Discretization of processes at stopping times and Uncertainty quantification of stochastic approximation limits

Uladzislau Stazhynski

université PARIS-SACLAY

NNT : 2018SACLX088

Thèse de doctorat

Discrétisation de processus à des temps d'arrêt et Quantification d'incertitudes pour des limites d'approximation stochastique

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Ecole Polytechnique

Ecole doctorale n°574 Ecole doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 12 Décembre 2018, par

M. ULADZISLAU STAZHYNSKI

Composition du Jury :

Gilles Pagés Professeur, Sorbonne Université	Président de Jury
Philip Protter Professeur, Columbia University	Rapporteur
Fabien Panloup Professeur, Université d'Angers	Rapporteur
Josselin Garnier Professeur, Ecole Polytechnique	Examinateur
Mathieu Rosenbaum Professeur, Ecole Polytechnique	Examinateur
Gersende Fort Directeur de recherche, Institut de Mathématique de Toulouse	Examinateur
Emmanuel Gobet Professeur, Ecole Polytechnique	Directeur de thèse

Discretization of processes at stopping times and Uncertainty quantification for stochastic approximation limits

Uladzislau Stazhynski
PhD Thesis under the supervision of Emmanuel Gobet, Ecole Polytechnique

January 6, 2019

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Emmanuel Gobet. I met Emmanuel two years before the start of the PhD while preparing a master thesis. His advice and guidance had a great positive impact on my education and career path. I appreciated his deep understanding of how to conduct and present mathematical research, his excellent management skills as an advisor, his personal maturity and ability to build friendly and supportive relationships with students. Emmanuel also helped me a lot in all of the stages of my thesis preparation and I thank him for careful revision of the manuscript and valuable discussions. I learnt a lot while working with Emmanuel and I could not have imagined having a better advisor for my PhD.

I am grateful to Philip Protter and Fabien Panloup for accepting to be referees, for their time and their valuable feedback on my manuscript. Also I would like to thank all the members of the jury: Josselin Garnier, Mathieu Rosenbaum, Gilles Pagés and Gersende Fort.

I would like to thank ENS Ulm for providing financial support for my studies as an International Selection student and also for my PhD studies. I also thank Chaire de Risques Financiers for supporting financially my participation at conferences and summer schools, as well as my consulting project on risk management in cooperation with Société Générale.

I am grateful to all the stuff of Centre de Mathématiques Appliquées, in particular, to Vincent Chapellier, Wilfried Edmond, Nasséra Naar and Alexandra Noiret for helping with administrative procedures. Also I would like to thank all the services of Ecole Polytechnique including the Graduate School, the library and the Magnan restaurant.

During my studies I had a great opportunity to work with some of the outstanding researchers and industry professionals. I would to like to thank all of my colleagues and co-authors including Gersende Fort, Stéphane Crépey, David Barrera, Babacar Diallo, Nicolas Landon, Pierre-Emmanuel Levy Dit Vehel and Christophe Bernard. I am also grateful to many of the university professors for giving excellent courses from which I learnt a lot, including Raphaël Cerf, Jean-François Le Gall, Nicolas Curien and also all of the professors of the Paris 6 Probability and Finance master program.

Also I would like to thank all of my friends, especially Sergey, Ivan, Maxime, Alexandre and Ruslan for their genuine support.

Last but most important, I would like to thank my wife and my best friend Daria for her patience, trust and continued encouragement.

Contents

Introduction and thesis summary

Contents

This thesis studies two separate subjects: Discretization of processes at stopping times and Uncertainty quantification for stochastic approximation limits. The introduction to each of them is given separately in Sections 1 and 2. Parts I and II contains 4 and 2 chapters respectively based on the following papers (published or in revision):

(1) E. Gobet and U. Stazhynski. Optimal discretization of stochastic integrals driven by general Brownian semimartingale. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques, Vol. 54, No. 3, pp. 1556-1582,* (2018).

(2) E. Gobet and U. Stazhynski. Model-adaptive optimal discretization of stochastic integrals. In revision for *Stochastics,* (2018).

(3) E. Gobet and U. Stazhynski. Central limit theorem for discretization errors based on stopping time sampling. *Submitted preprint,* (2018).

(4) E. Gobet and U. Stazhynski. Parametric inference for diffusions observed at stopping times. *Submitted preprint,* (2018).

(5) S. Crepey, G. Fort, E. Gobet and U. Stazhynski. Uncertainty quantification for stochastic approximation limits using chaos expansion. In revision for *SIAM Journal of Uncertainty Quantification,* (2018).

(6) U. Stazhynski. Uncertainty quantification for stochastic approximation limits: L^2-convergence rate. *Preprint,* (2018).

1 Part I: Discretization of processes at stopping times

1.1 Introduction to discretization of processes

Discretization problems play a fundamental role in the applications of continuous-time stochastic processes. Indeed, since only discrete data can be observed, simulated and processed, discretizated versions of such processes are usually used in practice. In this regard, the quantification of the errors related to discretization is of great importance.

In this work we study discretization problems for a class of models called the Itô processes (see [RY99, p. 298] for the definition) and their various generalizations. An Itô process $(S_t)_{0\le t\le T}$ on a given filtered probability space has the form

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}B_s, \quad t \in [0,T]$$

where $(B_t)_{0\le t\le T}$ is a Brownian motion, $(b_t)_{0\le t\le T}$ and $(\sigma_t)_{0\le t\le T}$ are adapted processes verifying suitable assumptions (though a more general finite variation part may be considered, see Chapter 1). This class of models is widely used in many applications in finance, insurance, economy, biology, population dynamics, random mechanics and physics.

We consider the high frequency fixed horizon asymptotic framework for discretization problems. Namely, we assume that the time interval $[0,T]$ is fixed and for each $n \ge 0$ a finite discretization scheme $\mathcal{T}^n = \{0 = \tau_0^n < \cdots < \tau_{N_T^n}^n = T\}$ is given with the number of discretization times N_T^n (possibly random) going to infinity as $n \to +\infty$. The goal may be either to quantify or to optimize (in a suitable asymptotic sense) the error produced from the substitution of S_t by $S_{\varphi(t)}$, where $\varphi(t)$ is the largest discretization time τ_i^n before t, into a given procedure.

Here we discuss several applications where discretization problems naturally arise.

The first class of problems is related to statistics. Deploying continuous-time stochastic models in applications requires an estimation of various statistics or model parameters based on discrete observations of a single process trajectory. A standard example is the quadratic variation of a 1-dimensional Itô process S_t given by $\int_0^T \sigma_t^2 \mathrm{d}t$ for which the classical estimator has the form $\sum_{i=1}^{N_T^n} (S_{\tau_i^n} - S_{\tau_{i-1}^n})^2$. A more complicated setting is the parametric inference for diffusion processes, where the diffusion coefficient $\sigma_t = \sigma(t, S_t, \theta)$ depends on an unknown parameter θ to be estimated based on discrete observations of the process (see Sections 1.9-1.10). In these applications the estimation error is usually expressed in terms of the discretization error for the underlying process S. This means that the analysis of the estimators consistency and asymptotic normality boils down to the corresponding analysis of discretization errors.

The second class of problems studies the optimization of the tracking of a target whose dynamics is modeled by an Itô process. Here the goal is to optimally choose the discretization times for the rebalancing of a stochastic system in order to minimize certain criteria

expressing the deviation from the continuous target. Continuous rebalancing is typically not possible due to various costs of intervention and adjustment (so-called transaction costs). Examples of applications in finance include option delta-hedging (see [Fuk11a, GL14a]) and index tracking (see [PS04]) among others. In such problems the optimal disretization times usually depend on the process trajectory in an adaptive way and thus are given by random stopping times. A particular case considered in Chapters 1-2 is the quadratic variation minimization for stochastic integrals (see Section 1.4 for details).

Finally, another group of problems is related to the process simulation via discretization schemes and the subsequent analysis of the discretization error in Monte Carlo simulations, see e.g. [FO15] and references therein. However we do not consider this type of problems in our work and focus solely on the other two explained above.

1.2 Random discretization schemes

Discretization based on equidistant times, i.e. for $\tau_i^n = \frac{iT}{n}$, is a well studied subject, see e.g. [Roo80, JP98, HM05, GT09, GT01, MZ06] among others, see also [JP12] and references therein. However, in practice, the discretization times are quite often not regularly spaced. The nature of the irregularity itself may be quite different, depending on the setting. Moreover, the discretization times may be random as well, which makes the analysis even more complicated.

Concerning statistical estimation problems, possible reasons for random observation times may be that *i)* some data is missing; *ii)* the observations are more frequent during certain periods of time, or when the observed process is in certain regions of space; *iii)* the observations occur randomly, e.g. according to the arrival times of a Poisson-like process or stopping times related to the process itself, and others. Many works in this direction report a non-negligible impact of the randomness of the discretization times on the asymptotic properties of the errors compared to the classical deterministic case. For example, [ASM03] observes considerable effect of random sampling on the estimators in the setting of parametric inference for diffusions. In [LZZ13] the authors note that taking into account the endogenous randomness of the observation grids, when it exists, may substantially improve the performance of the integrated volatility estimator.

In the problem of the discretization times optimization for optimal tracking, random discretization grids appear naturally as optimal rebalancing times, and thus play the key role in the analysis, see e.g. [Fuk11a, GL14a].

The importance of random discretization schemes in high frequency finance was, in particular, emphasized in [DGM$^+$01, Section 1.1] and [ASJ14, Chapter 9], see also [Fuk10, FR12, RR10, RR12].

We distinguish the following two levels of generality when considering irregular discretization schemes:

1. For all i the time τ_i^n depends only on $\mathcal{F}_{\tau_{i-1}^n}$ (where $(\mathcal{F}_t)_{0\le t\le T}$ is some fixed filtration) and some extra independent noise. This group contains, in particular, all deterministic, *strongly predictable* (i.e. τ_i^n is $\mathcal{F}_{\tau_{i-1}^n}$-measurable with no independent noise) and random independent times.

2. More general stopping times with respect to a given filtration. This setting presumes that endogenous random noise may trigger discretization times. A benchmark example is the discretization of the process at its own exit times from some specified domains.

While the first setting, including strongly predictable and random independent times, is better studied, the second one is more difficult for the analysis and constitutes the primary focus of this work. Random discretization schemes given by exit times appear naturally in the problem of the tracking error optimization (see [Fuk11b, Fuk11a, GL14a]). Chapters 1-2 are devoted to this problem in the context of optimal discretization of stochastic integrals (detailed discussion is given in Section 1.4).

The availability of the data only at stopping times may be an intrinsic property of a model that aims to explain certain observations that are irregularly spaced in time. Quite recently a number of papers appeared in this direction. In [RR10, RR12] the authors construct a financial high-frequency price model which combines microstructure noise, including rounding noise, and sampling at transaction times on the basis of suitably defined hitting times, and then estimate the integrated volatility. They also provide the asymptotic analysis of their estimator. An even more complicated setting is when the observations of different components of a multidimensional process are random and in addition not synchronized. This is a typical situation in some financial applications (see e.g. [HY08]). As additional motivation for stopping time discretization grids we also refer to [GW02] for empirical evidence about the connection of volatility and inter-transaction duration in finance, and [Fuk10] for modeling bid or ask quotation data and tick time sampling.

In Chapters 1-4 our goal is to extend the current research on the processes discretization based on random stopping times, regarding both the applications in statistics and in tracking error optimization. The rest of this section contains more detailed introductions to each of the problems considered with a literature review, and also a summary of the results of each chapter. In Section 1.3 we introduce the class of random grid sequences under study. Techniques related to this class of grids are the main driving force of the proofs in Chapters 1-4. In Section 1.4 we continue with the introduction to the problem of the quadratic variation minimization for stochastic integrals based on general Brownian semimartingales. Our contribution to this problem is summarized in Sections 1.5-1.6. In Section 1.7 we discuss the background results on Central Limit Theorems (CLTs) for discretization errors. This is followed by a summary of our work on the CLT for discretization errors based on random stopping time grids. Finally, in Section 1.9 we discuss the problem of parametric inference for diffusion processes. We conclude with the presentation of our work on the parametric estimation of diffusions based on observations at stopping times in Section 1.10.

1.3 A class of random discretization grids

In this section we present the class of random discretization grids under study. It has been introduced in [GL14a] as the class of admissible grids for optimal discretization of a stochastic integral. The tools developed in [GL14a] play the key role in all the aspects of our analysis in Chapters 1-4. This class is essentially defined through the two assumptions below. For a process S, a sequence of discretization grids $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ with $\mathcal{T}^n := \{\tau_i^n : 0 \leq$

$i \leq N_T^n\}$, some positive deterministic sequence $(\varepsilon_n)_{n\geq 0}$, such that $\sum_n \varepsilon_n^2 < +\infty$, and $\rho_N \geq 1$ consider the following assumptions:

($\mathbf{A}_S^{\text{osc.}}$): The following non-negative random variable is a.s. finite:

$$\sup_{n\geq 0}\left(\varepsilon_n^{-2}\sup_{1\leq i\leq N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}|S_t - S_{\tau_{i-1}^n}|^2\right) < +\infty. \tag{1.1}$$

($\mathbf{A}_N$): For a given parameter $\rho_N \geq 1$ (verifying certain assumptions, in particular $\rho_N < 4/3$) the following non-negative random variable is a.s. finite:

$$\sup_{n\geq 0}(\varepsilon_n^{2\rho_N} N_T^n) < +\infty. \tag{1.2}$$

Assumption ($\mathbf{A}_S^{\text{osc.}}$) means that the oscillation of the process S between two successive times obeys a scaling rule; it implicitly implies that the time step related to successive times is small enough in some sense described by ε_n. On the other hand, ($\mathbf{A}_N$) states that the number of random times is not too large at some scale, refraining for instance an accumulation of stochastic times.

Now for arbitrary $\varepsilon_n \to 0$ we consider the class of discretization grid sequences $\{\mathcal{T}^n : n \geq 0\}$ such that for any subsequence $\iota(n)$ there exists another subsequence $\iota' \circ \iota(n)$ such that $\{\mathcal{T}^{\iota'\circ\iota(n)} : n \geq 0\}$ verifies ($\mathbf{A}_S^{\text{osc.}}$) and ($\mathbf{A}_N$) with $(\varepsilon_{\iota'\circ\iota(n)})_{n\geq 0}$. Such a definition is motivated by the subsequence principle that we later use to pass from a.s. convergences to the corresponding convergences in probability (see Lemma 2.2.2).

In particular, this class contains most of the discretization grids considered in the previous works and that we can imagine from application point of view. To emphasize its generality we present below several large families of random grids that it contains (for a justification see Remark 1.2.2 and the discussion in Section 3.2.2)

1. $\mathcal{T} = \{\mathcal{T}^n : n \geq 0\}$ where each $\mathcal{T}^n = \{\tau_i^n : 0 \leq i \leq N_T^n\}$ is a sequence of stopping times (with N_T^n possibly random) and such that

$$C^{-1}\varepsilon_n^{\frac{2}{(1-\rho)}} \leq \min_{1\leq i\leq N_T^n}\Delta\tau_i^n \leq \max_{1\leq i\leq N_T^n}\Delta\tau_i^n \leq C\varepsilon_n^{\frac{2}{(1-\rho)}}, \quad n \geq 0, \quad \text{a.s.},$$

   for an a.s. finite positive random variable $C > 0$ and a parameter $\rho > 0$. This example contains, in particular, the sequences of deterministic or strongly predictable discretization times for which the time steps are controlled from below and from above and for which the step size tends to zero.

2. Poisson random times with the random noise independent of $\mathcal{F}_T$ but with a stochastic $\mathcal{F}$-adapted intensity. More precisely for a continuous adapted positive process $(\lambda_t)_{0\leq t\leq T}$ we consider $\mathcal{T}^n = \{\tau_i^n : 1 \leq i \leq N_T^n\}$ given by the jump time of a Poisson process with intensity $(\varepsilon_n^{-2\rho_N}\lambda_t)_{0\leq t\leq T}$.

3. Consider a sequence of adapted random processes $\{D_t^n : 0 \leq t \leq T\}$ where each D_t^n is

an open set such that

$$B(0, C_1\varepsilon_n) \subset D_t^n \subset B(0, C_2\varepsilon_n)$$

for some a.s. finite positive random variables C_1, C_2, here $B(0,r)$ denotes the ball centered at 0 with radius r. Define the sequence of strategies $\mathcal{T} = \{\mathcal{T}^n : n \geq 0\}$ with $\mathcal{T}^n = \{\tau_i^n : 0 \leq i \leq N_T^n\}$ as follows: $\tau_0^n = 0$ and for $i \geq 1$

$$\tau_i^n = \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin D_{\tau_{i-1}^n}^n\} \wedge T.$$

In other words, we consider exit times of random sets of size ε_n (more complicated examples may be found in Section 3.2.2).

As we may see, the class of discretization grids under study is quite universal and contains practically all the types of discretization grids that may appear interesting in practice. Chapters 1-4, along with their principal contributions, develop powerful techniques for the analysis of such discretization grid sequences, that provide a solid background for treating future problems in the discretization of processes.

1.4 Optimal discretization of stochastic integrals

This is an introductory section to Chapters 1-2. We consider the problem of finding a finite sequence of optimal stopping times $\mathcal{T}^n = \{0 = \tau_0^n < \tau_1^n < \cdots < \tau_{N_T^n}^n = T\}$ which minimizes the renormalized quadratic variation of the discretization error for the stochastic integral given by

$$Z_t^n = \int_0^t v(s, S_s) \cdot \mathrm{d}S_s - \sum_{\tau_{i-1}^n < t} v(\tau_{i-1}^n, S_{\tau_{i-1}^n}) \cdot (S_{\tau_i^n \wedge t} - S_{\tau_{i-1}^n}), \tag{1.3}$$

where S is a d-dimensional continuous Brownian semimartingale and $v(t,x)$ is a $\mathbb{R}^d$-valued continuous function. Here $T > 0$ is fixed and the number of stopping times N_T^n is allowed to be random.

Under some mild assumptions on the model, and for deterministic or strongly predictable grids, the discretization error Z_T^n after suitable renormalization converges in distribution to a mixture of Gaussian random variables (see [Roo80, KP91, JP12]). A natural candidate for minimization criterion in this case is the product $N_T^n \langle Z^n \rangle_T$. In particular, in the case where a CLT holds for $\sqrt{N_T^n} Z_T^n$, the limit $\lim_n N_T^n \langle Z^n \rangle_T$ represents the asymptotic (conditional) variance of the limit distribution (see e.g. Chapter 3).

The study of minimization problems for stochastic integral discretization has been initiated by [Fuk11a] in dimension $d = 1$, but instead of $N_T^n \langle Z^n \rangle_T$ the author considers a criterion in expectation for both terms, i.e. $\mathbb{E}(N_T^n)\,\mathbb{E}(\langle Z^n \rangle_T)$.

The pathwise minimization of $\lim_n N_T^n \langle Z^n \rangle_T$ has been addressed in a multi-dimensional setting $d \geq 1$, in [GL14a]. They define the class of admissible discretization strategies as those verifying $(\mathbf{A}_S^{\text{osc.}})$-$(\mathbf{A}_N)$. For S a local martingale and under certains conditions of v (essentially its Jacobian matrix $D_x v$ is invertible) the authors exibit a lower bound on $\liminf_n N_T^n \langle Z^n \rangle_T$ across the class of admissible grid sequences. An exhaustive discussion of this problem in the setting of hedging in finance, as well as a review of the existing literature on the subject may be found in [Fuk11a, GL14a].

In [GL14a] the authors show that the discretization grids giving optimal (or arbitrarily close to optimal) performance have the form of exit times from random ellipsoids. Namely for an explicitly specified continuous adapted process $(\Lambda_t)_{0\le t\le T}$ taking values in the set of symmetric positive definite $d\times d$ matrices, an optimal sequence $\mathcal{T}:=\{(\tau_i^n)_{0\le i\le N_T^n}: n\ge 0\}$ of grids may be written as

$$\begin{cases}\tau_0^n:=0,\\ \tau_i^n:=\inf\{t>\tau_{i-1}^n:(S_t-S_{\tau_{i-1}^n})^{\mathsf{T}}\Lambda_{\tau_{i-1}^n}(S_t-S_{\tau_{i-1}^n})\ge\varepsilon_n^2\}\wedge T,\end{cases}$$

As proved in [GL14a], the optimal sequence $\mathcal{T}$ is admissible and attains the optimal value of $\lim_n N_T^n\langle Z^n\rangle_T$ among the whole class of admissible grids. These results is a starting point of our work in Chapters 1-2 which is presented in the next two sections.

1.5 Summary of results in Chapter 1

In Chapter 1 we consider the optimal discretization problem presented in Section 1.4 and prove optimality results in a much larger setting than previously afforded in the literature.

First, we allow S to be a general Brownian semimartingale $S=A+M$ with A a general finite variation adapted continuous process with some Holder properties, while in [GL14a] the process S is essentially a local Brownian martingale ($A=0, M=\int_0^{\cdot}\sigma_s\mathrm{d}B_s$). For this generalized model we prove the following important results:

- In Theorem 1.3.4 we show that the sets of admissible strategies (admissibility w.r.t. a process S is again defined as verifying $(\mathbf{A}_S^{\text{osc.}})$-$(\mathbf{A}_N)$) for the semimartingale S and for its local martingale part M are the same. The result is non-trivial and is proved via a continuation scheme with a subsequent application of the BDG inequality. Since [GL14a] assumed local martingale condition, Theorem 1.3.4 is of primary importance: it allows to apply the results, previously established in [GL14a], to our generalized setting.

- In Theorem 1.3.10 we show that the discretization strategy based on hitting times of random ellipsoids of the form

$$\begin{cases}\tau_0^n:=0,\\ \tau_i^n:=\inf\{t>\tau_{i-1}^n:(S_t-S_{\tau_{i-1}^n})^{\mathsf{T}}H_{\tau_{i-1}^n}(S_t-S_{\tau_{i-1}^n})\ge\varepsilon_n^2\}\wedge T,\end{cases}$$

  is admissible under suitable assumptions. This extends [GL14a, Proposition 2.4].

- Theorem 1.4.2 is one of the main results in Chapter 1 and provides a uniform lower bound (which is sharp, as we show later) for $\liminf_{n\to+\infty} N_T^n\langle Z^n\rangle_T$ over the entire class of admissible strategies, given by

$$\liminf_{n\to+\infty} N_T^n\langle Z^n\rangle_T\ge\left(\int_0^T\mathrm{Tr}(X_t)\mathrm{d}t\right)^2\qquad\text{a.s.}$$

for an explicitly defined process X_t. This is an important extension of [GL14a, Theorem 3.1] to the semimartingale case.

Second, the martingale part of S can be degenerate in our setting, whereas a stronger a.s. ellipticity of σ_t is considered in [GL14a]. Namely we do not require that the inverse σ_t^{-1} exists. Also $D_x v(t, S_t)$ may be not invertible in our work.

For this generalized model we prove the following important result: Theorem 1.5.2 shows that the strategy of the form

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\intercal \Lambda^{(n)}_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^2 \varepsilon_n\} \wedge T, \end{cases}$$

where $\Lambda_t^{(n)}$ is a suitable perturbation of $\Lambda_t := (\sigma_t^\dagger)^\intercal X_t \sigma_t^\dagger$ ($\mathcal{M}^\dagger$ denotes the pseudo-inverse matrix of $\mathcal{M}$), and X_t is explicitly given as a solution of certain non-linear matrix equation, attains the lowest possible value of $\lim_n N_T^n \langle Z^n \rangle_T$ across the entire class of admissible random grid sequences.

The proof of Theorem 1.5.2 is non-trivial, since due to possible degeneracy of σ_t we use pseudo-inverses and lose certain continuity properties of the optimal strategy. In addition, our strategy attain the exact optimal limit while in [GL14a] only a μ-optimal strategy has been established which is arbitrarily close to the optimum. In Section 1.5.3 we provide a numerical test that confirms the optimal performance of the strategy given by Theorem 1.5.2.

The ability to treat the non-elliptic case is fundamental for applications as well:

- First, it allows to consider partially degenerate models which arise in various applications such as random mechanics (see Subsection 1.5.3 for examples).
- Second, it provides a robustness result for the optimal strategy studied in [GL14a]: namely Theorem 1.5.2 shows that even if σ_t is close to being degenerate, this will not effect the performance of the optimal strategy. This is an important consideration in financial applications related to option hedging developed in [GL14a] (see the discussion in Section 1.5.3).

1.6 Summary of results in Chapter 2

In Chapter 2 we continue the study of the optimal discretization problem for stochastic integrals with respect to Brownian semimartingales started in Chapter 1. Our goal here is to construct an adaptive version of the optimal discretization algorithm from Chapter 1 that does not require any prior knowledge about the model.

In the previous works optimal sequences $\{\mathcal{T}^n : n \geq 0\}$ strongly depend on the model for S, in particular on the volatility σ-process. As a difference, in Chapter 2 we suppose that no prior knowledge about the diffusion coefficient of the underlying process S is given. We do not assume neither a diffusion model for S nor a parametric form for σ. The process S of the form $A + \int_0^\cdot \sigma_s \mathrm{d}B_s$ is quite arbitrary and we only suppose that it satisfies some mild regularity and non-degeneracy assumptions. A model-adaptive version of the optimal

discretization algorithm designed in Chapter 1 is needed in order to make the latter algorithm applicable. Another important question is the robustness of the optimal discretization with respect to the estimation error of σ. The optimal strategy may be given as

$$\tau_i^n = \varphi\left(\sigma_{\tau_{i-1}^n}, D_x v(\tau_{i-1}^n, S_{\tau_{i-1}^n}), (S_t - S_{\tau_{i-1}^n})_{t \geq \tau_{i-1}^n}\right),$$

where φ represents a quite complex non-linear dependence. The robustness analysis of this dependence requires substantial effort and is critical for applications.

In Chapter 2 we investigate this issue and prove the following important results:

- In Theorem 2.2.4 we state sufficient assumptions on a general sequence of estimators σ_t^n of σ_t ensuring the optimality of the resulting sequence of strategies. Namely let σ_t^n verify (for a parameter $\delta > 0$) the condition

$$\varepsilon_n^{-\delta/2} \sup_{0 \leq t \leq T} |(\sigma_t \sigma_t^\mathsf{T})^{1/2} - \sigma_t^n| \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

  Then for a suitable perturbation $[\Lambda_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta}$ of Λ_t used in Chapter 1 to construct the optimal strategy, the sequence $\{\mathcal{T}^n\}_{n \geq 0}$, where $\mathcal{T}^n = (\tau_i^n)_{0 \leq i \leq N_T^n}$ is given by

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} [\Lambda_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta} (S_t - S_{\tau_{i-1}^n}) \geqslant \varepsilon_n^{2+\delta}\} \wedge T, \end{cases} \tag{1.4}$$

  attains the optimal lower bound on $\lim_n N_T^n \langle Z^n \rangle_T$.

- We interpret the assumptions of Theorem 2.2.4 on the estimator sequence σ_t^n for a general class of weighted moving average estimators and specify some sufficient joint conditions on the lookback estimation period and the frequency of estimation times in order to preserve the asymptotic optimality of the strategy. In particular, under certain assumptions and for some deterministic sequence $(\alpha_n)_{n \geq 0}$ and general kernel functions $K_\gamma(\cdot)$ we prove in Theorem 2.4.1 that σ_t^n given by

$$\sigma_t^n = (\Sigma_t^n + \alpha_n \operatorname{Id}_d)^{1/2},$$

  where (for some admissible observation grid $\{\tilde{\tau}_i^n\}_{i \geq 0}$)

$$\Sigma_t^n = \sum_{\tilde{\tau}_i^n < t} K_{\gamma_n}(\tilde{\tau}_{i-1}^n - t) \Delta S_{\tilde{\tau}_i^n} \Delta S_{\tilde{\tau}_i^n}^\mathsf{T}, \tag{1.5}$$

  verifies the assumptions of Theorem 2.2.4 and thus yields an optimal discretization strategy attaining the uniform lower bound on $\liminf_n N_T^n \langle Z^n \rangle_T$ across the class of admissible strategies.

- We also provide a counter-example (see (2.1.7)) which shows that the knowledge of σ_t is important for the optimal strategy construction, while a misspecification of σ_t leads to a suboptimal performance. In addition, we support our claim by a numerical example in Section 2.5.

1.7 Central Limit Theorems for discretization errors

In the high frequency fixed horizon asymptotic framework for discretization problems, a common goal is to analyze the asymptotic behavior of the discretization error as the grid meshsize goes to 0. In particular, one typically wants to establish a Central Limit Theorem (CLT) for the renormalized discretization error process. Classical results on CLTs for regular grids may be found in [JP12].

Here we are particularly interested in existing works studying random discretization grids. Many of them are restricted to certain specific cases in terms of the model, the dimension ($d = 1$ or $d > 1$), the discretization error term or the discretization grid. In particular, a group of works analyze the discretization error for strongly predictable (up to extra independent noise) discretization grids. Among them [ASM04, DG04] study the problem of statistical estimation for diffusions and [BNS05, BNGJ$^+$06, MZ06, KP08] deal with the estimation of integrated variance and the power variation estimation in a non-parametric setting. We also mention [Roo80, GT01, HM05, GT09], more detail may be found in the introduction to Chapter 3.

The closest to our setting are the recent works considering more general stopping time grids. In [Fuk11b] the author provides a quite general CLT result in dimension 1 for stopping time grids. However the class of random grids under study is described in an implicit way through a set of assumptions. The limit distribution also depends on the processes whose existence is given as an assumption. Verification of the assumptions for a particular random discretization schemes may require a substantial effort. Generalization of this work to the general multidimensional case seems non-trivial both in terms of the extension of the central limit theorem (in particular, characterization of the limit distribution) given the abstract assumptions on the moments, and even more in term of the determination of the class of random grids verifying these assumptions. For example, natural candidates for endogenously generated discretization times are exit times from random domains, whose analysis is much more complicated in multidimensional setting (while in dimension 1 such a domain is given by the two boundary points). We also mention [LMR$^+$14] where a Central Limit Theorem (CLT) for estimating the integrated volatility in dimension 1 is established assuming the convergence in probability of renormalized quarticity and tricity. Here again the authors do not characterize the stopping times for which these convergences hold. Moreover, the result is only one-dimensional and studies a specific application.

In Chapter 3, we aim at closing this gap in the existing literature on the subject. Namely we want to establish a CLT for a multidimensional general discretization error term and a multidimensional process for an explicitly described class of random discretization grids with explicit limit characterization. We aim at providing a result that would be sufficiently general in terms of the random discretization grids considered and with assumptions that would be immediate to check for a specific model.

1.8 Summary of results in Chapter 3

For a given sample path of a stochastic process S on a time interval $[0, T]$ and a sequence of random discretization grids $\mathcal{T}^n := \{\tau_0^n = 0 < \tau_1^n < \cdots < \tau_{N_T^n}^n = T\}$ given by stopping times,

we consider an m-dimensional error term $\mathcal{E}_t^n$ given by $\mathcal{E}_t^n = \mathcal{E}_t^{n,1} + \mathcal{E}_t^{n,2}$, where

$$\mathcal{E}_t^{n,1} = \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} \mathcal{M}_{\tau_{i-1}^n}(S_s - S_{\tau_{i-1}^n})\mathrm{d}s, \qquad \mathcal{E}_t^{n,2} = \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} (S_s - S_{\tau_{i-1}^n})^{\mathsf{T}} \mathcal{A}_{\tau_{i-1}^n} \mathrm{d}B_s, \tag{1.6}$$

here $\mathcal{M}$ and $\mathcal{A}$ are continuous adapted processes with values in $\mathrm{Mat}_{m,d}(\mathbb{R})$ and $\mathrm{Mat}_{d,d}(\mathbb{R}) \otimes \mathbb{R}^m$ respectively (so that $\mathcal{A}_t$ maps bilinearly $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ to $x^{\mathsf{T}} \mathcal{A}_t y \in \mathbb{R}^m$, for details see Section 3.2.3).

The error term given by (1.6) appears in such important applications as:

- strategies for quadratic variation minimization with application to optimal hedging in finance;
- error analysis for integrated variance estimation based on random times observations;
- parametric estimation for diffusion processes observed at random times;

see the introduction of Chapter 3 for a detailed discussion.

The goal of Chapter 3 is to prove a functional central limit theorem for the sequence of the renormalized discretization error processes $(\sqrt{N_t}\mathcal{E}_t^n)_{0 \le t \le T}$, where $N_t^n := \#\{i \ge 1 : \tau_i^n \le t\}$.

In Chapter 3 we consider a quite general concrete class of random discretization grids (i.e. specified directly by its definition and not by abstract assumptions) given as follows. Let $\{(D_t^n)_{0 \le t \le T} : n \ge 0\}$ be a sequence of adapted processes with values in the set of domains in $\mathbb{R}^d$ (the details are in Section 3.2.2). In particular, we assume the convergence (in a suitable sense) to an adapted continuous domain-valued process $(D_t)_{0 \le t \le T}$. Let $(U_{i,n})_{n,i \in \mathbb{N}}$ be an i.i.d. family of random variables $U := \{U_{n,i} : i, n \in \mathbb{N}\}$ with $U_{n,i} \sim \mathcal{U}(0,1)$, that are independent of $\mathcal{F}_T$. Let $G : (t, \omega, u) \in [0,T] \times \Omega \times [0,1] \mapsto \mathbb{R}^+ \cup \{+\infty\}$ be a $\mathcal{P} \otimes \mathcal{B}([0,1])$-measurable mapping, where $\mathcal{P}$ denotes the σ-field of predictable sets of $[0,T] \times \Omega$, to simplify we write $G_t(u)$. We consider the following class of discretization grids $\mathcal{T} := \{\mathcal{T}^n : n \ge 0\}$ with $\mathcal{T}^n = \{\tau_i^n : 0 \le i \le N_T^n\}$ given by

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin \varepsilon_n D_{\tau_{i-1}^n}^n\} \wedge (\tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}) \wedge T, \end{cases} \tag{1.7}$$

where $(\Delta_{n,i})_{n,i \in \mathbb{N}}$ is a term representing some negligible contribution.

This class of discretization grids, in particular, allows coupling of endogenous noise generated by hitting times and extra independent noise given, for example, by a Poisson point process with stochastic intensity (see an example in Section 3.2.2).

The advantages of our setting include the following:

- we consider a general filtration, which allows models with regime switching, see Example 3.2.1;
- our framework allows both multidimensional process and multidimensional error term, as opposed to 1-dimensional setting considered in the previous works (e.g. [Fuk11b]);

- domain exit times represent complicated objects in multiple dimension as opposed to dimension 1;
- our model for the process S is given by quite general Bronwian semimartingales satisfying mild regularity assumptions and includes a variety of models used in practice, such as diffusion processes, path dependent processes, stochastic volatility models, etc. (see Example 3.2.1);
- we allow domains with corners (such as bounded intersections of half-spaces, i.e. polyhedrons).

Theorem 3.2.7 constitutes the main result of Chapter 3 and is given as follows. For explicitly defined processes m_t, Q_t and $\mathcal{K}_t$ we prove (under quite mild assumptions) the following $\bar{\mathcal{F}}$-stable functional convergence of $(\sqrt{N_t^n}\mathcal{E}_t^n)_{0\leq t\leq T}$ in distribution:

$$\sqrt{N_t^n}\mathcal{E}_t^n \underset{[0,T]}{\overset{d}{\Longrightarrow}} \sqrt{\int_0^t m_s^{-1}ds}\left(\int_0^t \mathcal{M}_s Q_s \mathrm{d}s + \int_0^t Q_s^{\intercal}\mathcal{A}_s \mathrm{d}B_s + \int_0^t \mathcal{K}_s^{1/2}\mathrm{d}W_s\right), \tag{1.8}$$

where W is an m-dimensional Brownian motion defined on an extended probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and independent of B.

The proof of Theorem 3.2.7 consists of the following two blocks, each of which is itself a valuable contribution:

- Theorem 3.3.1 show a CLT of the type (1.8) for general stopping time grids satisfying suitable assumptions. This is the first result of this type in the multidimensional case. Moreover the assumptions are well adapted for verification and a tractable characterization of the limit distribution is provided.
- Propositions 3.3.4, 3.3.5 provide a weak error bound for the domain exit time of an Itô process with respect to the perturbation of its diffusion coefficient and the domain. They provide very delicate analysis of the weak errors that allows to pass from local estimation for a single exit time to a global estimation for a sequence of discretization grids, based on such domain exit times, under weak assumptions.

An important and direct application of our result is the case of time grids given by hitting times of random ellipsoids. Such grids naturally appear in [GL14a] and Chapters 1-2 as optimal discretization strategies regarding the minimization of the quadratic variation criteria for multidimensional models and play important role in the problem of hedging error optimization in finance (see [Fuk11a]). Theorem 3.2.7, in particular, justifies the use of $\lim_n N_T^n\langle Z^n\rangle_T$ as a minimization criteria since this limit appears to be the variance of the asymptotic limit in the CLT. Another important application developed in Chapter 4 is the parametric inference of diffusions observed at stopping times.

1.9 Parametric inference for diffusions observed at random times

Parametric inference for stochastic processes differs from the classical setting of finite dimensional i.i.d. observations and is more complex. Usually only discrete observations of a

single sample process trajectory are available. A classical estimation approach is based on the approximation of the process transition densities between the observation times resulting in so called approximate maximum likelihood estimators (AMLEs). In the high-frequency fixed horizon framework the number of observations N over a fixed interval $[0, T]$ is supposed to be large, and we are interested in the asymptotic properties of the estimators as N goes to infinity.

Estimation usually requires the knowledge of the diffusion coefficient σ at the observation times, which requires a Markovian assumption since we only observe the process S. This restrict the class of models under study to the diffusion processes of the form

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma(s, S_s, \xi^\star)\mathrm{d}B_s, \qquad t \in [0, T], \quad S_0 \in \mathbb{R}^d, \tag{1.9}$$

where $\xi^\star$ is an unknown parameter.

A number of works study the problem of inference for diffusions. For general references, see the books [Sør04, Fuc13] and the lecture notes [Jac07].

The nonparametric estimation of the diffusion coefficient $\sigma(.)$ is investigated in [FZ93] for equidistant observations times on a fixed time interval. In [GCJ93] the authors consider the problem of the parametric estimation of a multidimensional diffusion under regular deterministic observation grids. They construct consistent sequences of estimators of the unknown parameter based on the minimization of certain contrasts and prove the weak convergence of the error renormalized at the rate $\sqrt{n}$ to a mixed Gaussian variable, where n is the number of observations. We also mention [GCJ94], see the discussion in the introduction to Chapter 4. The problem of achieving minimal variance estimator is investigated using the local asymptotic mixed normality (LAMN) property, see e.g. [CY90, Chapter 5] for the definition: this LAMN property is established in [Doh87] for one-dimensional S, and in [Gob01] for higher dimensions using Malliavin calculus techniques, when the n observation times are equidistant on a fixed interval. These latter results show the optimality of Gaussian AMLEs that achieve consistency with minimal variance.

