Stochastic approximation, Markovian perturbation and their applications

Noufel Frikha

Habilitation à Diriger des Recherches

Spécialité : **Mathématiques Appliquées**

présentée
par

Noufel Frikha

Stochastic approximation, Markovian perturbation of stochastic processes and their applications.

soutenue publiquement le 23 Novembre 2017 devant le jury composé de :

Mme	Annie Millet,	Professeur à l'université Paris 1,	Examinateur
M.	Vlad Bally,	Professeur à l'université Paris Est,	Examinateur
M.	Dan Crisan,	Professeur à l'Imperial College,	Rapporteur
M.	François Delarue,	Professeur à l'université Nice-Sophia Antipolis,	Examinateur
M.	Gilles Pagès,	Professeur à l'université Pierre et Marie Curie,	Examinateur
M.	Denis Talay,	Directeur de recherche à l'INRIA,	Rapporteur
M.	Peter Tankov,	Professeur à l'ENSAE,	Rapporteur interne

Maturité de l'homme:
cela veut dire avoir retrouvé le sérieux qu'enfant, on mettait dans ses jeux.

Friedrich Nietzsche, par delà le bien et le mal, maximes et intermèdes.

Remerciements

En tout premier lieu, je tiens à remercier içi les rapporteurs de mon mémoire d'HDR, Dan Crisan, Denis Talay et Peter Tankov qui me font le grand honneur d'avoir accepté d'évaluer mes travaux. Je vous suis particulièrement reconnaissant pour vos remarques et pour l'intérêt dont vous avez témoigné pour mon travail. Je remercie chaleureusement les autres membres du jury, Annie Millet, Vlad Bally, François Delarue et Gilles Pagès. Votre présence aujourd'hui ainsi que votre bienveillance appellent ma très grande reconnaissance.

Je remercie particulièrement Annie Millet. La grande clarté de son cours de Master "Méthodes de Monte Carlo", son écoute ainsi que sa disponibilité sont pour moi des exemples. Elle a su donner à chacun de nos entretiens une note de gentillesse qui m'a beaucoup touchée. Je remercie Peter Tankov pour sa constante disponibilité, l'intérêt dont il a fait preuve à l'égard de l'ensemble de mes travaux et ses précieux conseils. Je remercie également Vlad Bally dont les travaux sur le calcul de Malliavin appliqué aux équations différentielles stochastiques et l'interprétation probabiliste de la méthode parametrix m'ont beaucoup apportés ces dernières années. Je tiens à remercier chaleureusement François Delarue dont la qualité et la profondeur des travaux, notamment ceux sur les jeux à champs moyens, éveillent constamment ma curiosité. Enfin, en dernier lieu, je souhaite exprimer içi toute ma gratitude à Gilles Pagès, directeur de thèse exemplaire, pour son enthousiasme et humour communicatifs. Je ferai de mon mieux pour suivre vos traces avec mes futurs élèves en thèse!

Je tiens également à remercier içi l'ensemble de mes collaborateurs mathématiciens, Olivier Bardou, Vincent Lemaire, Abass Sagna, Stéphane Menozzi, Max Fathi, Lorick Huang, Arturo Kohatsu-Higa et Libo Li. Vous m'avez beaucoup apporté, chacun à votre propre manière, et ce travail vous doit beaucoup. Je suis sûr que nous aurons l'occasion de nous retrouver et de retravailler ensemble!

Je tiens à exprimer ma sincère reconnaissance à l'ensemble de mes collègues du LPMA, et particulièrement l'équipe MathFiProNum. Je vous remercie pour votre bienveillance et sympathie quotidiennes. C'est aussi grâce à vous qu'il m'est si agréable de venir travailler jour après jour.

Pour terminer, je souhaite exprimer toute ma gratitude à ma (petite) famille et mes amis ainsi qu'à tous ceux qui ont su me soutenir ces dernières années.

Outline of the manuscript

This document presents and put in perspective my research work carried out after my Phd Thesis since my position as an associate professor at the university Paris Diderot (Paris VII) started in September 2012.

This work consists in two topics that interact with one another. The first part concerns stochastic approximation algorithm with some applications in mathematical finance. The second part focuses on Markovian perturbation of stochastic processes with several applications such that probabilistic representation, unbiased Monte Carlo simulation, density estimates, asymptotic expansions of some Markov semigroups, weak approximation error of some discretisation schemes.

The first four chapters deals with stochastic approximation theory. The first chapter is an introduction. We briefly present two problems arising in numerical probability. The first one deals with stochastic approximation algorithm, also referred in the literature as Robbins-Monro algorithms, which are simulation based recursive schemes that are widely used in stochastic optimization. We review results about $a.s.$ convergence, strong and weak convergence rates and present some applications in mathematical finance. The second one concerns several problems related to the discretisation of stochastic processes. We present the main results available in the literature and the different tools to obtain them, with a special emphasis on the weak discretisation error.

Chapter 2 deals with some contributions to risk management in financial and energy markets. It summarises the results contained in **[5]** and **[6]**. First, we study theoretical and computational aspects of risk minimisation in financial market models operating in discrete time **[6]**. To define the risk, we consider a class of convex risk measures defined on L^p in terms of shortfall risk. A dynamic programming principle in a non-Markovian framework is established, under mild assumptions, namely the absence of arbitrage opportunity and the non-degeneracy of the price process. Moreover, we provide a probabilistic algorithm to compute the optimal dynamic strategy using optimal quantization and analyse the global error between the true value function and the approximated one.

As an alternative method to simulation based recursive importance sampling procedure to estimate the optimal change of measure for Monte Carlo simulations, we propose an algorithm which combines (vector and functional) optimal quantization with Newton-Raphson zero search procedure **[5]**. We analyse the error induced by the proposed procedure and illustrate its efficiency on several examples.

In Chapter 3, we focus on non-asymptotic concentration bounds, deviation estimates and related transport entropy inequalities for stochastic approximation schemes **[7]**, **[8]**. Our results include the deviation between the expectation of a given function of an Euler like discretization scheme of some diffusion process at a fixed deterministic time and its empirical mean obtained by the Monte-Carlo procedure as well as the deviation between the value at a given time of a stochastic approximation algorithm and its target. Under suitable assumptions both concentration bounds turn out to be Gaussian **[7]**. In order to obtain different non-Gaussian concentration regimes, one has to quantify the contribution of the diffusion term to the concentration regime **[8]**. We also derive a general non-asymptotic deviation bound for the difference between a function of the trajectory of a continuous Euler scheme associated to a diffusion process and its mean. Finally, we obtain non-asymptotic bound for stochastic approximation with averaging of trajectories. Notably, it is shown that averaging a stochastic approximation algorithm with a slow decreasing step sequence provide the optimal concentration rate.

In Chapter 4, we present a class of multi-level and multi-step Richardson Romberg extrapolation methods for stochastic approximation algorithms **[9]**, **[10]**. In a first part, we extend the scope of the multilevel Monte Carlo method recently introduced for the expectation of a given function of some diffusion process to the framework of stochastic optimisation by means of stochastic approximation algorithm. A central limit theorem is proved and the possible optimal choices of step size sequence are provided **[9]**. In a second part, we also extend the well-known Richardson-Romberg extrapolation method for Monte Carlo linear estimator to the framework of stochastic approximation algorithm **[10]**. We illustrate the method to the estimation of the quantile of diffusion processes. For both methods, numerical results confirm the theoretical analysis and show a significant reduction in the initial computational cost.

The next chapters are related to the second part of my research work which focuses on Markovian perturbation of stochastic processes and its application.

Chapter 5 is dedicated to two approximation problems **[12]**, **[11]**. We begin by a brief presentation of the parametrix

method "à la Mc Kean and Singer" in the framework of diffusion processes. This perturbation method is a classical technique in PDE theory to construct the fundamental solution to an elliptic or parabolic PDE. Our presentation here is based on probabilistic arguments. We also briefly present its discrete counterpart when one considers discretisation schemes of stochastic processes.

In the second part of this chapter, we consider a discrete version of the parametrix technique that we apply in order to study the weak approximation error of a one-dimensional skew diffusion with bounded measurable drift and Hölder diffusion coefficient by an Euler-type scheme, which consists of iteratively simulating skew Brownian motions with constant drift. First, we establish two sided Gaussian bounds for the density of this approximation scheme. Then, a bound for the difference between the densities of the skew diffusion and its Euler approximation is obtained under the same assumption **[12]**.

In the third part, we study the asymptotic expansion of a Markov semigroup with respect to a small parameter **[11]**. The method proposed combines a perturbation approach of Markov semigroup and Malliavin's calculus. We study two applications. In the first one, heat kernel expansions of some hypo-elliptic diffusion processes are revisited. In the second one, we obtain an asymptotic expansion of the transition density of a skew diffusion process with small skew.

Chapter 6 is divided into two parts. In the first part, we focus on the regularity of the law associated to the first hitting time of a threshold by a one-dimensional uniformly elliptic diffusion process and to the associated process stopped at the threshold **[13]**. We present explicit expressions for the corresponding transition densities and studies their regularity properties up to the boundary under mild assumptions on the coefficients. Some Gaussian upper-estimates for these densities (and their derivatives) are also established. Our construction allows to obtain a probabilistic representation that can be used for the construction of an unbiased Monte Carlo simulation method and also to derive some integration by parts formula.

In the second part, we investigate the weak existence and uniqueness for some SDEs with coefficients depending on some of its path-functionals. A general method is developed in order to deal with a process whose law is singular with respect to the Lebesgue measure. We illustrate our approach on two examples: a diffusion process with coefficients depending on its running local time and a diffusion process with coefficients depending on its running maximum. We also construct the associated transition density and establish some Gaussian upper-estimates.

List of research works presented in this manuscript:

Publications related to my Phd Thesis:

[1] with O. Bardou, G. Pagès: Recursive Computation of Value-at-Risk and Conditional Value-at-Risk using MC and QMC, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, Part 3, 193-208, 2008.

[2] with O. Bardou, G. Pagès: Computing VaR and CVaR using Stochastic Approximation and Adaptive Unconstrained Importance Sampling , *Monte Carlo Methods and Applications*, 15(3):173-210, 2009.

[3] with O. Bardou, G. Pagès: CVaR hedging using quantization based stochastic approximation algorithm, *Mathematical Finance*, 26(1):184-229, 2016.

[4] with V. Lemaire: Joint modelling of Gas and Electricity spot prices, *Applied Mathematical Finance*, 20(1), 69-93, 2013.

Publications after my Phd Thesis:

[5] with A. Sagna: Quantization based Recursive Importance Sampling, *Monte Carlo Methods and Applications*, 18(4):287-326, 2012.

[6] Shortfall risk minimization in discrete time financial market models, *SIAM Journal on Financial Mathematics*, Volume 5, Issue 1, 384-414, 2012.

[7] with S. Menozzi: Concentration Bounds for Stochastic Approximations, *Electronic Communications in Probability*, Volume 17, no. 47, 1-15, 2012.

[8] with M. Fathi: Transport-Entropy inequalities and deviation estimates for stochastic approximation schemes, *Electronic Journal of Probability*, Volume 18, no. 67, 1-36, 2013.

[9] Multi-level stochastic approximation algorithms, *The Annals of Applied Probability*, Volume 26, Issue 2, 933-985, 2016.

[10] with L. Huang: A multi-step richardson-romberg extrapolation method for stochastic approximation, *Stochastic Processes and their Applications*, Volume 125, Issue 11, 4066-4101, 2015.

[11] with A. Kohatsu-Higa: A Parametrix approach for asymptotic expansion of Markov semigroups with applications to multi-dimensional diffusion processes, 50 pages, submitted, 2015.

[12] On the weak approximation of a skew diffusion by an Euler-type scheme, accepted for publication in *Bernoulli*, 28 pages, 2016.

[13] with A. Kohatsu-Higa, L. Li: On the first hitting times of one dimensional elliptic diffusions, 41 pages, submitted, 2016.

[14] with L. Li: Weak existence and uniqueness for some stochastic differential equations with path-functional coefficients, 26 pages, submitted, 2017.

Contents

Chapter 1

Introduction : On two problems arising in numerical probability

In this introductory chapter, we briefly present two problems arising in numerical probability. The first part deals with a brief presentation of the stochastic approximation algorithm theory. Since their introduction in the seminal paper of H. Robbins and S. Monro [117] , these simulation based procedures are widely used in various fields such that stochastic optimisation or zero search problems. The stochastic approximation theory provides various theorems that guarantee the $a.s.$, weak and/or strong convergence rates towards the desired target. We present here some results related to the $a.s.$ convergence, known as the Robbins-Siegmund theorem and the ordinary differential equation (ODE) method. We provide the key points for their proofs. Then, we review strong and weak convergence rates. Such algorithm satisfies a central limit theorem (CLT) with a rate that depends on the step sequence. Finally, we present some applications in mathematical finance.

In the second part, we briefly present several problems related to the discretisation of diffusion-jump type stochastic processes. We distinguish between strong and weak approximation since these two problems are of different nature and require different mathematical tools for their study. We first investigate the strong approximation error which is related to the error between the paths of the process and of its approximation scheme. Then, we focus on the weak approximation error which is the error between the law of the process and of its approximation scheme. This error is often estimated on a class of test functions. Since the pioneering paper of Talay and Tubaro [122], we know that there is a strong connection between the weak approximation error and the solution of the underlying parabolic linear partial differential equation (PDE) via the Feynman-Kac representation.

We try to gather major results obtained the past thirty years since they will be useful in order to understand some of the next chapters.

We will focus on the weak approximation and the mathematical tools needed in order to study the resulting error. More precisely, we will distinguish between the two following cases: when the test function and the coefficients of the stochastic differential equation are smooth; when the process is non-degenerated. This will be the topic of Section 5.2 in Chapter 5 which deals with the weak approximation error of a skew diffusion by an Euler-type scheme.

We briefly present the Multi-level Monte Carlo method developed by Giles [48] which allows to significantly increase the computational efficiency of the Monte Carlo method for the computation of the expectation of a non-simulatable random variable. We extend the scope of this method to the framework of stochastic optimisation by means of stochastic approximation algorithm in Chapter 4.

We conclude this chapter by presenting an unbiased simulation technique for stochastic differential equations based on a probabilistic representation of the parametrix technique that has been recently developed Bally and Kohatsu-Higa [7]. In Chapter 6, we will propose a similar probabilistic representation for hitting times and associated killed process of one-dimensional elliptic diffusions.

1.1 Stochastic approximation algorithm

1.1.1 Introduction

In applied mathematics, one often has to numerically solve some optimisation or zero search problems. The former problem can often be cast into the latter one when one faces a convex optimisation problem. Some commonly encountered examples are the computation of the quantile at level $\alpha \in (0,1)$ of probability distribution, in mathematical finance the extraction of implicit volatility or the minimisation of a convex function with respect to a parameter (e.g. quantity of assets in a portfolio, minimisation of the variance, etc).

In a deterministic framework, the zero-search recursive procedure has the following dynamics

$$\theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n), \quad n \geqslant 0 \tag{1.1}$$

where $h : \mathbb{R}^d \to \mathbb{R}^d$ is a continuous function satisfying a sub-linear growth assumption and $(\gamma_n)_{n\geqslant 1}$ is a non-negative non-increasing step sequence such that $\sum_{n\geqslant 1} \gamma_n = +\infty$.

Under an appropriate *mean-reverting assumption* on the function h, one shows that the deterministic sequence $(\theta_n)_{n\geqslant 1}$ defined by (1.1) is bounded and eventually converges toward a zero θ^* of h. More precisely, assuming that the following *mean-reverting assumption*

$$\langle \theta - \theta^*, h(\theta) \rangle \geqslant 0, \quad \forall \theta \neq \theta^* \tag{1.2}$$

and that $\{h = 0\} = \{\theta^*\}$, then one proves that $(\theta_n)_{n\geqslant 1}$ converges towards θ^* as $n \to \infty$. An example of such procedure is the Newton-Raphson zero-search algorithm obtained by choosing $h := Dh^{-1}h$ and $\gamma_n = 1$.

However, in a stochastic optimization framework as described in the examples mentioned above, one does not have straightforward access to the value of $h(\theta)$ because h is given by the expectation of some given random function, say

$$h(\theta) := \mathbb{E}[H(\theta, U)]$$

where $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ is a measurable function and U is an $\mathbb{R}^q$-valued random variable with law $\mu \in \mathcal{P}(\mathbb{R}^q)$. Here it is implicitly assumed that one can easily simulate according to the law μ and has direct access to the value of the function H. One simple idea could be to simply approximate the above dynamics (1.1) by using at each iterate a Monte Carlo approximation of $h(\theta_n)$. A more interesting idea consists in doing both simultaneously by using on the one hand $H(\theta_n, U^{(n+1)})$ instead of $h(\theta_n)$ and on the other hand by letting the step sequence goes to zero not too fast to take into account the additional variance induced by this randomisation. Hence, one is led to introduce the following dynamics

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, U^{(n+1)}), \quad n \geqslant 0 \tag{1.3}$$

where $(U^{(n)})_{n\geqslant 1}$ is an i.i.d. sequence of random vector distributed according to μ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and θ_0 is an $\mathbb{R}^d$-valued random vector independent of the sequence $(U^{(n)})_{n\geqslant 1}$.

Example:
As a first simple example, consider the empirical average $(\bar{U}^n)_{n\geqslant 1}$ of *i.i.d.* and integrable samples $(U^{(n)})_{n\geqslant 1}$ defined by $\bar{U}^n := \frac{1}{n}\sum_{k=1}^n U^{(k)}$, $\bar{U}^0 = 0$. We observe that it satisfies the following dynamics

$$\bar{U}^{n+1} = \bar{U}^n - \frac{1}{n+1}\left(\bar{U}^n - U^{(n+1)}\right) = \bar{U}^n - \frac{1}{n+1} H(\bar{U}^n, U^{(n+1)}), \quad n \geqslant 0, \tag{1.4}$$

where $H(\theta, U) := \theta - U$ and $h(\theta) := \theta - \mathbb{E}[U]$, so that $\theta^* = \mathbb{E}[U]$. Moreover, it is well-known that the sequence $(\bar{U}^n)_{n\geqslant 1}$ given by (1.4) converges *a.s.* to the unique zero of h. When one knows how to sample from the distribution μ, this procedure is easy to implement contrary to its deterministic counterpart (1.1). We see that stochastic approximation appears as a natural extension of the Monte Carlo method to solve optimisation or zero search problems when the function of interest h writes as the integral of some function against a law μ from which it is easy to sample.

Let us make an important remark. Writing $H(\theta_n, U^{(n+1)}) = h(\theta_n) + \Delta M_{n+1}$ where $\Delta M_{n+1} := H(\theta_n, U^{(n+1)}) - h(\theta_n)$ is a martingale increment with respect to the natural filtration $\{\mathcal{F}_n\}$, $\mathcal{F}_n := \sigma(\theta_0, U^{(1)}, \cdots, U^{(n)})$, $n \geqslant 1$ induced by the algorithm, we observe that the dynamics (1.3) can be seen as a perturbed Euler scheme (with decreasing step) of the ordinary differential equation (ODE)

$$\dot{\theta} = -h(\theta) \tag{1.5}$$

From this connection it may be motivating to use some ODE techniques such that existence of a *Lyapunov function* in order to study the asymptotic behaviour of the recursive scheme (1.3). A Lyapunov function for the ODE (1.5) is a function $L : \mathbb{R}^d \to \mathbb{R}_+$ such that any solution $t \mapsto \theta(t)$ satisfies $t \mapsto L(\theta(t))$ is non-increasing. Assuming that L is continuously differentiable, it is mainly equivalent to the condition $\langle \nabla L, h \rangle \geqslant 0$ since $dL(\theta(t)) = \langle \nabla L(\theta(t)), \theta(t) \rangle = -\langle \nabla L, h \rangle(\theta(t))dt$. Finding a Lyapunov function may be a difficult task and there is no result that guarantees its existence! Usually it requires a deep knowledge and understanding of the underlying dynamic (1.5). However, if such function does exist, the system (1.5) is said to be dissipative. We will have in mind two commonly encountered situations:

- when one faces a convex optimisation problem, the recursive scheme (1.3) is usually a stochastic gradient algorithm. The function L is the object of interest that is one has $h = \nabla L$ and one is computing a minimum of L.
- Unfortunately, as is so often the case when dealing with stochastic approximation algorithm, the function L is not identified *a priori* and one has to find a Lyapunov function associated to the function h, if such function exists!

1.1.2 Standard *a.s.* convergence results

We now turn our attention to results related to the *a.s.* and/or L^p convergence of the recursive procedure (1.3). We first provide a general result known as *Robbins-Siegmund Lemma* based on the existence of an *a priori* Lyapunov function from which we will deduce two corollaries that ensures the *a.s.* (and L^p) convergence of the sequence $(\theta_n)_{n\geqslant 1}$. We refer the interested reader to the monographs of Benveniste and al. [19], Duflo [39] and Kushner and Yin [81] among others for developments and a more complete overview in stochastic approximation theory.

Theorem 1.1.1 (Robbins-Siegmund Lemma)**.** *Assume that there exists a continuously differentiable function* $L : \mathbb{R}^d \to \mathbb{R}_+$ *such that* ∇L *is Lipschitz-continuous and*

$$|\nabla L|^2 \leqslant C(1+L) \tag{1.6}$$

and such that the mean-reverting assumption *holds*

$$\langle \nabla L, h\rangle \geqslant 0. \tag{1.7}$$

Assume furthermore that H *satisfies the following* sub-linear growth assumption

$$\forall \theta \in \mathbb{R}^d, \quad \mathbb{E}[|H(\theta,U)|^2]^{\frac{1}{2}} \leqslant C(1+L(\theta))^{\frac{1}{2}}, \tag{1.8}$$

for some positive constant C*.*

Let $(\gamma_n)_{n\geqslant 1}$ *be a positive step sequence satisfying the following* decreasing step assumption

$$\sum_{n\geqslant 1}\gamma_n = \infty \quad \textit{and} \quad \sum_{n\geqslant 1}\gamma_n^2 < \infty. \tag{1.9}$$

Finally assume that θ_0 *is independent of the innovation sequence* $(U^{(n)})_{n\geqslant 1}$ *and satisfies* $\mathbb{E}[L(\theta_0)] < \infty$*. Then, the sequence* $(\theta_n)_{n\geqslant 1}$ *defined by the recursive procedure* (1.3) *satisfies the following properties:*

1. *the sequence* $(\theta_n - \theta_{n-1})_{n\geqslant 1}$ *converges a.s. and in* $L^2(\mathbb{P})$ *to zero as* n *goes infinity,*
2. *the sequence* $(L(\theta_n))_{n\geqslant 1}$ *is* $L^1(\mathbb{P})$*-bounded,*
3. $L(\theta_n) \to L_\infty \in L^1(\mathbb{P})$ *a.s. as* $n \to \infty$*.*
4. $\sum_{n\geqslant 1}\gamma_n\langle \nabla L, h\rangle(\theta_{n-1}) < \infty \quad$ *a.s.*

Proof. The key ingredient of the proof is the convergence theorem for non-negative super-martingale. We only sketch the proof. Since ∇L is Lipschitz-continuous, by the fundamental theorem of calculus, one has

$$\begin{aligned} L(\theta_{n+1}) &\leqslant L(\theta_n) + \gamma_{n+1}\langle \nabla L(\theta_n), \theta_{n+1}-\theta_n\rangle + C|\theta_{n+1}-\theta_n|^2 \\ &= L(\theta_n) - \gamma_{n+1}\langle \nabla L, h\rangle(\theta_n) - \gamma_{n+1}\langle \nabla L(\theta_n), \Delta M_{n+1}\rangle + C\gamma_{n+1}^2|H(\theta_n, U^{(n+1)}|^2 \end{aligned} \tag{1.10}$$

where we recall that $\Delta M_{n+1} = H(\theta_n, U^{(n+1)}) - h(\theta_n)$ is an $\{\mathcal{F}_n\}$-martingale increment. Using the above inequality together with assumptions (1.7), (1.8) and (1.9), one proves that $L(\theta_n) \in L^1(\mathbb{P})$ so that $\Delta M_{n+1} \in L^2(\mathbb{P})$, $n \geqslant 0$. The key idea is to introduce the following sequence

$$S_n := \frac{L(\theta_n) + \sum_{k=1}^n \gamma_k\langle \nabla L, h\rangle(\theta_{k-1}) + C\sum_{k\geqslant n+1}\gamma_k^2}{\prod_{k=1}^n(1+C\gamma_k^2)}, \quad n \geqslant 1$$

with $S_0 = L(\theta_0)$. From assumption (1.7) and using (1.10), one proves that $(S_n)_{n\geqslant 0}$ is a non-negative super-martingale so that it converges *a.s.* toward an integrable random variable S_∞. As a by-product of this convergence, one deduces the L^1-boundedness of $(L(\theta_n))_{n\geqslant 0}$ and that $\sum_{n\geqslant 1}\gamma_n\langle \nabla L, h\rangle(\theta_{n-1}) < \infty$ *a.s.* Then one proves that $L(\theta_n) \to L_\infty$, *a.s.* and $\sum_{n\geqslant 0}\mathbb{E}[|\theta_{n+1}-\theta_n|^2] < \infty$. □

Remark 1.1.1. *We also want to mention that when the innovation sequence* $(U^{(n)})_{n\geqslant 1}$ *is not i.i.d. but only* $\{\mathcal{F}_n\}$*-adapted, the recursive procedure* (1.3) *can be written*

$$\theta_{n+1} = \theta_n - \gamma_{n+1}h(\theta_n) - \gamma_{n+1}(\Delta M_{n+1} + r_{n+1})$$

where $r_{n+1} := h(\theta_n) - \mathbb{E}[H(\theta_n, U^{(n+1)}|\mathcal{F}_n]$ *is an* $\mathcal{F}_{n+1}$*-measurable remainder term. In this case, assuming that*

$$\sum_{n\geqslant 1}\gamma_n|r_n|^2 < \infty, \quad a.s.$$

one still proves that the conclusions of the Robbins-Siegmund Lemma hold.

The next result allows one to conclude that the recursive scheme (1.3) converges toward its target θ^*.

Corollary 1.1.2 (Robbins-Monro & Stochastic gradient algorithms)**.**

(1) [Robbins-Monro algorithm] Assume that the function of interest h satisfies

$$\forall \theta \neq \theta^*, \quad \langle \theta - \theta^*, h(\theta) \rangle > 0, \tag{1.11}$$

and

$$\forall \theta \in \mathbb{R}^d, \quad \mathbb{E}[|H(\theta, U)|^2]^{\frac{1}{2}} \leqslant C(1 + |\theta|) \tag{1.12}$$

with $\theta_0 \in L^2(\mathbb{P})$ independent of the innovation sequence $(U^{(n)})_{n\geqslant 1}$. Assume that the step sequence $(\gamma_n)_{n\geqslant 1}$ satisfies (1.9)*. Then, the sequence $(\theta_n)_{n\geqslant 1}$ defined by* (1.3) *converges a.s. to θ^*, which satisfies $\{h = 0\} = \{\theta^*\}$.*

(2) [Stochastic Gradient algorithm] Assume that $h = \nabla L$ for a convex and continuously differentiable function L satisfying (1.6)*, $\lim_{|\theta|\to\infty} L(\theta) = \infty$, $\{\nabla L = 0\} = \{\theta^*\}$. Assume that $\mathbb{E}[|H(\theta, U)|^2] \leqslant C(1 + L(\theta))$ and that $L(\theta_0) \in L^1(\mathbb{P})$. Assume that the step sequence $(\gamma_n)_{n\geqslant 1}$ satisfies* (1.9)*. Then, $\theta^* = \arg\min_{\theta\in\mathbb{R}^d} h(\theta)$ and $(\theta_n)_{n\geqslant 1}$ converges a.s. toward θ^*.*

Remark 1.1.2. *There is also an alternative to the stochastic gradient algorithm known as the Kiefer-Wolfowitz (K.W.) procedure. In this approach, the function of interest h is given by $h(\theta) = \mathbb{E}[\partial_\theta H(\theta, U)]$. In some practical implementations, the local gradient $\partial_\theta H(\theta, U)$ may be difficult to simulate because of Malliavin weights, log-likelihood ratio, etc, whereas the computation of $H(\theta, U)$ is easy. The K.W. algorithm combines the recursive stochastic approximation algorithm with a finite difference approach to differentiation. Based on the following approximation of the local gradient*

$$\partial_{\theta_i} H(\theta, U) \approx \frac{H(\theta + \eta^i e_i, U) - H(\theta - \eta^i e_i, U)}{2\eta^i}$$

where $(e_i)_{1\leqslant i\leqslant d}$ denotes the canonical basis of $\mathbb{R}^d$ and $|\eta| << 1$, $(\eta^i)_{1\leqslant i\leqslant d} \in \mathbb{R}^d$, one introduces the following recursive scheme

$$\theta^i_{n+1} = \theta^i_n - \gamma_{n+1} \frac{H(\theta_n + \eta^i_{n+1}, U^{(n+1)}) - H(\theta_n - \eta^i_{n+1}, U^{(n+1)})}{2\eta^i_{n+1}}, \quad 1 \leqslant i \leqslant d, \tag{1.13}$$

where $(\gamma_n)_{n\geqslant 1}$ is a step sequence satisfying (1.9) *and $(\eta^i_n)_{n\geqslant 1}$, $1 \leqslant i \leqslant d$, are d non-negative step sequences satisfying for $1 \leqslant i \leqslant d$,*

$$\eta^i_n \to 0, \quad \sum_{n\geqslant 1} \eta^i_n = \infty \quad \text{and} \quad \sum_{n\geqslant 1} (\gamma_n/\eta^i_n) < \infty.$$

We refer to the monograph Benveniste and al. [19] for a result about the a.s. convergence of (1.13)*.*

We conclude by the *ODE method* which establishes a link between the recursive procedure (1.3) and the underlying ODE (1.5) as mentioned above. The main improvement provided by the ODE method is to study the asymptotic behaviour of the sequence $(\theta_n)_{n\geqslant 0}$ (assumed to be bounded a priori) from a sequence function $(\theta^{(k)})_{k\geqslant n}$, $n \geqslant 1$ defined by linear interpolation. Importantly, we will not require the mean reverting assumption (1.11) or the existence of a Lyapunov function that separates the target θ^*. However, the convergence of the procedure will be related to the asymptotic behaviour of the underlying ODE.

We set $\Gamma_0 = 0$ and $\Gamma_n = \sum_{k=1}^n \gamma_k$. Then we define the cádlàg step function $(\theta^{(0)}_t)_{t\geqslant 0}$ by

$$\theta^{(0)}_t := \theta_n, \quad t \in [\Gamma_n, \Gamma_{n+1}).$$

and the sequence of time shifted function

$$\theta^{(n)}_t := \theta^{(0)}_{\Gamma_n + t}, \quad t \in \mathbb{R}_+.$$

For every $t \in \mathbb{R}_+$, we also set $N(t) := \max\{k : \Gamma_k \leqslant t\}$. For every $t \in [\Gamma_n, \Gamma_{n+1})$, the dynamics (1.3) can be written as follows

$$\theta^{(0)}_t = \theta^{(0)}_0 - \int_0^{\Gamma_n} h(\theta^{(0)}_s) ds - \sum_{k=1}^n \gamma_k \Delta M_k$$

which in turn satisfies

$$\begin{aligned} \forall t \in \mathbb{R}_+, \quad \theta^{(0)}_t &= \theta^{(0)}_0 - \int_0^{\Gamma_{N(t)}} h(\theta^{(0)}_s) ds - \sum_{k=1}^{N(t)} \gamma_k \Delta M_k \\ &= \theta^{(0)}_0 - \int_0^t h(\theta^{(0)}_s) ds + \int_{\Gamma_{N(t)}}^t h(\theta^{(0)}_s) ds - \sum_{k=1}^{N(t)} \gamma_k \Delta M_k. \end{aligned}$$

Now by the definition of the shifted function $\theta^{(n)}$, one gets

$$\begin{aligned}\theta_t^{(n)} &= \theta_{\Gamma_n}^{(0)} + \theta_{\Gamma_n+t}^{(0)} - \theta_{\Gamma_n}^{(0)} \\ &= \theta_{\Gamma_n}^{(0)} - \int_{\Gamma_n}^{\Gamma_n+t} h(\theta_s^{(0)})ds + \mathscr{R}^n(t)\end{aligned}$$

where $\mathscr{R}^n(t) := \int_{\Gamma_{N(\Gamma_n+t)}}^{\Gamma_n+t} h(\theta_s^{(0)})ds - \sum_{k=n+1}^{N(\Gamma_n+t)} \gamma_k \Delta M_k$ is a remainder term.

The next result establishes a connection between the asymptotic behaviour of $(\theta_n)_{n\geqslant 0}$ and the asymptotic of the sequence of functions $(\theta^{(n)})_{n\geqslant 0}$.

Proposition 1.1.3. *Let $L : \mathbb{R}^d \to \mathbb{R}_+$ be a continuously differentiable function such that ∇L is Lipschitz-continuous, $\lim_{|\theta|\to+\infty} L(\theta) = +\infty$ and satisfying (1.6). If $|h| \leqslant C(1+L)^{\frac{1}{2}}$ and $\langle \nabla L, h\rangle$ is a non-negative lower semi-continuous function then the limiting values of $(\theta_n)_{n\geqslant 1}$ is a connected component of $\{\langle \nabla L, h\rangle = 0\}$.*

1.1.3 Convergence rate

We now present some results related to the convergence rate of the recursive scheme (1.3) in the *i.i.d.* framework. In standard settings, a stochastic algorithm converges to its target at rate $\sqrt{\gamma_n}$, which suggests to choose $\gamma_n = \gamma/n$, $\gamma > 0$, under assumption (1.9). Under a strong mean-reverting assumption, we will see that this is a sharp *non-asymptotic* $L^2(\mathbb{P})$, which is consistent with the Monte Carlo method. In Chapter 3, we will see that this rate also appears when one investigates concentration inequalities, that is non-asymptotic deviation estimates between $(\theta_n)_{n\geqslant 1}$ and its target. Then, we present the standard central limit theorem (CLT) which establishes that the renormalised sequence $\sqrt{\gamma_n^{-1}}(\theta_n - \theta^*)$ converges in distribution to a normal distribution with mean 0 and a covariance matrix based on $\mathbb{E}[H(\theta^*, U)H(\theta^*, U)^t]$.

Proposition 1.1.4. *Assume that the assumptions of Corollary 1.1.2 (1) (Robbins-Monro algoirthm) are satisfied. Assume that there exists $\alpha > 0$ such that the strong mean reverting assumption holds*

$$\forall \theta \in \mathbb{R}^d, \quad \langle \theta - \theta^*, h(\theta)\rangle \geqslant \alpha|\theta - \theta^*|. \tag{1.14}$$

Then, there exists a positive constant C, such that one has

$$\mathbb{E}[|\theta_n - \theta^*|^2] \leqslant C\gamma_n, \quad n \geqslant 1,$$

for the two following step sequences:

- $\gamma_n = \gamma/n^\beta$, $\beta \in (1/2, 1)$, $n \geqslant 1$.
- $\gamma_n = \gamma/n$, *with* $\gamma > \frac{1}{2\alpha}$. $n \geqslant 1$.

The above result suggests a weak rate given by $\sqrt{\gamma_n^{-1}}$. The study of the asymptotic distribution of the renormalised sequence $\sqrt{\gamma_n^{-1}}(\theta_n - \theta^*)$ led to a considerable and expanding literature starting from some pioneering (independent) works by Bouton and Kushner in the early 1980 until recently, see e.g. Fort [43] or Lelong [90]. We give here a result obtained by Pelletier [113] which has as the main advantage to be local in the sense that the CLT holds on the set of convergence of the algorithm to an equilibrium which makes possible a straightforward application to multi-target algorithms.

Theorem 1.1.5 (Central limit theorem)**.** *Consider the sequence $(\theta_n)_{n\geqslant 1}$ defined by (1.3) and let $\theta^* \in \{h = 0\}$ be one possible target. Assume that the following assumptions hold:*

(1) Attractivity: *θ^* is a strong attractor for the underlying ODE, that is h is twice continuously differentiable at θ^* and $Dh(\theta^*)$ is an Hurwitz matrix, i.e. all its eigenvalues have positive real parts. We denote by λ_m the eigenvalue with the lowest real part.*

(2) Regularity of *H: The function H satisfies the following regularity and growth control property*

$$\theta \mapsto \mathbb{E}[H(\theta, U)H(\theta, U)^t] \text{ is continuous at } \theta^* \text{ and } \theta \mapsto \mathbb{E}[|H(\theta, U)|^{2+\delta}] \text{ is locally bounded on } \mathbb{R}^d$$

for some $\delta > 0$.

(3) Non-degenerate asymptotic variance*: The covariance matrix*

$$\Sigma^* := \mathbb{E}[H(\theta^*, U)(H(\theta^*, U))^t]$$

is a positive definite matrix.

(4) Choice of the step sequence*: The step sequence is given by*

$$\gamma_n = \frac{\gamma}{a + n^\beta}, \quad a, \gamma > 0$$

for some $\beta \in (\frac{1}{2}, 1]$*, with the additional constraint, when* $\beta = 1$*,*

$$\gamma > \frac{1}{2\mathcal{R}e(\lambda_m)}. \tag{1.15}$$

Then, the a.s. convergence is ruled on the event $\{\theta_n \to \theta^*\}$ *by the following* stable *central limit theorem*

$$\gamma_n^{-1/2}(\theta_n - \theta^*) \Rightarrow \mathcal{N}(0, \Gamma), \tag{1.16}$$

where

$$\Gamma := \int_0^\infty (e^{-(Dh(\theta^*) - \frac{I_d}{2c})s})^t \Sigma^* e^{-(Dh(\theta^*) - \frac{I_d}{2c})s} ds$$

with $c = \infty$ *if* $\beta \neq 1$ *and* $c = \gamma$ *if* $\beta = 1$*. Moreover, the convergence is stable, that is, for every bounded continuous function* f *and every* $A \in \mathcal{F}_\infty$*, one has*

$$\mathbb{E}[\mathbf{I}_{\{\theta_n \to \theta^*\} \cap A} f(\sqrt{\gamma_n^{-1}}(\theta_n - \theta^*))] \to \mathbb{E}[\mathbf{I}_{\{\theta_n \to \theta^*\} \cap A} f(\Gamma^{\frac{1}{2}} \zeta))], \quad n \to \infty$$

where $\zeta \sim \mathcal{N}(0, I_d)$*.*

We see that the best rate of convergence is achieve for $\beta = 1$ but we have an additional constraint (1.15) on the constant γ depending on the unknown matrix $Dh(\theta^*)$. In practical implementations, one does not "observe" such constraint but we remark that the algorithm (1.3) spends most of its time exploring the state space. The CLT regime occurs when the algorithm gets trapped after the mentioned search phase.

In order to circumvent this (theoretical) constraint on the step sequence, one can use the *averaging principle*, also known as Ruppert & Polyak principle in stochastic approximation community, in order to achieve (for free!) the best convergence rate. The original motivation was to smoothen the behaviour of the original procedure by considering the empirical mean of the past values up to the nth iteration rather than the computed value at the nth iteration. If one devices this averaging procedure with a slow decreasing step sequence $(\gamma_n)_{n \geqslant 1}$ then one reaches for free the best possible rate of convergence.

More precisely, we consider the step sequence $\gamma_n = (\gamma/(a+n))^\beta$, with $\beta \in (1/2, 1)$ and $n \geqslant 1$ and define the empirical mean

$$\bar{\theta}_n := \frac{\theta_0 + \cdots + \theta_n}{n+1} = \bar{\theta}_{n-1} - \frac{1}{n}(\bar{\theta}_{n-1} - \theta_n), \quad n \geqslant 1,$$

of the sequence $(\theta_n)_{n \geqslant 0}$ defined by (1.3). We observe that by Cesàro's lemma, one has $\bar{\theta}_n \to \theta^*$ *a.s.* if $\theta_n \to \theta^*$ *a.s.* as $n \to \infty$. Then, under the assumptions (1), (2), (3) of Theorem 1.1.5, one has

$$\sqrt{n}(\bar{\theta}_n - \theta^*) \Rightarrow \mathcal{N}(0, \Gamma^*), \quad n \to +\infty$$

where

$$\Gamma^* := Dh(\theta^*)^{-1} \Sigma^* (Dh(\theta^*)^t)^{-1}$$

is the optimal asymptotic covariance matrix.

From a practical point of view, one should not start averaging the initial procedure at the very beginning but wait until the exploration phase is finished.

We have seen some basic results related to *a.s.* convergence and convergence rate of stochastic approximation, whilst other interesting aspects have purposely not been cited. For instance, we will mention *traps*, that is parasitic equilibrium such saddle points local maxima, see e.g. Brandière and Duflo [25], Pemantle [114] for some aspects on how an excited enough parasitic equilibrium point is *a.s.* not a possible limit point for a stochastic approximation algorithm. We also mention stochastic approximation with *averaging innovation* where the innovation sequence $(U^{(n)})_{n \geqslant 1}$ is no longer *i.i.d.* but satisfies some mild ergodic or simply averaging assumption, thus allowing for deterministic sequence *Quasi Monte Carlo stochastic approximation* or to plug exogenous data, like market data, see Laruelle & Pagès [87] and the references therein.

Example: A Robbins-Monro algorithm for quantile estimation

We conclude this section by a simple illustration of the mentioned results to the recursive computation of the α-quantile (or VaR_α) of a real-valued random variable X, with law μ, defined as a solution θ^* to the equation:

$$\mathbb{P}(X \leqslant \theta) = \alpha \quad \Leftrightarrow \quad h(\theta) := \mathbb{E}[H(\theta, X)] = 0, \quad \alpha \in (0,1)$$

with $H(\theta, x) := \mathbf{I}_{\{x \leqslant \theta\}} - \alpha$ and $h(\theta) = \mathbb{P}(X \leqslant \theta) - \alpha$. Since X is finite, a solution to the above equation always exists. If the probability distribution function of X is increasing, this solution is unique. For sake of simplicity, we will assume that this is case. Let $(\gamma_n)_{n \geqslant 1}$ be a deterministic positive step sequence satisfying (1.9). The dynamics of the Robbins-Monro algorithm is given by

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, X^{(n+1)}), \quad n \geqslant 0$$

where $\theta_0 \in L^2(\mathbb{P})$ and is independent of the i.i.d. innovation sequence $(X^{(n)})_{n \geqslant 1}$ (with common law μ). We introduce the Lyapunov function $L(\theta) := \frac{1}{2}|\theta - \theta^*|^2$ and remark that

$$\langle \nabla L, h \rangle(\theta) = (\theta - \theta^*)(\mathbb{P}(X \leqslant \theta) - \mathbb{P}(X \leqslant \theta^*)) > 0, \quad \forall \theta \neq \theta^*$$

since $\theta \mapsto \mathbb{P}(X \leqslant \theta)$ is increasing. Moreover, the function H being bounded assumption (1.8) is clearly satisfied. Hence, by Corollary 1.1.2 the sequence $(\theta_n)_{n \geqslant 1}$ converges *a.s.* toward its target θ^*.

In order to investigate the weak rate of convergence, we assume that X admits a continuously differentiable density p_X satisfying $p_X(\theta^*) > 0$. Our aim is to apply Theorem 1.1.5. Assumptions (1) and (2) are clearly satisfied. Moreover, by the very definition of θ^*, one has $\Sigma^* = \mathbb{E}[H(\theta^*, X)^2] = \alpha(1-\alpha) > 0$. In order to achieve the optimal asymptotic rate we set $\gamma_n = \gamma/(a+n)$, with $a > 0$ and $\gamma > 1/(2p_X(\theta^*))$. From Theorem 1.1.5, one gets

$$\sqrt{n}(\theta_n - \theta^*) \Rightarrow \mathcal{N}(0, \Gamma), \quad n \to \infty$$

with

$$\Gamma := \alpha(1-\alpha)\frac{\gamma^2}{2\gamma p_X(\theta^*) - 1}.$$

This asymptotic variance is minimal for $\gamma = \frac{1}{p_X(\theta^*)}$ with an optimal value $\Gamma = \alpha(1-\alpha)/p_X^2(\theta^*)$. However, in practical implementations, one does not have access to the value of θ^* or the density of X. As a consequence, one is naturally led to consider the averaged form of the algorithm although it is completely satisfactory for a practical point of view. When one tries to implement the stochastic approximation scheme $(\theta_n)_{n \geqslant 1}$, one observes a chaotic behaviour with a slow convergence, specially when $\alpha \approx 1$, say $\alpha > 95\%$. In **[1]** and **[2]**, we proposed a R.M. algorithm to compute the quantile and the expected shortfall and improved its performance by combining it with a stochastic approximation based importance sampling technique.

In Chapter 2, we propose an application to the recursive computation of the shortfall L^p-risk of a financial portfolio and the recursive computation of risk minimising strategies.

We refer the interested reader to the Phd dissertation of Laruelle [86] for some applications of the stochastic approximation theory to the optimal splitting of orders across liquidity pools such that dark pools, to multi-arm clinical trials by urn models. We also mention the monograph of Pagès [110] for stochastic approximation based procedures for parameter implicitation and calibration of financial models, or the computation of optimal quantizers by e.g. the Competitive Learning Vector Quantization (CLVQ) method.

1.2 Discretisation and simulation of stochastic processes

1.2.1 Motivation

In this section, we consider an $\mathbb{R}^d$-valued stochastic differential equation (SDE in short) with dynamics

$$X_t = x + \int_0^t b(s, X_{s-})ds + \int_0^t \sigma(s, X_{s-})dZ_s, \quad t \geqslant 0, \tag{1.17}$$

where the coefficients $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^q$ are bounded measurable in time, Lipschitz-continuous in space (uniformly in time) and where $(Z_t)_{t \geqslant 0}$ is a q-dimensional Lévy process with Lévy-Khintchine decomposition given by

$$\mathbb{E}[\exp(i\langle u, Z_t\rangle)] = \exp(t\psi(u)), \quad \psi(u) = \exp(-\frac{|u|^2}{2} + \int_{\mathbb{R}^q} (\exp(i\langle u, z\rangle) - 1 - i\langle u, z\rangle \mathbf{I}_{\{|z| \leqslant 1\}})\nu(dz)). \tag{1.18}$$

In other words, the SDE (1.17) is driven by a process Z that is given by the sum of a q-dimensional Brownian motion and pure jump process N of Lévy measure ν.

Such dynamics appears in various fields. This is so because of the strong connection between SDEs and PDEs (see e.g. Bass [14], Friedman [45]) via Feynman-Kac representation formulas. In various situations, the unique classical solution of a linear parabolic or elliptic PDE can be expressed as the expectation of a (path) functional of the SDE (1.17).

When Z is a q-Brownian motion, the diffusion process (1.17) is commonly used to provide probabilistic representation of the heat equation in non-homogeneous media. It is also commonly used to model the dynamics of assets in financial and energy markets. In **[4]**, we proposed a joint model of Gas and electricity spot prices based on Ornstein processes with parameterized diffusion coefficients that captures the most salient statistical properties observed on data. When Z is a pure jump process, the dynamics (1.17) allows to reproduce some stylised features observed in energy markets such that spikes in the spot prices, see e.g. Benth and al. [18]. In mathematical finance, the price of an option is usually seen as the solution of a parabolic PDE in which the terminal condition is given by the payoff.

In order to compute a quantity related to the law of the process with dynamics (1.17), which is not known in most applications, one usually considers a discrete approximation scheme. A natural scheme that can be implemented very easily is the Euler-Maruyama scheme that we now describe. On the interval $[0,T]$, $T>0$, for a given time step $h=T/N>0$, $N\in\mathbb{N}^*$, we introduce the time grid $t_i=ih$, $i=0,\cdots,N$ and the continuous Euler scheme $(X_t^h)_{t\in[0,T]}$ defined for $t\in[0,T]$ by

$$X_t^h = x+\int_0^t b(\phi(s),X_{\phi(s)}^h)ds+\int_0^t \sigma(\phi(s),X_{\phi(s)}^h)dZ_s,\quad t\geqslant 0, \tag{1.19}$$

where $\phi(s)=\sup\{t_i : t_i\leqslant s\}$. In other words, the Euler scheme is obtained by freezing the coefficients to the current discretisation point in each small interval $[t_i,t_{i+1}]$. Let us mention that in the case where Z has Brownian and jump parts, it is not always possible to simulate the increments $(Z_t-Z_s)_{0\leqslant s\leqslant t\leqslant T}$ of the innovation process so that it is not always possible to implement the scheme (1.19). However, it is still possible to provide an analysis of the error, see Section 1.2.3. In many important cases, one can implement this scheme for instance when Z is a stable process or Gamma process, etc.

