- [CGMY02] P. Carr, H. Geman, D. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation. *The Journal of Business*, 75:305–332, 2002.

- [CIR85] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.
- [CK08] F. Chaabane and A. Kebaier. Théorèmes limites avec poids pour les martingales vectorielles à temps continu. *ESAIM Probab. Stat.*, 12:464–491, 2008.
- [CKP10] K. Christensen, S. Kinnebrock, and M. Podolskij. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics*, 159(1):116–133, 2010.
- [CL09] M. Craddock and K. A. Lennox. The calculation of expectations for classes of diffusion processes by Lie symmetry methods. *Ann. Appl. Probab.*, 19(1):127–157, 2009.
- [CLG88] H. F. Chen, G. Lei, and A. J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27(2):217–231, 1988.
- [CM00] F. Chaabane and F. z. Maaouia. Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Probab. Statist.*, 4:137–189, 2000.
- [CPV13] K. Christensen, M. Podolskij, and M. Vetter. On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. *J. Multivariate Anal.*, 120:59–84, 2013.
- [CR07] S. Cohen and J. Rosiński. Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1):195–210, 2007.
- [CT04] R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.
- [CZ86] H. F. Chen and Y. M. Zhu. Stochastic approximation procedures with randomly varying truncations. *Sci. Sinica Ser. A*, 29(9):914–926, 1986.
- [De 14] M. d. R. De Chaumaray. Large deviations for the squared radial Ornstein-Uhlenbeck process. *ArXiv e-prints*, July 2014.
- [Dia13] E. H. A. Dia. Error bounds for small jumps of Lévy processes. *Adv. in Appl. Probab.*, 45(1):86–105, 2013.
- [DL16] S. Dereich and S. Li. Multilevel Monte Carlo for Lévy-driven SDEs: central limit theorems for adaptive Euler schemes. *Ann. Appl. Probab.*, 26(1):136–185, 2016.
- [DN01] D. Darrell and G. Nicolae. Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, 57(1):41–59, 2001.
- [DR15] K. Debrabant and A. Rößler. On the acceleration of the multi-level Monte Carlo method. *J. Appl. Probab.*, 52(2):307–322, 2015.

- [Duf96] M. Duflo. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996.
- [Fil01] D. Filipović. A general characterization of one factor affine term structure models. *Finance Stoch.*, 5(3):389–412, 2001.
- [FK12] H. Fathallah and A. Kebaier. Weighted limit theorems for continuous-time vector martingales with explosive and mixed growth. *Stoch. Anal. Appl.*, 30(2):238–257, 2012.
- [FM12] N. Frikha and S. Menozzi. Concentration bounds for stochastic approximations. *Electron. Commun. Probab.*, 17:no. 47, 15, 2012.
- [FT91] E. Fournié and D. Talay. Application de la statistique des diffusions à un modèle de taux d’intérêt. *Finance*, 12(2):79–111, 1991.
- [GG14] A. Gnoatto and M. Grasselli. The explicit Laplace transform for the Wishart process. *J. Appl. Probab.*, 51(3):640–656, 2014.
- [Gil08] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.
- [Gio17] D. Giorgi. *Théorèmes limites pour estimateurs Multilevel avec et sans poids. Comparaison et applications*. PhD thesis, Université Paris VI, 2017.
- [GJR14] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *ArXiv e-prints*, October 2014.
- [GJY03] A. Göing-Jaeschke and M. Yor. A survey and some generalizations of Bessel processes. *Bernoulli*, 9(2):313–349, 2003.
- [GL94] J. G. Gaines and T. J. Lyons. Random generation of stochastic area integrals. *SIAM J. Appl. Math.*, 54(4):1132–1146, 1994.
- [Gob01] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7(6):899–912, 2001.
- [Gob02] E. Gobet. LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):711–737, 2002.
- [GS14] M. B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.*, 24(4):1585–1620, 2014.
- [Hei01] S. Heinrich. Multilevel monte carlo methods. *Lecture Notes in Computer Science, Springer-Verlag*, 2179(1):58–67, 2001.
- [HP02] C. Houdré and N. Privault. Concentration and deviation inequalities in infinite dimensions via covariance representations. *Bernoulli*, 8(6):697–720, 2002.