Several works are dedicated to the inference problem with observations at random times, but under quite restrictive assumptions on those times. More precisely, in [ASM03, DG04] the authors assume that the time increment $\tau_i^n - \tau_{i-1}^n$ depends only on the information up to τ_{i-1}^n and on extra independent noise. A similar condition is considered in [HJY11], and it can take the form of strongly predictable times (τ_i^n is known at time τ_{i-1}^n). In [ASM04], the time increments are simply independent and identically distributed.

The above works consider only deterministic, strongly predictable of random independent grids. However, as we argue in Section 1.2, the case of more general random observation times given by stopping times is important in applications and must be investigated as well. To the best of our knowledge this setting has not yet been studied in the literature, except for [LMR+14] (in the non-parametric setting) where a Central Limit Theorem (CLT) for estimating the integrated volatility in dimension 1 is established assuming the convergence in probability of renormalized quarticity and tricity (however, the authors do not characterize the stopping times for which these convergences hold). One reason for this lack of studies in the literature is essentially that the necessary tools for the analysis of the stopping

time discretization grids for multidimensional processes were not available until recently. In particular, the study of the asymptotic normality for a sequence of estimators requires a general central limit theorem for discretization errors based on such grids. Such a result has been very recently obtained (see Chapter 3) in a concrete setting (i.e. for explicitly defined class of grids, and not given by abstract assumptions, as opposed to [LMR+14]), in several dimensions (as a difference with above references) and with a tractable limit characterization. Note that in [Fuk11b], the derivation of CLT is achieved in the context of general stopping times, but the limit depends on implicit conditions that are hardly tractable except in certain situations (notably in dimension 1).

In Chapter 4 we aim at constructing a consistent sequence of estimators $(\xi^n)_{n\geq 0}$ of the true parameter $\xi^\star$ and provide its asymptotic analysis in the case of random observation grids given by general stopping times. In particular, our setting covers those considered in the previous works on the subject and allows new more general observation stopping times, which is an important progress in the subject.

1.10 Summary of results in Chapter 4

We consider a d-dimensional Brownian semimartingale $(S_t)_{0\leq t\leq T}$ of the form (1.9), and a sequence of observation grids $\mathcal{T}^n = \{\tau_i^n : 0 \leq i \leq N_T^n\}$ verifying the assumptions $(\mathbf{A}_S^{\text{osc.}})$ and $(\mathbf{A}_N)$ given in Section 1.3.

Our goal is to construct for each $n \geq 0$ an estimator ξ^n of $\xi^\star$ based only on the knowledge of $\{\tau_i^n, S_{\tau_i^n} : 0 \leq i \leq N_T^n\}$. We also suppose that no additional information about the distribution properties of τ_i^n is provided (see the discussion in Section 4.1.2).

Although the distribution of S_τ, as τ is a stopping time, may be quite different from Gaussian, we are inspired by the same approach. However, we provide a slightly different interpretation of the same minimization criteria. We also generalize the criteria to account for non-equidistant distribution of the discretization points over $[0, T]$.

Denote $p_\Sigma(x) := (2\pi)^{-d/2}(\det\Sigma)^{-1/2}\exp\left(-\frac{1}{2}x^\mathsf{T}\Sigma^{-1}x\right)$ the density of a centered d-dimensional Gaussian variable $\mathcal{N}_d(0,\Sigma)$ with the covariance matrix Σ (assumed to be non-degenerate). Denote the Kullback-Leibler (KL) divergence between the variables $\mathcal{N}_d(0,\Sigma_1)$ and $\mathcal{N}_d(0,\Sigma_2)$ by

$$D_{\text{KL}}(\Sigma_1,\Sigma_2) := \int_{\mathbb{R}^d} p_{\Sigma_1}(x)\log\frac{p_{\Sigma_1}(x)}{p_{\Sigma_2}(x)}\mathrm{d}x. \tag{1.10}$$

For some continuous weight function $\omega : [0,T]\times\mathbb{R}^d \to]0,+\infty[$ set $\omega_t := \omega(t, S_t)$; the process $(\omega_t)_{0\leq t\leq T}$ is continuous adapted positive. Owing to a suitable identification assumption we show that the minimization of $\int_0^T D_{\text{KL}}(c_t(\xi^\star), c_t(\xi))\omega_t\mathrm{d}t$ under suitable assumptions yields the true parameter $\xi^\star$. Next we consider $U^\star(\cdot)$, given by

$$U^\star(\xi) := \int_0^T \left(\log(\det c_t(\xi)) + \text{Tr}(\sigma_t(\xi^\star)^\mathsf{T} c_t^{-1}(\xi)\sigma_t(\xi^\star))\right)\omega_t\mathrm{d}t,$$

and show that $\int_0^T D_{\text{KL}}(c_t(\xi^\star), c_t(\xi))\omega_t\mathrm{d}t = \frac{1}{2}U^\star(\xi) + C_0$, where C_0 is independent of ξ. The term $\int_0^T \text{Tr}(\sigma_t(\xi^\star)^\mathsf{T} c_t^{-1}(\xi)\sigma_t(\xi^\star))\omega_t\mathrm{d}t$ represents a quadratic variation. Thus we define the

following discretized version of $U^\star(\cdot)$, that uses only $\{\tau_i^n, S_{\tau_i^n} : 0 \leq i \leq N_T^n\}$,

$$U^n(\xi) := \sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \log\left(\det c_{\tau_{i-1}^n}(\xi)\right)(\tau_i^n - \tau_{i-1}^n) + \sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \Delta S_{\tau_i^n}^{\mathsf{T}} c_{\tau_{i-1}^n}^{-1}(\xi) \Delta S_{\tau_i^n}. \quad (1.11)$$

The random function $U^n(.)$ plays the role of a contrast function: it is asymptotically equal to $U^\star(.)$, for which the minimum is achieved at $\xi^\star$. In the case of regular grids and $\omega_t = 1$ the contrast (1.11) coincides with [GCJ93, eq. (3)].

We define the sequence of estimators $(\xi^n)_{n\geq 0}$ as follows:

$$\xi^n := \mathrm{Argmin}_{\xi\in\Xi} U^n(\xi) \quad (1.12)$$

(if the minimizing set of $U^n(\cdot)$ is not a single point we take any of its elements).

Note that the user is free to choose the form of the process ω_t. While the rigorous optimization of the choice of ω_t given only the observations $\{\tau_i^n, S_{\tau_i^n} : 0 \leq i \leq N_T^n\}$ is complicated, it seems reasonable to increase ω_t on the time intervals where the observation frequency is higher. We have not investigated furthermore in this direction.

In Chapter 4 we prove the results for the sequence of estimators $(\xi^n)_{n\geq 0}$ given by (1.12) following below. These results are new in the setting of general observation grids given by stopping times described in Section 1.3.

- Theorem 4.2.1 states that for the sequence estimators $(\xi^n)_{n\geq 0}$ given by (1.12) we have

$$\xi^n \xrightarrow[n\to+\infty]{\mathbb{P}} \xi^\star.$$

- In Theorem 4.2.2 we prove that under suitable assumptions we may write

$$\varepsilon_n^{-\rho_N}(\xi^n - \xi^\star) = (\mathcal{H}_T^{-1} + o_n^{\mathbb{P}}(1))\varepsilon_n^{-\rho_N} Z_T^n + o_n^{\mathbb{P}}(1),$$

  where ρ_N is given in ($\mathbf{A}_N$), $o_n^{\mathbb{P}}(1) \xrightarrow[n\to+\infty]{\mathbb{P}} 0$ and $\mathcal{H}_T$ is explicitly defined, and the term Z_t^n has the form of the discretization error studied in Chapter 3, i.e.

$$Z_T^n := \int_0^T \Delta S_t^{\mathsf{T}} \mathcal{A}_{\varphi(t)} \mathrm{d}B_t + \int_0^T \mathcal{M}_{\varphi(t)} \Delta S_t \mathrm{d}t$$

  for certain explicitly given processes $\mathcal{M}_t$ and $\mathcal{A}_t$. In particular, this result allows to derive a CLT for the estimator sequence by a direct application of a CLT for the discretization error process Z_t^n, such as the one developed in Chapter 3, as well as other works. As a consequence, our work provides sufficient results enabling the derivation of the CLT for the sequence of estimators $(\xi^n)_{n\geq 0}$ in a very general setting in terms of the (random) observation times, not previously available in the literature.

- In the case of 1-dimensional parameter ξ and when the asymptotic mixed normality holds without the bias term, Theorem 4.2.6 states a universal lower bound on the asymptotic variance of our sequence of estimators among the class of discretization grids given in Section 1.3 and prove the tightness of this bound. Namely we provide

a random variable $V_T^{opt.}$ such that if (for an arbitrary observation grid sequence $\mathcal{T}^n$) we have $\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{d} \mathcal{N}(0, V_T)$, and under some additional assumptions, we automatically get $V_T \geq V_T^{opt.}$ a.s. In addition, we provide a sequence of grids $\{\mathcal{T}^n : n \geq 0\}$ for which the limit variance is arbitrarily close to $V_T^{opt.}$ (in a suitable sense). To the best of our knowledge, this is the first result of this type in the parametric inference for diffusions (see also the discussion of the difference of our framework with [GCJ94] in the introduction of Chapter 4).

2 Part II: Uncertainty quantification for stochastic approximation limits

Part II of this thesis is devoted to the problem of model uncertainty quantification for Stochastic Approximation (SA) limits. SA is used to find zeros of a function $z \mapsto h(z)$ for which no closed-form formula is available and which may only be expressed as an expectation $h(z) := \mathbb{E}[H(z, V)]$. The expectation is taken with respect to some random variable V which models the stochastic system under study. In many applied problems the exact specification of the distribution of V is not known, and it is reasonable to consider the model for V as uncertain (see Section 2.2 for the motivation). We express this through a parametric dependence $V \sim \mu(\theta, \mathrm{d}v)$ where $\theta \in \Theta$ is an uncertain parameter following some distribution π. In this setting the function h, as well as its zero $z^\star$, will depend on θ, so that $z^\star = \phi^\star(\theta)$ for some function $\phi^\star(\cdot)$. Our goal is to compute $\phi^\star(\cdot)$ so that we can efficiently quantify the probability distribution of the SA limit $\phi^\star(\theta)$ given the probability distribution π for θ. We choose the chaos expansion approach and assume that $\phi^\star(\cdot)$ belongs to the Hilbert space of square integrable functions with respect to π. We design an SA procedure in increasing dimension (dubbed as the USA algorithm) for computing the basis coefficients of the chaos expansion of $\phi^\star(\cdot)$ on an orthogonal basis of this Hilbert space. This results in a sequence of approximations $\phi^k(\cdot)$ converging in a suitable sense to $\phi^\star(\cdot)$.

Chapters 5-6 are devoted to the design and the convergence analysis of the USA algorithm. Section 2.1 presents an introduction to Stochastic Approximation method. In Section 2.2 we formulate the problem of model uncertainty for SA limits and provide motivational examples. The chaos expansion approach is discussed in Section 2.3. Construction of the USA algorithm is given in Section 2.4. Section 2.5 provides the summary of our results on the convergence of the USA in Chapter 5 and the L^2-convergence rate of the USA obtained in Chapter 6.

2.1 Introduction to Stochastic Approximation algorithms

Stochastic Approximation method is used to find zeros of a function $z \mapsto h(z)$ expressed as an expectation. We assume that $h(z)$ represents an average of the values of some known function $H(z, V)$ over the random scenarios given by a random variable V. The goal is to calculate numerically a solution to $h(z) = \mathbb{E}[H(z, V)] = 0$ assuming that i.i.d. simulations of V are available. Classical deterministic methods combined with Monte Carlo approximation of each value of $h(z)$ would be too computationally demanding. Stochastic Approximation

was developed to solve this problem more efficiently.

Initiated by Robbins and Monro [RM51] and Kiefer and Wolfowitz [KW52] in the early 1950s, the theory of stochastic approximation algorithms has been the subject of extensive research, both theoretical and applied. It comprises the study of important theoretical issues in the analysis of dynamically defined stochastic processes and has a large number of applications. The SA method is now mainstream in such areas as optimization, parameter estimation, signal processing, adaptive control, Monte Carlo optimization of stochastic systems (see [KY97a, BMP90]), stochastic gradient descent methods in machine learning (see e.g. [BC05, SSS08, BCN17]), adaptive Monte Carlo sampler (see e.g. [HST01, AT08, FMP11, FJLS16, FS00, DVA98]), and efficient tail computations [BFP09], among others.

A common application of SA is where h is the gradient of a convex function c given by an expectation, i.e.

$$h(z) = \nabla_z c(z) = \nabla_z \mathbb{E}[C(z, V)].$$

In this case SA corresponds to the minimization of c and is called Stochastic Gradient Descent. Remark that in order to use classical SA we need to have $\nabla_z \mathbb{E}[C(z, V)] = \mathbb{E}[\nabla_z C(z, V)]$ and $H := \nabla_z C$ to be known. If only the function C is known one may apply a slightly different Kiefer-Wolfowitz procedure ([KW52]) using finite differences.

The basic paradigm of SA is a stochastic difference equation of the form

$$z^{n+1} = z^n + \gamma_{n+1} H(z^n, V_n).$$

Here, z is a parameter of a system, and the random vector $H(z^n, V_n)$ is an observation of a sample scenario of the system (produced by a simulation of $V_n \sim V$) with the parameter set to z^n. One recursively adjusts the parameter so that the goal is met asymptotically. The principal idea developed in [RM51] is that, if the step sizes γ_n in the parameter updates are allowed to go to zero in an appropriate way as $n \to +\infty$, then there is an implicit averaging that eliminates the randomness effects in the long run.

A lot of SA convergence results have been proved in various settings. To give an example we refer to [BFP09, Theorem 2.2] since this version is the closest to our analysis in Chapter 5. Concerning the rate of convergence for SA, classical results show the asymptotic normality for the renormalized sequence $\gamma_n^{1/2}(z^n - z^\star)$ where $z^\star = \lim_n z^n$ (see e.g. [Duf97, Chapter 2]). Applying in addition the Polyak-Ruppert averaging procedure (see e.g. [PJ92]) we can increase the convergence rate up to $1/n$ for any $\gamma_n = 1/n^a, a \in (1/2, 1)$.

We remark that the Monte Carlo method (aiming to calculate $\mathbb{E}[V]$) is a special case of SA with $\gamma_n = \frac{1}{n}$ and $H(z, v) = z - v$.

2.2 Uncertainty Quantification for SA limits: Motivation

The SA setting described in Section 2.1 involves a random variable V, so that the problem writes as $\mathbb{E}[H(z, V)] = 0$. In many applications, the choice of the model for V is of great importance. An exact specification of the model that describes some real-world phenomena must be chosen before the SA method is applied. Quite often it is chosen from a parametric family of distributions $\{\mu(\theta, \mathrm{d}v) : \theta \in \Theta \subset \mathbb{R}^d\}$, so that the parameter θ must be pre-estimated or set by an expect opinion. Obviously, a perfect specification of θ is rarely

possible. In some cases, where we lack information about θ, it is reasonable to assume that the model for V is uncertain. This may be expressed via additional randomness of the parameter θ.

Here we present several problems that are solved by SA, and for which the model uncertainty is relevant:

- Minimization of expected cost (or risk; or utility maximization) under model uncertainty. In this case V models a stochastic system and z corresponds to the parameter determining the strategy of interaction with this system. Further we have $H(z, v) := \nabla_z C(z, v)$ where C is some cost function. The goal is to find a strategy z in order to minimize the expected cost $\mathbb{E}[C(z, V)]$. Under suitable assumptions this will write as $\mathbb{E}[\nabla_z C(z, V)] = 0$ and thus may be solved by SA. In this case the model uncertainty problem for V is highly relevant. Examples of this setting include portfolio optimization (here one wants to optimally chose portfolio weights in order to minimize some expected risk, the variable V is the random market distribution over some future period, see e.g. [GMIC14]) and many other applied problems.

- SA may be used to calculate quantiles of a distribution, also known as Value-at-Risk (VaR) in finance (and more generally to calculate a pair of risk measures VaR and CVaR which are widely used, see [BFP09] for details). In financial applications V represents a future random value of some portfolio for which the choice of the distribution is not easy. Often we lack information about it and we need an efficient way to compute the risk measures for a family of models to analyze the model risk. In particular, such analysis is required by financial regulators.

- In some applications the Bayesian approach is used to choose the model for V. Here one considers a parametric family of distributions $\{\mu(\theta, \mathrm{d}v), \theta \in \Theta\}$ with some prior law on θ. After the observation of the data the law of θ is updated to some posterior distribution π. In this case the randomness of θ naturally yields model uncertainty for V.

We express the model uncertainty through a parametric dependence $V \sim \mu(\theta, \mathrm{d}v)$ of the distribution of V and suppose that θ follows some law π on Θ. In this case, the solution $z^\star$ of $\mathbb{E}[H(z, V)] = 0$ where $V \sim \mu(\theta, \mathrm{d}v)$ for a fixed θ, depends on θ, so that $z^\star = \phi^\star(\theta)$ for some function $\phi^\star(\cdot)$. Our main goal is to quantify the distribution of the solution (or the SA limit) given by $\{\phi^\star(\theta) : \theta \sim \pi\}$ via an approximation of $\phi^\star(\cdot)$ in a suitable functional space.

We remark that in the examples above a simpler alternative might be to increase the model complexity for V to incorporate the randomness of θ, i.e. to set V to follow the marginal distribution of $(\theta, V) \sim \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta)$. This problem may then be solved by standard methods. However such approach will give much less information, typically just one solution (representing some average value along different model specifications), while we are more interested in quantifying the entire distribution of $\phi^\star(\theta)$.

2.3 Problem formulation. Chaos expansion approach

Let us formalize the problem described in Section 2.2. We start from an equation $\mathbb{E}[H(z,V)] = 0$, where V is some random variable with values in a metric space $\mathcal{V}$, and impose the uncertainty in the form of a parametric dependence $V \sim \mu(\theta, \mathrm{d}v)$ (for some transition kernel μ from Θ to $\mathcal{V}$). The parameter θ itself follows some probability distribution $\pi(\mathrm{d}\theta)$ on $\Theta \subset \mathbb{R}^d$ which is known. We also allow uncertainty in the function H, through a dependency in θ, thus we take $H : \mathbb{R}^q \times \mathcal{V} \times \Theta \to \mathbb{R}^q$. We denote by L_2^π the Hilbert space of square integrable functions with respect to π. The uncertainty quantification problem for SA limits may be now formally given as follows:

$$\text{Find } \phi^\star \text{ in } L_2^\pi \text{ such that } \int_{\mathcal{V}} H(\phi^\star(\theta), v, \theta)\mu(\theta, \mathrm{d}v) = 0, \qquad \pi\text{-a.s.} \tag{2.1}$$

A naive way to access the distribution $\{\phi^\star(\theta), \theta \sim \pi\}$ would be to simulate $\theta \sim \pi$ and then, for each simulated value θ_i, run an SA procedure with the model parameter set to θ_i that will result in an approximation $\widehat{\phi^\star(\theta_i)}$ of $\phi^\star(\theta_i)$.

A more clever approach is to approximate the function $\phi^\star(\cdot)$ in a suitable way, so that further only the simulation of $\theta \sim \pi$ is needed. This may be done via the chaos expansion, which dates back to Wiener [Wie38] and has been developed in the fields of engineering and uncertainty quantification in the 2000s (see [GS03, LK10] and references therein). This technique, also known as the spectral method, consists of projecting the unknown function $\phi^\star : \Theta \mapsto \mathbb{R}^q$ on an orthonormal basis $\{\theta \mapsto B_i(\theta), i \in \mathbb{N}\}$ of the L^2 space with respect to the distribution π and computing the coefficients of its decomposition

$$\phi^\star = \sum_{i \geq 0} u_i^\star B_i. \tag{2.2}$$

In the most common case where $B_0 \equiv 1$, once the coefficients $\{u_i^\star, i \geq 0\}$ have been computed, the expectation and the variance-covariance matrix of $\{\phi^\star(\theta), \theta \sim \pi\}$ are available for free as

$$\mathbb{E}_{\theta\sim\pi}[\phi^\star(\theta)] = u_0^\star \text{ and } \mathbb{V}\mathrm{ar}_{\theta\sim\pi}(\phi^\star(\theta)) = \sum_{i \geq 1} u_i^\star (u_i^\star)^\top.$$

In the case of a polynomial basis, higher order moments are also usually computable explicitly, see [LK10, Appendix C].

The naive nested SA approach to calculate the coefficients $u_i^\star$'s is to simulate $\theta_1, \cdots, \theta_N \sim \pi$ and then for each θ_i run SA to get $\widehat{\phi^\star(\theta_i)}$ so that we get an approximation of $u_i^\star$ given by

$$\hat{u}_i := \frac{1}{N} \sum_{i=1}^{N} \widehat{\phi^\star(\theta_i)} B_i(\theta_i). \tag{2.3}$$

However, such method involves nested calculation and is inefficient. This may be easily seen on a simple example where SA is reduced to MC (a particular "linear" case with $H(z,v) = z - v$). This naive method will result in a 2-stage MC procedure which thus converges 2 times slower. The right approach here is to approximate the average $\int_{\Theta\times\mathcal{V}} v\mu(\theta, \mathrm{d}v)B_i(\theta)\pi(\mathrm{d}\theta)$

directly using i.i.d. simulations $(\theta, V) \sim \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta)$.

In this regard we expect that the naive approach (2.3) may be also largely improved in the general SA case and aim at designing an algorithm to calculate the coefficient $u_i^\star$'s using an efficient mix of the simulation of $\theta \sim \pi$ and the simulations feeding the SA algorithm. In the introduction of Chapter 5 we argue that a procedure in increasing dimension (i.e. with progressively increased truncation level) is needed due to non-linearity of the setting. Such chaos expansion methods are in general hard to analyze. Even in the case of an explicitly known function $\phi^\star$ (here finding individual coefficients $u_i^\star$ in (2.2) is straightforward by MC simulation) the global convergence of a method where more and more coefficients are computed by MC is subject to a nontrivial tuning of the speeds at which the number of coefficients and the number of simulations go to infinity (see [GS14]). Thus designing such a method in the case of general SA and analyzing its convergence is a non-trivial problem.

2.4 The USA algorithm

In Chapter 5 we design an SA procedure for computing the coefficients of $\phi^\star = \sum_{i\geq 0} u_i^\star B_i$ so that each iteration lies in finite dimensional subspace of the Hilbert space L_2^π, while the dimension of these subspaces goes to infinity.

Below we present a slightly simplified version of this procedure (called the USA algorithm, Uncertainty for Stochastic Approximation) that solves the problem (2.1). It is a fully constructive, detailed, and easy to implement. More details on the motivation and the construction of this algorithm may be found in Chapter 5.

Input: Sequences $\{\gamma_k, k \geq 1\}$, $\{m_k, k \geq 1\}$, $\{M_k, k \geq 1\}$, $K \in \mathbb{N}$, $\{u_i^0, i = 0, \dots, m_0\}$
for $k = 0$ **to** $K - 1$, **do**
  sample $(\theta_{k+1}^s, V_{k+1}^s), s = 1 \dots, M_{k+1}$, under the distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$; for $i > m_{k+1}$ define $u_i^k = 0$
  for $i = 0$ **to** m_{k+1}, **do**
    $u_i^{k+1} = u_i^k - \gamma_{k+1} M_{k+1}^{-1} \sum_{s=1}^{M_{k+1}} H\left(\sum_{j=0}^{m_k} u_j^k B_j(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) B_i(\theta_{k+1}^s)$
Output: The vector $\{u_i^K, i = 0, \dots, m_K\}$.

Algorithm 1: The USA algorithm for the coefficients of the basis decomposition of $\phi^\star$.

The inputs of the algorithm are: a positive stepsize sequence $\{\gamma_k, k \geq 1\}$, two integer valued sequences $\{m_k, k \geq 1\}$ and $\{M_k, k \geq 1\}$ corresponding to the number of non-zero coefficients in the approximation of $\phi^\star$ and to the number of Monte Carlo draws of the pair $(\theta, V) \sim \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$ at each iteration k, an initial value $u_0 \in \mathbb{R}^{m_0}$ and a total number of iterations K.

The output of the algorithm is a sequence $u^K = \{u_i^K, i \leq m_K\}$ approximating a solution $u^\star$. The corresponding approximation ϕ^K of a solution $\phi^\star$ to the problem (2.1) is then given by

$$\phi^K := \sum_{i=0}^{m_K} u_i^K B_i. \tag{2.4}$$

Further this approximation allows to easily calculate any statistics of the uncertain SA limit

$\phi^\star(\theta)$ using only the simulation of $\theta \sim \pi$.

2.5 Summary of results in Chapters 5 and 6

In Chapter 5, along with a more detailed motivation for the construction of the USA Algorithm 1, we prove its a.s. and $L^p, (p < 2)$, convergence.

In Theorem 5.3.5 we show under explicit and tractable assumptions that for the sequence $\{\phi^k : k \geq 0\}$ given by (2.4) we have

$$\lim_{k\to\infty} \left\|\phi^k - \phi^\infty\right\|_\pi = 0 \text{ a.s.}, \qquad \forall p \in (0,2) \ \lim_{k\to\infty} \mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^p\right] = 0.$$

where ϕ^∞ is some random variable on the set of solutions of the problem (2.1) (or simply $\phi^\infty = \phi^\star$ is the solution $\phi^\star$ to (2.1) is unique).

Algorithm 1 is an SA in the infinite dimensional Hilbert space l_2, which make the convergence analysis quite non-trivial. We argue that our result is original and is not covered by the previous papers on the SA in Hilbert spaces. There exists a number of works on infinite dimensional SA. We are only interested in SA in increasing dimension (so that all iterates lie in a finite dimensional subspace, and the algorithm is implementable). This is generally known as the sieve approach. The sieve-type SA procedures were studied in [Nix84, Gol88, Yin92] (in a particular case of independent noise, i.e. $H(z,V) = \tilde{H}(z) + V$). In [CW02] the authors derive results on the convergence and asymptotic normality for SA with growing dimension in a more general setting. However, the above-mentioned works are proved under fairly abstract conditions. Many of the assumptions in [YZ90] and [CW02] are hard to check. Also [Nix84, CW02] consider a noise term of the form $\tilde{H}(\phi^k, V_{k+1})$ with $\tilde{H} : \mathcal{H} \times \mathcal{V} \to \mathcal{H}$ and a single distribution for V_n's. By contrast, in our case, $H(\phi^k(\cdot), V_{k+1}, \cdot)$ can only be simulated θ by θ, as the distribution of V_{k+1} may depend on θ. In addition, some key assumptions of these works are not verified in our setting, see Chapter 5 for a more complete discussion and counter-examples (e.g. Remark 5.3.3).

In Section 5.5 we provide extensive numerical analysis of our algorithm, including a detailed discussion of the choice of its design parameters.

In Chapter 6 we analyze the L^2-convergence rate of the sequence $\{\phi^k, k \geq 0\}$ given by the USA algorithm. Our main result in Theorem 6.3.1 explicitly provides $\alpha > 0$ such that for some constant $C_\alpha > 0$ we have for all $k \geq 0$

$$\mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right] \leq C_\alpha \gamma_k^\alpha. \tag{2.5}$$

Control of the form γ_k^α is motivated by similar results in the finite dimensional case, where typically the SA squared error is proved to be of order $O(\gamma_k)$, i.e. $\alpha = 1$ (see e.g. [Duf97, Chapter 2]).

In finite-dimensional results on the SA convergence rate the convergence speed itself typically depends only on the step-size sequence γ_k. In our setting, however, the exponent α in (2.5) will depend non-trivially on the model, the regularity of $\phi^\star$, the choice of the basis functions and the design parameters of the USA algorithm. The knowledge of this dependence plays an important role in the correct tuning of the algorithm to guarantee the

L^2-convergence with the best possible rate, given the model specification. We illustrate how the obtained results justify the optimality of the heuristic choice of the dimension growth speed used in Section 6.3.2.

Again the previous works on the sieve-type infinite dimensional SA cannot be applied to the convergence rate analysis of the USA, see the introduction of Chapter 6. This emphasizes the novelty of our results.

Beyond model uncertainty, applications of our approach include sensitivity analysis, with respect to θ, or quasi-regression in the sense of reconstructing a whole unknown function, for instance in the context of nested Monte Carlo computations involving a nonlinear inner function $\theta \mapsto \phi^\star(\theta)$.

Introduction en français

Contents

Cette thèse étudie deux sujets différents : Discrétisation de processus à des temps d'arrêt et Quantification d'incertitudes pour des limites d'approximation stochastique. L'introduction de chaque partie est présentée indépendamment dans les Sections 3 et 4. Les Parties I et II contiennent respectivement 4 et 2 chapitres, qui se basent sur les articles suivants (publiés ou en révision):

(1) E. Gobet and U. Stazhynski. Optimal discretization of stochastic integrals driven by general Brownian semimartingale. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques, Vol. 54, No. 3, pp. 1556-1582,* (2018).

(2) E. Gobet and U. Stazhynski. Model-adaptive optimal discretization of stochastic integrals. En révision pour *Stochastics,* (2018).

(3) E. Gobet and U. Stazhynski. Central limit theorem for discretization errors based on stopping time sampling. *Soumis,* (2018).

(4) E. Gobet and U. Stazhynski. Parametric inference for diffusions observed at stopping times. *Soumis,* (2018).

(5) S. Crepey, G. Fort, E. Gobet and U. Stazhynski. Uncertainty quantification for stochastic approximation limits using chaos expansion. En révision pour *SIAM Journal of Uncertainty Quantification,* (2018).

(6) U. Stazhynski. Uncertainty quantification for stochastic approximation limits: L^2-convergence rate. *En préparation,* (2018).

3 Partie I: Discrétisation de processus à des temps d'arrêt

3.1 Introduction à la discrétisation de processus

Les problèmes de discrétisation jouent un rôle fondamental dans les applications des processus stochastiques à temps continu. En effet, comme seules des données discrètes peuvent être observées, traitées et simulées, des versions discrétisées de tels processus sont souvent utilisées en pratique. À cet égard, la quantification des erreurs reliées à la discrétisation est très importante. Dans ce travail, on étudie les problèmes de discrétisation pour une classe de modèles appelés processus d'Itô (voir la définition dans [RY99, p. 298]) et pour certaines de leurs généralisations. Un processus d'Itô $(S_t)_{0\le t\le T}$ sur un espace de probabilité filtré donné s'écrit comme

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}B_s, \quad t \in [0,T]$$

où $(B_t)_{0\le t\le T}$ est un mouvement brownien, $(b_t)_{0\le t\le T}$ et $(\sigma_t)_{0\le t\le T}$ sont des processus adaptés vérifiant des hypothèses convenables (bien que des termes à variation finie plus généraux puissent être pris en considération, voir le Chapitre 1). Cette classe de modèles est indispensable dans de nombreuses disciplines, y compris la finance, l'assurance, l'économie, la biologie, la dynamique de la population et la physique. On se place dans le cadre de haute fréquence et d'horizon fini pour les problèmes de discrétisation. Plus précisément, on suppose que l'intervalle de temps $[0,T]$ est fixe et pour tout $n \ge 0$ on se donne un schéma de discrétisation finie $\mathcal{T}^n = \{0 = \tau_0^n < \cdots < \tau_{N_T^n}^n = T\}$ avec un temps de discrétisation N_T^n (possiblement aléatoire) tendant vers l'infini quand $n \to +\infty$. Le but est soit de quantifier, soit d'optimiser (dans un sens asymptotique bien choisi) l'erreur produite par la substitution de S_t par $S_{\varphi(t)}$ dans une procédure donnée, où $\varphi(t)$ est le plus grand temps de discrétisation τ_i^n avant t.

Ci-dessous, on présente quelques applications pour lesquelles le traitement des problèmes de discrétisation est indispensable. La première classe de problèmes est liée aux statistiques. L'utilisation des modèles stochastiques à temps continu en pratique demande l'estimation de différentes statistiques ou de paramètres du modèle à partir d'observations d'une seule trajectoire du processus. Un exemple standard est la variation quadratique d'un processus d'Itô de dimension 1 donnée par $\int_0^T \sigma_t^2 \mathrm{d}t$ dont l'estimateur classique s'écrit comme $\sum_{i=1}^{N_T^n} (S_{\tau_i^n} - S_{\tau_{i-1}^n})^2$. Un cadre plus compliqué est l'estimation statistique des processus de diffusion, où le coefficient de diffusion $\sigma_t = \sigma(t, S_t, \theta)$ dépend d'un paramètre inconnu θ à estimer sur la base d'observations discrètes du processus (voir les Sections 3.9-3.10). Dans ces applications, l'erreur d'estimation fréquemment s'écrit en fonction de l'erreur de discrétisation pour le processus S. Cela signifie que l'analyse de la consistance et celle de la normalité asymptotique des estimateurs ramènent à l'analyse des erreurs de discrétisation correspondantes. La seconde classe de problèmes étudie l'optimisation du tracking d'une cible dont la dynamique est modélisée par un processus d'Itô. Ici, le but est de choisir de manière optimale

les temps de discrétisation pour le réajustement du système stochastique afin de minimiser certains critères exprimant la déviation de la cible continue. Le réajustement en continu est typiquement impossible à cause des coûts d'intervention variés (coûts de transaction). Les exemples d'applications en finance incluent la couverture d'options (voir [Fuk11a, GL14a]) et le tracking d'indices financiers (voir [PS04]), entre autres. Dans les problèmes de ce type, les temps de discrétisation optimaux dépendent de la trajectoire du processus de manière adaptative. Ainsi ils sont donnés par des temps d'arrêt aléatoires. Un problème particulier considéré dans les Chapitres 1-2 est la minimisation de la variation quadratique pour des intégrales stochastiques (voir la Section 3.4 pour les détails).

Enfin, un autre groupe de problèmes s'intéresse à la simulation de processus par des schémas de discrétisation et l'analyse ultérieure de l'erreur de discrétisation dans des simulations de Monte-Carlo (voir par exemple [FO15]). Cependant, on ne considère pas ce type de problèmes dans notre travail et on se concentre uniquement sur les deux classes de questions discutées précédemment.

3.2 Schémas de discrétisation aléatoires

La discrétisation basée sur des temps équidistants, c.à.d. pour $\tau_i^n = \frac{iT}{n}$, est un sujet bien fourni, voir par exemple [Roo80, JP98, HM05, GT09, GT01, MZ06] parmi d'autres, voir aussi [JP12] et les références y contenues. Cependant, en pratique, des temps de discrétisation sont assez souvent espacés de façon irrégulière. La nature de cette irrégularité peut être différente selon le cadre. De plus, les temps de discrétisation peuvent aussi être aléatoires, ce qui rend l'analyse encore plus compliquée. Concernant les problèmes d'estimation statistique, bien des raisons peuvent entraîner le caractère aléatoire des temps d'observation : *i)* une partie des données est manquante ; *ii)* les observations sont plus fréquentes durant certaines périodes de temps ou quand le processus se trouve dans certaines régions de l'espace ; *iii)* les observations arrivent aléatoirement selon les temps d'un processus de type Poisson ou des temps d'arrêt reliés au processus lui-même, et d'autres. De nombreux travaux dans cette direction signalent un effet non négligeable du caractère aléatoire des temps de discrétisation sur le comportement asymptotique des erreurs par rapport au cas déterministe classique. Par exemple, [ASM03] observe un impact considérable d'échantillonnage aléatoire sur les estimateurs dans le cadre de l'estimation paramétrique de diffusions. Dans [LZZ13] les auteurs remarquent qu'en prenant en compte le caractère aléatoire et endogène des grilles d'observation (lorsqu'il en existe), ils arrivent à améliorer considérablement la performance de l'estimateur de volatilité intégrée. Dans le problème d'optimisation des temps de discrétisation pour le tracking optimal, on retrouve naturellement des grilles de discrétisation aléatoires comme des temps optimaux du réajustement, et par conséquent, ils jouent un rôle clé dans l'analyse, voir [Fuk11a, GL14a]. L'importance des schémas de discrétisation aléatoires en finance haute fréquence a été souligné en particulier dans [DGM+01, Section 1.1] et [ASJ14, Chapitre 9], voir aussi [Fuk10, FR12, RR10, RR12]. On distingue les deux niveaux suivants de généralité lorsque l'on considère des schémas de discrétisation irrégulière :

1. Pour tout i le temps τ_i^n dépend uniquement de $\mathcal{F}_{\tau_{i-1}^n}$ (où $(\mathcal{F}_t)_{0\le t\le T}$ est une filtration fixée) et d'autres bruits indépendants. Ce groupe inclut, en particulier, toutes les

grilles déterministes, *fortement prévisibles* (c.à.d. τ_i^n est $\mathcal{F}_{\tau_{i-1}^n}$-mesurable, pas de bruit indépendant) et les temps aléatoires indépendants du processus.

2. Des temps d'arrêt généraux par rapport à une filtration donnée. Ce cadre présume que du bruit aléatoire endogène peut déclencher les temps de discrétisation. Un exemple de référence est la discrétisation d'un processus par ses temps d'atteinte de domaines spécifiés.

Le premier cadre, qui inclut les temps fortement prévisibles et les temps indépendants du processus, est mieux étudié, alors que le second est assez récent et plus compliqué pour l'analyse. Il constitue l'objet principal de notre travail. Des schémas de discrétisation aléatoires donnés par des temps d'atteinte sont étroitement liés au problème de l'optimisation de l'erreur de tracking (voir [Fuk11b, Fuk11a, GL14a]). Les Chapitres 1-2 sont consacrés à ce problème dans le contexte de discrétisation optimale d'intégrales stochastiques (une discussion détaillée est donnée dans la Section 3.4). La disponibilité des données seulement à des temps d'arrêt peut être une propriété intrinsèque d'un modèle dont le but est d'expliquer certaines observations qui arrivent de manière irrégulière. Assez récemment, plusieurs études ont été faites dans cette direction. Dans [RR10, RR12] les auteurs développent un modèle du prix financier en haute fréquence qui combine le bruit de microstructure, y compris l'erreur d'arrondi et l'échantillonnage à des temps de transaction basés sur des temps d'atteinte bien choisis. Ensuite, ils estiment la volatilité intégrée. Ils étudient aussi les propriétés asymptotiques de leur estimateur. Un cadre encore plus compliqué se présente lorsque les instants d'observation des différents composants d'un processus multidimensionnel sont aléatoires et qu'en plus, ils ne sont pas synchronisés. C'est un cadre typique dans certaines applications en finance (voir e.g. [HY08]). Pour une motivation supplémentaire pour des grilles de discrétisation aléatoires, on renvoie le lecteur à [GW02] où les auteurs fournissent des preuves empiriques de la connexion entre la volatilité et la durée entre transactions en finance, et à [Fuk10] qui étudie des modèles de données bid-ask du prix et l'échantillonnage à des temps de transactions. Dans les Chapitres 1-4 notre but est d'étendre la recherche actuelle sur la discrétisation de processus basée sur des temps d'arrêt aléatoires dans plusieurs directions, y compris les applications en statistiques et l'optimisation de l'erreur de tracking. Le reste de cette section contient les introductions plus détaillées pour chacun des problèmes étudiés avec une revue de la littérature, ainsi qu'un résumé des résultats de chaque chapitre. Dans la Section 3.3, on introduit la classe des suites de grilles aléatoires étudiée. Les techniques liées à cette classe sont indispensables pour notre analyse dans les Chapitres 1-4. Dans la Section 3.4 on continue avec l'introduction dans le problème de minimisation de la variation quadratique pour des intégrales stochastiques basées sur des semimartingales browniennes générales. On résume nos contributions à ce problème dans les Sections 3.5-3.6. Dans la Section 3.7 on discute les résultats précédents sur les théorèmes centraux limites pour les erreurs de discrétisation. Ensuite on résume notre travail sur le TCL pour des erreurs de discrétisation basées sur des grilles aléatoires. On conclut par une présentation de nos résultats sur l'estimation paramétrique de diffusions basée sur des observations à des temps d'arrêt dans la Section 3.10.