We will present two different results related to the approximation of $(X_t)_{t\in[0,T]}$ solution to the SDE (1.17) by the scheme $(X_t^h)_{t\in[0,T]}$ given by (1.19). Approximation of stochastic processes goes back to the works of Stroock and Varadhan. They investigated approximation of the distribution of solutions to martingale problems by Markov chains. We refer the interested reader to Chapter 11 in [120] and to [119] for Dirichlet boundary problems.

The pionnering works of Milstein [103], [104] investigated the strong approximation error, that is the L^p-error between the path of $(X_t)_{t\in[0,T]}$ and the path of its approximation $(X_t^h)_{t\geqslant 0}$. Then, Misltein [105] and Talay [121] proved that the weak error, that is the error between the law of the random variable X_T and the law of X_T^h, is of order h under some regularity assumptions.

1.2.2 Strong error

The strong error associated to the process (1.17) and its approximation scheme (1.19) is the quantity $\mathbb{E}[\sup_{0\leqslant t\leqslant T}|X_t-X_t^h|^p]^{\frac{1}{p}}$ for $p>0$. When the underlying innovation process Z is an $\mathbb{R}^q$-valued Brownian motion, one has the following result:

Proposition 1.2.1. *Suppose that the coefficients b and σ of the SDE (1.17) satisfy the following regularity condition:*

$$\forall (s,t)\in[0,T]^2,\quad \forall (x,y)\in\mathbb{R}^d,\quad |b(t,x)-b(s,y)|+\|\sigma(s,x)-\sigma(t,y)\|\leqslant C(|t-s|^\alpha+|x-y|) \tag{1.20}$$

where $\alpha\in[0,1]$ and for some positive constant C. Then, for all $p>0$, there exists a positive constant C_p such that

$$\mathbb{E}[\sup_{0\leqslant t\leqslant T}|X_t-\bar{X}_t^h|^p]^{\frac{1}{p}}\leqslant C_p(1+|x|)\,h^{\alpha\wedge\frac{1}{2}}. \tag{1.21}$$

As a consequence, we remark that the Hölder regularity time exponent α rules the strong convergence rate of the Euler approximation scheme as soon as $\alpha<1/2$. The key ingredient for the proof of Proposition 1.2.1 is to use a Gronwall Lemma to upper-bound $\mathbb{E}[\sup_{0\leqslant t\leqslant T}|X_t-\bar{X}_t^h|^p]^{1/p}$ by the L^p-norm of the increments $X_s-X_{\phi(s)}$ of the diffusion. As a straighforward corollary, we deduce that the Euler approximation scheme (1.19) converges $a.s.$ toward the diffusion process at a rate $\beta\in[0,\alpha\wedge\frac{1}{2})$, namely

$$\forall\beta\in[0,\alpha\wedge\frac{1}{2}),\quad h^{-\beta}\sup_{0\leqslant t\leqslant T}|X_t-\bar{X}_t^h|\to 0,\quad a.s.$$

as $h\downarrow 0$. The proof follows from the L^p-convergence Proposition 1.2.1 and an argument using Borel-Cantelli Lemma.

In order to compensate this slow convergence rate Milstein introduced in [104] a new scheme in order to achieve the usual strong error of order h that appears in the deterministic framework. In dimension $d = q = 1$ its expression is simple and it can be easily implemented provided that b and σ have enough regularity. In higher dimension, some theoretical and simulation problems make its use more questionable. The key idea is to use an Itô-Taylor formula in the stochastic integral between to discrete times of the grid (which behaves like $\sqrt{h}$ and is the leading term in the strong approximation error) in order to introduce higher order terms in the dynamics of the scheme (1.19). We refer to the book of Kloeden and Platten [71] for an overview about higher order schemes.

Similar arguments can be used in order to quantify the strong error with the same rate of convergence when the innovation process Z is a square integrable Lévy process, see e.g. Kohatsu-Higa and Protter [72]. We also mention the work of Bruti Liberati and Platen [27] where the authors investigate the strong approximation error for pure jump SDEs, propose jump-adapted time discretizations and higher order scheme (in the strong sense) when the intensity of the jump process is finite $\nu(\mathbb{R}^q) < \infty$.

1.2.3 Weak error

In many situations, like the pricing of European options in mathematical finance, an Euler discretisation scheme $(X_t^h)_{t\in[0,T]}$ such that (1.19) is introduced to evaluate quantities writing as $\mathbb{E}[F((X_t)_{t\in[0,T]})]$ where $F : \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \to \mathbb{R}$ is a path functional. In order to evaluate such quantities, one is led to use the Monte Carlo estimator

$$E_M^h := \frac{1}{M}\sum_{i=1}^{M} F((X_t^{h,i})_{t\in[0,T]}), \quad M \geqslant 1$$

where $((X_t^{h,i})_{t\in[0,T]})_{1\leqslant i\leqslant M}$ are i.i.d. samples of the path of the Euler scheme $(X_t^h)_{t\in[0,T]}$ defined by (1.19).

The global error between $\mathbb{E}[F((X_t)_{t\in[0,T]})]$, the quantity to estimate, and its implementable approximation E_M^h can be decomposed as follows:

$$\begin{aligned}\mathcal{E}_{glob} &:= (\mathbb{E}[F((X_t)_{t\in[0,T]})] - \mathbb{E}[F((X_t^h)_{t\in[0,T]}]) + (\mathbb{E}[f((X_t^h)_{t\in[0,T]})] - E_M^h) \\ &:= \mathcal{E}_W^h + \mathcal{E}_S^M. \end{aligned} \tag{1.22}$$

The term $\mathcal{E}_S^M$ corresponds to the statistical error. If $F((X_t^h)_{t\in[0,T]}) \in L^2(\mathbb{P})$, the statistical error is of order $\sigma^h/\sqrt{M}$ in $L^2(\mathbb{P})$ and the renormalised statistical error satisfies a CLT. One may also be interested in non-asymptotic deviation estimates of this quantity in order to device non-asymptotic confidence interval. We will come back to this particular topic in Chapter 3.

The term $\mathcal{E}_W^h$ corresponds to the *weak error* and has been widely investigated in the literature since the seminal work of Talay and Tubaro [122] for the case $F((X_t)_{t\in[0,T]}) = F(X_T)$. One may be interested in quantifying this error for other type of functionals. We will briefly present some results related to the two functionals:

1. $F((X_t)_{t\in[0,T]}) = F(X_T)$
2. $F((X_t)_{t\in[0,T]}) = F(\tau \wedge T, X_{\tau\wedge T})$, where $\tau := \inf\{t \geqslant 0 : X_t \notin D\}$ for a domain $D \subset \mathbb{R}^d$.

The first case corresponds to the standard Cauchy problem since the quantity to estimate is related to the unique classical solution of linear parabolic PDE with terminal fonction F.

The second case corresponds to the Cauchy-Dirichlet problem since the quantity to estimate is related to the solution of linear parabolic PDE with Dirichlet conditions.

Other type of functional can be discussed such that $F(\max_{0\leqslant t\leqslant T} X_t, X_T)$ or $F(\int_0^T \mathbf{I}_{\{X_t\geqslant 0\}} dt, X_T)$ both in dimension $d = q = 1$.

Still in the one-dimensional diffusion case $d = q = 1$, we mention the works of Alfonsi, Jourdain and Kohatsu-Higa [2], [3] where the Wasserstein distance between the laws of $(X_t)_{t\in[0,T]}$ and $(X_t^h)_{t\in[0,T]}$ defined by

$$\mathcal{W}_1(\mathcal{L}(X), \mathcal{L}(X^h)) := \sup_{F:\mathcal{C}([0,T],\mathbb{R})\to\mathbb{R},\, Lip(F)\leqslant 1} |\mathbb{E}[F(X^h)] - \mathbb{E}[F(X)]|$$

is investigated by means of a pathwise optimal transport technique. Here $Lip(F)$ denotes the Lipschitz constant of F (with respect to the sup norm on $\mathcal{C}([0,T],\mathbb{R})$). It is proved that this quantity is upper-bounded by $h^{\frac{2}{3}-\varepsilon}$, $\forall \varepsilon > 0$, which is an intermediate rate between the strong error and the usual weak error rates. In the multi-dimensional diffusion case, when the coefficients b and σ are time-homogeneous $\mathcal{C}^2$, bounded together with their derivatives up to the order 2 and uniform ellipticity holds, the ρ-Wasserstein distance is shown to be of order $\sqrt{\log(N)/N}$.

The Cauchy problem for diffusion processes

For the *standard* continuous Euler scheme (1.19), the weak discretisation error is of order h, provided the coefficients b, σ and the test function F are $\mathcal{C}^4$ with the same polynomial growth assumption. The key idea [121], [122] consists in observing that the function $u(t,x) = \mathbb{E}[F(X_T)|X_t = x] = \mathbb{E}[F(X_T^{t,x})]$ (note that $(X_t)_{t\in[0,T]}$ is a Strong Markov process under this assumption) is a classical solution to the parabolic linear PDE

$$\begin{cases} (\partial_t + \mathcal{L}_t)u = 0, \quad (t,x) \in [0,T) \times \mathbb{R}^d \\ u(T,x) = F(x), \end{cases} \tag{1.23}$$

where $\mathcal{L}_t$ is the infinitesimal generator associated to (1.17), namely $\mathcal{L}_t F(x) = \langle b(t,x), \nabla F(x)\rangle + \frac{1}{2} Tr(a(t,x) D_x^2 F(x))$, for $F \in \mathcal{C}_0^\infty(\mathbb{R}^d)$, where $a(t,x) = (\sigma\sigma^*)(t,x)$. Under the mentionned assumption on the coefficients and test function F, one gets that $u \in \mathcal{C}^{2,4}([0,T]\times\mathbb{R}^d)$ with the following controls on the derivatives: $\forall (t,x) \in [0,T]\times\mathbb{R}^d$, $|\partial_t^i \partial_x^j u(t,x)| \leqslant C_T(1+|x|^{p_{i,j}})$. We remark that the last controls allow to deduce that $|\partial_t^i \partial_x^j u(t_i, X_{t_i}^h)| \in L^p(\mathbb{P})$ for any $p \geqslant 1$. The next step is to introduce the solution of the PDE (1.23) as follows

$$\mathcal{E}_W^h := \mathbb{E}[F(X_T)] - \mathbb{E}[F(X_T^h)] = -(\mathbb{E}[u(T, X_T^h)] - u(0,x)) = -\sum_{i=0}^{N-1} \mathcal{E}_W^{h,i}$$

with $\mathcal{E}_W^{h,i} := \mathbb{E}[u(t_{i+1}, X_{t_{i+1}}^h) - u(t_i, X_{t_i}^h)]$. From these representation, one applies Itô-Taylor formula around the point $(t_i, X_{t_i}^h)$ for $i = 0, \cdots, N-1$ using the regularity of u. In order to prove that $\mathcal{E}_W^{h,i} = \mathcal{O}(h^2)$, it is necessary to obtain an expansion of order two in time and 4 in space, the term of order 3 in space vanishes because of the third moment of the underlying noise is zero.

Let us also mention here that when the Gaussian increments of the scheme are replaced by more general (possibly discrete) random variables having the same covariance matrix and odd moments up to order 5 as the standard Gaussian vector of $\mathbb{R}^q$, it can be checked that the error expansion at order h still holds for b, σ, F smooth enough.

If we keep on that direction by doing an expansion at order 3 in time and 6 in space, observing that the odd moments of the underlying Gaussian vector vanishes, one gets

$$\mathcal{E}_W^h = C_T^1 h + \mathcal{O}(h^2),$$

where C_T^1 is an explicit constant independent of h. According to Talay and Tubaro [122], when b, σ are $\mathcal{C}^\infty$ with bounded derivatives of all orders and F is $\mathcal{C}^\infty$ with polynomial growth together with its derivatives, then for each integer $L \geqslant 1$, the expansion

$$\mathcal{E}_W^h = \sum_{\ell=1}^{L} C_T^\ell h^\ell + \mathcal{O}(h^{L+1}) \tag{1.24}$$

holds.

On can take advantage of the previous expansion in order to perform a *Richardson-Romberg extrapolation*. This idea was originally introduced in [122] and then deeply investigated by Pagès in [109]. One considers the strong solution X^2 of a "copy" of equation (1.17), driven by a second Brownian motion W^2 defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. One may always choose such a Brownian motion by enlarging the space Ω if necessary.

Then one considers the continuous Euler scheme with a twice smaller step $h/2$, denoted by $(X_t^{2,h/2})_{t\in[0,T]}$ of the diffusion X^2. Then combining the two time discretisation error expansions associated to $X^{1,h}$ (related to the first Brownian motion W^1) and $X^{2,h/2}$, we get

$$\mathbb{E}[F(X_T)] = \mathbb{E}[2F(X_T^{2,h/2}) - F(X_T^{1,h})] - \frac{C_T^2}{2} h^2 + \mathcal{O}(h^3)$$

so that the global quadratic error becomes

$$\mathbb{E}\left[\left(\mathbb{E}[F(X_T)] - \frac{1}{M}\sum_{i=1}^{M} 2F(X_T^{2,h/2,(i)}) - F(X_T^{1,h,(i)})\right)^2\right] = (\frac{C_T^2}{2}h^2)^2 + \frac{Var(2F(X_T^{2,h/2}) - F(X_T^{1,h}))}{M} + \mathcal{O}(h^5)$$

If one follows a naive approach by simulating $X^{2,h/2}$ and $X^{1,h/2}$ independently, this corresponds to two independent Brownian motion W^1, W^2, then one increases the variance by a factor 5 (asymptotically) since $Var(2F(X_T^{2,h/2}) - F(X_T^{1,h})) = 4Var(2F(X_T^{2,h/2}) + Var(F(X_T^{1,h})) \to 5Var(F(X_T))$ as $h \downarrow 0$. It is shown in [109] that the consistent choice $W^1 \equiv W^2$ leading to consistent Brownian increments for the two schemes is asymptotically optimal among all possible choice of Brownian motions W^1 and W^2. We will see in Chapter 4 an extension of the Richardson-Romberg extrapolation method to stochastic optimisation by means of stochastic approximation algorithm.

Anyway in order to analyse the weak discretisation error we will always proceed this way. As mentioned above the key ingredient is to have controls on u and its derivatives and to proceed using Itô-Taylor's formula. This is for that reason that we require regularity on b, σ and F. However, in practical situations the test function F is not regular. For instance, in mathematical finance, one often deals with payoff F that are not regular, e.g. Call option $F(x) = (x-K)_+$ or digital option $F(x) = \mathbf{I}_{\{x \geqslant K\}}$.

One can weaken the regularity assumption on F if one can benefit from the smoothness of the underlying transition density. More precisely, if the process $(X_t)_{t\in[0,T]}$ solution to (1.17) admits a smooth transition density, one can relax the smoothness assumption on F. Indeed, in this case one has $u(t,x) = \int_{\mathbb{R}^q} F(y)p(t,T,x,y)dy$ where $y \mapsto p(t,T,x,y)$ is the density function of the random vector $X_T^{t,x}$. When the coefficients are time homogeneous, the density is given by $p(T-t,x,y)$.

However, in order to benefit from this regularising property one has to assume that the operator $\mathcal{L}$ satisfies some non-degeneracy assumption. One usually considers two type of assumptions: the Hörmander assumption **(H)** and uniform ellipticity assumption **(UE)** of $a = \sigma\sigma^t$.

The Hörmander condition is a sufficient (and is known to be nearly necessary) condition for a second-order linear Kolmogorov PDE with smooth coefficients to be hypoelliptic, that is, $P_t\mathcal{C}_b(\mathbb{R}^d,\mathbb{R}) \subseteq \mathcal{C}_b^\infty(\mathbb{R}^d,\mathbb{R})$, for all $t \in (0,\infty)$, where $(P_t)_{t\geqslant 0}$ denotes the associated Markov semigroup. Note that in the non-hypoelliptic regime, even in the case of smooth coefficients, there exist counterexamples to regularity preservation. In Hairer, Hutzenthaler, Jentzen [59], the *roughening effect* of the noise for non-hypoelliptic linear PDE is exhibited. From the perspective of numerical probability, this has the consequence that numerical approximations may converge (in the strong and weak sense) without any arbitrarily small polynomial rate.

Assumption (H): The coefficients b and σ are $\mathcal{C}^\infty$ and time homogeneous, with bounded derivatives of order larger than 1 (importantly the functions b and σ are not supposed to be bounded). Let $A_0 = b - \frac{1}{2}\sigma.\nabla\sigma$ and $A_i = \sigma_i$, $i = 1,\cdots,q$ where $\sigma = (\sigma_1,\cdots,\sigma_q)$. For a multi-index $\alpha = (\alpha_1,\cdots,\alpha_r) \in \{0,\cdots,q\}^r$ we define the vector fields A_i^α by $A_i^\varnothing = A_i$ for $1 \leqslant i \leqslant q$ and $A_i^{(\alpha,j)} = [A_j, A_i^\alpha]$ for $0 \leqslant j \leqslant q$ where for two smooth $\mathbb{R}^d$-valued vector fields V, W, $[V,W] = \nabla WV - \nabla VW$, $\nabla V = (\partial_j V^i)_{1\leqslant i,j\leqslant d}$, stands for their Lie bracket. The Hörmander condition holds if the vector space spanned by A_i^α, $1 \leqslant i \leqslant q$, α being a multi-index, at the point x is $\mathbb{R}^d$.

Under **(H)**, the density function $y \mapsto p(T-t,x,y)$ of $X_T^{t,x}$ exists and is smooth. Moreover, if the vector fields $A_0,\cdots,A_q$ satisfies the *uniform Hörmander* assumption **(UH)**, that is if there exists an integer L and $\lambda > 0$ such that

$$\forall x \in \mathbb{R}^d, \quad \inf_{u\in\mathcal{S}^{d-1}} \sum_{|\alpha|\leqslant L} \langle A^\alpha(x), u\rangle^2 \geqslant \lambda \tag{1.25}$$

Kusuoka and Stroock [83] provided the following controls on the density. For any integers m, k and any multi-index α and β such that $2m + |\alpha| + |\beta| \leqslant k$, there exist an integer M and a non-decreasing function $T \mapsto C(T)$ such that the following inequality holds:

$$|\partial_t^m \partial_x^\alpha \partial_y^\beta p(t,x,y)| \leqslant \frac{C(T)(1+|x|^M)}{t^q(1+|y-x|^2)^k} \exp(-c\frac{|y-x|^2}{t(1+|x|^2}). \tag{1.26}$$

Moreover, the inverse of the Malliavin covariance matrix denoted by $\Gamma_t(x)$ is bounded in any $L^p(\mathbb{P})$. More precisely, for any $p \geqslant 1$, there exists a non-decreasing function $C(T)$ and an integer k such that

$$\mathbb{E}[\|\Gamma_t\|^p]^{\frac{1}{p}} \leqslant C(T)\frac{1+|x|^k}{t^{dL}} \tag{1.27}$$

In [10], Bally and Talay take benefit from (1.26) and (1.27) in order to extend (1.24) for hypo-elliptic diffusions without any regularity on the test function F, which is supposed to be measurable with exponential growth at infinity. The main quantities that appears in the expansion of the weak discretisation error are $Q_i^h = \mathbb{E}[\partial_x^\alpha u(t_i, X_{t_i}^h)G(X_{t_i}^h)]$ for a smooth function G with polynomial growth. Roughly speaking, when t_i is close to 0, say $i \leqslant N/2$, then one uses (1.26) in order to control Q_i^h. When t_i is close to T, one has to use Malliavin's integration by parts formula with respect to the random variable $X_{t_i}^h$ which is expected to be non-degenerated with high-probability. We are led to compare Γ_t^h and Γ_t, where Γ_t^h is the inverse of the Malliavin covariance matrix of X^h at time t. This can be done using a localisation argument.

Let us also mention that in general the Euler scheme does not admit a density under assumption **(UH)**. In [11], Bally and Talay introduced a perturbed Euler scheme which admits a transition density p^h and proved an expansion of the error $p - p^h$.

Assumption (UE): The diffusion matrix a is bounded and uniformly elliptic, that is there exists $\bar{a}$ such that $\forall(t,x)[0,T]\times\mathbb{R}^d, \forall u \in \mathbb{R}^d$, $\bar{a}^{-1}|u|^2 \leqslant \langle a(t,x)u,u\rangle \leqslant \bar{a}|u|^2$.

When b and σ are Hölder continuous in space (uniformly in time) and bounded, the unique (weak) solution to (1.17) (driven by a Brownian motion W) satisfies the following Gaussian bound $\forall T > 0$, $\exists C > 1$, $\forall(t,x,y) \in$

$[0,T)\times\mathbb{R}^d\times\mathbb{R}^d, |p(t,T,x,y)|\leqslant C\exp(-C^{-1}|y-x|^2/t)$. The latter bound follows from the *parametrix method*, see Friedman [44] or McKean and Singer [95]. The parametrix technique is a classical perturbation method used in PDE theory that was originally formulated in order to give an expansion in infinite series of iterated kernels of the fundamental solution of an elliptic or parabolic PDE of order $2p$, $p\in\mathbb{N}^*$, with time inhomogeneous coefficients. More precisely, one has

$$p(t,T,x,y)=\sum_{n\geqslant 0}(\tilde{p}\otimes H^n)(t,T,x,y) \tag{1.28}$$

where $\tilde{p}(t,T,x,y)$ is the Gaussian density of the system (1.17) where we removed the drift b and fixed the diffusion coefficient to the terminal point y. This system is often referred to the *frozen* or *parametrix* process. The space-time convolution operator $\otimes$ is defined by $(f\otimes g)(t,T,x,y):=\int_t^T ds\int_{\mathbb{R}^d}dzf(t,s,x,z)g(s,T,z,y)$ and H is a kernel. The keystone of this approach is that the underlying parametrix kernel H enjoys a regularising property in the sense that $(\tilde{p}\otimes H^n)(t,T,x,y)$ is explicitly controlled by the term of an absolutely (and uniformly) convergent series times a Gaussian density. We will come back to this point latter on in the second part of this manuscript in Chapter 5.

This method has been successfully applied to many equations and various situations. Its merit and success are due to its flexibility as it can be invoked for a wide variety of PDEs both for theoretical goals such as density estimates [34], martingale problems [97], strong existence and uniqueness for SDE [29] and for numerical approximations [75]. Let us be more precise on the latter topic since it will be used in Chapter 5. If we replace in the dynamics of the continuous Euler scheme (1.19) the Brownian increments by a sequence of *i.i.d.* random variables with common law μ which admits a density with respect to the Lebesgue measure (with the same moments as the Gaussian density up to a certain order), then the new scheme admits a transition density p^N which admits a representation in infinite series

$$p^N(t,T,x,y)=\sum_{n\geqslant 1}(\tilde{p}\otimes_N H^{N,n})(t,T,x,y) \tag{1.29}$$

where $y\mapsto\tilde{p}^N(t,T,x,y)$ is the density of the recursive scheme where we removed the drift b and fixed the diffusion coefficient to the terminal point y. The kernel H^N and the operator $\otimes_N$ are discrete version of the kernel H and the space time convolution operator $\otimes$. In [75], Konakov and Mammen followed these approach. By comparing the two series (1.28) and (1.29) they obtained local limit theorems for the difference $(p-p^N)(t,T,x,y)$ at a rate $h^{\frac{1}{2}}$.

When one considers the true continuous Euler scheme with Brownian increments, an expansion of the error for $p-p^N$ can be obtained. More precisely, Konakov and Mammen in [76] proved the following expansion

$$\forall L\in\mathbb{N}^*,\quad (p-p^N)(t,T,x,y)=\sum_{\ell=1}^{L}\frac{\pi_\ell^N(t,T,x,y)}{N^\ell}+\frac{1}{N^{L+1}}\mathscr{R}_N(t,T,x,y) \tag{1.30}$$

with the following Gaussian controls

$$\exists C_{T-t}>1,\ \forall(x,y)\in(\mathbb{R}^d)^2, |\pi_\ell^N(t,T,x,y)|+|\mathcal{R}_N(t,T,x,y)|\leqslant C_{T-t}\exp(-c\frac{|y-x|^2}{T-t})$$

under the assumptions **(UE)** and assuming that the coefficients b, σ are $\mathcal{C}^\infty$ with bounded derivatives of all order. We will present in Chapter 5 some results related to the weak approximation error of a skew diffusion with bounded measurable drift and Hölder diffusion coefficient by an Euler-type scheme. We obtained a similar error bound for the difference between the densities of the skew diffusion and its Euler approximation under mild smoothness assumption on the coefficients.

Under stronger smoothness assumption on the coefficients, Gobet and Labart [53] quantified the small time asymptotics of the difference $p-p^N$. More precisely, using Malliavin Calculus techniques, they established the bound $|(p-p^N)(t,T,x,y)|\leqslant Ch(T-t)^{-\frac{d+1}{2}}\exp(-c|y-x|^2/(T-t))$, where $C,\ c>1$ are constants independent of t,T.

The main advantage of (1.30) compared to the previous approach is that one can considerably weaken the assumption on the test function F. Guyon [58] proved that (1.24) holds if F is a tempered distributions, e.g. a Dirac mass or derivatives of a Dirac mass. We will see in Chapter 4 that this point is very important in order to obtain an expansion of the implicit discretisation error for solutions to inverse problems. It is also important to quantify this kind of error when one considers the computation of greeks in option hedging.

Recent research works have shown a rising interest in weak higher order approximation schemes starting with the pioneering work of Kusuoka [82]. The method relies on the splitting of the infinitesimal generator in order to use composition techniques of ODEs flow see Ninomiya and Victoir [107] and very recently Bally and Rey [9]. In the latter work, the authors obtained estimates of the error in total variation distance for the Ninomiya Victoir scheme using Malliavin calculus techniques under a Doeblin's condition on the underlying noise.

An interesting problem, which has not been investigated (to the best of our knowledge), will be to study the existence of a transition density for higher order scheme like the Ninomiya scheme (as well as Gaussian lower and upper-bounds) and to see if expansions similar to (1.30) holds. For instance, one may try to follow the approach initiated by Konakov and Mammen [76].

The Cauchy problem in the case of SDE driven by jump process

The approach of Talay and Tubaro [122] still works when one adds jumps in the innovation process Z appearing in the dynamics (1.17). The key idea still consists in controlling the derivatives of the function u, the unique classical solution to linear integro-differential PDE

$$\begin{cases} (\partial_t + \mathcal{L}_t)u(t,x) + \\ \int_{\mathbb{R}^q}(u(t,x+\sigma(t,x)z - u(t,x) - \langle \nabla u(t,x), \sigma(t,x)z\rangle \mathbf{I}_{\{|z|\leqslant 1\}}) = 0, \quad (t,x) \in [0,T)\times\mathbb{R}^d \\ u(T,x) = F(x), \end{cases} \tag{1.31}$$

Like in the Brownian case, under regularity assumptions on the coefficients and terminal condition F and under integrability condition of the Lévy measure, Protter and Talay [115] derived an expansion of the weak discretisation error similar to (1.24).

One important issue that does not appear in the Brownian framework is the simulation of the increments of the Lévy process Z. In many cases such that stable, Gamma, compound poisson process, one knows how simulate the increments but in general one does not know how to do. Jacod and al. [66] proposed an implementable Euler scheme where the increments $(Z_{t_{i+1}} - Z_{t_i})_{0\leqslant i\leqslant N-1}$ are approximated by a sequence $(U_i^h)_{0\leqslant i\leqslant N-1}$ of i.i.d random variables satisfying $|\mathbb{E}[g(Z_h) - g(U_1^h)]| \leqslant Ch\epsilon_h$ for $g \in \mathcal{C}^4(\mathbb{R}^q)$. Under some integrability conditions on ν, they obtained the bound $|\mathbb{E}[F(X_T) - F(X_T^h)]| \leqslant C(\epsilon_h \vee h)$ and derived an expansion of the error similar to (1.24) under stronger smoothness assumption on the coefficients b and σ.

When the innovation is a non-degenerated stable process, Konakov and Menozzi in [77] extended the parametrix machinery developed in [76] and successfully derived an error expansion similar to (1.30).

Let us also mention the work of Kohatsu-Higa and Tankov [74] where the authors study the weak approximation for a jump-adapted discretisation scheme of Lévy driven SDEs. The approach consists in building upon adaptive non-uniform discretisation based on the jump times of the driving process coupled with suitable approximations of the solutions between these jump times.

The Cauchy-Dirichlet problem for diffusion processes

For a Brownian driven SDE with dynamics (1.17), let us consider the first exit time τ of the domain $D \subset \mathbb{R}^d$ defined by $\tau = \inf\{t \geqslant 0 : X_t \notin D\}$. The computation of $\mathbb{E}[F(\tau\wedge T, X_{\tau\wedge T})]$ is an important issue that appears in various fields such as mathematical finance for the pricing and hedging of look-back options. More generally, it is the probabilistic representation of the classical solution of the following Cauchy-Dirichlet problem

$$\begin{cases} (\partial_t + \mathcal{L}_t)u(t,x) = 0, & (t,x) \in [0,T)\times D, \\ u(t,x) = F(t,x), & (t,x) \in [0,T)\times\partial D \cup \{T\}\times\bar{D} \end{cases} \tag{1.32}$$

One is naturally led to consider the approximation $\mathbb{E}[F(\tau^h \wedge T, \Pi_{\bar{D}}(X^h_{\tau^h\wedge T}))]$ where $\tau^h := \inf\left\{t_i : X^h_{t_i} \notin D\right\}$and $\Pi_{\bar{D}}$ is the projection on $\bar{D}$. Then, using the same decomposition as in the standard Cauchy problem, one has

$$\mathcal{E}_W^h = \mathbb{E}[F(\tau^h \wedge T, \Pi_{\bar{D}}(X^h_{\tau^h\wedge T})] - \mathbb{E}[F(\tau\wedge T, X_{\tau\wedge T})] = \mathbb{E}[u(\tau^h \wedge T, \Pi_{\bar{D}}(X_{\tau^h\wedge T}))] - u(0,x).$$

Then, following a similar strategy as in the standard Cauchy problem, one has to apply Itô-Taylor formula and use the regularity of the solution u to (1.32).

If the domain D is sufficiently smooth and under non-degeneracy assumption (namely **(UE)** or **(H)**), one is able to prove that $\mathcal{E}_W^h$ is of order $h^{\frac{1}{2}}$. More precisely, Gobet [50], [51] proved that $\mathcal{E}_W^h = \mathcal{O}(h^{\frac{1}{2}})$ under some regularity assumption on F. Then, Gobet and Menozzi [54] sharpened this result by proving a lower bound of order $h^{1/2}$ and a first order expansion that writes $\mathcal{E}_W^h = C_1 h^{1/2} + o(h^{1/2})$.

1.2.4 Multi-level Monte Carlo simulation

The multi-level Monte Carlo method introduced by Heinrich [61] for parametric integration and recently developed by Giles [48] for Monte Carlo simulation as an extension of the two level method of Kebaier [70] allows to minimise the simulation cost of the quantity $\mathbb{E}[Y]$ where Y is an $\mathbb{R}$-valued random variable that can only be approximated by a sequence $(Y^n)_{n\geqslant 1}$ of easily simulatable random variables as the *bias parameter* n goes to infinity with a *weak discretisation error* or *bias* $\mathbb{E}[Y] - \mathbb{E}[Y^n]$ of order $n^{-\alpha}$, $\alpha > 0$.

This method is now very popular, notably in mathematical finance where as mentioned in the previous sections the random variable of interest is given by a functional of an SDE with dynamics (1.17). However, it has been widely applied to various problems of numerical probability, see Giles [47], Dereich [36], Giles, Higham and Mao [49] among others. We refer the interested reader to the webpage: `http://people.maths.ox.ac.uk/gilesm/mlmc_community.html` for further developments.

Let us be more specific. In this context, the standard Monte Carlo method uses the statistical estimator $M^{-1} \times \sum_{j=1}^{M} Y^{n,(j)}$ where the $(Y^{n,(j)})_{1 \leqslant j \leqslant M}$ are M independent copies of Y^n. Given the order of the weak error, a natural question is to find the optimal choice of the sample size M to achieve a global error. If the weak discretisation error is of order $n^{-\alpha}$ (for instance $\alpha = 1$ for the standard Cauchy problem and $\alpha = \frac{1}{2}$ for the Cauchy Dirichlet problem) then for a total error of order $n^{-\alpha}$ ($\alpha \in [1/2, 1]$), the minimal computation necessary for the standard Monte Carlo algorithm is obtained for $M = n^{2\alpha}$, see Duffie and Glynn [38]. So, if the computational cost required to simulate one sample of Y^n is of order n (as it is the case for the Euler scheme (1.19)) then the optimal computational cost of the Monte Carlo method is $C_{MC} = C \times n^{2\alpha+1}$, for a positive constant $C > 0$.

In order to reduce the complexity of the computation, the principle of the multi-level Monte Carlo method introduced by Giles [48] consists in using the telescopic sum

$$\mathbb{E}[Y^{m^L}] = \mathbb{E}[Y^1] + \sum_{\ell=1}^{L} \mathbb{E}[Y^{m^\ell} - Y^{m^{\ell-1}}],$$

for an integer $m > 1$ satisfying $m^L = n^\alpha$. For each level $\ell \in \{1, \cdots, L\}$ the numerical computation of $\mathbb{E}[Y^{m^\ell} - Y^{m^{\ell-1}}]$ is achieved by the standard Monte Carlo method using N_ℓ independent samples of $(Y^{m^{\ell-1}}, Y^{m^\ell})$. An important point is that the random samples Y^{m^ℓ} and $Y^{m^{\ell-1}}$ are perfectly correlated. Then the expectation $\mathbb{E}[Y^n]$ is approximated by the following multi-level estimator

$$E_{MLMC} := \frac{1}{N_0} \sum_{j=1}^{N_0} Y^{1,(j)} + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} \left(Y^{m^\ell,(j)} - Y^{m^{\ell-1},(j)} \right),$$

where for each level ℓ, $(Y^{m^\ell,(j)})_{1 \leqslant j \leqslant N_\ell}$ is a sequence of i.i.d. random variables with the same law as Y^{m^ℓ}. The variance of the new estimator is the sum of the local variance:

$$Var(E_{MLMC}) = \sum_{\ell=0}^{L} N_\ell^{-1} V_\ell, \quad V_\ell := Var(Y^{m^\ell} - Y^{m^{\ell-1}}),\ \ell = 0, \cdots, L$$

with the convention $Y^{m^{-1}} \equiv 0$. The computational cost is proportional to $\sum_{\ell=0}^{L} N_\ell m^{-\ell}$.

A simple optimisation problem shows that the variance of the multi-level estimator is minimised for a fixed computational cost by choosing N_ℓ proportional to $\sqrt{V_\ell m^{-\ell}}$. In the case we have in mind corresponding to the Euler discretisation scheme with Lipschitz-continuous coefficients, one has $Y^{m^\ell} = F(X_T^{m^\ell})$ for a Lipschitz continuous function F so that $V_\ell = \mathcal{O}(m^{-\ell})$ and N_ℓ is proportional to the variance V_ℓ (at least asymptotically). This leads us to choose N_ℓ satisfying $N_\ell^{-1} V_\ell = \mathcal{O}(n^{-2\alpha}(L+1)^{-1})$. Then the variance of E_{MLMC} is $\mathcal{O}(n^{-2\alpha})$ and the biais is still $n^{-\alpha}$! Moreover, in order to achieve a global error of order $n^{-\alpha}$, the computational cost of the new estimator is $\mathcal{O}(n^{-2\alpha} L^2) = \mathcal{O}(n^{-2\alpha}(\log(n))^2)$. Based on this heuristic, we clearly see that the new multi-level Monte Carlo estimator outperforms the basic Monte Carlo estimator in terms of computational complexity.

More recently, Ben Alaya and Kebaier [1] proposed a different analysis to obtain the optimal choice of the parameters that relies on a Lindeberg-Feller CLT for the multi-level Monte Carlo algorithm.

More importantly, Lemaire and Pagès [93] proposed a multi-level Richardson-Romberg estimator which combines the higher order bias cancellation of the multi-step Richardson Romberg method investigated in [109] and the variance control of the multi-level Monte Carlo method briefly exposed above. Notably, in standard frameworks like the discretisation of diffusion processes, for a global error of order $n^{-\alpha}$, the computational cost is of order $\mathcal{O}(n^{-2\alpha} \log(n))$.

In Chapter 4, we extend the scope of the multi-level Monte Carlo method to the framework of stochastic optimisation by means of stochastic approximation algorithm. The optimal choice of the parameters is done by proving a CLT for the new estimators. We also present some numerical results that confirm the theoretical analysis and show a significant reduction in the initial computational cost.

1.2.5 Unbiased simulation of SDE

Unbiased simulation methods of SDEs offers an alternative to the Multi level Monte Carlo for the computation of $\mathbb{E}[F(X_T)]$ where $(X_t)_{t \in [0,T]}$ is solution to (1.17). We differentiate here between exact methods and unbiased methods. Exact methods allows to sample a path, at a finite set of points, with the exact distribution of the SDE, while unbiased means that we can estimate $\mathbb{E}[F(X_T)]$ without any bias, which is enough for many applications such that the pricing and hedging of financial claims.

An example of an exact simulation method is given in Beskos and al. [20] in dimension $d = q = 1$ based on the Lamperti transform of the original SDE. For some extensions to the SABR and the Heston models we refer the reader to Broadie and Kaya [26], Chen and al. [30]. However, all these methods cannot be easily extended to a general SDE.

The unbiased simulation technique for the multi-dimensional SDE has been introduced by Bally and Kohatsu-Higa [7] and developed by Anderson and Kohatsu-Higa [4] for multi-dimensional diffusions with hölder continuous coefficients. This method is based on a probabilistic interpretation of the parametrix method introduced by E. Lévi to construct the fundamental solution of elliptic and parabolic PDE. The proposed method is also based on the Euler scheme although in the proposed simulation scheme the time partition is random.

We now briefly expose the main guidelines for multi-dimensional diffusion processes with infinitesimal general $\mathcal{L} = \frac{1}{2}\sum_{i,j} a_{i,j}\partial_{i,j} + \sum_i b_i\partial_i$ where a is assumed to be uniformly elliptic. The coefficients a and b are smooth, say $a_{i,j} \in \mathcal{C}_b^2(\mathbb{R}^d)$ and $b_i \in \mathcal{C}_b^1(\mathbb{R}^d)$. We know that the unique solution to (1.17) defines a strong Markov semigroup $(P_t)_{t\geqslant 0}$ defined for all bounded measurable maps by $P_t f(x) = \mathbb{E}[f(X_t)]$. Under these assumptions it is also known that $P_t f \in \mathcal{C}_b^2(\mathbb{R}^d)$ for all $t > 0$, see e.g. Friedmann [44].

We introduce the approximation process $\bar{X}$ defined by $\bar{X}_t = x + b(x)t + \sigma(x)W_t$, that is the process obtained from (1.17) by freezing the drift and the diffusion coefficients at the initial point x. We introduce accordingly the collection of linear maps $(\bar{P}_t)_{t\geqslant 0}$ defined for all bounded measurable functions f by $\bar{P}_t f(x) = \mathbb{E}[f(\bar{X}_t)] = \int f(y)g(a(x)t, y-b(x)t-x)$, where, for a symmetric positive definite matrix a, $g(a,x)$ is the probability density function of $\mathcal{N}(0,a)$ taken at x. In particular, one has $\partial_t \bar{P}_t f(x) = \bar{P}_t \bar{\mathcal{L}} f$ for every $t > 0$. As a consequence, using an integration by parts formula the following one step formula holds

$$P_t f(x) = \bar{P}_t f(x) + \int_0^t \partial_s(\bar{P}_{t-s}P_s f(x))ds = \bar{P}_t f(x) + \int_0^t \bar{P}_{t-s}(\mathcal{L}-\bar{\mathcal{L}})P_s f(x)ds = \bar{P}_t f(x) + \int_0^t \mathcal{S}_{t-s}P_s f(x)ds \quad (1.33)$$

where we introduced the operator $S_t f(x) = \int_{\mathbb{R}^d} f(y)\theta_t(x,y)g(a(x)t, y-b(x)t-x)dy$ with

$$\begin{aligned}
\theta_t(x,y) &= \frac{1}{2}\sum_{i,j=1}^d \kappa_t^{i,j}(x,y) - \sum_{i=1}^d \rho_t^i(x,y),\\
\kappa_t^{i,j}(x,y) &= \partial_{i,j}^2 a_{i,j}(y) + \partial_j a_{i,j}(y)H^i(a(x)t, y-x-b(x)t) + \partial_i a^{i,j}(y)H^j(a(x)t, y-x-b(x)t)\\
&\quad + (a_{i,j}(y) - a_{i,j}(x))H^{i,j}(a(x)t, y-x-b(x)t),\\
\rho_t^i(x,y) &= \partial_i b^i(y) + (b^i(y) - b^i(x))H^i(a(x)t; y-x-b(x)t).
\end{aligned}$$

Here $H^i(a,x) = -(a^{-1}x)^i$ and $H^{i,j}(a,x) = (a^{-1}x)^i(a^{-1}x)^j - (a^{-1})^{i,j}$ are the Hermite polynomials of order 1 and 2 associated to the Gaussian density $g(a,x)$. From the *space-time inequality* $\forall x \in \mathbb{R}$, $|x|^p e^{-qx^2} \leqslant ((p/(2qe))^{p/2}$, valid for any $p, q > 0$, one easily gets $|\mathcal{S}_{t-s}f|_\infty \leqslant C(t-s)^{-\frac{1}{2}}|f|_\infty$, for a constant C depending only on the coefficients a, b and their derivatives.

Iterating the first step expansion (1.33), one proves that the following expansion holds

$$P_t f(x) = \sum_{n\geqslant 0}\int_{\Delta_n^*(t)} d\mathbf{s}_n \prod_{j=0}^{n-1} \mathcal{S}_{s_{j+1}-s_j}\bar{P}_{t-s_n}f(x), \quad (1.34)$$

where $\Delta_n^*(t) := \{\mathbf{s}_n = (s_1, \cdots, s_n) \in [0,t]^n : 0 < s_1 < \cdots < s_n < t\}$ and $s_0 = 0$.

In order to provide a probabilistic representation of the above series, for a time partition $\pi : 0 = s_0 < s_1 < \cdots < s_n < s_{n+1} = t$, we introduce the following discrete time process

$$X_{s_{i+1}}^\pi = X_{s_i}^\pi + b(X_{s_i}^\pi)(s_{i+1}-s_i) + \sigma(X_{s_i}^\pi)(W_{s_{i+1}} - W_{s_i}), \quad i = 0, \cdots, n.$$

Then, the following intermediary probabilistic representation holds

$$P_t f(x) = \sum_{n\geqslant 0}\int_{\Delta_n(t)} d\mathbf{s}_n \mathbb{E}[f(X_t^\pi)\prod_{i=0}^{n-1}\theta_{s_{j+1}-s_j}(X_{s_j}^\pi, X_{s_{j+1}}^\pi)], \quad (1.35)$$

and it remains to provide a probabilistic interpretation of the time integrals. To this end, we let $N(t)$ be a Poisson process with intensity parameter $\lambda > 0$ and define $N \equiv N(t)$. Let $\tau_1, \cdots, \tau_N$ be the event times of the Poisson process and set $\tau_0 = 0$ and $\tau_{N+1} = t$. Since, conditional on N, the event times are distributed as a uniform statistic, $\mathbb{P}(N = n, \tau_1 \in ds_1, \cdots, \tau_n \in ds_n) = \lambda^n e^{-\lambda t}$, for $\mathbf{s}_n \in \Delta_n^*(t)$ and the following probabilistic representation holds

$$P_t f(x) = \mathbb{E}[f(X_t)] = e^{\lambda t}\mathbb{E}[f(X_t^\pi)\prod_{i=0}^{N-1}\theta_{\tau_{j+1}-\tau_j}(X_{\tau_j}^\pi, X_{\tau_{j+1}}^\pi)]. \quad (1.36)$$

Let us note that the right-hand side of (1.36) can be sampled exactly, this gives us an unbiased simulation method for the computation of $\mathbb{E}[f(X_T)]$. The simulation algorithm consists in independently simulating the event times

of a Poisson process and an Euler scheme with a time partition given by the simulated random times. However, the method suffers from a poor convergence rate since the variance of the associated estimator may be infinite. An importance sampling on the time steps using a Beta or Gamma distribution is proposed in [4] which allows to achieve finite variance.

Another method is proposed in [7] and [4] which relies on a backward Euler scheme. The backward method can be applied for Hölder continuous coefficients. In this case, the Euler scheme is ran backward on the random time grid.

In Chapter 6, we will extend the above probabilistic representation to the one dimensional Cauchy-Dirichlet problem. More precisely, we obtain probabilistic representation of the process $(\tau \wedge t, X_{\tau \wedge t})_{t \geqslant 0}$ and obtain partial integration by parts formula that can be explicitly simulated without any bias.

We also mention the work of Henry-Labordère and al. [62] where the authors proposed a probabilistic representation for a class of semi-linear parabolic PDEs using a branching process. Usually this class of PDEs is solved using probabilistic numerical schemes of Backward SDEs. Here the algorithm is forward and backward regressions are no longer needed.

Part I : Stochastic approximation algorithm and applications

Chapter 2

Some contributions to risk management in financial and energy markets

In this chapter we present the two research works **[5]** and **[6]**.

In **[6]**, which can be seen as a continuation of **[3]**, we study the theoretical and computational aspects of risk minimisation in financial and energy market models operating in discrete time. The risk is defined by considering a class of convex risk measures defined on $L^p(\mathbb{P})$ in terms of shortfall risk **[6]**. Existence of an optimal strategy is proved under the absence of arbitrage and the non-degeneracy of the price process in a non-Markovian framework. We propose a Robbins-Monro algorithm to compute the shortfall risk of a given real-valued random variable. Then, we propose and analyse a Newton-Raphson optimisation algorithm to compute the optimal static risk minimising strategy. Finally, in a dynamic framework a quantisation based stochastic approximation algorithms to compute the corresponding risk measures as well as the optimal dynamic strategies is proposed.

In the research work **[5]**, in collaboration with A. Sagna of University Evry Val d' Essonne, we study a quantisation based Newton-Raphson algorithm to compute the optimal importance sampling (IS) parameter. The approach proposed here can be seen as a robust and automatic deterministic counterpart of recursive IS (by translation of the mean) by means of stochastic approximation algorithm, as proposed by Lemaire and Pagès [92] which may require tuning of the step sequence and a good knowledge of the payoff function in practice.

2.1 Risk minimisation in financial or energy markets operating in discrete time

In **[6]**, we investigate the risk minimisation problem of a portfolio loss with maturity T described by an $\mathbb{R}$-valued random variable L which contains an *observable* but *non-tradable* source of risk. More precisely we consider an energy market operating at discrete trading dates $t_0 = 0 < t_1 < \cdots < t_N = T$. We have d assets available for trade with price process $X = \left(X^1, \cdots, X^d\right)$ and $X^i = \left(X^i_\ell\right)_{0 \leqslant \ell \leqslant M}$ for $i = 1, \cdots, d$.

In our framework, the source of market incompleteness comes from the presence in L of a state process Z that is *observable* but *not available* for trade. Thus, it induces a source of risk that is not completely hedgeable. Typically, in the electricity market, the loss L suffered by an energy company may be due to an anormal annual electricity (or gas) consumption. This consumption depends on the temperature, which is an *observable* but non tradable source of risk. In this example the process $(Z_\ell)_{1 \leqslant \ell \leqslant M}$ can be considered as the temperature which may influence not only the loss but the assets available for trade, *i.e.* electricity prices of spot and forward contracts (which are in this example the only assets available for hedge). The probability space is equipped with a filtration $\mathbb{F} = (\mathcal{F}_\ell)_{0 \leqslant \ell \leqslant M}$. Intuitively, $\mathcal{F}_\ell$ represents the observable information at time t_ℓ by all investors, so that $\mathcal{F}_\ell = \sigma\{X_i, Z_i; 0 \leqslant i \leqslant \ell\}$.

The gains from a trading strategy θ with an initial investment of 0 are described by the discrete stochastic integral $\sum_{\ell=1}^N \theta_{\ell-1}.\Delta X_\ell$, where we denote by ΔX_ℓ the increments $X_\ell - X_{\ell-1}$ and $\theta = (\theta_\ell)_{0 \leqslant \ell < M} \in \mathcal{A}_\mathcal{F}$, where $\mathcal{A}_\mathcal{F} = \left\{\theta = (\theta_\ell)_{0 \leqslant \ell \leqslant M-1} \mid \theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell, \mathbb{P}),\ \ell = 0, \cdots,\ M-1\right\}$ is the set of admissible strategies, $L^0_{\mathbb{R}^d}(\mathcal{F}_\ell, \mathbb{P})$ denoting the space of all $\mathcal{F}$-mesurable and $\mathbb{P} - a.s.$ finite random variables with values in $\mathbb{R}^d$. In order to measure the risk, we consider a class of risk measures defined on $L^p(\mathbb{P})$ in terms of shortfall risk. It is a standard class of convex risk measure as mentioned by Föllmer and Schied [41] and Kaina and Rüschendorf [69].