- [IW81] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 1981.
- [Jeg82] P. Jeganathan. On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A*, 44(2):173–212, 1982.
- [JK17] B. Jourdain and A. Kebaier. Non-asymptotic error bounds for The Multilevel Monte Carlo Euler method applied to SDEs with constant diffusion coefficient. *ArXiv e-prints*, August 2017.
- [JL09] B. Jourdain and J. Lelong. Robust adaptive importance sampling for normal random vectors. *Annals of Applied Probability*, 19 (5):1687–1718, 10 2009.
- [JMS16] Y. Jiao, C. Ma, and S. Scotti. Alpha-CIR Model with Branching Processes in Sovereign Interest Rate Modelling. *ArXiv e-prints*, February 2016.
- [JP98] J. Jacod and P. Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, 26(1):267–307, 1998.
- [JS03] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2003.
- [Kaw09] R. Kawai. Optimal importance sampling parameter search for Lévy processes via stochastic approximation. *SIAM J. Numer. Anal.*, 47(1):293–307, 2008/09.
- [Kaw13] R. Kawai. Local asymptotic normality property for Ornstein-Uhlenbeck processes with jumps under discrete sampling. *J. Theoret. Probab.*, 26(4):932–967, 2013.
- [Keb05] A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4):2681–2705, 2005.
- [KHLY17] A. Kohatsu-Higa, A. Lejay, and K. Yasuda. Weak rate of convergence of the Euler–Maruyama scheme for stochastic differential equations with non-regular drift. *J. Comput. Appl. Math.*, 326:138–158, 2017.
- [KHNT14] A. Kohatsu-Higa, E. Nualart, and N. K. Tran. LAN property for a simple Lévy process. *C. R. Math. Acad. Sci. Paris*, 352(10):859–864, 2014.
- [KHNT17] A. Kohatsu-Higa, E. Nualart, and N. K. Tran. LAN property for an ergodic diffusion with jumps. *Statistics*, 51(2):419–454, 2017.
- [KHP02] A. Kohatsu-Higa and R. Pettersson. Variance reduction methods for simulation of densities on Wiener space. *SIAM J. Numer. Anal.*, 40(2):431–450 (electronic), 2002.

- [KKH08] A. Kebaier and A. Kohatsu-Higa. An optimal control variance reduction method for density estimation. *Stochastic Process. Appl.*, 118(12):2143–2180, 2008.
- [KKK10] R. Korn, E. Korn, and G. Kroisandt. *Monte Carlo methods and models in finance and insurance*. Chapman & Hall/CRC Financial Mathematics Series. CRC Press, Boca Raton, FL, 2010.
- [KL17] A. Kebaier and J. Lelong. Coupling Importance Sampling and Multilevel Monte Carlo using Sample Average Approximation. *Methodology and Computing in Applied Probability*, (in press), 2017.
- [KLSr12] M. Kessler, A. Lindner, and M. Sørensen, editors. *Statistical methods for stochastic differential equations*, volume 124 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2012. Revised papers from the 7th Séminaire Européen de Statistique on Statistics for Stochastic Differential Equations Models held in Cartagena, May 7–12, 2007.
- [KP91] T. G. Kurtz and P. Protter. Wong-Zakai corrections, random evolutions, and simulation schemes for SDEs. In *Stochastic analysis*, pages 331–346. Academic Press, Boston, MA, 1991.
- [KP95] P. E. Kloeden and E. Platen. Numerical methods for stochastic differential equations. In *Nonlinear dynamics and stochastic mechanics*, CRC Math. Model. Ser., pages 437–461. CRC, Boca Raton, FL, 1995.
- [KP16] A. Kyprianou and S. Palau. Extinction properties of multi-type continuous-state branching processes. *ArXiv e-prints*, April 2016.
- [KR08] M. Keller-Ressel. *Affine Processes - Theory and Applications in Finance*. PhD thesis, Vienna University of Technology, 2008.
- [KRS08] M. Keller-Ressel and T. Steiner. Yield curve shapes and the asymptotic short rate distribution in affine one-factor models. *Finance Stoch.*, 12(2):149–172, 2008.
- [KS85] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 32(1):1–76, 1985.
- [Kut04] Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2004.
- [KY03] H. J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [LC60] L. Le Cam. Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. california Publ. Statist.*, 3:37–98, 1960.