3.3 Une classe de grilles aléatoires de discrétisation

Dans cette section, on présente une classe de grilles aléatoires de discrétisation étudiée dans ce travail. Elle a été introduite dans [GL14a] comme la classe des grilles admissibles pour la discrétisation optimale d'intégrales stochastiques. Les techniques développées dans [GL14a] sont fondamentales pour tous les aspects de notre analyse dans les Chapitres 1-4. Cette classe est essentiellement définie par les deux hypothèses ci-dessous. Pour un processus S, une suite de grilles de discrétisation $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ avec $\mathcal{T}^n := \{\tau_i^n : 0 \leq i \leq N_T^n\}$, une suite déterministe strictement positive $(\varepsilon_n)_{n\geq 0}$, telle que $\sum_n \varepsilon_n^2 < +\infty$, et $\rho_N \geq 1$ on considère les hypothèses suivantes :

($\mathbf{A}_S^{\text{osc.}}$): La variable aléatoire suivante est p.s. finie :

$$\sup_{n\geq 0}\left(\varepsilon_n^{-2}\sup_{1\leq i\leq N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}|S_t - S_{\tau_{i-1}^n}|^2\right) < +\infty. \tag{3.1}$$

($\mathbf{A}_N$): Pour un paramètre $\rho_N \geq 1$ donné (qui vérifie certaines hypothèses, en particulier $\rho_N < 4/3$) la variable aléatoire suivante est p.s. finie:

$$\sup_{n\geq 0}(\varepsilon_n^{2\rho_N} N_T^n) < +\infty. \tag{3.2}$$

L'hypothèse ($\mathbf{A}_S^{\text{osc.}}$) signifie que les oscillations entre deux temps successifs suivent une règle de scaling ; elle implique que le pas entre temps successifs est suffisamment petit dans un sens défini par ε_n. Par ailleurs, ($\mathbf{A}_N$) dit que le nombre des temps aléatoires n'est pas très grand à une échelle donnée, ce qui exclut par exemple le cas d'accumulation de temps stochastique. Maintenant, pour tout $\varepsilon_n \to 0$ on considère une classe contenant les suites de grilles de discrétisation $\{\mathcal{T}^n : n \geq 0\}$ telles que pour toute sous-suite $\iota(n)$ il existe une autre sous-suite $\iota' \circ \iota(n)$ pour laquelle $\{\mathcal{T}^{\iota'\circ\iota(n)} : n \geq 0\}$ vérifie ($\mathbf{A}_S^{\text{osc.}}$) et ($\mathbf{A}_N$) avec $(\varepsilon_{\iota'\circ\iota(n)})_{n\geq 0}$. Cette forme particulière de la définition est motivée par le principe de sous-suites que l'on utilise plus tard pour passer de convergences p.s. à des convergences correspondantes en probabilité (voir le Lemme 2.2.2). En particulier, cette classe contient la plupart des grilles de discrétisation considérées dans les travaux précédents et des grilles que l'on peut imaginer du point de vue d'applications. Pour souligner sa généralité, on présente ci-dessous plusieurs familles assez larges des grilles aléatoires qui sont incluses dans cette classe (pour une justification, voir la remarque 1.2.2 et la discussion dans la Section 3.2.2)

1. Les suites $\mathcal{T} = \{\mathcal{T}^n : n \geq 0\}$ où chaque $\mathcal{T}^n = \{\tau_i^n : 0 \leq i \leq N_T^n\}$ est une grille formée des temps d'arrêt (avec N_T^n possiblement aléatoires) et telle que

$$C^{-1}\varepsilon_n^{\frac{2}{(1-\rho)}} \leq \min_{1\leq i\leq N_T^n}\Delta\tau_i^n \leq \max_{1\leq i\leq N_T^n}\Delta\tau_i^n \leq C\varepsilon_n^{\frac{2}{(1-\rho)}}, \quad n \geq 0, \quad \text{p.s.},$$

pour une variable aléatoire p.s. finie $C > 0$ et un paramètre $\rho > 0$. Cet exemple contient, en particulier, les suites de temps déterministes et fortement prévisibles pour lesquelles le pas de temps est contrôlé en haut et en bas et tend vers zéro.

2. Les temps aléatoires de Poisson avec du bruit indépendant de $\mathcal{F}_T$ mais avec une intensité stochastique $\mathcal{F}$-adaptée. Plus précisément, pour un processus continu adapté strictement positif $(\lambda_t)_{0\le t\le T}$ on considère $\mathcal{T}^n = \{\tau_i^n : 1 \le i \le N_T^n\}$ donné par les temps des sauts d'un processus de Poisson avec l'intensité $(\varepsilon_n^{-2\rho_N}\lambda_t)_{0\le t\le T}$.

3. On se donne une suite des processus aléatoires adaptés $\{D_t^n : 0 \le t \le T\}$ où D_t^n est un ensemble ouvert tel que

$$B(0, C_1\varepsilon_n) \subset D_t^n \subset B(0, C_2\varepsilon_n)$$

pour des variables aléatoires p.s. finies $C_1, C_2 > 0$; ici par $B(0,r)$ on note la boule centrée en 0 avec un rayon r. On définit une suite de stratégies $\mathcal{T} = \{\mathcal{T}^n : n \ge 0\}$ avec $\mathcal{T}^n = \{\tau_i^n : 0 \le i \le N_T^n\}$ de la manière suivante : $\tau_0^n = 0$ et pour tout $i \ge 1$

$$\tau_i^n = \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin D_{\tau_{i-1}^n}^n\} \wedge T.$$

Autrement dit, on considère des temps d'atteinte d'ensembles aléatoires de taille ε_n (des exemples plus compliqués peuvent être trouvés dans la Section 3.2.2).

Comme on peut le voir, la classe des grilles de discrétisation étudiée est assez universelle ; elle contient en effet à peu près tous les types de grilles de discrétisation qui peuvent sembler intéressants en pratique. Les Chapitres 1-4, en plus de leurs contributions principales, développent des outils puissants permettant d'analyser les grilles de discrétisation de ce type, ce qui fournit une base solide pour les études futures d'autres problèmes dans la discrétisation de processus.

3.4 Discrétisation optimale d'intégrales stochastiques

Cette section est une introduction pour les Chapitres 1-2. On considère le problème qui a pour l'ambition de trouver une suite finie des temps d'arrêt optimaux. $\mathcal{T}^n = \{0 = \tau_0^n < \tau_1^n < \cdots < \tau_{N_T^n}^n = T\}$ qui minimise la variation quadratique renormalisée de l'erreur de discrétisation pour une intégrale stochastique donnée par

$$Z_t^n = \int_0^t v(s, S_s)\cdot \mathrm{d}S_s - \sum_{\tau_{i-1}^n < t} v(\tau_{i-1}^n, S_{\tau_{i-1}^n})\cdot(S_{\tau_i^n \wedge t} - S_{\tau_{i-1}^n}), \tag{3.3}$$

où S est une semi-martingale brownienne continue de dimension d et $v(t,x)$ est une fonction continue à valeurs dans $\mathbb{R}^d$. Ici $T > 0$ est fixé et le nombre de temps d'arrêt N_T^n peut être aléatoire. Avec des conditions faibles sur le modèle, et pour des grilles déterministes et fortement prévisibles, l'erreur de discrétisation Z_T^n après une renormalisation bien choisie converge en loi vers un mélange de gaussiennes (voir [Roo80, KP91, JP12]). Un choix naturel du critère de minimisation dans ce cas est le produit $N_T^n\langle Z^n\rangle_T$. En particulier, dans le cas où $\sqrt{N_T^n}Z_T^n$ vérifie un TCL, la limite $\lim_n N_T^n\langle Z^n\rangle_T$ constitue la variance asymptotique (conditionnelle) de la loi limite (voir e.g. Chapitre 3). L'étude des problèmes de minimisation pour la discrétisation d'intégrales stochastiques a été initiée par [Fuk11a] en dimension $d = 1$, mais au lieu de $N_T^n\langle Z^n\rangle_T$ les auteurs considèrent un critère en espérance pour les

deux termes, c.à.d. $\mathbb{E}\left(N_T^n\right)\mathbb{E}\left(\langle Z^n\rangle_T\right)$. La minimisation trajectorielle de $\lim_n N_T^n\langle Z^n\rangle_T$ a été adressée dans un cadre multidimensionnel $d \geq 1$, dans [GL14a]. Les auteurs définissent une classe des stratégies de discrétisation admissibles comme celles vérifiant $(\mathbf{A}_S^{\text{osc.}})$-$(\mathbf{A}_N)$. Pour une martingale locale S et sous certaines hypothèses sur v (essentiellement son Jacobien $D_x v$ est inversible), ils fournissent une borne inférieure sur $\liminf_n N_T^n\langle Z^n\rangle_T$ pour toute la classe des suites de grilles admissibles. Une discussion complète de ce problème dans le cadre de couverture d'options en finance, ainsi qu'une présentation des travaux précédents, peuvent être trouvées dans [Fuk11a, GL14a]. Dans [GL14a] les auteurs exhibent des grilles de discrétisation optimales (ou arbitrairement proches de l'optimum) ayant une forme de temps d'atteinte d'ellipsoïdes aléatoires. Plus précisément, pour un processus continu adapté explicite $(\Lambda_t)_{0\leq t\leq T}$ à valeurs dans l'ensemble des matrices symétriques définies positives de taille $d\times d$, une suite optimale de grilles $\mathcal{T} := \{(\tau_i^n)_{0\leq i\leq N_T^n} : n \geq 0\}$ s'écrit comme

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T}\Lambda_{\tau_{i-1}^n}(S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^2\} \wedge T, \end{cases}$$

Comme démontré dans [GL14a], la suite optimale $\mathcal{T}$ est admissible et elle atteint la valeur optimale de $\lim_n N_T^n\langle Z^n\rangle_T$ parmi toutes les grilles admissibles. Ces résultats sont un point de départ de notre travail dans les Chapitres 1-2 qui est résumé dans les deux sections suivantes.

3.5 Résumé des résultats du Chapitre 1

Dans le Chapitre 1 on considère le problème de discrétisation optimale introduit dans la Section 3.4. On démontre des résultats d'optimalité dans un cadre beaucoup plus large par rapport à la littérature précédente. Premièrement, on permet S d'être une semi-martingale brownienne de la forme $S = A + M$ où A est un processus continu adapté général à variation finie qui satisfait certaines propriétés de Holder, alors que dans [GL14a] le processus S est essentiellement une martingale locale brownienne ($A = 0, M = \int_0^\cdot \sigma_s \mathrm{d}B_s$). Pour ce modèle généralisé, on démontre les résultats suivants :

- Dans le Théorème 1.3.4 on montre que les ensembles des stratégies admissibles (admissibilité par rapport à un processus S est définie par $(\mathbf{A}_S^{\text{osc.}})$-$(\mathbf{A}_N)$) pour une semi-martingale S et pour sa partie martingale locale M sont les mêmes. Le résultat est non trivial, il est démontré par un schéma de continuation avec une application successive de l'inégalité de BDG. Car [GL14a] suppose la condition de martingale locale, le Théorème 1.3.4 est primordial : il permet d'appliquer les résultats établis récemment dans [GL14a] à notre cadre plus général.

- Dans le Théorème 1.3.10 on montre que la stratégie de discrétisation basée sur les temps d'atteinte d'ellipsoïdes qui a la forme

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} H_{\tau_{i-1}^n}(S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^2\} \wedge T, \end{cases}$$

  est admissible sous certaines hypothèses. Ce résultat généralise [GL14a, Proposition

2.4].

- Le Théorème 1.4.2 est un des résultats principaux du Chapitre 1 ; il exhibe une borne inférieure uniforme sur $\liminf_{n\to+\infty} N_T^n\langle Z^n\rangle_T$ (qui est exacte comme démontré plus tard) parmi toutes les stratégies admissibles. Cette borne est donnée par

$$\liminf_{n\to+\infty} N_T^n\langle Z^n\rangle_T \geq \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2 \qquad \text{p.s.}$$

pour un processus X_t explicitement défini. C'est une extension importante de [GL14a, Théorème 3.1] au cas de semimartingales.

Deuxièmement, la partie martingale de S peut être dégénérée dans notre cadre, alors qu'une hypothèse plus forte d'ellipticité p.s. de σ_t est supposée dans [GL14a]. Autrement dit, on ne demande pas que l'inverse σ_t^{-1} existe. Aussi $D_x v(t, S_t)$ peut être non inversible dans notre cadre. Pour ce modèle généralisé, on démontre le résultat suivant : le Théorème 1.5.2 montre que la stratégie de discrétisation de la forme

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Lambda^{(n)}_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^2 \varepsilon_n\} \wedge T, \end{cases}$$

où $\Lambda_t^{(n)}$ est une perturbation bien choisie de $\Lambda_t := (\sigma_t^\dagger)^\mathsf{T} X_t \sigma_t^\dagger$ (par $\mathcal{M}^\dagger$ on note la matrice pseudo-inverse de $\mathcal{M}$), et X_t est une solution d'une certaine équation matricielle non linéaire donnée explicitement, atteint la valeur la plus petite possible de $\lim_n N_T^n\langle Z^n\rangle_T$ parmi toutes les suites des grilles aléatoires admissibles. La preuve du Théorème 1.5.2 est non trivial, car à cause de l'absence d'inversibilité de σ_t, on utilise des matrices pseudo-inverses de sorte que l'on perd certaines propriétés de continuité de la stratégie optimale. De plus, notre stratégie atteint exactement la limite optimale, tandis que dans [GL14a] seulement une stratégie μ-optimale a été établie (qui peut être arbitrairement proche de l'optimum). Dans la Section 1.5.3 on illustre par un test numérique la performance optimale de la stratégie exhibée dans le Théorème 1.5.2. La possibilité de traiter le cas non elliptique est aussi fondamentale pour les applications :

- Premièrement, elle permet de considérer des modèles partiellement dégénérés qui apparaissent dans des applications variées comme, par exemple, la mécanique aléatoire (voir la Section 1.5.3 pour des exemples).

- Deuxièmement, elle implique un résultat de robustesse de la stratégie optimale étudiée dans [GL14a] : le Théorème 1.5.2 montre que même si σ_t est proche d'être dégénérée, cela ne va pas impacter la performance de la stratégie optimale. C'est une considération importante dans les applications financières reliées à la couverture d'options, développées dans [GL14a] (voir la discussion dans la Section 1.5.3).

3.6 Résumé des résultats du Chapitre 2

Dans le Chapitre 2 on continue l'étude du problème de discrétisation optimale pour des intégrales stochastiques par rapport à des semimartingales browniennes, commencée dans le Chapitre 1. Ici notre but est de construire une version adaptative, qui ne demande aucune information sur le modèle de l'algorithme optimal de discrétisation présenté dans le Chapitre 1. Dans les travaux précédents, les suites optimales $\{\mathcal{T}^n : n \geq 0\}$ dépendent fortement du modèle pour S, en particulier, du processus de volatilité σ. Contrairement à cela, dans le Chapitre 2 on ne suppose aucune information sur le coefficient de diffusion du processus S. On ne suppose pas de modèle de diffusion pour S, ni de forme paramétrique de σ. Le processus S de la forme $A + \int_0^{\cdot} \sigma_s \mathrm{d}B_s$ est assez arbitraire et on suppose seulement qu'il vérifie certaines conditions faibles de régularité et d'inversibilité. Une version adaptative au modèle de l'algorithme de discrétisation optimale, développé dans Chapitre 1 est nécessaire pour rendre cet algorithme applicable en pratique. Une autre question importante est la robustesse de la discrétisation optimale par rapport à l'erreur de l'estimation de σ. La stratégie optimale s'écrit comme

$$\tau_i^n = \varphi\left(\sigma_{\tau_{i-1}^n}, D_x v(\tau_{i-1}^n, S_{\tau_{i-1}^n}), (S_t - S_{\tau_{i-1}^n})_{t \geq \tau_{i-1}^n}\right),$$

où φ représente une dépendance complexe et non linéaire. L'analyse de la robustesse de cette dépendance nécessite un effort considérable ; il est aussi critique pour les applications.

Dans le Chapitre 2 on étudie cette question et démontre les résultats suivants :

- Dans le Théorème 2.2.4 on établit des conditions suffisantes pour une suite générale d'estimateurs σ_t^n de σ_t qui garantissent l'optimalité de la suite de stratégies correspondantes. Plus précisément, supposons que σ_t^n vérifie (pour un paramètre $\delta > 0$) la condition

$$\varepsilon_n^{-\delta/2} \sup_{0 \leq t \leq T} |(\sigma_t \sigma_t^{\mathsf{T}})^{1/2} - \sigma_t^n| \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

  Alors pour une perturbation bien choisie $[\Lambda_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta}$ de Λ_t (utilisé dans le Chapitre 1 pour construire la stratégie optimale) la suite $\{\mathcal{T}^n\}_{n \geq 0}$, où $\mathcal{T}^n = (\tau_i^n)_{0 \leq i \leq N_T^n}$ est donné par

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^{\mathsf{T}} [\Lambda_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta} (S_t - S_{\tau_{i-1}^n}) \geqslant \varepsilon_n^{2+\delta}\} \wedge T, \end{cases} \tag{3.4}$$

  atteint la borne inférieure optimale sur $\lim_n N_T^n \langle Z^n \rangle_T$.

- On interprète les conditions du Théorème 2.2.4 imposées sur la suite d'estimateurs σ_t^n pour une classe générale d'estimateurs de la moyenne glissante pondérée et spécifie certaines conditions jointes sur la période d'estimation et la fréquence des temps d'estimation afin de préserver l'optimalité asymptotique de la stratégie. En particulier, sous des hypothèses convenables et pour une certaine suite $(\alpha_n)_{n \geq 0}$ et des fonctions générales de noyau $K_\gamma(\cdot)$ on démontre dans le Théorème 2.4.1 que σ_t^n donné par

$$\sigma_t^n = (\Sigma_t^n + \alpha_n \, \mathrm{Id}_d)^{1/2},$$

où (pour une grille d'observation admissible $\{\tilde{\tau}_i^n\}_{i\geq 0}$)

$$\Sigma_t^n = \sum_{\tilde{\tau}_i^n < t} K_{\gamma_n}(\tilde{\tau}_{i-1}^n - t)\Delta S_{\tilde{\tau}_i^n}\Delta S_{\tilde{\tau}_i^n}^{\mathsf{T}}, \tag{3.5}$$

vérifie les hypothèses du Théorème 2.2.4. Cela produit donc une stratégie de discrétisation optimale qu'atteint la borne inférieure uniforme sur $\liminf_n N_T^n \langle Z^n \rangle_T$ parmi toutes les stratégies admissibles.

- On fournit aussi un contre-exemple (voir (2.1.7)) qui montre que la connaissance de σ_t est importante pour la construction de la stratégie optimale, alors qu'un choix de σ_t erroné peut impliquer une performance sous-optimale. En plus, on soutient notre assertion par un exemple numérique dans la Section 2.5.

3.7 Théorème Central Limite pour des erreurs de discrétisation

Dans le cadre asymptotique de haute fréquence et d'horizon fixe pour les problèmes de discrétisation, la démarche classique est d'analyser le comportement asymptotique de l'erreur de discrétisation quand le pas de discrétisation tend vers 0. En particulier, on cherche à établir un Théorème Central Limite (TCL) pour les processus de l'erreur de discrétisation renormalisés. Des résultats classiques sur le TCL pour des grilles régulières peuvent être trouvés dans [JP12]. Ici on est particulièrement intéressé par les travaux précédents qui étudient des grilles de discrétisation aléatoires. Beaucoup d'entre eux sont restreints aux cas spécifiques en ce qui concerne le modèle, la dimension ($d = 1$ or $d > 1$), le terme de l'erreur de discrétisation ou la grille de discrétisation. En particulier, un groupe de travaux (de chercheurs ?) analyse l'erreur de discrétisation pour des grilles fortement prévisibles ou aléatoires indépendantes du processus. Parmi eux, [ASM04, DG04] étudient le problème de l'estimation statistique pour des diffusions et [BNS05, BNGJ+06, MZ06, KP08] traitent l'estimation de la variance intégrée et l'estimation de la variation de puissance dans un cadre non paramétrique. On cite aussi [Roo80, GT01, HM05, GT09], et pour plus de détails, on renvoie le lecteur à l'introduction du Chapitre 3. Les plus proches de notre cadre sont les travaux qui considèrent des grilles plus générales données par des temps d'arrêt. Dans [Fuk11b], l'auteur montre un TCL assez général en dimension 1 pour des grilles formées de temps d'arrêt. Cependant, la classe de grilles considérée est décrite de manière implicite par un ensemble d'hypothèses. La loi limite dépend aussi des processus dont l'existence est donnée comme hypothèse. La vérification de ces hypothèses pour un schéma de discrétisation particulier demande un effort considérable. La généralisation de ce travail pour le cas général multidimensionnel semble non triviale en ce que concerne l'extension du TCL (en particulier, la caractérisation de la loi limite) sous des hypothèses abstraites sur les moments des incréments du processus, ainsi que la détermination de la classe de grilles aléatoires qui vérifient ces hypothèses. Par exemple, des candidats naturels pour les temps de discrétisation générés de manière endogène sont des temps d'atteinte de domaines aléatoires, dont l'analyse est beaucoup plus compliquée dans le cadre multidimensionnel (alors qu'en dimension 1, un tel domaine est donné par les deux points du bord). On mentionne aussi [LMR+14] où un TCL pour l'estimateur de la volatilité intégrée en dimension 1 a été établi en supposant la

convergence en probabilité de *quarticity* et *tricity* renormalisées. Ici encore, les auteurs ne caractérisent pas les temps d'arrêt pour lesquels ces convergences ont lieu. En plus le résultat est unidimensionnel et il étudie une application particulière. Dans le Chapitre 3, on a pour objet de rectifier des lacunes dans la littérature existante sur le problème. Plus précisément, on veut établir un TCL pour une erreur de discrétisation multidimensionnelle générale et pour un processus multidimensionnel. De plus, on cherche à avoir une description explicite de la classe des grilles aléatoires considérées, et également une caractérisation explicite de la loi limite. Ainsi, notre but est de produire un résultat qui serait suffisamment général au niveau des grilles de discrétisation et avec des hypothèses qui seraient immédiates à vérifier pour des modèles particuliers.

3.8 Résumé des résultats du Chapitre 3

Pour une trajectoire donnée d'un processus S sur un intervalle de temps $[0,T]$ et une grille aléatoire de discrétisation $\mathcal{T}^n := \{\tau_0^n = 0 < \tau_1^n < \cdots < \tau_{N_T^n}^n = T\}$ formée de temps d'arrêt, on considère une erreur de discrétisation $\mathcal{E}_t^n$ de dimension m de la forme $\mathcal{E}_t^n = \mathcal{E}_t^{n,1} + \mathcal{E}_t^{n,2}$, où

$$\mathcal{E}_t^{n,1} = \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} \mathcal{M}_{\tau_{i-1}^n}(S_s - S_{\tau_{i-1}^n})\mathrm{d}s, \qquad \mathcal{E}_t^{n,2} = \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} (S_s - S_{\tau_{i-1}^n})^{\mathsf{T}} \mathcal{A}_{\tau_{i-1}^n} \mathrm{d}B_s, \tag{3.6}$$

ici $\mathcal{M}$ et $\mathcal{A}$ sont des processus continus adaptés à valeurs dans $\mathrm{Mat}_{m,d}(\mathbb{R})$ et $\mathrm{Mat}_{d,d}(\mathbb{R}) \otimes \mathbb{R}^m$ respectivement (de sorte que $\mathcal{A}_t$ est une application bilinéaire qui envoie $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ vers $x^{\mathsf{T}}\mathcal{A}_t y \in \mathbb{R}^m$, voir la Section 3.2.3) pour les détails. Le terme d'erreur donné par (3.6) apparaît dans les applications suivantes :

- les stratégies de minimisation de la variation quadratique avec une application à la couverture optimale d'options ;
- l'analyse d'erreur de l'estimation de la variance intégrée basée sur des observations à des temps d'arrêt ;
- l'estimation paramétrique de diffusions observées à des temps d'arrêt ;

voir l'introduction du Chapitre 3 pour plus de détails. Le but du Chapitre 3 est de démontrer un théorème central limite fonctionnel pour une suite de processus d'erreurs de discrétisation renormalisés $(\sqrt{N_t}\mathcal{E}_t^n)_{0\le t\le T}$, où $N_t^n := \#\{i \ge 1 : \tau_i^n \le t\}$. Dans le Chapitre 3 on considère une classe concrète assez générale de grilles de discrétisation aléatoires (c.à.d. spécifiée directement par sa définition et non pas par des hypothèses abstraites) donnée comme suit. Soit $\{(D_t^n)_{0\le t\le T} : n \ge 0\}$ une suite des processus adaptés à valeurs dans l'ensemble de domaines dans $\mathbb{R}^d$ (voir les détails dans la Section 3.2.2). En particulier, on suppose la convergence dans un sens approprié vers un processus continu adapté $(D_t)_{0\le t\le T}$ à valeur dans les domaines. Soit $(U_{i,n})_{n,i\in\mathbb{N}}$ une famille i.i.d. de variables aléatoires $U := \{U_{n,i} : i,n \in \mathbb{N}\}$ avec $U_{n,i} \sim \mathcal{U}(0,1)$, qui sont indépendantes de $\mathcal{F}_T$. Soit $G : (t,\omega,u) \in [0,T] \times \Omega \times [0,1] \mapsto \mathbb{R}^+ \cup \{+\infty\}$ une application $\mathcal{P} \otimes \mathcal{B}([0,1])$-mesurable, où par $\mathcal{P}$ on note la tribu des ensembles

prévisibles de $[0,T]\times\Omega$, pour simplifier on écrit $G_t(u)$. On considère une classe de grilles de discrétisation $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ avec $\mathcal{T}^n = \{\tau_i^n : 0 \leq i \leq N_T^n\}$ donnée par

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin \varepsilon_n D_{\tau_{i-1}^n}^n\} \wedge (\tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}) \wedge T, \end{cases} \tag{3.7}$$

où $(\Delta_{n,i})_{n,i\in\mathbb{N}}$ est un terme d'erreur négligeable. Cette classe de grilles de discrétisation permet, en particulier, un couplage du bruit endogène généré par des temps d'atteinte et du bruit indépendant donné, par exemple, par un processus de Poisson avec une intensité stochastique (voir par exemple la Section 3.2.2). Les avantages de notre cadre sont, en particulier, que :

- on considère une filtration générale qui permet des modèles avec un changement de régime, voir Exemple 3.2.1;
- notre cadre permet des processus et des termes d'erreurs multidimensionnels, contrairement aux résultats de dimension 1 dans les travaux précédents (e.g. [Fuk11b]);
- les temps d'atteinte des domaines constituent des objets compliqués en plusieurs dimensions à la différence de la dimension 1 ;
- notre cadre pour le processus S comprend des semimartingales browniennes assez générales qui satisfont certaines hypothèses faibles de régularité ; il inclut beaucoup de modèles utilisés en pratique comme, par exemple, des processus de diffusion, des processus dépendants de trajectoire, des modèles à volatilité stochastique, etc. (voir Exemple 3.2.1);
- on permet des domaines avec des coins (comme des intersections bornées de demi-espaces, c.à.d. des polyèdres).

Le Théorème 3.2.7 constitue le résultat principal du Chapitre 3 qui est donné comme suit : Pour des processus explicites m_t, Q_t et $\mathcal{K}_t$ on démontre (sous des hypothèses faibles), la convergence en loi $\bar{\mathcal{F}}$-stable fonctionnelle de $(\sqrt{N_t^n}\mathcal{E}_t^n)_{0\leq t\leq T}$:

$$\sqrt{N_t^n}\mathcal{E}_t^n \underset{[0,T]}{\overset{d}{\Longrightarrow}} \sqrt{\int_0^t m_s^{-1} ds}\left(\int_0^t \mathcal{M}_s Q_s \mathrm{d}s + \int_0^t Q_s^\intercal \mathcal{A}_s \mathrm{d}B_s + \int_0^t \mathcal{K}_s^{1/2}\mathrm{d}W_s\right), \tag{3.8}$$

où W est un mouvement brownien de dimension m défini sur un espace de probabilité étendu $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ qui est indépendant de B. La preuve du Théorème 3.2.7 consiste des deux blocs, dont chacun représente une contribution importante :

- Le Théorème 3.3.1 montre un TCL du type (3.8) pour des grilles générales données par des temps d'arrêt sous des hypothèses convenables. C'est le premier résultat de ce type dans un cadre multidimensionnel. En plus, les hypothèses sont bien adaptées à la vérification, et une caractérisation traitable de la loi limite est fournie.
- Les Propositions 3.3.4, 3.3.5 exhibent une borne sur l'erreur faible pour les temps d'atteinte de domaines par un processus d'Itô, par rapport à une perturbation de

son coefficient de diffusion et du domaine. Ils effectuent une analyse très délicate de l'erreur faible qui permet de passer d'une estimation locale pour un seul temps d'arrêt à une estimation globale pour une suite de grilles de discrétisation formée de tel temps d'arrêt, sous des hypothèses peu restrictives.

Une application directe et importante de notre résultat est le cas des grilles de temps donnés par des temps d'atteinte d'ellipsoïdes aléatoires. Les grilles de ce type apparaissent dans [GL14a] et les Chapitres 1-2 comme des stratégies optimales de discrétisation pour la minimisation de la variation quadratique, dans un cadre multidimensionnel. Ils jouent un rôle important dans le problème d'optimisation de l'erreur de couverture d'options en finance (voir [Fuk11a]). Le Théorème 3.2.7, en particulier, justifie l'utilisation de $\lim_n N_T^n \langle Z^n \rangle_T$ comme un critère de minimisation car cette limite se révèle être la variance de la loi limite dans le TCL. Une autre application importante développée dans le Chapitre 4 est l'estimation paramétrique de diffusions observées à des temps d'arrêt.

3.9 L'estimation paramétrique de diffusions observées à des temps d'arrêt

L'estimation paramétrique de processus stochastiques est une tâche plus difficile par rapport au cadre classique des observations i.i.d.. Typiquement, seules des observations discrètes d'une seule trajectoire du processus sont disponibles. Une approche classique d'estimation est basée sur l'approximation des densités de transition du processus entre les temps d'observation, et utilisée pour construire des estimateurs de vraisemblance approchée. Dans le cadre de haute fréquence et d'horizon fini, le nombre d'observations N sur un intervalle fixe $[0, T]$ est supposé large, et on s'intéresse aux propriétés asymptotiques d'estimateurs quand N tend vers l'infini. L'estimation typiquement (ou type ?) demande la connaissance du coefficient de diffusion σ aux temps d'observation, lequel, à son tour, nécessite une hypothèse de Markov car on observe uniquement le processus S. Cela restreint la classe de modèles étudiés aux processus de diffusion de la forme

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma(s, S_s, \xi^\star) \mathrm{d}B_s, \qquad t \in [0, T], \quad S_0 \in \mathbb{R}^d, \tag{3.9}$$

où $\xi^\star$ est un paramètre inconnu. De nombreux travaux étudient le problème d'estimation de diffusions. Pour des références générales, voir les livres [Sør04, Fuc13] et les notes de cours [Jac07]. L'estimation non paramétrique de coefficient de diffusion $\sigma(.)$ est étudiée dans [FZ93] pour des temps d'observation équidistants sur un intervalle fixe. Dans [GCJ93] les auteurs considèrent le problème d'estimation paramétrique pour des diffusions multidimensionnelles et des grilles d'observations régulières. Ils construisent une suite consistante d'estimateurs du paramètre inconnu basée sur la minimisation de certaines fonctions appelées *contrastes* et démontrent la convergence faible de l'erreur renormalisée au taux $\sqrt{n}$ vers un mélange de gaussiennes, où n est le nombre d'observations. On cite aussi [GCJ94], voir la discussion dans l'introduction du Chapitre 4. Le problème d'estimation à variance minimale a été étudié en utilisant la propriété de normalité mélangée asymptotique locale (local asymptotic mixed normality ou LAMN en anglais), voir e.g. [CY90, Chapitre 5] pour la définition : la propriété LAMN a été établie dans [Doh87] pour S unidimensionnel, et dans [Gob01]

pour plusieurs dimensions en utilisant le calcul de Malliavin, quand n temps d'observation sont équidistants sur un intervalle fixe. Ces derniers résultats montrent l'optimalité des estimateurs gaussien qui atteignent la consistance avec la variance optimale. Certains travaux sont dédiés au problème d'estimation basée sur des observations à des temps aléatoires, mais sous des hypothèses assez restrictives sur ces temps. Plus précisément dans [ASM03, DG04] les auteurs supposent que l'incrément $\tau_i^n - \tau_{i-1}^n$ du temps dépend seulement de l'information avant τ_{i-1}^n et d'autres bruits indépendants. Une condition similaire a été considérée dans [HJY11], et il peut y avoir une forme des temps fortement prévisible (τ_i^n est connu à temps τ_{i-1}^n). Dans [ASM04], les incréments de temps sont simplement indépendants identiquement distribués. Les travaux précédents considèrent seulement des grilles déterministes, fortement prévisibles ou aléatoires indépendantes du processus. Cependant, suivant les arguments dans la Section 3.2, le cas des temps d'observation plus généraux donnés par des temps d'arrêt est important dans les applications et doit aussi être examiné. À notre connaissance, ce type de problèmes n'était pas encore étudié dans la littérature, sauf dans [LMR+14] (dans un cadre non paramétrique)où un TCL pour l'estimation de la volatilité intégrée en dimension 1 a été établi en supposant la convergence en probabilité de *quarticity* et *tricity* renormalisées (mais les auteurs ne caractérisent pas les temps d'arrêt pour lesquels ces convergences ont lieu). Une raison qui explique la carence d'études dans la littérature est essentiellement que les techniques nécessaires pour l'analyse des grilles de discrétisation aléatoires pour des processus multidimensionnels n'étaient pas disponibles jusqu'à récemment. En particulier, l'étude de la normalité asymptotique pour des suites d'estimateurs demande un TCL pour des erreurs de discrétisation basées sur telles grilles. Un résultat de ce type a été établi récemment (voir le Chapitre 3) dans un cadre concret (c.à.d. pour une classe de grilles définie explicitement, et non pas par des hypothèses abstraites, contrairement à [LMR+14]), en plusieurs dimensions (contrairement aux travaux précédents) et avec la loi limite explicite. Notons que dans [Fuk11b], la dérivation du TCL est atteinte dans le contexte des temps d'arrêt généraux, mais la limite dépend des conditions implicites qui sont difficiles à vérifier sauf certains cas particuliers (notamment en dimension 1). Dans le Chapitre 4 notre but est de construire une suite consistante d'estimateurs $(\xi^n)_{n\geq 0}$ du vrai paramètre $\xi^\star$ et d'effectuer son analyse asymptotique dans le cas de grilles de discrétisation données par des temps aléatoires. En particulier, notre cadre couvre les grilles considérées dans les travaux précédents sur le sujet mais permet aussi des temps d'observation plus généraux donnés par des temps d'arrêt, ce qui constitue un avancement important dans le sujet.

3.10 Résumé des résultats du Chapitre 4

On se donne une semi-martingale brownienne $(S_t)_{0\leq t\leq T}$ de dimension d de la forme (3.9), et une suite de grilles d'observation $\mathcal{T}^n = \{\tau_i^n : 0 \leq i \leq N_T^n\}$ qui vérifie les hypothèses $(\mathbf{A}_S^{\text{osc.}})$ et $(\mathbf{A}_N)$ introduites dans la Section 3.3. Notre but est de construire pour tout $n \geq 0$ un estimateur ξ^n de $\xi^\star$ utilisant seulement la connaissance de $\{\tau_i^n, S_{\tau_i^n} : 0 \leq i \leq N_T^n\}$. On suppose aussi qu'aucune information supplémentaire concernant les propriétés de la loi de τ_i^n n'est fournie (voir la discussion dans la Section 4.1.2). Bien que la loi de S_τ, pour τ un temps d'arrêt, puisse être assez différente de la loi gaussienne, on est inspiré par la même approche. Cependant, on présente une interprétation un peu différente du même critère de minimisa-

tion. On généralise aussi le critère pour prendre en compte l'irrégularité de distribution des points de discrétisation sur $[0,T]$. Notons $p_\Sigma(x) := (2\pi)^{-d/2}(\det\Sigma)^{-1/2}\exp\left(-\frac{1}{2}x^\mathsf{T}\Sigma^{-1}x\right)$ la densité de la loi gaussienne centrée $\mathcal{N}_d(0,\Sigma)$ de dimension d avec la matrice de covariance Σ (supposée être non dégénérée). Notons la divergence de Kullback-Leibler (KL) entre les variables $\mathcal{N}_d(0,\Sigma_1)$ et $\mathcal{N}_d(0,\Sigma_2)$ par

$$D_{\mathrm{KL}}(\Sigma_1,\Sigma_2) := \int_{\mathbb{R}^d} p_{\Sigma_1}(x)\log\frac{p_{\Sigma_1}(x)}{p_{\Sigma_2}(x)}\mathrm{d}x. \tag{3.10}$$

Pour une fonction de poids continue $\omega : [0,T]\times\mathbb{R}^d \to]0,+\infty[$ on note $\omega_t := \omega(t,S_t)$; le processus $(\omega_t)_{0\le t\le T}$ est continu adapté positif. Sur la base d'une condition d'identifiabilité appropriée, on montre que la minimisation de $\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star),c_t(\xi))\omega_t\mathrm{d}t$ sous certaines hypothèses donne le vrai paramètre $\xi^\star$. Ensuite on considère $U^\star(\cdot)$, donné par

$$U^\star(\xi) := \int_0^T \left(\log(\det c_t(\xi)) + \mathrm{Tr}(\sigma_t(\xi^\star)^\mathsf{T} c_t^{-1}(\xi)\sigma_t(\xi^\star))\right)\omega_t\mathrm{d}t,$$

et on montre que $\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star),c_t(\xi))\omega_t\mathrm{d}t = \frac{1}{2}U^\star(\xi) + C_0$, où C_0 est indépendant de ξ. Le terme $\int_0^T \mathrm{Tr}(\sigma_t(\xi^\star)^\mathsf{T} c_t^{-1}(\xi)\sigma_t(\xi^\star))\omega_t\mathrm{d}t$ représente une variation quadratique. Ainsi on définit la version discrétisée de $U^\star(\cdot)$ suivante, qui utilise seulement $\{\tau_i^n, S_{\tau_i^n} : 0\le i\le N_T^n\}$,

$$U^n(\xi) := \sum_{\tau_{i-1}^n<T} \omega_{\tau_{i-1}^n}\log\left(\det c_{\tau_{i-1}^n}(\xi)\right)(\tau_i^n-\tau_{i-1}^n) + \sum_{\tau_{i-1}^n<T}\omega_{\tau_{i-1}^n}\Delta S_{\tau_i^n}^\mathsf{T} c_{\tau_{i-1}^n}^{-1}(\xi)\Delta S_{\tau_i^n}. \tag{3.11}$$

La fonction aléatoire $U^n(.)$ joue un rôle de contraste : elle est asymptotiquement égale à $U^\star(.)$, pour laquelle le minimum est atteint en $\xi^\star$. Dans le cas des grilles régulières et $\omega_t = 1$ le contraste (3.11) coïncide avec [GCJ93, eq. (3)]. On définit une suite d'estimateurs $(\xi^n)_{n\ge0}$ comme suit :

$$\xi^n := \mathrm{Argmin}_{\xi\in\Xi} U^n(\xi) \tag{3.12}$$

(si l'ensemble minimisant de $U^n(\cdot)$ n'est pas un singleton, on prend n'importe lequel de ces éléments). Notons que l'utilisateur est libre de choisir la forme du processus ω_t. L'optimisation rigoureuse du choix de ω_t en utilisant uniquement $\{\tau_i^n, S_{\tau_i^n} : 0\le i\le N_T^n\}$ est compliquée ; néanmoins, en pratique, il semble raisonnable de faire augmenter ω_t sur les intervalles de temps où la fréquence des observations est plus élevée. On n'a pas réalisé d'autres études dans cette direction. Dans le Chapitre 4 on démontre les résultats pour la suite d'estimateurs $(\xi^n)_{n\ge0}$ donnée par (3.12) présentés ci-dessous. Ces résultats sont nouveaux dans le cadre de grilles d'observation aléatoires données par des temps d'arrêt introduites dans la Section 3.3.