To be more precise about this risk measure, let $\ell : \mathbb{R} \to \mathbb{R}_+$ be a non-decreasing convex loss function, not identically constant, hence satisfying $\lim_{x \to +\infty} \ell(x) = +\infty$. Assume that $\lim_{x \to -\infty} l(x) = \underline{l} < l(x_0)$. For some $x_0 \in \mathbb{R}$, we define the acceptance set

$$\mathcal{A} := \{L \in L^p : \mathbb{E}[\ell(-L)] \leqslant \ell(x_0)\},$$

and the corresponding *shortfall convex risk measure* $\rho : L^p(\mathbb{P}) \to \mathbb{R} \cup \{+\infty\}$

$$\rho(L) := \inf\{\xi \in \mathbb{R} : L + \xi \in \mathcal{A}\} = \inf\{\xi \in \mathbb{R} : \mathbb{E}[\ell(-L-\xi)] \leqslant l(x_0)\}, \tag{2.1}$$

Under the assumption on the function ℓ, the infimum in (2.1) exists. Assuming that $\forall M > 0,\ \mathbb{E}[l(-L+M)] < +\infty$, one has $\rho(L) \in \mathbb{R}$. Moreover, it is the lowest solution of the equation:

$$\mathbb{E}[\ell(-L-\xi)] = l(x_0). \tag{2.2}$$

Classical examples are the *entropic risk measure* defined by setting $\ell(x) = e^{\lambda x}$, $x \in \mathbb{R}$, $\lambda > 0$, $x_0 \in \mathbb{R}$, $\rho(L) = \frac{1}{\lambda}\log\mathbb{E}\left[e^{-\lambda(L+x_0)}\right]$ and the *L^p-shortfall risk measure*, defined by setting $\ell(x) = x^p\mathbf{I}_{\{x\geqslant 0\}}$, $x \in \mathbb{R}$, $p \geqslant 1$, $x_0 > 0$.

The basic problem for the investor with a given portfolio loss L is to minimise the risk

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}} \rho\left(L - \sum_{n=1}^{N}\theta_{n-1}.\Delta X_n\right) = \inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\inf\left\{\xi \in \mathbb{R} : \mathbb{E}\left[\ell(-L + \sum_{n=1}^{N}\theta_{n-1}.\Delta X_n - \xi)\right] \leqslant \ell(x_0)\right\}. \tag{2.3}$$

The main difference between the CVaR risk minimisation problem investigated in **[3]** and the shortfall risk minimisation problem (2.3) is that a shortfall risk measure does not write as an expectation contrary to the CVaR but appears as the *level* of a function which can be written as the expectation of a loss. Hence the arguments developed in **[3]** do not work.

Under the absence of arbitrage, we prove that the following two steps strategy works:

- the first step consists in solving for every $\xi \in \mathbb{R}$, the following stochastic control problem,

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\ell(-L + \sum_{n=1}^{N}\theta_{n-1}.\Delta X_n - \xi)\right] \tag{2.4}$$

  By performing a non-trivial dynamic programming argument and using a measurable selection theorem, see e.g. p. 266 Lemma 5 in Jacod and Shiryayev [67] or Dellacherie and Meyer [35] Theorem 82, p. 252, we prove the existence of optimal strategies for the risk minimisation problem (2.4) under the absence of arbitrage property (NA). [1]The optimal strategy is computed by a backward induction

$$\varphi_N(\xi,\theta_1,\cdots,\theta_{N-1}) := \ell(-L + \sum_{n=1}^{N}\theta_{n-1}.\Delta X_n - \xi),\quad a.s.$$

  for all $\xi \in \mathbb{R},\ \theta \in \mathcal{A}_{\mathcal{F}}$; for $0 \leqslant t < N$ and for all $\xi \in \mathbb{R}$, $(\theta_1,\cdots,\theta_t) \in L^0_{\mathcal{F}_0}(\mathbb{P}) \times \cdots \times L^0_{\mathcal{F}_{t-1}}(\mathbb{P})$,

$$\varphi_t(\xi,\theta_1,\cdots,\theta_t) := \operatorname*{ess\,inf}_{\theta_{t+1}\in\Theta_t} \mathbb{E}\left[\varphi_{t+1}(\xi,\theta_1,\cdots,\theta_t)\middle|\mathcal{F}_t\right],\quad a.s. \tag{2.5}$$

- once an optimal solution to (2.4) is computed (if any), say $\theta^*(\xi)$, $\xi \in \mathbb{R}$, the second step consists in finding the lowest solution ξ^* of the following equation

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\ell(-L + \sum_{n=1}^{N}\theta_{n-1}.\Delta X_n - \xi)\right] = \mathbb{E}\left[\ell(-L + \sum_{n=1}^{N}\theta^*_{n-1}(\xi).\Delta X_n - \xi)\right] = \ell(x_0). \tag{2.6}$$

Then we are concerned by the problem of computing the shortfall risk (2.1) of a random variable X and the optimal strategy for the risk minimization problem (2.3). In order to estimate the shortfall risk, which is a solution to equation (2.2), we propose a Robbins-Monro stochastic algorithm. Assuming that there exists a positive function φ such that $\mathbb{E}\left[\ell^2(-L-\xi)\right] \leqslant C(1+\varphi(\xi))$, $\xi \in \mathbb{R}$, we implement

$$\xi_{n+1} = \xi_n + \gamma_{n+1}H(\xi_n, L^{(n+1)}),\quad n \geqslant 0,\quad H(\xi,x) := \frac{\ell(-x-\xi) - l(x_0)}{(1+\varphi(\xi))^{\frac{1}{2}}} \tag{2.7}$$

where $(L^{(n)})_{n\geqslant 1}$ is an i.i.d. sequence of random variables with the same distribution as L, independent of $\xi_0 \in L^2(\mathbb{P})$ and $(\gamma_n)_{n\geqslant 1}$ is a step sequence satisfying the classical condition (1.9). We prove that the sequence $(\xi_n)_{n\geqslant 0}$ *a.s.* converges to a random variable $\rho(L)$ taking values in the set $\mathcal{T}^* := \{\xi : \mathbb{E}[\ell(-L-\xi)] = -\ell(x_0)\}$ as $n \to \infty$.

[1] Denoting $V_t^{v,\theta} = v + \sum_{\ell=1}^{t}\theta_\ell.\Delta X_\ell$, one has (NA) : $\forall\theta \in \mathcal{A}_{\mathcal{G}}$, $\left(V_M^{0,\theta} \geqslant 0\ a.s. \Longrightarrow\ V_M^{0,\theta} = 0\ a.s.\right)$

Moreover, if one implements the recursive scheme (2.7) with a step sequence $\gamma_n = \frac{a}{(b+n)^\beta}$, $a, b > 0$, $\frac{1}{2} < \beta < 1$ and computes the empirical mean $\bar{\xi}_{n+1} = \frac{1}{n+1}\sum_{k=0}^{n} \xi_k$, $n \geqslant 0$ then according to the Ruppert and Polyak averaging principle one has

$$\sqrt{n}(\bar{\xi}_n - \rho(X)) \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}\left(0, \sigma_*^2\right), \quad \sigma_*^2 := \frac{\mathbb{E}\left[(\ell(-L-\rho(L)))^2\right] - \ell^2(x_0)}{\mathbb{E}\left[\ell'(-L-\rho(L))\right]^2}.$$

Similarly to CVaR hedging we first investigate the case of static hedging, that is, the holder of the portfolio only uses one step strategy. In this case, the problem (2.4) and (2.6) can be written

$$\inf_{\theta\in\mathbb{R}^d} \mathbb{E}\left[\ell(-L+\theta.X-\xi)\right] = \ell(x_0), \quad \text{with } X := X_N - X_0. \tag{2.8}$$

which amounts to the computation of $(\xi^*, \theta^*) \in \mathbb{R}\times\mathbb{R}^d$ of the unique solution (for simplicity ℓ is assumed to be increasing and strictly convex) to the following systems of $d+1$ non-linear equations of $d+1$ unknowns

$$\mathbb{E}\left[\ell(-H+\theta.X-\xi)\right] = \ell(x_0), \quad \mathbb{E}\left[X\ell'\left(-H+\theta.X-\xi\right)\right] = 0. \tag{2.9}$$

In order to solve this system, we implement the following Newton-Raphson's optimisation algorithm

$$z_{k+1} = z_k - Dh_n(z_k)^{-1}h_n(z_k), \quad k \geqslant 0, \quad z_0 := (\xi_0, \theta_0) \tag{2.10}$$

where we approximated h and Dh using Monte Carlo simulation with n samples

$$h_n(\xi,\theta) = \left(\frac{1}{n}\sum_{k=1}^{n} \ell(-L^{(k)}+\theta.X^{(k)}-\xi) - l(x_0), \frac{1}{n}\sum_{k=1}^{n} L^{(k)}\ell'\left(-L^{(k)}+\theta.X^{(k)}-\xi\right)\right)$$

and

$$Dh_n(\xi,\theta) = \begin{pmatrix} -\frac{1}{n}\sum_{k=1}^{n}\ell'(-L^{(k)}+\theta.X^{(k)}-\xi) & \frac{1}{n}\sum_{k=1}^{n} X^{(k)}\ell'\left(-L^{(k)}+\theta.X^{(k)}-\xi\right) \\ -\frac{1}{n}\sum_{k=1}^{n} X^{(k)}\ell''(-L^{(k)}+\theta.X^{(k)}-\xi) & \frac{1}{n}\sum_{k=1}^{n} X^{(k)}(X^{(k)})^T\ell''\left(-L^{(k)}+\theta.X^{(k)}-\xi\right) \end{pmatrix}$$

$(L^{(k)}, X^{(k)})_{1\leqslant k\leqslant n}$ being an i.i.d. sequence of random vectors having the distribution of (H, X). The recursive algorithm (2.10) is known to converge to (ξ_n^*, θ_n^*) unique solution to $h_n(\xi,\theta) = 0$ if $z_0 := (\xi_0, \theta_0)$ is sufficiently close to (ξ_n^*, θ_n^*). Moreover, we prove that $(\xi_n^*, \theta_n^*)_{n\geqslant 1}$ *a.s.* converges to (ξ^*, θ^*) as $n \to \infty$.

Finally, in the dynamic framework we propose a probabilistic algorithm to compute the optimal strategy using optimal Markovian quantization. As argued in Bally and Pagès [8] and Pagès and al. [111], Markovian quantization algorithms are suited for stochastic control problem in high dimensions. It consists in approximating the underlying Markov process by a process taking its values in optimized finite grids which take into account the fine structure of the Markov dynamics. We suppose that $\mathcal{F}_t := \sigma((X_s, Z_s), s \leqslant t)$, $t = 1, \cdots, N$ and restrict consideration to cases of strategies taking values in a bounded domain $A \subset \mathbb{R}^d$. Like in the CVaR hedging framework, the process Z is *observable* but not *available for trade*. We also assume that the loss is given by $L = \phi(x)$, $x \in \mathbb{R}^{d+q}$. We introduce the notation $V_t = \sum_{n=1}^{t} \theta_{n-1}.\Delta X_n$, $t \geqslant 1$ and $V_0 = -\xi$.

The dynamics of the discretized process $X := (X_t, Z_t)_{0\leqslant t\leqslant N}$ is given by a Gaussian Euler scheme with step h that is

$$X_{t+1} = X_t + b(X_t)h + \sigma(X_t)h^{1/2}\varepsilon_{t+1} := G_h(X_t, \varepsilon_{t+1}), \ t = 0, \cdots, N-1, \tag{2.11}$$

where $\varepsilon_{t+1} = h^{-\frac{1}{2}}(W_{t+h} - W_t) \sim \mathcal{N}(0, I_{d+q})$ independent of $\mathcal{F}_t$. Given a strategy $\theta \in \mathcal{A}_{\mathcal{F}}$, the dynamics of $(V_t)_{t\geqslant 0}$ is given by:

$$V_{t+1} = V_t + \theta_t.(X_{t+1} - X_t) = V_t + \theta_t.B(X_{t+1} - X_t) := H_h(X_t, V_t, \theta_t, \varepsilon_{t+1}), \tag{2.12}$$

where $B = (I_d, 0_{d\times q})$. The dynamic programming principle that we established can be solved recursively by the backward algorithm:

$$\overline{\varphi}_N(x,v) = l(-\phi(x)+v), \quad (x,v) \in \mathbb{R}^{d+q}\times\mathbb{R}, \tag{2.13}$$

$$\overline{\varphi}_t(x,v) = \inf_{\theta\in A} \mathbb{E}\left[\overline{\varphi}_{t+1}(X_{t+1}^x, V_{t+1}^{x,v,\theta})\Big| (X_t, V_t) = (x,v)\right], (x,v) \in \mathbb{R}^{d+q}\times\mathbb{R} \tag{2.14}$$

for $t = 0, \cdots, N-1$, where $X_{t+1}^x := G_h(x, \varepsilon_{t+1})$ and $V_{t+1}^{x,v,\theta} := H_h(x, v, \theta, \varepsilon_{t+1})$. At step 0, one has to solve $\overline{\varphi}_0(X_0, -\xi) = l(x_0)$. In the spirit of the (multi-dimensional) Markov chain approximation method for solving non-linear problems (see *e.g.* [8], [111]), we consider an approximating dynamic of $(X_t, V_t)_{0\leqslant t\leqslant N}$ by a process $(\hat{X}_t, \hat{V}_t)_{0\leqslant t\leqslant N}$ taking its value in a finite grid.

Let $\Gamma_V := (2\delta)\mathbb{Z} \cap \bar{B}(0,R) = \{v \in \mathbb{R} : v = 2\delta z, \text{ for some } z \in \mathbb{Z}, |v| \leqslant R\}$ be a fixed bounded lattice grid on $\mathbb{R}$. The projection on Γ_V according to the closest neighbour rule is denoted π_V. At each time t, we consider a finite grid

Table 2.1: Shortfall risk estimation and One Step risk hedging of Spark Spread option

l	No hedging	One step hedging	θ^*
l_{exp}	-18.60 $\pm$0.02119	-21.71 $\pm$0.01809	15.98
l_2	-2.79 $\pm$0.00349	-5.83 $\pm$0.01440	4.68
l_3	-2.05 $\pm$0.00197	-3.69 $\pm$0.01253	2.56
l_4	-1.73 $\pm$0.00132	-2.71 $\pm$0.00983	1.54

$\Gamma_t = \left\{x_t^1, \cdots, x_t^{N_t}\right\}$ on the state space $\mathbb{R}^{d+q}$, for the uncontrolled process X. The projection on the grid Γ_t will be denoted π_t. We consider a quantized Euler scheme $\hat{X}$ and a finite state space Markov chain $\hat{V}$ both defined by

$$\hat{X}_{t+1} = \pi_{t+1}\left(G_h(\hat{X}_t, \epsilon_{t+1})\right), \quad \hat{V}_{t+1} = \pi_V\left(H_h(\hat{X}_t, \hat{V}_t, \theta_{t+1}, \hat{X}_{t+1}\right), \quad t = 0, \cdots, N-1$$

and following [111], we approximate (2.13) and (2.14), by the following stochastic control problem

$$\hat{\varphi}_N(x,v) = l(-\phi(x)+v), \quad (x,v) \in \Gamma_N \times \Gamma_V, \tag{2.15}$$

$$\hat{\varphi}_t(x,v) = \inf_{\theta\in A} \mathbb{E}\left[\hat{\varphi}_{t+1}(\hat{X}_{t+1}, \hat{V}_{t+1}^{\theta})\Big| (\hat{X}_t, \hat{V}_t) = (x,v)\right] = \inf_{\theta\in A} \mathbb{E}\left[\hat{\varphi}_{t+1}(\hat{X}_{t+1}^{x}, \hat{V}_{t+1}^{x,v,\theta})\right], (x,v) \in \Gamma_t \times \Gamma_V \tag{2.16}$$

where $\hat{X}_{t+1}^x = \pi_{t+1}\left(G_h(x, \epsilon_{t+1})\right)$ and $\hat{V}_{t+1}^{x,v,\theta} = \pi_V\left(H_h(x,v,\theta,\hat{X}_{t+1}^x)\right)$, $t = 0, \cdots, N-1$.

Now, using the optimal dispatching of the grid sizes $(N_t)_{1\leqslant t\leqslant N}$ for the Markovian quantization of the process X and Theorem 3.1 in Pagès and al. [111], under some mild assumptions on the coefficients b and σ, for all $v \in \Gamma_V$ one has

$$|\overline{\varphi}_0(v) - \hat{\varphi}_0(v)| \leqslant C_1 \frac{N^{2-\frac{1}{\hat{p}}}}{(M/N)^{1/(d+q)}} + C_2 N\delta + C_3(\bar{p})\frac{N}{R^{\bar{p}-1}}, \tag{2.17}$$

where $M := N_1 + \cdots + N_N$, $\hat{p} := \max(2, p_1+1)$, $\bar{p} > 1$, C_1, C_2 and $C_3(\bar{p})$ are positive constants. Therefore, we deduce the convergence of the value function approximated by optimal quantization toward the value function of the original control problem and the convergence of $\hat{\xi}^*$ toward ξ^* as $1/M$, $N\delta$ and $N/R^{\bar{p}-1}$ go to zero.

As a simple numerical illustration of our approach we consider a portfolio composed of an exchange option between gas and electricity (called Spark Spread) of maturity $T = 1$ year. The payoof of such option is defined by $L = (S_T^e - h_R S_T^g - C)_+$. Since Electricity has very limited storage possibilities, the holder of this option hedges by trading only a gas forward contract of maturity T. We choose to model the price $(S_t^e)_{0\leqslant t\leqslant T}$ of the electricity spot contract and the price $(F_{t,T}^g)_{0\leqslant t\leqslant T}$ of the gas forward contract of maturity T by two correlated geometric Brownian motions, the correlation factor between the two Brownian motions is $\rho = 0.5$. The parameters are: heat rate $h_R = 4$BTU/kWh (BTU: British Thermal Unit), the generation costs $C = 3$\$/MWh, the two volatilities $\sigma_g = 0.4$, $\sigma_e = 0.8$, the correlation factor $\rho = 0.5$ and the electricity and gas initial spot prices are $S_0^e = 40$\$/MWh, $S_0^g = 3$\$/MMBTU. In this section, we only consider the one-step hedging and we refer the reader to **[6]** for numerical results in the dynamic framework. Results about shortfall risk estimations and one step hedging are summarised in Table 2.1. The first column corresponds to the loss function. Four loss functions are considered: $l_{exp}(x) = e^{\lambda x}$, $\lambda = 1/50$ and $x_0 = 0$; $l_p(x) = x^p \mathbf{I}_{\{x\geqslant 0\}}$, $x_0 = 1$. The second column corresponds to the risk without any hedging using the Robbins-Monro algorithm (2.7) with its confidence interval at level 95%. The third column is the estimate of the shortfall risk with a one step risk minimization strategy and the fourth is the estimate of the optimal strategy θ^*.

2.2 Quantisation based recursive importance sampling

In this section, we present **[5]** related to a quantisation based recursive importance sampling (IS) technique. We follow the idea of Pagès and Lemaire [92] who proposed a recursive IS stochastic gradient algorithm for the computation of $\mathbb{E}[F(X)]$ where $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, |.|_E)$ is a random vector taking values in a Banach space E and $F : E \to \mathbb{R}$ is a Borel function such that $F(X) \in L^2(\mathbb{P})$. We present briefly the procedure proposed in [92]. For sake of simplicity we confine the presentation to the case $E = \mathbb{R}^d$. Note that one can also consider a functional space of paths of a process $X = (X_t)_{t\in[0,T]}$ solution to the SDE (1.17).

The IS is done by translating the distribution of the underlying random vector. The main idea of importance sampling by translation applied to the computation of $\mathbb{E}[F(X)]$ is to use the invariance by translation of the Lebesgue measure, for every $\theta \in \mathbb{R}^d$,

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X+\theta)\frac{p(X+\theta)}{p(X)}\right], \tag{2.18}$$

where $p : \mathbb{R}^d \to \mathbb{R}$ is probability density function of X. Among all these random variables with the same expectation we want to select the one with the lowest variance, *i.e.* the one with lowest quadratic norm

$$Q(\theta) := \mathbb{E}\left[F^2(X+\theta)\frac{p^2(X+\theta)}{p^2(X)}\right] = \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right], \quad \theta \in \mathbb{R}^d. \tag{2.19}$$

Assuming that $\theta \mapsto \log p(x-\theta)$ is log-concave, $x \in \mathbb{R}^d$, and $\lim_{|\theta|\to\infty} p(x-\theta) = 0$ then one shows that Q is convex and goes to infinity at infinity so that $\arg\min Q$ is non-empty. Assume now that p is differentiable on $\mathbb{R}^d$. Under additional growth and integrability conditions that we do not detail here, Q is differentiable on $\mathbb{R}^d$ with a gradient given by

$$\nabla Q(\theta) := \mathbb{E}\left[F(X-\theta)^2 \underbrace{\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}}_{W(\theta,X)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}\right]. \tag{2.20}$$

Now if the growth of the function F is controlled by G (say G grows exponentially), under some integrability assumption, setting

$$\widetilde{W}(\theta,X) := \frac{e^{-2\rho|\theta|^b}}{1+G(-\theta)^{2c}}W(\theta,X), \quad K(\theta,X) := F(X-\theta)^2\widetilde{W}(\theta,X), \tag{2.21}$$

where ρ and c are some positive constants, one proves that the function K satisfies the linear growth assumption (1.12) of the Robbins-Monro theorem, see Corollary 1.1.2.

For instance, when $X \sim \mathcal{N}(0, I_d)$, taking $\rho = \frac{1}{2}$, $b = 2$, one gets

$$K(\theta,x) = \frac{1}{1+G(-\theta)^{2c}}F^2(x-\theta)(2\theta - x).$$

Coming back to the general case, the mean function of K has the same zeroes as ∇Q, namely one has

$$\left\{\theta \in \mathbb{R}^d \mid \mathbb{E}\left[K(\theta,X)\right] = 0\right\} = \left\{\theta \in \mathbb{R}^d \mid \nabla Q(\theta) = 0\right\}.$$

Since Q is convex $h := \nabla Q$ satisfies (1.11) and the same holds for the function $h : \theta \mapsto h(\theta) = \mathbb{E}[K(\theta,X)]$. Hence, according to the Robbins-Monro algorithm the sequence $(\theta_n)_{n\geqslant 0}$ defined by

$$\theta_{n+1} = \theta_n - \gamma_{n+1}K(\theta_n, X^{(n+1)}),\ \theta_0 \in \mathbb{R}^d, \tag{2.22}$$

a.s. converges to an $\arg\min Q$-valued (square integrable) random variable θ^*.

In the infinite dimensional setting when $E = \mathcal{C}([0,T],\mathbb{R}^d)$ (equipped with the supremum norm) in order to deal with the case where X is a continuous path-dependent diffusion process, a recursive scheme similar to (2.22) is proposed in [92]. However, one relies on the Girsanov theorem instead of the invariance by translation of the Lebesgue measure in order to derive a representation similar to (2.18).

From a numerical point of view, the procedure (2.22) suffers from several drawbacks. First the tuning of the algorithm needs a good knowledge of the behaviour of F at infinity. Moreover, in practical implementations, recursive importance sampling methods using stochastic approximation algorithm need specific tuning of the step sequence and the convergence may be slow even when the dimension of the underlying random vector X is low. As a consequence, sometimes one may prefer deterministic optimisation procedures when it is possible.

In **[5]**, we propose a deterministic version of the IS algorithm investigated in [92]. The main idea can be summed up as follows. Assume that X with probability density function p, supposed to be two times continuously differentiable, can be approximated by a discrete random variable taking N values and whose distribution is known. Let us denote by $(x_k)_{1\leqslant k\leqslant N}$ its possible values and by $(p_k)_{1\leqslant k\leqslant N}$ the associated probabilities. For every $\theta \in \mathbb{R}^d$, the function Q given by (2.19) as well as ∇Q and $\nabla^2 Q$ can be easily approximated by

$$\begin{aligned} Q_N(\theta) &= \sum_{k=1}^N p_k\, F^2(x_k)\frac{p(x_k)}{p(x_k-\theta)}, \quad \nabla Q_N(\theta) = \sum_{k=1}^N p_k\, F^2(x_k)\frac{p(x_k)}{p^2(x_k-\theta)}\nabla p(x_k-\theta), \\ \nabla^2 Q_N(\theta) &= \sum_{k=1}^N p_k\, F^2(x_k)\frac{p(x_k)}{p(x_k-\theta)}\left(2\frac{\nabla p(x_k-\theta)\nabla p(x_k-\theta)^t}{p^2(x_k-\theta)} - \frac{\nabla^2 p(x_k-\theta)}{p(x_k-\theta)}\right). \end{aligned} \tag{2.23}$$

Hence, one may devise a classical Newton-Raphson algorithm, namely

$$\theta_{k+1}^N = \theta_k^N - \nabla^2 Q_N(\theta_k^N)^{-1}\nabla Q_N(\theta_k^N), \quad \theta_0 \in \mathbb{R}^d, \quad k \geqslant 0 \tag{2.24}$$

to approximate the optimal (variance) minimiser θ^* of Q. Starting from that idea, Jourdain and Lelong [68] proposes to replace X, where X is assumed to be a d-dimensional standard normal random vector, by a close discrete random variable obtained by a Monte Carlo simulation. The (unique) minimum of Q is approximated by the (unique) zero θ^N of ∇Q_N which can be computed using the deterministic Newton-Raphson algorithm (2.24) involving ∇Q_N and $\nabla^2 Q_N$. Moreover, several asymptotic properties are addressed.

Here, we propose to replace the Monte Carlo phase by a vector or functional quantization phase of X. Optimal quadratic quantisation of the random vector X consists basically of finding the best approximation (in $L^2(\mathbb{P})$) of X by a discrete random variable $\widehat{X}^N$ (the quantization of X) taking at most N values. When X takes its values in $\mathbb{R}^d$, one speaks about vector quantization. When X takes its values in an infinite dimensional Hilbert space like $L^2([0,T],dt)$ endowed with the usual norm $|u|_2 = \left(\int_0^T u(t)^2 dt\right)^{\frac{1}{2}}$, one speaks about functional quantization.

For $N \geqslant 1$ and $x := (x_1, \cdots, x_N) \in E^N$ (E will either be $\mathbb{R}^d$ or $L^2([0,T],dt)$ in what follows) an N-tuple referred as an $N-$ quantizer, we let $\mathrm{Proj}_x : E \to \{x_1, \cdots, x_N\}$ be a Borel projection following the nearest neighbour rule. Then, the Borel partition $C = \{C_1, \cdots, C_N\}$ of E defined by $C_k :=\mathrm{Proj}_x^{-1}(x_k)$, $k = 1, \cdots, N$ and satisfying

$$\mathrm{Proj}_x^{-1}(\{x_k\}) \subset \left\{ y \in E, |x_k - y|_E = \min_{1 \leqslant j \leqslant N} |x_j - y|_E \right\}, \quad 1 \leqslant k \leqslant N,$$

is called a Voronoi tessellation of E induced by x and one defines accordingly the Voronoi quantization of X induced by x as follows

$$\widehat{X} = \mathrm{Proj}_x(X) = \sum_{k=1}^{N} x_k \mathbf{I}_{\{X \in C_k\}}.$$

Note that the discrete random variable $\widehat{X}$ is the best $L^p(\mathbb{P})$-approximation of X among all measurable random variable taking values in $x := \{x_1, \cdots, x_N\}$. For any fixed N-quantizer x, we associate the $L^p(\mathbb{P})$-mean error $\mathbb{E}[|X - \widehat{X}|^p]^{1/p}$ induced by x and computing an optimal quantization of X consists in finding an $N-$tuple $x \in E^N$ which minimises the L^p-mean error over E^N. It amounts to minimising the function $\Lambda_N^X : (x_1, \cdots, x_N) \mapsto \mathbb{E}[|X - \widehat{X}|^p]^{1/p} = \mathbb{E}[(\min_{1 \leqslant i \leqslant N} |X - x_i|_E)^p]^{1/p}$.

One shows that, if $X \in L_E^p(\mathbb{P})$, the function Q_N^X reaches its minimum (at least) at one N-tuple x^* called an optimal N-quantizer. This minimum is in general not unique, except in some cases, *e.g.* when $d = 1$ and the density of X is log-concave. Moreover, the L^p-mean quantization error $e_{N,p}^X := \min_{E^N} Q_N^X$ converges toward 0 as $N \uparrow \infty$ and for "non-singular" $\mathbb{R}^d$-valued random vectors, the rate of convergence of $e_{N,p}^X$ is ruled by the so-called Zador Theorem (see Graf and Luschgy [57]) which asserts that $\lim_{N \to +\infty} N^{1/d} e_{N,p}^X = q_p(d) \left(\int f(x)^{d/(d+p)} dx\right)^{(d+p)/dp}$ if $X \in L_{\mathbb{R}^d}^{p+\delta}(\mathbb{P})$ where f denotes the density of the absolutely continuous part of $\mathbb{P}_X$ and $q_p(d)$ a positive constant depending on p and d.

The main idea now is that we know that $\widehat{X}$ is close to X in distribution and if one has a numerical access to the optimal N-quantizer x with the associated weights sequence $(p_k = \mathbb{P}(X \in C_k(x)))_{1 \leqslant k \leqslant N}$ of the quantisation $\widehat{X}$ then the computation of the functions Q_N, ∇Q_N and $\nabla^2 Q_N$ are dramatically faster compared to any other simulation based approach. For instance, a sharply optimised database of quantizers of standard univariate and multivariate Gaussian distributions is available on the web site `www.quantize.maths-fi.com` [112] for download. To an N-quantizer x we associate the unique minimum $\hat{\theta}^N$ of Q_N, that is, the unique solution to the equation $\nabla Q_N(\theta) = 0$. One expects that $\hat{\theta}^N \to \theta^*$ as the size N of its associated N-quantizer x goes to infinity.

Under the assumptions that the payoff function F is Lipschitz-continuous, the density p is strictly log-concave with $\lim_{|x| \to +\infty} p(x) = 0$ then under some growth condition on the first and second derivatives of p that we do not detail here, one gets

$$\widehat{\theta}^N \to \theta^*, \, Q_N(\widehat{\theta}^N) \to Q(\theta^*) \, as \, N \to +\infty$$

as soon as $\mathbb{E}[|X - \widehat{X}^N|^2] \to 0$ as $N \to +\infty$.

In order to obtain a convergence rate, we assume that $\nabla p/p$ is η-Hölder. For instance, for the multi-dimensional Gaussian law $\eta = 1$. Under these additional assumption, one obtains the following convergence rate $|\widehat{\theta}^N - \theta^*| = \mathcal{O}\left(\left\|X - \widehat{X}^N\right\|_2^{\alpha}\right)$ as $N \to \infty$. For instance, when X is a d-dimensional Gaussian vector, one has $|\widehat{\theta}^N - \theta^*| \leqslant C N^{-1/d}$.

We also extend our approach to the infinite dimensional setting, *i.e.* the case of path-dependent diffusion. Instead of the invariance by translation of the Lebesgue measure, we rely on the Girsanov transform to play the role of mean translator. To be more precise, we consider a d-dimensional Itô process X solution to the SDE

$$X_t = x + \int_0^t b(s, X^s) ds + \int_0^t \sigma(s, X^s) dW_s \tag{2.25}$$

where $W = (W_t)_{t \in [0,T]}$ is a q-dimensional Brownian motion and where $X^t := (X_{t \wedge s})_{s \in [0,T]}$ (with $X = X^T$) is the stopped process at time t. We assume that a strong solution to the SDE (2.25) exists. In order to devise a quantized

IS procedure for $\mathbb{E}[F(X)]$, F being a Borel map defined on $\mathcal{C}\left([0,T],\mathbb{R}^d\right)$ we consider a translation functional given by $\theta \in L^2_{T,q} := L^2([0,T],\mathbb{R}^q)$ which is slightly less general than the one used in [92]. From the Girsanov Theorem it follows that for every $\theta \in L^2_{T,q}$

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X^{(\theta)})e^{-\int_0^T \langle \theta_s, \mathrm{d}W_s\rangle - \frac{1}{2}\|\theta\|^2_{L^2_{T,q}}}\right]$$

where $X^{(\theta)}$ denotes the unique strong solution to (2.25) with drift $b+\sigma\theta$. Among all these estimators we want to select the one with the lowest quadratic norm so that we want to solve the following minimization problem

$$\min_{\theta\in L^2_{T,q}} Q(\theta) \quad \text{where} \quad Q(\theta) := \mathbb{E}\left[F^2(X^{(\theta)})e^{-2\int_0^T \langle \theta_s, \mathrm{d}W_s\rangle - \|\theta\|^2_{L^2_{T,q}}}\right] = \mathbb{E}\left[F^2(X)e^{-\int_0^T \langle \theta_s, \mathrm{d}W_s\rangle + \frac{1}{2}\|\theta\|^2_{L^2_{T,q}}}\right].$$

Note that Q is strictly convex and $\lim_{\|\theta\|^2_{L^2_{T,q}}\to\infty} Q(\theta) = +\infty$ so that it admits a unique minimum. Moreover, Q is twice differentiable at any $\theta \in L^2_{T,q}$ and $\langle DQ(\theta), \psi\rangle_{L^2_{T,q}}$, $(D^2Q(\theta)\psi, \phi)$, for $\psi, \phi \in L^2_{T,q}$ admit a representation as an expectation.

In view of a numerical implementation of Newton-Raphson's algorithm to estimate a minimum of Q, we are led to consider a (non-trivial) finite dimensional subspace E of $L^2_{T,p}$ spanned by an orthonormal basis $(e_1,\cdots,e_m)$. Like for the finite dimensional framework, our procedure will be based on the representation (as an expectation) of the first and second differential of Q on E combined with functional quantization of the solution to (2.25).

The functional quantization of (2.25) is based on the Karhunen-Loève (K-L) decomposition $W \overset{L^2_T}{=} \sum_{n\geqslant 1}\sqrt{\lambda_n}\xi_n e_n$ where $(e_n)_{n\geqslant 1}$ is the K-L orthonormal basis, λ_n are the corresponding eigenvalues of the covariance operator associated to the one-dimensional Brownian motion W and $\left\{\xi_n := \langle W, e_n\rangle/\sqrt{\lambda_n},\ n\geqslant 1\right\}$, is an *i.i.d.* sequence of random variables with standard normal distribution. One considers an optimal N_n-quantization ($N_n \geqslant 1$) of ξ_n, denoted $\hat{\xi}_n$ where $\hat{\xi}_n := \mathrm{Proj}_{x_n}(\xi_n)$, $x_n := \left(x_1^{N_n},\cdots,x_{N_n}^{N_n}\right)$ is the unique optimal N_n-quantizer of the normal distribution and $N_1\times\cdots\times N_n \leqslant N$, with $N_1,\cdots,N_n \geqslant 1$. For large enough n, we set $N_n = 1$, $\hat{\xi}_n = 0$ (which is the optimal 1-quantization of $\mathcal{N}(0,1)$) and we define the product quantizer by (the finite sum) $\widehat{W}_t = \sum_{n\geqslant 1}\sqrt{\lambda_n}\hat{\xi}_n e_n(t)$. The product quantizer χ that produces the above Voronoi quantization $\widehat{W}$ is defined by $\chi_{\underline{i}}(t) = \sum_{n\geqslant 1}\sqrt{\lambda_n}x_{i_n}e_n(t)$, $\underline{i} = (i_1,\cdots,i_n,\cdots) \in \prod_{n\geqslant 1}\{1,\cdots,N_n\}$ in the sense that $\mathbb{P}\left(\widehat{W} = \chi_{\underline{i}}\right) = \prod_{n\geqslant 1}\mathbb{P}\left(\xi_n \in C_{i_n}(x_n)\right)$ with an approximation rate $\left\|W - \widehat{W}^N\right\|_2 \leqslant C_T(\log N)^{-\frac{1}{2}}$. Then, for each $\underline{i}$, one solves the following ODEs

$$dx_{\underline{i}}(t) = \left(b(x_{\underline{i}}(t)) - \frac{1}{2}\sigma\sigma'(x_{\underline{i}}(t))\right)dt + \sigma(x_{\underline{i}}(t))d\chi_{\underline{i}}(t), \tag{2.26}$$

and set $\widetilde{X}^N_t = \sum_{\underline{i}} x_{\underline{i}}(t)\mathbf{I}_{\{C_{\underline{i}}(\chi)\}}(W)$, $N\geqslant 1$. Assuming that b, σ are continuously differentiable, this non-Voronoï N-functional quantization of (2.25) produces an error rate $\left\|X - \widetilde{X}^N\right\|_2 \leqslant C_T(\log N)^{-\frac{1}{2}}$. We approximate Q by $\widetilde{Q}_N$, the m-tuple $\langle DQ(\theta), e_i\rangle_{L^2_{T,q}}$ by $J(\theta)_i = \left\langle D\widetilde{Q}_N(\theta), e_i\right\rangle_{L^2_{T,q}}$, $1\leqslant i\leqslant m$ and the $m\times m$ symmetric positive definite matrix $\left(D^2Q(\theta)e_i, e_j\right)$ by $H(\theta)_{i,j} := \left(D^2\widetilde{Q}_N(\theta)e_i, e_j\right)$, $1\leqslant i\leqslant m$, $1\leqslant j\leqslant m$, for every $\theta\in E$ where we replaced the random variables (X,W) by their quantized version $(\widetilde{X}^N, \widehat{W}^N)$ in the representation as an expectation of Q, DQ and D^2Q. Hence, we compute the minimum $\widetilde{\theta}^N$ of $\widetilde{Q}_N$ by devising the following Newton-Raphson algorithm

$$\widetilde{\theta}^N_{k+1} = \widetilde{\theta}^N_k - H(\widetilde{\theta}^N_k)^{-1}J(\widetilde{\theta}^N_k), \quad k\geqslant 0, \quad \widetilde{\theta}^N_0 \in E \ \text{ given.} \tag{2.27}$$

We briefly illustrate the performance of the algorithm in the infinite dimensional setting by considering three different basis of $L^2([0,1],\mathbb{R})$

- a polynomial basis composed of the shifted Legendre polynomials $(\tilde{P}_n)_{n\geqslant 0}$ defined by

$$\forall n\geqslant 0,\ \forall t\in[0,1], \quad \tilde{P}_n(t) = P_n(2t-1) \ \text{ where } \ P_n(t) = \frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\left(\left(t^2-1\right)\right). \tag{ShLeg}$$

- the Karhunen-Loève basis defined by

$$\forall n\geqslant 0,\ \forall t\in[0,1], \quad e_n(t) = \sqrt{2}\ \sin\left(\left(n+\frac{1}{2}\right)\pi t\right). \tag{KL}$$

- the Haar basis which is defined by

$$\forall n \geqslant 0,\ \forall k = 0, ..., 2^n - 1,\ \forall t \in [0,1],\ \psi_{n,k}(t) = 2^{\frac{k}{2}}\psi(2^k t - n), \quad \psi(t) = \begin{cases} 1 & \text{if } t \in [0, \frac{1}{2}) \\ -1 & \text{if } t \in [\frac{1}{2}, 1) \\ 0 & \text{otherwise.} \end{cases} \tag{Haar}$$

We consider a Down&In Call option of strike K and barrier L. This option is activated when the underlying process X moves down and hits the barrier L. The payoff function at maturity T is defined by

$$F(X) = (X_T - K)_+ \mathbf{1}_{\{\min_{0 \leqslant t \leqslant T} X_t \leqslant L\}}.$$

A standard approach to price the option is to consider the continuous Euler scheme $\bar{X}$ of step $t_k = k\frac{T}{M}$ obtained by extrapolation of the Brownian Motion between two instants of discretization. For every $t \in [t_k, t_{k+1}]$, we can write

$$\bar{X}_t = \bar{X}_{t_k} + b(\bar{X}_{t_k})(t - t_k) + \sigma(\bar{X}_{t_k})(W_t - W_{t_k}),\ \bar{X}_0 = x_0 \in \mathbb{R}.$$

By pre-conditioning,

$$\mathbb{E}\left[(\bar{X}_T - K)_+ \mathbf{1}_{\left\{\min_{0 \leqslant t \leqslant T} \bar{X}_t \leqslant L\right\}}\right] = \mathbb{E}\left[(\bar{X}_T - K)_+ \left(1 - \prod_{p=0}^{M-1} p(\bar{X}_{t_p}, \bar{X}_{t_{p+1}})\right)\right], \tag{2.28}$$

where $p(x_p, x_{p+1}) = \mathbb{P}\left(\min_{t_p \leqslant t \leqslant t_{p+1}} \bar{X}_t \geqslant L \,\middle|\, (\bar{X}_{t_p}, \bar{X}_{t_{p+1}}) = (x_p, x_{p+1})\right)$ is the probability of non-exit from $[L, +\infty)$ of some Brownian bridge. Using the law of the Brownian bridge (see e.g. Gobet [52]), we can write

$$p(x_p, x_{p+1}) = 1 - \mathbb{P}\left(\min_{t \in [0, t_1]} W_t \leqslant \frac{L - x_p}{\sigma(x_p)} \,\middle|\, W_{t_1} = \frac{x_{p+1} - x_p}{\sigma(x_p)}\right) = \begin{cases} 1 - e^{-\frac{2(L - x_p)(L - x_{p+1})}{(t_{p+1} - t_p)\sigma^2(x_p)}}, & L \leqslant \min(x_p, x_{p+1}), \\ 0 & \text{, otherwise.} \end{cases} \tag{2.29}$$

Hence we run ou algorithm with the probabilistic representation appearing in the right-hand side of (2.28). We consider the local volatility model $b(x) = rx$, $\sigma(x) = \sigma x^{1+\beta}/(1 + x^2)^{\frac{1}{2}}$, with $r = 4\%$, $\sigma = 5$, $x = 100$, $\beta = 0.5$. The number of Monte Carlo simulations n is 5.10^4 in every case. We set the optimal product quantizer at level $d_N = 966$ which corresponds to the optimal decomposition $N_1 = 23$, $N_2 = 7$, $N_3 = 3$, $N_4 = 2$ for the product quantization $(\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3, \hat{\xi}_4)$. The numerical results are summarised in Table 2.2. In Figure 2.1, the optimal variance reducers for the local volatility model are depicted

Basis	**m**	**Price MC**	**Variance MC**	**Price QIS**	**Variance QIS**
Constant	1	0.684	25.80	0.673	15.33
Legendre	2	0.711	27.92	0.662	4.58
(ShLeg)	4	0.683	25.66	0.684	3.46
	8	0.686	26.59	0.685	3.35
Karhunen-Loève	2	0.680	25.38	0.696	5.22
(KL)	4	0.702	26.39	0.683	6.53
	8	0.687	26.39	0.688	5.80
Haar	2	0.648	24.90	0.673	8.15
(Haar)	4	0.671	25.15	0.692	5.46
	8	0.709	30.17	0.700	5.29

Table 2.2: Down&In Call option in Local volatility model with $X_0 = 100$, $K = 115$, $L = 65$, $T = 1$, $\sigma = 5$, $r = 4\%$, $\beta = 0.5$, $d_N = 966$, $n = 5.10^4$, $M = 100$.

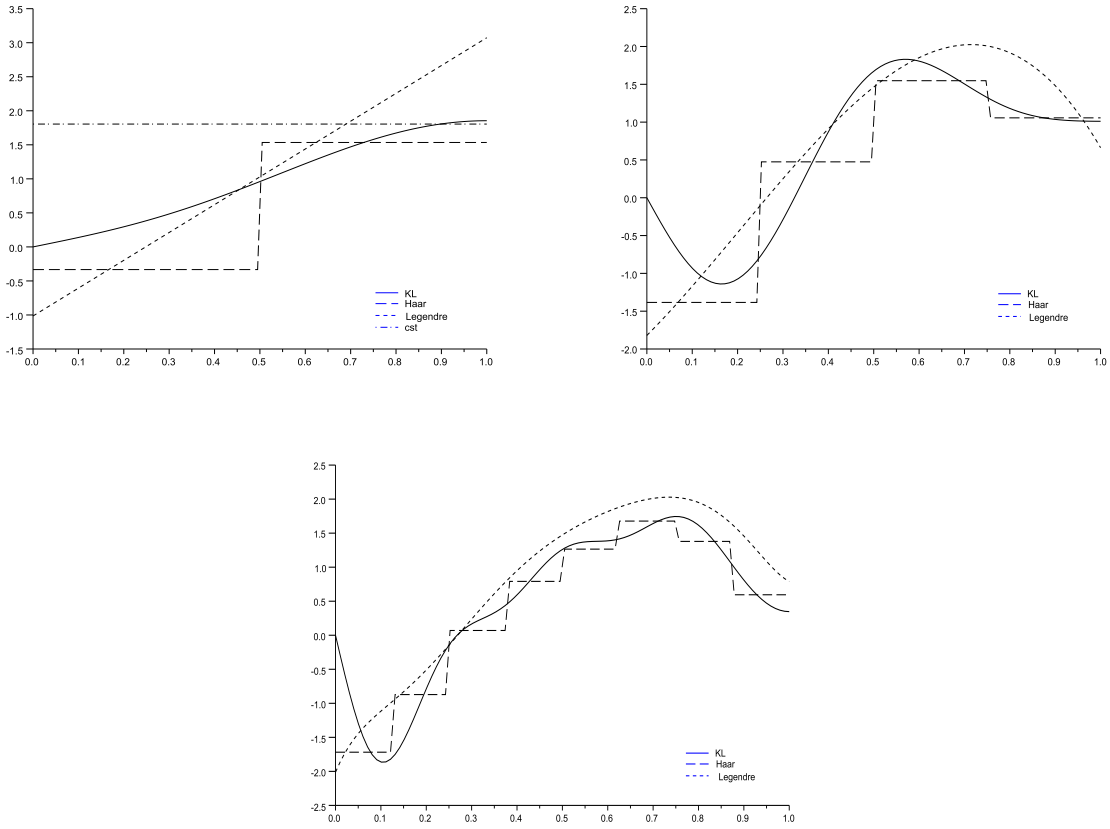

Figure 2.1: Down & In Call option: Optimal θ^N obtained by our algorithm in the case of the Local Volatility model for different basis and several values of m ($m = 2$ for the left upper curves, $m = 4$ for the right upper curves and $m = 8$ for the lower curves).

Chapter 3

Concentration inequalities and deviation estimates for stochastic approximation schemes

In this chapter, we present **[7]** which is a joint work with S. Menozzi from the university of Evry Val d'Esonne and **[8]** which is a joint work with M. Fathi from the university Paul Sabatier of Toulouse. Both works are related to concentration inequalities and deviation estimates for stochastic approximation schemes. Here we are interested in two kinds of discrete stochastic approximation schemes. The first one refers to the law of an Euler like discretisation scheme of a diffusion process at a fixed deterministic date and the second one concerns the law of a stochastic approximation algorithm at a given time-step.

Concentration inequalities consists in establishing *a non-asymptotic* Gaussian **[7]** or sub-Gaussian **[8]** control of the probability that the law of the scheme deviates from its target. More precisely, we consider the deviations between the expectation of a given function of an Euler like discretisation scheme of some diffusion process at a fixed deterministic time and its empirical mean obtained by the Monte Carlo procedure. We then consider some estimates concerning the deviation between the value at a given time-step of a stochastic approximation algorithm and its target.

Much work on non asymptotic deviation estimates for the law of the Euler discretisation scheme has now been conducted. In the ergodic framework and for a constant diffusion coefficient Gaussian controls have been obtained by Malrieu and Talay [94]. For the *standard* Euler scheme, a first attempt to establish two-sided Gaussian bounds for the statistical error can be found in Lemaire and Menozzi [91] under some non-degeneracy conditions and up to a systematic bias. Very rencently, Honoré, Menozzi and Pagès [63] obtained non-asymptotic Gaussian concentration bounds for the difference between the invariant measure ν of an ergodic Brownian diffusion process and the empirical distribution of an approximating scheme with decreasing time step along a suitable class of smooth enough test functions f such that $f - \nu(f)$ is a *coboundary* of the infinitesimal generator of the process.