- [LCY90] L. Le Cam and G. L. Yang. *Asymptotics in statistics*. Springer Series in Statistics. Springer-Verlag, New York, 1990. Some basic concepts.
- [Lel08] J. Lelong. Almost sure convergence for randomly truncated stochastic algorithms under verifiable conditions. *Statist. Probab. Lett.*, 78(16):2632–2636, 2008.
- [LL11] B. Lapeyre and J. Lelong. A framework for adaptive Monte Carlo procedures. *Monte Carlo Methods Appl.*, 17(1):77–98, 2011.
- [LLP13] S. Laruelle, C. Lehalle, and G. Pagès. Optimal posting price of limit orders: learning by trading. *Math. Financ. Econ.*, 7(3):359–403, 2013.
- [LP10] V. Lemaire and G. Pagès. Unconstrained recursive importance sampling. *Ann. Appl. Probab.*, 20(3):1029–1067, 2010.
- [LP17] V. Lemaire and G. Pagès. Multilevel Richardson-Romberg extrapolation. *Bernoulli*, 23(4A):2643–2692, 2017.
- [LS77] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes. I,II*. Springer-Verlag, New York, 1977. General theory, Translated by A. B. Aries, Applications of Mathematics, Vol. 5.
- [LS01] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes. I,II*. Applications of Mathematics (New York). Springer-Verlag, Berlin, expanded edition, 2001. Applications, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.
- [LT01] B. Lapeyre and E. Temam. Competitive Monte carlo methods for the pricing of Asian options. *Journal of Computational Finance*, 5(1), 2001.
- [Mai12] H. Mai. *Drift estimation for jump diffusions*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2012.
- [May12] E. Mayerhofer. Wishart Processes and Wishart Distributions: An Affine Processes Point of View. *ArXiv e-prints*, January 2012.
- [Mil74] G. N. Milsteein. Approximate integration of stochastic differential equations. *Teor. Veroyatnost. i Primenen.*, 19:583–588, 1974.
- [Mil95] G. N. Milstein. *Numerical integration of stochastic differential equations*, volume 313 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.
- [MT05] F. Maaouia and A. Touati. Identification of multitype branching processes. *Ann. Statist.*, 33(6):2655–2694, 2005.
- [Nua95] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

- [NV08] S. Ninomiya and N. Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance*, 15(1-2):107–121, 2008.
- [Ogi14] T. Ogihara. Parametric Inference for Nonsynchronously Observed Diffusion Processes in the Presence of Market Microstructure Noise. *ArXiv e-prints*, December 2014.
- [OR97] L. Overbeck and T. Rydén. Estimation in the Cox-Ingersoll-Ross model. *Econometric Theory*, 13(3):430–461, 1997.
- [Ove98] L. Overbeck. Estimation for continuous branching processes. *Scand. J. Statist.*, 25(1):111–126, 1998.
- [Pag01] G. Pagès. Sur quelques algorithmes récurrents pour les probabilités numériques. *ESAIM Probab. Statist.*, 5:141–170 (electronic), 2001.
- [Pag07] G. Pagès. Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods Appl.*, 13(1):37–70, 2007.
- [Pin72] M. A. Pinsky. Limit theorems for continuous state branching processes with immigration. *Bull. Amer. Math. Soc.*, 78:242–244, 1972.
- [PV09] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658, 2009.
- [PY81] J. Pitman and M. Yor. Bessel processes and infinitely divisible laws. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, volume 851 of *Lecture Notes in Math.*, pages 285–370. Springer, Berlin, 1981.
- [PY82] J. Pitman and M. Yor. A decomposition of Bessel bridges. *Z. Wahrsch. Verw. Gebiete*, 59(4):425–457, 1982.
- [RW01] T. Rydén and M. Wiktorsson. On the simulation of iterated Itô integrals. *Stochastic Process. Appl.*, 91(1):151–168, 2001.
- [RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [Sat13] K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation.
- [Tou91] A. Touati. Sur la convergence en loi fonctionnelle de suites de semimartingales vers un mélange de mouvements browniens. *Teor. Veroyatnost. i Primenen.*, 36(4):744–763, 1991.

- [Tra17] N. K. Tran. LAN property for an ergodic Ornstein-Uhlenbeck process with Poisson jumps. *Comm. Statist. Theory Methods*, 46(16):7942–7968, 2017.
- [TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509, 1990.
- [Wik01] M. Wiktorsson. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Ann. Appl. Probab.*, 11(2):470–487, 2001.
- [Yos92] N. Yoshida. Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41(2):220 – 242, 1992.
- [ZMAS05] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472):1394–1411, 2005.