- Le Théorème 4.2.1 déclare que pour la suite d'estimateurs $(\xi^n)_{n\ge0}$ donnée par (3.12) on a
$$\xi^n \xrightarrow[n\to+\infty]{\mathbb{P}} \xi^\star.$$

- Dans le Théorème 4.2.2 on montre que sous des hypothèses convenables, on a une

représentation

$$\varepsilon_n^{-\rho_N}(\xi^n - \xi^\star) = (\mathcal{H}_T^{-1} + o_n^{\mathbb{P}}(1))\varepsilon_n^{-\rho_N} Z_T^n + o_n^{\mathbb{P}}(1),$$

où ρ_N est donné par ($\mathbf{A}_N$), $o_n^{\mathbb{P}}(1) \xrightarrow[n\to+\infty]{\mathbb{P}} 0$ et $\mathcal{H}_T$ sont définis explicitement, et le terme Z_t^n a une forme de l'erreur de discrétisation étudiée dans le Chapitre 3, c.à.d.

$$Z_T^n := \int_0^T \Delta S_t^\intercal \mathcal{A}_{\varphi(t)} \mathrm{d}B_t + \int_0^T \mathcal{M}_{\varphi(t)} \Delta S_t \mathrm{d}t$$

pour certains processus $\mathcal{M}_t$ et $\mathcal{A}_t$ définis explicitement. En particulier, ce résultat permet de déduire un TCL pour la suite d'estimateurs (3.12) par une application directe d'un TCL pour l'erreur de discrétisation Z_t^n, comme celui dans le Chapitre 3, ainsi que pour d'autres résultats. Par conséquent, notre travail fournit des résultats suffisants qui permettent de déduire un TCL pour une suite d'estimateurs $(\xi^n)_{n\geq 0}$ dans un cadre très général en termes de temps d'observation aléatoires, lequel cadre n'était pas disponibles avant dans la littérature.

- Dans le cas de paramètre ξ de dimension 1 et quand la normalité asymptotique est vérifiée sans terme du biais, le Théorème 4.2.6 exhibe une borne inférieure universelle sur la variance asymptotique de notre suite d'estimateurs parmi toute la classe de grilles de discrétisation introduite dans la Section 3.3. On montre aussi que cette borne est tendue. Notamment on exhibe une variable aléatoire $V_T^{opt.}$ telle que si (pour une suite de grilles d'observation assez arbitraire $\mathcal{T}^n$) on a $\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{d} \mathcal{N}(0, V_T)$, et sous certaines hypothèses, on obtient automatiquement que $V_T \geq V_T^{opt.}$ p.s. En plus, on donne une suite de grilles $\{\mathcal{T}^n : n \geq 0\}$ pour laquelle la variance limite est arbitrairement proche de $V_T^{opt.}$ dans un sens approprié. À notre connaissance, c'est le premier résultat de ce type dans l'estimation paramétrique de diffusions (voir aussi la discussion sur la différence de notre cadre par rapport à [GCJ94] dans l'introduction du Chapitre 4).

4 Partie II : Quantification d'incertitudes pour des limites d'approximation stochastique

La Partie II de cette thèse s'intéresse au problème de la quantification d'incertitudes de modèle pour des limites d'approximation stochastique. L'approximation stochastique (Stochastic Approximation ou SA en anglais) est utilisée pour trouver des zéros d'une fonction $z \mapsto h(z)$ pour laquelle il n'y a pas de formule fermée, et qui est disponible sous une forme d'espérance $h(z) := \mathbb{E}[H(z, V)]$. L'espérance est prise par rapport à une variable aléatoire V qui modélise le système stochastique étudié. Dans de nombreux problèmes appliqués, la spécification exacte de la distribution de V est inconnue, et il est raisonnable de considérer le modèle comme incertain (voir la Section 4.2 pour la motivation). Cette incertitude peut être exprimée par une dépendance paramétrique $V \sim \mu(\theta, \mathrm{d}v)$ où $\theta \in \Theta$ est un paramètre incertain qui suit une certaine loi π. Dans ce cadre la fonction h, ainsi que son zéro $z^\star$, va dépendre de θ, ou autrement dit $z^\star = \phi^\star(\theta)$ pour une fonction $\phi^\star(\cdot)$. Notre but est de calculer $\phi^\star(\cdot)$

de sorte que l'on puisse quantifier de manière efficace la loi de la limite de SA $\phi^\star(\theta)$ étant donné la loi π de θ. Pour accomplir cette tâche, on choisit la méthode d'expansion de chaos. On suppose que $\phi^\star(\cdot)$ appartient à l'espace de Hilbert des fonctions carrées intégrables par rapport à π. Ensuite on développe une procédure d'approximation stochastique en dimension croissante (surnommée l'algorithme USA, Uncertainty for Stochastic Approximation) pour le calcul des coefficients de l'expansion de chaos de $\phi^\star(\cdot)$ dans une base orthogonale de l'espace de Hilbert. Cela nous fournit une suite d'approximations $\phi^k(\cdot)$ convergeant vers $\phi^\star(\cdot)$ par rapport aux normes de l'espace de Hilbert. Les Chapitres 5-6 sont consacrés au développement et à l'analyse de convergence de l'algorithme USA. La Section 4.1 présente une introduction à la méthode d'approximation stochastique. Dans la Section 4.2 on formule le problème d'incertitude de modèle pour des limites d'approximation stochastique et on donne des exemples de motivation. On discute l'approche d'expansion de chaos dans la Section 4.3. La construction de l'algorithme USA est présentée dans la Section 4.4. La Section 4.5 contient un résumé de nos résultats sur la convergence de USA dans le Chapitre 5 et le taux de convergence dans L^2 obtenu dans le Chapitre 6.

4.1 Introduction aux algorithmes d'approximation stochastique

La méthode d'approximation stochastique est utilisée pour trouver des zéros d'une fonction $z \mapsto h(z)$ donnée sous une forme d'espérance. On suppose que $h(z)$ représente la moyenne des valeurs d'une fonction connue $H(z, V)$ sur des scénarios aléatoires produits par une variable aléatoire V. Le but est de calculer numériquement une solution de $h(z) = \mathbb{E}[H(z, V)] = 0$ en supposant que des simulations i.i.d. de V sont disponibles. Des méthodes classiques déterministes en combinaison avec une approximation par Monte-Carlo de chaque valeur de $h(z)$ peuvent s'avérer trop coûteuses en temps de calcul. L'approximation stochastique a été développée pour résoudre ce problème de manière plus efficace. Initiée par Robbins et Monro [RM51] et Kiefer et Wolfowitz [KW52] dans les années 1950, la théorie d'algorithmes d'approximation stochastique était un sujet de recherche intensive, théorique comme appliquée. La méthode d'approximation stochastique est maintenant activement utilisée dans de nombreux domaines comme l'optimisation, l'estimation paramétrique, le traitement des signaux, le contrôle adaptatif, l'optimisation des systèmes stochastiques par Monte-Carlo (voir [KY97a, BMP90]), les méthodes de descente du gradient stochastique dans l'apprentissage (voir e.g. [BC05, SSS08, BCN17]), l'échantillonneur de Monte- Carlo adaptatif (voir e.g. [HST01, AT08, FMP11, FJLS16, FS00, DVA98]), et le calcul efficace des queues de distribution, parmi d'autres. Une application standard de l'approximation stochastique est le cas où h représente le gradient d'une fonction convexe c donnée par une espérance, c.à.d.

$$h(z) = \nabla_z c(z) = \nabla_z \mathbb{E}[C(z, V)].$$

Dans ce cas SA correspond à la minimisation de c et s'appelle descente du gradient stochastique. Notons que pour utiliser le SA classique, il faut que $\nabla_z \mathbb{E}[C(z, V)] = \mathbb{E}[\nabla_z C(z, V)]$ et que $H := \nabla_z C$ soient connus. Si seule la fonction C est connue, on peut appliquer la méthode de Kiefer-Wolfowitz ([KW52]) qui utilise les différences finies.

Le paradigme de base de SA est l'équation de différence stochastique de la forme

$$z^{n+1} = z^n + \gamma_{n+1} H(z^n, V_n).$$

Ici, z est un paramètre du système ; le vecteur aléatoire $H(z^n, V_n)$ est une observation du scénario du système (produite par une simulation de $V_n \sim V$) avec le paramètre fixé à z^n. On ajuste récursivement le paramètre afin d'atteindre le but asymptotiquement. L'idée principale développée dans [RM51] est que, si le pas γ_n dans les mises à jour du paramètre tend vers 0 de manière appropriée quand $n \to +\infty$, cela produit de la moyennisation implicite qui élimine les effets aléatoires à long terme. De nombreux résultats sur la convergence de SA ont été démontrés dans des cadres différents. Pour donner un exemple, on renvoie le lecteur à [BFP09, Théorème 2.2] car cette version est la plus proche de notre analyse dans le Chapitre 5. Pour ce qui concerne le taux de convergence de SA, les résultats classiques montrent la normalité asymptotique de la suite renormalisée $\gamma_n^{1/2}(z^n - z^\star)$ où $z^\star = \lim_n z^n$ (voir par exemple [Duf97, Chapitre 2]). En utilisant en plus la procédure de moyennisation de Polyak-Ruppert (voir [PJ92]) on peut atteindre le taux de convergence de $1/n$ pour tout $\gamma_n = 1/n^a, a \in (1/2, 1)$. On remarque que la méthode de Monte-Carlo (ayant pour le but de calculer $\mathbb{E}[V]$) est un cas particulier de SA avec $\gamma_n = \frac{1}{n}$ et $H(z, v) = z - v$.

4.2 Quantification d'incertitudes pour des limites d'approximation stochastique: motivation

Le cadre de SA introduit dans la Section 4.1 comprend une variable aléatoire V, pour laquelle le problème principal s'écrit comme $\mathbb{E}[H(z, V)] = 0$. Dans beaucoup d'applications, le choix du modèle pour V est très important. Une spécification exacte du modèle expliquant un phénomène du monde réel doit être choisie avant l'utilisation de la méthode de SA. Assez souvent elle est choisie dans une famille paramétrique de distributions $\{\mu(\theta, \mathrm{d}v) : \theta \in \Theta \subset \mathbb{R}^d\}$, de sorte que le paramètre θ doit être pré-estimé ou fixé par un avis d'expert. Évidemment, une spécification parfaite de θ est rarement possible. Dans certains cas, où on manque d'informations sur le paramètre θ, il est raisonnable de supposer que le modèle de V est incertain. Cette incertitude peut être exprimée par le caractère aléatoire du paramètre θ. Ici on présente quelques problèmes résolus par SA, pour lesquels l'incertitude de modèle constitue un problème pertinent :

- Minimisation de coûts espérés (aussi minimisation de risque ; maximisation d'utilité) sous l'incertitude de modèle. Dans ce cas V modélise un système stochastique et z correspond au paramètre déterminant la stratégie d'interaction avec le système. Ensuite on a $H(z, v) := \nabla_z C(z, v)$ où C est une fonction de coût. Le but est de trouver une stratégie z qui minimise le coût espéré $\mathbb{E}[C(z, V)]$. Sous des hypothèses convenables, ce problème s'écrit comme $\mathbb{E}[\nabla_z C(z, V)] = 0$ et donc peut être résolu par SA. Dans ce cas l'incertitude du modèle est très pertinente. Un exemple d'un tel cadre, parmi beaucoup d'autres, est l'optimisation de portefeuille (où on veut choisir de manière optimale les poids de portefeuille afin de minimiser certains risques espérés ; la variable V représente la distribution aléatoire future du marché, voir par exemple [GMIC14]).

- SA peut être utilisé pour calculer une quantité de distributions, qui porte aussi le nom de Value-at-Risk (VaR) en finance (et plus généralement pour calculer une paire de mesures de risque VaR et CVaR qui sont largement utilisées, voir [BFP09] pour les détails). Dans des applications financières, V représente une valeur future de portefeuille pour laquelle le choix de la loi n'est pas simple. Souvent on manque d'informations et on a besoin d'une approche efficace pour calculer les mesures de risque pour une famille de modèles afin d'analyser le risque du modèle. En particulier, une telle analyse est exigée par les régulateurs financiers.

- Dans certaines applications une approche bayésienne est utilisée pour choisir le modèle de V. Ici on considère une famille paramétrique de lois $\{\mu(\theta, \mathrm{d}v), \theta \in \Theta\}$ avec une loi a priori sur θ. Après avoir observé les données, la loi de θ donne jour à une certaine loi a posteriori π. Dans ce cas le caractère aléatoire de θ implique l'incertitude de modèle pour V.

On exprime l'incertitude de modèle par une dépendance paramétrique $V \sim \mu(\theta, \mathrm{d}v)$ de la distribution de V et on suppose que θ suit une loi π sur Θ. Dans ce cas, la solution $z^\star$ de $\mathbb{E}[H(z, V)] = 0$ où $V \sim \mu(\theta, \mathrm{d}v)$ pour θ fixé, dépend de θ, ce qui implique $z^\star = \phi^\star(\theta)$ pour une fonction $\phi^\star(\cdot)$. Notre but principal est de quantifier la loi de la solution (ou la limite de SA) $\{\phi^\star(\theta) : \theta \sim \pi\}$ par une approximation de $\phi^\star(\cdot)$ dans un espace fonctionnel bien choisi. On note que dans les exemples ci-dessus, une alternative plus simple pourrait être d'augmenter la complexité du modèle de V en faisant V suivre la loi marginale de $(\theta, V) \sim \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta)$. Ce problème peut être ensuite résolu par des méthodes standard. Cependant cette approche donne beaucoup moins d'informations, typiquement une seule valeur (qui représente une sorte de moyenne des solutions parmi les différents choix du modèle), alors que l'on est intéressé par la quantification de la loi complète de $\phi^\star(\theta)$.

4.3 Formulation du problème. Expansion de chaos

Maintenant on va formaliser le problème introduit dans la Section 4.2. On commence par l'équation $\mathbb{E}[H(z, V)] = 0$, où V est une variable aléatoire à valeurs dans un espace métrique $\mathcal{V}$. On impose de l'incertitude sous une forme de dépendance paramétrique $V \sim \mu(\theta, \mathrm{d}v)$ (pour un noyau de transition μ de Θ à $\mathcal{V}$). Le paramètre θ lui-même suit une loi de probabilité connue $\pi(\mathrm{d}\theta)$ sur $\Theta \subset \mathbb{R}^d$. On permet aussi de l'incertitude sur la fonction H, sous une forme de dépendance en θ, de sorte que l'on prend $H : \mathbb{R}^q \times \mathcal{V} \times \Theta \to \mathbb{R}^q$. On note par L_2^π l'espace de Hilbert des fonctions carrées intégrables par rapport à π. Le problème de la quantification d'incertitudes pour des limites de SA s'écrit maintenant formellement comme

$$\text{Find } \phi^\star \text{ in } L_2^\pi \text{ such that } \int_{\mathcal{V}} H(\phi^\star(\theta), v, \theta)\mu(\theta, \mathrm{d}v) = 0, \qquad \pi\text{-a.s.} \tag{4.1}$$

Une approche naïve pour accéder à la distribution $\{\phi^\star(\theta), \theta \sim \pi\}$ est de simuler $\theta \sim \pi$ et ensuite, pour chaque valeur simulée θ_i, de lancer une procédure d'approximation stochastique avec le paramètre du modèle fixé à θ_i qui donne une approximation $\widehat{\phi^\star(\theta_i)}$ de $\phi^\star(\theta_i)$. Une technique plus intelligente est d'approcher la fonction $\phi^\star(\cdot)$ de telle manière qu'ensuite seulement, la simulation de $\theta \sim \pi$ soit nécessaire. Cela peut être fait par l'expansion de chaos

qui remonte à Wiener [Wie38] et qui a été développée dans les domaines d'ingénierie et de la quantification d'incertitudes dans les années 2000 (voir [GS03, LK10] et les références y contenues). Cette technique, connue aussi comme la méthode spectrale, consiste en la projection de la fonction inconnue $\phi^\star : \Theta \mapsto \mathbb{R}^q$ sur une base orthogonale $\{\theta \mapsto B_i(\theta), i \in \mathbb{N}\}$ de l'espace L^2 par rapport à la loi π et le calcul des coefficients de sa décomposition.

$$\phi^\star = \sum_{i \geq 0} u_i^\star B_i. \tag{4.2}$$

Dans le cas le plus commun où $B_0 \equiv 1$, une fois que les coefficients $\{u_i^\star, i \geq 0\}$ sont calculés, l'espérance et la matrice de la variance-covariance de $\{\phi^\star(\theta), \theta \sim \pi\}$ sont disponibles directement sous une forme

$$\mathbb{E}_{\theta\sim\pi}[\phi^\star(\theta)] = u_0^\star \text{ and } \mathbb{V}\mathrm{ar}_{\theta\sim\pi}(\phi^\star(\theta)) = \sum_{i\geq 1} u_i^\star (u_i^\star)^\top.$$

Dans le cas de base polynomiale, les moments plus élevés sont aussi calculables, voir [LK10, Appendix C].

Une méthode naïve imbriquée pour calculer les coefficients $u_i^\star$ peut consister en une simulation de $\theta_1, \cdots, \theta_N \sim \pi$ et ensuite, pour chaque θ_i, d'un tour de SA pour obtenir $\widehat{\phi^\star(\theta_i)}$. Ainsi on retrouve une approximation de $u_i^\star$ donnée par

$$\hat{u}_i := \frac{1}{N} \sum_{i=1}^{N} \widehat{\phi^\star(\theta_i)} B_i(\theta_i). \tag{4.3}$$

Cependant, cette méthode demande des calculs imbriqués, et par conséquent, elle est peu efficace. On peut l'illustrer à partir d'un exemple simple où SA est réduit à la méthode de Monte-Carlo (un cas particulier "linéaire" où $H(z, v) = z - v$). Dans cet exemple, la méthode naïve entraîne une procédure de MC en deux étapes qui convergent deux fois plus lentement. La façon appropriée de faire le calcul dans ce cas serait d'approcher la moyenne $\int_{\Theta\times\mathcal{V}} v\mu(\theta, \mathrm{d}v)B_i(\theta)\pi(\mathrm{d}\theta)$ directement par des simulations i.i.d. de $(\theta, V) \sim \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta)$. À cet égard, on attend que l'approche naïve (4.3) peut être aussi largement améliorée dans le cas général ; notre but est de développer un algorithme pour calculer les coefficients $u_i^\star$ en utilisant un mélange efficace de la simulation de $\theta \sim \pi$ et les simulations de l'algorithme d'approximation stochastique. Dans l'introduction du Chapitre 5 on montre qu'une procédure en dimension croissante (c.à.d. avec le niveau de troncature progressivement augmenté) est nécessaire à cause de la non linéarité du problème. De telles méthodes d'expansion de chaos sont difficiles à analyser. Même dans le cas où la fonction $\phi^\star$ est connue (ici le calcul des coefficients individuels $u_i^\star$ dans (4.2) est évident par Monte- Carlo) la convergence globale de la méthode où le nombre de coefficients croît vers l'infini nécessite un réglage de la vitesse de croissance de leur nombre et du nombre de simulations. Le développement d'une telle méthode dans le cas plus général de SA et l'analyse de sa convergence sont donc un problème ambitieux.

4.4 L'algorithme USA

Dans le Chapitre 5 on développe une procédure d'approximation stochastique pour calculer les coefficients de $\phi^\star = \sum_{i\geq 0} u_i^\star B_i$ de sorte que chaque itération appartient à un espace de dimension finie et la dimension de ces espaces tend vers l'infini. Ici on présente une version un peu simplifiée de cette procédure surnommée l'algorithme USA, qui résout le problème (4.1). Elle est complètement constructive et facile à réaliser. Plus de détails sur la motivation pour la construction de cet algorithme peuvent être trouvés dans le Chapitre 5.

Input: Sequences $\{\gamma_k, k \geq 1\}$, $\{m_k, k \geq 1\}$, $\{M_k, k \geq 1\}$, $K \in \mathbb{N}$, $\{u_i^0, i = 0, \ldots, m_0\}$
for $k = 0$ **to** $K - 1$, **do**
  sample $(\theta_{k+1}^s, V_{k+1}^s), s = 1 \ldots, M_{k+1}$, under the distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$; for $i > m_{k+1}$ define $u_i^k = 0$ **for** $i = 0$ **to** m_{k+1}, **do**
    $u_i^{k+1} = u_i^k - \gamma_{k+1} M_{k+1}^{-1} \sum_{s=1}^{M_{k+1}} H\left(\sum_{j=0}^{m_k} u_j^k B_j(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) B_i(\theta_{k+1}^s)$
Output: The vector $\{u_i^K, i = 0, \ldots, m_K\}$.

Algorithm 2: L'algorithme USA pour le calcul des coefficients de la décomposition de $\phi^\star$.

Les entrées de l'algorithme sont : une suite positive de pas $\{\gamma_k, k \geq 1\}$, deux suites à valeurs entières positives $\{m_k, k \geq 1\}$ et $\{M_k, k \geq 1\}$ qui correspond au nombre de coefficients non nuls dans l'approximation de $\phi^\star$ et le nombre de simulations de Monte-Carlo du couple $(\theta, V) \sim \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$ à chaque itération k, une valeur initiale $u_0 \in \mathbb{R}^{m_0}$ et un nombre total d'itérations K. La sortie de l'algorithme est une suite $u^K = \{u_i^K, i \leq m_K\}$ qui approche la solution $u^\star$. L'approximation correspondante ϕ^K de la solution $\phi^\star$ du problème (4.1) est donc donnée par

$$\phi^K := \sum_{i=0}^{m_K} u_i^K B_i. \tag{4.4}$$

Ensuite cette approximation permet de calculer facilement toutes les statistiques de la limite incertaine de SA $\phi^\star(\theta)$ en utilisant seulement la simulation de $\theta \sim \pi$.

4.5 Résumé des résultats des Chapitres 5 and 6

Dans le Chapitre 5, en plus de la motivation plus détaillée de la construction de l'algorithme USA 2, on démontre sa convergence p.s. et dans $L^p, (p < 2)$. Dans le Théorème 5.3.5 on montre sous des hypothèses explicites et claires que pour la suite $\{\phi^k : k \geq 0\}$ produite par (4.4) on a

$$\lim_{k\to\infty} \left\|\phi^k - \phi^\infty\right\|_\pi = 0 \text{ p.s.}, \qquad \forall p \in (0,2) \ \lim_{k\to\infty} \mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^p\right] = 0.$$

où ϕ^∞ est une variable aléatoire à valeur dans l'ensemble des solutions du problème (4.1) (ou tout simplement $\phi^\infty = \phi^\star$ est la solution $\phi^\star$ de (4.1) s'il est unique). L'algorithme 2 est une procédure d'approximation stochastique dans l'espace de Hilbert de dimension infinie l_2, ce qui rend l'analyse de la convergence relativement non triviale. On montre que nos résultats sont originaux car ils ne sont pas couverts par les travaux existants sur le SA

dans des espaces de Hilbert. Parmi eux, on est uniquement intéressé par SA en dimension croissante (mais où chaque itération est de dimension finie, ce qui rend l'algorithme réalisable en pratique). Ce type de SA a été étudié dans [Nix84, Gol88, Yin92] (dans un cas particulier du bruit indépendant, c.à.d. $H(z,V) = \tilde{H}(z)+V$). Plus récemment, dans [CW02] les auteurs montrent des résultats de convergence et de normalité asymptotique pour SA en dimension croissante dans un cadre plus général. Cependant, ces résultats sont démontrés sous des conditions assez abstraites. De plus [Nix84, CW02] considèrent un terme de bruit de la forme $\tilde{H}(\phi^k, V_{k+1})$ avec $\tilde{H} : \mathcal{H} \times \mathcal{V} \to \mathcal{H}$ et une seule loi pour les V_n. Par contre, dans notre cas, $H(\phi^k(\cdot), V_{k+1}, \cdot)$ peut être simulé seulement θ par θ, car la loi de V_{k+1} peut dépendre de θ. Enfin, certaines hypothèses clés des travaux ne sont pas vraies dans notre cadre, voir le Chapitre 5 pour une discussion plus complète et des contre-exemples (par exemple dans la Remarque 5.3.3). Dans la Section 5.5 on effectue une analyse numérique détaillée, incluant la discussion du réglage des paramètres pour la mise en pratique de l'algorithme USA. Dans le Chapitre 6 on étudie le taux de convergence dans L^2 de la suite $\{\phi^k, k \geq 0\}$ produite par l'algorithme USA. Notre résultat principal dans le Théorème 6.3.1 donne explicitement $\alpha > 0$ tel que pour une constante $C_\alpha > 0$ on a pour tout $k \geq 0$

$$\mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right] \leq C_\alpha \gamma_k^\alpha. \tag{4.5}$$

Un contrôle de la forme γ_k^α a été motivé par des résultats similaires dans le cas de dimension finie où typiquement l'erreur carrée de SA est d'ordre $O(\gamma_k)$, c.à.d. $\alpha = 1$ (voir par exemple [Duf97, Chapitre 2]). Dans les résultats de dimension finie sur le taux de convergence de SA, la vitesse de convergence dépend uniquement de la suite de pas γ_k. Dans notre cadre, l'exposant α dans (4.5) va dépendre de façon non triviale du modèle, de la régularité de $\phi^\star$, du choix des fonctions de base et des paramètres de l'algorithme USA. La connaissance de cette dépendance joue un rôle important dans le réglage correct de l'algorithme afin de garantir la convergence dans L^2 avec le meilleur taux possible, étant donné le modèle. On montre comment les résultats obtenus justifient l'optimalité du choix heuristique de la vitesse de croissance de dimension dans la Section 6.3.2. Encore une fois, on conclut que les travaux précédents sur le SA en dimension infinie ne peuvent pas être appliqués à l'analyse de taux de convergence de USA, voir l'introduction du Chapitre 6. Cela souligne la nouveauté de nos résultats.

En dehors de la quantification d'incertitudes, notre approche peut être appliquée à l'analyse de sensibilité par rapport à θ, ou la quasi-régression dans le sens de reconstruction complète d'une fonction, par exemple dans le contexte du calcul de Monte-Carlo imbriqué qui comprend une fonction intérieure non linéaire $\theta \mapsto \phi^\star(\theta)$.

Part I

Discretization of processes at stopping times

Chapter 1

Optimal discretization of stochastic integral driven by general Brownian semimartingale

Contents

1.1 Introduction

Statement of the problem. In this work we consider the problem of finding a finite sequence of optimal stopping times $\mathcal{T}^n = \{0 = \tau_0^n < \tau_1^n < \cdots < \tau_{N_T^n}^n = T\}$ which minimizes the renormalized quadratic variation of the discretization error of the stochastic integral

$$Z_s^n = \int_0^s v(t, S_t) \cdot \mathrm{d}S_t - \sum_{\tau_{i-1}^n < s} v(\tau_{i-1}^n, S_{\tau_{i-1}^n}) \cdot (S_{\tau_i^n \wedge s} - S_{\tau_{i-1}^n}), \tag{1.1.1}$$

where S is a d-dimensional continuous Brownian semimartingale and $v(t,x)$ is a $\mathbb{R}^d$-valued continuous function. Here $T \in (0,+\infty)$ is fixed. The number of stopping times N_T^n is allowed to be random.

The almost sure minimization of Z_T^n is hopeless since after suitable renormalization and under some mild assumptions on the model, Z_T^n weakly converges to a mixture of Gaussian random variables (see [Roo80][KP91][JP12]). Alternatively we aim at minimizing a.s. the product

$$N_T^n \langle Z^n \rangle_T. \tag{1.1.2}$$

The choice of this minimization criterion is inspired by the fact that in many particular cases with deterministic discretization times, we have $\mathbb{E}\left(\langle Z^n \rangle_T\right) \sim \text{Const}/N_T^n$ as $N_T^n \to +\infty$. For example, in the one-dimensional Brownian motion case with $v(t,x) = x$ the value of $\mathbb{E}(\langle Z^n \rangle_T)$ for the regular mesh of size n may be calculated exactly and is equal to $\frac{1}{2n}$. For more general S and v satisfying fractional regularity conditions [GG04], the error $\mathbb{E}(\langle Z^n \rangle_T)$ is still of magnitude Cst/n by appropriately choosing n deterministic times on $[0,T]$.

Background results. The problem of optimizing the discretization times was initially considered in a different framework: simulation of diffusion processes. In [HMGR01] the authors study the optimal discretization times for the simulation of a one-dimensional diffusion X via the Euler/Milshtein schemes, where the discretization times adapt to the local properties of every single trajectory. They consider three different schemes and analyze their L_2 errors (in time and ω):

a) *A simplified Adaptive scheme* $\hat{X}_h^{**}$, for which the sequence of discretization times $(\tau_i)_{1 \le i \le \nu}$ is such that each τ_i is a measurable function of the previously simulated values of the Brownian motions $W_{\tau_1}, \ldots, W_{\tau_{i-1}}$, and Euler and Milshtein schemes with two appropriate time scales are combined to approximate X. This method is of varying cardinality since the number ν of times is random. Observe that $(\tau_i)_i$ are stopping times but they belong to the subclass of strongly predictable times (see [JP12, Chapter 14]), along which moments of martingale increments are easier to compute.

b) *An Adaptive scheme* $\hat{X}_h^*$ *with discretization times of fixed cardinality.* To control the number of times, a first monitoring of an approximation of X is considered in order to decide where to refine the discretization whilst maintaining a given number of time points. Therefore, the discretization times are somehow anticipative and they are not stopping times.

c) *An Adaptive scheme* $\hat{X}_h$ *with path-independent step-size Control*, as a variant of $\hat{X}_h^*$ where the monitoring is made in mean and not on the specific path X to simulate.

In [HMGR01, Theorem 1], the authors prove the asymptotic superiority of $\hat{X}_h^{**}$ over the two other schemes and [HMGR01, Theorem 2] states the asymptotic optimality of each scheme within its own class. For the latter optimality result, the criterion used for the optimization is the renormalized L_2-error. Despite the similarities between our current work and theirs, there are significant differences that we shall stress. First, we consider discretization of stochastic integrals and not of diffusion processes, therefore the objectives are quite different. Second,

we study the case of general multi-dimensional continuous Brownian semimartingale whereas [HMGR01] handles the case of diffusion in $d = 1$ and [MG02, Chapter III] deals with $d \geq 1$ under commutative noise assumption. Third, we allow optimization over a quite large class of stopping times, see examples of Remark 1.2.2 illustrating this fact.

Besides, the study of minimization problems for stochastic integral discretization has been initiated by [Fuk11a] in dimension $d = 1$, but instead of (1.1.2) the author considers a criterion in expectation for both terms, i.e. $\mathbb{E}\left(N_T^n\right)\mathbb{E}\left(\langle Z^n\rangle_T\right)$. However, if $n \to +\infty$ denotes an asymptotic parameter (defined later), observe that

$$\liminf_{n\to+\infty} \mathbb{E}\left(N_T^n\right)\mathbb{E}\left(\langle Z^n\rangle_T\right) \underset{\text{Cauchy}-\text{Schwarz ineq.}}{\geq} \liminf_{n\to+\infty} \left[\mathbb{E}\left(\sqrt{N_T^n\langle Z^n\rangle_T}\right)\right]^2 \tag{1.1.3}$$

$$\underset{\text{Fatou lemma}}{\geq} \left[\mathbb{E}\left(\sqrt{\liminf_{n\to+\infty} N_T^n\langle Z^n\rangle_T}\right)\right]^2. \tag{1.1.4}$$

Since the solution to the problem of a.s. minimizing (1.1.2) exists (see Theorem 1.5.2) and is such that N_T^n and $\langle Z^n\rangle_T$ are asymptotically proportional (see the limits (1.5.14) and (1.5.15)), the above inequalities can be turned into equalities (with a little of technical work) and therefore, we get for free a solution to minimizing asymptotically $\mathbb{E}\left(N_T^n\right)\mathbb{E}\left(\langle Z^n\rangle_T\right)$, however with substantially more information.

The pathwise minimization of (1.1.2) has been addressed in a multi-dimensional setting $d \geq 1$, in [GL14a]: the authors assume that S is a local martingale and the lower bound is achieved under stringent conditions of v (essentially its Jacobian matrix $D_x v$ is invertible). These assumptions are restrictive and we aim at relaxing the hypotheses and strengthening the optimality results. This requires to develop new arguments presented in this work.

As an extra motivation for this theoretical study, we refer to the recent work of Hairer et al. [HHJ15], which highlights that discretization schemes for stochastic differential equations using deterministic grid may surprisingly converge very slowly in L_2-norm. Actually any slow rate is possible [JMY16]. These amazing results give a strong incentive for studying discretization problems with stochastic grids and pathwise criterion. Applications of the current results to pathwise-optimal discretization of SDEs are left to future research.

Our contributions. In the current work, we prove optimality results in a much larger setting than previously afforded in the literature.

- First, we allow S to be a general Brownian semimartingale $S = A + M$, while in [GL14a] S is essentially a local Brownian martingale ($A = 0, M = \int_0^{\cdot} \sigma_s \mathrm{d}B_s$). Actually, considering the existence of the *finite variation term* A modifies a priori significantly the definition of admissible discretization strategies (see the definition $\mathbf{(A}_S^{\text{osc.}}\mathbf{)}$ later) and restricts the set of available tools to analyze them. Our first contribution is to establish that admissible strategies for the semimartingale S and for its local martingale part M are the same: see Theorem 1.3.4. This is a non-trivial result. This allows to transfer a priori estimates available in the martingale case (Lemmas 1.3.2 and 1.3.3) to our extended setting, this is instrumental for the subsequent analysis.

- Second, the martingale part of S can be degenerate in our setting, whereas a stronger a.s. ellipticity (on σ) is considered in [GL14a]. This allows to consider partially degenerate models like
$$S_t = (\tilde{S}_t, \int_0^t \tilde{S}_s \mathrm{d}s)$$
or other SDEs with vanishing diffusion coefficient (see Subsection 1.5.3 for examples). Also $D_x v(t, S_t)$ may be not invertible in our work. This second set of improvements requires a quite delicate analysis, which constitutes the core of this work. Actually the possible degeneracy lets us lose some continuity property (in particular because we need to consider the inverse σ^{-1}) and some convergence properties. To overcome these issues, we assume that in a sense, σ_t and $D_x v(t, S_t)$ are not zero simultaneously: for a precise statement, see Assumption ($\mathbf{H}_C$) or a weaker Assumption ($\mathbf{H}_\Lambda$). These are quite mild conditions.

  The ability to treat the non-elliptic case is fundamental for applications as well:

  (a) Regarding financial applications, see for example [Fuk11a, GL14a], minimizing $\langle Z^n \rangle_T$ is related to better hedge market risks. In that context, the treatment of degenerate case appears to be important. Though the covariance matrix of a group of asset returns is usually non-degenerate, it may have some very small eigenvalues [BP11]. The reason is that typically a large portfolio of financial assets is driven by a smaller number of significant factors, while the other degrees of freedom represent low-variance noise. Thus the inversion of the covariance matrix is often seen as undesirable by practitioners, if no robustness analysis is provided. Our study of the degenerate case justifies in a way the robustness of the optimal discretization algorithm when the diffusion coefficient is degenerate or close to being degenerate.

  (b) Some important examples of diffusion models with degenerate diffusion coefficient come as well from random mechanics, see [KS12] for an overview. Typically, a body is modeled by its position X and its velocity V: it is subjected to random forces, so that due to the second Newton law of motion, its dynamics writes
$$\begin{cases} X_t = X_0 + \int_0^t V_s \mathrm{d}s, \\ V_t = V_0 + \int_0^t \phi(X_s, V_s)\mathrm{d}s + \int_0^t \psi(X_s, V_s)\mathrm{d}W_s. \end{cases} \tag{1.1.5}$$
  In [LPC11], these equations describe the response of structural systems subjected to severe environmental loads (like earthquakes, strong winds, recurrent waves...). The authors study examples like seismic-excited ten-storey building (see [LPC11, Section 5]) where they propose to optimally control the structure by activating tendons, in order to compensate external forces. They derive a continuous-time optimal control, but in practice, only discrete-time controls can be applied. Our study gives a theoretical framework to determine when to apply the controls in order to minimize the deviation from optimally-controlled building.
  In [Tal02], the author studies the approximation of stochastic Hamiltonian sys-

tems of the form (1.1.5). The author emphasizes the technical difficulty of the analysis coming from the polynomial growth of the coefficients and the degeneracy of the infinitesimal generators. In our context of optimal discretization problem, our a.s. analysis allows for arbitrary growth conditions on the coefficients.