Such concentration inequalities are strongly connected to transport-entropy inequalities. In Section 3.1, we present this connection and recall basic definitions and properties. For a complete overview and recent developments in the theory of transport inequalities, the reader may refer to the recent survey of Gozlan and Leonard [56] and the book of Villani [126]. We establish Gaussian concentration inequalities in **[7]** for the two specific problems mentioned above under suitable assumptions. The key tool consists in exploiting accurately the concentration properties of the increments of the schemes and proving that this concentration transfers to the law of the scheme. We will explain this point in Section 3.2. Finally, in Section 3.3 we explain the results obtained in **[8]** which improve and complete those obtained in **[7]**. The key point is to properly quantify the contribution of the diffusion term to the concentration regime. We also derive a non-asymptotic bound for stochastic approximation with averaging of trajectories, in particular we prove that averaging a stochastic approximation algorithm with a slow decreasing step sequence gives rise to optimal concentration rate.

3.1 A brief presentation of transport inequalities and concentration inequalities

In this section, we present the connection between transport inequalities and concentration inequalities. We also recall some basic definitions and properties that will be needed in the two next sections. We denote by $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$.

For $p \geqslant 1$, we consider the set $\mathcal{P}_p(\mathbb{R}^d)$ of probability measures with finite moment of order p. The Wasserstein

metric $W_p(\mu,\nu)$ of order p between two probability measures $\mu,\nu\in\mathcal{P}_p(\mathbb{R}^d)$ is defined by

$$W_p^p(\mu,\nu)=\inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\pi(dx,dy):\ \pi\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d),\ \pi_0=\mu,\ \pi_1=\nu\right\}$$

where π_0 and π_1 are two probability measures standing for the first and second marginals of $\pi\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$. For $\mu\in\mathcal{P}(\mathbb{R}^d)$, we define the relative entropy w.r.t $\nu\in\mathcal{P}(\mathbb{R}^d)$ as

$$H(\mu,\nu)=\int_{\mathbb{R}^d}\log\left(\frac{d\mu}{d\nu}\right)d\mu$$

if $\mu\ll\nu$ and $H(\mu,\nu)=+\infty$ otherwise. We are now in position to define the notion of transport-entropy inequality. Here as below, $\Phi:\mathbb{R}_+\to\mathbb{R}_+$ is a convex, increasing function with $\Phi(0)=0$.

Definition 3.1.1. A probability measure μ on $\mathbb{R}^d$ satisfies a transport-entropy inequality with function Φ if for all $\nu\in\mathcal{P}(\mathbb{R}^d)$, one has

$$\Phi(W_1(\nu,\mu))\leqslant H(\nu,\mu).$$

For the sake of simplicity, we will write that μ satisfies T_Φ.

The following proposition comes from Corollary 3.4. in Gozlan and Leonard [56].

Proposition 3.1.1. *The following propositions are equivalent:*

- *The probability measure μ satisfies T_Φ.*
- *For all 1-Lipschitz function $f:\mathbb{R}^d\to\mathbb{R}$, one has*

$$\forall\lambda\geqslant 0,\quad\int\exp(\lambda f)d\mu\leqslant\exp\left(\lambda\int f d\mu+\Phi^*(\lambda)\right),$$

where Φ^* is the monotone conjugate of Φ defined on $\mathbb{R}_+$ as $\Phi^*(\lambda)=\sup_{\rho\geqslant 0}\{\lambda\rho-\Phi(\rho)\}$. Note that since Φ is a non-negative convex function which is non-decreasing with $\Phi(0)<\infty$, the monotone conjugate of Φ^* is Φ.

As an important example we will say that μ satisfies the *Gaussian concentration property* if there exists $\beta>0$ such that $\Phi(\lambda)=\lambda^2/\beta$, $\lambda\geqslant 0$ so that $\Phi^*(\lambda)=\lambda^2\beta/4$. In this chapter we will simply say that μ satisfies $GC(\beta)$.

From the Markov exponential inequality and $GC(\beta)$, one derives $\mathbb{P}[f(Y_1)-\mathbb{E}[f(Y_1)]\geqslant r]\leqslant\exp(-\lambda r+\frac{\beta\lambda^2}{4}),\forall\lambda,r\geqslant 0$. An optimisation over λ gives that this probability is bounded by $\exp(-\frac{r^2}{\beta})$.

A practical criterion for $GC(\beta)$ to hold is given by Bolley and Villani [24]. If there exists $\varepsilon>0$ s.t. $\int\exp(\varepsilon|y|^2)\mu(dy)<\infty$, then μ satisfies $GC(\beta)$ for some $\beta:=\beta(\varepsilon)$. The two claims are actually equivalent.

In **[7]** and **[8]**, $GC(\beta)$ is the only crucial property we require on the underlying innovations of the approximation schemes.

Such transport-entropy inequalities are very attractive especially from a numerical point of view since they are related to the concentration of measure phenomenon which allows to establish non-asymptotic deviation estimates. In particular, some *non-asymptotic confidence intervals* can be explicitly computed. This can be very useful in practice when the computational ressources are constrained.

The three next results put an emphasis on this point. Suppose that $(X^{(n)})_{n\geqslant 1}$ is a sequence of i.i.d. $\mathbb{R}^d$-valued random variables with common law μ.

Corollary 3.1.2. *If μ satisfies T_Φ then for all 1-Lipschitz function $f:\mathbb{R}^d\to\mathbb{R}$ and for all $r\geqslant 0$, for all $M\geqslant 1$, one has*

$$\mathbb{P}\left(|\frac{1}{M}\sum_{k=1}^M f(X^{(k)})-\mathbb{E}[f(X^{(1)})]|\geqslant r\right)\leqslant 2\exp(-M\Phi(r))$$

Deriving non-asymptotic deviation bounds for $W_1(\mu_M,\mu)$ is of interest for many applications in the fields of numerical probability and statistic. In its present form, next result is due to Gozlan and Leonard [55], Theorem 12.

Proposition 3.1.3. *If μ satisfies T_Φ then the empirical measure μ_M defined as $\mu_M=\frac{1}{M}\sum_{k=1}^M\delta_{X^{(k)}}$ satisfies the following concentration bound*

$$\mathbb{P}\left(W_1(\mu_M,\mu)\geqslant\mathbb{E}[W_1(\mu_M,\mu)]+r\right)\leqslant\exp\left(-M\Phi(r)\right).$$

where for $x\in\mathbb{R}^d$, δ_x stands for the Dirac mass at point x.

The quantity $\mathbb{E}[W_1(\mu_M,\mu)]$ will go to zero as M goes to infinity, by convergence of empirical measures, but we still need quantitative bounds. The next result is an adaptation of Theorem 10.2.1 in [116] on similar bounds but for the distance W_2. A proof can be found in the appendix of **[8]**.

Proposition 3.1.4. *Assume that μ has a finite moment of order $d+3$. Then, one has*

$$\mathbb{E}[W_1(\mu_M,\mu)] \leqslant C(d,\mu)M^{-1/(d+2)}$$

where

$$C(d,\mu) := 4\sqrt{d} + 2\sqrt{\int_{\mathbb{R}^d}(1+|x|^{d+1})^{-1}dx}\sqrt{2^{-2d}+2^{3-d}\int|y|^{d+3}\mu(dy)+2^{3-d}d(d+3)!}.$$

This bound is not optimal in general, but has the advantage of having very explicit constants. In the case of a distribution with compact support, it has been shown in [12], Section 7, that $\mathbb{E}[W_1(\mu_M,\mu)]$ is of order $O(M^{-1/d})$, and that this is the optimal exponent in d when $d \geqslant 3$.
In view of Kantorovich-Rubinstein duality formula, namely

$$W_1(\mu,\nu) = \sup\left\{\int f d\mu - \int f d\nu : [f]_1 \leqslant 1\right\}$$

where $[f]_1$ denotes the Lipschitz-modulus of f, the latter result provides the following concentration bounds

$$\forall r \geqslant 0,\ \forall M \geqslant 1,\quad \mathbb{P}\left(\sup_{f:[f]_1\leqslant 1}\left(\frac{1}{M}\sum_{k=1}^{M} f(X^{(k)}) - \mathbb{E}[f(X^{(1)})]\right) \geqslant C(d,\mu)M^{-1/(d+2)} + r\right) \leqslant \exp\left(-M\Phi(r)\right).$$

Similar results were first obtained for different concentration regimes by Bolley, Guillin and Villani [23] relying on a non-asymptotic version of Sanov's Theorem. Some of these results have also been derived by Boissard [22] using concentration inequalities, and were also extended to ergodic Markov chains up to some contractivity assumptions in the Wasserstein metric on the transition kernel.

Some applications are proposed in Bolley, Guillin and Villani [23]. Such results can indeed provide non-asymptotic deviation bounds for the estimation of the density of the invariant measure of a Markov chain. Let us note that the (possibly large) constant $C(d,\mu)$ appears as a trade-off to obtain uniform deviations over all Lipschitz functions.

In the two next sections we will present some transport-entropy inequalities and, as a consequence, some non-asymptotic deviation estimates for the laws at a given time step of two kinds of discrete-time and d-dimensional stochastic evolution scheme of the form

$$X_{n+1} = X_n + \gamma_{n+1}H(n,X_n,U_{n+1}),\ n \geqslant 0, X_0 = x \in \mathbb{R}^d, \tag{3.1}$$

where $(\gamma_n)_{n\geqslant 1}$ is a deterministic positive sequence of time steps, $(U_i)_{i\in\mathbb{N}*}$ is an i.i.d. $\mathbb{R}^q$-valued sequence of random variables defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ with law μ and the function $H : \mathbb{N}\times\mathbb{R}^d\times\mathbb{R}^q \to \mathbb{R}^d$ is a measurable function satisfying for all $x \in \mathbb{R}^d$, for all $n \in \mathbb{N}$, $H(n,x,.) \in \mathcal{L}^1(\mu)$, and $\mu(du)$-a.s., $H(n,.,u)$ is continuous. We will also assume that μ satisfies the *Gaussian concentration property* $GC(\beta)$.

As a consequence of the transport-entropy inequalities obtained for the laws at a given time step of Euler like schemes and stochastic approximation algorithm, we will derive in Section 3.3 non-asymptotic deviation bounds in the Wasserstein metric.

3.2 Gaussian concentration inequalities

In this section, we present the results obtained in **[7]** in collaboration with S. Menozzi from the university of Evry Val d'Essonne related to Gaussian concentration inequalities for two kinds of recursive scheme with dynamics (3.1). In Section 3.3, we will extend those results to other regimes ranging from exponential to Gaussian.

We consider the unique strong solution X to the SDE (1.17) where the underlying innovation process Z is a q-dimensional Brownian motion $(W_t)_{t\geqslant 0}$. In **[7]**, we assume that the coefficients satisfy the mild smoothness condition:

(A) The coefficients b,σ are uniformly Lipschitz continuous in space uniformly in time, σ is bounded.

Note that we do not assume any non-degeneracy condition on σ in **(A)**.

We do not consider the associated continuous Euler Maruyama scheme (1.19) but rather define an Euler like scheme. For a fixed time step $\Delta = T/N,\ N \in \mathbb{N}^*$, and the time grid $t_i := i\Delta$, for all $i = 0,\cdots,N$, we consider the following dynamics

$$X_0^\Delta = x,\quad X_{t_{i+1}}^\Delta = X_{t_i}^\Delta + b(t_i,X_{t_i}^\Delta)\Delta + \sigma(t_i,X_{t_i}^\Delta)\Delta^{1/2}U^{(i+1)},\quad i = 0,\cdots,N-1 \tag{3.2}$$

where $(U^{(i)})_{i\in\mathbb{N}^*}$ is a sequence of $\mathbb{R}^q$-valued i.i.d. random variables with law μ satisfying: $\mathbb{E}[U^{(1)}] = 0_q$, $\mathbb{E}[U^{(1)}(U^{(1)})^*] = I_q$, where $(U^{(1)})^*$ denotes the transpose of the column vector $U^{(1)}$ and $0_q, I_q$ respectively denote the zero vector of $\mathbb{R}^q$ and the identity matrix of $\mathbb{R}^q \otimes \mathbb{R}^q$. We also assume that μ satisfies $GC(\beta)$ for some $\beta > 0$. The main advantage of such a situation is that it includes the case of the standard Euler scheme where $U^{(1)} \overset{\mathrm{d}}{=} \mathcal{N}(0, I_q)$ and the case of the Bernoulli law where $U^{(1)} \overset{\mathrm{d}}{=} (B_1, \cdots, B_q)$, $(B_k)_{1\leqslant k\leqslant q}$ are i.i.d random variables with law $\mu = \frac{1}{2}(\delta_{-1} + \delta_1)$, both satisfying $GC(\beta)$ with $\beta = 2$.

As already mentioned, for practical purposes, we are interested in non-asymptotical controls of the *statistical error*, which is given by $\frac{1}{M}\sum_{i=1}^M f(X_T^{\Delta,(i)}) - \mathbb{E}[f(X_T^{\Delta})]$ where the $((X_T^{\Delta,(i)})^i)_{1\leqslant i\in\leqslant M}$ are independent copies of the scheme (3.2) starting at x at time 0 and evaluated at time T. Precisely, for a *fixed* M and a given threshold $r > 0$, one would like to give bounds on the quantity $\mathbb{P}[|\frac{1}{M}\sum_{i=1}^M f(X_T^{\Delta,(i)}) - \mathbb{E}[f(X_T^{\Delta})]| > r]$.

Our first main result obtained in **[7]** shows that when the innovations satisfy $GC(\beta)$, the Gaussian concentration property transfers to the statistical error.

Theorem 3.2.1 (Concentration Bounds for an Euler-like scheme)**.** *Denote by X_T^{Δ} the value at time T of the scheme* (3.2)*. Assume that the common law of the i.i.d. sequence $(U_i)_{i\in\mathbb{N}^*}$ in* (3.2) *satisfies $GC(\beta)$ for some $\beta > 0$ and that the coefficients b, σ satisfy* **(A)***. Let f be a real valued uniformly Lipschitz continuous function on $\mathbb{R}^d$. For all $M \in \mathbb{N}^*$ and all $r \geqslant 0$, one has*

$$\mathbb{P}[|\frac{1}{M}\sum_{i=1}^M f(X_T^{\Delta,(i)}) - \mathbb{E}[f(X_T^{\Delta})]| \geqslant r] \leqslant 2\exp(-\frac{r^2 M}{T\Psi(T, f, b, \sigma, q)}),$$
$$\Psi(T, f, b, \sigma, q) := 4\beta[f]_1^2|\sigma|_\infty^2 \exp\left(2([b]_1 + c[\sigma]_1(1 \vee c[\sigma]_1))T\right),$$

where q is the dimension of the underlying Brownian motion W and $c := c(q)$.

Note that contrary to Lemaire and Menozzi [91] we do not need any non-degeneracy condition on the diffusion coefficient and we got rid off the systematic bias.

The strategy to prove the above concentration inequality is similar to the first step of the analysis of the discretisation error of Talay and Tubaro [122] presented in Section 1.2.3. Indeed, introducing the functions $v^{\Delta}(t_i, x) := \mathbb{E}[f(X_T^{\Delta})|X_{t_i}^{\Delta} = x]$, $0 \leqslant i \leqslant N$, $x \in \mathbb{R}^d$, we use the Markov property satisfied by the scheme and the following telescopic sum

$$\begin{aligned} f(X_T^{\Delta}) - \mathbb{E}[f(X_T^{\Delta})] &= \sum_{i=1}^N v^{\Delta}(t_i, X_{t_i}^{\Delta}) - \mathbb{E}[v^{\Delta}(t_i, X_{t_i}^{\Delta})|X_{t_{i-1}}^{\Delta}] \\ &= \sum_{i=1}^N f_i^{\Delta}(X_{t_{i-1}}^{\Delta}, \sqrt{\Delta}U^{(i)}) - \mathbb{E}[f_i^{\Delta}(X_{t_{i-1}}^{\Delta}, \sqrt{\Delta}U^{(i)})|X_{t_{i-1}}^{\Delta}] \end{aligned} \tag{3.3}$$

where $f_i^{\Delta}(x, u) := \mathbb{E}[f(X_T^{\Delta})|X_{t_i}^{\Delta} = x + b(t_{i-1}, x)\Delta + \sigma(t_{i-1}, x)u]$, $1 \leqslant i \leqslant N$, $(x, u) \in \mathbb{R}^d \times \mathbb{R}^q$. In Talay and Tubaro [122], $v^{\Delta}(t_i, X_{t_i}^{\Delta})$ is replaced by $v(t_i, X_{t_i}^{\Delta}) = \mathbb{E}[f(X_T)|X_{t_i} = x]_{|x=X_{t_i}^{\Delta}}$, that is the expectation involving the diffusion at time T starting from the current value of the scheme at t_i. Direct computations based on B-D-G inequality and Gronwall's Lemma show that $u \mapsto f_i^{\Delta}(x, u)$ is Lipschitz-continuous with $[f_i^{\Delta}(x, .)]_1 \leqslant (\Psi(T, f, b, \sigma, q))^{1/2}$, $x \in \mathbb{R}^d$.

Since $U^{(i)}$ satisfies $GC(\beta)$, from Proposition 3.1.1 it follows that

$$\mathbb{E}[\exp(\lambda(f_i^{\Delta}(X_{t_{i-1}}^{\Delta}, \sqrt{\Delta}U^{(i)}) - \mathbb{E}[f_i^{\Delta}(X_{t_{i-1}}^{\Delta}, \sqrt{\Delta}U^{(i)})|X_{t_{i-1}}^{\Delta}]))|X_{t_{i-1}}^{\Delta}] \leqslant \exp(\beta([f_i^{\Delta}(X_{t_{i-1}}^{\Delta}, .)]_1\sqrt{\Delta}\lambda)^2/4), \quad i = 1, \cdots, N. \tag{3.4}$$

Now using the Markov exponential inequality, the decomposition (3.3) and the tower property of condition expectation one proves Theorem 3.2.1 for $M = 1$. The bound for all positive integer M follows from a simple tensorization argument for independent random variables satisfying the Gaussian concentration property.

Besides our consideration for the Euler scheme, we also consider a Robbins-Monro algorithm with dynamics

$$\theta_{n+1} = \theta_n - \gamma_{n+1}H(\theta_n, U^{(n+1)}),\ n \geqslant 0,\ \theta_0 \in \mathbb{R}^d, \tag{3.5}$$

where $(U^{(n)})_{n\geqslant 1}$ is an i.i.d. $\mathbb{R}^q$-valued sequence of random variables with law μ satisfying $GC(\beta)$ for some $\beta > 0$ and $(\gamma_n)_{n\geqslant 1}$ is a sequence of non-negative deterministic steps satisfying the usual assumption (1.9). We assume that the following conditions on the function H and the step sequence $(\gamma_n)_{n\geqslant 1}$ in (4.1) are in force:

(HL) The map $(\theta, u) \in \mathbb{R}^d \times \mathbb{R}^q \mapsto H(\theta, u)$ is uniformly Lipschitz continuous.

(HUA) The map $h : \theta \in \mathbb{R}^d \mapsto \mathbb{E}[H(\theta, U)]$ is continuously differentiable in θ and there exists $\underline{\lambda} > 0$ s.t. $\forall \theta \in \mathbb{R}^d$, $\forall \xi \in \mathbb{R}^d$, $\underline{\lambda}|\xi|^2 \leqslant \langle Dh(\theta)\xi, \xi\rangle$ (*Uniform Attractivity*).

As we have seen in Section 1.1.3 of the introductory chapter, in order to derive a Central Limit Theorem for the sequence $(\theta_n)_{n\geqslant 1}$ as described in Duflo [39] or Kushner and Yin [81], it is commonly assumed that the matrix $Dh(\theta^*)$ is *uniformly attractive*. In our current framework, this local condition on the Jacobian matrix of h at the equilibrium is replaced by the uniform assumption **(HUA)**. In the framework of stochastic gradient algorithm, this assumption is also known as *uniform convexity*.

Here the global error between the stochastic approximation procedure θ_n at a given time step n and its target θ^* can be decomposed as follows:

$$\begin{aligned} z_n := |\theta_n - \theta^*| &= (|\theta_n - \theta^*| - \mathbb{E}[|\theta_n - \theta^*|]) + \mathbb{E}[|\theta_n - \theta^*|] \\ &:= \mathcal{E}_{Emp}(\gamma, n, H, \underline{\lambda}, \alpha) + \delta_n \end{aligned} \tag{3.6}$$

where $\delta_n := \mathbb{E}[|\theta_n - \theta^*|]$.

The term $\mathcal{E}_{Emp}(\gamma, n, H, \underline{\lambda}, \alpha)$ corresponding to the difference between the absolute value of the error at time n and its mean can be viewed as an *empirical error*. The term δ_n is an *intrinsic bias* coming from the *non-linear nature* of the algorithm. Note that from Proposition 1.14 in Section 1.1.3 under **(HL)** and **(HUA)** one has

$$\delta_N \leqslant \mathbb{E}[|\theta_N - \theta^*|^2]^{\frac{1}{2}} \leqslant C(\gamma_N)^{\frac{1}{2}}, \quad N \geqslant 1$$

for the two following step sequences: $\gamma_n = c/n^\beta$, $\beta \in (1/2, 1)$, and $\gamma_n = c/n$, with $c > 1/(2\underline{\lambda}$, $n \geqslant 1$.

An explicit expression for the constant C is provided in **[7]**. As for the Euler scheme, we prove that the Gaussian concentration property transfers to the empirical error under **(HL)** and **(HUA)**.

Theorem 3.2.2 (Concentration Bounds for Robbins-Monro algorithms)**.** *Assume that the function H of the recursive procedure $(\theta_n)_{n\geqslant 0}$ (with starting point $\theta_0 \in \mathbb{R}^d$) defined by* (4.1) *satisfies* **(HL)** *and* **(HUA)**, *and that the step sequence $(\gamma_n)_{n\geqslant 1}$ satisfies the usual step assumption* (1.9)*. Suppose that the law of the innovation satisfies $GC(\beta)$, $\beta > 0$. Then, for all $N \in \mathbb{N}^*$ and all $r \geqslant 0$,*

$$\mathbb{P}\left(|\theta_N - \theta^*| \geqslant r + \delta_N\right) \leqslant \exp\left(-\frac{r^2}{\beta[H]_1^2 \Pi_N \sum_{k=1}^N \gamma_k^2/\Pi_k}\right)$$

where $\Pi_N := \prod_{k=0}^{N-1}\left(1 - 2\underline{\lambda}\gamma_{k+1} + [H]_1^2\gamma_{k+1}^2\right)$.

Concerning the choice of the step sequence $(\gamma_n)_{n\geqslant 1}$ and its impact on the concentration rate and bias, we obtain the following results:

- If we choose $\gamma_n = \frac{c}{n}$, with $c > 0$. Then, $\Gamma_N = \sum_{k=1}^N \gamma_k = c\log(N) + c_1' + r_N$, $c_1' > 0$ and $r_N \to 0$, so that $\Pi_N = (N^{-2c\underline{\lambda}})$.
  - If $c < \frac{1}{2\underline{\lambda}}$, we obtain $\Pi_N \sum_{k=1}^N \gamma_k^2/\Pi_k = \mathcal{O}(N^{-2c\underline{\lambda}})$.
  - If $c > \frac{1}{2\underline{\lambda}}$, a comparison between the series and the integral yields $\Pi_N \sum_{k=1}^N \gamma_k^2/\Pi_k = (N^{-1})$.
- If we choose $\gamma_n = \frac{c}{n^\rho}$, $c > 0$, $\frac{1}{2} < \rho < 1$, then one has $\Pi_N \sum_{k=1}^N \gamma_k^2 \Pi_k^{-1} = o(N^{-\rho+\epsilon})$, $\epsilon \in (0, 1-\rho)$.

As for the Euler like scheme, the strategy consists in introducing again a telescopic sum of conditional expectations. Denoting for all $i \in \mathbb{N}$, $\mathcal{F}_i := \sigma(U^{(j)},\ j \leqslant i)$, i.e. $(\mathcal{F}_i)_{i\in\mathbb{N}}$ is the natural filtration of the algorithm, we write for all $n \in \mathbb{N}^*$:

$$\begin{aligned} \mathcal{E}_{Emp}(\gamma, N, H, \underline{\lambda}, \alpha) = |z_N| - \mathbb{E}[|z_N|] &= \sum_{i=1}^N \mathbb{E}[|z_N||\mathcal{F}_i] - \mathbb{E}[|z_N||\mathcal{F}_{i-1}] \\ &= \sum_{i=1}^N v_i^\gamma(\theta_i) - \mathbb{E}[v_i^\gamma(\theta_i)|\mathcal{F}_{i-1}], \\ &= \sum_{i=1}^N f_i^\gamma(\theta_{i-1}, U^{(i)}) - \mathbb{E}[f_i^\gamma(\theta_{i-1}, U^{(i)})|\mathcal{F}_{i-1}], \end{aligned}$$

where we used the Markov property for the second equality and we introduced the notations $v_i^\gamma(\theta) := \mathbb{E}[|\theta_n - \theta^*||\theta_i = \theta]$, $1 \leqslant i \leqslant n$, $\theta \in \mathbb{R}^d$, $f_i^\gamma(\theta, u) = v_i^\gamma(\theta - \gamma_i H(\theta, u))$. The stability of the Gaussian concentration property is then derived using that the f_i^γ are Lipschitz with respect to the variable u and similar arguments to those employed for the Euler like scheme.

3.3 From Exponential to Gaussian concentration regimes

In **[7]**, we obtained Gaussian concentration inequalities for the law of an Euler like scheme of a diffusion process and for the deviation of a Robbins-Monro algorithm from its target. It should be noted that it is the boundedness of the diffusion coefficient σ and the uniform Lipschitz continuity of the function $u \mapsto H(\theta, u)$, $\theta \in \mathbb{R}^d$, that give rise to the Gaussian concentration regime for the deviation of the empirical error. In this section, we present the results obtained **[8]** in collaboration with M. Fathi from the university Paul Sabatier that extend and complete those exposed in the previous section. The key point is to quantify to properly quantify the contribution of the diffusion term to the concentration regime. We also obtain non-asymptotic bound for stochastic approximation with averaging of trajectories, in particular we prove that averaging a stochastic approximation algorithm with a slow decreasing step sequence gives rise to optimal concentration rate.

We consider the Euler-like scheme $(X^{\Delta}_{t_i})_{0\leqslant i\leqslant N}$ with dynamics (3.2). The innovation sequence $(U^{(i)})_{1\leqslant i\leqslant N}$ satisfies $GC(\beta)$ for some $\beta > 0$. In **[8]**, we replace assumption **(A)** by the following smoothness and domination assumptions

(HS) The coefficients b, σ are uniformly Lipschitz continuous in space uniformly in time.

($\mathbf{HD}_\alpha$) There exists a $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^*_+)$ function V satisfying $\exists C_V > 0, |\nabla V|^2 \leqslant C_V V$, $\eta := \frac{1}{2}\sup_{x\in\mathbb{R}^d} \left\|\nabla^2 V(x)\right\| < +\infty$ and $\exists \alpha \in (0, 1]$, such that for all $x \in \mathbb{R}^d$,

$$\exists C_b > 0, \quad \sup_{t\in[0,T]} |b(t,x)|^2 \leqslant C_b V(x), \quad , \ \exists C_\sigma > 0, \quad \sup_{t\in[0,T]} Tr(a(t,x)) \leqslant C_\sigma V^{1-\alpha}(x).$$

where $a = \sigma\sigma^*$.

The idea behind assumption **($\mathbf{HD}_\alpha$)** is to parameterize the growth of the diffusion coefficient in order to quantify its contribution to the concentration regime. Indeed, under **(HS)** and **($\mathbf{HD}_\alpha$)**, with $\alpha \in [1/2, 1]$ we derive non-asymptotic deviation bounds for the statistical error ranging from exponential (if $\alpha = 1/2$) to Gaussian (if $\alpha = 1$) regimes.

Theorem 3.3.1 (Transport-Entropy inequalities for Euler like schemes)**.** *Denote by X^{Δ}_T the value at time T of the scheme (3.2) and by μ^{Δ}_T its law. Denote the Lipschitz modulus of b and σ by $[b]_1$ and $[\sigma]_1$ respectively. Assume that the common law of the innovation sequence $(U^{(i)})_{1\leqslant i\leqslant N}$ in (3.2) satisfies $GC(\beta)$ for some $\beta > 0$ and that the coefficients b, σ satisfy* **(HS)** *and* **($\mathbf{HD}_\alpha$)** *for some $\alpha \in [\frac{1}{2}, 1]$.*

Then, μ^{Δ}_T satisfies $T_{\Phi^*_\alpha}$ with $\Phi^*_\alpha(\lambda) = \sup_{\rho\geqslant 0}\{\lambda\rho - \Phi_\alpha(\rho)\}$ and one has:

- *If $\alpha \in (\frac{1}{2}, 1]$, for all $\rho \geqslant 0$*
$$\Phi_\alpha(\rho) = \Psi_\alpha(T, \Delta, b, \sigma, x)(\rho^2 \vee \rho^{\frac{2\alpha}{2\alpha-1}}),$$
- *If $\alpha = \frac{1}{2}$, for all $\rho \in [0, \varphi(T, b, \sigma, \Delta)^{-1/2}\lambda_{3.2})$*
$$\Phi_{1/2}(\rho) = K_{3.2}\frac{(\rho\varphi(T, b, \sigma, \Delta)^{1/2}/\lambda_{3.2})^2}{1 - (\rho\varphi(T, b, \sigma, \Delta)^{1/2}/\lambda_{3.2})}.$$

Moreover, we have $\Psi_\alpha(T, \Delta, b, \sigma, x) = K_{3.1}(\varphi(T, b, \sigma, \Delta)^2 \vee \varphi(T, b, \sigma, \Delta)^{\frac{\alpha}{2\alpha-1}})$, $\varphi(T, b, \sigma, \Delta) = C_\sigma\beta\frac{(1+C(\Delta)\Delta)}{4C(\Delta)}e^{3C(\Delta)T}$, $C(\Delta) := 2[b]_1 + [\sigma]^2_1 + \Delta[b]^2_1$ where $K_{3.1}$, $\lambda_{3.2}$ and $K_{3.2}$ are explicit constants.

- In the case $\alpha \in (\frac{1}{2}, 1]$, one easily gets the following explicit formula:
  - If $\lambda \in [0, 2\Psi]$, then $\Phi^*_\alpha(\lambda) = \frac{1}{4\Psi}\lambda^2$;
  - If $\lambda \in [\frac{2\alpha}{2\alpha-1}\Psi, +\infty)$, then $\Phi^*_\alpha(\lambda) = \frac{1}{2\alpha}\left(\frac{2\alpha-1}{2\alpha\Psi}\right)^{2\alpha-1}\lambda^{2\alpha}$;
  - If $\lambda \in (2\Psi, \frac{2\alpha}{2\alpha-1}\Psi)$,then $\Phi^*_\alpha(\lambda) = \lambda - \Psi$.

  Let us note that the linear behavior of Φ^*_α on a small interval is due to the fact that Φ_α is not $\mathcal{C}^1$.
- In the case $\alpha = 1/2$, tedious but simple computations show that
$$\Phi^*_{1/2}(\lambda) = \left(\left(1 + \frac{\lambda_{3.2}}{K_{3.2}\varphi(T, b, \sigma, \Delta)^{1/2}}\lambda\right)^{\frac{1}{2}} - 1\right)^2.$$

We also consider the Robbins-Monro algorithm with dynamics (4.1) under the assumption that the innovation sequence $(U^{(i)})_{1\leqslant i\leqslant N}$ satisfies $GC(\beta)$ for some $\beta > 0$. Instead of **(HL)** we will make the following assumptions:

(HL) For all $u \in \mathbb{R}^q$, the function $H(.,u)$ is Lipschitz-continuous with a Lipschitz modulus having linear growth in the variable u, that is:

$$\exists C_H > 0,\ \forall u \in \mathbb{R}^q,\quad \sup_{(\theta,\theta')\in(\mathbb{R}^d)^2} \frac{|H(\theta,u)-H(\theta',u)|}{|\theta-\theta'|} \leqslant C_H(1+|u|).$$

(HLS)$_\alpha$ (*Lyapunov Stability-Domination*) There exists a $\mathcal{C}^2(\mathbb{R}^d,\mathbb{R}_+^*)$ function L satisfying $\exists C_L > 0, |\nabla L|^2 \leqslant C_L L$, $\eta := \frac{1}{2}\sup_{x\in\mathbb{R}^d}\left\|\nabla^2 L(x)\right\| < +\infty$ such that

$$\forall \theta \in \mathbb{R}^d,\ \ \langle \nabla L(\theta), h(\theta)\rangle \geqslant 0, \quad \text{and} \quad \exists C_h > 0,\ \forall \theta\in\mathbb{R}^d,\ |h(\theta)|^2 \leqslant C_h L(\theta).$$

and $\exists \alpha \in (0,1]$,

$$\exists C_\alpha > 0,\ \forall \theta\in\mathbb{R}^d,\quad \sup_{(u,u')\in(\mathbb{R}^q)^2} \frac{|H(\theta,u)-H(\theta,u')|}{|u-u'|} \leqslant C_\alpha L^{\frac{1-\alpha}{2}}(\theta).$$

Assumption **(HLS)**$_\alpha$ is similar to assumption **(HD)**$_\alpha$ for the Euler like scheme and allows to quantify the contribution of the diffusion term to the concentration regime of the empirical error $\mathcal{E}_{Emp}(\gamma,N,H,\underline{\lambda},\alpha)$ of the Robbins-Monro scheme.

Theorem 3.3.2 (Transport-Entropy inequalities for stochastic approximation algorithms)**.** *Let $N \in \mathbb{N}^*$. Assume that the function H of the recursive procedure $(\theta_n)_{0\leqslant n\leqslant N}$ (with starting point $\theta_0 \in \mathbb{R}^d$) defined by* (4.1) *satisfies* **(HL)**, **(HUA)** *and* **(HLS)**$_\alpha$ *for $\alpha \in [\frac{1}{2},1]$, and that the step sequence $\gamma = (\gamma_n)_{n\geqslant 0}$ satisfies* (1.9)*. Suppose that the law of the innovation satisfies $GC(\beta)$, $\beta > 0$. Denote by μ_N^γ the law of θ_N.*

Then, μ_N^γ satisfies $T_{\Phi_\alpha^*}$ with $\Phi_{\alpha,N}^*(\lambda) = \sup_{\rho\geqslant 0}\{\lambda\rho - \Phi_{\alpha,N}(\rho)\}$ and one has:

- *If $\alpha \in (\frac{1}{2},1]$, for all $\rho \geqslant 0$*
$$\Phi_{\alpha,N}(\rho) = \varphi_\alpha(\gamma,H,\theta_0)(C_N^\gamma\rho^2 \vee C_N^{\gamma,\alpha}\rho^{\frac{2\alpha}{2\alpha-1}}).$$
- *If $\alpha = \frac{1}{2}$, for all $\rho \in [0,\lambda_{4.1}/\tilde{s}_N)$,*
$$\Phi_{1/2,N}(\rho) = 2\varphi_{1/2}(\gamma,H,\theta_0)C_N^\gamma \frac{(\rho/\lambda_{4.1})^2}{1-(\rho\tilde{s}_N/\lambda_{4.1})}.$$

Moreover the three concentration rate sequences are defined for $N \in \mathbb{N}^*$ by

$$C_N^\gamma := \sum_{k=0}^{N-1}\gamma_{k+1}^2\frac{\Pi_{1,N}}{\Pi_{1,k}},$$
$$C_N^{\gamma,\alpha} := \sum_{k=0}^{N-1}\gamma_{k+1}^{\frac{2\alpha}{2\alpha-1}}\left(\frac{\Pi_{1,N}}{\Pi_{1,k}}\right)^{\frac{2\alpha}{2\alpha-1}}((k+1)\log^2(k+4))^{\frac{1-\alpha}{2\alpha-1}}$$
$$\tilde{s}_N := \max_{0\leqslant k\leqslant N-1}(k+1)^{1/2}\log(k+4)\gamma_{k+1}\left(\frac{\Pi_{1,N}}{\Pi_{1,k}}\right)^{\frac{1}{2}}\exp\left(\sum_{p=0}^{N-1}\frac{1}{(p+1)\log^2(p+4)}\right)$$

with $\Pi_{1,N} := \prod_{k=0}^{N-1}(1-2\underline{\lambda}\gamma_{k+1}+C_{H,\mu}\gamma_{k+1}^2)$ and where $C_{H,\mu}$ and $\varphi := \varphi_\alpha(\gamma,H,\theta_0)$ are explicitly constants.

As in the case of Euler like schemes, for $\alpha \in (\frac{1}{2},1]$, we have:

- if $\lambda \in [0, 2\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}]$, then $\Phi_{\alpha,N}^*(\lambda) = \lambda^2/(4\varphi C_N^\gamma)$;
- If $\lambda \in [\frac{2\alpha}{2\alpha-1}\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}, +\infty)$, then $\Phi_{\alpha,N}^*(\lambda) = \frac{1}{2\alpha}\left(\frac{2\alpha-1}{2\alpha\varphi}\right)^{2\alpha-1}(\lambda^{2\alpha}/(C_N^{\gamma,\alpha})^{2\alpha-1})$;
- If $\lambda \in (2\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}, \frac{2\alpha}{2\alpha-1}\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}})$, then $\Phi_{\alpha,N}^*(\lambda) = (\frac{C_N^\gamma}{C_N^{\gamma,\alpha}})^{\frac{2\alpha-1}{2(1-\alpha)}}\lambda - \varphi\frac{(C_N^\gamma)^{\frac{\alpha}{1-\alpha}}}{(C_N^{\gamma,\alpha})^{\frac{2\alpha-1}{1-\alpha}}}$.

For $\alpha = \frac{1}{2}$, we obtain the following explicit bound for the Legendre transform of $\Phi_{1/2,N}$

$$\forall \lambda \geqslant 0,\quad \Phi_{1/2,N}^*(\lambda) = \frac{2\varphi C_N^\gamma}{\tilde{s}_N^2}\left(\left(1+\frac{\tilde{s}_N\lambda_{4.1}\lambda}{2\varphi C_N^\gamma}\right)^{\frac{1}{2}}-1\right)^2$$

Hence, for $N \geqslant 1$ being fixed, the following simple asymptotic behaviours can be easily derived:

- When λ is small, $\Phi^*_{1/2,N}(\lambda) \sim \lambda^2_{4.1}\lambda^2/(2\varphi C_N^\gamma)$;
- When λ goes to infinity, $\Phi^*_{1/2}(\lambda) \sim \lambda_{4.1}\lambda/\tilde{s}_N$.

Concerning the choice of the step sequence $(\gamma_n)_{n\geqslant 1}$ and its impact on the concentration rate and bias, we obtain the following results:

- If we choose $\gamma_n = \frac{c}{n}$, with $c > 0$.
  - If $c < \frac{1}{2\underline{\lambda}}$, one has $C_N^\gamma = \mathcal{O}(N^{-2c\underline{\lambda}})$, $C_N^{\gamma,\alpha} = \mathcal{O}(N^{-\frac{2\alpha}{2\alpha-1}c\underline{\lambda}})$, $\tilde{s}_N = (N^{-c\underline{\lambda}})$.
  - If $c > \frac{1}{2\underline{\lambda}}$, one has $C_N^\gamma = \mathcal{O}(N^{-1})$, $C_N^{\gamma,\alpha} = \mathcal{O}((\log(N))^{2\frac{1-\alpha}{2\alpha-1}}N^{-\frac{\alpha}{2\alpha-1}})$, $\tilde{s}_N = \mathcal{O}(\log(N)N^{-\frac{1}{2}})$.
- If we choose $\gamma_n = \frac{c}{n^\rho}$, $c > 0$, $\frac{1}{2} < \rho < 1$, then one has $C_N^\gamma = o(N^{-\rho+\epsilon})$, $\epsilon \in (0, 1-\rho)$, $C_N^{\gamma,\alpha} = o(N^{-\frac{(\rho-(1-\alpha))}{2\alpha-1}-\epsilon})$ and $\tilde{s}_N = \mathcal{O}\left(\log(N)N^{-(\rho-\frac{1}{2})}\right)$.

We see that the optimal convergence rate $\mathcal{O}(N^{-1})$ and $\mathcal{O}((\log(N))^{2\frac{1-\alpha}{2\alpha-1}}N^{-\frac{\alpha}{2\alpha-1}})$ is obtained by choosing $\gamma_n = c/n$ for $c > 1/(2\underline{\lambda})$. Let us notice that we find the same critical level for the constant c as in the CLT for stochastic algorithms, see Theorem 1.1.5. Indeed, if $c > \frac{1}{2\mathcal{R}e(\lambda_{min})}$ where λ_{min} denotes the eigenvalue of $Dh(\theta^*)$ with the smallest real part then we know that the optimal rate $\sqrt{N}$ is achieved in the CLT.

In order to get rid of this constraint, we also proved in **[8]** a transport entropy inequality for the law $\bar{\mu}_N^\gamma$ of the empirical mean $(\bar{\theta}_n)_{0\leqslant n\leqslant N}$ of the sequence $(\theta_n)_{0\leqslant n\leqslant N}$ defined by (4.1) when the step sequence is given by $\gamma_n = c/n^\rho$, $\frac{1}{2} < \rho < 1$. In particular, we prove that averaging the original Robbins-Monro dynamics with this slow decreasing step allows to achieve for free the optimal concentration rates, that is, the one obtained with $\gamma_n = c/n$ with $c > 1/(2\underline{\lambda})$.

We briefly explain the strategy used in **[8]** to prove this new transport entropy. The first step is similar to the one presented in the previous section. We consider the telescopic sum (3.3). However, under **($\mathbf{HD}_\alpha$)** the bound (3.4) becomes

$$\mathbb{E}[\exp(\lambda(f_i^\Delta(X^\Delta_{t_{i-1}}, \sqrt{\Delta}U^{(i)}) - \mathbb{E}[f_i^\Delta(X^\Delta_{t_{i-1}}, \sqrt{\Delta}U^{(i)})|X^\Delta_{t_{i-1}}]))|X^\Delta_{t_{i-1}}] \leqslant \exp(\beta\Psi_\alpha(T,\Delta,b,\sigma,x)\Delta\frac{\lambda^2}{4}V^{1-\alpha}(X^\Delta_{t_{i-1}})),\ i = 1,\cdots,N. \tag{3.7}$$

Let $M_N = \sum_{i=1}^N v^\Delta(t_i, X^\Delta_{t_i}) - \mathbb{E}[v^\Delta(t_i, X^\Delta_{t_i})|X^\Delta_{t_{i-1}}]$, $N \geqslant 1$ and $M_0 = 0$. From (3.7) and Hölder inequality, one gets

$$\begin{aligned}
\mathbb{E}[\exp(\lambda M_N)] &\leqslant \mathbb{E}[\exp(\lambda M_{N-1})\exp(\beta\Psi_\alpha(T,\Delta,b,\sigma,x)\Delta\frac{\lambda^2}{4}V^{1-\alpha}(X^\Delta_{t_{N-1}}))] \\
&\leqslant \mathbb{E}[\exp(\lambda p M_{N-1})]^{\frac{1}{p}}\mathbb{E}[\exp(q\beta\Psi_\alpha(T,\Delta,b,\sigma,x)\Delta\frac{\lambda^2}{4}V^{1-\alpha}(X^\Delta_{t_{N-1}}))]^{\frac{1}{q}} \\
&\leqslant \prod_{k=0}^{N-1}\left(\mathbb{E}\left[\exp\left(\frac{\Psi_\alpha(T,\Delta,b,\sigma,x)\beta}{4}\lambda^2 qp^{2k}\Delta V^{1-\alpha}(X^\Delta_{t_{N-k-1}})\right)\right]^{\frac{1}{q}}\right)^{\frac{1}{p^k}}
\end{aligned}$$

where we used an induction argument for the last inequality. Hence, since (p, q) are conjugate, one gets

$$\mathbb{E}\left[\exp(\lambda(f(X_T^\Delta) - \mathbb{E}\left[f(X_T^\Delta)\right])\right] \leqslant \exp\left(\sup_{0\leqslant n\leqslant N}\log\left(\mathbb{E}\left[e^{\frac{C_{\sigma\beta}}{4}\lambda^2\Delta qp^{2N}(1+C(\Delta)\Delta)^N V^{1-\alpha}(X^\Delta_{t_n})}\right]\right)\right).$$

Finally, by stability arguments on the dynamics (3.2) with an innovation sequence satisfying $GC(\beta)$, $\beta > 0$ one gets

$$\sup_{0\leqslant n\leqslant N}\log\left(\mathbb{E}\left[\exp(\lambda^2 V^{1-\alpha}(X^\Delta_{t_n}))\right]\right) \leqslant \begin{cases} K_{3.1}(\lambda \vee \lambda^{\frac{\alpha}{2\alpha-1}}) & \alpha \in (\frac{1}{2}, 1], \\ K_{3.2}\frac{(\lambda/\lambda_{3.2})^2}{1-(\lambda/\lambda_{3.2})}, & \alpha = \frac{1}{2}. \end{cases}$$

A suitable choice of p and q allows to conclude. The strategy of proof for the Robbins-Monro algorithm follows similar lines of arguments.

3.4 Perspective

An interesting problem is to investigate the case of non-regular test function f. In practical situations, notably in mathematical finance, one often has to deal with an irregular payoff function f such as indicator functions. One possibility to address this problem is to allow σ to be non-degenerated so that the Euler scheme admits a smooth

density. Then, one has to investigate the connection between the lack of regularity of f, the regularity of the density of the scheme and the resulting Lipschitz modulus of the function $y \mapsto f_i^{\Delta}(x, y)$.

Another interesting problem is to investigate concentration inequalities for the multi-level Monte Carlo estimator presented in Section 1.2.4. One expects that if the test function f is Lipschitz-continuous then the Gaussian concentration regime still holds for this estimator but with a convergence rate given by the CLT established in [1]. The same problem may be considered for the multi-level stochastic approximation schemes presented in the next chapter.

Chapter 4

Improving the computational efficiency of stochastic approximation algorithms

In this chapter, we present **[9]** and **[10]**, the latter being a joint work with L. Huang from the engineering school INSA of Toulouse. Both works aim to improve the computational efficiency for stochastic approximation schemes when one does not know how to directly sample from the exact distribution of the underlying innovation sequence. We extend the scope of the multi-level Monte Carlo and the multi-step Richardson Romberg methods briefly presented in Section 1.2.4 and Section 1.2.3.

The paradigm for both methods is described in Section 4.1. In Section 4.2 we present the results obtained in **[9]** related to multi-level methods for stochastic optimisation by means of stochastic approximation schemes. Contrary to the $L^2(\mathbb{P})$ analysis initiated by Giles [48], our methodology is based on establishing CLT for the proposed multi-level stochastic approximation estimators as proposed by Ben Alaya and Kebaier [1] in the standard Monte Carlo framework. In Section 4.3 we present the results obtained in **[10]** where we show that the principle of the multi-step Richardson-Romberg extrapolation method for Monte Carlo linear estimator as investigated in Pagès [109] can be extended to the framework of stochastic optimisation by means of stochastic approximation algorithm. For both methods we will see in particular that the new estimator outperforms the standard stochastic approximation estimator in terms of computational cost. Numerical results are presented in order to confirm the analysis.

4.1 Motivation

As was already mentioned in Section 1.1 of the introductory chapter, in many fields of applied mathematics such as computational finance, one often has to face some stochastic optimisation or zero search problems. We have seen some of them in Chapter 2. The stochastic approximation theory provides some recursive schemes in order find the zero of the function $h : \mathbb{R}^d \to \mathbb{R}^d$ defined by $h(\theta) := \mathbb{E}[H(\theta, U)]$, $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$, based on *i.i.d.* samples of the $\mathbb{R}^q$-valued random vector U.

Now assume, as it is the case of the Monte Carlo method, that the random vector U is not directly simulatable (at a reasonable cost) but can only be approximated by another sequence of easily simulatable random vectors $((U^n)^p)_{p\geqslant 1}$, which strongly approximates U as $n \to +\infty$ with a standard weak discretization error (or bias) $\mathbb{E}[f(U)] - \mathbb{E}[f(U^n)]$ of order $n^{-\alpha}$ for a specific class of functions. The computational cost required to simulate one sample of U^n is assumed to be of order n that is $Cost(U^n) = K \times n$ for some positive constant K. One standard situation corresponds to the case of a discretization of an SDE by means of an Euler-Maruyama scheme with n time steps.

The important point here is that the function h is generally neither known nor computable (at least at reasonable cost) and since the random variable U cannot be simulated, estimating θ^* using a Robbins-Monro algorithm is not possible. Therefore, two steps are needed to compute θ^*:

- the first step consists in approximating the zero θ^* of h by the zero $\theta^{*,n}$ of h^n defined by $h^n(\theta) := \mathbb{E}[H(\theta, U^n)]$, $\theta \in \mathbb{R}^d$. It induces *an implicit weak error* which writes

$$\mathcal{E}_D(n) := \theta^* - \theta^{*,n}.$$

Let us note that $\theta^{*,n}$ appears as a proxy of θ^* and one would naturally expect that $\theta^{*,n} \to \theta^*$ as the bias parameter n tends to infinity.