- Third, we provide a strategy $\mathcal{T}^n$ attaining the lower bound, while in [GL14a], only a μ-optimal strategy (with μ small) is designed. Informally, the natural candidate for optimality is a sequence of hitting times by S of random ellipsoids which characteristics depend on $D_x v$ and S. However, in general and in particular because of the degenerate setting on σ_t and $D_x v(t, S_t)$, this strategy is not admissible (ellipsoids may be flat or infinite). Alternatively, we prove that a suitable perturbation makes the strategy admissible, without altering its asymptotic optimality.

Our main result (Theorem 1.5.2) states that an optimal strategy is of the form

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Lambda^{(n)}_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) \geq \widetilde{\varepsilon}_n\} \wedge T, \end{cases}$$

for a sequence $\widetilde{\varepsilon}_n \to 0$, where $\Lambda_t^{(n)}$ is a suitable perturbation of $\Lambda_t := (\sigma_t^\dagger)^\mathsf{T} X_t \sigma_t^\dagger$ (where $\mathcal{M}^\dagger$ is the pseudo-inverse matrix of $\mathcal{M}$), and X_t is the symmetric non-negative definite matrix solution to the equation

$$2\,\mathrm{Tr}(X_t)X_t + 4X_t^2 = \sigma_t^\mathsf{T}(D_x v(t, S_t))^\mathsf{T} \sigma_t \sigma_t^\mathsf{T} D_x v(t, S_t) \sigma_t.$$

Additionally the asymptotic lower bound to (1.1.2) is $\left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2$.

Organisation of the chapter. In Section 1.2, we define the model and the admissible strategies under study. In Section 1.3, we state and establish crucial properties of admissible strategies. The minimization of (1.1.2) is studied in Section 1.4, and designing an optimal strategy is made in Section 1.5. We also present a few examples and a numerical experiment in Subsection 1.5.3. Technical results are postponed to Appendix.

Notation used throughout this work.

- We denote by $x \cdot y$ the scalar product between two vectors x and y and by $|x| = (x \cdot x)^{\frac{1}{2}}$ the Euclidean norm of x. The induced norm of a $m \times d$ -matrix is denoted by $|A| := \sup_{x \in \mathbb{R}^d : |x|=1} |Ax|$.

- The transposition of a matrix A is denoted by A^T; we denote by $\mathrm{Tr}(A)$ the trace of a square matrix A; Id_d stands for the identity matrix of size d.

- $\mathcal{S}^d(\mathbb{R}), \mathcal{S}^d_+(\mathbb{R})$ and $\mathcal{S}^d_{++}(\mathbb{R})$ are respectively the sets of symmetric, symmetric non-negative definite and symmetric positive-definite $d \times d$ matrices with real coefficients.

- For $A \in \mathcal{S}^d(\mathbb{R})$ we denote $\Lambda(A) = (\lambda_1(A), \dots, \lambda_d(A))$ the eigenvalues of A placed in decreasing order, we set $\lambda_{\min}(A) := \lambda_d(A)$ and $\lambda_{\max}(A) := \lambda_1(A)$.
- We denote by $\mathrm{Diag}(a_1, \dots, a_d)$ the square matrix of size d with diagonal entries $a_1, \dots, a_d$.
- For the partial derivatives of a function $f(t,x)$ we write

$$D_t f(t,x) = \frac{\partial f}{\partial t}(t,x),\ D_{x_i} f(t,x) = \frac{\partial f}{\partial x_i}(t,x),\ D^2_{x_i x_j} f(t,x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(t,x).$$

- For a $\mathbb{R}^d$-valued semimartingale S we denote $\langle S \rangle_t$ its matrix of cross-variations $(\langle S^i, S^j \rangle_t)_{1 \le i,j \le d}$.
- We sometimes write f_t for $f(t, S_t)$ where S is a semimartingale and f is some function.
- For a given sequence of stopping times $\mathcal{T}^n$, the last stopping time before $t \le T$ is defined by $\phi(t) = \max\{\tau^n_j : \tau^n_j < t\}$. We omit to indicate the dependence on n. Furthermore for a process $(f_t)_{0 \le t \le T}$ we write $\Delta f_t := f_t - f_{\phi(t-)}$. Besides we set $\Delta_t := t - \phi(t-)$ and $\Delta \tau^n_i := \tau^n_i - \tau^n_{i-1}$.
- C_0 stands for a a.s. finite non-negative random variable, which may change from line to line.

1.2 Model and strategies

1.2.1 Probabilistic model: assumptions

Let $T > 0$ and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$ be a filtered probability space supporting a d-dimensional Brownian motion $B = (B^i)_{1 \le i \le d}$ defined on $[0,T]$, where $(\mathcal{F}_t)_{0 \le t \le T}$ is the $\mathbb{P}$-augmented natural filtration of B and $\mathcal{F} = \mathcal{F}_T$. Let

$$(\alpha, \theta_\sigma) \in (\frac{1}{2}, 1] \times (0, 1] \tag{1.2.1}$$

be some regularity parameters and let $(S_t)_{0 \le t \le T}$ be a d-dimensional continuous semimartingale of the form

$$S_t = A_t + M_t, \qquad 0 \le t \le T, \tag{1.2.2}$$

where the processes A and M satisfy the following hypotheses.

($\mathbf{H}_A$) The process A is continuous, adapted and of finite variation, and satisfies

$$|A_t - A_s| \le C_0 |t-s|^\alpha \qquad \forall s,t \in [0,T] \quad \text{a.s..} \tag{$\mathbf{H}_A$}$$

($\mathbf{H}_M$) The process M is a continuous local martingale of the form

$$M_t = \int_0^t \sigma_s \mathrm{d}B_s, \quad 0 \le t \le T, \tag{$\mathbf{H}_M$}$$

where σ is a continuous adapted $d \times d$-matrix valued process, such that the value σ_t is a.s. non-zero for any $t \in [0,T]$, and

$$|\sigma_t - \sigma_s| \leq C_0 |t-s|^{\theta_\sigma/2} \qquad \forall s,t \in [0,T] \quad \text{a.s..}$$

Furthermore, we assume that the function v, involved in (1.1.1), is a $\mathcal{C}^{1,2}([0,T) \times \mathbb{R}^d)$ function with values in $\mathbb{R}^d$. For applications like in [GL14a], we shall allow its derivatives in uniform norm (in space) to explode as $t \to T$, whilst remaining bounded a.s. in an infinitesimal tube centered at $(t, S_t)_{0 \leq t < T}$. This is stated precisely in what follows.

($\mathbf{H}_v$) Let $\mathcal{D} \in \{D_{x_j}, D^2_{x_j x_k}, D_t : 1 \leq j,k \leq d\}$, then

$$\mathbb{P}\left(\lim_{\delta \to 0} \sup_{0 \leq t < T} \sup_{|x - S_t| \leq \delta} |\mathcal{D} v(t,x)| < +\infty \right) = 1. \tag{$\mathbf{H}_v$}$$

1.2.2 Class $\mathcal{T}^{\text{adm.}}$ of admissible sequences of strategies

Now we define the class of strategies under consideration. As the optimality in our problem is achieved asymptotically as a parameter $n \to +\infty$, a strategy is naturally indexed by $n \in \mathbb{N}$: a *strategy* is a finite sequence of increasing stopping times

$$\mathcal{T}^n := \{\tau^n_0 = 0 < \cdots < \tau^n_i < \cdots < \tau^n_{N^n_T} = T\}, \quad \text{with } N^n_T < +\infty \text{ a.s..}$$

We now define the appropriate asymptotic framework. Let $(\varepsilon_n)_{n \in \mathbb{N}}$ be a sequence of positive deterministic real numbers such that

$$\sum_{n \geq 0} \varepsilon_n^2 < +\infty.$$

In the following, all convergences are taken as $n \to +\infty$. The above summability enables to derive a.s. convergence results: alternatively, had we assumed only $\varepsilon_n \to 0$, using a subsequence-based argument (see [GL14b, Section 2.2]) we would get convergences in probability.

On the one hand the parameter ε_n controls the oscillations of S between two successive stopping times in $\mathcal{T}^n$.

($\mathbf{A}^{\text{osc.}}_S$) The following non-negative random variable is a.s. finite:

$$\sup_{n \geq 0} \left(\varepsilon_n^{-2} \sup_{1 \leq i \leq N^n_T} \sup_{t \in (\tau^n_{i-1}, \tau^n_i]} |S_t - S_{\tau^n_{i-1}}|^2 \right) < +\infty.$$

Here the lower argument in the assumption ($\mathbf{A}^{\text{osc.}}_{\cdot}$) refers explicitly to the process at hand. On the other hand $\varepsilon_n^{-2\rho_N}$ (for some $\rho_N \geq 1$) upper bounds up to a constant the number of stopping times in the strategy $\mathcal{T}^n$.

($\mathbf{A}_N$) The following non-negative random variable is a.s. finite:

$$\sup_{n\geq 0}(\varepsilon_n^{2\rho_N} N_T^n) < +\infty.$$

In the above, ρ_N is a given parameter satisfying

$$1 \leq \rho_N < \left(1+\frac{\theta_\sigma}{2}\right) \wedge \frac{4}{3} \wedge \left(\frac{1}{2}+\alpha\right). \tag{1.2.3}$$

where (α, θ_σ) are given in (1.2.1).

Definition 1.2.1. *A sequence of strategies $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ is admissible for the process S and the parameters $(\varepsilon_n)_{n\in\mathbb{N}}$ and ρ_N if it fulfills the hypotheses $(\mathbf{A}_S^{\text{osc.}})$ and $(\mathbf{A}_N)$. The set of admissible sequences is denoted by $\mathcal{T}_S^{\text{adm.}}$.*

The larger ρ_N, the wider the class of strategies under consideration.

Remark 1.2.2. *The notion of admissible sequence is quite general, in particular, it includes the following two wide families of random grids.*

i) Let $\rho \in (0,1)$ and let $(\varepsilon_n)_{n\geq 0}$ be a deterministic sequence such that $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. Consider $\mathcal{T} = \{\mathcal{T}^n\}_{n\geq 0}$ where each $\mathcal{T}^n = (\tau_i^n)_{0\leq i\leq N_T^n}$ is a sequence of stopping times (with N_T^n possibly random) and such that

$$C^{-1}\varepsilon_n^{\frac{2}{(1-\rho)}} \leq \min_{1\leq i\leq N_T^n}\Delta\tau_i^n \leq \max_{1\leq i\leq N_T^n}\Delta\tau_i^n \leq C\varepsilon_n^{\frac{2}{(1-\rho)}}, \quad n\geq 0, \quad a.s.,$$

for an a.s. finite positive random variable $C > 0$. This example contains in particular the sequences of deterministic grids for which the time steps are controlled from below and from above (like those of [HMGR01] used for building $\hat{X}_h^{**}$ mentioned in introduction), and for which the step size tends to zero fast enough.
Let us check $(\mathbf{A}_S^{\text{osc.}})$and $(\mathbf{A}_N)$. First, note that S is a.s. Hölder continuous on $[0,T]$ with exponent $\frac{1-\rho}{2}$: this is a consequence of $(\mathbf{H}_A)$ for the finite-variation component A and of [BY82a, Theorem 5.1] for the martingale component M under the assumption $(\mathbf{H}_M)$. Therefore, a.s. for each $n \geq 0$

$$\sup_{1\leq i\leq N_T^n}\sup_{t\in[\tau_{i-1}^n,\tau_i^n]}|S_t - S_{\tau_{i-1}^n}| \leq C_S\Big[\max_{1\leq i\leq N_T^n}\Delta\tau_i^n\Big]^{\frac{1-\rho}{2}} \leq C_S C^{\frac{1-\rho}{2}}\varepsilon_n.$$

Furthermore,

$$N_T^n \leq \frac{T}{\min_{1\leq i\leq N_T^n}\Delta\tau_i^n} \leq TC\varepsilon_n^{-\frac{2}{(1-\rho)}}$$

so that $(\mathbf{A}_N)$ is verified with $2\rho_N = 2(1-\rho)$ provided that we take ρ small enough to satisfy the upper bound (1.2.3). Thus the sequence of strategies $\mathcal{T}$ is admissible for $(\varepsilon_n)_{n\geq 0}$ and ρ_N given above.

ii) Consider a sequence of adapted random processes $\{D_t^n : 0 \le t \le T\}$ *where each* D_t^n *is an open set such that*

$$B(0, C_1\varepsilon_n) \subset D_t^n \subset B(0, C_2\varepsilon_n)$$

for some a.s. finite positive random variables C_1, C_2*, here* $B(0,r)$ *denotes the ball centered at 0 with radius* r*. Here again the deterministic sequence* $(\varepsilon_n)_{n\ge 0}$ *is such that* $\sum_{n\ge 0}\varepsilon_n^2 < +\infty$*. Define the sequence of strategies* $\mathcal{T} = \{\mathcal{T}^n\}_{n\ge 0}$ *with* $\mathcal{T}^n = (\tau_i^n)_{0\le i\le N_T^n}$ *as follows:* $\tau_0^n = 0$ *and for* $i \ge 1$

$$\tau_i^n = \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin D_{\tau_{i-1}^n}^n\} \wedge T.$$

In other words, we consider exit times of random sets of size ε_n*. The assumption* $(\mathbf{A}_S^{\text{osc.}})$ *follows from the definition of* $\mathcal{T}^n$*:*

$$\sup_{1\le i\le N_T^n}\ \sup_{t\in[\tau_{i-1}^n,\tau_i^n]} |S_t - S_{\tau_{i-1}^n}| \le C_2\varepsilon_n.$$

Further to check $(\mathbf{A}_N)$*, we write (using Proposition 1.3.9)*

$$C_1^2\varepsilon_n^2 N_T^n \le C_1^2\varepsilon_n^2 + \sum_{\tau_{i-1}^n<T} |\Delta S_{\tau_i^n}|^2 \underset{n\to+\infty}{\to} \mathrm{Tr}(\langle S\rangle_T) < +\infty \qquad a.s.$$

This proves the admissibility of $\mathcal{T}$*. A particular case is the ellipsoid exit times, see [GL14a, Proposition 2.4].*

1.3 General results for admissible strategies

This section gathers preliminary results, needed to establish the subsequent main results. In Subsection 1.3.1, we recall without proof some estimates about the mesh size $\sup_{1\le i\le N_T^n}\Delta\tau_i^n$ of the time grid $\mathcal{T}^n$ simultaneously for any n, as well as bounds on (local) martingales depending on n. This is preparatory for Subsection 1.3.2 where we establish an important result: in our setting, admissible sequences of strategies for S and M are the same. Last in Subsection 1.3.3, we establish the a.s. convergence of weighted quadratic variations under some mild assumptions, which are crucial to derive our new optimality results.

1.3.1 Control of $\Delta\tau^n$ and martingale increments

We start from a simple and efficient criterion for a.s. convergence of continuous local martingales.

Lemma 1.3.1 ([GL14a, Corollary 2.1])**.** *Let* $p > 0$*, and let* $\{(K_t^n)_{0\le t\le T} : n \ge 0\}$ *be a sequence of continuous scalar local martingales vanishing at zero. Then*

$$\sum_{n\ge 0}\langle K^n\rangle_T^{p/2} < +\infty \ \text{ a.s.} \iff \sum_{n\ge 0}\sup_{0\le t\le T}|K_t^n|^p < +\infty \ \text{ a.s..}$$

The useful application is the sense $\Rightarrow$: by controlling the summability of quadratic varia-

tions, we obtain the non trivial a.s. convergence of $\sup_{0\leq t\leq T}|K_t^n|$ to 0. This kind of reasoning is used in this work.

The next two lemmas yield controls of $\Delta\tau_i$ and of martingales increments for an admissible sequence of strategies. In view of the Brownian motion scaling property one might guess that an admissible sequence of strategies $\mathcal{T}=\{\mathcal{T}^n : n\geq 0\}$ yields stopping times increments of magnitude roughly equal to ε_n^2. More generally, we can study in a similar way the increments of martingales. Here we give a rigorous statement of these heuristics.

Lemma 1.3.2 ([GL14a, Corollary 2.2]). *Assume* $(\mathbf{H}_M)$ *and let* $\mathcal{T}=\{\mathcal{T}^n : n\geq 0\}$ *be a sequence of strategies. Let* $\rho>0$*, then the following hold:*

(i) Assume $\mathcal{T}$ *satisfies* $(\mathbf{A}_M^{\text{osc.}})$*, then*

$$\sup_{n\geq 0}(\varepsilon_n^{\rho-1}\sup_{1\leq i\leq N_T^n}\Delta\tau_i^n)<+\infty \quad \text{a.s..}$$

(ii) Assume $\mathcal{T}$ *satisfies* $(\mathbf{A}_M^{\text{osc.}})$-$(\mathbf{A}_N)$*, then*

$$\sup_{n\geq 0}(\varepsilon_n^{\rho-2}\sup_{1\leq i\leq N_T^n}\Delta\tau_i^n)<+\infty \quad \text{a.s..}$$

Lemma 1.3.3 ([GL14a, Corollary 2.3]). *Assume* $(\mathbf{H}_M)$*. Let* $((K_t^n)_{0\leq t\leq T})_{n\geq 0}$ *be a sequence of* $\mathbb{R}^d$*-valued continuous local martingales such that* $\langle K^n\rangle_t=\int_0^t\kappa_r^n\mathrm{d}r$ *for a measurable adapted* κ^n *satisfying the following inequality: there exist a non-negative* a.s. *finite random variable* C_κ *and a deterministic parameter* $\theta\geq 0$ *such that*

$$0\leq|\kappa_r^n|\leq C_\kappa(|\Delta M_r|^{2\theta}+|\Delta_r|^\theta)\qquad\forall 0\leq r<T,\quad\forall n\geq 0,\quad\text{a.s..}$$

Finally, let $\rho>0$*, then the following assertions hold.*

(i) Assume $\mathcal{T}$ *satisfies* $(\mathbf{A}_M^{\text{osc.}})$*, then*

$$\sup_{n\geq 0}(\varepsilon_n^{\rho-(1+\theta)/2}\sup_{1\leq i\leq N_T^n}\sup_{\tau_{i-1}^n\leq t\leq\tau_i^n}|\Delta K_t^n|)<+\infty \quad \text{a.s..}$$

(ii) Assume $\mathcal{T}$ *satisfies* $(\mathbf{A}_M^{\text{osc.}})$-$(\mathbf{A}_N)$*, then*

$$\sup_{n\geq 0}(\varepsilon_n^{\rho-(1+\theta)}\sup_{1\leq i\leq N_T^n}\sup_{\tau_{i-1}^n\leq t\leq\tau_i^n}|\Delta K_t^n|)<+\infty \quad \text{a.s..}$$

1.3.2 The admissible sequences of strategies for S and M coincide

We now aim at proving the following Theorem.

Theorem 1.3.4. *Let* S *be a semimartingale of the form* (1.2.2) *and satisfying* $(\mathbf{H}_A)$-$(\mathbf{H}_M)$*. Then a sequence of strategies* $\mathcal{T}=\{\mathcal{T}_n : n\geq 0\}$ *is admissible for* S *if and only it is admissible for* M *with the same parameter* ρ_N*: in other words, if* $\mathcal{T}$ *satisfies* $(\mathbf{A}_N)$*,*

$$(\mathbf{A}_M^{\text{osc.}})\Leftrightarrow(\mathbf{A}_S^{\text{osc.}}).$$

Rephrased differently, defining admissible sequence of strategies based on the martingale M is robust to perturbation by adding to M a finite variation process A, satisfying α-Hölder regularity with $\alpha > 1/2$.

Proof. For convenience in the proof, we adopt the short notation

$$|\Delta\tau^n|_\infty := \sup_{1\le i\le N_T^n} \Delta\tau_i^n, \qquad |\Delta U|_\infty := \sup_{1\le i\le N_T^n} \sup_{\tau_{i-1}^n\le t\le\tau_i^n} |\Delta U_t|,$$

for any process U.

Proof of $\Rightarrow$. Suppose first that $\mathcal{T} = \{\mathcal{T}_n : n \ge 0\}$ is admissible for S. Let us prove that it is admissible for M, i.e. the assumption $(\mathbf{A}_M^{\text{osc.}})$ is satisfied. We proceed in several steps.

▷ *Step 1. Preliminary bound.* From $|M_t - M_s| \le |S_t - S_s| + |A_t - A_s|$ and $(\mathbf{H}_A)$, we get

$$|\Delta M|_\infty \le |\Delta S|_\infty + C_0|\Delta\tau^n|_\infty^\alpha \le C_0(\varepsilon_n + |\Delta\tau^n|_\infty^\alpha). \tag{1.3.1}$$

Using Itô's formula and $(\mathbf{H}_M)$, we obtain that for any $0 \le s < t \le T$

$$0 \le t - s \le C_E^{-1}\int_s^t \mathrm{Tr}(\sigma_r\sigma_r^\mathsf{T})\mathrm{d}r = C_E^{-1}\sum_{j=1}^d (\langle S^j\rangle_t - \langle S^j\rangle_s) \tag{1.3.2}$$

$$= C_E^{-1}\sum_{j=1}^d \left((S_t^j - S_s^j)^2 - 2\int_s^t (S_r^j - S_s^j)\mathrm{d}S_r^j\right), \tag{1.3.3}$$

where $C_E := \inf_{t\in[0,T]} \mathrm{Tr}(\sigma_t\sigma_t^\mathsf{T}) > 0$ a.s.. Hence

$$\Delta t \le C_E^{-1}\left(C_0\varepsilon_n^2 + 2\sum_{j=1}^d \left|\int_{\phi(t)}^t \Delta S_r^j \mathrm{d}A_r^j\right| + 2\sum_{j=1}^d \left|\int_{\phi(t)}^t \Delta S_r^j \mathrm{d}M_r^j\right|\right). \tag{1.3.4}$$

Using that A is of finite variation and $(\mathbf{A}_S^{\text{osc.}})$, we get the crude estimate

$$\sum_{j=1}^d \left|\int_{\phi(t)}^t \Delta S_r^j \mathrm{d}A_r^j\right| \le C_0\varepsilon_n. \tag{1.3.5}$$

Now consider the local martingale $K_t^{n,j} = \varepsilon_n^{\frac{2}{p}-1}\left(\int_0^t \Delta S_r^j \mathrm{d}M_r^j\right)$ for some $p > 0$. We have

$$\sum_{n\ge0} \langle K^{n,j}\rangle_T^{\frac{p}{2}} = \sum_{n\ge0} \varepsilon_n^{2-p}\left(\int_0^T |\Delta S_r^j|^2 \mathrm{d}\langle M^j\rangle_r\right)^{\frac{p}{2}} \le C_0\sum_{n\ge0}\varepsilon_n^2 < +\infty \qquad \text{a.s.},$$

which by Lemma 1.3.1 implies that $\sum_{n\ge0}\sup_{0\le t\le T}|K_t^{n,j}|^p < +\infty$ a.s., and thus we have

$\sup_{n\geq 0}\sup_{0\leq t\leq T}|K_t^{n,j}|<+\infty$ a.s.. This reads

$$\sup_{0\leq t\leq T}\left|\int_0^t \Delta S_r^j \mathrm{d}M_r^j\right| \leq C_0\varepsilon_n^{1-\frac{2}{p}} = C_0\varepsilon_n^{1-\delta}, \tag{1.3.6}$$

where $\delta = 2/p$ is an arbitrary positive number. Plugging this and (1.3.5) into (1.3.4) yields

$$|\Delta\tau^n|_\infty \leq C_0(\varepsilon_n^2+\varepsilon_n+\varepsilon_n^{1-\delta}) \leq C_0\varepsilon_n^{1-\delta}. \tag{1.3.7}$$

The above is analogous to Lemma 1.3.2-(i) but under the assumption $(\mathbf{A}_S^{\text{osc.}})$. Combined with (1.3.1), we then deduce

$$|\Delta M|_\infty \leq C_0\varepsilon_n^{\alpha(1-\delta)} \tag{1.3.8}$$

for any given $\delta\in(0,1)$.

▷ *Step 2.* We prove the following lemma, which gives the basis for a continuation argument (Step 3): once we have estimated $|\Delta M|_\infty$ with some order w.r.t. ε_n, we obtain automatically a slightly better order, up to reaching the order 1, as required by $(\mathbf{A}_M^{\text{osc.}})$.

Lemma 1.3.5. *Suppose that for some $\beta>0$*

$$\sup_{n\geq 0}\left(\varepsilon_n^{-\beta}\sup_{1\leq i\leq N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}|\Delta M_t|^2\right)<+\infty \quad \text{a.s..} \tag{1.3.9}$$

Then for any $\rho>0$

$$\sup_{n\geq 0}\left(\varepsilon_n^{-(\beta-\rho)}\sup_{1\leq i\leq N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}\sum_{j=1}^d\Delta\langle M^j\rangle_t\right)<+\infty \quad \text{a.s..}$$

Proof. Let $p>0$. Consider the following two sequences of processes:

$$U_t^n = \varepsilon_n^{2-\beta p+2\rho_N}\sum_{\tau_{i-1}^n<t}\left|\sum_{j=1}^d\Delta\langle M^j\rangle_{\tau_i^n\wedge t}\right|^p,$$
$$V_t^n = \varepsilon_n^{2-\beta p+2\rho_N}\sum_{\tau_{i-1}^n<t}\sup_{s\in(\tau_{i-1}^n,\tau_i^n\wedge t]}|\Delta M_s|^{2p}.$$

We aim at proving that $\sum_{n\geq 0}U_T^n<+\infty$ a.s. using Lemma 1.A.1 in Appendix. First, $\sum_{n\geq 0}V_T^n$ converges a.s.: indeed using $(\mathbf{A}_N)$ and (1.3.9) we obtain

$$\sum_{n\geq 0}V_T^n \leq C_0\sum_{n\geq 0}\varepsilon_n^{2-\beta p+2\rho_N}N_T^n\sup_{1\leq i\leq N_T^n}\sup_{s\in(\tau_{i-1}^n,\tau_i^n]}|\Delta M_s|^{2p} \leq C_0\sum_{n\geq 0}\varepsilon_n^2<+\infty.$$

Second observe that for any n, $t\mapsto V_t^n$ is a.s. non-decreasing. Last it remains to verify the relation of domination of Lemma 1.A.1-(iii). Let $k\in\mathbb{N}$, let θ_k be defined as in the quoted lemma. On the set $\{\tau_{i-1}^n<t\wedge\theta_k\}$ from a conditional version of the multidimensional BDG

inequality we have

$$\mathbb{E}\left(\left|\sum_{j=1}^{d}\Delta\langle M^j\rangle_{\tau_i^n\wedge t\wedge\theta_k}\right|^p\middle|\mathcal{F}_{\tau_{i-1}^n}\right)\le c_p\mathbb{E}\left(\sup_{\tau_{i-1}^n<s\le\tau_i^n\wedge t\wedge\theta_k}|\Delta M_s|^{2p}\middle|\mathcal{F}_{\tau_{i-1}^n}\right).$$

Then it follows that

$$\begin{aligned}\mathbb{E}\left(U_{t\wedge\theta_k}^n\right)&=\varepsilon_n^{2-\beta p+2\rho_N}\sum_{i=1}^{+\infty}\mathbb{E}\left(1_{\tau_{i-1}^n<t\wedge\theta_k}\mathbb{E}\left(\left|\sum_{j=1}^{d}\Delta\langle M^j\rangle_{\tau_i^n\wedge t\wedge\theta_k}\right|^p\middle|\mathcal{F}_{\tau_{i-1}^n}\right)\right)\\&\le c_p\mathbb{E}\left(V_{t\wedge\theta_k}^n\right).\end{aligned}$$

Hence by Lemma 1.A.1, we obtain that $\sum_{n\ge0}U_T^n$ converges a.s. and thus $\sup_{n\ge0}U_T^n<+\infty$ a.s..

Now write $\varepsilon_n^{2-\beta p+2\rho_N}\sup_{1\le i\le N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}\left|\sum_{j=1}^{d}\Delta\langle M^j\rangle_t\right|^p\le U_T^n$, which implies

$$\sup_{n\ge0}\left(\varepsilon_n^{(2+2\rho_N)/p-\beta}\sup_{1\le i\le N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}\left|\sum_{j=1}^{d}\Delta\langle M^j\rangle_t\right|\right)<+\infty\qquad\text{a.s..}$$

To conclude, choose $p=\frac{2+2\rho_N}{\rho}$ to get the desired result. □

▷ *Step 3. Continuation scheme.* Take $\delta>0$, as in (1.3.8), set $d_0=\alpha(1-\delta)$ and $\rho_0=\dfrac{(2\alpha-1)d_0}{2\alpha}>0$. Consider the sequence $(d_m)_{m\ge0}$ given by $d_{m+1}=2\alpha d_m-\alpha\rho_0$ for $m\ge0$. Assume for a while that

$$d_{m+1}-d_m\ge\alpha\rho_0,\tag{1.3.10}$$

and let us show by induction that, for any $m\ge0$,

$$|\Delta M|_\infty\le C_0\varepsilon_n^{\min(d_m,1)}.\tag{1.3.11}$$

The case $m=0$ stems directly from (1.3.8). Now suppose that (1.3.11) holds for m. If $d_m\ge1$, since $d_{m+1}\ge d_m$ owing to (1.3.10), (1.3.11) is valid for $m+1$. If $d_m<1$, then we have $|\Delta M|_\infty\le C_0\varepsilon_n^{d_m}$ and using Lemma 1.3.5 we obtain

$$\left|\sum_{j=1}^{d}\Delta\langle M^j\rangle\right|_\infty\le C_0\varepsilon_n^{2d_m-\rho_0}.$$

Consequently (1.3.2) gives $|\Delta\tau^n|_\infty\le C_0\varepsilon_n^{2d_m-\rho_0}$ which, combined with (1.3.1), yields

$$|\Delta M|_\infty\le C_0\varepsilon_n^{\min(1,\alpha(2d_m-\rho_0))}.$$

This finishes the proof of (1.3.11) for $m+1$. It remains to show (1.3.10) by induction. For

$m = 0$ we get $d_1 = 2\alpha d_0 - \alpha\rho_0$ and thus

$$d_1 - d_0 = (2\alpha - 1)d_0 - \frac{(2\alpha - 1)d_0}{2} = \frac{(2\alpha - 1)d_0}{2} = \alpha\rho_0.$$

Suppose that (1.3.10) is true for all $m < k$ and let us extend to $m = k$. We write

$$d_{m+1} - d_m = (2\alpha - 1)d_m - \frac{(2\alpha - 1)d_0}{2} \geq (2\alpha - 1)d_0 - \frac{(2\alpha - 1)d_0}{2} = \alpha\rho_0,$$

using that $d_m \geq d_0$ by the induction assumption. We are done.
▷ *Step 4. Conclusion.* In view of (1.3.10), $(d_m)_{m\geq 0}$ becomes larger than 1 for some m, for which (1.3.11) simply writes $|\Delta M|_\infty \leq C_0\varepsilon_n$. $(\mathbf{A}_M^{\text{osc.}})$ is proved. □

Proof of $\Leftarrow$. Now suppose that the sequence $\mathcal{T}$ is admissible for M. Let us prove the admissibility of $\mathcal{T}$ for the process S. Again it is enough to verify the assumption $(\mathbf{A}_S^{\text{osc.}})$. Similarly to the decomposition (1.3.1), we have

$$|\Delta S|_\infty \leq |\Delta M|_\infty + |\Delta A|_\infty \leq C_0(\varepsilon_n + |\Delta\tau^n|_\infty^\alpha).$$

From Lemma 1.3.2-(ii), for any $\gamma > 0$, we have $|\Delta\tau^n|_\infty \leq C_0\varepsilon_n^{2-\gamma}$ a.s.. Since $\alpha > 1/2$, we can choose γ such that $(2-\gamma)\alpha > 1$ and for such γ we deduce $|\Delta S|_\infty \leq C_0(\varepsilon_n + \varepsilon_n^{(2-\gamma)\alpha}) \leq C_0\varepsilon_n$. The proof is complete. □

Remark 1.3.6.

- *Theorem 1.3.4 implies that if a sequence of strategies fulfills* $(\mathbf{A}_N)$*, we do not need to emphasize anymore the dependence of the assumption* $(\mathbf{A}_\cdot^{\text{osc.}})$ *on a particular process* M *or* S*; in that case, we will write simply* $(\mathbf{A}^{\text{osc.}})$ *and will refer to admissible sequence of strategies* $\mathcal{T}^{\text{adm.}} := \mathcal{T}_M^{\text{adm.}} = \mathcal{T}_S^{\text{adm.}}$.
- *In addition, we can use all the results for admissible sequences of strategies based on the local martingale* M *and* $(\mathbf{A}_M^{\text{osc.}})$ *(as those from [GL14a]): in particular, for any admissible sequences of strategies (for* M *or* S*), we have* $\sup_{1\leq i\leq N_T^n} |\Delta\tau_i^n| \leq C_0\varepsilon_n^{2-\gamma}$ *for any* $\gamma > 0$.

A direct consequence of Lemma 1.3.2-(ii), $(\mathbf{H}_A)$ and Theorem 1.3.4 is the following.

Corollary 1.3.7. *Let* S *be a semimartingale of the form* (1.2.2) *and satisfying* $(\mathbf{H}_A)$-$(\mathbf{H}_M)$. *If* $\mathcal{T} \in \mathcal{T}^{\text{adm.}}$*, then for any* $\rho > 0$,

$$\sup_{1\leq i\leq N_T^n} \sup_{\tau_{i-1}^n \leq t \leq \tau_i^n} |\Delta A_t| \leq C_0\varepsilon_n^{2\alpha-\rho}.$$

1.3.3 Convergence results for quadratic variation

We first recall a convergence result about weighted discrete quadratic M-variations corresponding to $\mathcal{T} = \{\mathcal{T}^n, n \geq 0\}$.

Proposition 1.3.8. *[GL14a, Proposition 2.3] Assume* ($\mathbf{H}_M$) *and let* $\mathcal{T}$ *be a sequence of strategies satisfying* ($\mathbf{A}_M^{\text{osc.}}$). *Let* $(H_t)_{0\le t<T}$ *be a continuous adapted* $d\times d$*-matrix process such that* $\sup_{t\in[0,T)}|H_t|<+\infty$ a.s., *and let* $(K_t)_{0\le t\le T}$ *be a* $\mathbb{R}^d$*-valued continuous local martingale such that* $\langle K\rangle_t=\int_0^t\kappa_r\mathrm{d}r$ *with* $\sup_{t\in[0,T]}|\kappa_t|<+\infty$ a.s.. *Then*

$$\sum_{\tau_{i-1}^n<T}\Delta K_{\tau_i^n}^{\mathsf{T}}H_{\tau_{i-1}^n}\Delta K_{\tau_i^n}\overset{\text{a.s.}}{\to}\int_0^T\mathrm{Tr}(H_t\mathrm{d}\langle K\rangle_t).$$

We now establish an extension to the semimartingale S.

Proposition 1.3.9. *Let* S *be a semimartingale of the form* (1.2.2) *and satisfying* ($\mathbf{H}_A$)-($\mathbf{H}_M$), *and let* $\mathcal{T}$ *be a sequence of strategies satisfying* ($\mathbf{A}_S^{\text{osc.}}$). *Let* $(H_t)_{0\le t<T}$ *be as in Proposition 1.3.8. Then*

$$\sum_{\tau_{i-1}^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}H_{\tau_{i-1}^n}\Delta S_{\tau_i^n}\overset{\text{a.s.}}{\to}\int_0^T\mathrm{Tr}\left(H_t\mathrm{d}\langle M\rangle_t\right).$$

Proof. From Itô's lemma, the difference between the above left hand side and the right one is equal to

$$\int_0^T\Delta S_t^{\mathsf{T}}(H_{\varphi(t)}+H_{\varphi(t)}^{\mathsf{T}})\mathrm{d}S_t+\int_0^T\mathrm{Tr}([H_{\varphi(t)}-H_t]\mathrm{d}\langle M\rangle_t). \tag{1.3.12}$$

Due to ($\mathbf{H}_M$), the second term is bounded by $C_0\int_0^T|H_{\varphi(t)}-H_t|\mathrm{d}t$: it converges to 0 by an application of the dominated convergence theorem. Indeed, H is continuous and bounded on $[0,T)$ and the mesh size goes to 0 under ($\mathbf{A}_S^{\text{osc.}}$) (see (1.3.7) which is established under ($\mathbf{A}_S^{\text{osc.}}$) and without using ($\mathbf{A}_N$)). Next, decompose the first term of (1.3.12) into stochastic integrals w.r.t. A and M. On the one hand, A is of finite variation, thus

$$\left|\int_0^T\Delta S_t^{\mathsf{T}}(H_{\varphi(t)}+H_{\varphi(t)}^{\mathsf{T}})\mathrm{d}A_t\right|\le C_0\sup_{1\le i\le N_T^n}\sup_{\tau_{i-1}^n\le t\le\tau_i^n}|\Delta S_t|\sup_{t\in[0,T)}|H_t|\overset{\text{a.s.}}{\to}0 \tag{1.3.13}$$

in view of ($\mathbf{A}_S^{\text{osc.}}$). On the other hand, $\int_0^T\Delta S_t^{\mathsf{T}}(H_{\varphi(t)}+H_{\varphi(t)}^{\mathsf{T}})\mathrm{d}M_t\overset{\text{a.s.}}{\to}0$ by proceeding very similarly to the proof of (1.3.6). □

In the next theorems we identify an important admissible sequence of strategies, namely hitting times by S of random ellipsoids parametrized by a matrix process $(H_t)_{0\le t<T}$ (or a perturbation of it). This extends [GL14a, Proposition 2.4] to hitting times of S and to possibly degenerate H. This more general construction of ellipsoids is a significant improvement, and crucial for the subsequent optimality results.

Theorem 1.3.10. *Let* S *be a semimartingale of the form* (1.2.2) *and satisfying* ($\mathbf{H}_A$)-($\mathbf{H}_M$), *and let* $(H_t)_{0\le t<T}$ *be a continuous adapted symmetric non-negative definite* $d\times d$ *matrix process, such that* a.s.

$$0<\inf_{0\le t<T}\lambda_{\min}(H_t)\le\sup_{0\le t<T}\lambda_{\max}(H_t)<+\infty.$$

The strategy $\mathcal{T}^n$ *given by*

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} H_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^2\} \wedge T, \end{cases}$$

defines a admissible sequence of strategies.