- the second step consists in approximating $\theta^{*,n}$ by $M \in \mathbb{N}^*$ steps of the following SA scheme

$$\theta^n_{p+1} = \theta^n_p - \gamma_{p+1} H(\theta^n_p, U^{n,(p+1)}),\ 0 \leqslant p \leqslant M-1, \tag{4.1}$$

where $(U^{n,(p)})_{1\leqslant p\leqslant M}$ is an i.i.d. sequence of random variables with the same law as U^n, θ_0^n is independent of the innovation of the algorithm with $\sup_{n\geqslant 1}\mathbb{E}[|\theta_0^n|^2] < +\infty$ and $\gamma = (\gamma_p)_{p\geqslant 1}$ is a sequence of non-negative deterministic and decreasing steps satisfying the usual step assumption (1.9). This induces a *statistical error* which writes

$$\mathcal{E}_S(n, M, \gamma) := \theta^{*,n} - \theta_M^n.$$

The global error between θ^*, the quantity to estimate, and its implementable approximation θ_M^n can be decomposed as follows:

$$\begin{aligned}\mathcal{E}_{glob}(n, M, \gamma) &= \theta^* - \theta^{*,n} + \theta^{*,n} - \theta_M^n \\ &:= \mathcal{E}_D(n) + \mathcal{E}_S(n, M, \gamma).\end{aligned}$$

Concerning the asymptotic behaviour of the implicit weak error $\mathcal{E}_D(n)$, the following result is established in **[9]**.

Proposition 4.1.1 (Implicit discretisation error)**.** *For all $n \in \mathbb{N}^*$, assume that h and h^n satisfy the mean reverting assumption* (1.11) *of Corollary 1.1.2. Moreover, suppose that $(h^n)_{n\geqslant 1}$ converges locally uniformly towards h. Then, one has*

$$\theta^{*,n} \to \theta^* \quad as \quad n \to +\infty.$$

Moreover, suppose that h and h^n, $n \geqslant 1$, are continuously differentiable and that $Dh(\theta^*)$ is non-singular. Assume that $(Dh^n)_{n\geqslant 1}$ converges locally uniformly to Dh. If there exists $\alpha \in \mathbb{R}^*$ such that

$$\forall \theta \in \mathbb{R}^d, \ \lim_{n\to+\infty} n^\alpha (h^n(\theta) - h(\theta)) = \mathcal{E}(h, \alpha, \theta),$$

then, one has

$$\lim_{n\to+\infty} n^\alpha (\theta^{*,n} - \theta^*) = -Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*).$$

Hence we see that the standard weak rate of convergence transfers to the implicit weak error under mild assumptions. The above proposition is the first step toward an expansion of $\theta^{*,n} - \theta^*$ in powers of $n^{-\alpha}$. This will be discussed in further details in Section 4.3 for the development of a multi-step Richardson-Romberg method.

Regarding the *statistical error* $\mathcal{E}_S(n, M, \gamma) := \theta^{*,n} - \theta_M^n$, it is well-known that under standard assumptions, the Robbins-Monro theorem, guarantees that $\lim_M \mathcal{E}_S(n, M, \gamma) = 0$ *a.s.* for each fixed $n \in \mathbb{N}^*$, see Corollary 1.1.2. Moreover, under mild technical conditions, a CLT holds at rate $\gamma^{-1/2}(M)$, that is, for each fixed $n \in \mathbb{N}^*$, $\gamma^{-1/2}(M)\mathcal{E}_S(n, M, \gamma)$ converges in distribution to a normally distributed random variable with mean zero and finite covariance matrix, see Theorem 1.1.5. In particular if we set $\gamma(p) = \gamma_0/p$, $\gamma_0 > 0$, $p \geqslant 1$, the weak convergence rate is $\sqrt{M}$ provided that $2\mathcal{R}e(\lambda_{min})\gamma_0 > 1$ where λ_{min} denotes the eigenvalue of $Dh(\theta^*)$ with the smallest real part. Moreover, the Ruppert and Polyak averaging principle allows one to achieve the optimal rate of convergence for free by devising a standard Robbins-Monro algorithm with step sequence $\gamma(p) = \gamma_0/p^\rho$, $\frac{1}{2} < \rho < 1$ and computing the empirical mean of the trajectory $(\theta_p^n)_{1\leqslant p\leqslant M}$.

Given the order of the implicit weak error and a step sequence γ satisfying (1.9) a natural question is to find the optimal balance between the value of n and the number M of steps in the recursive scheme (4.1) in order to achieve a given global error. This problem was originally investigated by Duffie and Glynn [38] for the standard Monte Carlo method. The error between θ^* and the approximation θ_M^n writes $\theta_M^n - \theta^* = \theta_M^n - \theta^{*,n} + \theta^{*,n} - \theta^*$ suggesting to select $M = \gamma^{-1}(1/n^{2\alpha})$, where γ^{-1} is the inverse function of γ, when the *weak error* is of order $n^{-\alpha}$.

However, due to the non-linearity of the SA algorithm (4.1), the methodology developed in [38] does not apply in our context. The key ingredients to tackle this question consists in *linearising* the dynamic of $(\theta_p^n)_{1\leqslant p\leqslant M}$ around its target $\theta^{*,n}$, quantifying the contribution of the non linearities in the space variable θ_p^n and the innovations and finally exploiting stability arguments from SA schemes. The next result is established in **[9]**.

Theorem 4.1.2 (Optimal tradeoff)**.** *Suppose that the assumptions of Proposition 4.1.1 are satisfied and that h satisfies the assumptions of Theorem 1.1.5. Assume that h^n is twice continuously differentiable with Dh^n Lipschitz continuous uniformly in n and that it satisfies the strong mean-reverting assumption* (1.14) *with a constant $\underline{\lambda}$ (instead of α) that does not depend on n. Then, under some additional assumptions that we omit for sake of simplicity, one has*

$$n^\alpha \left(\theta^n_{\gamma^{-1}(1/n^{2\alpha})} - \theta^*\right) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*) + \mathcal{N}(0, \Sigma^*),$$

where

$$\Sigma^* := \int_0^\infty \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right)^T \mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T] \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right) ds \tag{4.2}$$

with $\zeta = 0$ if $\gamma(n) = \gamma_0/n^\rho$, $\frac{1}{2} < \rho < 1$ and $\zeta = 1/(2\gamma_0)$ if $\gamma(n) = \gamma_0/n$, $\gamma_0 > 1/(2\underline{\lambda})$.

Define the empirical mean sequence $(\bar{\theta}_p^n)_{p\geqslant 1}$ *of the sequence* $(\theta_p^n)_{p\geqslant 1}$ *by setting*

$$\bar{\theta}_p^n = \frac{\theta_0 + \theta_1^n + \cdots + \theta_p^n}{p+1} = \bar{\theta}_{p-1}^n - \frac{1}{p+1}\left(\bar{\theta}_{p-1}^n - \theta_p^n\right),$$

with a the step sequence $\gamma_p = \gamma_0/p^\rho$ *with* $\rho \in (1/2, 1)$. *Then, one has*

$$n^\alpha\left(\bar{\theta}_{n^{2\alpha}}^n - \theta^*\right) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h,\alpha,\theta^*) + \mathcal{N}\left(0, Dh(\theta^*)^{-1}\mathbb{E}[H(\theta^*,U)H(\theta^*,U)^T](Dh(\theta^*)^{-1})^T\right), \quad n\to\infty.$$

Optimising with respect to the usual choice of the step sequence, the minimal computational cost (to achieve an error of order $n^{-\alpha}$) is given by $C_{\mathrm{SA}} = K\times n\times\gamma^{-1}(1/n^{2\alpha})$ and is optimal for $\gamma(p) = \gamma_0/p$, $p\geqslant 1$, provided that $\gamma_0 > \frac{1}{2\alpha}$ or for the Ruppert and Polyak averaged algorithm, leading to a complexity of order $n^{2\alpha+1}$. In this case, we see that the computational cost is similar to the one achieved by the classical Monte Carlo algorithm for the computation of $\mathbb{E}[f(X_T)]$

In Section 4.2, we present a multi-level estimator which allows to significantly increase the efficiency the standard stochastic algorithm designed according to the optimal tradeoff.

4.2 Multi-level stochastic approximation algorithm

The multi-level stochastic approximation method uses the following decomposition $\theta^{*,n} = \theta^{*,1} + \sum_{\ell=1}^L(\theta^{*,m^\ell} - \theta^{*,m^{\ell-1}})$ with $m^L = n$, $m\geqslant 2$, and implements $L+1$ stochastic schemes with a sequence of bias parameter $(m^\ell)_{0\leqslant\ell\leqslant L}$. More precisely, the target θ^* is estimated by the quantity

$$\Theta_n^{ml} = \theta_{M_0}^1 + \sum_{\ell=1}^L\left(\theta_{M_\ell}^{m^\ell} - \theta_{M_\ell}^{m^{\ell-1}}\right).$$

It is important to point out here that for each level ℓ the couple $(\theta_{M_\ell}^{m^\ell}, \theta_{M_\ell}^{m^{\ell-1}})$ is computed using M_ℓ i.i.d. copies of $(U^{m^{\ell-1}}, U^{m^\ell})$. Moreover the random variables $U^{m^{\ell-1}}$ and U^{m^ℓ} use two different bias parameter but are perfectly correlated. Finally, for two different levels, the SA schemes are based on independent samples.

Theorem 4.2.1. *Suppose that* h *and* $(h^n)_{n\in\mathbb{N}}$ *satisfy the assumptions of Theorem 4.1.1. Assume that* h^n *is twice continuously differentiable in a neighbourhood of* θ^*, *with* Dh^n *Lipschitz-continuous uniformly in* n. *Assume that the following CLT holds:*

$$m^{\ell\rho}(U^{m^\ell} - U^{m^{\ell-1}}) \overset{stably}{\Longrightarrow} V^m \quad as \quad \ell\to+\infty$$

where V^m *is an* $\mathbb{R}^q$*-valued random variable eventually defined on an extension* $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$ *of* $(\Omega,\mathcal{F},\mathbb{P})$.

Suppose that $\tilde{\mathbb{E}}[(D_xH(\theta^*,U)V^m - \tilde{\mathbb{E}}[D_xH(\theta^*,U)V^m])(D_xH(\theta^*,U)V^m - \tilde{\mathbb{E}}[D_xH(\theta^*,U)V^m])^T]$ *is a positive definite matrix. Assume that the step sequence is given by* $\gamma_p = \gamma(p)$, $p\geqslant 1$, *where* γ *is a positive function defined on* $[0,+\infty[$ *decreasing to zero, satisfying one of the following assumptions:*

- γ *varies regularly with exponent* $(-a)$, $a\in(1/2,1)$, *that is, for any* $x>0$, $\lim_{t\to+\infty}\gamma(tx)/\gamma(t) = x^{-a}$.
- *for* $t\geqslant 1$, $\gamma(t) = \gamma_0/t$ *and* γ_0 *satisfies* $\underline{\lambda}\gamma_0 > 1$.

Suppose that ρ *satisfies one of the following assumptions:*

- *if* $\rho\in(0,1/2)$, *then assume that* $\alpha > 2\rho$, $\underline{\lambda}\gamma_0 > \alpha/(\alpha-2\rho)$ *(if* $\gamma(t) = \gamma_0/t$*) and*

$$\exists\beta>\rho,\quad \forall\theta\in\mathbb{R}^d,\ \sup_{n\geqslant 1} n^\beta\|Dh^n(\theta) - Dh(\theta)\| < +\infty.$$

  In this case we set:

$$M_0 = \gamma^{-1}(1/(n^{2\alpha}\log(n))),\quad M_l = \gamma^{-1}(m^{\ell\frac{(1+2\rho)}{2}}(m^{\frac{1-2\rho}{2}}-1)/(n^{2\alpha}(n^{\frac{(1-2\rho)}{2}}-1))),\ \ell = 1,\cdots,L.$$

- *if* $\rho = 1/2$, *then assume that* $\alpha = 1$, $\theta_0^{m^\ell} = \theta_0$, $\ell = 1,\cdots,L$, *with* $\mathbb{E}[|\theta_0|^2] < +\infty$ *and*

$$\exists\beta>1/2,\quad \forall\theta\in\mathbb{R}^d,\ \sup_{n\geqslant 1} n^\beta\|Dh^n(\theta) - Dh(\theta)\| < +\infty.$$

  In this case we set

$$M_0 = \gamma^{-1}(1/(n^2\log(n))),\quad M_l = \gamma^{-1}(m^\ell\log(m)/(n^2\log(n)(m-1))),\ \ell = 1,\cdots,L.$$

Then one has

$$n^{\alpha}(\Theta_n^{ml} - \theta^*) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h, 1, \theta^*) + \mathcal{N}(0, \Sigma^*), \quad n \to +\infty$$

with

$$\Sigma^* := \int_0^{\infty} \left(e^{-s(Dh(\theta^*)-\zeta I_d)}\right)^T \tilde{\mathbb{E}}\left[\left(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V]\right)\left(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V]\right)^T\right] e^{-s(Dh(\theta^*)-\zeta I_d)} ds.$$

Remark 4.2.1. *The previous CLT for the multi-level stochastic approximation estimator of θ^* holds if the standard weak error (and thus the implicit weak error), is of order $1/n^{\alpha}$ and the strong rate error is of order $1/n^{\rho}$ with $\alpha > 2\rho$ or $\alpha = 1$ and $\rho = 1/2$. Due to the non-linearity of the scheme, which leads to annoying remainder terms in the Taylor's expansions used in the proof, those results do not seem to easily extend to a weak discretisation error of order $1/n^{\alpha}$ with $\alpha < 1$ and $\rho = 1/2$ or a faster strong convergence rate $\rho > 1/2$. Moreover, for the same reason this result does not seem to extend to the empirical sequence associated to the multi-level estimator according to the Ruppert & Polyak averaging principle.*

The assumption related to the CLTs satisfied by the two sequences $(U^{(n)} - U)_{n \geqslant 1}$ and $(U^{m^{\ell}} - U^{m^{\ell-1}})_{\ell \geqslant 1}$ includes the case of the value at time T of a SDE, namely $U = X_T$ approximated by its continuous Euler-Maruyama scheme $U^n = X_T^n$ with n steps. Under mild assumption on the coefficients of the SDE, one has $\rho = 1/2$. Moreover, U may depend on the whole path of the process. For instance, one may have $U = L_T^0(X)$, the symmetricc local time at level 0 of the one-dimensional continuous diffusion process X and the approximations may be given by

$$U^n = \sum_{i=1}^{[nt]} f\left(u_n X_{\frac{i-1}{n}}, \sqrt{n}\left(X_{\frac{i}{n}} - X_{\frac{i-1}{n}}\right)\right).$$

Then under some assumptions on the function f and the coefficients b, σ, the weak and strong rate of convergence are $\rho = 1/4$, see Jacod [65] for more details. Let us note that we do not know what happens in our results when $\rho > 1/2$ which includes the case of higher order schemes for discretisation schemes of SDE, like the Milstein scheme.

The CLT presented in Theorem 4.2.1 shows that in order to obtain an error of order $1/n^{\alpha}$, one has to set $M_0 = \gamma^{-1}(1/n^{2\alpha})$ and $M_l = \gamma^{-1}(m^{\ell\frac{(1+2\rho)}{2}}(m^{\frac{1-2\rho}{2}} - 1)/(n^{2\alpha}(n^{\frac{(1-2\rho)}{2}} - 1)))$, if $\rho \in (0, 1/2)$ or $M_0 = \gamma^{-1}(1/n^2)$ and $M_l = \gamma^{-1}(m^{\ell}\log(m)/(n^2\log(n)(m-1)))$ if $\rho = 1/2$, $\ell = 1, \cdots, L$ with $L = \log(n)/\log(m)$. In both cases the complexity of the multi-level stochastic approximation method is given by

$$C_{\text{ML-SA}}(\gamma) = C \times \left(\gamma^{-1}(1/(n^{2\alpha}\log(n))) + \sum_{\ell=1}^{L} M_{\ell}(m^{\ell} + m^{\ell-1})\right). \tag{4.3}$$

In order to compute this complexity, we distinguish the two following cases:

- If $\gamma(p) = \gamma_0/p$ then the optimal complexity is given by

$$C_{\text{ML-SA}}(\gamma) = C\left(n^{2\alpha}\log(n) + \frac{n^2(n^{\frac{(1-2\rho)}{2}} - 1)}{m^{\frac{1-2\rho}{2}} - 1}\sum_{\ell=1}^{L} m^{-\frac{(1+2\rho)}{2}\ell}(m^{\ell} + m^{\ell-1})\right) = \mathcal{O}(n^{2\alpha}n^{1-2\rho}),$$

  if $\rho \in (0, 1/2)$ under the constraint $\underline{\lambda}\gamma_0 > \alpha(\alpha - 2\rho)$ and

$$C_{\text{ML-SA}}(\gamma) = C\left(n^2\log(n) + n^2(\log n)^2\frac{m^2 - 1}{m(\log m)^2}\right) = \mathcal{O}(n^2(\log(n))^2),$$

  if $\rho = 1/2$ under the constraint $\underline{\lambda}\gamma_0 > 1$. These computational costs are similar to those achieved by the multi-level Monte Carlo method for the computation of $\mathbb{E}[f(X_T)]$, see Giles [48] and Ben Alaya and Kebaier [1]. As discussed in Giles [48], this complexity attains a minimum near $m = 7$.

- If we choose $\gamma(p) = \gamma_0/p^a$, $\frac{1}{2} < a < 1$ then simple computations show that the computational cost is given by

$$C_{\text{ML-SA}}(\gamma) = C\left(n^{\frac{2\alpha}{a}}\log^{\frac{1}{a}}(n) + n^{\frac{2}{a}}(n^{1-2\rho} - 1)^{\frac{1}{a}}\sum_{\ell=1}^{L} m^{-\frac{(1+2\rho)}{a}\ell}(m^{\ell} + m^{\ell-1})\right) = \mathcal{O}(n^{\frac{2\alpha}{a}}n^{\frac{1-2\rho}{a}}),$$

  if $\rho \in (0, 1/2)$ and

$$C_{\text{ML-SA}}(\gamma) = C \times \left(n^{\frac{2}{a}}\log^{\frac{1}{a}}(n) + n^{\frac{2}{a}}(\log n)^{\frac{1}{a}}\frac{(m-1)^{\frac{1}{a}}(m+1)}{m(\log m)^{\frac{1}{a}}}\sum_{\ell=1}^{L} m^{-\ell(\frac{1}{a}-1)}\right) = \mathcal{O}(n^{\frac{2}{a}}(\log n)^{\frac{1}{a}})$$

  if $\rho = 1/2$.

Hence we see that the optimal complexity is achieved for $\gamma(p) = \gamma_0/p$ under a constraint on γ_0 involving $\underline{\lambda}$. We also obtained a CLT for the statistical Romberg stochastic approximation estimator that is defined by $\Theta_n^{sr} = \theta_{M_1}^{n^\beta} + \theta_{M_2}^n - \theta_{M_2}^{n^\beta}$. The couple $(\theta_{M_2}^n, \theta_{M_2}^{n^\beta})$ is computed using i.i.d. copies of (U^{n^β}, U^n), the random variables U^{n^β} and U^n being perfectly correlated. Moreover, the random variables used to obtain $\theta_{M_1}^{n^\beta}$ are independent to those used for the computation of $(\theta_{M_2}^n, \theta_{M_2}^{n^\beta})$. The optimal parameters are $M_1 = \gamma^{-1}(1/n^{2\alpha})$ and $M_2 = \gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))$. When $\gamma(p) = \gamma_0/p$, a constraint on γ_0 involving $\underline{\lambda}$ appears. Here we can circumvent this issue by considering the *averaged version* of the last estimator namely $\bar{\Theta}_n^{sr} = \bar{\theta}_{M_3}^{n^\beta} + \bar{\theta}_{M_4}^n - \bar{\theta}_{M_4}^{n^\beta}$, with $M_3 = n^{2\alpha}$ and $M_4 = n^{2\alpha-2\rho\beta}$, for any $\beta \in (0,1)$ (note that $\rho \leqslant \alpha$ so that $\rho\beta < \alpha$). The optimal choice is $\beta^* = 1/(1+2\rho)$ leading to an optimal complexity $C_{\text{SR-SA}}(\gamma) = C_{\text{SR-RP}}(\gamma) = C' n^{2\alpha + \frac{1}{1+2\rho}}$.

As a numerical illustration of these results we consider the following toy example. We turn our attention to the computation of the level of the function $\theta \mapsto e^{-rT}\mathbb{E}(X_T - \theta)_+$ (European call option), $X_T = x\exp((r - \frac{\sigma^2}{2})T + \sigma W_T)$, which is given by the Black-Scholes formula.

Therefore, we first fix a value θ^* (the target of our procedure) and compute the corresponding level $\ell = e^{-rT}\mathbb{E}(X_T - \theta^*)_+$ by the Black-Scholes formula. Accordingly, we set $h(\theta) = \mathbb{E}[H(\theta, X_T)]$ with $H(\theta, x) = \ell - (x-\theta)_+$. We use the following values for the parameters: $x = 100,\ r = 0.05,\ \sigma = 0.4,\ T = 1$. We plot in Figure 4.1 the behaviors of $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ for $n = 100, \cdots, 500$. The value of $h^n(\theta^*)$ is approximated by its Monte Carlo estimator and $\theta^{*,n}$ is estimated by θ_M^n, both estimators being computed with $M = 10^8$ samples. The variance of the Monte Carlo estimator ranges from 9.73×10^6 for $n = 100$ to 9.39×10^7 for $n = 500$.

To compare the three methods in terms of computational costs, we compute the different estimators, namely $\theta_{\gamma^{-1}(1/n^2)}^n$ where $(\theta_p^n)_{p\geqslant 1}$ is given by (4.1), Θ_n^{sr} and Θ_n^{ml} for a set of $N = 200$ values of the target θ^* equidistributed on the interval $[90, 110]$ and for different values of n. For each value n and for each method we compute its complexity as the number of operations and the root-mean-squared error which is given by

$$\text{RMSE} = \left(\frac{1}{N}\sum_{k=1}^N (\Theta_k^n - \theta_k^*)^2\right)^{1/2}$$

where $\Theta_k^n = \theta_{\gamma^{-1}(1/n^2)}^n$, Θ_n^{sr} or Θ_n^{ml} is the considered estimator. For each given n, we provide a couple (RMSE, Complexity) which is plotted on Figure 4.2. Let us note that the multi-level SA estimator has been computed for different values of m (ranging from $m = 2$ to $m = 7$) and different values of L. We set $\gamma(p) = \gamma_0/p$, with $\gamma_0 = 2$, $p \geqslant 1$, so that $\beta^* = 1/2$.

We briefly comment Figure 4.2. The curves of the statical Romberg SA and the multi-level SA methods are displaced below the curve of the SA method. Therefore, for a given error, the complexity of both methods are much lower than the one of the crude SA. The difference in terms of computational cost becomes more significant as the RMSE is small, which corresponds to large values of n. The difference between the statistical Romberg and the multi-level SA method is not significant for small values of n, *i.e.* for a RMSE between 1 and 0.1. For a RMSE lower than 5.10^{-2}, which corresponds to a number of steps n greater than about 600-700, we observe that the multi-level SA procedure becomes much more effective than both methods. For a RMSE fixed around 1 (which corresponds to $n = 100$ for the SA algorithm and Statiscal Romberg SA), one divides the complexity by a factor of approximately 5 by using the statistical romberg SA. For a RMSE fixed at 10^{-1}, the computational cost gain is approximately equal to 10 by using either the statistical romberg SA algorithm or the multi-level SA one. Finally, for a RMSE fixed at $5.5.10^{-2}$, the complexity gain achieved by using the multi-level SA procedure instead of the statistical Romberg one is approximately equal to 5.

4.3 Multi-step Richardson-Romberg for stochastic approximation algorithm

This section is devoted to the work **[10]** related to a Multi-step Richardson-Romber method for the stochastic approximation algorithms introduced in **[9]**. Our first result is an expansion of the *implicit discretisation error* $\theta^{*,n} - \theta^*$ presented in the previous expansion. We will denote by **[H-k]** the following set of assumptions:

1. For all $\theta \in \mathbb{R}^d$,
$$h(\theta) - h^n(\theta) = \frac{\Lambda_1^0(\theta)}{n^\alpha} + \cdots + \frac{\Lambda_k^0(\theta)}{n^{\alpha k}} + o\left(\frac{1}{n^{\alpha k}}\right). \tag{4.4}$$

2. $h, h^n \in \mathcal{C}^k(\mathbb{R}^d, \mathbb{R}^d)$ and for all $l \leqslant k-1$, for all $\theta \in \mathbb{R}^d$,
$$D^l h^n(\theta) - D^l h(\theta) = \frac{\Lambda_1^l(\theta)}{n^\alpha} + \cdots + \frac{\Lambda_{k-l}^l(\theta)}{n^{\alpha(k-l)}} + o\left(\frac{1}{n^{\alpha(k-l)}}\right). \tag{4.5}$$

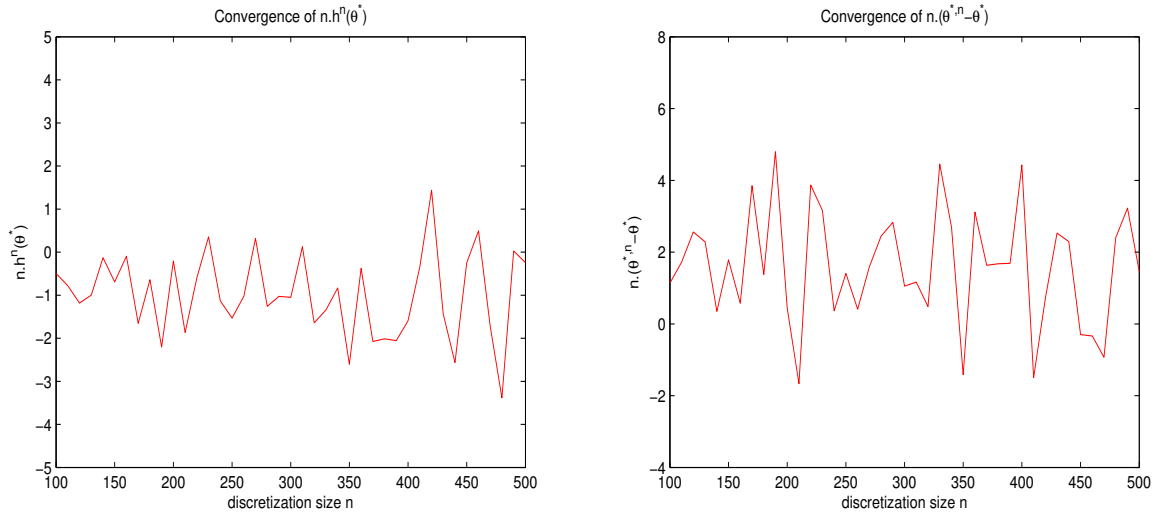

Figure 4.1: On the left: Weak discretization error $n \mapsto nh^n(\theta^*)$. On the right: Implicit discretization error $n \mapsto n(\theta^{*,n} - \theta^*)$, $n = 100, \cdots, 500$.

where for all $\theta \in \mathbb{R}^d$, $\Lambda_1^l(\theta), \cdots, \Lambda_{k-l}^l(\theta)$ and $o(n^{-\alpha(k-l)})$ are multilinear maps from $(\mathbb{R}^d)^l$ to $\mathbb{R}^d$.

3. For all $1 \leqslant l \leqslant k$, $(D^l h^n)_{n \geqslant 1}$ converges locally uniformly towards $D^l h$.

4. $Dh(\theta^*)$ is invertible.

Under **(H-1)** we have seen in Proposiiton 4.1.1 that a first order expansion of $\theta^{*,n} - \theta^*$ holds, that is, $\theta^{*,n} - \theta^* = C_1 n^{-\alpha} + o(n^{-\alpha})$. In **[10]**, we obtained a generalisation of this first result.

Theorem 4.3.1. *Assume that $\theta^{*,n} \to \theta^*$, $n \to +\infty$, and that **[H-p]** holds for some $p \in \mathbb{N}^*$. Then, $\theta^{*,n} - \theta^*$ has an expansion up to order p, that is, the following expansion holds:*

$$\theta^{*,n} - \theta^* = \frac{C_1}{n^\alpha} + \cdots + \frac{C_p}{n^{\alpha p}} + o\left(\frac{1}{n^{\alpha p}}\right).$$

From the previous expansion, one is able to develop a multi-step Richardson-Romberg extrapolation method for stochastic approximation algorithm. The idea is the same as for the Richardson-Romberg method in the context of Monte Carlo linear estimator for the computation of $\mathbb{E}[f(X_T)]$, see Section 1.2.3 for a brief description and Pagès [109] for a complete overview.

We proceed as follows. Let $R \geqslant 2$ be an integer. To devise a SA estimator whose target has an implicit discretisation error of order $n^{-\alpha R}$ as $n \to +\infty$, we introduce a sequence of R random vectors $\{U^{rn}, 1 \leqslant r \leqslant R\}$, $n \in \mathbb{N}^*$. It is assumed that this sequence satisfies $U^{rn} \xrightarrow{\mathbb{P}} U^r$ with $U^r \overset{d}{=} U$, $1 \leqslant r \leqslant R$, all variables being defined on the same probability space. If assumption **[H-R]** holds then for $r = 1, \cdots, R$ one gets

$$\theta^{*,rn} = \theta^* + \sum_{p=1}^{R-1} \frac{C_p}{r^{\alpha p}} \frac{1}{n^{\alpha p}} + \frac{C_R}{r^{\alpha R}} \frac{1}{n^{\alpha R}} (1 + \varepsilon_r(n)), \quad \varepsilon_r(n) \to 0.$$

By solving a *Vandermonde system*, one finds an $Rd \times d$ weight matrix $\mathbf{w} = (\mathbf{w}_1, \cdots, \mathbf{w}_R)^T$, $\mathbf{w}_i$ being a $d \times d$ matrix $i = 1, \cdots, R$ such that $\sum_{r=1}^R \mathbf{w}_r = I_d$ and

$$\sum_{r=1}^R \mathbf{w}_r \theta^{*,rn} = \theta^* + C_R \frac{1}{n^{\alpha R}} \frac{(-1)^{R-1}}{R!^\alpha} (1 + \varepsilon_{R+1}(n)) \tag{4.6}$$

Hence, we removed the $R-1$ first terms of the expansion by considering a suitable linear combination of the $(\theta^{*,rn})_{1 \leqslant r \leqslant R}$. We now approximate the new target $\sum_{r=1}^R \mathbf{w}_r \theta^{*,rn}$, by means of $M \in \mathbb{N}^*$ steps of R stochastic algorithms with dynamics

$$\theta_{p+1}^{rn} = \theta_p^{rn} - \gamma_{p+1} H(\theta_p^{rn}, U^{rn,(p+1)}),\ 0 \leqslant p \leqslant M-1,\ r = 1, \cdots, R, \tag{4.7}$$

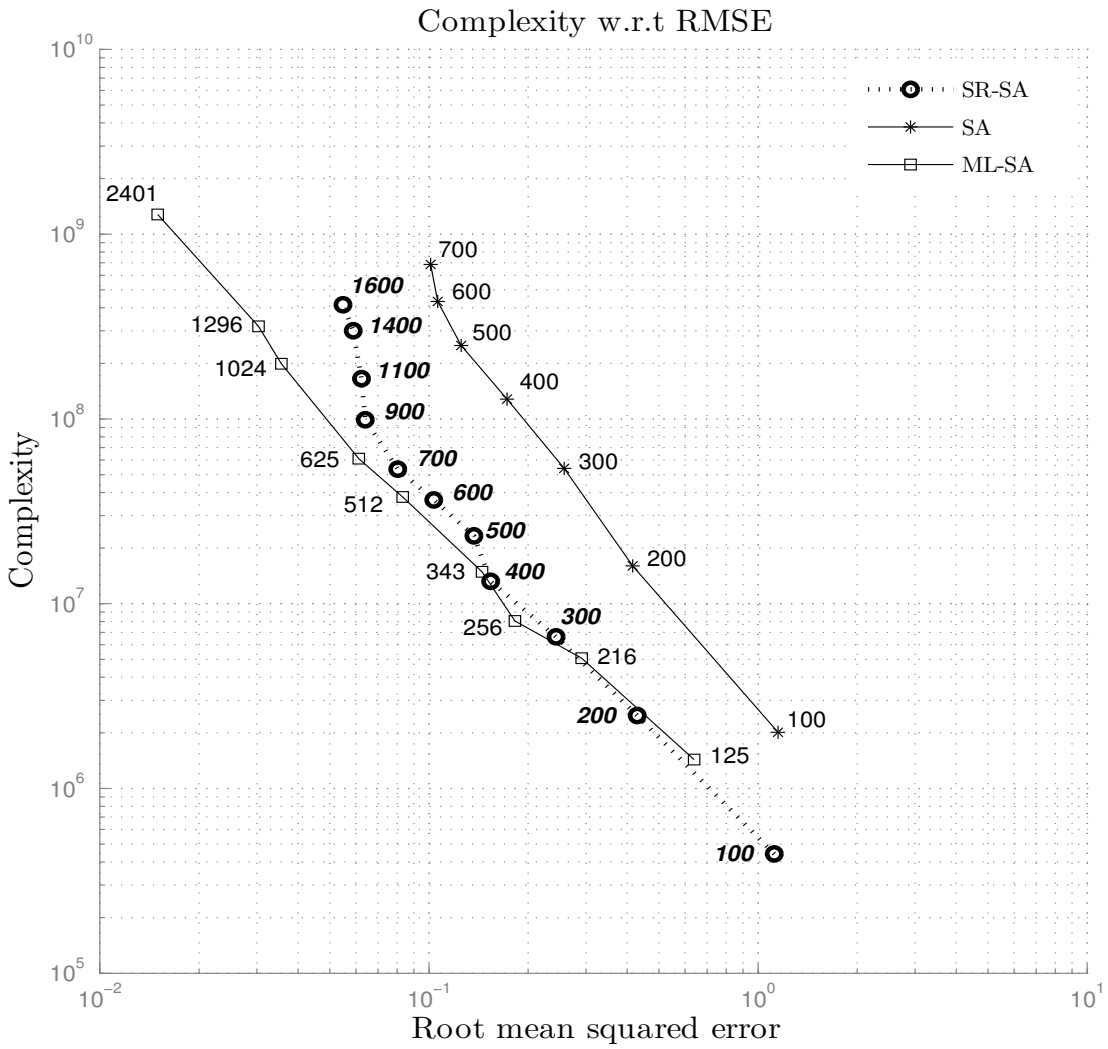

Figure 4.2: Complexity with respect to RMSE.

where $(U^{rn,(p)}, r = 1, \cdots, R)_{1 \leqslant p \leqslant M}$ is an i.i.d sequence with the same law as $(U^{rn}, r = 1, \cdots, R)$, θ_0^{rn}, $r = 1 \cdots, R$ are the initial conditions independent of the innovation sequence satisfying $\sup_{n \geqslant 1} \mathbb{E}|\theta_0^n|^2 < +\infty$ and the sequence $(\gamma_p)_{p \geqslant 1}$ satisfies (1.9). Now the new *statistical error* of the Richardson-Romberg extrapolation estimator writes

$$\mathcal{E}_S^{R-R}(n, M) := \sum_{r=1}^{R} \mathbf{w}_r(\theta^{*,rn} - \theta_M^{rn}).$$

Under mild assumptions, we prove that this statistical error satisfies

$$\mathbb{E}[|\mathcal{E}_S^{R-R}|] \leqslant \nu_R \gamma^{1/2}(M) \left(1 + \phi_1^R(n) + \phi_2^R(M)\right)$$

where $\nu_R := C\mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r H(\theta^*, U^r)\right|^2\right]^{1/2}$, ϕ_1^R, ϕ_2^R are two positive functions satisfying: $\phi_1^R(n) \to 0$ and $\phi_2^R(M) \to 0$ respectively as $M \to +\infty$, $n \to +\infty$ and ϕ_2^R is non-increasing. Hence, the global error of the Richardson-Romberg estimator is bounded by $\mathbb{E}[|\mathcal{E}_{glob}^{R-R}|] \leqslant \mu_R n^{-\alpha R}(1 + |\epsilon_{R+1}(n)|) + \nu_R \gamma^{1/2}(M)(1 + \phi_1^R(n) + \phi_2^R(M)$ with $\mu_R := \frac{|C_R|}{(R!n^R)^\alpha}$.

In practical situations, one is looking for an efficient estimator among the family $\left\{\sum_{r=1}^R \mathbf{w}_r \theta_M^{rn}, (n, M) \in (\mathbb{N}^*)^2\right\}$. To be more precise, one is interested in minimising the computational cost for a given $L^1(\mathbb{P})$-error $\varepsilon > 0$. We assume that the cost of a single simulation of U^n is proportional to n and is given by $K \times n$, where K is a generic positive constant independent of n. It notably corresponds to the case of discretisation schemes of a stochastic process. In the case of the Richardson-Romberg method for SA, at each step $p = 1, \cdots, M$ of the procedure, for $r = 1, \cdots, R$, one has to simulate the random vector $(U^n, U^{2n}, \cdots, U^{Rn})$ so that the global computational cost is given by

$$\text{Cost(R-R)} := KM \sum_{r=1}^{R} rn = KMn\frac{R(R+1)}{2}.$$

Hence the problem of interest writes

$$(n(\epsilon), M(\epsilon)) = \arg \min_{\mu_R n^{-\alpha R}(1+|\epsilon_{R+1}(n)|)+\nu_R\gamma^{1/2}(M)(1+\phi_1^R(n)+\phi_2^R(M) \leqslant \varepsilon} \text{Cost(R-R)}.$$

We solve the above computational optimisation problem asymptotically. More precisely, if $\gamma(p) = \gamma_0/p^\beta$, $\gamma_0 > 0$, $p > 0$, $\beta \in (1/2, 1]$. The multi-step Richardson-Romberg estimator of order R satisfies

$$\inf_{\mu_R n^{-\alpha R}(1+|\epsilon_{R+1}(n)|)+\nu_R\gamma^{1/2}(M)(1+\phi_1^R(n)+\phi_2^R(M)) \leqslant \varepsilon} \text{Cost(R-R)} \sim K\frac{R(R+1)}{2}\gamma_0^{\frac{1}{\beta}}\nu_R^{\frac{2}{\beta}}\mu_R^{\frac{1}{\alpha R}}\frac{1}{\varepsilon^{\frac{2}{\beta}+\frac{1}{\alpha R}}}\left(1+\frac{2\alpha R}{\beta}\right)^{\frac{1}{\alpha R}}\left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}}$$

as $\varepsilon \to 0$. Eventually this asymptotically optimal bound may be achieved with parameters satisfying:

$$n(\varepsilon) \sim \left(\frac{2\alpha R}{\beta} + 1\right)^{\frac{1}{\alpha R}} \mu_R^{\frac{1}{\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} \quad \text{and} \quad M(\varepsilon) \sim \gamma_0^{\frac{1}{\beta}} \nu_R^{\frac{2}{\beta}} \left(1 + \frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}} \varepsilon^{-\frac{2}{\beta}} \quad \text{as } \varepsilon \to 0. \tag{4.8}$$

Let us note that when one decides to implement the Richardson-Romberg extrapolation scheme with an innovation satisfying $U^r = U$ a.s. $r = 1, \cdots, R$ then one has $H(\theta^*, U^r) = H(\theta^*, U)$ *a.s.* for every $r = 1, \cdots, R$ which in turn yields

$$\mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r H(\theta^*, U^r)\right|^2\right] = \mathbb{E}\left[\left|(\sum_{r=1}^R \mathbf{w}_r) H(\theta^*, U)\right|^2\right] = \mathbb{E}\left[|H(\theta^*, U)|^2\right].$$

Hence we clearly see that this choice leads to a control in the L^1-norm of the statistical error of the multi-step Richardson-Romberg SA estimator. On the opposite considering mutually independent innovations U^r lead to an explosion of the previous control with respect to R. Using the exact expression of the weights, one has

$$\begin{aligned}
\mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r H(\theta^*, U^r)\right|^2\right] &= \left(\sum_{r=1}^R \frac{r^{2\alpha R}}{\prod_{j=0}^{r-1}(r^\alpha - j^\alpha)^2 \prod_{j=r+1}^R (j^\alpha - r^\alpha)^2}\right) \mathbb{E}\left[|H(\theta^*, U)|^2\right] \\
&\geqslant \left(\frac{R^R}{R!}\right)^{2\alpha} \mathbb{E}\left[|H(\theta^*, U)|^2\right] \\
&\sim \left(\frac{e^R}{\sqrt{2\pi}\sqrt{R}}\right)^{2\alpha} \mathbb{E}\left[|H(\theta^*, U)|^2\right] \quad \text{as } R \to +\infty,
\end{aligned}$$

For instance when one is concerned with the discretisation of a Brownian diffusion, the first aforementioned case consists in implementing the Richardson-Romberg method with R Euler schemes devised with the *same Brownian motion* W namely $W^r = W,\ r = 1, \cdots, R$ whereas the second case consists in implementing the method with mutually independent Brownian motions W^r. The optimality of this choice is discussed in Pagès [109].

Finally, we illustrate our results by applying the Richardson-Romberg method to the estimation of the ℓ-quantile ($\ell \in (0,1)$) $\theta^* = \inf\{\theta \in \mathbb{R} : \mathbb{P}_x(X_1^d \leqslant \theta) \geqslant \ell\}$ where X is a d-dimensional stochastic process solution to a stochastic differential equation (1.17) where the underlying process Z is a symmetric α-stable, for $\alpha \in (0,2]$. Note that the case $\alpha = 2$ corresponds to the standard Brownian motion.

The mean function h is defined by $h(\theta) = \mathbb{E}[H(\theta, U)]$, $H(\theta, u) = \mathbf{I}_{\{u \leqslant \theta\}}$ and $U = X_1^d$. We approximate U by $U^n = X_1^{n,d}$ where X^n is the continuous Euler scheme associated to X.

In order to prove an expansion in powers of $1/n$ for the implicit discretisation error we first remark that $(h - h^n)(\theta) = \mathbb{P}^x(X_1^d \leqslant \theta) - \mathbb{P}^x(X_1^{n,d} \leqslant \theta)$ so that $\frac{d^k}{d\theta^k}h(\theta) - \frac{d^k}{d\theta^k}h^n(\theta) = \frac{\partial^{k-1}}{\partial\theta^{k-1}}p^{X^d}(0,1,x,\theta) - \frac{\partial^{k-1}}{\partial\theta^{k-1}}p_n^{X^{n,d}}(0,1,x,\theta)$. Hence, one has to rely on the smoothness of the law of X_1^d and $X_1^{n,d}$ in order to verify **[H-k]**.

We rely on the parametrix expansion for the error $p^X(0,1,x,\theta) - p_n^{X^n}(0,1,x,\theta)$ between the densities of the process X and its Euler approximation X^n obtained by Konakov and Mammen [76] in the case of Brownian SDEs and by Konakov and Menozzi [77] in the case of stable SDEs (see Section 1.2.3 for a brief presentation) when the coefficients b, σ of the SDE are assumed to be smooth and bounded. Moreover, we assume that the diffusion matrix is uniformly elliptic and the spectral measure of Z when $\alpha < 2$ is non-degenerated. When $\alpha \leqslant 1$, we put $b = 0$.

Note that we require more than an expansion of the error between the two densities since we also need an expansion of their derivatives in order to verify **[H-k]**. As a simple extension of the results of Theorem 1.1 in Konakov and Mammen in [76], for the Brownian case, and Theorem 21 in Konakov and Menozzi in [77], for the Stable case, we derive in **[10]** an expansion in powers of n^{-1} for $\partial_\theta p^X(0,1,x,\theta) - \partial_\theta p_n^{X^n}(0,1,x,\theta)$. Hence, **[H-k]** is satisfied and we obtain an expansion of $\theta^{*,n} - \theta^*$ in powers of n^{-1} at any order. Our result notably extend the error bound for the computation of the quantile of a diffusion process obtained by Talay and Zheng [123].

To illustrate the method we consider a geometric Brownian motion $(X_t)_{t\in[0,T]}$. Its dynamics and ℓ-quantile are given by

$$\forall t \in [0,T],\ X_t = x_0 \exp((r - \sigma^2/2)t + \sigma W_t), \quad \theta^* = x_0 \exp((r - \sigma^2/2)t + \sigma\sqrt{t}\phi^{-1}(\ell))$$

where W is a one dimensional standard Brownian motion, ϕ is the distribution function of the standard normal distribution $\mathcal{N}(0,1)$. We use the following values for the parameters: $x_0 = 100$, $r = 0.05$, $\sigma = 0.4$, $T = 1$, $\ell = 0.7$. The reference Black-Scholes quantile is $\theta^* = 119.69$. We set $\gamma(p) = \gamma_0/p$ with $\gamma_0 = 60$.

We plot in Figure 4.3 the behaviours of $\sum_{r=1}^R \mathbf{w}_r \theta^{*,rn} - \theta^*$ for $R = 2, 3, 4$ and $n = 2, \cdots, 15$. We estimate $\theta^{*,rn}$ by θ_M^{rn}, with $M = 10^6$ samples for $R = 2$ and $M = 10^8$ samples for $R = 3, 4$ using *consistent Brownian increments*, that is, the same Brownian motion for each R. We clearly see that the Richardson-Romberg estimator efficiency increases with R and the method gives satisfying results with $R = 3, 4$ for small values of n.

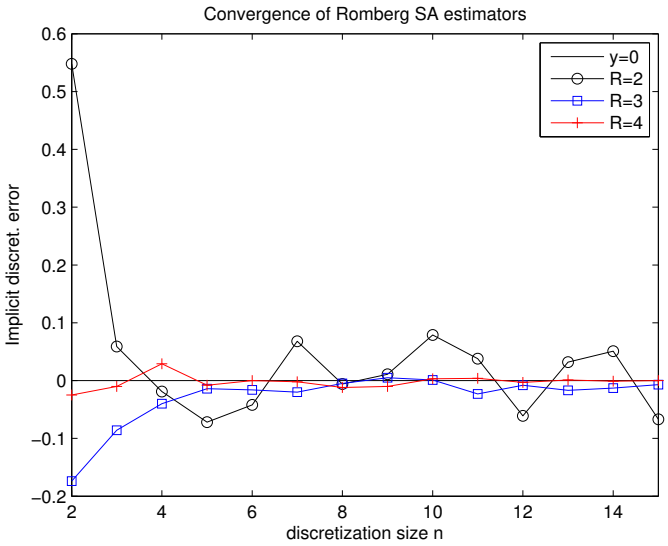

Figure 4.3: Richardson Romberg SA estimators: $\sum_{r=1}^R \mathbf{w}_r \theta^{*,rn} - \theta^*$ with respect to $n = 2, \cdots, 15$ for $R = 2, 3, 4$.

In order to illustrate the computational efficiency of the Richardson-Romberg estimator we set the optimal param-

Target accuracy: ε	L^1-error	time (s)	R	n	M
5.00×10^{-1}	3.21×10^{-1}	0.9×10^{1}	2	14	8.69×10^{5}
2.50×10^{-1}	4.80×10^{-2}	5.15×10^{1}	2	20	3.48×10^{6}
1.25×10^{-1}	4.32×10^{-2}	1.70×10^{2}	3	8	1.21×10^{7}
6.25×10^{-2}	3.48×10^{-2}	7.92×10^{2}	3	10	4.85×10^{7}

Table 4.1: Richardson-Romberg SA estimators for the quantile at level ℓ of a geometric Brownian motion with a target accuracy $\varepsilon = 2^{-p}$, $p = 1, \cdots, 4$.

Target accuracy: ε	L^1-error	time (s)	n	M
5.00×10^{-1}	2.09×10^{-1}	$1,09 \times 10^{2}$	235	1.25×10^{6}
2.50×10^{-1}	3.84×10^{-2}	8.18×10^{2}	469	5.01×10^{6}
1.25×10^{-1}	3.48×10^{-2}	7.09×10^{3}	938	2.00×10^{7}
6.25×10^{-2}	2.91×10^{-2}	5.25×10^{4}	1876	8.01×10^{7}

Table 4.2: Crude SA estimators for the quantile at level ℓ of a geometric Brownian motion with a target accuracy $\varepsilon = 2^{-p}$, $p = 1, \cdots, 4$.

eters according to (4.8), namely

$$n(\varepsilon) = \lceil \left(\frac{2\alpha R}{\beta} + 1 \right)^{\frac{1}{\alpha R}} \mu_R^{\frac{1}{\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} \rceil \quad \text{and} \quad M(\varepsilon) = \lceil \gamma_0^{\frac{1}{\beta}} \nu_R^{\frac{2}{\beta}} \left(1 + \frac{\beta}{2\alpha R} \right)^{\frac{2}{\beta}} \varepsilon^{-\frac{2}{\beta}} \rceil.$$

The target accuracy ε for the L^1-error has been set at $\varepsilon = 2^{-p}$, $p = 1, \cdots, 4$. The L^1-error is estimated using 400 runs of the algorithm. The results are summarized in Table 4.1 for the Richardson-Romberg extrapolation SA method and in Table 4.2 for the crude SA method.[1] Note that as expected the L^1-error is always lower than the specified ε for our estimators. Using the Richardson-Romberg SA scheme instead of the crude SA method leads to a gain in terms of CPU-time varying from 12 (for $\varepsilon = 5.00 \times 10^{-1}$) to 66 (for $\varepsilon = 6.25 \times 10^{-2}$).

4.4 Perspective

From a theoretical point of view it would be interesting to obtain a CLT for the multi-level Robbins-Monro algorithm when one uses the averaging principle of Ruppert-Polyak since we only obtain the optimal computational cost under a constraint on γ_0 when $\gamma(p) = \gamma_0/p$. One expects to remove this constraint by devising a Robbins-Monro algorithm devised with a low decreasing step sequence $\gamma(p) = \gamma_0/p^a$, $\frac{1}{2} < a < 1$ and computing the empirical mean along its trajectories.

From a practical point of view, we investigated multi-level and multi-step Richardson-Romberg methods. It would be interesting to combine them in order to obtain a multi-level Richardson-Romberg method for stochastic approximation as it is done by Pagès and Lemaire [93] for the estimation of $\mathbb{E}[f(U)]$ by the Monte Carlo method. One expects to achieve a complexity of order $\mathcal{O}(n^2 \log(n))$ with a target accuracy of order $\mathcal{O}(n^{-1})$, that is, to remove a $\log(n)$ factor in the multi-level optimal complexity.

[1] The computations were performed on a computer with 4 multithreaded(16) octo-core processors (Intel(R) Xeon(R) CPU E5-4620 @ 2.20GHz).

Part II : Markovian perturbation of stochastic processes and applications

Chapter 5

Weak approximation and asymptotic expansion of some Markov semigroups

In this chapter we present **[11]** and **[12]**. The first work, in collaboration with A. Kohatsu-Higa, deals with asymptotic expansions of Markov semigroups with respect to a small parameter. The key tools are the parametrix method and Malliavin's calculus. We revisit heat kernel expansions of some hypo-elliptic diffusion processes. We also discuss some extensions to skew diffusions with non-regular coefficients.

In the second work, we are interested in the weak approximation error of a skew diffusion with bounded measurable drift and Hölder continuous diffusion coefficient by an Euler-type scheme which consists of iteratively simulating skew Brownian motions with constant drift. Here, we rely on a discrete version of the parametrix method "à la McKean and Singer" [95]. This approach has been successfully applied in the Brownian case by Konakov and Mammen [75] in order to derive some local limit theorems for Markov chain approximations and to establish weak approximation expansions for the Euler scheme of a diffusion [76] when the coefficients of the underlying scheme are smooth. We also refer to Konakov and Menozzi [77] for an extension to stable driven SDEs still in the case of smooth coefficients. However, little attention has so far been given to the case of SDEs with non-smooth coefficients.

We first briefly present in Section 5.1.1 a probabilistic point of view of the parametrix method "à la McKean and Singer". For simplicity of the presentation, we consider the case of time-homogeneous diffusion process. In Section 5.1.2, we present a discrete version for the approximation of Lévy driven SDEs.

In Section 5.2, we illustrate the effectiveness of the parametrix method for the weak approximation of a skew diffusion with non-smooth coefficients. In Section 5.3, we present a general approach for asymptotic expansion of a Markov semigroup $(P_t^\varepsilon)_{t\geqslant 0}$ with respect to the small parameter ε by means of the parametrix method and Malliavin's calculus and illustrate it on two examples.

5.1 A brief presentation of the parametrix method "à la Mc Kean and Singer"

5.1.1 The case of diffusion processes

Let $T > 0$. We consider the (strong) time-homogeneous Markov process $(X_t)_{t\in[0,T]}$ taking values in $\mathbb{R}^d$ with infinitesimal generator defined by:

$$\mathcal{L}f(x) = \langle b(x), \nabla f(x)\rangle + \frac{1}{2}Tr(a(x)D_x^2 f(x)), \quad x \in \mathbb{R}^d \tag{5.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^q$ are bounded measurable functions, $a(.) = (\sigma\sigma^*)(.)$ is η-Hölder continuous and uniformly elliptic, that is, $\forall (x,\xi) \in (\mathbb{R}^d)^2$, $\langle a(x)\xi, \xi\rangle \geqslant \underline{a}|\xi|^2$. Under these assumptions, its is known that the related martingale problem is *well-posed*, see Stroock and Varadhan [120], in other words, the SDE with $\mathcal{L}$ as infinitesimal generator admits a unique weak solution. We denote by $(P_t)_{t\geqslant 0}$ the Markov semigroup defined on bounded measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ by $P_t f(x) = \mathbb{E}[f(X_t)|X_0 = x]$ where $(X_t)_{t\geqslant 0}$ is the unique weak solution to the SDE with $\mathcal{L}$ as infinitesimal generator. The idea of the parametrix method is to approximate the dynamics of X by a simple process $\bar{X}^z$, referred as parametrix process, obtained by removing the drift and freezing the diffusion coefficient to a fixed point $z \in \mathbb{R}^d$. More precisely, one has $\bar{X}_t^z = x + \sigma(z)W_t$, where $(W_t)_{t\in[0,T]}$ stands for a q-dimensional Brownian motion, with infinitesimal generator $\bar{\mathcal{L}}^z f(x) = \frac{1}{2}Tr(a(z)D_x^2 f(x))$. We denote by $(\bar{P}_t^z)_{0\leqslant t\leqslant T}$ the associated Markov semigroup. For $t > 0$, the (transition) density of the parametrix process taken at time t starting from x at time 0 is $y \mapsto \bar{p}^z(t, y - x)$, where $\bar{p}^z(t, x)$ is the Gaussian density with mean 0 and covariance matrix $a(z)t$ taken at x. We also

denote by $(\delta_\varepsilon^z)_{\varepsilon>0}$ a sequence of positive mollifier converging weakly to the Dirac mass at z. For instance, one may consider $\delta_\varepsilon^z(y) = g(\varepsilon, y-z)$ with the notation $g(c,x) := (2\pi c)^{-\frac{d}{2}} \exp(-|y-z|^2/(2c))$. Let $r, \varepsilon > 0$ and $f \in \mathcal{C}_b(\mathbb{R}^d)$. By Itô's lemma, for $s \in [0,t]$, one has

$$\begin{aligned}\bar{P}_{t-s}^z \bar{P}_r^z \delta_\varepsilon^z f(X_s) &= \bar{P}_{t+r}^z \delta_\varepsilon^z f(x) + \int_0^s (\partial_v + \mathcal{L}) \bar{P}_{t-v+r}^z \delta_\varepsilon^z f(X_v) dv + M_s \\ &= \bar{P}_{t+r}^z \delta_\varepsilon^z f(x) + \int_0^s (\mathcal{L} - \bar{\mathcal{L}}^z) \bar{P}_{t-v+r}^z \delta_\varepsilon^z f(X_v) dv + M_s \end{aligned} \tag{5.2}$$

where $(M_s)_{s\in[0,t]}$ is a square integrable martingale. Taking expectations in both sides of the previous equality and letting $s \to t$ and then integrating with respect to z, one gets

$$\int P_t \bar{P}_r^z \delta_\varepsilon^z f(x) dz = \int \bar{P}_{t+r}^z \delta_\varepsilon^z f(x) dz + \int_0^t \mathbb{E}[\int\int \delta_\varepsilon^z(y) f(y) \mathcal{S}_{t-v+r}^z(X_v, y) dz dy] dv. \tag{5.3}$$

where we introduced the notations:

$$\mathcal{S}_t^z f(x) := \int f(y) \mathcal{S}_t^z(x,y) dy, \quad \mathcal{S}_t^z(x,y) := \left\{ \frac{1}{2}(a_{i,j}(x) - a_{i,j}(z)) H_2^{i,j}(a(z)t, y-x) - b_i(x) H_1^i(a(z)t, y-x) \right\} g(a(z)t, y-x)$$

with $H_2^{i,j}(a,x) = (g(a,x))^{-1} \partial_{x_i,x_j}^2 g(a,x)$, $H_1^i(a,x) = (g(a,x))^{-1} \partial_{x_i} g(a,x)$. Passing to the limit as $\varepsilon \downarrow 0$ then in $r \downarrow 0$ and omitting technical details, we get

$$P_t f(x) = \hat{P}_t f(x) + \int_0^t P_v \hat{\mathcal{S}}_{t-v} f(x) dv, \quad t \in [0,T] \tag{5.4}$$

with $\hat{P}_t f(x) := \int f(y) g(a(y)t, y-x) dy$ and $\hat{\mathcal{S}}_t f(x) := \int f(y) \hat{\mathcal{S}}_t(x,y) dy$, $\hat{\mathcal{S}}_t(x,y) := \hat{S}_t^y(x,y)$. Importantly, using the η-Hölder regularity of a, one has the following control: $\forall t \in (0,T]$, $|\hat{\mathcal{S}}_t(x,y)| \leqslant C t^{-1+\frac{\eta}{2}} g(ct, y-x)$ for some positive constants $C := C(a,b,T)$, $c := c(a)$.

The above lines of reasoning allows to select the point z where the coefficients of the parametrix process are fixed as the terminal point where the transition density is evaluated. This argument will be referred as the *diagonalisation argument*. In Section 6.2 of Chapter 6, we will see that this argument still works when the probability measure of the parametrix process is not absolutely continuous with respect to the Lebesgue measure. Assuming that b is continuous, one may iterate the first step expansion (5.4):

$$P_t f(x) = \sum_{n=0}^{N-1} I_t^n f(x) + \mathscr{R}_t^N f(x) \tag{5.5}$$

where $I_t^n f(x) := \int_{\Delta_n(t)} \hat{P}_{s_n} \hat{S}_{s_{n-1}-s_n} \cdots \hat{\mathcal{S}}_{t-s_1} f(x) d\mathbf{s}_n$, $I_t^0 f(x) := \hat{P}_t f(x)$ and $\mathscr{R}_t^N f(x) := \int_{\Delta_N(t)} P_{s_N} \hat{S}_{s_{N-1}-s_N} \cdots \hat{\mathcal{S}}_{t-s_1} f(x) d\mathbf{s}_N$. Using repeatedly the previous control on the kernel $\hat{S}_t(x,y)$, one gets

$$|\mathscr{R}_t^N f(x)| \leqslant |f|_\infty \int_{\Delta_N(t)} d\mathbf{s}_N \prod_{n=0}^{N-1} C_T (s_n - s_{n+1})^{-(1-\frac{\eta}{2})} = |f|_\infty C_T^N t^{N\eta/2} \frac{\Gamma(\eta/2)^N}{\Gamma(1+N\eta/2)},$$

where Γ denotes the Gamma function. From the asymptotics of the Gamma function, one may let N goes to infinity in (5.5) and deduces that the series $\sum_{n\geqslant 0} I_t^n f(x)$ converges absolutely and uniformly for $t \in [0,T]$. From the previous expansion, one may prove the existence of the transition density for the process $(X_t)_{t\in[0,T]}$. Indeed, from Fubini's theorem, one has

$$P_t f(x) = \sum_{n\geqslant 0} I_t^n f(x) = \int f(z_0) \sum_{n\geqslant 0} \left(\int_{\Delta_n(T)} d\mathbf{s}_n \int_{(\mathbb{R}^d)^n} d\mathbf{z}_n \; g(a(z_n)s_n, z_n - x) \, \hat{\mathcal{S}}_{t-s_1}^{z_0}(z_1, z_0) \prod_{i=1}^{n-1} \hat{\mathcal{S}}_{s_i - s_{i+1}}^{z_i}(z_{i+1}, z_i) \right) dz_0$$

so that the transition density $(t,x,y) \mapsto p(t,x,y)$ associated to $(P_t)_{t\geqslant 0}$ exists and is given by

$$p(t,x,y) = \sum_{n\geqslant 0} \int_{\Delta_n(T)} d\mathbf{s}_n \int_{(\mathbb{R}^d)^n} d\mathbf{z}_n \; g(a(z_n)s_n, z_n - x) \, \hat{\mathcal{S}}_{t-s_1}^{z_0}(z_1, y) \prod_{i=1}^{n-1} \hat{\mathcal{S}}_{s_i - s_{i+1}}^{z_i}(z_{i+1}, z_i). \tag{5.6}$$

Moreover, from the asymptotics of the Gamma function, one deduces the Gaussian upper-bound: $\exists C := C(a, |b|_\infty, T)$, $c := c(a) > 0$ such that $\forall (t,x,y) \in (0,T] \times (\mathbb{R}^d)^2$, $p(t,x,y) \leqslant C g(ct, y-x)$. The case of a bounded measurable drift can be handled by an approximation argument: take $(b_n)_{n\geqslant 0}$ such that b_n is continuous, $\sup_{n\geqslant 1} |b_n|_\infty \leqslant |b|_\infty$ and $b_n \to b$ a.e. Denote by $(P_t^n)_{t\geqslant 0}$ the semigroup associated to this approximation. By a compactness argument and uniqueness of the martingale problem, $\left\{(X_t^n)_{t\in[0,T]}, n \geqslant 0\right\}$ converges weakly to $(X_t)_{t\geqslant 0}$ and one can pass to the limit in the corresponding parametrix series. When b is Hölder continuous, one can prove that the series expansion for $(t,x) \mapsto p(t,x,y)$ satisfies the Backward Kolmogorov equation, namely: $(\partial_t + \mathcal{L}) p(t,x,y) = 0$ with $p(t,.,y) \to \delta_y(.)$ as $t \downarrow 0$, see e.g. Friedman [44].

5.1.2 A discrete version of the parametrix method

In the previous section, we have exposed a probabilistic point of view of the parametrix method of McKean and Singer [95] for a time-homogeneous diffusion process. We illustrate now its robustness by showing that it can be easily adapted for approximation schemes. In this section, we consider an $\mathbb{R}^d$-valued process $(X_t)_{t\in[0,T]}$ with generator $\mathcal{L}$ given by (5.1).

For a positive integer N, we define the time step $h = T/N$ and the corresponding uniform time grid $t_i = ih$, $i = 0, \cdots, N$. A natural approximation scheme of the dynamics of $(X_t)_{t\in[0,T]}$ by a Markov chain is given by the following scheme:

$$X^N_{t_{i+1}} = X^N_{t_i} + b(X^N_{t_i})h + \sigma(x')\zeta^N_{i+1}, \quad i = 0, \cdots, N-1, \quad X^N_0 = x \tag{5.7}$$

where $(\zeta^N_i)_{1\leqslant i\leqslant N}$ are i.i.d. random variables such that their common law approximate the law of the true increments W_h and such that ζ^N_1 has a density. If σ is non-degenerated then the one step transition admits a density and so does $X^N_{t_i}$ for $i = 1, \cdots, N$. The density of the Markov chain approximation scheme taken at time $t_{j'}$ starting from x at time t_j will be denoted $y \mapsto p_N(t_j, t_{j'}, x, y)$. In the spirit of the previous section, we introduce a parametrix process denoted by $(\tilde{X}^{N,x'}_{t_i})_{j\leqslant i\leqslant j'}$ associated to $(X^N_{t_i})_{j\leqslant i\leqslant j'}$ which consists in removing the drift part and freezing the diffusion coefficient in the dynamics of (5.7) at the terminal point x' where the density of $(X^N_{t_i})_{j\leqslant i\leqslant j'}$ is evaluated. Its dynamics is given by

$$\tilde{X}^{N,x'}_{t_{i+1}} = \tilde{X}^{N,x'}_{t_i} + \sigma(\tilde{X}^{N,x'}_{t_i})\zeta^N_{i+1}, \quad i = j, \cdots, j'-1, \quad \tilde{X}^{N,x'}_j = x \tag{5.8}$$

and its transition density is denoted by $\tilde{p}^{x'}_N$. To simplify the notations, we will write $\tilde{p}_N(t_j, t_{j'}, x, x')$ for $\tilde{p}^{x'}_N(t_j, t_{j'}, x, x')$. We also introduce the discrete counterpart of the infinitesimal generators considered so far. For $g : \mathcal{C}_b(\mathbb{R}^d)$, $j = 0, \cdots, j'-1$, we define

$$\mathcal{L}^N_{t_j} g(x) = \frac{\mathbb{E}[g(X^N_{t_j+h})|X^N_{t_j} = x] - g(x)}{h}, \quad \text{and} \quad \tilde{\mathcal{L}}^N_{t_j} g(x) = \frac{\mathbb{E}[g(\tilde{X}^{N,x'}_{t_j+h})|\tilde{X}^N_{t_j} = x] - g(x)}{h} \tag{5.9}$$

for $0 \leqslant j < j' \leqslant N$, the discrete kernel $H_N(t_j, t_{j'}, x, x') = \left(\left(\mathcal{L}^N_{t_j} - \tilde{\mathcal{L}}^N_{t_j}\right)\tilde{p}_N(t_j + h, t_{j'}, ., x')\right)(x)$ and finally the discrete time-space convolution type operator $\otimes_N$ as follows

$$(g \otimes_N f)(t_j, t_{j'}, x, x') = \sum_{i=j}^{j'-1} h \int_{\mathbb{R}^d} g(t_j, t_i, x, z) f(t_i, t_{j'}, z, x') dz$$

with the convention that $\sum_{i=j}^{j'-1} \cdots = 0$ if $j \geqslant j'$. From (5.9), it is easily seen that

$$H_N(t_j, t_{j'}, x, x') = h^{-1} \int_{\mathbb{R}^d} (p_N - \tilde{p}^{t_{j'}, x'}_N)(t_j, t_{j+1}, x, z)\tilde{p}_N(t_{j+1}, t_{j'}, z, x')dz.$$

After some simple computations that we omit, see e.g. Lemma 3.6 in Konakov and Mammen [75], for $0 \leqslant j < j' \leqslant N$, the following expansion holds

$$\forall (x, x') \in (\mathbb{R}^d)^2, \; p_N(t_j, t_{j'}, x, x') = \sum_{r=0}^{j'-j} \left(\tilde{p}_N \otimes_N H^{(r)}_N\right)(t_j, t_{j'}, x, x') \tag{5.10}$$

where we define the convolution as follows: $g \otimes_N H^{(0)}_N = g$ and for $r \geqslant 1$, $g \otimes_N H^{(r)}_N = (g \otimes_N H^{(r-1)}_N) \otimes_N H_N$ and use the convention $\tilde{p}(t_{j'}, t_{j'}, ., x') = \tilde{p}(t_{j'}, t_{j'}, ., x') = \delta_{x'}(.)$ in the computation of $\tilde{p} \otimes_N H^{(r)}_N$. From (5.10) and (5.6), the densities of X_t and of its approximation scheme share a similar representation. In order to control the weak approximation error $p - p_N$ between the two densities, one has to quantify the following errors:

1. the error coming from the approximation of the "true" Brownian increment by the sequence $(\zeta^N_i)_{1\leqslant i\leqslant N}$, in other words the error $\tilde{p} - \tilde{p}_N$.
2. the error coming from the discretisation of the space-time convolution operator appearing in (5.6).
3. the contribution when one replaces the continuous kernel $\hat{\mathcal{S}}_{t-s}(.,.)$ by its discrete counterpart $H_N(s, t, ., .)$.

When the coefficients are smooth and time-inhomogeneous, following these steps, Konakov and Mammen [75] proved an error bound for the difference $p - p_N$, namely $|(p - p_N)(0, t, x, x')| \leqslant Ch^{1/2}t^{-d/2}(1 + (|x' - x|/t^{1/2})^S)^{-1}$ for some S depending of the integrability properties of the underlying innovation sequence $(\zeta^N_i)_{1\leqslant i\leqslant N}$. It is the approximation coming from step 1 that dominates in the error bound.

When one directly samples the sequence $(\zeta_i^N)_{1\leqslant i\leqslant N}$ according to the true law $h^{\frac{1}{2}}\mathcal{N}(0,I_d)$, that is, when one removes the error coming from step 1, one obtains an error bound of order h. An expansion of the error in powers of h has been established by Konakov and Mammen [76], as mentioned in equation (1.30) in Section 1.2.3. All these results are obtained under enough regularity on the coefficients b and σ.

When the coefficients b and σ are not smooth, there are only few results. Very recently, in the diffusion case, Konakov and al. [78] derived an upper bound for $p-p_N$ of order $h^{\frac{\eta}{2}-C\psi(h)}$, where $\psi(h)$ is a slowly varying factor that goes to zero as $h\to 0$ under the assumption that the coefficients b and σ are η-Hölder continuous. Their strategy consists in introducing perturbed dynamics of the considered SDE and its scheme by suitably mollifying the coefficients of both dynamics and to quantify the distance between the densities and their respective perturbations.

We also mention the work of Mikulevicius and Platen [101] who established an error bound for the weak approximation error $\mathbb{E}[f(X_T)-f(X_T^N)]$ of order $h^{\eta/2}$ provided $f\in\mathcal{C}^{2+\eta}([0,T]\times\mathbb{R}^d)$ and the coefficients b and σ are η-Hölder continuous in space and $\eta/2$-Hölder continuous in time. We also refer the reader to [100] and [102] for some recent extensions of this result to the case of Lévy driven SDEs.

In the next section we present the results obtained in **[12]** where we proved an error bound for the difference between the densities of the skew diffusion and its Euler approximation of order $h^{\eta/2}$ under the assumption that b is bounded measurable and σ is η-Hölder continuous and uniformly elliptic.

5.2 Weak approximation of a skew diffusion by an Euler-type scheme

We consider the unique weak solution of the following $\mathbb{R}$-valued SDE $(X_t)_{t\geqslant 0}$ with dynamics

$$X_t = x+\int_0^t b(X_s)ds+\int_0^t \sigma(X_s)dW_s+(2\alpha-1)L_t^0(X),\quad \alpha\in(0,1) \tag{5.11}$$

where $W=(W_t)_{t\geqslant 0}$ is a one dimensional Brownian motion defined on a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geqslant 0},\mathbb{P})$ satisfying the usual assumptions and $L^0(X)$ is the symmetric local time of X at the origin.

When $b=0$ and $\sigma=1$, the solution to (5.11) is called the skew Brownian motion. Harrison and Shepp [60] proved that if $|2\alpha-1|\leqslant 1$ then there is a unique strong solution and if $|2\alpha-1|>1$, there is no solution. The case $\alpha=1$ corresponds to reflecting Brownian motion.

Here we assume that the coefficients of the SDE (5.11) satisfies the following assumptions:

(HR) The drift b is bounded measurable and $a=\sigma^2$ is η-Hölder continuous, for some $\eta\in(0,1]$. That is, there exists a positive constant L such that

$$\sup_{x\in\mathbb{R}}|b(x)|+\sup_{(x,y)\in\mathbb{R}^2,x\neq y}\frac{|a(x)-a(y)|}{|x-y|^\eta}<L.$$

(HE) The diffusion coefficient is uniformly elliptic that is there exists $\lambda>1$ such that for every $x\in\mathbb{R}$, $\lambda^{-1}<a(x)<\lambda$. Since σ is continuous, without loss of generality, we may assume that σ is positive.

The previous assumptions guarantee the existence of a unique weak solution to (5.11). Moreover, for any $(t,x)\in\mathbb{R}_+^*\times\mathbb{R}$, X_t admits a density $y\mapsto p(0,t,x,y)$, which is continuous on $\mathbb{R}^*$ and satisfies a Gaussian upper-bound. We refer to Kohatsu-Higa and al. [73] for more details, see also Le Gall [88]. We also refer the interested reader to the recent survey of Lejay [89] and the references therein for various applications of such equation.

In particular, in [73], a representation of the transition density $p(0,t,x,y)$ by means of the parametrix method "à la McKean and Singer" similar to the expansion (5.6) is obtained. However, the parametrix process is chosen to be the skew Brownian motion, that is, $\bar{X}_t^z=x+\sigma(z)W_t+(2\alpha-1)L_t^0(\bar{X}^z)$ for which the transition density $\bar{p}^z(0,t,x,y)$ is defined by

Case 1: For $x\geqslant 0$, one has

$$\bar{p}^z(0,t,x,y)=\{g(a(z)t,y-x)+(2\alpha-1)g(a(z)t,y+x)\}\,\mathbf{I}_{\{y\geqslant 0\}}+2(1-\alpha)g(a(z)t,y-x)\mathbf{I}_{\{y<0\}}.$$

Case 2: For $x<0$, one has

$$\bar{p}^z(0,t,x,y)=\{g(a(z)t,y-x)+(1-2\alpha)g(a(z)t,y+x)\}\,\mathbf{I}_{\{y<0\}}+2\alpha g(a(z)t,y-x)\mathbf{I}_{\{y\geqslant 0\}}.$$

Note also that the infinitesimal generators $\mathcal{L}$ of (5.11) and $\bar{\mathcal{L}}^z$ of $\bar{X}^z$ are respectively given by $\mathcal{L}=\frac{1}{2}a(x)\partial_x^2+b(x)\partial_x$ and $\bar{\mathcal{L}}^z f(x)=\frac{1}{2}a(z)\partial_x^2$ acting on $\mathcal{D}(\alpha)$ the set of bounded continuous functions $f:\mathbb{R}\to\mathbb{R}$ with bounded continuous derivatives f' and f'' on $\mathbb{R}^*$ such that $f'(0+)$ and $f'(0-)$ exists, are finite and satisfies $\alpha f'(0+)=(1-\alpha)f'(0-)$. Instead of Itô's lemma, one has to use the Itô-Tanaka formula in (5.2).

To approximate equation (5.11) on the time interval $[0,T]$, $T>0$, we introduce the Euler type scheme $(X_t^N)_{t\in[0,T]}$ associated to the time step $h=T/N$, $N\in\mathbb{N}^*$ and time grid $t_i=ih$, $i=0,\cdots,N$, defined by $X_0^N=x$ and for all $t_i\leqslant t\leqslant t_{i+1}$

$$X_t^N = X_{t_i}^N + b(X_{t_i}^N)(t-t_i) + \sigma(X_{t_i}^N)(W_t-W_{t_i}) + (2\alpha-1)L_{t-t_i}^0(X^N(t_i,X_{t_i}^N)). \tag{5.12}$$

Observe that the above scheme does not correspond to a standard Euler-Maruyama approximation scheme since we do not discretise the part corresponding to the local time in the dynamics (5.11). However, its computation only requires to be able to simulate exactly the skew Brownian motion with a constant drift at time $t-t_i$. This process is known to be exactly simulatable and we refer to Etoré and Martinez [40] for the exact expression of its density. In partiular, under assumptions **(HR)** and **(HE)**, the transition between t_i and $t\in(t_i,t_{i+1}]$ admits a positive density so that the discretisation scheme (5.12) admits a positive transition density between two times t_i and t_j, that we will denote $p_N(t_i,t_j,x,y)$, for any $0\leqslant t_i<t_j\leqslant T$, $(x,y)\in\mathbb{R}\times\mathbb{R}^*$. In particular, a Gaussian upper-bound has been established in [73] for the transition density of the skew diffusion (5.11) under **(HR)** and **(HE)**. Our first result is to prove similar Aronson's estimate, that is a Gaussian upper estimate but also a lower bound hold for the discretisation scheme (5.12).

Theorem 5.2.1. *(Two sided Gaussian estimates for the scheme) Under* **(HR)** *and* **(HE)**, *there exist two constants* $C:=C(T,b,\sigma), c:=c(\lambda,\eta)>1$ *such that for every* $0\leqslant j<i\leqslant N$,

$$\forall (x,y)\in\mathbb{R}\times\mathbb{R}^*,\ C^{-1}g(c^{-1}(t_i-t_j),y-x)\leqslant p_N(t_j,t_i,x,y)\leqslant Cg(c(t_i-t_j),y-x).$$

The proof relies on a parametrix expansion of the density $p_N(t_j,t_i,x,y)$ as explained in the previous sections. In particular, one proves that the series representation (5.10) holds for the transition density of the scheme (5.12) with $\tilde{p}_N(t_j,t_{j'},x,x')=\tilde{p}(t_j,t_{j'},x,x')=\bar{p}^{x'}(t_j,t_{j'},x,x')$. The lower bound is obtained by a chaining argument as usually done in this context, see Chapter VII in Bass [14], Kusuoka and Stroock [84] and Lemaire and Menozzi [91] in the case of discretisation schemes of uniformly elliptic diffusions and of some degenerate Kolmogorov processes.

Our second result is an error bound for the difference between the densities of the skew diffusion (5.11) and its approximation scheme (5.12).

Theorem 5.2.2 (Error bound on the difference between the densities)**.** *Assume that* **(HR)** *and* **(HE)** *hold. Then, there exists a constant* $c:=c(\lambda,\eta)>1$ *such that, for all* $0\leqslant t_j<t_i\leqslant N$, *one has*

$$\forall (x,y)\in\times\mathbb{R}\times\mathbb{R}^*,\ |(p-p_N)(t_j,t_i,x,y)|\leqslant C(T,b,\sigma)h^{\eta/2}g(c(t_i-t_j),y-x)$$

where $T\mapsto C(T,b,\sigma)$ *is a non-decreasing positive function.*

Observe that the weak rate $h^{\eta/2}$ is coherent with previous results obtained in the literature for the weak approximation error $\mathbb{E}[f(X_T)-f(X_T^N)]$ where X is given by a Brownian SDE and X^N is its Euler scheme when b and σ are η-Hölder continuous in space and $\eta/2$-Hölder continuous in time, see [101]. In Konakov and al. [78], the error bound on the densities is proved to be of order $h^{\frac{\eta}{2}-C\psi(h)}$, $\psi(h)=\log_3(h^{-1})/\log_2(h^{-1})$ when the coefficients b and σ are η-Hölder continuous. Moreover, in their results, the singularity in time is given by $C(t_i-t_j)^{-(1-\eta/2)\eta/2}$ whereas this quantity does not appear in the previous theorem so it is tighter in this sense. For $\alpha=1/2$, which corresponds to the case of (time homogeneous) diffusion process (since the local time part vanishes), compared to [78], our result notably removes the slowly varying factor $\psi(h)$ and shows that the drift plays no role in the approximation of the density since we only require b to be a bounded measurable function. This phenomenon is not surprising since one of the advantages of the parametrix method is the removal of the drift part in the analysis. Eventually, it should be possible to extend our strategy of proof to the case of multi-dimensional Brownian diffusion.

In order to prove the error bound of Theorem 5.2.2, our strategy is the following. The main point is to compare the parametrix series of the transition $p(t_j,t_i,x,y)$ obtained in [73] and the parametrix series of the Euler scheme (5.12) which differs on account of the discrete nature of time-space convolution operator $\otimes_N$ and the discrete smoothing kernel H_N. In order to do this, we introduce for $0\leqslant j<j'\leqslant N$,

$$\forall (x,x')\in\mathbb{R}\times\mathbb{R}^*,\ p_N^d(t_j,t_{j'},x,x')=\sum_{r\geqslant 0}\tilde{p}\otimes_N H^{(r)}(t_j,t_{j'},x,x'). \tag{5.13}$$

Arguments similar to those used in Section 5.1.1 show that the series (5.13) converge absolutely and uniformly on $\mathbb{R}\times\mathbb{R}^*$ and that p_N^d satisfies the following Gaussian upper-bound:

$$\forall (x,x')\in\mathbb{R}\times\mathbb{R}^*, p_N^d(t_j,t_{j'},x,x')\leqslant C(T,|b|_\infty,a)g(c(t_{j'}-t_j),x'-x). \tag{5.14}$$

Indeed, by induction on r, one proves that

$$\forall (x,x')\in\mathbb{R}\times\mathbb{R}^*, |\tilde{p}\otimes_N H^{(r)}(t_j,t_{j'},x,x')|\leqslant C^r(t_{j'}-t_j)^{r\eta/2}\prod_{i=1}^r B\left(1+\frac{(i-1)\eta}{2},\frac{\eta}{2}\right)g_{c(t_{j'}-t_j)}(x'-x) \tag{5.15}$$

where $C := C(\lambda,\eta)(|b|_\infty T^{\frac{1-\eta}{2}}+1)$ and $B(m,n)=\int_0^1 dv(1-v)^{m-1}v^{n-1}$ is the Beta function. Then, (5.14) follows from the previous bound. We omit technical details.

The idea is now to decompose the global error as follows:

$$(p-p_N)(t_j,t_i,x,y)=(p-p_N^d)(t_j,t_i,x,y)+(p_N^d-p_N)(t_j,t_i,x,y).$$

A similar decomposition has been used in [76] when the coefficients b and σ are smooth. The smoothness of the coefficients notably allows to use Taylor expansions to express the transition density p as the fundamental solution of the underlying parabolic PDE and to use integration by parts (that may be expressed as the duality relation satisfied by the infinitesimal generator when seen as a differential operator) in order to equilibrate time singularities. Obviously, these arguments do not work here under the mild smoothness assumption **(HR)** so that computations become more delicate. In a first step, we express the difference $p_N^d-p_N$ in an infinite parametrix series that involves the difference between the two kernels H and H_N. Then, the symmetric Itô-Tanaka formula allows to express the difference $H-H_N$ as the difference of the kernel H between two consecutive discretisation times plus a remainder term $\mathscr{R}_N$. We then study the weak approximation rate induced by $\mathscr{R}_N$. The delicate point here lies in the non-differentiability of $x\mapsto\partial_x\tilde{p}(t_j,t_{j'},x,x')$ at zero caused by the presence of the local time part in the dynamics of X^N and $\tilde{X}^N$ which prevents us to use (again) the Itô-Tanaka formula. In a second step, in order to study $p-p_N^d$, we express this difference as an infinite parametrix series that involves the difference between the two convolution operators $\otimes$ and $\otimes_N$. Then, we notably use a (kind of) Lipschitz property in time for the transition density p and a smoothing procedure for the drift part.

5.3 Asymptotic expansion of Markov semigroups and application to diffusion processes

In this section, we present the results obtained in **[11]** in collaboration with A. Kohatsu-Higa. We develop a general framework for asymptotic expansion of a Markov semigroup $(P_t^\varepsilon)_{t\geqslant 0}$ with respect to a small parameter ε based on the parametrix technique and Malliavin's calculus. The main idea can be summed up as follows. Assume that a first order expansion for the infinitesimal generator $\mathcal{L}_t^\varepsilon$ of $(P_t^\varepsilon)_{t\geqslant 0}$ holds, that is $\mathcal{L}_t^\varepsilon=\mathcal{L}_t^0+\varepsilon\mathcal{L}_t^{1,\varepsilon}$, $t\geqslant 0$, where $\mathcal{L}_t^0$ is the infinitesimal generator of some strongly continuous Feller semigroup $(P_{s,t}^0)_{0\leqslant s\leqslant t}$, which will be referred to as the *proxy semigroup*. The basic argument in order to obtain the expansion in powers of ε consists in writing

$$P_{0,t}^\varepsilon f(x)-P_{0,t}^0 f(x)=\int_0^t\partial_s(P_{0,s}^\varepsilon P_{s,t}^0)f(x)ds=\int_0^t P_{0,s}^\varepsilon(\mathcal{L}_s^\varepsilon-\mathcal{L}_s^0)P_{s,t}^0 f(x)ds=\varepsilon\int_0^t P_{0,s}^\varepsilon\mathcal{L}_s^{1,\varepsilon}P_{s,t}^0 f(x)ds \tag{5.16}$$

where we assumed that $P_{s,t}^0 f\in\mathrm{Dom}(\mathcal{L}_s^\varepsilon)$ for $f\in\mathrm{Dom}(\mathcal{L}_s^0)$ and $s\in[0,t]$. In order to iterate the above expansion, one needs an expansion of $\mathcal{L}^\varepsilon$, namely we assume $\mathcal{L}^\varepsilon=\mathcal{L}^0+\varepsilon\mathcal{L}^1+\cdots+\varepsilon^{\ell+1}\mathcal{L}^{\ell+1,\varepsilon}$. More precisely, for every $f\in\mathcal{C}_c^\infty(\mathbb{R}^d)\subset\cap_{t\geqslant 0}\mathrm{Dom}(\mathcal{L}_t^\varepsilon)\cap\mathrm{Dom}(\mathcal{L}_t^0)\cdots\cap\mathrm{Dom}(\mathcal{L}_t^{\ell+1,\varepsilon})$, we assume that

$$\mathcal{L}_t^\varepsilon f(x)=\mathcal{L}_t^0 f(x)+\varepsilon\mathcal{L}_t^1 f(x)+\cdots+\varepsilon^{\ell+1}\mathcal{L}_t^{\ell+1,\varepsilon}f(x). \tag{5.17}$$

Then, under some integrability conditions that we omit, for all $(t,x)\in\mathbb{R}_+\times\mathbb{R}^d$, the following expansion holds

$$P_{0,t}^\varepsilon f(x)=P_{0,t}^0 f(x)+\sum_{p=1}^{\ell}\varepsilon^p I_p(f)(x)+\varepsilon^{\ell+1}(R_\ell^{1,\varepsilon}+R_\ell^{2,\varepsilon}) \tag{5.18}$$

with

$$I_p(f)(x):=\sum_{k=1}^{p}\sum_{\substack{\alpha=(\alpha_1,\ldots,\alpha_k)\\ \|\alpha\|=p}}\int_{\Delta_k(t)}d\mathbf{s}_k P_{0,s_k}^0\mathcal{L}_{s_k}^{\alpha_k}P_{s_k,s_{k-1}}^0\cdots P_{s_2,s_1}^0\mathcal{L}_{s_1}^{\alpha_1}P_{s_1,t}^0 f(x)$$

and $\limsup_{\varepsilon\to 0}\left\{|R_\ell^{1,\varepsilon}|+|R_\ell^{2,\varepsilon}|\right\}<\infty$. The norm of the index $\alpha=(\alpha_1,\cdots,\alpha_\ell)$, $\alpha_i\in\{1,\cdots,\ell\}$, is defined as $\|\alpha\|=\sum_{j=1}^\ell\alpha_j$.

The main application discussed in this work concerns semigroups generated by diffusion processes which are defined as solutions of SDEs. In order to extend the expansion of $P_{0,t}^\varepsilon f$ to the class of bounded Borel functions, we will rely on some integration by parts formula on the Wiener space. We now explain the main idea. We will use the standard notations of Malliavin calculus. We refer to Ikeda and Watanabe [64] and Nualart [108] for a complete exposition of this topic. We consider the unique strong solution Z of the following SDE

$$Z_t=Z_s+\int_s^t b(u,Z_u)du+\int_s^t\sigma_\ell(u,Z_u)dW_u^\ell;\ Z_0=z\in\mathbb{R}^d,(s,t)\in[0,T]^2 \tag{5.19}$$

with coefficients $b(t,.), \sigma_\ell(t,.) \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$, $t \in [0,T]$, $\ell = 1, \cdots, q$, with bounded derivatives of all orders greater than one. We also assume that $b(t,0), \sigma_1(t,0), \cdots, \sigma_q(t,0)$ are bounded on $[0,T]$. The corresponding time inhomogeneous Markov semigroup is given by $P_{s,t}f(x) := \mathbb{E}[f(Z_t(s,x))]$, $0 \leqslant s \leqslant t \leqslant T$, for any given bounded measurable function f.

It is well known that the above non-homogeneous diffusion defines a $\mathcal{C}^\infty$ non-homogeneous stochastic flow, see e.g. Kunita [80]. Therefore we may define flow derivatives of the type $\partial^k_{Z^{\alpha_1}_{s_1}, \ldots, Z^{\alpha_k}_{s_k}} Z_t$ for any $\mathbf{s}_k \in [0,t]^k$ and any multi-index $\alpha \in \{1, \cdots, d\}^k$, $k \in \mathbb{N}^*$. Similarly, we also define accordingly the multilinear form $\partial^i_{Z_s} Z_t$, $i \in \mathbb{N}$. Notice that we can therefore understand $\partial^k_{Z^{\alpha_1}_{s_1}, \ldots, Z^{\alpha_k}_{s_k}} Z_t$ as a (random) function of the vector $(Z_{s_1}, \cdots Z_{s_k})$.

For $j \in \mathbb{N}$, $s \in [0,T]$, $\varepsilon \in [0,1]$, we consider coefficients $B^j_s(\varepsilon,.) \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$ and $A^j_s(\varepsilon,.) \in \mathcal{C}^\infty(\mathbb{R}^d, S^d(\mathbb{R}))$ such that for any multi-index α of elements of $\{1, \cdots, d\}$ and any $j \in \mathbb{N}$,

$$\sup_{s \in [0,T], \varepsilon \in [0,1]} \left\{ |\partial_\alpha B^j_s(\varepsilon, x)| + |\partial_\alpha A^j_s(\varepsilon, x)| \right\} \leqslant Q_\alpha(|x|) \tag{5.20}$$

for some polynomial function Q_α. We also define the corresponding non-homogeneous derivative operators

$$\mathcal{L}^{j,\varepsilon}_s f(x) = B^j_s(\varepsilon, x) \nabla_x f(x) + \frac{1}{2} A^j_s(\varepsilon, x) \partial^2_x f(x) \tag{5.21}$$

and note that we are using product notation: $b \nabla_x f(x) = \sum_{i=1}^d b_i \partial_{x_i} f(x)$ and $A \partial^2_x f(x) = \sum_{i,j=1}^d A_{i,j} \partial^2_{x_i, x_j} f(x)$ for $f \in \mathcal{C}^2(\mathbb{R}^d)$, $b \in \mathbb{R}^d$ and $A \in S^d(\mathbb{R})$. With this notation together with the existence of smooth flows the interchange of derivatives and expectations is straightforward, that is for any $f \in \mathcal{C}^\infty_b(\mathbb{R}^d)$, one has

$$\mathcal{L}^{j,\varepsilon}_s P_{s,t} f(x) = \mathbb{E}\left[\mathcal{L}^{j,\varepsilon}_{Z_s} f(Z_t) \middle| Z_s = x \right], \text{ with } \mathcal{L}^{j,\varepsilon}_{Z_s} f(Z_t) = B^j_s(\varepsilon, Z_s) \nabla_{Z_s} f(Z_t) + \frac{1}{2} A^j_s(\varepsilon, Z_s) \partial^2_{Z_s} f(Z_t) \tag{5.22}$$

Clearly, this formula transfers derivatives operators applied to semigroups into differentiation along the flow generated by the diffusion. From here, we can also obtain the following generalised transfer formula:

$$y \mapsto \mathcal{L}^{\alpha_p,\varepsilon}_{s_p} P_{s_p, s_{p-1}} \cdots P_{s_2,s_1} \mathcal{L}^{\alpha_1,\varepsilon}_{s_1} P_{s_1,t} f(y) = \mathbb{E}\left[\mathcal{L}^{\alpha_p,\varepsilon}_{Z_{s_p}} \cdots \mathcal{L}^{\alpha_1,\varepsilon}_{Z_{s_1}} f(Z_t) | Z_{s_p} = y \right]$$

with

$$\mathcal{L}^{\alpha_p,\varepsilon}_{Z_{s_p}} \cdots \mathcal{L}^{\alpha_1,\varepsilon}_{Z_{s_1}} f(Z_t) = \sum_{\beta \in \mathcal{A}_0(2p)} \partial_\beta f(Z_t) \Theta^{\beta,\alpha}_\varepsilon(\mathbf{s}_p, Z_{s_p}).$$

We use the notation $\mathcal{A}_0(p)$ for the set of multi-indices $\alpha = (\alpha_1, \cdots, \alpha_\ell) \in \{1, \cdots, d\}^\ell$, with length $\ell \in \{1, \cdots, p\}$. Also $\left\{ y \mapsto \Theta^{\beta,\alpha}_\varepsilon(\mathbf{s}_p, y), \beta \in \mathcal{A}_0(2p), \varepsilon \in [0,1] \right\}$ is a family of smooth functions such that for any $\beta \in \mathcal{A}_0(2p)$, $\varepsilon \in [0,1]$ and any tuple γ of elements of $\{1, \cdots, d\}$, $\Theta^{\beta,\alpha}_\varepsilon : \Delta_p(t) \times \mathbb{R}^d \to \mathbb{D}^\infty$ satisfies

1. $\Theta^{\beta,\alpha}_\varepsilon(\mathbf{s}_p, .) \in \mathcal{C}^\infty(\mathbb{R}^d)$ and $[\partial^\gamma \Theta^{\beta,\alpha}_\varepsilon / \partial y^\gamma](.,.)$ is continuous on $\Delta_p(t) \times \mathbb{R}^d$ a.s.
2. $\forall (k,p) \in \mathbb{N} \times [1, \infty)$, there exists a polynomial $Q_{k,p}$ function such that

$$\forall y \in \mathbb{R}^d, \sup_{\mathbf{s}_p \in \Delta_p(t), \varepsilon \in [0,1]} \left\| [\partial^\gamma \Theta^{\beta,\alpha}_\varepsilon / \partial y^\gamma](\mathbf{s}_p, y) \right\|_{k,p} < Q_{k,p}(|y|).$$

Moreover, if one assumes that the Malliavin covariance matrix Γ_{Z_T} of Z_T is invertible and that $\det(\Gamma_{Z_T})^{-1}$ has moments of all orders. Then, making use of the Malliavin integration by parts formula, for any multi-index α of length p and any $\mathbf{s}_p \in \Delta_p(T)$, one has

$$P_{0,s_p} \mathcal{L}^{\alpha_p,\varepsilon}_{s_p} P_{s_p, s_{p-1}} \cdots P_{s_2,s_1} \mathcal{L}^{\alpha_1,\varepsilon}_{s_1} P_{s_1,T} f(x) = \sum_{\beta \in \mathcal{A}_0(2p)} \mathbb{E}[f(Z_T(x)) H_\beta(Z_T(x), \Theta^{\beta,\alpha}_\varepsilon(\mathbf{s}_p, Z_{s_p}(x)))]. \tag{5.23}$$

We apply this approach to a d-dimensional diffusion processes $(X^\varepsilon_t)_{t \geqslant 0}$ solution of an SDE with dynamics:

$$X^\varepsilon_t = x_0 + \int_0^t b(\varepsilon, X^\varepsilon_s) ds + \varepsilon \int_0^t \sigma_\ell(X^\varepsilon_s) dW^\ell_s, \ (\varepsilon, t) \in [0,1] \times [0,T], \tag{5.24}$$

where $(W_t = W^1_t, \cdots, W^q_t)_{t \in [0,T]}$ is a q-dimensional Brownian motion, $\sigma = (\sigma_1, \cdots, \sigma_q)$ is a system of $\mathbb{R}^d$-valued functions all defined on $\mathbb{R}^d$ and $b : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\varepsilon \in [0,1]$. Our aim is to derive an asymptotic expansion of the density $p^\varepsilon(0,T,x_0,y)$ of (5.24) starting from x_0 at time 0 and evaluated at terminal point y at time T as $\varepsilon \downarrow 0$.

Small noise as well as short time asymptotics can be derived from the heat kernel expansion of (5.24). As the first part of our main results, we obtain the expansion at any order of $p^\varepsilon(0,T,x_0,y)$ by combining the parametrix approach exposed above with Malliavin's calculus techniques.

Short time expansions, that is, the case $b(\varepsilon,.)=\varepsilon^2 b(.)$, have been investigated by Molchanov [106] and Azencott [6] for elliptic operators. An application of the Malliavin calculus to this problem was first discussed by Bismut [21] under the so-called (H2) assumption which is weaker than the ellipticity assumption. Then, Watanabe [127] obtained a similar expansion through the asymptotics of generalized Wiener functionals. Following [6] and [21], Ben Arous [16], [17], combined the Laplace method on Wiener space with the Malliavin calculus to derive short-time asymptotics for hypo-elliptic diffusion processes.

A first order expansion for the transition density of (5.24) has already been considered in Deuschel and al. [37] using the Laplace method on Wiener space, as initially developed by Ben Arous in [16] and [17] in the small time case. Here, following our approach, we derive the full expansion of $p^\varepsilon(0,t,x_0,y)$ at any order under mild conditions, including unicity of the minimiser and a non-degeneracy condition **(ND)** (see below for precise statement) in the spirit of Bismut's hypothesis.