The proof is given later. The condition $\sup_{0\leq t<T} \lambda_{\max}(H_t) < +\infty$ ensures that none of the corresponding ellipsoids $\mathcal{E}_t := \{x^\mathsf{T} H_t x \leq c\}$ with $c > 0$ are flat in some directions, it allows to derive a bound on the number of hitting times N_T^n as in $(\mathbf{A}_N)$. The non-degeneracy condition $\lambda_{\min}(H_t) > 0$ (i.e. $\mathcal{E}_t$ is bounded) is important to control the increments ΔS as in $(\mathbf{A}_S^{\text{osc.}})$. Without this latter condition, we need to perturb the above sequence of strategies. To this purpose, let $\chi(.)$ be a smooth function such that

$$\mathbf{1}_{(-\infty,1/2]} \leq \chi(.) \leq \mathbf{1}_{(-\infty,1]}, \tag{1.3.14}$$

and for $\mu > 0$ set $\chi_\mu(x) = \chi(x/\mu)$.

Theorem 1.3.11. *Let* S *be a semimartingale of the form* (1.2.2) *and satisfying* $(\mathbf{H}_A)$-$(\mathbf{H}_M)$. *Assume that* ρ_N *defined in* (1.2.3) *is such that* $\rho_N > 1$, *and let* $\delta \in (0, 2(\rho_N - 1)]$. *Let* $(H_t)_{0\leq t<T}$ *be an adapted symmetric non-negative definite* $d \times d$ *matrix process, such that*

(i) there exists a random variable C_H, *positive and finite* a.s., *such that*

$$\lambda_{\max}(H_t) \leq C_H, \quad \forall t \in [0, T), \quad \text{a.s.}.$$

(notice that H *is not necessarily continuous).*

Define a sequence of processes $H^{(n)}$ *by*

$$H_t^{(n)} = H_t + \varepsilon_n^\delta \chi_{\varepsilon_n^\delta}(\lambda_{\min}(H_t))\, \mathrm{Id}_d\,.$$

Then the strategy $\mathcal{T}^n$ *defined by*

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} H^{(n)}_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^{2+\delta}\} \wedge T, \end{cases} \tag{1.3.15}$$

forms a sequence $\mathcal{T} = \{\mathcal{T}^n : n \geq 0\}$ *satisfying the assumption* $(\mathbf{A}_S^{\text{osc.}})$. *If in addition the following convergence holds*

(ii)

$$\sum_{\tau_{i-1}^n < T} \Delta S^\mathsf{T}_{\tau_i^n} H_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \overset{\text{a.s.}}{\to} \int_0^T \mathrm{Tr}(H_t \mathrm{d}\langle M\rangle_t),$$

then the sequence $\mathcal{T}$ *satisfies also the assumption* $(\mathbf{A}_N)$, *that is* $\mathcal{T} \in \mathcal{T}^{adm.}$.

Proof of Theorem 1.3.11. First let us prove that $\mathcal{T}_n$ is a.s. of finite size for any $n \in \mathbb{N}$. The definition of $H_t^{(n)}$ implies that

$$\lambda_{\max}(H_t^{(n)}) \leq C_H + \sup_{n\geq 0} \varepsilon_n^{\delta} < +\infty, \quad \forall t \in [0,T) \quad \text{a.s..}$$

Define the event $\mathcal{N}^n := \{\omega : N_T^n(\omega) = +\infty\}$. For $\omega \in \mathcal{N}^n$ the infinite sequence $(\tau_i^n(\omega))$ is increasing and bounded, thus converges. Hence on $\mathcal{N}^n \cap E_S$, with

$$E_S = \{(S_t)_{t\in[0,T]} \text{ is continuous and } C_H < +\infty\},$$

we have

$$\begin{aligned} 0 < \varepsilon_n^{2+\delta} &= (S_{\tau_i^n} - S_{\tau_{i-1}^n})^{\mathsf{T}} H^{(n)}_{\tau_{i-1}^n} (S_{\tau_i^n} - S_{\tau_{i-1}^n}) \\ &\leq \left(C_H + \sup_{n\geq 0} \varepsilon_n^{\delta}\right) |S_{\tau_i^n} - S_{\tau_{i-1}^n}|^2 \xrightarrow{i\to+\infty} 0, \end{aligned}$$

which is impossible. Hence $\mathbb{P}(\mathcal{N}^n \cap E_S) = 0$, but $\mathbb{P}(E_S) = 1$ thus $\mathbb{P}(\mathcal{N}^n) = 0$.

Next we show that $\mathcal{T}$ satisfies $(\mathbf{A}_S^{\text{osc.}})$. From the definition of $H_t^{(n)}$ it is straightforward that

$$\lambda_{\min}(H_t^{(n)}) \geq \frac{\varepsilon_n^{\delta}}{2}, \quad \forall t \in [0,T).$$

Thus

$$\begin{aligned} &\varepsilon_n^{-2} \sup_{1\leq i\leq N_T^n} \sup_{\tau_{i-1}^n \leq t \leq \tau_i^n} |\Delta S_t|^2 \\ &\leq \left(\inf_{t\in[0,T)} \lambda_{\min}(H_t^{(n)})\right)^{-1} \varepsilon_n^{-2} \sup_{1\leq i\leq N_T^n} \sup_{\tau_{i-1}^n \leq t \leq \tau_i^n} (\Delta S_t^{\mathsf{T}} H^{(n)}_{\tau_{i-1}^n} \Delta S_t) \leq 2\varepsilon_n^{-\delta}\varepsilon_n^{-2}\varepsilon_n^{2+\delta} = 2, \end{aligned}$$

which validates the assumption $(\mathbf{A}_S^{\text{osc.}})$.

Finally assume that in addition *(ii)* holds and let us show that the sequence of strategies $\mathcal{T}$ satisfies the assumption $(\mathbf{A}_N)$. Writing $N_T^n = 1 + \sum_{1\leq i\leq N_T^n - 1} 1$ and using $2 + \delta \leq 2\rho_N$, we observe that (for n large enough so that $\varepsilon_n \leq 1$)

$$\varepsilon_n^{2\rho_N} N_T^n \leq \varepsilon_n^{2+\delta} N_T^n \leq \varepsilon_n^{2+\delta} + \sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} H_{\tau_{i-1}^n} \Delta S_{\tau_i^n} + \sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} (H^{(n)}_{\tau_{i-1}^n} - H_{\tau_{i-1}^n}) \Delta S_{\tau_i^n}. \tag{1.3.16}$$

Now by (ii) we have

$$\sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} H_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \overset{\text{a.s.}}{\to} \int_0^T \mathrm{Tr}(H_t \mathrm{d}\langle M\rangle_t) \overset{\text{a.s.}}{<} +\infty$$

(the contribution $i = N_T^n$ does not change the convergence). Besides from the definition of

$H^{(n)}$ we get

$$\left| \sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} (H^{(n)}_{\tau^n_{i-1}} - H_{\tau^n_{i-1}}) \Delta S_{\tau_i^n} \right| \leq \varepsilon_n^{\delta} \sum_{\tau_i^n < T} |\Delta S_{\tau_i^n}|^2 \overset{\text{a.s.}}{\to} 0, \tag{1.3.17}$$

using $\delta > 0$ and Proposition 1.3.9 (valid since $(\mathbf{A}_S^{\text{osc.}})$ is in force now). We have proved that the r.h.s. of (1.3.16) converges a.s. to a finite random variable, which completes the verification of the assumption $(\mathbf{A}_N)$. □

Proof of Theorem 1.3.10. This is an adaptation of the previous proof. First, with the same arguments we prove that $\mathcal{T}_n$ is a.s. of finite size for any $n \in \mathbb{N}$. Second, the verification of $(\mathbf{A}_S^{\text{osc.}})$ stems from

$$\varepsilon_n^{-2} \sup_{1 \leq i \leq N_T^n} \sup_{\tau^n_{i-1} \leq t \leq \tau_i^n} |\Delta S_t|^2 \leq \left(\inf_{t \in [0,T)} \lambda_{\min}(H_t) \right)^{-1}.$$

Third, for n large enough so that $\varepsilon_n \leq 1$, we write

$$\varepsilon_n^{2\rho_N} N_T^n \leq \varepsilon_n^2 N_T^n \leq \varepsilon_n^2 + \sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} H_{\tau^n_{i-1}} \Delta S_{\tau_i^n}$$

and we conclude to $(\mathbf{A}_N)$ using Proposition 1.3.9 and the continuity and boundedness of H. □

1.4 Asymptotic lower bound on the discretization error

Let S be a semimartingale of the form (1.2.2) and let v be the function appearing in the discretization error (1.1.1), and satisfying $(\mathbf{H}_v)$. The main result of the section is Theorem 1.4.2: this is an extension to the semimartingale case of the asymptotic lower bound on the discretization error, proved in [GL14a, Theorem 3.1] in the martingale case.

The discretization error Z^n defined in (1.1.1) can be decomposed into a martingale part and a finite variation part:

$$Z_s^n = \int_0^s (v(t, S_t) - v(\phi(t), S_{\phi(t)})) \cdot \mathrm{d}M_t + \int_0^s (v(t, S_t) - v(\phi(t), S_{\phi(t)})) \cdot \mathrm{d}A_t.$$

The analysis is partially derived from a smart representation of $\langle Z^n \rangle_T$ as a sum of squared random variables and an adequate application of Cauchy-Schwarz inequality. The derivation of such a representation is based on applying the Itô formula to a *suitable function* and identifying the bounded variation term. While it is straightforward in dimension one, a multidimensional version of this result requires to solve the following matrix equation.

Lemma 1.4.1. *Let c be a $d \times d$-matrix with real-valued entries. Then the equation*

$$2 \operatorname{Tr}(x) x + 4x^2 = cc^{\mathsf{T}} \tag{1.4.1}$$

admits exactly one solution $x(c) \in \mathcal{S}_+^d(\mathbb{R})$. *Moreover, the mapping* $c \mapsto x(c)$ *is continuous.*

The proof of the above lemma directly follows from [GL14a, Lemma 3.1] applied for $(cc^\mathsf{T})^{1/2}$ (i.e. the symmetric non-negative definite square root of cc^T). Now we state the main result.

Theorem 1.4.2 (Lower bound)**.** *Assume* $(\mathbf{H}_A)$, $(\mathbf{H}_M)$, $(\mathbf{H}_v)$ *and let* $\mathcal{T}$ *be an admissible sequence of strategies (satisfying* $(\mathbf{A}_N)$ *and* $(\mathbf{A}^{\text{osc.}})$*). Let* X *be the continuous adapted symmetric non-negative definite matrix process solution of* (1.4.1) *with* $c = \sigma^\mathsf{T}(D_x v)^\mathsf{T}\sigma$, *i.e.*

$$X_t := x(\sigma_t^\mathsf{T}(D_x v_t)^\mathsf{T}\sigma_t), \qquad \text{for} \quad 0 \le t < T. \tag{1.4.2}$$

Then we have

$$\liminf_{n\to+\infty} N_T^n \langle Z^n \rangle_T \ge \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t \right)^2 \qquad \text{a.s..}$$

Proof. The martingale part of the discretization error can be written

$$\int_0^s (v(t, S_t) - v(\phi(t), S_{\phi(t)})) \cdot \mathrm{d}M_t =: \int_0^s (D_x v_{\phi(t)} \Delta S_t)\mathrm{d}M_t + R_s^n. \tag{1.4.3}$$

Therefore the quadratic variation of Z^n is given by

$$\begin{aligned} \langle Z^n \rangle_T &= \int_0^T \Delta S_t^\mathsf{T}(D_x v_{\phi(t)})^\mathsf{T}\mathrm{d}\langle M\rangle_t D_x v_{\phi(t)} \Delta S_t + e_{1,T}^n \\ &= \int_0^T \Delta M_t^\mathsf{T}(D_x v_{\phi(t)})^\mathsf{T}\mathrm{d}\langle M\rangle_t D_x v_{\phi(t)} \Delta M_t + e_{1,T}^n + e_{0,T}^n, \end{aligned} \tag{1.4.4}$$

where

$$\begin{aligned} e_{0,T}^n &:= \int_0^T \Delta A_t^\mathsf{T}(D_x v_{\phi(t)})^\mathsf{T}\mathrm{d}\langle M\rangle_t D_x v_{\phi(t)}(\Delta S_t + \Delta M_t), \\ e_{1,T}^n &:= \langle R^n \rangle_T + 2\langle \int_0^{\cdot} (D_x v_{\phi(t)} \Delta M_t) \cdot \mathrm{d}M_t, R^n \rangle_T. \end{aligned}$$

Now in the first contribution of $\langle Z^n \rangle_T$ in (1.4.4), we seek an expression involving only the Brownian motion B and not the local martingale M: hence we replace ΔM_t by $\sigma_{\phi(t)}\Delta B_t$ and $\mathrm{d}\langle M\rangle_t$ by $\sigma_{\phi(t)}\sigma_{\phi(t)}^\mathsf{T}\mathrm{d}t$, which leads to

$$\langle Z^n \rangle_T = \int_0^T \Delta B_t^\mathsf{T}(\sigma_{\phi(t)}^\mathsf{T}(D_x v_{\phi(t)})^\mathsf{T}\sigma_{\phi(t)}\sigma_{\phi(t)}^\mathsf{T} D_x v_{\phi(t)}\sigma_{\phi(t)})\Delta B_t \mathrm{d}t + e_{0,T}^n + e_{1,T}^n + e_{2,T}^n,$$

where

$$\begin{aligned} e_{2,T}^n := &\int_0^T \Delta M_t^\mathsf{T}(D_x v_{\phi(t)})^\mathsf{T}\Delta(\sigma_t\sigma_t^\mathsf{T}) D_x v_{\phi(t)} \Delta M_t \mathrm{d}t \\ &+ \int_0^T (\Delta M_t + \sigma_{\phi(t)}\Delta B_t)^\mathsf{T}(D_x v_{\phi(t)})^\mathsf{T}\sigma_{\phi(t)}\sigma_{\phi(t)}^\mathsf{T} D_x v_{\phi(t)}(\Delta M_t - \sigma_{\phi(t)}\Delta B_t)\mathrm{d}t. \end{aligned}$$

Denote $C_t = \sigma_t^{\mathsf{T}}(D_x v_t)^{\mathsf{T}}\sigma_t$. We seek a smart representation of the main term of $\langle Z^n\rangle_T$ in the form

$$\sum_{\tau_{i-1}^n<T} (\Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n})^2, \tag{1.4.5}$$

where X is a suitable measurable adapted symmetric $d\times d$-matrix process. For such a process X, the Itô formula on each interval $[\tau_{i-1}^n, \tau_i^n]$ yields

$$\sum_{\tau_{i-1}^n<T} (\Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n})^2 = \int_0^T \Delta B_t^{\mathsf{T}}(2\,\mathrm{Tr}(X_{\phi(t)})X_{\phi(t)} + 4X_{\phi(t)}^2)\Delta B_t \mathrm{d}t + 4\int_0^T \Delta B_t^{\mathsf{T}} X_{\phi(t)} \Delta B_t \Delta B_t^{\mathsf{T}} X_{\phi(t)} \mathrm{d}B_t.$$

Now take X as stated in the theorem. Clearly $X_t \in \mathcal{S}_+^d(\mathbb{R})$ owing to Lemma 1.4.1. The continuity of the mapping $c \mapsto x(c)$ also ensures that X is continuous and adapted, as $\sigma^{\mathsf{T}}(D_x v)^{\mathsf{T}}\sigma$ is. Then a simplified representation of $\langle Z^n\rangle_T$ readily follows:

$$\langle Z^n\rangle_T = \sum_{\tau_{i-1}^n<T} (\Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n})^2 + e_{0,T}^n + e_{1,T}^n + e_{2,T}^n + e_{3,T}^n, \tag{1.4.6}$$

where

$$e_{3,T}^n := -4\int_0^T \Delta B_t^{\mathsf{T}} X_{\phi(t)} \Delta B_t \Delta B_t^{\mathsf{T}} X_{\phi(t)} \mathrm{d}B_t.$$

Using Cauchy-Schwarz inequality and $X_t \in \mathcal{S}_+^d(\mathbb{R})$, we obtain

$$N_T^n \sum_{\tau_{i-1}^n<T} (\Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n})^2 \geq \left(\sum_{\tau_{i-1}^n<T} \Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n}\right)^2.$$

The process X_t is a.s. continuous on $[0,T)$, with $\sup_{t\in[0,T)} |X_t| < +\infty$ a.s., and thus the assumptions of Proposition 1.3.8 are satisfied for $(H,K) = (X,B)$. Therefore

$$\sum_{\tau_i^n<T} \Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n} \overset{\text{a.s.}}{\to} \int_0^T \mathrm{Tr}(X_t)\mathrm{d}t.$$

To summarize we have obtained that

$$\liminf_{n\to+\infty} \left(N_T^n \langle Z^n\rangle_T - N_T^n(e_{0,T}^n + e_{1,T}^n + e_{2,T}^n + e_{3,T}^n)\right) \geq \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2 \quad \text{a.s..}$$

To complete the proof, it is enough to show that $N_T^n(e_{0,T}^n + e_{1,T}^n + e_{2,T}^n + e_{3,T}^n) \overset{\text{a.s.}}{\to} 0$. In view of the assumption $(\mathbf{A}_N)$ it is sufficient to prove that

$$\varepsilon_n^{-2\rho_N} e_{i,T}^n \overset{\text{a.s.}}{\to} 0 \quad \text{for } i = 0,1,2,3. \tag{1.4.7}$$

Contribution $e^n_{0,T}$. Owing to Corollary 1.3.7, we obtain immediately that

$$|e^n_{0,T}| \leq C_0 \int_0^T |\Delta A_t|(|\Delta S_t| + |\Delta M_t|)\mathrm{d}t \leq C_0 \varepsilon_n^{1+2\alpha-\rho},$$

for any $\rho > 0$, which implies $\varepsilon_n^{-2\rho_N} e^n_{0,T} \to 0$ since $\rho_N < \frac{1}{2} + \alpha$.
Contribution $e^n_{1,T}$. To handle it, we need the following lemma; its proof follows that of [GL14a, Lemma 3.2], with minor adaptations (see Appendix 1.A.1).

Lemma 1.4.3. *Under the assumptions* $(\mathbf{H}_A)$, $(\mathbf{H}_M)$, $(\mathbf{H}_v)$, $(\mathbf{A}^{\text{osc.}})$ *and* $(\mathbf{A}_N)$, *we have* $\varepsilon_n^{2-4\rho_N}\langle R^n\rangle_T \overset{\text{a.s.}}{\to} 0$, *where* R^n *is defined in* (1.4.3).

Now to show that $\varepsilon_n^{-2\rho_N} e^n_{1,T} \to 0$, use the above lemma and $(\mathbf{A}^{\text{osc.}}_M)$ to get

$$\varepsilon_n^{-2\rho_N}|e^n_{1,T}| \leq \varepsilon_n^{-2\rho_N}(\langle R^n\rangle_T + 2C_0\varepsilon_n(\langle R^n\rangle_T)^{1/2} = o(\varepsilon_n^{2\rho_N-2}) + o(1) \overset{\text{a.s.}}{\to} 0.$$

Contributions $e^n_{2,T}$ *and* $e^n_{3,T}$. The proof is similar to that of [GL14a, Theorem 3.1], we skip the details. □

1.5 Optimal strategy

1.5.1 Preliminaries, pseudo-inverses

Now our main purpose is to provide, in notation of Theorem 1.4.2, an optimal discretization strategy, i.e. an admissible strategy $\mathcal{T}$ for which

$$\lim_{n\to+\infty} N^n_T\langle Z^n\rangle_T = \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2 \qquad \text{a.s..}$$

Notice that an existence result is proved in [GL14a, Theorem 3.3], only under the conditions that σ is invertible, that $v(t,x) = \nabla_x u(t,x)$ with

$$\inf_{0\leq t<T} \lambda_{\min}(D^2_{xx}u(t,S_t)) > 0 \quad \text{a.s.}$$

and that $A = 0$ (martingale case). Our aim here is to relax these three conditions, and to extend the ideas of this aforementioned theorem to our general setting.

Actually, the main difficulty comes from the possible degeneracy of σ. First recall the definition and some properties of pseudo-inverse matrix (a.k.a. Moore-Penrose generalized inverse).

Definition 1.5.1 (pseudo-inverse of a matrix)**.** *Let* $\mathcal{M}$ *be a real-valued* $d \times d$*-matrix. Consider the singular value decomposition of* $\mathcal{M}$

$$\mathcal{M} = U \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V^\intercal,$$

where U, V are both orthogonal matrices, and D is a diagonal matrix containing the (positive) singular values of $\mathcal{M}$ on its diagonal. Then the pseudo-inverse of M is the $d \times d$- matrix defined as

$$\mathcal{M}^\dagger = V \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^\mathsf{T}.$$

We recall the following well-known properties, which can be easily checked from Definition 1.5.1:

$$\begin{cases} \mathcal{M}\mathcal{M}^\dagger\mathcal{M} = \mathcal{M}, \qquad \mathcal{M}^\dagger\mathcal{M}\mathcal{M}^\dagger = \mathcal{M}^\dagger, \\ \text{the matrices } \mathcal{M}\mathcal{M}^\dagger \text{ and } \mathcal{M}^\dagger\mathcal{M} \text{ are symmetric.} \end{cases} \tag{1.5.1}$$

1.5.2 Main result

We wish to design optimal stopping times in terms of the process S to allow better tractability. Inspired by [GL14a], a good candidate is then the sequence $\{\mathcal{T}^n : n \geq 0\}$ where $\mathcal{T}^n$ is defined as:

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Lambda_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^2\} \wedge T, \end{cases} \tag{1.5.2}$$

where $\Lambda_t := (\sigma_t^{-1})^\mathsf{T} X_t \sigma_t^{-1}$ with X given by (1.4.2).

Such a sequence turns out to be optimal when S is a martingale and under some additional assumptions (see [GL14a, Theorem 3.3]). The problems with this definition can arise if σ_t is not invertible, or if Λ_t is degenerate for some values of t (then we have difficulties to verify $(\mathbf{A}^{\text{osc.}})$). To overcome these problems we use $\sigma_t^\dagger$ instead of σ_t^{-1}. Furthermore we take $\Lambda_t^{(n)}$ equal to a small perturbation of Λ_t depending on ε_n, such that $\Lambda_t^{(n)}$ is always non-degenerate.

We need one additional assumption.

($\mathbf{H}_\Lambda$) Let $(X_t)_{0 \leq t < T}$ be defined in (1.4.2) and consider the $\mathcal{S}_+^d(\mathbb{R})$-valued process defined by

$$\Lambda_t := (\sigma_t^\dagger)^\mathsf{T} X_t \sigma_t^\dagger, \quad \forall t \in [0, T). \tag{$\mathbf{H}_\Lambda$}$$

There exists a non-negative random variable $c_{(1.5.3)}$, finite a.s., such that

$$0 \leq \mathrm{Tr}(\Lambda_t) \leq c_{(1.5.3)}, \quad \forall t \in [0, T), \quad \text{a.s..} \tag{1.5.3}$$

Note that $\sigma^\dagger$ may be discontinuous, so Λ may be too. Recall (see (1.3.14)) that $\chi(.)$ stands for a continuous function such that $\mathbf{1}_{(-\infty,1/2]} \leq \chi(.) \leq \mathbf{1}_{(-\infty,1]}$, and for $\mu > 0$, we set $\chi_\mu(x) = \chi(x/\mu)$. Now we state the precise definition of an optimal sequence of strategies.

Theorem 1.5.2 (Optimal strategy)**.** *Assume that ($\mathbf{H}_A$), ($\mathbf{H}_M$), ($\mathbf{H}_v$), ($\mathbf{H}_\Lambda$) are in force. Let ρ_N satisfy (1.2.3) with $\rho_N > 1$, and let $\delta \in (0, 2(\rho_N - 1)]$. For each $n \in \mathbb{N}$, define the process $(\Lambda_t^{(n)} : t < T)$ by*

$$\Lambda_t^{(n)} = \Lambda_t + \varepsilon_n^\delta \chi_{\varepsilon_n^\delta}(\lambda_{\min}(\Lambda_t))\, \mathrm{Id}_d$$

where Λ is given in ($\mathbf{H}_\Lambda$), and define the strategy $\mathcal{T}^n_{\varepsilon_n^\delta}$ by

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Lambda^{(n)}_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^{2+\delta}\} \wedge T. \end{cases} \tag{1.5.4}$$

Then the sequence of strategies $\mathcal{T} = \{\mathcal{T}^n_{\varepsilon_n^\delta} : n \geq 0\}$ is admissible for the parameter ρ_N (in the sense of Definition 1.2.1 and Theorem 1.3.4) and is asymptotically optimal, i.e.

$$\lim_{n\to+\infty} N_T^n \langle Z^n \rangle_T = \left(\int_0^T \mathrm{Tr}(X_t) \mathrm{d}t \right)^2 \qquad \text{a.s..}$$

To conclude this subsection, we provide a condition simpler than ($\mathbf{H}_\Lambda$), the proof is postponed to the end of this section.

Proposition 1.5.3. *Assume that ($\mathbf{H}_A$), ($\mathbf{H}_M$), ($\mathbf{H}_v$) are in force, and assume that $v \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d)$ so that $D_x v_t$ and X_t can be defined continuously up to $t = T$. If the matrix*

$$C_t := \sigma_t^\mathsf{T} (D_x v_t)^\mathsf{T} \sigma_t \neq 0 \tag{$\mathbf{H}_C$}$$

for all $t \in [0,T]$ a.s., then ($\mathbf{H}_\Lambda$) holds.

1.5.3 Examples

About the assumptions ($\mathbf{H}_\Lambda$) and ($\mathbf{H}_C$)

Recall that under our assumptions, X is a.s. uniformly bounded on $[0,T)$. Thus in order to satisfy ($\mathbf{H}_\Lambda$), it is enough to have $\sigma^\dagger$ a.s. uniformly bounded on $[0,T)$. We provide a (non-exhaustive) list of such examples.

a) σ_t is invertible for any t a.s.: then $\sigma_t^\dagger = \sigma_t^{-1}$ is clearly bounded on $[0,T]$.

b) We can also afford degenerate cases: for instance if σ_t is constant in time (but possibly with $\mathrm{rank}(\sigma_t) < d$), then $\sigma_t^\dagger$ is also constant in time (and thus bounded).

c) The previous principle can be generalized to the time-dependent case $\sigma_t = \begin{pmatrix} \Sigma_t & 0 \\ 0 & 0 \end{pmatrix}$ where Σ_t is a square matrix, a.s. invertible at any time: indeed $\sigma_t^\dagger = \begin{pmatrix} \Sigma_t^{-1} & 0 \\ 0 & 0 \end{pmatrix}$ is bounded on $[0,T]$.

Now, we argue that checking ($\mathbf{H}_C$) may be sometimes much simpler than the verification of ($\mathbf{H}_\Lambda$). Let us give a non-trivial example where $\sigma^\dagger$ is not continuous a.s. For the i-th component of S, take a squared δ_i-dimensional radial Ornstein-Uhlenbeck process with parameter $-\lambda_i$, which is the strong solution to

$$S_t^i = S_0^i + \int_0^t (\delta_i - \lambda_i S_s^i) \mathrm{d}s + 2 \int_0^t \sqrt{S_s^i} \mathrm{d}B_s^i,$$

where $S_0^i > 0$, $\delta_i \geq 0$, $\lambda_i \in \mathbb{R}$ (see [GJY03]). The matrix σ_t is diagonal and its i-th element is equal to $2\sqrt{S_t^i}$. It is easy to check that ($\mathbf{H}_A$) and ($\mathbf{H}_M$) hold (in particular $\sigma_t \neq 0$ for all t a.s.). The pseudo-inverse $\sigma_t^\dagger$ is diagonal with i-th element equal to $[2\sqrt{S_t^i}]^{-1}\mathbf{1}_{S_t^i>0}$. Assume now that one of the δ_i is strictly smaller than 2: then the associated component S^i has a positive probability to hit 0 before T. As a consequence, with positive probability, $\sigma^\dagger$ is unbounded on $[0,T]$ and it is not clear anymore to check directly ($\mathbf{H}_\Lambda$). Alternatively, assume (again to simplify) that $D_x v_t \in \mathcal{S}_{++}^d(\mathbb{R})$. Then $C_t \neq 0$: indeed, $C_t \in \mathcal{S}_+^d(\mathbb{R})$ and $\mathrm{Tr}(C_t) = \mathrm{Tr}(D_x v_t \sigma_t \sigma_t^\mathsf{T}) > 0$ since $\sigma_t \sigma_t^\mathsf{T} \neq 0$ and $D_x v_t$ is invertible.

A numerical example

We consider a two-dimensional example, defined by

$$S_t = \begin{pmatrix} B_t^1 + 0.3 B_t^2 \\ \int_0^t B_s^1 \mathrm{d}s \end{pmatrix}.$$

It corresponds to a constant (degenerate) matrix

$$\sigma_t = \begin{pmatrix} 1 & 0.3 \\ 0 & 0 \end{pmatrix}.$$

For the function v we take

$$v(t,x) = \begin{pmatrix} \cos(3x_1) \\ \cos(3x_2) \end{pmatrix},$$

and we set $T = 1$. According to the previous paragraph, ($\mathbf{H}_\Lambda$) is satisfied and an optimal sequence of strategies is given by Theorem 1.5.2. To assess the efficiency of an arbitrary admissible sequence of strategies we set

$$\alpha_n := \frac{N_T^n \langle Z^n \rangle_T}{\left(\int_0^T \mathrm{Tr}(X_t) \mathrm{d}t \right)^2} \qquad \text{and} \qquad \beta_n := \frac{\sqrt{N_T^n} Z_T^n}{\int_0^T \mathrm{Tr}(X_t) \mathrm{d}t}.$$

From Theorem 1.4.2 we must have $\liminf_{n \to +\infty} \alpha_n \geqslant 1$ a.s., while for the optimal sequence the equality holds. The normalized error β_n is also important in practice, however we cannot in general asymptotically control a.s. this quantity. But it is easy to believe that the values of β_n are smaller for strategies where the corresponding values of α_n are smaller, at least in mean. We will illustrate this heuristics in the following.

To simulate the process S on $[0,1]$ we use a thin uniform time mesh with $\bar{n} = 10000$ points. The same mesh is later used to calculate the true value of the stochastic integral and the optimal lower bound equal to $\left(\int_0^T \mathrm{Tr}(X_t) \mathrm{d}t \right)^2$. The hitting times are calculated as well on this mesh. Using this thin grid induces a discrete-time sampling error but by taking $\bar{n}$ quite large as we do, we guess that this error can be neglected in our subsequent results.

We simulate 25 trajectories of the process S on $[0,1]$. Further we test the optimal discretization strategy and the regular deterministic discretization on these trajectories, for different discretization parameters ε_n.

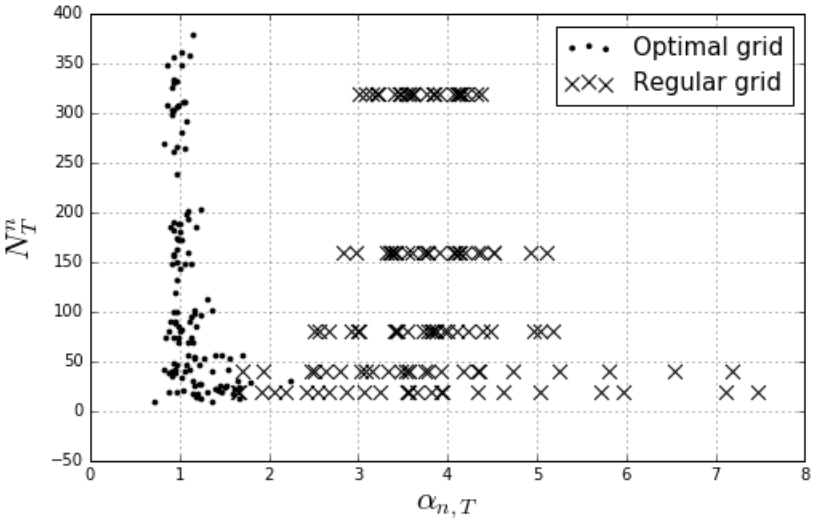

Figure 1.1: The values $\alpha_{n,opt}$ and $\alpha_{n,det}$ with respect to N_T^n.

a) To test the performance of the optimal discretization we take 5 different values of ε_n, namely $0.2, 0.14, 0.1, 0.07, 0.05$, and apply the strategy given in Theorem 1.5.2.

b) Further we test the performance of the deterministic discretization strategy with N_T^n equidistant times, for $N_T^n = 20, 40, 80, 160, 320$ (the values of N_T^n are empirically chosen as approximately equal to the average number of discretization times in the optimal algorithm for the values of ε_n given above).

We denote $(\alpha_{n,opt}, \beta_{n,opt})$ and $(\alpha_{n,det}, \beta_{n,det})$ the pairs (α_n, β_n) respectively for the optimal and the regular deterministic strategy.

Regarding further details of implementation, we refer to [GL14a, Proof of Lemma 3.1] for the detailed construction of the solution to the matrix equation (1.4.1). For the computation of the pseudo-inverse matrix in ($\mathbf{H}_\Lambda$), this is straightforward since σ_t is constant. For the perturbation procedure appearing in (1.5.4), we take $\delta = 0.6 \leq 2(\rho_N - 1) < \frac{2}{3}$ and the function $\chi(x) = \sin(\pi(x \vee 1/2) \wedge 1)$.

Figure 1.1 shows the values of $\alpha_{n,opt}$ and $\alpha_{n,det}$ with respect to the number of the discretization times N_T^n for the optimal and the regular discretization in all the tests belonging to 5 different groups. We observe that the values $\alpha_{n,opt}$ become less and less dispersed and converges to 1 as N_T^n increases ($\varepsilon_n \to 0$), which confirms the theoretical results. In particular, from $N_T^n = 80$ the quality of the algorithm is already good and it largely outperforms the regular discretization.

Figure 1.2 illustrates the pairs (α_n, β_n) for the same 25 simulations, where $\varepsilon_n = 0.05$ was used for the optimal discretization and $N_T^n = 320$ was used for the regular deterministic strategy (i.e. the last group of the tests). As expected from Theorems 1.4.2 and 1.5.2, we observe the inequality $\alpha_{n,opt} < \alpha_{n,det}$ and the limit $\alpha_{n,opt} \approx 1$. Moreover, the inequality $|\beta_{n,opt}| < |\beta_{n,det}|$ holds as well for 21 of the 25 simulations. The empirical variances of the

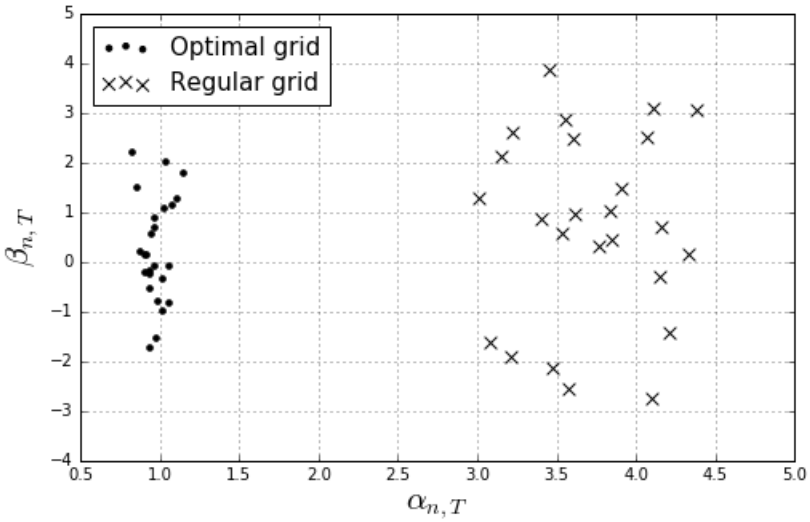

Figure 1.2: The pairs $(\alpha_{n,det}, \beta_{n,det})$ and $(\alpha_{n,opt}, \beta_{n,opt})$ are represented by crosses and points respectively.

values of $\beta_{n,opt}$ and $\beta_{n,det}$ are equal to 1.07 and 3.52 respectively, which is nearly the same ratio as for the corresponding values of α_n: this observation is coherent with the possible property of Central Limit Theorem for β_n, where the limiting distribution would be a mixture of Gaussian distributions with variance roughly equal to α_n. This latter property is just a conjecture which is delicate to prove and left for further research. Anyway, this observation confirms that the almost sure minimization of the limit of α_n helps to reduce the variance of β_n as expected.

1.5.4 Proof of Theorem 1.5.2

The proof is divided into several steps. Assumptions of Theorem 1.5.2 are in force in all this subsection.

Step 1: a reverse relation between X and Λ

Proposition 1.5.4. *The following equality holds*

$$X_t = (\sigma_t)^{\mathsf{T}} \Lambda_t \sigma_t, \qquad \forall t \in [0,T) \quad \text{a.s..} \tag{1.5.5}$$

Proof. We are going to establish the above relation for any given t, with probability 1: however, the reader can check that the negligible set can be the same for all t (as for the definitions of σ, X, Λ) because the arguments used are of deterministic nature.

If σ_t is invertible, $\sigma_t^{\dagger} = \sigma_t^{-1}$ and obviously $X_t = (\sigma_t)^{\mathsf{T}} \Lambda_t \sigma_t$ in view of the definition ($\mathbf{H}_\Lambda$).