In the sequel we will suppose that the following assumption is in force:

($\mathbf{H}_1$) $b_i \in \mathcal{C}^\infty([0,1]\times\mathbb{R}^d,\mathbb{R})$ and $\sigma_{i,j}\in\mathcal{C}^\infty(\mathbb{R}^d,\mathbb{R})$, $(i,j)\in\{1,\cdots,d\}\times\{1,\cdots,q\}$. Moreover, for all $\varepsilon\in[0,1]$ and for all integer n, the functions $\partial_\varepsilon^n b_i(\varepsilon,.),\sigma_{i,j}(.)$ have bounded derivatives of all orders greater or equal to one.

(WH) $\text{span}\{\sigma_i(x_0), i=1,\cdots,q; [\sigma_i,\sigma_j](x_0), 0\leqslant i,j\leqslant q;\cdots\}=\mathbb{R}^d$, where we set $\sigma_0(.)=b(0,.)$ for convenience.

We consider the process $(X_t^{\varepsilon,h})_{t\in[0,T]}$ with dynamics

$$\forall(\varepsilon,t)\in[0,1]\times[0,T],\ X_t^{\varepsilon,h}=x_0+\int_0^t b(\varepsilon,X_s^{\varepsilon,h})ds+\varepsilon\int_0^t\sigma_\ell(X_s^{\varepsilon,h})dW_s^\ell+\int_0^t\sigma_\ell(X_s^{\varepsilon,h})\dot{h}_s^\ell ds \tag{5.25}$$

where h is an element of the Cameron-Martin space H. Under **($\mathbf{H}_1$)**, for fixed x and $h\in H$, for all integer N, $X_t^{\varepsilon,h}$ has the asymptotic expansion

$$X_t^{\varepsilon,h}=X_t^{0,h}+\sum_{k=1}^N\varepsilon^k\Delta_0^k X^{\cdot,h}(t)+\varepsilon^{N+1}\Delta_\varepsilon^{N+1}X^{\cdot,h}(t) \tag{5.26}$$

with $\Delta_0^k X^{\cdot,h}(t)=\frac{1}{k!}\partial_\varepsilon^k X_t^{\varepsilon,h}|_{\varepsilon=0}$, residue $\Delta_\varepsilon^{N+1}X^{\cdot,h}(t)=\frac{1}{N!}\int_0^1 d\lambda(1-\lambda)^N\partial_\alpha^{N+1}X_t^{\alpha,h}|_{\alpha=\lambda\varepsilon}$ and the convention $\sum_{k=1}^0...=0$.

Moreover, from the previous definition, for all integer N, for all integer $p\geqslant 1$, one has

$$\sup_{r\in[0,1]}\mathbb{E}[\sup_{0\leqslant t\leqslant T}|\Delta_r^{N+1}X_t^{\cdot,h}|^p]<\infty. \tag{5.27}$$

In particular, the limit equation as ε goes to zero of the previous system is given by

$$\begin{aligned}
X_t^{0,h}&=x_0+\int_0^t b(0,X_s^{0,h})ds+\int_0^t\sigma_\ell(X_s^{0,h})\dot{h}_s^\ell ds,\\
X_t^{1,h}&=\int_0^t Y_t^h(Y_s^h)^{-1}\partial_\varepsilon b(0,X_s^{0,h})ds+\int_0^t Y_r^h(Y_s^h)^{-1}\sigma_\ell(X_s^{0,h})dW_s^\ell,\\
Y_t^h&=I_d+\int_0^t\partial_x b(0,X_s^{0,h})Y_s^h ds+\int_0^t\partial_x\sigma_\ell(X_s^{0,h})Y_s^h\dot{h}_s^\ell ds.
\end{aligned}$$

A crucial assumption to derive the expansion of the density is that the set $\mathcal{K}_{x_0,y}:=\left\{h\in H: X_T^{0,h}=y\right\}$ is non-empty. It is satisfied under the strong Hörmander condition (see e.g. Coron [31])

(SH) $\forall x\in\mathbb{R}^d,\ Lie\left\{\sigma_1,\cdots,\sigma_q\right\}(x)=\mathbb{R}^d$,

Now let $\bar{h}\in\mathcal{K}_{x,y}$ which minimises the norm that is $\|\bar{h}\|_H^2=\inf_{h\in\mathcal{K}_{x,y}}\|h\|_H^2$. A standard weak-compactness argument shows that such an $\bar{h}$ always exists (see e.g. Bismut [21] or Watanabe [127]). In the spirit of Bismut [21], we introduce the following non-degeneracy assumption:

(ND) For every $u\in\mathbb{R}^q$,

$$\mathcal{L}\left\{\sigma_1(x_0),\cdots,\sigma_\ell(x_0),[\sigma_1,V](x_0),\cdots,[\sigma_\ell,V](x_0)\right\}=\mathbb{R}^d,\quad \text{for } V(x_0)=\sum_{\ell=1}^q u_\ell\sigma_\ell(x_0)+b(0,x_0)$$

Assuming that **(H$_1$)** and **(ND)** are satisfied and that $h \in H \cap \mathcal{C}^2([0,T])$ then the Malliavin covariance matrix $\Gamma_{\Delta_\varepsilon^1 X_T^{\varepsilon,h}}$ is uniformly non degenerate in the sense that

$$\limsup_{\varepsilon\to 0} \| \det(\Gamma_{\Delta_\varepsilon^1 X_T^{\varepsilon,h}})^{-1}\|_p < \infty. \tag{5.28}$$

Once we have fixed the minimiser on the set $\mathcal{K}_{x_0,y}$ of $\|\bar h\|_H^2$ and in the interests of simplifying notations we omit the variable denoting this element of the Cameron-Martin space in all processes. Therefore, we will have, e.g., $X^{\varepsilon,\bar h} \equiv \bar X^{0,\varepsilon} \equiv \bar X^\varepsilon$, $\Delta_\varepsilon^i \bar X^{\cdot,\bar h} \equiv \bar X^{i,\varepsilon}$, for $\varepsilon \in [0,1]$, $X^{0,\bar h} \equiv \bar X^0$, $\Delta_0^i \bar X^{\cdot,\bar h} \equiv \bar X^i$ and $Y^{\bar h} \equiv \bar Y$.

In particular, the "deterministic" Malliavin covariance matrix $\Gamma_{\bar X_T^1}$ of $\bar X_T^1$ defined by

$$(\Gamma_{\bar X_T^1})_{i,j} = \sum_{\ell=1}^q \int_0^T (\bar Y_T(\bar Y_s)^{-1}\sigma_\ell(\bar X_s^0))_i (\bar Y_T(\bar Y_s)^{-1}\sigma_\ell(\bar X_s^0))_j ds,\ (i,j)\in\{1,\cdots,d\}^2 \tag{5.29}$$

is invertible under **(ND)**.

Moreover, in order to characterise some properties of the minimiser, one uses the Lagrange multiplier method which ensures that

$$\exists \lambda \in \mathbb{R}^d \text{ s.t. for a.a. } t\in[0,T],\ \dot{\bar h}_t^\ell = \langle \lambda, \bar Y_T(\bar Y_t)^{-1}\sigma_\ell(\bar X_t^0)\rangle,\ \ell = 1,\cdots,q. \tag{5.30}$$

Using the dynamics of $(X_t^{1,h})_{t\in[0,T]}$ with (5.30), we deduce the following crucial identity

$$\int_0^T \langle \dot{\bar h}_s, dW_s\rangle = \langle \lambda, \int_0^T \bar Y_T(\bar Y_s)^{-1}\sigma_\ell(\bar X_s^0) dW_s^\ell\rangle = \langle \lambda, \bar X_T^1\rangle - \langle \lambda, \int_0^T \bar Y_T(\bar Y_s)^{-1}\partial_\varepsilon b(0,\bar X_s^0) ds\rangle. \tag{5.31}$$

Finally we introduce the following condition

(H$_2$) $\mathbb{E}\left[\exp\langle \lambda, \bar X_T^2\rangle | \bar X_T^1 = 0\right] < \infty$.

We now give a guideline of the arguments to prove the main result (Theorem 5.3.1) of this section. Let $\delta > 0$ to be fixed later on. Under **(WH)**, the density of X_T^ε may be decomposed as follows

$$p^\varepsilon(0,T,x_0,y) = \mathbb{E}[\delta_y(X_T^\varepsilon)] = \mathbb{E}[\psi_\delta(Z_T^{1,\varepsilon})\delta_y(X_T^\varepsilon)] + \mathbb{E}[(1-\psi_\delta(Z_T^{1,\varepsilon}))\delta_y(X_T^\varepsilon)] \tag{5.32}$$

where $\psi \in \mathcal{C}^\infty(\mathbb{R})$ is a bump function such that $0 \leqslant \psi(x) \leqslant 1$, $\psi(x) = 1$ on $|x| \leqslant 1/2$ and $\psi(x) = 0$ for $|x| > 1$, for $\delta > 0$, $\psi_\delta(x) := \psi(x/\delta)$ and we introduced the following process $Z_t^{1,\varepsilon} = \int_0^t |X_s^\varepsilon - \bar X_s^0|^2 ds,\ t\in[0,T]$. The above decomposition is necessary in order to deal with the remainder terms in the expansion in powers of ε of the first term appearing in the right-hand side of (5.32).

Using Malliavin's integration by parts and a large deviation arguments, one can prove that $0 \leqslant \mathbb{E}[(1-\psi_\delta(Z_T^{1,\varepsilon}))\delta_y(X_T^\varepsilon)] \leqslant C\exp(-(\|\bar h\|_H^2 + c)/(2\varepsilon^2))$, for some positive constant c, C, so that, as it will be clear later on, this term only appears in the remainder of the asymptotic expansion (at any order).

We now focus on the first term appearing in the right hand side of (5.32). By the Girsanov Theorem with drift equals to $\bar h/\varepsilon$, $\bar h \in \mathcal{K}_{x_0,y}$, (5.26) with $N = 0, 1$ and (5.31), one successively deduces

$$\begin{aligned}
\mathbb{E}[\psi_\delta(Z_T^{1,\varepsilon})\delta_y(X_T^\varepsilon)] &= \exp\left(-\frac{\|\bar h\|_H^2}{2\varepsilon^2}\right)\mathbb{E}\left[\exp\left(-\frac{1}{\varepsilon}\int_0^T \langle \bar h_s, dW_s\rangle\right)\psi_\delta(\bar Z_T^{1,\varepsilon})\delta_y(\bar X_T^\varepsilon)\right] \\
&= \varepsilon^{-d}\exp\left(-\frac{\|\bar h\|_H^2}{2\varepsilon^2}\right)\mathbb{E}\left[\exp\left(-\frac{1}{\varepsilon}\int_0^T \langle \bar h_s, dW_s\rangle\right)\psi_\delta(\bar Z_T^{1,\varepsilon})\delta_0(\bar X_T^{1,\varepsilon})\right] \\
&= \varepsilon^{-d}\exp\left(-\frac{\|\bar h\|_H^2}{2\varepsilon^2}\right)\mathbb{E}\left[\exp\left(-\frac{1}{\varepsilon}\int_0^T \langle \lambda, \bar Y_T(\bar Y_s)^{-1}\sigma_\ell(\bar X_s^0) dW_s^\ell\rangle\right)\psi_\delta(\bar Z_T^{1,\varepsilon})\delta_0(\bar X_T^1 + \varepsilon \bar X_T^{2,\varepsilon})\right] \\
&= \varepsilon^{-d}\exp\left(-\frac{\|\bar h\|_H^2}{2\varepsilon^2}\right)\exp\left(\frac{1}{\varepsilon}\langle \lambda, \int_0^T \bar Y_T(\bar Y_s)^{-1}\partial_\varepsilon b(0,\bar X_s^0) ds\rangle\right)\mathbb{E}\left[e^{-\frac{\langle \lambda, \bar X_T^1\rangle}{\varepsilon}}\psi_\delta(\bar Z_T^{1,\varepsilon})\delta_0(\bar X_T^1 + \varepsilon \bar X_T^{2,\varepsilon})\right] \\
&= \varepsilon^{-d}\exp\left(-\frac{\|\bar h\|_H^2}{2\varepsilon^2}\right)\exp\left(\frac{1}{\varepsilon}\langle \lambda, \int_0^T \bar Y_T(\bar Y_s)^{-1}\partial_\varepsilon b(0,\bar X_s^0) ds\rangle\right)\mathbb{E}\left[e^{\langle \lambda, \bar X_T^{2,\varepsilon}\rangle}\psi_\delta(\bar Z_T^{1,\varepsilon})\delta_0(\bar X_T^{1,\varepsilon})\right].
\end{aligned} \tag{5.33}$$

In the above formula, we extracted the main term of the asymptotic expansion. To conclude we prove that $\mathbb{E}\left[e^{\langle \lambda, \bar X_T^{2,\varepsilon}\rangle}\psi_\delta(\bar Z_T^{1,\varepsilon})\delta_0(\bar X_T^{1,\varepsilon})\right]$ admits an expansion in powers of ε using the framework developed above. Roughly

speaking we introduce the Markov semigroup $(P_t^\varepsilon)_{t\in[0,T]}$ associated to the trivariate process $(\bar{X}_t^{1,\varepsilon}, \bar{Z}_t^{1,\varepsilon}, \bar{X}_t^{2,\varepsilon})_{t\in[0,T]}$. We prove that its infinitesimal generator admits an expansion in powers of ε. We use the integration by parts formula (5.23) w.r.t the non-degenerate process $\bar{X}^1$ in order to get rid of the derivatives on the test function. In order to deal with the remainder term in the expansion, one has to use a localised version of the usual Malliavin's integration by parts and the asymptotic non-degeneracy (5.28). Finally we obtain the following expansion.

Theorem 5.3.1. *Let $x_0, y \in \mathbb{R}^d$. Assume that* **($\mathbf{H}_1$)**, **(WH)**, **($\mathbf{H}_2$)** *and* **(ND)** *hold. Suppose that $\mathcal{K}_{x_0,y} \neq \varnothing$ (which is guaranteed under* **(SH)***) and that the minimizer $\bar{h} = \arg\min_{h\in\mathcal{K}_{x_0,y}} \|h\|_H^2$ is unique. Then, for all $\ell \in \mathbb{N}$, one has*

$$p^\varepsilon(0,T,x,y) = \varepsilon^{-d}\exp\left(-\frac{\|\bar{h}\|_H^2}{2\varepsilon^2}\right)\exp\left(\frac{1}{\varepsilon}\langle\lambda, \int_0^T \bar{Y}_T(\bar{Y}_s)^{-1}\partial_\varepsilon b(0,\bar{X}_s^0)ds\rangle\right)\left(c_0(x,y)+\varepsilon c_1(x,y)+\cdots+\varepsilon^\ell c_\ell(x,y)+\mathcal{O}(\varepsilon^{\ell+1})\right).$$

We can compute the exact expressions of the coefficients. One notably has

$$c_0(x,y) = \mathbb{E}\left[\delta_0(\bar{X}_T^1)e^{\langle\lambda,\bar{X}_T^2\rangle}\right],$$
$$c_1(x,y) = \mathbb{E}\left[\delta_0(\bar{X}_T^1)e^{\langle\lambda,\bar{X}_T^2\rangle}\sum_{\beta_1=1}^{2d}\left\{\lambda^{\tilde{\beta}_1}\tilde{H}_{\tilde{\beta}_1}(\bar{X}_T^1,\Theta^{\beta_1}(s_1,\bar{X}_{s_1}),\langle\lambda,\bar{X}_t^2\rangle)+\sum_{\beta_2=1}^{2d}\lambda^{\tilde{\beta}_2}\tilde{H}_{\tilde{\beta}}(\bar{X}_T^1,\Theta^\beta(s_1,\bar{X}_{s_1}),\langle\lambda,\bar{X}_T^2\rangle)\right\}\right].$$

where, $(s,x)\mapsto\Theta^{\beta_1}(s,x)$, $\Theta^\beta(s,x)$ are two explicit smooth random functions and we interpret $\tilde{\beta}_1 = \varnothing$ if $\beta_1 \in \{d+1,\cdots,2d\}$ and that $H_\varnothing \equiv 1$.

As a second application we consider a one dimensional Markov semigroup with singular coefficients for which we derive an asymptotic expansion with respect to a small parameter ε. We consider the following skew diffusion with small skew

$$X_t^\varepsilon = x + \int_0^t b(X_s^\varepsilon)ds + \int_0^t \sigma(X_s^\varepsilon)dW_s + \varepsilon L_t^0(X^\varepsilon),\ \varepsilon\in[0,1). \tag{5.34}$$

where $W=(W_t)_{t\geqslant 0}$ is a one-dimensional standard Brownian motion and $L^0(X^\varepsilon) = (L_t^0(X^\varepsilon))_{t\geqslant 0}$ is the symmetric local time of X^ε at the origin. We work under assumptions **(HR)** and **(HE)** presented in Section 5.2 which guarantees the existence of a unique weak solution to (5.34). The first order approximation is the semigroup $(P_t^0)_{t\geqslant 0}$ associated to $\mathcal{L}_s^0 = \mathcal{L}^0 = \frac{1}{2}a\partial_x^2 + b\partial_x$. For $f\in\mathcal{C}_c^\infty(\mathbb{R})$, we prove the following first step expansion

$$P_t^\varepsilon f(x) = P_t^0 f(x) + \varepsilon\int_0^t\int_{\mathbb{R}}\frac{p^\varepsilon(0,s,x,0+)+p^\varepsilon(0,s,x,0-)}{2}\partial_x p^0(s,t,0,y)f(y)dyds$$

or equivalently,

$$p^\varepsilon(0,t,x,y) = p^0(0,t,x,y) + a(0)\varepsilon\int_0^t ds_1\frac{p^\varepsilon(0,s_1,x,0+)+p^\varepsilon(0,s_1,x,0-)}{2}\partial_x p^0(s_1,t,0,y).$$

We understand that the expansion (5.17) is satisfied in a generalised sense with $\mathcal{L}^{1,\varepsilon} = \mathcal{L}^1 = a(0)\delta_0\partial_x$. Then, using some sharp estimates on $\lim_{\eta\downarrow 0}\frac{1}{2}(\partial_x p^0(s_1,t,0,\eta)+\partial_x p^0(s_1,t,0,-\eta))$, we prove that we can iterate the above first step expansion.

Theorem 5.3.2. *Under* **(HR)** *and* **(HE)**, *for all $\varepsilon\in[0,1)$, the transition density of* (5.34) *can be expressed as*

$$\forall(t,x,y)\in\mathbb{R}_+^*\times\mathbb{R}\times\mathbb{R}^*,\ p^\varepsilon(0,t,x,y) = p^0(0,t,x,y)+\sum_{p\geqslant 1}(a(0)\varepsilon)^p\int_{\Delta_p(t)}d\mathbf{s}_p\prod_{\ell=1}^{p-1}\partial_x p^0(s_{\ell+1},s_\ell,0,0)p^0(0,s_p,x,0)\partial_x p^0(s_1,t,0,y)$$

with $s_0 = t$. Alternatively, for all positive integer n, one has

$$\forall(t,x,y)\in\mathbb{R}_+^*\times\mathbb{R}\times\mathbb{R}^*,\ p^\varepsilon(0,t,x,y) = p^0(0,t,x,y)+\sum_{p=1}^{n}(a(0)\varepsilon)^p\int_{\Delta_p(t)}d\mathbf{s}_p\prod_{\ell=1}^{p-1}\partial_x p^0(s_{\ell+1},s_\ell,0,0)p^0(0,s_p,x,0)\partial_x p^0(s_1,t,0,y)$$
$$+\varepsilon^{n+1}t^{(n+1)\frac{\eta}{2}-\frac{1}{2}}R_n^{1,\varepsilon}$$

with $\limsup_{\varepsilon\to 0}|R_n^{1,\varepsilon}| < +\infty$.

Chapter 6

One-dimensional elliptic diffusions and some of their path-functionals

In this last chapter we present the results of **[13]** in collaboration with A. Kohatsu-Higa from Ritsumeikan university and L. Li from the university of New South Wales and **[14]** in collaboration with L. Li. In both works we are interested in Markov perturbation problems related to some path-functionals of a one-dimensional elliptic diffusion.

In Section 6.1, we expose the results obtained in **[13]** related to the joint law of the first hitting time of a threshold by a one-dimensional uniformly elliptic diffusion process and to the associated process stopped at the threshold. Our methodology relies on the parametrix method that we apply to the associated Markov semigroup. In the spirit of Bally and Kohatsu-Higa [7], see Section 1.2.5 for a brief account, we obtain a probabilistic representation that may be useful for the construction of an unbiased Monte Carlo path simulation method. Some integrations by parts formula are also established under mild assumptions on the coefficients.

In Section 6.2, we present the results obtained in **[14]** related to the weak existence and uniqueness as well as density estimates for SDEs with coefficients depending on some path-functionals of the process. We consider two examples: a diffusion process with coefficients depending on its running symmetric local time and a diffusion process with coefficients depending on its running maximum. In each example, we also prove the existence of the associated transition density and establish some Gaussian upper-estimates.

6.1 Hitting time and killed process associated to a one-dimensional diffusion

In this section, we expose the results obtained in **[13]** in collaboration with A. Kohatsu-Higa and L. Li related to the first hitting time of a one-dimensional diffusion process with dynamics

$$X_t^x = x + \int_0^t b(X_s^x)ds + \int_0^t \sigma(X_s^x)dW_s,\ x \in \mathbb{R} \tag{6.1}$$

where $(W_t)_{t\geqslant 0}$ stands for a one-dimensional Brownian motion on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$. We denote by $\tau^x := \inf\{t \geqslant 0, X_t^x \geqslant L\}$ the first hitting time of the level L (or equivalently the exit time of the open set $(-\infty, L)$) by the one-dimensional process X and by $(X^x_{\tau_t^x})_{t\geqslant 0}$ the associated killed process. Here we introduced the notation $\tau_t^x := \tau^x \wedge t$. In order to study the joint law of $(\tau_t^x, X_{\tau_t^x}^x)_{t\geqslant 0}$, we introduce the collection of linear maps $(P_t)_{t\geqslant 0}$, acting on $\mathcal{B}_b(\mathbb{R})$, as follows

$$\forall (u,x) \in \mathbb{R}_+ \times \mathbb{R}, \quad P_t h(u,x) = \mathbb{E}\left[h(u+\tau_t^x, X^x_{\tau_t^x})\right]. \tag{6.2}$$

In a multi-dimensional setting and for a domain D such that ∂D is smooth and noncharacteristic, Cattiaux [28] developed a Malliavin's calculus approach to prove that the semigroup associated to a process killed when it hits the boundary ∂D admits an infinitely differentiable kernel under a restricted Hörmander condition on the vector fields. Gaussian bounds on this kernel are also established in small time. We also refer the reader to Ladyzenskaja and al. [85], Friedman [44] and Garroni and Menaldi [46] for constructions of Green functions related to a class of Cauchy-Dirichlet value problems in a uniformly elliptic setting using a partial differential equation framework. In particular, in Chapter IV in the latter reference, the authors introduced the notion of Poisson and Green functions related to a similar Cauchy-Dirichlet problem and in Chapter VI existence and Gaussian upper-bounds are established when the coefficients b and σ are Hölder continuous and $a = \sigma^2$ is uniformly elliptic. As a slight extension of the latter result, in

[13], we prove the existence of a transition density for $(P_t)_{t\geqslant 0}$, thus proving existence of a density for $(\tau_t^x, X^x_{\tau_t^x})$, $t>0$, under the assumption that b is bounded measurable and σ is Hölder continuous and a is uniformly elliptic. Some Gaussian upper-bounds for both densities are also established. In this section, we rather focus on the probabilistic representation for $(P_t)_{t\geqslant 0}$.

The fundamental idea in order to obtain a probabilistic representation for the joint law $(\tau_t^x, X^x_{\tau_t^x})$ is to see the operator P_t as a perturbation of the operator $\bar{P}_t$ defined by $\bar{P}_t h(u,x) = \mathbb{E}[h(u+\bar{\tau}_t^x, \bar{X}^x_{\bar{\tau}_t^x})]$ where $(\bar{X}_t)_{t\geqslant 0}$ is obtained from (6.7) by removing the drift and freezing the diffusion coefficient to the initial point x, namely $\bar{X}_t^x = x+\sigma(x)W_t$ and $\bar{\tau}^x$ is its associated first hitting time. Note that from the reflection principle of Brownian motion, the kernel associated to $\bar{P}_t$ is given by

$$\begin{aligned}\bar{P}_t h(u,x) &= \mathbf{I}_{\{x\geqslant L\}} h(u,x) + \mathbb{E}[h(u+t,\bar{X}_t^x)\mathbf{I}_{\{\bar{\tau}^x>t\}}] + \mathbb{E}[h(u+\bar{\tau}^x, L)\mathbf{I}_{\{\bar{\tau}^x\leqslant t\}}]\\ &= \mathbf{I}_{\{x\geqslant L\}} h(u,x) + S_t^x h(u,x) + K_t^x h(u,x)\\ &= \mathbf{I}_{\{x\geqslant L\}} h(u,x) + \mathbf{I}_{\{x<L\}} \int_u^{u+t}\int_{-\infty}^{L} h(s,z)\left[f^x_{\bar{\tau}}(x,s-u)\delta_L(dz)ds + \bar{q}_t^x(x,z)dz\delta_{u+t}(ds)\right].\end{aligned}$$

with

$$K_t^x h(u,x) = \mathbf{I}_{\{x<L\}}\int_0^t h(u+s,L) f^x_{\bar{\tau}}(x,s)ds, \quad f^y_{\bar{\tau}}(x,s) := \partial_s \mathbb{P}(\bar{\tau}^x\leqslant s) = \frac{L-x}{\sqrt{2\pi a(y)}s^{3/2}}\exp\left(-\frac{(L-x)^2}{2a(y)s}\right)\mathbf{I}_{\{x\leqslant L\}}$$

$$S_t^x h(u,x) = \mathbf{I}_{\{x<L\}}\int_{-\infty}^{L} h(u+t,z)\bar{q}_t^x(x,z)dz, \quad \bar{q}_t^y(x,z) := \left(g(a(y)t, z-x) - g(a(y)t, z+x-2L)\right)\mathbf{I}_{\{x\leqslant L\}}\mathbf{I}_{\{z\leqslant L\}},$$

and $g(a,x) = (2\pi a)^{-\frac{1}{2}}\exp(-\frac{1}{2}a^{-1}x^2)$. A direct application of Itô's lemma allows to characterise the infinitesimal generators $\mathcal{L}$ of P_t and $\bar{\mathcal{L}}^x$ of $\bar{P}_t$, namely for $h\in\mathcal{C}_b^{1,2}(\mathbb{R}_+\times(-\infty,L])$ and $(u,x)\in\mathbb{R}_+\times\mathbb{R}$

$$\begin{aligned}\frac{P_t h(u,x)-h(u,x)}{t} &\to \mathcal{L}h(u,x) := \mathbf{I}_{\{x<L\}}\left(b(x)\partial_2 h(u,x) + \frac{1}{2}a(x)\partial_2^2 h(u,x) + \partial_1 h(u,x)\right),\\ \frac{\bar{P}_t h(u,x)-h(u,x)}{t} &\to \bar{\mathcal{L}}^x h(u,x) = \mathbf{I}_{\{x<L\}}\left(\frac{1}{2}a(x)\partial_2^2 h(u,x) + \partial_1 h(u,x)\right)\end{aligned}$$

as $t\downarrow 0$. We assume that $a\in\mathcal{C}_b^2(\mathbb{R})$, $b\in\mathcal{C}_b^1(\mathbb{R})$. Roughly speaking, the first step of the forward parametrix method writes as follows

$$P_t h(u,x) = \bar{P}h(u,x) + \int_0^t \partial_s(\bar{P}_{t-s}P_s h(u,x))ds = \bar{P}h(u,x) + \int_0^t \bar{P}_{t-s}(\mathcal{L}-\bar{\mathcal{L}}^x)P_s h(u,x))ds$$

where $(\mathcal{L}-\bar{\mathcal{L}}^x)h(v,y) = \frac{1}{2}(a(y)-a(x))\partial_2^2 h(v,y) + b(y)\partial_2 h(v,y)$. Now, by an integration by parts formula, one can remove the derivatives on $P_s h$ and put them on the kernel of $\bar{P}_{t-s}$. Taking care of border terms, one gets

$$P_T h(u,x) - \bar{P}_T h(u,x) = \int_0^T \left(\bar{K}_{T-s}P_s h(u,x) + \bar{S}_{T-s}P_s h(u,x)\right)ds. \tag{6.3}$$

with

$$\bar{K}_t h(u,x) = \bar{K}_t(x,L)h(u+t,L), \quad \bar{K}_t(x,L) = \mathbf{I}_{\{x<L\}}\frac{(a(L)-a(x))}{2}\frac{2(L-x)}{a(x)t}g(a(x)t, L-x) = \mathbf{I}_{\{x<L\}}\frac{(a(L)-a(x))}{a(x)}f^x_{\bar{\tau}}(x,t),$$

$$\bar{S}_t h(u,x) = \int_{-\infty}^{L} h(u+t,z)\bar{S}_t(x,z)dz, \quad \bar{S}_t(x,z) = \mathbf{I}_{\{x<L\}}\left\{\frac{1}{2}\partial_z^2\left[(a(z)-a(x))\bar{q}_t^x(x,z)\right] - \partial_z\left[b(z)\bar{q}_t^x(x,z)\right]\right\}.$$

Moreover, one has the following controls of the two kernels: $|\bar{K}_t(x,L)|\leqslant C\frac{|L-x|^2}{t}g(a(x)t, L-x)\mathbf{I}_{\{x<L\}}\leqslant C\mathbf{I}_{\{x<L\}}g(ct, L-x)$ and $|\bar{S}_t(x,z)|\leqslant C_T\mathbf{I}_{\{x<L\}}\frac{1}{t^{\frac{1}{2}}}g(ct, z-x)$. In order to iterate the first step expansion (6.3), we remark that $\bar{K}_s h(u,L) = \bar{S}_s h(u,L) = 0$ so that $\bar{K}_{T-s_1}\bar{K}_{s_1-s_2}h(u,x) = \bar{K}_{T-s_1}\bar{S}_{s_1-s_2}h(u,x) = 0$. We also define the following family of operators for $(s_0,u,x)\in[0,T]\times\mathbb{R}_+\times(-\infty,L]$:

$$I^n_{s_0}h(u,x) = \begin{cases}\int_{\Delta_n(s_0)} d\mathbf{s}_n\left\{\left(\prod_{i=0}^{n-1}\bar{S}_{s_i-s_{i+1}}\right)\bar{P}_{s_n}h(u,x) + \left(\prod_{i=0}^{n-2}\bar{S}_{s_i-s_{i+1}}\right)\bar{K}_{s_{n-1}-s_n}h(u,x)\right\} & \text{if } n\geqslant 1,\\ \bar{P}_{s_0}h(u,x) & \text{if } n=0,\end{cases} \tag{6.4}$$

where $d\mathbf{s}_p = ds_1\cdots ds_p$ and for a fixed time $t\geqslant 0$, we denote by $\Delta_p(t) = \{\mathbf{s}_p\in[0,t]^p : s_{p+1} := 0\leqslant s_p\leqslant s_{p-1}\leqslant\cdots\leqslant s_1\leqslant t =: s_0\}$.

In **[13]**, we prove that the series $\sum_{n\geqslant 0} I_T^n h(u,x)$ converges absolutely and uniformly and iterating (6.3) one obtains

$$\forall h \in \mathcal{C}_b^{1,2}(\mathbb{R}_+ \times (-\infty, L]), \quad P_t h(u,x) = \sum_{n\geqslant 0} I_t^n h(u,x), \quad (t,u,x) \in [0,T] \times s\mathbb{R}_+ \times \mathbb{R}. \tag{6.5}$$

Actually we will work with the following expression of $I_t^n h(u,x)$. By Fubini's theorem and the change of variable $t_i = t - s_i$, $i = 0, \cdots, n$, one has

$$I_t^n h(u,x) = \int_{\Delta_n(t)} d\mathbf{s}_n \left(\prod_{i=0}^{n-1} \bar{S}_{s_i - s_{i+1}} \right) \bar{P}_{s_n} h(u,x) = \int_{\Delta_n^*(t)} d\mathbf{t}_n \left(\prod_{i=0}^{n-1} \bar{S}_{t_{i+1}-t_i} \right) \bar{P}_{t-t_n} h(u,x)$$

where we used the notation $\Delta_p^*(t) = \{\mathbf{s}_p \in [0,t]^p : s_0 := 0 \leqslant s_1 \leqslant s_2 \leqslant \cdots \leqslant s_p \leqslant t =: s_{p+1}\}$. We begin by providing a probabilistic interpretation of each kernel appearing in (6.4). First one has,

$$\begin{aligned}
\bar{S}_t(x,z) &= \mathbf{I}_{\{x<L\}} \left\{ \partial_z^2 \left[\frac{1}{2}(a(z) - a(x)) \bar{q}_t^x(x,z) \right] - \partial_z \left[b(z) \bar{q}_t^x(x,z) \right] \right\} \\
&= \mathbf{I}_{\{x<L\}} \left\{ (\frac{1}{2}a''(z) - b'(z)) \bar{q}_t^x(x,z) + (a'(z) - b(z)) \partial_z \bar{q}_t^x(x,z) + \frac{1}{2}(a(z) - a(x)) \partial_z^2 \bar{q}_t^x(x,z) \right\} \\
&= \bar{\theta}_t(x,z) g(a(x)t, z - x).
\end{aligned} \tag{6.6}$$

$$\begin{aligned}
\theta_t(x,z) &:= \mathbf{I}_{\{x<L\}} \Lambda_t(x,z) \left\{ \left(\frac{1}{2}a''(z) - b'(z) \right) + (a'(z) - b(z)) \mu_t^1(x,z) + \frac{1}{2}(a(z) - a(x)) \mu_t^2(x,z) \right\}, \\
\mu_t^1(x,z) &:= H_1(a(x)t, z - x) - \frac{1}{a(x)t} \frac{2(L-x)}{(\exp(-\frac{2(z-L)(L-x)}{a(x)t}) - 1)}, \\
\mu_t^2(x,z) &:= H_2(a(x)t, z - x) + \frac{1}{a^2(x)t^2} \frac{4(z-L)(L-x)}{(\exp(-\frac{2(z-L)(L-x)}{a(x)t}) - 1)}, \\
\Lambda_t(x,z) &:= \mathbb{P}\Big(\max_{0\leqslant v\leqslant t} W_v \leqslant \frac{L-x}{\sigma(x)} \,\Big|\, W_t = \frac{z-x}{\sigma(x)} \Big) = \left\{ 1 - \exp\left(-2 \frac{(L-x)(L-x-(z-x))}{ta(x)} \right) \right\} \mathbf{I}_{\{x\leqslant L\}} \mathbf{I}_{\{z\leqslant L\}}.
\end{aligned}$$

From the expression above, for $h \in \mathcal{B}_b(\mathbb{R}_+ \times (-\infty, L])$, one gets $\bar{S}_T h(u,x) = \mathbb{E}[h(u+T, \bar{X}_T^x) \bar{\theta}_T(x, \bar{X}_T^x)]$. For a given time partition $\pi : 0 = t_0 < t_1 < \cdots < t_N < t_{N+1} = T$, we introduce the Euler scheme $\bar{X}^\pi = (\bar{X}_{t_i}^\pi)_{0\leqslant i\leqslant N+1}$ with the following dynamics

$$\bar{X}_{t_{i+1}}^\pi = \bar{X}_{t_i}^\pi + \sigma(\bar{X}_{t_i}^\pi)(W_{t_{i+1}} - W_{t_i}), \quad \bar{X}_{t_0}^\pi = x$$

which in turn allows us to write the following (partial) probabilistic representation

$$\begin{aligned}
\left(\prod_{i=0}^{n-1} \bar{S}_{t_{i+1}-t_i} \right) \bar{P}_{T-t_n} h(u,x) &= \mathbb{E}[h(u+T, \bar{X}_T^\pi) \mathbf{I}_{\{t_n + \bar{\tau}^{\bar{X}_{t_n}^\pi} > T\}} \bar{\theta}_{t_n - t_{n-1}}(\bar{X}_{t_{n-1}}^\pi, \bar{X}_{t_n}^\pi) \cdots \bar{\theta}_{t_1}(x, \bar{X}_{t_1}^\pi)] \\
&\quad + \mathbb{E}[h(u + t_n + \bar{\tau}^{\bar{X}_{t_n}^\pi}, L) \mathbf{I}_{\{t_n + \bar{\tau}^{\bar{X}_{t_n}^\pi} \leqslant T\}} \bar{\theta}_{t_n - t_{n-1}}(\bar{X}_{t_{n-1}}^\pi, \bar{X}_{t_n}^\pi) \cdots \bar{\theta}_{t_1}(x, \bar{X}_{t_1}^\pi)] \\
&= \mathbb{E}[h(u + (t_n + \bar{\tau}^{\bar{X}_{t_n}^\pi}) \wedge T, \bar{X}_{(t_n + \bar{\tau}^{\bar{X}_{t_n}^\pi}) \wedge T}^\pi) \bar{\theta}_{t_n - t_{n-1}}(\bar{X}_{t_{n-1}}^\pi, \bar{X}_{t_n}^\pi) \cdots \bar{\theta}_{t_1}(x, \bar{X}_{t_1}^\pi)],
\end{aligned}$$

where $\bar{\tau}^x$ is a Lévy distributed random variable with scale parameter $c = (L-x)^2/a(x)$ and location parameter $\mu = 0$ which is independent of the underlying Brownian motion W. Similarly, for the second term appearing in (6.4), after noting that

$$\int_0^{s_{n-1}} \bar{K}_{s_{n-1}-s_n} h(u,x) ds_n = \mathbb{E}[h(u + \bar{\tau}^x, L) \frac{(a(L) - a(x))}{a(x)} \mathbf{I}_{\{\bar{\tau}^x \leqslant s_{n-1}\}}] =: \check{K}_{s_{n-1}} h(u,x)$$

we can write

$$\int_{\Delta_n(T)} d\mathbf{s}_n \left(\prod_{i=0}^{n-2} \bar{S}_{s_i - s_{i+1}} \right) \bar{K}_{s_{n-1}-s_n} h(u,x) = \int_{\Delta_{n-1}^*(T)} d\mathbf{t}_{n-1} \left(\prod_{i=0}^{n-2} \bar{S}_{t_{i+1}-t_i} \right) \check{K}_{T-t_{n-1}} h(u,x).$$

Similarly to the previous term, one obtains the following partial probabilistic representation

$$\begin{aligned}
\left(\prod_{i=0}^{n-2} \bar{S}_{t_{i+1}-t_i} \right) \check{K}_{T-t_{n-1}} h(u,x) &= \mathbb{E}[h(u + t_{n-1} + \bar{\tau}^{\bar{X}_{t_{n-1}}^\pi}, L) \mathbf{I}_{\{\bar{\tau}^{\bar{X}_{t_{n-1}}^\pi} \leqslant T - t_{n-1}\}} \frac{(a(L) - a(\bar{X}_{t_{n-1}}^\pi))}{a(\bar{X}_{t_{n-1}}^\pi)} \\
&\quad \times \bar{\theta}_{t_{n-1}-t_{n-2}}(\bar{X}_{t_{n-2}}^\pi, \bar{X}_{t_{n-1}}^\pi) \cdots \bar{\theta}_{t_1}(x, \bar{X}_{t_1}^\pi)].
\end{aligned}$$

Finally, in the spirit of Bally and Kohatsu-Higa [7], we let $(N(t))_{t\geqslant 0}$ be a simple Poisson process with intensity parameter $\lambda > 0$ and define $N \equiv N(t)$ for $t \in [0,T]$. Let $\zeta_1 < \zeta_2 < \cdots < \zeta_N$ be the jump times of the Poisson process and set $\zeta_0 = 0$, $\zeta_{N+1} = T$. We still denote by π the random time partition $\pi : \zeta_0 = 0 < \zeta_1 < \cdots < \zeta_{N+1} = T$ and denote by $\bar{X}^\pi = (\bar{X}^\pi_{\zeta_i})_{0\leqslant i\leqslant N+1}$ its associated Euler scheme. In **[13]**, we obtain the following theorem.

Theorem 6.1.1. *Let $T > 0$. Assume that $a \in \mathcal{C}^2_b(\mathbb{R})$, $b \in \mathcal{C}^1_b(\mathbb{R})$ and a is uniformly elliptic. Define the two sequences $(\Gamma_N(x))_{N\geqslant 0}$ and $(\bar{\Gamma}_N(x))_{N\geqslant 0}$ as follows*

$$\Gamma_N(x) = \begin{cases} \prod_{j=0}^{N-1} \lambda^{-1}\bar{\theta}_{\zeta_{j+1}-\zeta_j}(\bar{X}^\pi_{\zeta_j}, \bar{X}^\pi_{\zeta_{j+1}}) & \text{if } N \geqslant 1, \\ 1 & \text{if } N = 0, \end{cases}$$

and

$$\bar{\Gamma}_N(x) = \begin{cases} \frac{(a(L)-a(\bar{X}^\pi_{\zeta_{N-1}}))}{a(\bar{X}^\pi_{\zeta_{N-1}})} \prod_{j=0}^{N-2} \lambda^{-1}\bar{\theta}_{\zeta_{j+1}-\zeta_j}(\bar{X}^\pi_{\zeta_j}, \bar{X}^\pi_{\zeta_{j+1}}) & \text{if } N \geqslant 1, \\ 0 & \text{if } N = 0. \end{cases}$$

Then, for all $h \in \mathcal{B}_b(\mathbb{R}_+ \times \mathbb{R})$, for all $(u,x) \in \mathbb{R}_+ \times \mathbb{R}$, the following probabilistic representation holds

$$\mathbb{E}[h(\tau^x_t, X^x_{\tau^x_t})] = e^{\lambda t}\mathbb{E}\left[h((\zeta_{N(t)} + \bar{\tau}^{\bar{X}^\pi_{\zeta_{N(t)}}}) \wedge t, \bar{X}^\pi_{(\zeta_{N(t)}+\bar{\tau}^{\bar{X}^\pi_{\zeta_{N(t)}}})\wedge t})\Gamma_{N(t)}(x)\right]$$
$$+ e^{\lambda t}\mathbb{E}\left[h(\zeta_{N(t)-1} + \bar{\tau}^{\bar{X}^\pi_{\zeta_{N(t)-1}}}, L)\mathbf{I}_{\{\bar{\tau}^{\bar{X}^\pi_{\zeta_{N(t)-1}}} \leqslant t-\zeta_{N(t)-1}\}}\bar{\Gamma}_{N(t)}(x)\right].$$

Similarly, for all $(s,x,z) \in (0,t] \times (-\infty, L]^2$, the following probabilistic representation for the probability law of $(\tau^x_t, X^x_{\tau^x_t})$ is satisfied

$$p_t(0,x,ds,dz) = p^D_t(x,z)dz\delta_t(ds) + p^K(x,s)\delta_L(dz)ds$$

with, for all,

$$p^D_t(x,z) = e^{\lambda t}\mathbb{E}\left[\bar{q}^{\bar{X}^\pi_{\zeta_{N(t)}}}_{t-\zeta_{N(t)}}(\bar{X}^\pi_{\zeta_{N(t)}}, z)\Gamma_{N(t)}(x)\right],$$
$$p^K(x,s) = e^{\lambda t}\mathbb{E}\left[f^{\bar{X}^\pi_{\zeta_{N(t)}}}_{\bar{\tau}}(\bar{X}^\pi_{\zeta_{N(t)}}, t-\zeta_{N(t)})\mathbf{I}_{\{s\geqslant\zeta_{N(t)}\}}\Gamma_{N(t)}(x) + f^{\bar{X}^\pi_{\zeta_{N(t)-1}}}_{\bar{\tau}}(\bar{X}^\pi_{\zeta_{N(t)-1}}, s-\zeta_{N(t)-1})\mathbf{I}_{\{s\geqslant\zeta_{N(t)-1}\}}\bar{\Gamma}_{N(t)}(x)\right].$$

As a corollary, we also obtain the following integration by parts formula.

Corollary 6.1.2. *Suppose that the assumptions of the previous theorem hold. Let $h \in \mathcal{C}^1((-\infty, L])$ satisfying: there exist $c > 0$, such that for all $z \in (-\infty, L]$, $|h(z)| + |h'(z)| \leqslant c\exp(c|z|)$. Then, for all $x \in (-\infty, L)$, one has*

$$\mathbb{E}[h'(X^x_t)\mathbf{I}_{\{\tau^x > t\}}] = -e^{\lambda T}\mathbb{E}\left[h(\bar{X}^\pi_t)\Lambda_{T-\zeta_{N(t)}}(\bar{X}^\pi_{\zeta_{N(t)}}, \bar{X}^\pi_t)\mu^1_{t-\zeta_{N(t)}}(\bar{X}^\pi_{\zeta_{N(t)}}, \bar{X}^\pi_t)\Gamma_{N(t)}(x)\right].$$

An unbiased Monte Carlo method for evaluating $\mathbb{E}[h(\tau^x_t, X^x_{\tau^x_t})]$, $\mathbb{E}[h'(X^x_t)\mathbf{I}_{\{\tau^x>t\}}]$ or the probability law $p_T(0,x,dt,dz)$ of $(\tau^x_t, X^x_{\tau^x_t})$ stems from the probabilistic representations obtained in Theorem 6.1.1 and Corollary 6.1.2. The explosion of the variance may be an important issue that can induce poor convergence rate of the method as pointed out in [4] for unbiased simulation of multi-dimensional diffusions. In these situations, an importance sampling method on the time steps using a Beta or Gamma distribution may be used. In short, it would seem that this approximation will work well in the case of small parameters. A very close analysis could be carried out here but we do not intend to develop importance sampling schemes and refer the interested reader to [4] for some developments in the diffusion case.

6.2 Weak uniqueness and density estimates for SDEs with coefficients depending on some path-functionals

In this section, we expose the results obtained in **[14]** in collaboration with L. Li related to the weak existence and uniqueness as well as density estimates of SDEs with dynamics

$$X_t = x + \int_0^t b(X_s, A_s(X))ds + \int_0^t \sigma(X_s, A_s(X))dW_s, \ t \in [0,T] \tag{6.7}$$

where $(W_t)_{t\geqslant 0}$ stands for a one-dimensional Brownian motion and $(A_t(X))_{t\geqslant 0}$ is an $\mathbb{R}^{d-1}$-valued functional depending on the path X, $d \geqslant 2$. Some examples include its local and occupation times, its running maximum or minimum,

its first hitting time of a level, its running average, etc. From the point of view of applications, systems of the type (6.7) appear in many fields. Let us mention stochastic Hamiltonian systems where $A_t(X) = \int_0^t F(s, X_s)ds$, see e.g. Soize [118] for a general overview, Talay [124] for convergence to equilibrium or Barucci and al. [13] for an application to the pricing of Asian options. We also mention the work of Forde [42], where the author constructs a weak solution to the SDE (6.7) with $b \equiv 0$ and $A_t(X) = \max_{0 \leqslant s \leqslant t} X_s$ is the running maximum of X and investigates an application in mathematical finance.

In order to prove weak uniqueness for the SDE (6.7), we follow the recent approach of Bass and Perkins [15] for proving uniqueness for the martingale problem and illustrated it in the framework of non-degenerate, non-divergence and time-homogeneous diffusion operators under the assumption that the diffusion matrix is strictly positive and Hölder continuous. It has also been recently extended by Menozzi [98] for a class of multi-dimensional degenerate Kolmogorov equations that is the case of a multi-dimensional path functional $A = (A_t^1, \cdots, A_t^N)_{t \geqslant 0}$ given by: $A_t^1(X) = \int_0^t F_1(X_s, A_s(X))ds$, $A_t^2(X) = \int_0^t F_2(A_s^1, \cdots, A_s^N)ds$, $\cdots$, $A_t^N(X) = \int_0^t F_N(A_s^{N-1}, A_s^N)ds$, under an assumption of weak Hörmander type on the functions $(F_1, \cdots, F_N)$. The main new feature added here compared to previous works on this topic is that our technique enables us to deal with a process whose probability law is absolutely continuous with respect to a σ-finite measure.

For instance, if one considers the couple $(X_t, A_t(X))_{t \geqslant 0}$, $A_t(X) = L_t^0(X)$ being the symmetric local time at point 0 accumulated by X up to time t, it is easy to see that on $\{T_0 > t\}$, T_0 being the first hitting time of 0 by X, one has $L_t^0(X) = 0$ whereas on $\{T_0 \leqslant t\}$, the process may accumulate local time so that the probability law of the couple $(X_t, L_t^0(X))$ consists in two parts, one being singular with an atom in the local time part, the other one (hopefully) being absolutely continuous with respect to the Lebesgue measure on $\mathbb{R} \times \mathbb{R}_+$. Hence, we see that for such dynamics the situation is more challenging than in the standard diffusion setting. This new difficulty will be overcome by choosing two independent parametrix kernels, one for each part, and then by proceeding with a non-trivial selection of the freezing point according to the singular measure induced by the approximation process.