Now assume that $\operatorname{rank}(\sigma_t) < d$. By (1.5.1) we have

$$\sigma_t \sigma_t^\dagger \sigma_t = \sigma_t \tag{1.5.6}$$

and the matrix $\sigma_t^\dagger \sigma_t$ is symmetric. We choose an orthonormal basis $(e_i)_{1\le i\le d}$ under which the matrix $\sigma_t^\dagger \sigma_t$ is diagonal, i.e.

$$\sigma_t^\dagger \sigma_t = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_d \end{pmatrix}$$

for some $\alpha_1, \dots, \alpha_d$. If $\sigma_t^1, \dots, \sigma_t^d$ are the column vectors of σ_t (in the basis $(e_i)_{1\le i\le d}$), then from (1.5.6) we get

$$(\alpha_1 \sigma_t^1, \dots, \alpha_d \sigma_t^d) = (\sigma_t^1, \dots, \sigma_t^d). \tag{1.5.7}$$

For any $1 \le i \le d$ if $\sigma_t^i \neq 0$ then we must have $\alpha_i = 1$. On the other hand $k := \operatorname{rank}(\sigma_t^\dagger \sigma_t) \le \operatorname{rank}(\sigma_t) < d$. Hence by permuting the basis elements and using (1.5.6) we can write $\sigma_t^\dagger \sigma_t$ and σ_t in the form:

$$\sigma_t^\dagger \sigma_t = \begin{pmatrix} \mathrm{Id}_k & 0 \\ 0 & 0 \end{pmatrix}, \quad \sigma_t = \begin{pmatrix} \sigma_{1,t}^1 & \dots & \sigma_{1,t}^k & 0 & \dots & 0 \\ \sigma_{2,t}^1 & \dots & \sigma_{2,t}^k & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ \sigma_{d,t}^t & \dots & \sigma_{d,t}^k & 0 & \dots & 0 \end{pmatrix}. \tag{1.5.8}$$

We want to show that $X_t = (\sigma_t)^\mathsf{T} \Lambda_t \sigma_t$ which by the definition of Λ_t is equivalent to

$$X_t = (\sigma_t^\dagger \sigma_t)^\mathsf{T} X_t (\sigma_t^\dagger \sigma_t) = (\sigma_t^\dagger \sigma_t) X_t (\sigma_t^\dagger \sigma_t). \tag{1.5.9}$$

In view of (1.5.8) and since X is symmetric non-negative definite, the equality (1.5.9) is equivalent to the following system of equations:

$$e_i^\mathsf{T} X_t e_i = 0 \quad \text{for} \quad i = k+1, \dots, d, \tag{1.5.10}$$

where (e_i) are the vectors of the basis. We now prove (1.5.10). Let $i \in \{k+1, \dots, d\}$. From the definition of X_t we get

$$2 \operatorname{Tr}(X_t) X_t + 4 X_t^2 = \sigma_t^\mathsf{T} \tilde{C}_t \sigma_t, \tag{1.5.11}$$

where $\tilde{C}_t = (D_x v_t)^\mathsf{T} \sigma_t \sigma_t^\mathsf{T} D_x v_t$. From (1.5.8) it is clear that $\sigma_t e_i = 0$, thus Equation (1.5.11) yields

$$2 \operatorname{Tr}(X_t) e_i^\mathsf{T} X_t e_i + 4 e_i^\mathsf{T} X_t^2 e_i = 0.$$

Both X_t and X_t^2 are in $\mathcal{S}_+^d(\mathbb{R})$, thus both above terms are non-negative, therefore they are equal to 0. Either $\operatorname{Tr}(X_t) = 0$ (implying $X_t = 0$ and (1.5.10)), or $\operatorname{Tr}(X_t) > 0$ and $e_i^\mathsf{T} X_t e_i = 0$. In any case, (1.5.10) holds and we are done. □

Step 2: verification of $(\mathbf{A}_S^{\text{osc.}})$

The stopping times (1.5.4) define a sequence of strategies satisfying $(\mathbf{A}_S^{\text{osc.}})$: this is a consequence of Theorem 1.3.11-*(i)* with $H = \Lambda$. Indeed the existence of the finite random variable C_H stems from (1.5.3).

Step 3: verification of $(\mathbf{A}_N)$

We aim at showing

Proposition 1.5.5. *We have the following convergence*

$$\sum_{\tau_{i-1}^n<T} \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \overset{\text{a.s.}}{\to} \int_0^T \mathrm{Tr}(\Lambda_t \mathrm{d}\langle M\rangle_t) = \int_0^T \mathrm{Tr}(X_t)\mathrm{d}t.$$

Then, in view of Theorem 1.3.11-*(ii)*, we conclude that the sequence of strategies $\mathcal{T} = \{\mathcal{T}_{\varepsilon_n^\delta}^n : n \geq 0\}$ satisfies $(\mathbf{A}_N)$. Combined with Step 2, we have proved that this is an admissible sequence.

Observe that the above result is not a particular case of Proposition (1.3.9) since we do not know if Λ is continuous in time (it is likely not for degenerate σ). To handle this difficulty, we are going to leverage the reverse relation between X and Λ (Step 1), and the continuity of X.

Proof of Proposition 1.5.5. By Itô's lemma like for (1.3.12) and using that Λ is symmetric, we obtain

$$\sum_{\tau_{i-1}^n<T} \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} = 2\int_0^T \Delta S_t^{\mathsf{T}} \Lambda_{\phi(t)} \mathrm{d}S_t + \int_0^T \mathrm{Tr}(\Lambda_{\phi(t)} \mathrm{d}\langle M\rangle_t). \tag{1.5.12}$$

Then

$$\begin{aligned}\int_0^T \mathrm{Tr}(\Lambda_{\phi(t)} \mathrm{d}\langle M\rangle_t) = \int_0^T \mathrm{Tr}(\sigma_t^{\mathsf{T}} \Lambda_{\phi(t)} \sigma_t)\mathrm{d}t = \int_0^T \mathrm{Tr}(\sigma_{\phi(t)}^{\mathsf{T}} \Lambda_{\phi(t)} \sigma_{\phi(t)})\mathrm{d}t \\ + \int_0^T \mathrm{Tr}\left((\sigma_t - \sigma_{\phi(t)})^{\mathsf{T}} \Lambda_{\phi(t)} (\sigma_t + \sigma_{\phi(t)})\right) \mathrm{d}t.\end{aligned}$$

Observe that the first term on the r.h.s. above is equal to $\int_0^T \mathrm{Tr}(X_{\phi(t)})\mathrm{d}t$ owing to Proposition 1.5.4: since X is a.s. bounded continuous and the time step goes to 0 (see (1.3.7) valid under $(\mathbf{A}_S^{\text{osc.}})$), we easily obtain $\int_0^T \mathrm{Tr}(X_{\phi(t)})\mathrm{d}t \overset{\text{a.s.}}{\to} \int_0^T \mathrm{Tr}(X_t)\mathrm{d}t$.

The second term tends to 0 a.s. thanks to the continuity of σ and the uniform bound (1.5.3) on Λ. We have proved

$$\int_0^T \mathrm{Tr}(\Lambda_{\phi(t)} \mathrm{d}\langle M\rangle_t) \overset{\text{a.s.}}{\to} \int_0^T \mathrm{Tr}(X_t)\mathrm{d}t.$$

To complete the proof, in view of (1.5.12) it remains to show that

$$\int_0^T \Delta S_t^{\mathsf{T}} \Lambda_{\phi(t)} \mathrm{d}S_t \overset{\text{a.s.}}{\to} 0.$$

The a.s.-convergence to 0 of the contribution $\int_0^T \Delta S_t^{\mathsf{T}} \Lambda_{\phi(t)} \mathrm{d}A_t$ is proved as for (1.3.13), using $(\mathbf{A}_S^{\text{osc.}})$ and $(\mathbf{H}_\Lambda)$. The second contribution $K_T^n := \int_0^T \Delta S_t^{\mathsf{T}} \Lambda_{\phi(t)} \mathrm{d}M_t$ is a local martingale, which bracket is bounded by ε_n^2 up to a random finite constant (use again $(\mathbf{A}_S^{\text{osc.}})$ and $(\mathbf{H}_\Lambda)$). Consequently, an application of Lemma 1.3.1 with $p = 2$, ensures that $K_T^n \overset{\text{a.s.}}{\to} 0$. We are done. □

Final step: completion of proof of Theorem 1.5.2

So far, we have showed that the strategy $\mathcal{T} = \{\mathcal{T}^{(n)}_{\varepsilon_n^\delta} : n \geq 0\}$ is admissible. We now prove that

$$\lim_{n\to+\infty} N_T^n \langle Z^n \rangle_T = \left(\int_0^T \mathrm{Tr}(X_t) \mathrm{d}t \right)^2 \qquad \text{a.s..}$$

First, proceeding as (1.3.16), we write that $\varepsilon_n^{2+\delta} N_T^n$ equals

$$\varepsilon_n^{2+\delta} + \sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} + \sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} (\Lambda^{(n)}_{\tau_{i-1}^n} - \Lambda_{\tau_{i-1}^n}) \Delta S_{\tau_i^n}. \tag{1.5.13}$$

The first term converges to 0, as well as the last term (proceeding as for (1.3.17)), while the second one converges a.s. to $\int_0^T \mathrm{Tr}(\Lambda_t \mathrm{d}\langle M \rangle_t)$ (Proposition 1.5.5). To summarize, we have justified

$$\lim_{n\to+\infty} \varepsilon_n^{2+\delta} N_T^n = \int_0^T \mathrm{Tr}(\Lambda_t \mathrm{d}\langle M \rangle_t) = \int_0^T \mathrm{Tr}(X_t) \mathrm{d}t \qquad \text{a.s..} \tag{1.5.14}$$

Thus it remains to show that

$$\lim_{n\to+\infty} \varepsilon_n^{-(2+\delta)} \langle Z^n \rangle_T = \int_0^T \mathrm{Tr}(X_t) \mathrm{d}t \qquad \text{a.s..} \tag{1.5.15}$$

Starting from (1.4.6), write $\langle Z^n \rangle_T$ in the form

$$\langle Z^n \rangle_T = \sum_{\tau_{i-1}^n < T} (\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda^{(n)}_{\tau_{i-1}^n} \Delta S_{\tau_i^n})^2 + e^n_{0,T} + e^n_{1,T} + e^n_{2,T} + e^n_{3,T} + e^n_{4,T} + e^n_{5,T},$$

where $e^n_{0,T}, e^n_{1,T}, e^n_{2,T}, e^n_{3,T}$ are defined as in the proof of Theorem 1.4.2 and the other terms are defined as follows:

$$e^n_{4,T} := \sum_{\tau_{i-1}^n < T} (\Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n})^2 - \sum_{\tau_{i-1}^n < T} (\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n})^2,$$

$$e^n_{5,T} := \sum_{\tau_{i-1}^n < T} (\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n})^2 - \sum_{\tau_{i-1}^n < T} (\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda^{(n)}_{\tau_{i-1}^n} \Delta S_{\tau_i^n})^2.$$

First notice that for each $i \leq N_T^n - 1$ we have $\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n}^{(n)} \Delta S_{\tau_i^n} = \varepsilon_n^{2+\delta}$, thus

$$\begin{aligned}
&\varepsilon_n^{-(2+\delta)} \sum_{\tau_{i-1}^n < T} (\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n}^{(n)} \Delta S_{\tau_i^n})^2 \\
&= \sum_{\tau_i^n < T} \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n}^{(n)} \Delta S_{\tau_i^n} + \varepsilon_n^{-(2+\delta)} (\Delta S_T^{\mathsf{T}} \Lambda_{\tau_{N_T^n-1}^n}^{(n)} \Delta S_T)^2 \\
&\overset{\text{a.s.}}{\to} \int_0^T \mathrm{Tr}(\Lambda_t \mathrm{d}\langle M \rangle_t) = \int_0^T \mathrm{Tr}(X_t) \mathrm{d}t,
\end{aligned}$$

where the last convergence is derived similarly to that of (1.5.13).

Moreover, from (1.4.7) in the proof of Theorem 1.4.2, we already have (for ε_n small enough so that $\varepsilon_n \leq 1$ and since $2 + \delta \leq 2\rho_N$)

$$\varepsilon_n^{-(2+\delta)} e_{i,T}^n \leq \varepsilon_n^{-2\rho_N} e_{i,T}^n \overset{\text{a.s.}}{\to} 0 \qquad \text{a.s. for } i = 0, 1, 2, 3.$$

To complete the proof of Theorem 1.5.2, it remains only to prove that

$$\varepsilon_n^{-(2+\delta)} e_{i,T}^n \overset{\text{a.s.}}{\to} 0 \qquad \text{a.s. for } i = 4, 5.$$

We start with $e_{5,T}^n$:

$$\begin{aligned}
|\varepsilon_n^{-(2+\delta)} e_{5,T}^n| &\leq \sum_{\tau_{i-1}^n < T} (\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} + \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n}^{(n)} \Delta S_{\tau_i^n}) \\
&\qquad \times |\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} - \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n}^{(n)} \Delta S_{\tau_i^n}| \varepsilon_n^{-(2+\delta)} \\
&\leq \sum_{\tau_{i-1}^n < T} \varepsilon_n^{\delta} \chi_{\varepsilon_n^{\delta}}(\lambda_{\min}(\Lambda_{\tau_{i-1}^n})) |\Delta S_{\tau_i^n}|^2 |2\varepsilon_n^{-2-\delta} \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n}^{(n)} \Delta S_{\tau_i^n}| \\
&\leq 2\varepsilon_n^{\delta} \sum_{\tau_{i-1}^n < T} |\Delta S_{\tau_i^n}|^2 \overset{\text{a.s.}}{\to} 0
\end{aligned}$$

thanks to Proposition 1.3.9.

Finally, we analyse $e_{4,T}^n$. From its definition, Proposition 1.5.4 and ($\mathbf{H}_\Lambda$), we get

$$\begin{aligned}
&|\varepsilon_n^{-(2+\delta)} e_{4,T}^n| \\
&\leq \varepsilon_n^{-(2+\delta)} \sum_{\tau_{i-1}^n < T} \left| \Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n} - \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \right| \left(\Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n} + \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \right) \\
&\leq \varepsilon_n^{-(2+\delta)} c_{(1.5.3)} \sup_{1 \leq i \leq N_T^n} \sup_{t \in (\tau_{i-1}^n, \tau_i^n]} |\Delta S_t + \sigma_{\phi(t)} \Delta B_t| \left| \int_{\phi(t)}^t \Delta \sigma_s \mathrm{d}B_s + \Delta A_t \right| \\
&\qquad \times \left(\Delta B_{\tau_i^n}^{\mathsf{T}} X_{\tau_{i-1}^n} \Delta B_{\tau_i^n} + \Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \right). \qquad (1.5.16)
\end{aligned}$$

Now we apply twice Lemma 1.3.3-(ii), first taking $\theta = 0$ and second taking $\theta = \theta_\sigma$: it readily

follows that for any given $\rho > 0$, we have a.s. for any $n \in \mathbb{N}$

$$\sup_{1\le i\le N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}\Big(|\Delta M_t| + |\sigma_{\phi(t)}\Delta B_t|\Big) \le C_0\varepsilon_n^{1-\rho}, \tag{1.5.17}$$

$$\sup_{1\le i\le N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}\left|\int_{\phi(t)}^t \Delta\sigma_s \mathrm{d}B_s\right| \le C_0\varepsilon_n^{1+\theta_\sigma-\rho}. \tag{1.5.18}$$

Moreover by Corollary 1.3.7 we have

$$\sup_{1\le i\le N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}|\Delta A_t| \le C_0\varepsilon_n^{2\alpha-\rho}.$$

The last factor in the r.h.s. of (1.5.16) converges a.s. to a finite random variable (Propositions 1.3.8 and 1.5.5). Combining this with the above estimates, the inequality (1.5.16) becomes

$$|\varepsilon_n^{-(2+\delta)}e_{4,T}^n| \le C_0\varepsilon_n^{-2-\delta}\varepsilon_n^{1-\rho}(\varepsilon_n^{1+\theta_\sigma-\rho} + \varepsilon_n^{2\alpha-\rho}).$$

It is now easy to see that, since we have chosen $\delta < \theta_\sigma$ and $\delta < 2\alpha - 1$, we can take ρ small enough so that $\varepsilon_n^{-(2+\delta)}e_{4,T}^n \to 0$. The proof is finished. □

1.5.5 Proof of Proposition 1.5.3

Consider the equation solved by X_t (see (1.4.1) and (1.4.2)), and multiply it by $\sigma_t^\dagger$ from the right and by $(\sigma_t^\dagger)^\mathsf{T}$ from the left: it gives

$$2\,\mathrm{Tr}(X_t)(\sigma_t^\dagger)^\mathsf{T}X_t\sigma_t^\dagger + 4(\sigma_t^\dagger)^\mathsf{T}X_t^2\sigma_t^\dagger = (\sigma_t\sigma_t^\dagger)^\mathsf{T}\tilde{C}_t(\sigma_t\sigma_t^\dagger)$$

where $\tilde{C}_t = (D_x v_t)^\mathsf{T}\sigma_t\sigma_t^\mathsf{T}D_x v_t$. Take the trace, use that $(\sigma_t^\dagger)^\mathsf{T}X_t^2\sigma_t^\dagger \in \mathcal{S}_+^d(\mathbb{R})$, in order to obtain

$$2\,\mathrm{Tr}(X_t)\,\mathrm{Tr}\,(\Lambda_t) \le \mathrm{Tr}\left((\sigma_t\sigma_t^\dagger)^\mathsf{T}\tilde{C}_t(\sigma_t\sigma_t^\dagger)\right).$$

Recall the inequality $\mathrm{Tr}(\mathcal{S}\mathcal{S}') \le \mathrm{Tr}(\mathcal{S})\,\mathrm{Tr}(\mathcal{S}')$ for any non-negative definite symmetric matrices $\mathcal{S}$ and $\mathcal{S}'$. Thus, $\mathrm{Tr}((\sigma_t\sigma_t^\dagger)^\mathsf{T}\tilde{C}_t(\sigma_t\sigma_t^\dagger)) \le d^2\,\mathrm{Tr}(\tilde{C}_t)$ where we have used the easy inequality $\mathrm{Tr}(\sigma_t\sigma_t^\dagger) \le d$. Note that the above inequalities are of deterministic nature and therefore they hold for any t with probability 1 (the full set is the one allowing to define $X, \Lambda, \sigma, \tilde{C}$). Invoking $(\mathbf{H}_M)$ and $(\mathbf{H}_v)$ to control $\tilde{C}$, we deduce that there exists a non-negative random variable $\tilde{c}$, finite a.s., such that

$$\mathrm{Tr}(X_t)\,\mathrm{Tr}(\Lambda_t) \le \tilde{c}, \qquad \forall t \in [0,T] \quad \text{a.s.}. \tag{1.5.19}$$

Owing to the condition $(\mathbf{H}_C)$, $X_t \neq 0$ for any $t \in [0,T]$ a.s., and by continuity of X_t, we get that $\inf_{t\in[0,T]}\mathrm{Tr}(X_t) > 0$ a.s. and we conclude to $(\mathbf{H}_\Lambda)$ thanks to (1.5.19). □

1.A Appendix

1.A.1 Proof of the Lemma 1.4.3

In view of ($\mathbf{H}_v$) there exists $\Omega_{\mathcal{D}}$ with $\mathbb{P}(\Omega_{\mathcal{D}}) = 1$ such that for every $\omega \in \Omega_{\mathcal{D}}$ there is $\delta(\omega) > 0$ such that, for any $\mathcal{A} \in \{D_{x_j}, D^2_{x_j x_k}, D_t : 1 \le j, k \le d\}$,

$$\sup_{0\le t<T} \sup_{|x-S_t(\omega)|\le\delta(\omega)} |\mathcal{A}v(t,x)| < +\infty.$$

Since $\sup_{1\le i\le N_T^n} \Delta\tau_i^n \overset{\text{a.s.}}{\to} 0$ and S is continuous on $[0,T]$, there exists a set $\Omega_{\mathcal{C}}$ of full measure such that, for every $\omega \in \Omega_{\mathcal{C}}$, for n large enough we have

$$\sup_{0\le s,t\le T, |t-s|\le \sup_{1\le i\le N_T^n}\Delta\tau_i^n} |S_t(\omega) - S_s(\omega)| \le \delta(\omega).$$

Hence for $\omega \in \Omega_{\mathcal{C}} \cap \Omega_{\mathcal{D}}$, for n large enough, by a Taylor formula we obtain (the dependence on ω is further omitted, we assume $\omega \in \Omega_{\mathcal{C}} \cap \Omega_{\mathcal{D}}$)

$$\sup_{t\in(\tau_{i-1}^n,\tau_i^n]} |v(t,S_t) - v(\tau_{i-1}^n, S_{\tau_{i-1}^n}) - D_x v(\tau_{i-1}^n, S_{\tau_{i-1}^n})| \le C_0 \left(\Delta\tau_i^n + \sup_{t\in(\tau_{i-1}^n,\tau_i^n]} |\Delta S_t|^2\right).$$

Plugging this estimate into $\langle R^n\rangle_T$ we obtain that a.s., for n large enough,

$$\varepsilon_n^{2-4\rho_N}\langle R^n\rangle_T \le C_0\varepsilon_n^{2-4\rho_N} \sum_{\tau_{i-1}^n<T} \left((\Delta\tau_i^n)^3 + \Delta\tau_i^n \sup_{\tau_{i-1}^n\le t\le\tau_i^n} |\Delta S_t|^4\right).$$

We deduce that $\varepsilon_n^{2-4\rho_N}\langle R^n\rangle_T \overset{\text{a.s.}}{\to} 0$ since

- for any $\rho > 0$, $\varepsilon_n^{2-4\rho_N}\sum_{\tau_{i-1}^n<T}(\Delta\tau_i^n)^3 \le \varepsilon_n^{2-4\rho_N} N_T^n \sup_{1\le i\le N_T}(\Delta\tau_i^n)^3 \le C_0\varepsilon_n^{8-6\rho_N-\rho}$ by using Lemma 1.3.2-*(ii)*, thus it converges to 0 since $\rho_N < 4/3$,
- $\varepsilon_n^{2-4\rho_N}\sum_{\tau_{i-1}^n<T}\Delta\tau_i^n \sup_{\tau_{i-1}^n\le t\le\tau_i^n}|\Delta S_t|^4 \le C_0\varepsilon_n^{6-4\rho_N}T \to 0$ a.s..

We are done. □

1.A.2 Almost sure convergence using domination in expectation

The next result allows to prove the a.s. convergence of a dominated process U using that of a dominating process V, the domination relation being in expectation. Its use is crucial in our analysis.

Lemma 1.A.1 ([GL14a, Lemma 2.2]). *Let $\mathcal{C}_0^+$ be the set of non-negative continuous adapted processes, vanishing at $t = 0$. Let $(U^n)_{n\ge0}$ and $(V^n)_{n\ge0}$ be two sequences of processes in $\mathcal{C}_0^+$. Assume that*

(i) $t \mapsto V_t^n$ is a non-decreasing function on $[0,T]$, a.s.*;*

(ii) *the series* $\sum_{n\geq 0} V_T^n$ *converges* a.s.*;*

(iii) *there is a constant* $c \geq 0$ *such that, for every* $n \in \mathbb{N}$*,* $k \in \mathbb{N}$ *and* $t \in [0,T]$*, we have*

$$\mathbb{E}[U^n_{t\wedge\theta_k}] \leq c\mathbb{E}[V^n_{t\wedge\theta_k}]$$

with the stopping time $\theta_k := \inf\{s \in [0,T] : \bar{V}_s \geq k\}$[1] *setting* $\bar{V}_t = \sum_{n\geq 0} V_t^n$*.*

Then for any $t \in [0,T]$*, the series* $\sum_{n\geq 0} U_t^n$ *converges* a.s.*. As a consequence,* $U_t^n \overset{\text{a.s.}}{\to} 0$*.*

[1] with the usual convention $\inf \emptyset = +\infty$.

Chapter 2

Model-adaptive optimal discretization

Contents

2.1 Introduction

Statement of the problem. In this chapter we continue the study of the optimal discretization problem for stochastic integrals with respect to Brownian semimartingales initiated in [Fuk11a] and further developed in [GL14a] and Chapter 1. Our goal here is to construct an adaptive version of the optimal discretization algorithm from Chapter 1 that does not require any prior knowledge about the model.

Let $T > 0$ be fixed. We consider a Brownian semimartingale $(S_t)_{t\in[0,T]}$, valued in $\mathbb{R}^d$, of the form

$$S_t := M_t + A_t, \tag{2.1.1}$$

where M is written as a Brownian local martingale

$$M_t := \int_0^t \sigma_s \mathrm{d}B_s, \tag{2.1.2}$$

and A is adapted, Hölder-continuous and of finite variation.

The problem of optimal discretization consists of finding optimal stopping times to discretize the stochastic integral

$$\int_0^T v(t, S_t) \cdot \mathrm{d}S_t \tag{2.1.3}$$

for a $\mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$-function v, with the highest possible accuracy for a given number of discretization times. More precisely we aim at finding a sequence of strategies $\{\mathcal{T}^n\}_{n\geq 0}$ where each $\mathcal{T}^n$ is an increasing sequence of stopping times $\{\tau_i^n\}_{i\geq 0}$ that will achieve the lower bound of the limiting renormalized discretization error

$$\liminf_{n\to+\infty} N_T^n \langle Z^n \rangle_T, \tag{2.1.4}$$

within the class of admissible sequences of discretization strategies. Here N_T^n is the number of stopping times in $\mathcal{T}^n$, Z^n is the discretization error for the grid $\mathcal{T}^n$ defined by

$$Z_s^n := \int_0^s v(t, S_t) \cdot \mathrm{d}S_t - \sum_{\tau_{i-1}^n < s} v(\tau_{i-1}^n, S_{\tau_{i-1}^n}) \cdot (S_{\tau_i^n \wedge s} - S_{\tau_{i-1}^n}), \quad 0 \leq s \leq T, \tag{2.1.5}$$

and $\langle Z^n \rangle$ is the bracket process of the continuous semimartingale Z^n. In the aforementioned references, optimal sequences of $(\mathcal{T}^n)_{n\geq 0}$ are derived under some assumptions, but these strategies strongly depend on the model for S, in particular on the σ-process. As a difference, in this work we suppose that no prior knowledge about the diffusion coefficient of the underlying process S is given. We do not assume neither a diffusion model for S nor a parametric form for σ. The process S of the form (2.1.1)-(2.1.2) is quite arbitrary and we only suppose that it satisfies some mild regularity and non-degeneracy assumptions. Thus a model-adaptive version of the optimal discretization algorithm designed in Chapter 1 is needed in order to make the latter algorithm applicable.

Regarding applications, having at hand an algorithm able to adapt automatically to the model without fully identifying it, is quite useful; it is usually referred to as data-driven algorithm. The problem of discretizing (2.1.3) can be interpreted as a pursuit problem where the target evolves like $s \to \int_0^s v(t, S_t) \cdot \mathrm{d}S_t$ and S is a random path modeling the system. The interpretation is application-dependent: in random mechanics [KS86], the system is a controlled object subject to random forces; in finance [GL14a], the system is an investment portfolio. In these cases, S should be seen as a black-box process, which we don't know exactly the coefficients of.

The notion of admissible sequence $\mathcal{T}$ of discretization strategies is parametrized by a real number $\rho_N \geq 1$ and a sequence of positive numbers $(\varepsilon_n)_{n\geq 0}$ converging to zero, see Definition 2.2.1 for details. To get the a.s. optimality we need additional assumption of the square summability of $(\varepsilon_n)_{n\geq 0}$, otherwise the optimality holds in probability.

In Chapter 1 an optimal sequence of strategies is constructed under very mild assumptions on the model. The optimal stopping times are given in the form

$$\tau_i^n = \varphi\left(\sigma_{\tau_{i-1}^n}, D_x v(\tau_{i-1}^n, S_{\tau_{i-1}^n}), (S_t - S_{\tau_{i-1}^n})_{t \geq \tau_{i-1}^n}\right), \tag{2.1.6}$$

for some complicated function φ. These stopping times are interpreted as exit times of random ellipsoids. If we think that S comes from statistical data (as a black-box), the only unobservable process that is used to construct optimal strategies given by Theorem 1.5.2 is the diffusion coefficient σ. The rest is based on the observation of the semimartingale S. In practice a statistical estimator of the process σ must be used instead.

This rises the question of the robustness of the optimal discretization with respect to the estimation error of σ. In view of the lack of continuity of the function φ from (2.1.6), of its non-linearity and more generally because of its complicated structure, this problem is quite non-trivial. This issue is quite fundamental regarding the applications listed before. We exemplify now the importance of taking the right σ in the minimization of $N_T^n\langle Z^n\rangle_T$.

Example 2.1.1 (Non optimality for misspecified model)**.** *Consider the case $d = 2$, $S_t = M_t = \sigma B_t$ (i.e. $A \equiv 0$) with*

$$\sigma = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \text{ and } v(t, S^1, S^2) = \begin{pmatrix} \sqrt{14}\ S^1 \\ \sqrt{104}\ S^2 \end{pmatrix}$$

(the constants are chosen purely for the sake of analytical tractability). Suppose that the optimizer has no prior knowledge about the diffusion coefficient σ and he/she believes that σ is proportional to identity, say $\tilde{\sigma} = \mathrm{Id}_2$ instead of the true matrix σ as above. Then suppose that he/she constructs a discretization algorithm from Theorem 1.5.2 (for some $\varepsilon_n \to 0$ and $\rho_N \geq 1$) based on this assumption. Let $\tilde{N}_T^n\langle\tilde{Z}^n\rangle_T$ denote the renormalized error for this sequence of discretization strategies. Then the sequence is suboptimal in the sense that

$$\tilde{N}_T^n\langle\tilde{Z}^n\rangle_T \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} \textit{(Opt. Lower Bound)} + \beta T^2, \tag{2.1.7}$$

where $\beta > 0$ (the exact calculation of β is given in Equation (2.A.7)).

The first term on the above right hand side is the lower bound over all admissible strategies. The proof of (2.1.7) is postponed to Appendix 2.A.4. This clearly shows that a model misspecification (via taking an erroneous diffusion coefficient) likely leads to suboptimal results.

In this work we investigate this issue and our goal is to find sufficient assumptions on a general sequence of estimators σ_t^n of σ_t ensuring the optimality of the resulting sequence of strategies. Rather than assuming a particular parametric model for σ we only suppose Hölder-continuity and non-degeneracy of σ. In such a general framework (nonparametric, non-Markovian, and multidimensional), presumably the only accessible class of estimators for σ_t is the one based on a weighted moving average estimation (Kernel techniques). It assumes that, as intuitively expected, recently realized quadratic covariations are good predictions of the current value of the instantaneous covariances. In this work we prove an optimality result for a general class of weighted moving average estimators and specify some sufficient joint conditions on the lookback estimation period and the frequency of estimation in order to obtain the asymptotic optimality in the discretization problem (2.1.4)-(2.1.5).

Background results. The study of optimal discretization of stochastic integrals using random grids has been initiated by [Fuk11a] in the case of dimension $d = 1$, but instead of (2.1.4) the author considers a criterion in expectation for both terms, i.e. $\mathbb{E}\left(N_T^n\right)\mathbb{E}\left(\langle Z^n\rangle_T\right)$. The point of view of pointwise approximation of scalar SDEs is investigated in [MG04]. Optimal convergence rates of $\mathbb{E}\left(\langle Z^n\rangle_T\right)$ for deterministic grids are investigated in [GG04].

The pathwise minimization of (2.1.4) has been addressed in a multi-dimensional martingale setting $d \geq 1$ in [GL14a]. The authors have made a very useful observation that, although more mathematically demanding at the first sight, the use of *almost sure* criterion simplifies the analysis and makes it possible to tackle the multidimensional case. The a.s. framework has proved to be much more flexible and in this work too, it is crucial for the construction of adaptive optimal discretization schemes. In Chapter 1 the study of optimal discretization is generalized for non-elliptic σ and a general Hölder-continuous finite variation part. The framework of Chapter 1 covers most of the examples that are interesting in applications.

In our work we combine previous studies on optimal discretization with standard techniques of non-parametric estimation of the diffusion coefficient in a multidimensional setting. The problem of retrieving diffusion coefficient from observations of S is classic and has been studied by many authors. In particular, in [Hof97] and [FZ93] estimation techniques are derived in the non-parametric setting and for irregular samplings respectively. Although quite different in the mathematical tools, these two studies are the closest to our work.

Our contributions. In the current work, we prove optimality results for an adaptive discretization strategy that does not assume any prior knowledge on the diffusion coefficient model. This is the first result on adaptive strategies for optimal discretization problems. In particular,

- we prove that if σ is estimated by σ^n at some rate, a strategy of the form (2.1.6) but with σ^n instead of σ yields an optimal sequence for the problem (2.1.4). This reads as a robustness result w.r.t. the model.

- we prove the optimality of the strategies based on general weighted moving average estimators provided that certains conditions on the lookback period and the estimation frequency are fullfilled.

Optimality results in almost sure sense (Theorem 2.2.5(bis)) or in probability (Theorem 2.2.5) are derived. In [GL14a] and Chapter 1 only a.s. results are established.

Organisation of the chapter. In Section 2.2 we present the model under study and list the assumptions used throughout the chapter. In Section 2.3 we investigate the optimality of the discretization strategies for a general sequence of diffusion coefficient estimators. In Section 2.4 the same issues are considered for a particular class of weighted moving average estimators of $(\sigma_t)_{t\in[0,T]}$. A numerical experiment is presented in Section 2.5.

Notation used throughout the chapter.

- $x \cdot y$ stands for the scalar product between two vectors x and y, $|x| = (x \cdot x)^{\frac{1}{2}}$ denotes the Euclidean norm of x.
- For a given matrix A, A^T denotes its transpose, $\mathrm{Tr}(A)$ is its trace (when A is square), Id_d stands for the identity matrix of size d.
- We write $|A|$ for the spectral norm of a matrix A, i.e. the square-root of the largest eigenvalue of $A^\mathsf{T}A$. $\lambda_{\min}(A)$ is the square-root of the smallest eigenvalue of $A^\mathsf{T}A$.
- $\mathcal{S}^d, \mathcal{S}^d_+$ and $\mathcal{S}^d_{++}$ are respectively the sets of symmetric, symmetric non-negative definite and symmetric positive-definite $d \times d$ matrices with real coefficients.
- For $A \in \mathcal{S}^d_+$, we denote by $A^{1/2} \in \mathcal{S}^d_+$ the principal square root of A. Recall that for an orthogonal diagonalization $A = U^\mathsf{T}DU$ of A, the principal square root $A^{1/2}$ is given by $U^\mathsf{T}D^{1/2}U$ where $D^{1/2}$ is diagonal with entries $D_{11}^{1/2}, \dots, D_{dd}^{1/2}$.
- For $A \in \mathcal{S}^d_+$ and a real number $\mu > 0$ we define

$$[A]^\mu := A + \mu\chi(\lambda_{\min}(A)/\mu)\,\mathrm{Id}_d, \tag{2.1.8}$$

  where χ is a smooth function such that $\mathbf{1}_{(-\infty,1/2]} \le \chi(.) \le \mathbf{1}_{(-\infty,1]}$. We easily check that

$$\lambda_{\min}([A]^\mu) \ge \mu/2. \tag{2.1.9}$$

- In what follows, C_0 stands for a finite non-negative random variable, that may change from line to line.

2.2 Model and main results

2.2.1 Model and assumptions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space supporting a standard d-dimensional Brownian motion $(B_t)_{t\in[0,T]}$. Denote $(\mathcal{F}_t)_{t\in[0,T]}$ the filtration generated by B, augmented by the $\mathbb{P}$-null sets.

We consider a d-dimensional Brownian semimartingale $S = M + A$ of the form (2.1.1) and (2.1.2).

Here we state the assumptions on the processes under consideration.

Assumptions 1.

($\mathbf{H}_A$): *the process A is adapted, continuous, with finite variation and verifies for some $\alpha \in (1/2, 1]$*

$$|A_t - A_s| \le C_0|t-s|^\alpha \qquad \forall s,t \in [0,T] \quad a.s.$$

($\mathbf{H}_\sigma$): *the process $(\sigma_t)_{t\in[0,T]}$ is a continuous adapted $d\times d$-matrix process such that the value σ_t is a.s. non-zero for any $t \in [0,T]$. Moreover it verifies the following Hölder property for some $\theta_\sigma \in (0,1]$:*

$$|\sigma_t - \sigma_s| \le C_0|t-s|^{\theta_\sigma/2} \qquad \forall s,t \in [0,T] \quad a.s.$$

($\mathbf{H}_\sigma^{ellip.}$): *$\sigma_t$ is a non-degenerate matrix for all $t \in [0,T]$ a.s.*

($\mathbf{H}_v$): *the function v belongs to the class $\mathcal{C}^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R}^d)$.*

($\mathbf{H}_{Dv}$): *$D_x v_t := D_x v(t,S_t)$ is a non-zero matrix for all $t \in [0,T]$ a.s.*

We suppose Assumptions 1 to be verified in all subsequent sections of the chapter and the constants θ_σ and α are fixed from now on.

In particular, (1) and (1) imply that the inverse matrix process σ_t^{-1} is continuous, since the determinant and the adjugate matrix of σ_t are continuous (their components are given by polynomials of the components of σ_t) and the determinant is positive.

Remark that in Chapter 1 an optimal strategy is derived under more general assumptions on σ_t, namely without (1), by using at some places the Moore-Penrose pseudo-inverse of σ_t. However here we restrict to the case where σ_t is a.s. non-degenerate for all t, because the pseudo-inverse is discontinuous w.r.t. perturbations (occurring when σ is replaced by σ^n). Removing (1) in the current model-adaptive setting leads to delicate issues that are seemingly quite difficult to overcome.

2.2.2 Background results: optimal discretization when the model is known

First, we briefly recall the construction of the optimal sequence of strategies from Chapter 1 in the particular case of (1), giving almost sure convergence results. Then, we extend them to the convergence in probability sense.

Almost sure convergence

The following matrix equation plays an important role in the analysis:

$$2\,\mathrm{Tr}(x)x + 4x^2 = cc^\mathsf{T}, \tag{2.2.1}$$

here c is a $d\times d$-matrix and x is unknown. From [GL14a, Lemma 3.1], there exists a unique $x \in \mathcal{S}_+^d$ solution to (2.2.1). It can be shown that this solution is continuous in c.

The definition of an admissible sequence of strategies is parametrized by the choice of a sequence $(\varepsilon_n)_{n\geq 0}$ of positive real numbers and a real number $\rho_N \geq 1$. Here we assume

$$\sum_{n\geq 0} \varepsilon_n^2 < +\infty. \tag{2.2.2}$$

Definition 2.2.1. *A sequence of strategies $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$, where $\mathcal{T}^n := \{\tau_0^n = 0 < \cdots < \tau_i^n < \cdots < \tau_{N_T^n}^n = T\}$ is a finite sequence of stopping times, is admissible if the two following conditions hold:*

($\mathbf{A}^{\mathrm{osc.}}$): $\sup_{n\geq 0}\left(\varepsilon_n^{-1}\sup_{1\leq i\leq N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}|S_t - S_{\tau_{i-1}^n}|\right) < +\infty, \quad a.s.$

($\mathbf{A}_N$): $\sup_{n\geq 0}(\varepsilon_n^{2\rho_N} N_T^n) < +\infty, \quad a.s.$

The set of admissible sequences is denoted by $\mathcal{T}^{\mathrm{adm.}}$.

Let ρ_N be fixed such that

$$1 \leq \rho_N < \left(\frac{\theta_\sigma}{2} + 1\right) \wedge \frac{4}{3} \wedge \left(\alpha + \frac{1}{2}\right). \tag{2.2.3}$$

Observe that the larger ρ_N, the larger the family $\mathcal{T}^{\text{adm.}}$. However, to keep the presentation concise, in what follows the notion of admissibility is used without anymore reference to ρ_N and to the square-summable sequence $(\varepsilon_n)_{n\geq 0}$.