We first provide a general framework in order to establish weak uniqueness for SDEs with dynamics (6.7). The contribution of this part compared to the existing literature on this topic lies in the fact that we identify the main assumptions, namely **(H1)** and **(H2)** see below, needed to establish weak uniqueness, thus allowing for a general treatment.

We assume that there exists a weak solution (X, W), $\{\mathcal{F}_t\}$ to (6.7) and that the process $Y := (X_t, A_t(X))_{t \geqslant 0}$, starting from the initial point x at time 0, lives on a closed space $\mathcal{J} \subset \mathbb{R}^d$. The process Y induces a probability measure $\mathbb{P}^x$ (or simply denoted $\mathbb{P}$) on $\Omega = \mathcal{C}([0, \infty), \mathcal{J})$ which is endowed with the canonical filtration $(\mathcal{F}_t)_{t \geqslant 0}$. We consider the collection of linear maps $(P_t)_{t \geqslant 0}$ defined by $P_t f(x) := \mathbb{E}[f(X_t, A_t(X))]$ for $f \in \mathcal{B}_b(\mathcal{J})$.

We denote the approximation or the proxy process by $\bar{X}$, that is the solution of (6.7) with $b = 0$ and the diffusion coefficient σ evaluated at some fixed point $z \in \mathcal{J}$. Without going into details at this point, the key idea is to consider the process Y as a perturbation of the proxy $\bar{Y} = (\bar{X}_t, A(\bar{X}_t))_{t \geqslant 0}$ whose law is denoted by $\bar{p}_t(x, dy) = \bar{p}_t^z(x, dy)$. Accordingly, we define the collection of linear maps $(\bar{P}_t)_{t \geqslant 0}$ by $\bar{P}_t f(x) := \mathbb{E}[f(\bar{X}_t, A_t(\bar{X}))] = \int f(y)\bar{p}_t(x, dy)$ for $f \in \mathcal{B}_b(\mathcal{J})$.

We work under the following set of assumptions:

Assumptions (H1): Given the initial and frozen point $x, z \in \mathcal{J}$.

(i) (a) The proxy process $\bar{Y}^z$ is a Markov process with infinitesimal generator $\bar{\mathcal{L}}^z$.

(b) There exists a σ-finite measure $\nu(x, .)$ such that for all $t > 0$, the law of $\bar{Y}_t^z$ is absolutely continuous with respect to $\nu(x, .)$. More specifically, there exists a $\nu(x, dy)$-integrable function $(t, x, y) \mapsto \bar{p}_t^z(x, y)$ satisfying

$$\bar{p}_t^z(x, dy) = \bar{p}_t^z(x, y)\nu(x, dy) \tag{6.8}$$

and $\bar{P}_t^z f(x) = \int f(y)\bar{p}_t^z(x, y)\nu(x, dy)$ for all $f \in \mathcal{B}_b(\mathcal{J})$.

(ii) There exists a class of functions $\mathcal{D} \subset \mathrm{Dom}(\bar{\mathcal{L}}^z) \cap \mathcal{C}_b(\mathcal{J})$ and a linear operator $\mathcal{L}$ acting on $\mathcal{D}$ such that:

(a) For all $g \in \mathcal{C}_b^\infty(\mathcal{J})$, $\bar{P}_t^z g \in \mathcal{D}$.

(b) For all functions h such that: $s \mapsto h(s, .) \in \mathcal{C}^1(\mathbb{R}_+, \mathcal{D})$, the process

$$h(t, Y_t) - h(0, x) - \int_0^t \{\partial_1 h(s, Y_s) + \mathcal{L}h(s, Y_s)\}\, ds, \quad t \geqslant 0$$

is a continuous square integrable martingale under $\mathbb{P}^x$.

(c) There exists a *parametrix kernel* $\bar{\theta}_t$ with respect to the measure ν, that is a measurable map $(t, z, x, y) \mapsto \bar{\theta}_t^z(x, y)$ such that for all $g \in \mathcal{C}_b^\infty(\mathcal{J})$

$$(\mathcal{L} - \bar{\mathcal{L}}^z)\bar{P}_t^z g(x) = \int g(y)\bar{\theta}_t^z(x, y)\nu(x, dy), \quad t > 0. \tag{6.9}$$

(iii) For all $x, y \in \mathcal{J}$, the maps $(t,z) \mapsto \bar{p}_t^z(x,y)$ and $(t,z) \mapsto \bar{\theta}_t^z(x,y)$ are continuous on $(0,\infty) \times \mathcal{J}$.

(iv) For all $t > 0$, there exists some $\nu(x,dy)$-integrable functions $p_t^*(x,y)$, $\theta_t^*(x,y)$, a constant $\bar{\zeta} = \bar{\zeta}(z) \in \mathbb{R}$ and a positive constant C, eventually depending on t but in a non-decreasing way, such that

$$|\bar{p}_t^z(x,y)| \leqslant p_t^*(x,y), \quad |\bar{\theta}_t^z(x,y)| \leqslant \theta_t^*(x,y), \quad \text{and} \quad \int |\theta_t^*(x,y)|\nu(x,dy) \leqslant Ct^{\bar{\zeta}}. \tag{6.10}$$

For the case $z = y$, we assume that the parametrix kernel enjoys the following *smoothing property:* there exists $\zeta > -1$ and a positive constant C, eventually depending on t in a non-decreasing way, such that

$$\forall t > 0,\ \forall x \in \mathcal{J}, \quad \int |\bar{\theta}_t^y(x,y)|\nu(x,dy) \leqslant Ct^{\zeta}. \tag{6.11}$$

(v) For any $g \in \mathcal{C}_b(\mathcal{J})$, one has

$$\lim_{t \downarrow 0} \int g(y)\bar{p}_t^y(x,y)\nu(x,dy) = g(x).$$

For notational simplicity, we define for $t > 0$,

$$\begin{aligned}
\hat{p}_t(x,y) &:= \bar{p}_t^y(x,y), \\
\hat{P}_t f(x) &:= \int f(y)\hat{p}_t(x,y)\nu(x,dy) = \int f(y)\bar{p}_t^y(x,y)\nu(x,dy), \\
\mathcal{S}_t g(x) &:= \int g(y)\hat{\theta}_t(x,y)\nu(x,dy) = \int g(y)\bar{\theta}_t^y(x,y)\nu(x,dy).
\end{aligned}$$

Theorem 6.2.1. *Let $T > 0$. Assume that* **(H1)** *holds. Then, for any $g \in \mathcal{C}_b(\mathcal{J})$ and any $t \in [0,T]$, one has*

$$P_t g(x) = \hat{P}_t g(x) + \int_0^t P_s \mathcal{S}_{t-s} g(x)\, ds.$$

Moreover, if for $t > 0$, the function $x \mapsto \mathcal{S}_t g$, $g \in \mathcal{C}_b(\mathcal{J})$, belongs to $\mathcal{C}_b(\mathcal{J})$. Then one may iterate the first order formula

$$P_t g(x) = \hat{P}_t g(x) + \sum_{n \geqslant 1} I_t^n g(x), \ \text{ with } I_t^n g(x) = \int_{\Delta_n(t)} d\mathbf{s}_n \hat{P}_{s_n} \mathcal{S}_{s_{n-1}-s_n} \cdots \mathcal{S}_{t-s_1} g(x) \tag{6.12}$$

Moreover, the series (6.12) *converges absolutely and uniformly for $x \in \mathcal{J}$. The notation $\Delta_n(t)$ has been defined in the previous section, just after equation* (6.4).

We consider two weak solutions of the SDE (6.7) starting at time 0 from the same initial point $x \in \mathcal{J}$. Denote by $\mathbb{P}_1$ and $\mathbb{P}_2$ the two probability measures induced on the space $(\mathcal{C}([0,\infty),\mathcal{J}), \mathcal{B}(\mathcal{C}([0,\infty),\mathcal{J})))$. Define for $f \in \mathcal{B}_b(\mathcal{J})$, $\lambda > 0$ and $i = 1,2$

$$S_\lambda^i f(x) := \mathbb{E}_i[\int_0^\infty e^{-\lambda t} f(Y_t)dt] = \int_0^\infty e^{-\lambda t}\mathbb{E}_i[f(Y_t)]\, dt, \quad S_\lambda^\Delta f(x) := (S_\lambda^1 - S_\lambda^2)f(x), \quad \|S_\lambda^\Delta\| := \sup_{\|f\|_\infty \leqslant 1} |\mathcal{S}_\lambda^\Delta f|.$$

We notice that by **(H1)** (ii) b), $S_\lambda^i(\lambda - \mathcal{L})f(x) = f(x)$, $\forall f \in \mathcal{D}$, $i = 1,2$. For $z \in \mathcal{J}$, the resolvent of the process with frozen coefficients is defined by

$$\bar{R}_\lambda f(x) = \int_0^\infty e^{-\lambda t}\bar{P}_t f(x)dt, \qquad \forall f \in \mathcal{B}_b(\mathcal{J}) \tag{6.13}$$

and for $f \in \mathcal{D}$ one has $\bar{R}_\lambda(\lambda - \bar{\mathcal{L}})f = f$. We make the following assumptions:
Assumption (H2): For all $\lambda > 0$, one has $\bar{R}_\lambda \mathcal{D} \subset \mathcal{D}$ and for $f \in \mathcal{D}$,

$$(\lambda - \bar{\mathcal{L}})\bar{R}_\lambda f = \bar{R}_\lambda(\lambda - \bar{\mathcal{L}})f \quad \textit{and} \quad (\mathcal{L} - \bar{\mathcal{L}})\bar{R}_\lambda f(x) = \int_0^\infty e^{-\lambda t}(\mathcal{L} - \bar{\mathcal{L}})\bar{P}_t f(x)\, dt. \tag{6.14}$$

Theorem 6.2.2. *Assume* **(H1)** *and* **(H2)** *are satisfied, then weak uniqueness holds for the SDE* (6.7).

We sketch the proof and omit some technical details. Let $z \in \mathcal{J}$ and $r > 0$. We consider a sequence of non-negative mollifiers δ^z_ε, $\varepsilon > 0$, converging to the Dirac mass at z as $\varepsilon \downarrow 0$. Let us first observe that if $g \in \mathcal{C}^\infty_b(\mathcal{J})$ then $\bar{P}_r g \in \mathcal{D}$, and by **(H2)** one has

$$(\lambda - \mathcal{L})\bar{R}_\lambda \bar{P}_r \delta^z_\varepsilon g(x) = (\lambda - \bar{\mathcal{L}})\bar{R}_\lambda \bar{P}_r \delta^z_\varepsilon g(x) - (\mathcal{L} - \bar{\mathcal{L}})\bar{R}_\lambda \bar{P}_r \delta^z_\varepsilon g(x) = \bar{P}_r \delta^z_\varepsilon g(x) - (\mathcal{L} - \bar{\mathcal{L}})\bar{R}_\lambda \bar{P}_r \delta^z_\varepsilon g(x). \tag{6.15}$$

We integrate both sides of (6.15) with respect to dz and apply S^Δ_λ. For $i = 1, 2$, applying Fubini's theorem, one gets

$$S^i_\lambda \int (\lambda - \mathcal{L})\bar{R}^z_\lambda \bar{P}^z_r \delta^z_\varepsilon g \, dz = \int_0^\infty e^{-\lambda t} \mathbb{E}_i[\int (\lambda - \mathcal{L})\bar{R}^z_\lambda \bar{P}^z_r \delta^z_\varepsilon g(Y_t) dz] dt = \int S^i_\lambda (\lambda - \mathcal{L})\bar{R}^z_\lambda \bar{P}^z_r \delta^z_\varepsilon g dz$$

so that

$$S^\Delta_\lambda (\int \bar{P}^z_r \delta^z_\varepsilon g \, dz) - S^\Delta_\lambda (\int (\mathcal{L} - \bar{\mathcal{L}}^z)\bar{R}^z_\lambda \bar{P}^z_r \delta^z_\varepsilon g \, dz) = 0. \tag{6.16}$$

Passing to the limit as $\varepsilon \downarrow 0$ and $r \downarrow 0$, one gets $S^\Delta_\lambda g = S^\Delta_\lambda (\int_0^\infty e^{-\lambda t} \mathcal{S}_t g \, dt)$ and one can pick λ such that $|\int_0^\infty e^{-\lambda t} \mathcal{S}_t g dt| \leqslant |g|_\infty \int_0^\infty e^{-\lambda t} t^\zeta dt = |g|_\infty \frac{\Gamma(\zeta)}{\lambda^{1+\zeta}} < \frac{1}{2}|g|_\infty$. Hence, from the definition of $\|S^\Delta_\lambda\|$, we find that $|S^\Delta_\lambda g| = |S^\Delta_\lambda (\int_0^\infty e^{-\lambda t} \mathcal{S}_t g \, dt)| \leqslant \frac{1}{2}\|S^\Delta_\lambda\| |g|_\infty$. Taking the supremum over $|g|_\infty \leqslant 1$ yields $\|S^\Delta_\lambda\| \leqslant \frac{1}{2}\|S^\Delta_\lambda\|$ and, since $\|S^\Delta_\lambda\| < \infty$, we conclude that $S^\Delta_\lambda = 0$. By the uniqueness of the Laplace transform together with continuity w.r.t the variable t, $\mathbb{E}_1[g(Y_t)] = \mathbb{E}_2[g(Y_t)]$ for all $t \geqslant 0$ if g is bounded measurable. Now one can use the standard argument based on regular conditional probabilities to show that the finite dimensional distributions of the process $(Y_t)_{t \geqslant 0} = (X_t, A_t(X))_{t \geqslant 0}$ agree under $\mathbb{P}_1$ and $\mathbb{P}_2$.

In **[14]**, we considered the two following examples: $A_t(X) = L^0_t(X)$ where $L^0_t(X)$ is the symmetric local time at 0 accumulated by X at time t with state space $\mathcal{J} = \mathbb{R} \times \mathbb{R}_+$ and $A_t(X) = M_t := m_0 \vee \max_{0 \leqslant s \leqslant t} X_s$, $m_0 \geqslant x_0$, with state space $\mathcal{J} = \big\{(x, m) \in \mathbb{R}^2 : x \leqslant m\big\}$. We work under the following mild assumptions:

(R-η) The coefficients b and $a = \sigma^2$ are bounded measurable functions defined on $\mathcal{J}$. The diffusion coefficient a is η-Hölder continuous on $\mathcal{J}$.

(UE) There exists some constant $\underline{a} > 0$ such that $\forall z \in \mathcal{J}$, $\underline{a} \leqslant a(z)$.

In the following we detail the computations for the first case only, namely the diffusion process with its symmetric local time at zero. A similar analysis for the SDE with its running maximum is done in **[14]**. The set $\mathcal{D}$ is the class of function $f \in \mathcal{C}^{2,1}_b(\mathbb{R}\backslash\{0\} \times \mathbb{R}_+) \cap \mathcal{C}_b(\mathbb{R} \times \mathbb{R}_+)$ such that $\partial_1 f(0+, \ell) = \lim_{x \downarrow 0} \frac{f(x,\ell) - f(0,\ell)}{x}$ and $\partial_1 f(0-, \ell) = \lim_{x \uparrow 0} \frac{f(x,\ell) - f(0,\ell)}{x}$ exist, are finite and satisfy the following transmission condition:

$$\forall \ell \in \mathbb{R}_+, \quad \frac{\partial_1 f(0+, \ell) - \partial_1 f(0-, \ell)}{2} + \partial_2 f(0, \ell) = 0. \tag{6.17}$$

We define the linear operator $\mathcal{L}$ by $\mathcal{L}f(x, \ell) = b(x, \ell)\partial_1 f(x-, \ell) + \frac{1}{2}a(x, \ell)\partial^2_1 f(x-, \ell)$, $(x, \ell) \in \mathbb{R} \times \mathbb{R}_+$. As we need a chain rule formula for the process $(X_t, A_t(X))_{t \geqslant 0}$, we rely on the following generalisation of Itô's Lemma:

$$\begin{aligned} f(t, X_t, L^0_t(X)) = f(0, x, 0) &+ \int_0^t \big\{\partial_1 f(s, X_s, L^0_s(X)) + \mathcal{L}f(s, .)(X_s, L^0_s(X))\big\} ds \\ &+ \int_0^t \left\{ \frac{\partial_2 f(s, 0+, L^0_s(X)) - \partial_2 f(s, 0-, L^0_s(X))}{2} + \partial_3 f(s, 0, L^0_s(X)) \right\} dL^0_s(X) \\ &+ \int_0^t \sigma(X_s, L^0_s(X))\partial_2 f(s, X_s-, L^0_s(X)) dW_s \quad a.s. \end{aligned}$$

for $f \in \mathcal{C}^{1,2,1}(\mathbb{R}_+ \times \mathbb{R}\backslash\{0\} \times \mathbb{R}_+) \cap \mathcal{C}(\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+)$ such that $\partial_2 f(t, 0+, \ell) = \lim_{x \downarrow 0}(f(t, x, \ell) - f(t, 0, \ell))/x$ and $\partial_2 f(t, 0-, \ell) = \lim_{x \uparrow 0}(f(t, x, \ell) - f(t, 0, \ell))/x$ exist and are finite. This chain rule formula allows one to associate a martingale problem to the SDE (6.7) with $A_t(X) = L^0_t(X)$. This connection provides a natural way to establish weak existence by a compactness argument when b and σ are continuous, see e.g. Stroock and Varadhan [120]. Remark that one can remove the continuity assumption on b when σ is uniformly elliptic via the Girsanov transform method.

We now introduce the proxy process $\bar{X}_t := x_0 + \sigma(z_1)W_t$, $t \geqslant 0$, $z_1 = (x_1, \ell_1) \in \mathbb{R} \times \mathbb{R}_+$. For $f \in \mathcal{C}_b(\mathbb{R} \times \mathbb{R}_+)$, simple computations show that $\bar{P}_t f(x_0, \ell_0) = \mathbb{E}[f(\bar{X}_t, \ell_0 + L^0_t(\bar{X}))] = \int_{\mathbb{R} \times \mathbb{R}_+} f(x, \ell)\bar{p}_t(x_0, \ell_0, x, \ell)\nu(x_0, \ell_0, dx, d\ell)$ with $\bar{p}_t(x_0, \ell_0, x, \ell) := \bar{f}_t(x_0, x)\mathbf{I}_{\{\ell = \ell_0\}} + \bar{q}_t(x_0, \ell_0, x, \ell)\mathbf{I}_{\{\ell_0 < \ell\}}$ and

$$\begin{aligned} \bar{f}_t(x_0, x) &:= g(a(z_1)t, x - x_0) - g(a(z_1)t, x + x_0), \\ \bar{q}_t(x_0, \ell_0, x, \ell) &:= \frac{(|x| + |x_0| + \ell - \ell_0)}{a^{\frac{3}{2}}(z_1)t} g(a(z_1)t, |x| + |x_0| + \ell - \ell_0), \\ \nu(x_0, \ell_0, dx, d\ell) &:= \mathbf{I}_{\{\ell_0 < \ell\}} dx d\ell + \mathbf{I}_{\{x_0 x \geqslant 0\}} dx \delta_{\ell_0}(d\ell). \end{aligned}$$

and $\hat{f}_t(x_0,x) := g(a(x,\ell_0)t, x-x_0) - g(a(x,\ell_0)t, x+x_0)$, $\hat{q}_t(x_0,\ell_0,x,\ell) := \frac{(|x|+|x_0|+\ell-\ell_0)}{a^{\frac{3}{2}}(x,\ell))t} g(a(x,\ell)t, |x|+|x_0|+\ell-\ell_0)$.

We also observe that the measure $\nu(x_0,\ell_0,dx,d\ell)$ satisfies an important convolution type property in the sense that

$$\nu(x_0,\ell_0,dx',d\ell')\nu(x',\ell',dx,d\ell) = u(x_0,\ell_0,x,\ell,dx',d\ell')\nu(x_0,\ell_0,dx,d\ell) \tag{6.18}$$

As an application of the general methodology developed above, we obtain the following result.

Theorem 6.2.3. *If* **(R-η)** *and* **(UE)** *are satisfied for some $\eta \in (0,1]$ then weak uniqueness holds for the SDE* (6.7) *with $A_t(X) = L_t^0(X)$, $t \geqslant 0$.*

Moreover, the first step expansion obtained in Theorem 6.2.1 holds. However, since $(x_0,\ell_0) \mapsto \partial_{x_0}\bar{q}_t(x_0,\ell_0,x,\ell)$ is not continuous at zero, the function $(x_0,\ell_0) \mapsto \mathcal{S}_t g(x_0,\ell_0)$ shares the same property unless the drift b satisfies $b(0,\ell) = 0$, $\ell \in \mathbb{R}_+$. Unfortunately, one cannot iterate the first step expansion. We first remove the drift. Now the function $S_t g$ is continuous. Applying repeatedly (6.18) and Fubini's theorem, one gets

$$\int_{\Delta_n(T)} d\mathbf{s}_n \hat{P}_{s_n} \mathcal{S}_{s_{n-1}-s_n} \dots \mathcal{S}_{t-s_1} g(x_0,\ell_0) = \int_{\mathbb{R}\times\mathbb{R}_+} g(x,\ell) p_t^n(x_0,\ell_0,x,\ell)\nu(x_0,\ell_0,dx,d\ell)$$

with

$$p_t^n(x_0,\ell_0,x,\ell) := \begin{cases} \int_{\Delta_n(t)} d\mathbf{s}_n \int_{(\mathbb{R}\times\mathbb{R}_+)^n} \hat{p}_{s_n}(x_0,\ell_0,x_1,\ell_1)\times \\ \left[\prod_{i=1}^n \hat{\theta}_{s_{n-i}-s_{n-i+1}}(x_i,\ell_i,x_{i+1},\ell_{i+1}) u(x_0,\ell_0,x_{i+1},\ell_{i+1},dx_i,d\ell_i)\right] & n \geqslant 1, \\ \hat{p}_t(x_0,\ell_0,x,\ell) & n = 0 \end{cases} \tag{6.19}$$

where we omit the exact expression of $\hat{\theta}_t(x_0,\ell_0,x,\ell)$.

As our second main result, we prove that the transition density of $(X_t, \ell_0 + L_t^0(X))_{t\geqslant 0}$ exists and satisfies a Gaussian upper bound.

Theorem 6.2.4. *Let $T > 0$. Assume that* **(R-η)** *and* **(UE)** *hold for some $\eta \in (0,1]$. For $(t,x_0,\ell_0) \in (0,T]\times\mathbb{R}\times\mathbb{R}_+$, define the probability measure*

$$p_t(x_0,\ell_0,dx,d\ell) := p_t(x_0,\ell_0,x,\ell)\nu(x_0,\ell_0,dx,d\ell) = p_t^1(x_0,\ell_0,x,\ell)\,\mathbf{I}_{\{\ell_0<\ell\}} dx d\ell + p_t^2(x_0,\ell_0,x)\,\mathbf{I}_{\{xx_0\geqslant 0\}} dx \delta_{\ell_0}(d\ell)$$

with $p_t(x_0,\ell_0,x,\ell) := \sum_{n\geqslant 0} p_t^n(x_0,\ell_0,x,\ell)$, $p_t^1(x_0,\ell_0,x,\ell) := \sum_{n\geqslant 0} p_t^n(x_0,\ell_0,x,\ell)$, $\quad p_t^2(x_0,\ell_0,x) := \sum_{n\geqslant 0} p_t^n(x_0,\ell_0,x,\ell_0)$.

Then, both series defining $p_t^1(x_0,\ell_0,x,\ell)$ and $p_t^2(x_0,\ell_0,x)$ converge absolutely and uniformly for (x_0,ℓ_0), $(x,\ell) \in (\mathbb{R}\times\mathbb{R}_+)^2$. Moreover for $h \in \mathcal{C}_b(\mathbb{R}\times\mathbb{R}_+)$ the following representation for the semigroup holds,

$$P_t h(x_0,\ell_0) = \int_{\mathbb{R}\times\mathbb{R}_+} h(x,\ell)\, p_t(x_0,\ell_0,x,\ell)\nu(x_0,\ell_0,dx,d\ell).$$

Therefore, for all $(x_0,\ell_0) \in \mathbb{R}\times\mathbb{R}_+$, the function $(x,\ell) \mapsto p_t(x_0,\ell_0,x,\ell)$ is the probability density function of the random vector $(X_t^{x_0}, \ell_0 + L_t^0(X^{x_0}))$ with respect to the σ-finite measure $\nu(x_0,\ell_0,dx,d\ell)$, where $X_t^{x_0}$ is the solution taken at time t of the SDE (6.7) *(with $A_t(X) = L_t^0(X)$) starting from x_0 at time 0, $L_t^0(X^{x_0})$ being its running symmetric local time at time t.*

Finally, there exists some constants $C, c > 1$ such that for all (x_0,ℓ_0), $(x,\ell) \in \mathbb{R}\times\mathbb{R}_+$, the following Gaussian upper-bounds hold

$$p_t^1(x_0,\ell_0,x,\ell) \leqslant Ct^{-1/2} g(ct, |x|+|x_0|+\ell-\ell_0) \quad \text{and} \quad p_t^2(x_0,\ell_0,x) \leqslant C g(ct, x-x_0). \tag{6.20}$$

The proof is done in several steps. First, we assume that $b = 0$ and obtain the above series expansion. Then, we prove the existence of a transition density in the presence of a bounded measurable drift via Girsanov transform. This allows us to extend Theorem 6.2.1 for a non-continuous function g. Finally we repeat the same argument as above.

We point out that the proof of the convergence of the asymptotic expansion for the transition density is not standard in the current setting. Indeed, in the classical diffusion setting, the parametrix expansion of the transition density converges since the order of the singularity in time induced by the parametrix kernel $\hat{\theta}_t$ is of order $t^{-1+\frac{\eta}{2}}$, which is still integrable near 0. The main difficulty that appears here is when one wants to control the whole convolution appearing in the right-hand side of (6.19). More precisely, it lies in the cross-terms which are of a different nature, for instance when one convolutes the non-singular part in the convolution kernel $\hat{\theta}_{T-s_1}$ with the singular part in the convolution kernel $\hat{\theta}_{s_1-s_2}$. Standard arguments similar to those used in the standard diffusion case do not guarantee the convergence of the integral defining (6.19). To overcome this difficulty and show that the parametrix expansion for the transition density converges, one has to make use of the symmetry in the initial and terminal point of the density of the killed proxy process, in order to retrieve the integrability in time of the underlying convolution kernel.

6.3 Perspective

In **[13]**, we considered probabilistic representations for the first hitting time and the associated killed process of a one dimensional elliptic diffusion. We obtain some integration by parts for the killed diffusion process for the first derivative of a test function. One may be interested in establishing higher order integration by parts and also integration by parts formulas for the first hitting time. This is not a simple extension of Corollary (6.1.2) where we performed an integration by parts in the last interval $[\zeta_N, T]$. If we want to keep the Malliavin weights in $L^2(\mathbb{P})$, one has to perform the integration by parts on the whole time interval $[0, T]$ using the whole underlying noise.

One may be interested in establishing a probabilistic representation as well as integration by parts formulas for other path functionals such as the local time, the running maximum, the occupation time, etc as considered in Section 6.2. Using the parametrix expansions of the transition densities of $(X_t, A_t(X))_{t\geqslant 0}$, one can prove that the corresponding semigroup $P_t h$ is smooth when a and b are Hölder continuous. Then, one may proceed using the forward parametrix method as presented in Section 6.1.

More challenging extensions could include other type of processes. One may notably consider the case of a skew diffusion with path-dependent coefficients involving its local and occupation times, see Appuhamillage & al. [5] for an expression of the trivariate density $(B_t^{(\alpha)}, L_t^0(B^{(\alpha)}), \Gamma_t^0(B^{(\alpha)}))$, $t \geqslant 0$, where $(B_t^{(\alpha)})_{t\geqslant 0}$ is an α-skew Brownian motion or reflected SDEs. We are working on this topic with L. Li.

Another interesting point to investigate in that direction is to weaken the assumption on the coefficients, notably on a by assuming only uniform ellipticity and continuity, in order to establish weak uniqueness for the martingale problem using Calderón-Zygmund estimates, see Menozzi [99] for some degenerate equations of Kolmogorov type. Another related issue concerns pathwise uniqueness for the SDE (6.7). A possible strategy could be to extend the results of Zvonkin [128] and Veretennikov [125], Krylov and Röckner [79] established in the diffusion case to the case of a diffusion with a path-functional.

On may study a class of Hörmander equations with non-regular coefficients. The structure of the irregularity has to be carefully chosen so that the approximation process satisfies a *uniform Hörmander* condition or even a *UFG* condition. We refer to the work of Kusuoka and Stroock [84] for a careful analysis of the regularity properties of the unique solution to the related linear parabolic PDE by probabilistic tools. We also refer to the work of Crisan and Delarue [32] for an extension to degenerate semi-linear PDE. For instance, one may consider an equation of the form: $V_0 + (1/2)\sum_{i=1}^N V_i^2$, $V_0, \cdots, V_N$ being first order differential operators and the drift term V_0 being irregular. One simple example is $V_0 + \sum_{i=1}^N \varphi_i V_i + \sum_{\alpha\in\mathcal{A}_0(2)} \varphi_\alpha V_{[\alpha]}$, where $V_0 \in \mathcal{C}_\infty^\infty(\mathbb{R}^d)$ and $(\varphi_i)_{1\leqslant i\leqslant N}$ are bounded measurable functions, $(\varphi_\alpha)_{\alpha\in\mathcal{A}_0(2)}$ are $\mathcal{C}_b^1(\mathbb{R}^d)$. Questions of interests are: pathwise uniqueness, density estimates, regularity of the solution of the PDE, etc.

A challenging problem is to investigate probabilistic representation such as the one obtained in Section 6.1 for McKean-Vlasov equations. These equations exhibit a dependence w.r.t to the law of the process in its coefficients. Unlike the two examples investigated in **[14]**, the law of the underlying process is a deterministic object so that it is not natural to expect a regularising property w.r.t. this variable. We refer to the P.h.d. of Murray [96] and the recent work of Crisan and Murray [33] for some integration by parts formula and density estimates for such equations.

Bibliography

[1] M. Ben Alaya et A. Kebaier : Central limit theorem for the multilevel monte carlo euler method and applications to asian options. *Ann. Appl. Probab.*, 2012.

[2] A. Alfonsi, B. Jourdain et A. Kohatsu-Higa : Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme. *Ann. Appl. Probab.*, 24(3):1049–1080, 2014.

[3] A. Alfonsi, B. Jourdain et A. Kohatsu-Higa : Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme. *Electron. J. Probab.*, 20:no. 70, 31, 2015.

[4] P. Andersson et A. Kohatsu-Higa : Unbiased simulation of stochastic differential equations using parametrix expansions. *Bernoulli*, 23(3):2028–2057, 08 2017.

[5] T. Appuhamillage, V. Bokil, E. Thomann, E. Waymire et B. Wood : Occupation and local times for skew brownian motion with applications to dispersion across an interface. *Ann. Appl. Probab.*, 21(1):183–214, 02 2011.

[6] R. Azencott : Densité des diffusions en temps petit: développements asymptotiques. I. *In Seminar on probability, XVIII*, volume 1059 de *Lecture Notes in Math.*, pages 402–498. Springer, Berlin, 1984.

[7] V. Bally et Arturo Kohatsu-Higa : A probabilistic interpretation of the parametrix method. *Ann. Appl. Probab.*, 25(6):3095–3138, 12 2015.

[8] V. Bally et G. Pagès : Error analysis of the optimal quantization algorithm for obstacle problems. *Stochastic Process. Appl.*, 106(1):1–40, 2003.

[9] V. Bally et C. Rey : Approximation of markov semigroups in total variation distance. *Electron. J. Probab.*, 21:44 pp., 2016.

[10] V. Bally et D. Talay : The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Prob. Th. Rel. Fields*, 104-1:43–60, 1996.

[11] V. Bally et D. Talay : The law of the Euler scheme for stochastic differential equations, II. Convergence rate of the density. *Monte-Carlo methods and Appl.*, 2:93–128, 1996.

[12] F. Barthe et C. Bordenave : Combinatorial optimization over two random point sets. *To appear in Séminaire de Probabilités*, 2011.

[13] E. Barucci, S. Polidoro et V. Vespri : Some results on partial differential equations and Asian options. *Math. Models Methods Appl. Sci.*, 11(3):475–497, 2001.

[14] R. F. Bass : *Diffusions and Elliptic Operators*. Springer, 1997.

[15] R.F. Bass et E.A. Perkins : A new technique for proving uniqueness for martingale problems. *From Probability to Geometry (I): Volume in Honor of the 60th Birthday of Jean-Michel Bismut*, pages 47–53, 2009.

[16] G. Ben Arous : Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus. *Ann. Sci. École Norm. Sup. (4)*, 21(3):307–331, 1988.

[17] G. Ben Arous : Methodes de laplace et de la phase stationnaire sur l'espace de wiener. *Stochastics*, 25(3):125–153, 1988.

[18] F.E. Benth, J.S. Benth et S. Koekebakker : *Stochastic Modelling of Electricity and Related Markets*. Advanced series on statistical science & applied probability. World Scientific, 2008.

[19] A. Benveniste, M. Métivier et P. Priouret : *Adaptive algorithms and stochastic approximations*, volume 22 de *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.

[20] A. Beskos, O Papaspiliopoulos et Roberts G. O. : Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098, 2006.

[21] J.-M. Bismut : *Large deviations and the Malliavin calculus*, volume 45 de *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1984.

[22] E. Boissard : Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance. *Electron. J. of Probab.*, 16, 2011.

[23] F. Bolley, A. Guillin et C. Villani : Quantitative concentration inequalities for empirical measures on non-compact spaces. *Prob. Th. Rel. Fields*, 137:541–593, 2007.

[24] F. Bolley et C. Villani : Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse (6)*, 14(3):331–352, 2005.

[25] O. Brandière et M. Duflo : Les algorithmes stochastiques contournent-ils les pièges? *Annales de l'I.H.P.*, 32(3):395–427, 1996.

[26] M. Broadie et O Kaya : Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231, 2006.

[27] N. Bruti-Liberati et E. Platen : Strong approximations of stochastic differential equations with jumps. *Journal of Computational and Applied Mathematics*, 205(2):982 – 1001, 2007. Special issue on evolutionary problems.

[28] P. Cattiaux : Calcul stochastique et opérateurs dégénérés du second ordre. II. Problème de Dirichlet. *Bull. Sci. Math.*, 115(1):81–122, 1991.

[29] P. E. Chaudru de Raynal : Strong existence and uniqueness for degenerate SDE with Hölder drift. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(1):259–286, 2017.

[30] B. CHEN, C. W. OOSTERLEE et H. VAN DER WEIDE : A low-bias simulation scheme for the sabr stochastic volatility model. *International Journal of Theoretical and Applied Finance*, 15(02):1250016, 2012.

[31] J.-M. Coron : *Control and nonlinearity*, volume 136 de *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2007.

[32] D. Crisan et F. Delarue : Sharp gradient bounds for solutions of degenerate semi-linear partial differential equations. *Journal of functional analysis*, 263(10):3024–3101, 2012.

[33] D. Crisan et E. McMurray : Smoothing properties of mckean-vlasov sdes. *Probab. Theory Relat. Fields*, 2017.

[34] F. Delarue et S. Menozzi : Density estimates for a random noise propagating through a chain of differential equations. *J. Funct. Anal.*, 259(6):1577–1630, 2010.

[35] C. Dellacherie et P.-A. Meyer : *Probabilités et potentiel*. Hermann, Paris, 1975. Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.

[36] S. Dereich : Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.*, 21(1):283–311, 2011.

[37] J. D. Deuschel, P. K. Friz, A. Jacquier et S. Violante : Marginal density expansions for diffusions and stochastic volatility I: Theoretical foundations. *Comm. Pure Appl. Math.*, 67(1):40–82, 2014.

[38] D Duffie et P. Glynn : Efficient monte carlo simulation of security prices. *Ann. Appl. Probab.*, 5(4):897–905, 1995.

[39] M. Duflo : *Algorithmes stochastiques*, volume 23 de *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996.

[40] P. Étoré et M. Martinez : Exact simulation for solutions of one-dimensional stochastic differential equations with discontinuous drift. *ESAIM Probab. Stat.*, 18:686–702, 2014.

[41] H. Föllmer et A. Schied : Convex measures of risk and trading constraints. *Finance Stoch.*, 6(4):429–447, 2002.

[42] M. Forde : A diffusion-type process with a given joint law for the terminal level and supremum at an independent exponential time. *Stoch. Process. Appli.*, 121(12):2802 – 2817, 2011.

[43] G. Fort : Central limit theorems for stochastic approximation with controlled markov chain dynamics. *ESAIM: PS*, 19:60–80, 2015.

[44] A. Friedman : *Partial differential equations of parabolic type.* Prentice-Hall, 1964.

[45] A. Friedman : *Stochastic differential equations.* Chapmann-Hall, 1975.

[46] M. G. Garroni et J.-L. Menaldi : *Green functions for second order parabolic integro-differential problems*, volume 275 de *Pitman Research Notes in Mathematics Series.* Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1992.

[47] M. B. Giles : Improved multilevel Monte Carlo convergence using the Milstein scheme. *In Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

[48] M. B. Giles : Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[49] M. B. Giles, D. J. Higham et X. Mao : Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance Stoch.*, 13(3):403–413, 2009.

[50] E. Gobet : *Schémas d'Euler pour diffusion tuée. Application aux options barrière.* PhD Thesis, University Paris VII., 1998.

[51] E. Gobet : Euler schemes for the weak approximation of killed diffusion. *Stoch. Proc. Appl.*, 87:167–197, 2000.

[52] E. Gobet : Weak approximation of killed diffusion using Euler schemes. *Stoch. Process. Appli.*, 87(2):167–197, 2000.

[53] E. Gobet et C. Labart : Short time asymptotics of the density of the Euler scheme. *Electron. Comm. iProbab.*, 23, 2008.

[54] E. Gobet et S. Menozzi : Stopped diffusion processes: boundary corrections and overshoot. *Stoch. Process. Appli.*, 2009.

[55] N. Gozlan et C. Léonard : A large deviation approach to some transportation cost inequalities. *Prob. Th. Rel. Fields*, 139(1-2):235–283, 2007.

[56] N. Gozlan et C. Léonard : Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.

[57] S. Graf et H. Luschgy : *Foundations of Quantization for Probability Distributions.* Lect. Notes in Math. 1730, 2000.

[58] J. Guyon : Euler scheme and tempered distributions. *Stoch. Process. Appli.*, 116(6):877 – 904, 2006.

[59] M. Hairer, M Hutzenthaler et A. Jentzen : Loss of regularity for kolmogorov equations. *Ann. Probab.*, 43(2):468–527, 2015.

[60] J. M. Harrison et L. A. Shepp : On skew Brownian motion. *Ann. Probab.*, 9(2):309–313, 1981.

[61] S. Heinrich : *Multilevel Monte Carlo methods.* Springer, 2001. In Large-scale scientific computing.

[62] P. Henry-Labordère, X. Tan et N. Touzi : A numerical algorithm for a class of BSDEs via the branching process. *Stochastic Process. Appl.*, 124(2):1112–1140, 2014.

[63] I Honoré, S. Menozzi et G. Pagès : Non-asymptotic gaussian estimates for the recursive approximation of the invariant measure of a diffusion. *Preprint*, 2016.

[64] N. Ikeda et S. Watanabe : *Stochastic differential equations.* North Holland, 1980.

[65] J. Jacod : Rates of convergence to the local time of a diffusion. *Ann. I.H.P., section B*, 34:505–544, 1998.

[66] J. Jacod, T. G. Kurtz, S. Méléard et P. Protter : The approximate euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist*, 41–3:523–558, 2005.

[67] J. Jacod et A.N. Shiryayev : Local martingales and the fundamental asset pricing theorems in the discrete-time case. *Finance and Stochastics*, 2:259–273, 1998.

[68] B. Jourdain et J. Lelong : Robust adaptive importance sampling for normal vectors. *Ann. App.. Probab.*, 19(5):1687–1718, 2009.

[69] M. Kaina et L. Rüschendorf : On convex risk measures on l^p-spaces. *Mathematical Methods in Operations Research (MMR)*, 69:475–495, 2009.

[70] A. Kebaier : Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4):2681–2705, 2005.

[71] P. E. Kloeden et E. Platen : *Numerical solution of stochastic differential equations*, volume 23 de *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[72] A. Kohatsu-Higa et P. Protter : The Euler scheme for SDE's driven by semimartingales. *In Stochastic analysis on infinite-dimensional spaces (Baton Rouge, LA, 1994)*, volume 310 de *Pitman Res. Notes Math. Ser.*, pages 141–151. Longman Sci. Tech., Harlow, 1994.

[73] A. Kohatsu-Higa, D. Taguchi et J. Zhong : The parametrix method for skew diffusions. *Potential Analysis*, pages 1–31, 2016.

[74] A. Kohatsu-Higa et P. Tankov : Jump-adapted discretization schemes for lévy-driven sdes. *Stoch. Process. Appli.*, 120:2258–2285, 2010.

[75] V. Konakov et E. Mammen : Local limit theorems for transition densities of Markov chains converging to diffusions. *Prob. Th. Rel. Fields*, 117:551–587, 2000.

[76] V. Konakov et E. Mammen : Edgeworth type expansions for euler schemes for stochastic differential equations. *Monte Carlo Methods Appl.*, 8–3:271–285, 2002.

[77] V. Konakov et S. Menozzi : Weak error for stable driven stochastic differential equations: Expansion of the densities. *Journal of Theoretical Probability*, 24(2):454–478, 2011.

[78] V. Konakov et S. Menozzi : Weak error for the Euler scheme approximation of diffusions with non-smooth coefficients. *Electron. J. Probab.*, 22:Paper No. 46, 47, 2017.

[79] N.V. Krylov et M. Röckner : Strong solutions of stochastic equations with singular time dependent drift. *Probab. Th. Rel. fields*, 131(2):154–196, 2005.

[80] H. Kunita : *Stochastic flows and stochastic differential equations*, volume 24 de *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[81] H. J. Kushner et G. G. Yin : *Stochastic approximation and recursive algorithms and applications*, volume 35 de *Applications of Mathematics (New York)*. Springer-Verlag, New York, second édition, 2003. Stochastic Modelling and Applied Probability.

[82] S. Kusuoka : *Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus*, pages 69–83. Springer Japan, Tokyo, 2004.

[83] S. Kusuoka et D. Stroock : Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math*, 32:1–76, 1985.

[84] S. Kusuoka et D. Stroock : Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math*, 34:391–442, 1987.

[85] O.A. Ladyzenskaja, V.A. Solonnikov et N.N. Ural'ceva : *Linear and quasi-linear equations of parabolic type*. Vol.23 Trans. Math. Monog., AMS, Providence, 1968.

[86] S. Laruelle : Analyse d'algorithmes stochastiques appliqués à la finance,. *Phd dissertation*, 2011.

[87] S. Laruelle et G. Pagès : Stochastic approximation with averaging innovation applied to finance. *Monte Carlo Methods and Applications*, 18(1):1–52, 2012.

[88] J.-F. Le Gall : One-dimensional stochastic differential equations involving the local times of the unknown process. *In Stochastic analysis and applications (Swansea, 1983)*, volume 1095 de *Lecture Notes in Math.*, pages 51–82. Springer, Berlin, 1984.

[89] A. Lejay : On the constructions of the skew Brownian motion. *Probab. Surv.*, 3:413–466, 2006.

[90] Lelong, Jérôme : Asymptotic normality of randomly truncated stochastic algorithms. *ESAIM: PS*, 17:105–119, 2013.

[91] V. Lemaire et S. Menozzi : On some non asymptotic bounds for the euler scheme. *Electron. J. Probab.*, 15:1645–1681, 2010.

[92] V. Lemaire et G. Pagès : Unconstrained recursive importance sampling. *Ann. Appl. Probab.*, 20(3):1029–1067, 06 2010.

[93] V. Lemaire et G. Pagès : Multilevel richardson–romberg extrapolation. *Bernoulli*, 23(4A):2643–2692, 11 2017.

[94] F. Malrieu et D. Talay : Concentration inequalities for euler schemes. *Monte Carlo and Quasi-Monte Carlo Methods 2004, H. Niederreiter and D. Talay (Eds.)*, pages 355–37, 2006.

[95] H. P. McKean et I. M. Singer : Curvature and the eigenvalues of the Laplacian. *J. Differential Geometry*, 1:43–69, 1967.

[96] E. McMurray : Regularity of mckean-vlasov stochastic differential equations and applications, phd dissertation, 2015.

[97] S. Menozzi : Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. *Electron. Commun. Probab.*, 16:234–250, 2011.

[98] S. Menozzi : Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. *Electron. Commun. Probab.*, 16:234–250, 2011.

[99] S. Menozzi : Martingale problems for some degenerate kolmogorov equations. *Preprint*, 2015.

[100] R. Mikulevicius : On the rate of convergence of simple and jump-adapted weak Euler schemes for Lévy driven SDEs. *Stochastic Process. Appl.*, 122(7):2730–2757, 2012.

[101] R. Mikulevičius et E. Platen : Rate of convergence of the Euler approximation for diffusion processes. *Math. Nachr.*, 151:233–239, 1991.

[102] R. Mikulevičius et C. Zhang : Weak Euler approximation for Itô diffusion and jump processes. *Stoch. Anal. Appl.*, 33(3):549–571, 2015.

[103] G. N. Mil'shtein : Approximate integration of stochastic differential equations. *Teor. Verojatnost. i Primenen.*, 19:583–588, 1974.

[104] G. N. Mil'shtein : A method of second-order accuracy integration of stochastic differential equations. *Theory of Probability & Its Applications*, 23(2):396–401, 1979.

[105] G. N. Mil'shtein : Weak approximation of solutions of systems of stochastic differential equations. *Theory of Probability & Its Applications*, 30(4):750–766, 1986.

[106] S. A. Molčanov : Diffusion processes, and Riemannian geometry. *Uspehi Mat. Nauk*, 30(1(181)):3–59, 1975.

[107] Syoiti Ninomiya et Nicolas Victoir : Weak approximation of stochastic differential equations and application to derivative pricing. *Applied Mathematical Finance*, 15(2):107–121, 2008.

[108] D. Nualart : *Malliavin calculus and related topics*. Springer Verlag, 1995.

[109] G. Pagès : Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods Appl.*, 13(1):37–70, 2007.

[110] G. Pagès : Introduction to numerical probability for finance. *handout*, page 354.p, 2016.

[111] G. Pagès, H. Pham et J. Printems : *Handbook of Numerical Methods in Finance*, chapitre Optimal quantization methods and applications to numerical problems in finance, pages 253–298. Birkhauser, 2004.

[112] G. Pagès et J. Printems : http://www.quantize.maths-fi.com, web site devoted to optimal quantization, 2005.

[113] M. Pelletier : Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, 8(1):10–44, 02 1998.

[114] R. Pemantle : Non convergence to unstable points in urn models and stochastic approximations. *Ann. Prob.*, 18(2):698–712, 1990.

[115] P. Protter et D. Talay : The Euler scheme for lévy driven stochastic differential equations. *Ann. Probab.*, 1(25):393–423, 1997.

[116] S. T. Rachev et L. Rüschendorf : *Mass transportation problems. Vol. II.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications.

[117] H. Robbins et S. Monro : A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.

[118] C. Soize : *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17 de *Series on Advances in Mathematics for Applied Sciences*. World Scientific Publishing Co., Inc., River Edge, NJ, 1994.

[119] D. W. Stroock et S. R. S. Varadhan : Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971.

[120] D. W. Stroock et S. R. S. Varadhan : *Multidimensional diffusion processes*, volume 233 de *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin-New York, 1979.

[121] D. Talay : Discrétisation d'une équation différentielle stochastique et calcul approché d'espérances de fonctionnelles de la solution. *RAIRO Modél. Math. Anal. Numér.*, 20(1):141–179, 1986.

[122] D. Talay et L. Tubaro : Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. and App.*, 8-4:94–120, 1990.

[123] D. Talay et Z. Zheng : Approximation of quantiles of components of diffusion processes. *Stoch. Process. Appli.*, 109-1:23–46, 2004.

[124] Talay, D. : Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002.

[125] A. Yu. Veretennikov : On the criteria for existence of a strong solution to a stochastic equation. *Theory of Probability & Its Applications*, 27:441–449, 1982.

[126] C. Villani : *Optimal transport: Old and New.* Springer, 2008.

[127] S. Watanabe : Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. *Ann. Probab.*, 15(1):1–39, 1987.

[128] A. K. Zvonkin : A transformation of the state space of a diffusion process that removes the drift. *Math. USSR (Sbornik)*, 22:129–149, 1974.