Some optimal results in a.s. sense are established in Chapter 1. First, by Theorem 1.4.2, for any sequence in $\mathcal{T}^{\text{adm.}}$,

$$\liminf_{n\to+\infty} N_T^n \langle Z^n \rangle_T \geq \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2 \qquad a.s., \tag{2.2.4}$$

where X_t is the solution of (2.2.1) with $c = C_t := \sigma_t^\mathsf{T}(D_x v_t)^\mathsf{T}\sigma_t$ for all $0 \leq t \leq T$. In addition, the lower bound is attained by a sequence of hitting times of random ellipsoids, that are admissible as soon as $\rho_N > 1$. Namely, consider the $\mathcal{S}_+^d$-valued process defined by

$$\Lambda_t := (\sigma_t^{-1})^\mathsf{T} X_t \sigma_t^{-1} \qquad \text{for } \ 0 \leq t \leq T, \tag{2.2.5}$$

and choose δ such that $0 < \delta \leq 2(\rho_N - 1)$ (still under the condition $\rho_N > 1$). Then for a given $n \in \mathbb{N}$ define the strategy $\mathcal{T}_{\varepsilon_n^\delta}^n$ as $(\tau_i^n)_{0\leq i\leq N_T^n}$ verifying

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T}[\Lambda_{\tau_{i-1}^n}]^{\varepsilon_n^\delta}(S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^{2+\delta}\} \wedge T. \end{cases} \tag{2.2.6}$$

This reads as iterative exit times of random ellipsoids (parametrized by $[\Lambda_t]^{\varepsilon_n^\delta}$ defined as in (2.1.8)). From Theorem 1.5.2, the sequence of strategies $\mathcal{T} = \{\mathcal{T}_{\varepsilon_n^\delta}^n : n \geq 0\}$ is admissible and asymptotically optimal, i.e.

$$\lim_{n\to+\infty} N_T^n \langle Z^n \rangle_T = \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2 \qquad a.s. \tag{2.2.7}$$

Convergence in probability

Now we remove the square-summability condition (2.2.2), i.e. we only assume $\varepsilon_n \to 0$, and we extend the previous results to convergence in probability. This is made by using a subsequence principle stated below.

Lemma 2.2.2 ([Bil95, Theorem 20.5])**.** *Consider real-valued random variables. $\mathcal{X}_n \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} \mathcal{X}$ if, and only if, for any subsequence $(\mathcal{X}_{\iota(n)})_{n\geq 0}$ of $(\mathcal{X}_n)_{n\geq 0}$, we can extract another subsequence $(\mathcal{X}_{\iota\circ\iota'(n)})_{n\geq 0}$ such that $\mathcal{X}_{\iota\circ\iota'(n)} \underset{n\to+\infty}{\overset{a.s.}{\to}} \mathcal{X}$.*

With this lemma at hand, we are now in a position to clearly motivate the next definition of admissibility, which is suitable for convergences in probability.

Definition 2.2.3. *A sequence of strategies* $\{\mathcal{T}^n\}_{n\geq 0}$ *is called* $\mathbb{P}$-admissible *for a real number* $\rho_N \geq 1$ *and a sequence of positive real numbers* $(\varepsilon_n)_{n\geq 0}$ *converging to 0 if for any subsequence* $(\varepsilon_{\iota(n)})_{n\geq 0}$ *of* $(\varepsilon_n)_{n\geq 0}$ *, we can extract another subsequence* $(\varepsilon_{\iota\circ\iota'(n)})_{n\geq 0}$ *such that* $\sum_{n\geq 0}\varepsilon^2_{\iota\circ\iota'(n)} < \infty$ *and* $\{\mathcal{T}^{\iota\circ\iota'(n)}\}_{n\geq 0}$ *is admissible for* ρ_N *and* $(\varepsilon_{\iota\circ\iota'(n)})_{n\geq 0}$ *in the sense of Definition 2.2.1.*

An easy example of such $\mathbb{P}$-admissible sequence is, for instance, $\{\mathcal{T}^n\}_{n\geq 0}$ fulfilling both 2.2.1 and 2.2.1 (with $\varepsilon_n \to 0$ only).

Theorem 2.2.4. *Let Assumptions 1 hold and let* $\rho_N \geq 1$*. For any* $\mathbb{P}$*-admissible sequence of strategies* $\{\mathcal{T}^n\}_{n\geq 0}$*, we have*

$$\min\left[0, N_T^n\langle Z^n\rangle_T - \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2\right] \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} 0. \tag{2.2.8}$$

For $\rho_N > 1$*, the sequence defined in* (2.2.6) *is admissible and optimal:*

$$N_T^n\langle Z^n\rangle_T \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2. \tag{2.2.9}$$

Proof. Let ζ_n be equal to the left hand side of (2.2.8). To prove its convergence in probability to 0, it is enough (Lemma 2.2.2) to consider a.s. convergence along iterated subsequences. Let $\{\iota(n)\}_{n\geq 0}$ be a given subsequence, consider an extraction for which $\sum_{n\geq 0}\varepsilon^2_{\iota\circ\iota'(n)} < \infty$ and that makes $\{\mathcal{T}^{\iota\circ\iota'(n)}\}_{n\geq 0}$ admissible in the sense of Definition 2.2.1. We can then apply (2.2.4) to obtain $\zeta_n \underset{n\to+\infty}{\overset{a.s.}{\to}} 0$. Thus (2.2.8) is proved. The justification of (2.2.9) follows the same arguments. □

2.2.3 Assumptions and main results for adaptive optimal discretization

Let $\{\sigma_t^n : 0 \leq t \leq T\}_{n\geq 0}$ denote a sequence of cáglád adapted processes, valued in invertible $d \times d$-matrices. We will study the sequence where the n-th strategy is built in the same way as in (2.2.6) but using the approximation process σ_t^n in (2.2.5).

Let us introduce the following definition, that will be constantly used in the sequel:

$$C_t^n := (\sigma_t^n)^\mathsf{T}(D_x v_t)^\mathsf{T}\sigma_t^n; \tag{2.2.10}$$

X_t^n is the solution of (2.2.1) with $c = C_t^n$;

$$\Lambda_t^n := ((\sigma_t^n)^{-1})^\mathsf{T} X_t^n (\sigma_t^n)^{-1}. \tag{2.2.11}$$

In this subsection we prove the optimality of the sequence of discretization strategies based on a general sequence of estimators for σ^n provided that it converges[1] sufficiently fast to $(\sigma\sigma^*)^{1/2}$. These estimators may depend on the path of S. Our working assumptions are the following.

[1] Note that this would not be much realistic to assume convergence to σ since at the continuous time-limit, we can retrieve $\sigma\sigma^*$ only.

Assumptions 2.

- $(\varepsilon_n)_{n\geq 0}$ *is a positive sequence with* $\varepsilon_n \to 0$*;*
- ρ_N *is a real number such that* $\rho_N \in (1, (\theta_\sigma/2+1) \wedge (4/3) \wedge (\alpha+1/2))$*;*
- δ *is a real number such that* $\delta \in (0, 2(\rho_N - 1)]$*.*

Assumptions 3.

- $\sup_{n\geq 0}\left(\sup_{t\in[0,T]} |\sigma_t^n|\right)$ *is a.s. finite;*
- σ_t^n *is non-degenerate for all* $t \in [0,T]$ *and* $n \geq 0$ *a.s.*

Assumptions 4.

- *For any subsequence* $\{\iota(n)\}_{n\geq 0}$*, there exists a further subsequence* $\{\iota \circ \iota'(n)\}_{n\geq 0}$ *such that* $\sup_{n\geq 0}\left(\sup_{t\in[0,T]} |\sigma_t^{\iota\circ\iota'(n)}|\right)$ *is a.s. finite;*
- σ_t^n *is non-degenerate for all* $t \in [0,T]$ *and* $n \geq 0$ *a.s.*

Let $\{\mathcal{T}^n\}_{n\geq 0}$, where $\mathcal{T}^n = \{\tau_i^n\}_{i\geq 0}$, be a sequence of discretization strategies defined as follows: for any $n \in \mathbb{N}$

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} [\Lambda_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta} (S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^{2+\delta}\} \wedge T. \end{cases} \tag{2.2.12}$$

The sequence $(\varepsilon_n^\delta)_{n\geq 0}$ serves to measure the minimal convergence rate of σ^n to $(\sigma\sigma^\mathsf{T})^{1/2}$ in order to maintain global optimality results. Since δ can be taken arbitrary close to 0, the requirement on the estimation rate is quite mild. Now we state one of the main results of this work.

Theorem 2.2.5. *Let Assumptions 1 and 2 be verified. Suppose that the sequence* $\{\sigma_t^n : 0 \leq t \leq T\}_{n\geq 0}$ *satisfies Assumptions 4. Assume also that*

$$\varepsilon_n^{-\delta/2} \sup_{t\in[0,T]} |(\sigma_t \sigma_t^\mathsf{T})^{1/2} - \sigma_t^n| \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{2.2.13}$$

Then the sequence of strategies (2.2.12) *is* $\mathbb{P}$*-admissible and optimal in the sense that*

$$N_T^n \langle Z^n \rangle_T \xrightarrow[n\to+\infty]{\mathbb{P}} \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2.$$

In the case where $\sum_{n\geq 0} \varepsilon_n^2 < +\infty$ we can state an a.s. version of Theorem 2.2.5. In fact the rest of this section will be devoted to the proof of this next result with a.s. convergence. Then as for Theorem 2.2.4, we will deduce the convergence in probability for a general sequence ε_n such that $\varepsilon_n \to 0$.

Theorem 2.2.5(bis). *Let Assumptions 1 and 2 be verified, under the stronger condition that* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Suppose that the sequence* $\{\sigma_t^n : 0 \leq t \leq T\}_{n\geq 0}$ *satisfies Assumptions 3 and that*

$$\varepsilon_n^{-\delta/2} \sup_{t\in[0,T]} |(\sigma_t\sigma_t^\mathsf{T})^{1/2} - \sigma_t^n| \xrightarrow[n\to+\infty]{a.s.} 0. \tag{2.2.14}$$

Then the sequence of strategies (2.2.12) *is admissible and optimal in the sense that*

$$N_T^n\langle Z^n\rangle_T \xrightarrow[n\to+\infty]{a.s.} \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2.$$

The proofs are given in Section 2.3.

2.3 Proofs of optimality of adaptive discretization strategy

This section is dedicated to the proofs of Theorems 2.2.5 and 2.2.5(bis).

Outline. All the preliminary results are preparatory for the proof of Theorem 2.2.5(bis) regarding almost sure results; later (Section 2.3.3), the proof of Theorem 2.2.5 (convergence in probability) is deduced owing to the subsequence principle. We divide the proof in two stages.

First in Section 2.3.1 we aim at finding a sufficient speed of convergence for the approximated processes Λ^n (see (2.2.11)) that characterize the ellipsoids in (2.2.12), in order to ensure the optimality of the resulting sequence of discretization strategies. This involves a careful analysis of the error terms about the deviation of the renormalized discretization errors from the optimal sequence.

Second in Section 2.3.2 we give the complementary part of the arguments to complete the theorem proof: we establish the Lipschitz property w.r.t. $(\sigma\sigma^*)^{1/2}$ of the ellipsoid generating process Λ. Surprisingly this property holds under the general enough Assumptions 1-2-3 despite the non-linear nature of the dependence given from the solution of the matrix equation (2.2.1). This implies in particular that the sufficient speed of approximation is the same for diffusion coefficient estimators and for the ellipsoid generating processes.

For this section, we need some specific notation:

- for a given sequence of stopping times and a given function $(U_t)_{0\leq t\leq T}$, we set $\Delta U_t = U_t - U_{\tau_{i-1}^n}$ for $t \in (\tau_{i-1}^n, \tau_i^n]$.
- $o_n(1)$ stands for a random variable converging to 0 a.s as $n \to +\infty$.

2.3.1 Deviation of the discretization error from the optimum

In this subsection we prove a preliminary result on the optimality for an adaptive sequence of discretization strategies.

The next lemma states the admissibility of a sequence based on hitting times of random ellipsoids. It is a generalization of [GL14a, Proposition 2.4], its proof is given in Appendix 2.A.2.

Lemma 2.3.1. *Let Assumptions 1-2 be in force and assume that* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Let* $(H_t^n)_{n\geq 0}$ *be a sequence of adapted cágládmatrix processes defined on* $[0,T]$ *and valued in* $\mathcal{S}_+^d$. *Suppose that*

$$\sup_{t\in[0,T]}|H_t^n - H_t| \xrightarrow[n\to+\infty]{a.s.} 0, \tag{2.3.1}$$

where $(H_t)_{t\in[0,T]}$ *is a continuous adapted matrix process.*

Then the sequence of discretization strategies $\{\mathcal{T}^n\}_{n\geq 0}$, *where* $\mathcal{T}^n = \{\tau_i^n\}_i$ *is defined for any* $n\in\mathbb{N}$ *by*

$$\begin{cases}\tau_0^n := 0,\\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T}[H^n_{\tau_{i-1}^n}]^{\varepsilon_n^\delta}(S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^{2+\delta}\}\wedge T,\end{cases} \tag{2.3.2}$$

is admissible (in the sense of Definition 2.2.1).

The following lemma states uniform positivity for some sequences of processes that will be important in the subsequent proofs. For its proof, see Appendix 2.A.3.

Lemma 2.3.2. *Let Assumptions 1-2-3 be in force and assume that*

$$\sup_{t\in[0,T]}|\sigma_t^n - (\sigma_t\sigma_t^\mathsf{T})^{1/2}| \xrightarrow[n\to+\infty]{a.s.} 0. \tag{2.3.3}$$

We have a.s.

$$\inf_{n\geq 0}\left(\inf_{t\in[0,T]}\lambda_{\min}(\sigma_t^n)\right) > 0,$$

$$\inf_{t\in[0,T]}\mathrm{Tr}(X_t) > 0, \qquad \inf_{n\geq 0}\left(\inf_{t\in[0,T]}\mathrm{Tr}(X_t^n)\right) > 0.$$

We are now in a position to state and prove the main result of this subsection, which is instrumental for the subsequent analysis.

Proposition 2.3.3. *Let Assumptions 1-2-3 be in force and assume that* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Consider the discretization strategy (2.2.12): it defines an admissible sequence.*
Set $R_t^n := [\Lambda_t^n]^{\varepsilon_n^\delta} - \Lambda_t$ *and assume that*

$$\varepsilon_n^{-\delta/2}\sup_{t\in[0,T]}|R_t^n| \xrightarrow[n\to+\infty]{a.s.} 0. \tag{2.3.4}$$

Then we have

$$\left|N_T^n\langle Z^n\rangle_T - \left(\sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^\mathsf{T}\Lambda_{\tau_{i-1}^n}\Delta S_{\tau_i^n}\right)^2\right| \xrightarrow[n\to+\infty]{a.s.} 0. \tag{2.3.5}$$

Proof. Assumption (2.3.4) readily implies that

$$\sup_{t\in[0,T]}|\Lambda_t^n - \Lambda_t| \xrightarrow[n\to+\infty]{a.s.} 0.$$

Remark that the process X and thus Λ are a.s. continuous due to the continuity of the solution to the matrix equation (2.2.1). Thus by Lemma 2.3.1, the sequence of strategies defined in (2.2.12) is admissible.

Now we decompose $N_T^n\langle Z^n\rangle_T$ as

$$N_T^n\langle Z^n\rangle_T = (\varepsilon_n^{2+\delta}N_T^n)(\varepsilon_n^{-(2+\delta)}\langle Z^n\rangle_T),$$

and consider the two multipliers separately. First using (2.2.12) we write

$$\varepsilon_n^{2+\delta}N_T^n = \varepsilon_n^{2+\delta} + \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n} \tag{2.3.6}$$

$$= \varepsilon_n^{2+\delta} + \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau_{i-1}^n}\Delta S_{\tau_i^n} + \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}R_{\tau_{i-1}^n}^n\Delta S_{\tau_i^n}. \tag{2.3.7}$$

Since from the continuity of Λ and (2.3.4) we have

$$\sup_{t\in[0,T]}|\Lambda_t| < +\infty, \qquad \sup_{n\geq 0}\sup_{t\in[0,T]}|R_t^n| < +\infty, \tag{2.3.8}$$

by applying Proposition 2.A.1 we easily justify that the three terms on the right hand side of (2.3.7) are uniformly bounded in n, almost surely. We have proved that $\varepsilon_n^{2+\delta}N_T^n$ is an a.s. bounded sequence.

Second, from the equations (1.4.6)-(1.4.7), we can decompose $\langle Z^n\rangle_T$ as follows:

$$\langle Z^n\rangle_T = \sum_{\tau_{i-1}^n<T}(\Delta B_{\tau_i^n}^{\mathsf{T}}X_{\tau_{i-1}^n}\Delta B_{\tau_i^n})^2 + e_{0,T}^n + e_{1,T}^n + e_{2,T}^n + e_{3,T}^n, \tag{2.3.9}$$

$$\varepsilon_n^{-2\rho_N}\sup_{0\leq i\leq 3}|e_{i,T}^n| \overset{a.s.}{\underset{n\to+\infty}{\to}} 0. \tag{2.3.10}$$

Next, from [GL14a, Corollary 2.3(ii)] or Lemma 1.3.3, we get that for any $\rho > 0$ there exists an a.s. finite random variable C_0 such that

$$\sup_{t\in[0,T]}|\Delta B_t| \leq C_0\varepsilon_n^{1-\rho/4}. \tag{2.3.11}$$

This readily follows that

$$\sum_{\tau_{i-1}^n<T}(\Delta B_{\tau_i^n}^{\mathsf{T}}X_{\tau_{i-1}^n}\Delta B_{\tau_i^n})^2 \leq C_0N_T^n\varepsilon_n^{4-\rho}\sup_{t\in[0,T]}|X_t|^2.$$

The above upper bound goes to 0 a.s., by taking ρ small enough because of 2.2.1 and $\rho_N < \frac{4}{3}$. In view of (2.3.9)-(2.3.10) and the above, we have proved that

$$\langle Z^n\rangle_T \overset{a.s.}{\underset{n\to+\infty}{\to}} 0. \tag{2.3.12}$$

Furthermore, from (2.3.9), we can write

$$\langle Z^n\rangle_T = \sum_{\tau_i^n<T} (\Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n})^2 \tag{2.3.13}$$

$$+ \sum_{\tau_i^n<T} \left((\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2 - (\Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n})^2\right) \tag{2.3.14}$$

$$+ \sum_{\tau_i^n<T} \left((\Delta B_{\tau_i^n}^{\mathsf{T}}X_{\tau^n_{i-1}}\Delta B_{\tau_i^n})^2 - (\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2\right) \tag{2.3.15}$$

$$+ (\Delta B_T^{\mathsf{T}}X_{\tau^n_{N_T^n-1}}\Delta B_T)^2 + e^n_{0,T} + e^n_{1,T} + e^n_{2,T} + e^n_{3,T}. \tag{2.3.16}$$

We aim at analyzing $\varepsilon_n^{-(2+\delta)}\langle Z^n\rangle_T$. From (2.3.11) we deduce again that, for $\rho > 0$ small enough (and since $\delta < 2/3$),

$$\varepsilon_n^{-(2+\delta)}(\Delta B_T^{\mathsf{T}}X_{\tau^n_{N_T^n-1}}\Delta B_T)^2 \le C_0\varepsilon_n^{2-\delta-\rho} \underset{n\to+\infty}{\overset{a.s.}{\to}} 0.$$

The convergences (2.3.10) combined with $2+\delta \le 2\rho_N$ ensure

$$\varepsilon_n^{-(2+\delta)}(e^n_{0,T} + e^n_{1,T} + e_{2,T} + e^n_{3,T}) \underset{n\to+\infty}{\overset{a.s.}{\to}} 0. \tag{2.3.17}$$

Regarding the third sum on the right hand side of (2.3.16), we obtain

$$\varepsilon_n^{-(2+\delta)} \sum_{\tau_i^n<T} \left((\Delta B_{\tau_i^n}^{\mathsf{T}}X_{\tau^n_{i-1}}\Delta B_{\tau_i^n})^2 - (\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2\right) \underset{n\to+\infty}{\overset{a.s.}{\to}} 0$$

by the same argument as for $e^n_{4,T}$ in the proof of Theorem 1.5.2 (it only uses the admissibility of the sequence of strategies). We now handle the second sum on the right hand side of (2.3.16):

$$\begin{aligned}
&\varepsilon_n^{-(2+\delta)} \sum_{\tau_i^n<T} \left((\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2 - (\Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n})^2\right)\\
&= \varepsilon_n^{-(2+\delta)} \sum_{\tau_i^n<T} \left(\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n} - \Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n}\right)\\
&\qquad\qquad \times \left(\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n} + \Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n}\right)\\
&= -\varepsilon_n^{-(2+\delta)} \sum_{\tau_i^n<T} \Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n} \left(2\Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n} - \Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n}\right)\\
&= -\sum_{\tau_i^n<T} \Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n} \left(2 - \varepsilon_n^{-(2+\delta)}\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n}\right)\\
&= -2\sum_{\tau_i^n<T} \Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n} + \varepsilon_n^{-(2+\delta)} \sum_{\tau_i^n<T} (\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2.
\end{aligned}$$

Consequently, we get that

$$\begin{aligned}\varepsilon_n^{-(2+\delta)}\langle Z^n\rangle_T &= \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n} - 2\sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n}\\ &\quad+\varepsilon_n^{-(2+\delta)}\sum_{\tau_i^n<T}(\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2+o_n(1)\\ &= \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n} - \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n}\\ &\quad+\varepsilon_n^{-(2+\delta)}\sum_{\tau_i^n<T}(\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2+o_n(1).\end{aligned}$$

Therefore, combining the above with (2.3.7) and (2.3.12), it comes

$$\begin{aligned}N_T^n\langle Z^n\rangle_T &= \left(\sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n}\right)^2-\left(\sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n}\right)^2\\ &+\left(\sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}[\Lambda^n_{\tau^n_{i-1}}]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n}\right)\varepsilon_n^{-(2+\delta)}\sum_{\tau_i^n<T}(\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2+o_n(1).\end{aligned}$$

Thus, since the admissibility of $\{\mathcal{T}^n\}_{n\geq 0}$ implies

$$\varepsilon_n^{-2}\sum_{\tau_i^n<T}(\Delta S_{\tau_i^n}^{\mathsf{T}}R^n_{\tau^n_{i-1}}\Delta S_{\tau_i^n})^2\leq C_0\sup_{0\leq t\leq T}|R_t^n|^2\left(\sum_{\tau_i^n<T}|\Delta S_{\tau_i^n}|^2\right),$$

we finally derive

$$\begin{aligned}&\left|N_T^n\langle Z^n\rangle_T-\left(\sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}\Lambda_{\tau^n_{i-1}}\Delta S_{\tau_i^n}\right)^2\right|\\ &\leq\sup_{0\leq t\leq T}|R_t^n|^2\left(\sum_{\tau_i^n<T}|\Delta S_{\tau_i^n}|^2\right)^2\left(1+C_0\varepsilon_n^{-\delta}\sup_{n\geq 0}\sup_{t\in[0,T]}|[\Lambda_t^n]^{\varepsilon_n^\delta}|\right)+o_n(1).\end{aligned}$$

From (2.3.8) we get that

$$\sup_{n\geq 0}\sup_{t\in[0,T]}|[\Lambda_t^n]^{\varepsilon_n^\delta}|=\sup_{n\geq 0}\sup_{t\in[0,T]}|R_t^n+\Lambda_t|<+\infty.$$

In addition, Proposition 2.A.1 ensures the a.s.-convergence of $\sum_{\tau_i^n<T}|\Delta S_{\tau_i^n}|^2$. It remains to use the convergence (2.3.4) to complete the proof of (2.3.5). We are done. □

2.3.2 Verification of Assumption (2.3.4) of Proposition 2.3.3: bound for $|R_t^n|$

In the previous subsection we have estimated the deviation of the discretization error from the optimal one in terms of $|R_t^n|$. Our goal is to further express this error in terms of the estimation error for σ_t^n and specify the conditions on the estimators σ_t^n to get an optimal sequence of discretization strategies.

The purpose of this subsection is to prove the following result.

Proposition 2.3.4. *Let Assumptions 1-2-3 be in force and assume that* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Consider the discretization strategy* (2.2.12) *and assume*

$$\varepsilon_n^{-\delta/2}\sup_{t\in[0,T]}|\sigma_t^n-(\sigma_t\sigma_t^\mathsf{T})^{1/2}|\overset{a.s.}{\underset{n\to+\infty}{\to}}0. \tag{2.3.18}$$

Then we have

$$|\Lambda_t^n-\Lambda_t|\leq C_0|(\sigma_t\sigma_t^\mathsf{T})^{1/2}-\sigma_t^n|, \tag{2.3.19}$$

$$\varepsilon_n^{-\delta/2}\sup_{t\in[0,T]}|R_t^n|\overset{a.s.}{\underset{n\to+\infty}{\to}}0. \tag{2.3.20}$$

Proof. Since $R_t^n=[\Lambda_t^n]^{\varepsilon_n^\delta}-\Lambda_t$, we get $|R_t^n|\leq\varepsilon_n^\delta+|\Lambda_t^n-\Lambda_t|$, which shows that it is enough to prove (2.3.19) to derive (2.3.20).

To prove (2.3.19), we argue that this enough to consider the case where σ_t is symmetric. Indeed, observe that the matrix solution Λ_t computed with σ_t or $(\sigma_t\sigma_t^\mathsf{T})^{1/2}$ is the same, so that the bound (2.3.19) depends intrinsically on σ_t through $(\sigma_t\sigma_t^\mathsf{T})^{1/2}$. Therefore, assume hereafter that σ_t is symmetric, so that $\sigma_t=(\sigma_t\sigma_t^\mathsf{T})^{1/2}$. The proof of (2.3.19) requires a few linear algebra results that we recall below. For this proof, we use three convenient notations.

- For a given $A\in\mathcal{S}^d$, let $\alpha(A)$ denote the vector of the eigenvalues of A placed in decreasing order.
- For a square matrix A, $|A|_F=\sqrt{\mathrm{Tr}(A^\mathsf{T}A)}$ stands for the Frobenius norm of A (also known as the Hilbert-Schmidt norm).
- Let $\{(P_t^n)_{0\leq t\leq T}:n\geq 0\}$ be a sequence of non-negative measurable processes, we write

$$P_t^n\preceq|\sigma_t-\sigma_t^n|$$

if $P_t^n\leq C_0|\sigma_t-\sigma_t^n|,\forall t\in[0,T],\forall n\geq 0,\quad a.s.$.

Lemma 2.3.5 ([Wih09, Theorem 1.1])**.** *Let $f:I\to\mathbb{R}$ be a real function on a interval $I\subset\mathbb{R}$ which is Lipschitz with a constant C_I. Let $A,B\in\mathcal{S}^d$ such that their eigenvalues are in I. For a matrix $M\in\mathcal{S}^d$ with spectral decomposition $M=U^\mathsf{T}\,\mathrm{Diag}(\lambda_1,\dots,\lambda_d)U$ we denote*

$$f(M)=U^\mathsf{T}\,\mathrm{Diag}(f(\lambda_1),\dots,f(\lambda_d))U.$$

Then

$$|f(A)-f(B)|_F\leq C_I|A-B|_F.$$

Lemma 2.3.6 (Hoffman-Wielandt inequality, [AGZ09, Lemma 2.1.19]). *Let $A, B \in \mathcal{S}^d$. Then*

$$|\alpha(A) - \alpha(B)| \leq |A - B|_F.$$

We now go back to the proof of (2.3.19) (still in the symmetric case for σ_t): We extensively use Assumptions 3, the statements of Lemma 2.3.2 and we proceed in several steps.

1: $|C_t C_t^\mathsf{T} - C_t^n (C_t^n)^\mathsf{T}| \preceq |\sigma_t - \sigma_t^n|$.

From the definitions of C_t and C_t^n using a standard calculation we get

$$|C_t C_t^\mathsf{T} - C_t^n (C_t^n)^\mathsf{T}| = |\sigma_t^\mathsf{T} (D_x v_t)^\mathsf{T} \sigma_t \sigma_t^\mathsf{T} D_x v_t \sigma_t - (\sigma_t^n)^\mathsf{T} (D_x v_t)^\mathsf{T} \sigma_t^n (\sigma_t^n)^\mathsf{T} D_x v_t \sigma_t^n|$$
$$\leq \sum_{i=0}^{3} |D_x v_t|^2 |\sigma_t|^i |\sigma_t^n|^{3-i} |\sigma_t - \sigma_t^n|.$$

We are done.

2: $|\alpha(C_t C_t^\mathsf{T}) - \alpha(C_t^n (C_t^n)^\mathsf{T})| \preceq |\sigma_t - \sigma_t^n|$.

This directly follows from Lemma 2.3.6 and Step 1.

3: $|\mathrm{Tr}(X_t^n - X_t)| \preceq |\sigma_t - \sigma_t^n|$.

Here we need to recall part of the proof of [GL14a, Lemma 3.1], which gives the solution to the matrix equation (2.2.1). Consider the function $f : \mathbb{R}_d^+ \times \mathbb{R}^+ \to \mathbb{R}$, defined by

$$f(\beta, y) = (4 + d)y - \sum_{i=1}^{d} \sqrt{y^2 + 4\beta_i}.$$

For any $\beta \in \mathbb{R}_d^+$, there is a unique $y_\beta^f \geq 0$ solution to $f(\beta, y) = 0$. In addition $\mathrm{Tr}(X_t) = y^f_{\alpha(C_t C_t^\mathsf{T})} > 0$ and $\mathrm{Tr}(X_t^n) = y^f_{\alpha(C_t^n (C_t^n)^\mathsf{T})} > 0$. The first fundamental theorem of calculus yields, for any β, β',

$$|y_\beta^f - y_{\beta'}^f| \leq \frac{\sup_{\lambda \in [0,1]} \left| D_\beta f(\lambda\beta + (1-\lambda)\beta', y^f_{\lambda\beta + (1-\lambda)\beta'}) \right|}{\inf_{\lambda \in [0,1]} \left| D_y f(\lambda\beta + (1-\lambda)\beta', y^f_{\lambda\beta + (1-\lambda)\beta'}) \right|} |\beta - \beta'|.$$

Remark that for any $y > 0$ and any β

$$D_y f(\beta, y) = 4 + d - \sum_{i=1}^{d} \frac{y}{\sqrt{y^2 + 4\beta_i}} \geq 4,$$
$$|D_{\beta_j} f(\beta, y)| = \frac{2}{\sqrt{y^2 + 4\beta_j}} \leq \frac{2}{y}.$$

Therefore, $|y^f_\beta - y^f_{\beta'}| \le \dfrac{\sqrt{d}}{2\min(y^f_\beta, y^f_{\beta'})}\,|\beta-\beta'|$. It readily follows that

$$\begin{aligned}|\mathrm{Tr}(X_t - X^n_t)| &= \left|y^f_{\alpha(C_tC_t^\mathsf{T})} - y^f_{\alpha(C^n_t(C^n_t)^\mathsf{T})}\right| \\ &\le \frac{\sqrt{d}}{2\min(\mathrm{Tr}(X_t),\mathrm{Tr}(X^n_t))}\left|\alpha(C_tC_t^\mathsf{T}) - \alpha(C^n_t(C^n_t)^\mathsf{T})\right| \preceq |\sigma_t - \sigma^n_t|\end{aligned}$$

using Lemma 2.3.2 and Step 2.

4: $|X^n_t - X_t| \preceq |\sigma_t - \sigma^n_t|$.
From [GL14a, Equation (A.7)], we have

$$X_t = -\frac{1}{4}\mathrm{Tr}(X_t)\,\mathrm{Id}_d + \frac{1}{2}\left(\frac{\mathrm{Tr}(X_t)^2}{4}\,\mathrm{Id}_d + C_tC_t^\mathsf{T}\right)^{1/2}$$

and similarly for X^n_t. We now apply Lemma 2.3.5 with $f(x)=\sqrt{x}$: with the notation of this lemma, the eigenvalues of $\frac{\mathrm{Tr}(X_t)^2}{4}\,\mathrm{Id}_d + C_tC_t^\mathsf{T}$ and $\frac{\mathrm{Tr}(X^n_t)^2}{4}\,\mathrm{Id}_d + C^n_t(C^n_t)^\mathsf{T}$ take values in $I := [\min(\frac{\mathrm{Tr}(X_t)^2}{4}, \frac{\mathrm{Tr}(X^n_t)^2}{4}), +\infty)$ and thus, the Lipschitz constant of Lemma 2.3.5 is $C_{n,t,I} := 1/\min(\mathrm{Tr}(X_t), \mathrm{Tr}(X^n_t))$. Furthermore, by Lemma 2.3.2

$$\inf_{n\ge 0}\left(\inf_{t\in[0,T]}\mathrm{Tr}(X^n_t)\right) > 0, \quad \inf_{t\in[0,T]}\mathrm{Tr}(X_t) > 0$$

which implies $\sup_{n\ge0}\sup_{t\in[0,T]} C_{n,t,I} < +\infty$. As a consequence, we obtain

$$\begin{aligned}|X^n_t - X_t| &\preceq |\,\mathrm{Tr}(X^n_t - X_t)| + \left|\mathrm{Tr}(X^n_t)^2 - \mathrm{Tr}(X_t)^2\right| + \left|C^n_t(C^n_t)^\mathsf{T} - C_tC_t^\mathsf{T}\right| \\ &\preceq |\sigma_t - \sigma^n_t|\end{aligned}$$

where we have used Assumptions 1 and 3, Steps 1 and 3.

5: $\left|\sigma_t^\mathsf{T}(\Lambda^n_t - \Lambda_t)\sigma_t\right| \preceq |\sigma_t - \sigma^n_t|$.
We write

$$\begin{aligned}\sigma_t^\mathsf{T}(\Lambda^n_t - \Lambda_t)\sigma_t &= \sigma_t^\mathsf{T}((\sigma^n_t)^\mathsf{T})^{-1}X^n_t(\sigma^n_t)^{-1}\sigma_t - X_t \\ &= (\sigma_t^\mathsf{T}((\sigma^n_t)^\mathsf{T})^{-1}X_t(\sigma^n_t)^{-1}\sigma_t - X_t) + \ \sigma_t^\mathsf{T}((\sigma^n_t)^\mathsf{T})^{-1}(X^n_t - X_t)(\sigma^n_t)^{-1}\sigma_t.\end{aligned}$$

For the first term on the right hand side, we get

$$\begin{aligned}&\left|\sigma_t^\mathsf{T}((\sigma^n_t)^\mathsf{T})^{-1}X_t(\sigma^n_t)^{-1}\sigma_t - X_t\right| \\ &= \left|(\mathrm{Id}_d + (\sigma_t - \sigma^n_t)^\mathsf{T}((\sigma^n_t)^\mathsf{T})^{-1})X_t(\mathrm{Id}_d + (\sigma^n_t)^{-1}(\sigma_t - \sigma^n_t)) - X_t\right| \\ &= \Big|(\sigma_t - \sigma^n_t)^\mathsf{T}((\sigma^n_t)^\mathsf{T})^{-1}X_t + X_t(\sigma^n_t)^{-1}(\sigma_t - \sigma^n_t) \\ &\qquad + (\sigma_t - \sigma^n_t)^\mathsf{T}((\sigma^n_t)^\mathsf{T})^{-1}X_t(\sigma^n_t)^{-1}(\sigma_t - \sigma^n_t)\Big|\end{aligned}$$

$$\leq \left(2|X_t||(\sigma_t^n)^{-1}| + |\sigma_t - \sigma_t^n||(\sigma_t^n)^{-1}|^2|X_t|\right)|\sigma_t - \sigma_t^n|$$
$$\preceq |\sigma_t - \sigma_t^n|$$

using that by Assumptions 1, 3 and the continuity of X we have

$$\sup_{n\geq 0}\left(\sup_{t\in[0,T]}|\sigma_t^n|\right) < +\infty, \quad \sup_{t\in[0,T]}|\sigma_t| < +\infty, \quad \sup_{t\in[0,T]}|X_t| < +\infty$$

while Lemma 2.3.2 gives

$$\inf_{n\geq 0}\left(\inf_{t\in[0,T]}\lambda_{\min}(\sigma_t^n)\right) > 0.$$

Now for the second summand, it is clear from Step 4 that

$$\left|\sigma_t^\mathsf{T}((\sigma_t^n)^\mathsf{T})^{-1}(X_t^n - X_t)(\sigma_t^n)^{-1}\sigma_t\right| \preceq |\sigma_t - \sigma_t^n|.$$

We are done.

6: $|\Lambda_t^n - \Lambda_t| \preceq |\sigma_t - \sigma_t^n|$.

This directly follows from Step 5 and from the sub-multiplicativity of the matrix norm.

This finishes the proof of Proposition 2.3.4. □

2.3.3 Proof of Theorems 2.2.5(bis) and 2.2.5

Now we are ready to finish the proof of the announced theorems.

Proof of Theorem 2.2.5(bis)

We assume $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. First the sequence $\{\mathcal{T}^n\}_{n\geq 0}$ defined in (2.2.12) is admissible: this is the first part of Proposition 2.3.3.

The hypothesis (2.2.14) of Theorem 2.2.5(bis) implies that (2.3.20) of Proposition 2.3.4 holds. Therefore, we can apply the last part of Proposition 2.3.3 to write

$$\left|N_T^n\langle Z^n\rangle_T - \left(\sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^\mathsf{T}\Lambda_{\tau_{i-1}^n}\Delta S_{\tau_i^n}\right)^2\right| \underset{n\to+\infty}{\overset{a.s.}{\to}} 0.$$

The a.s.-limit of $(\cdots)^2$ is handled using Proposition 2.A.1 and we finally obtain

$$\lim_{n\to+\infty} N_T^n\langle Z^n\rangle_T = \left(\int_0^T \mathrm{Tr}(\Lambda_t\sigma_t\sigma_t^\mathsf{T})\mathrm{d}t\right)^2 = \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2 \qquad a.s.$$

This finishes the proof of Theorem 2.2.5(bis). □

Proof of Theorem 2.2.5

Here we assume only $\varepsilon_n \to 0$. Let $\{\mathcal{T}^n\}_{n\geq 0}$ be the sequence of strategies defined in (2.2.12). By the subsequence principle (Lemma 2.2.2), it is enough to show that for any subsequence $(\varepsilon_\iota(n))_{n\geq 0}$ of $(\varepsilon_n)_{n\geq 0}$, we can extract another subsequence $(\varepsilon_{\iota\circ\iota'(n)})_{n\geq 0}$ which is square summable and such that $\{\mathcal{T}^{\iota\circ\iota'(n)}\}_{n\geq 0}$ is admissible and

$$N_T^{\iota\circ\iota'(n)}\langle Z^{\iota\circ\iota'(n)}\rangle_T \underset{n\to+\infty}{\overset{a.s.}{\to}} \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2. \tag{2.3.21}$$

But these properties have been proved in Theorem 2.2.5(bis), which holds because Assumption 3 with the subsequence $(\iota\circ\iota'(n))_{n\geq 0}$ is satisfied (owing to Assumption 4). We are done. □

2.4 Rolling diffusion coefficient estimation

In this section we study the properties of the estimators $(\sigma_t^n)_{t\in[0,T]}$ based on the rolling covariance estimation of the increments of S with a general weighting kernel. We will find out the asymptotic properties of the lookback window size and the estimation grid size that will ensure the optimality of the resulting sequence of strategies (in the sense of verification of $\varepsilon_n^{-\delta/2}\sup_{t\in[0,T]}|(\sigma_t\sigma_t^\intercal)^{1/2} - \sigma_t^n| \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} 0$ as in Theorem 2.2.5).

The subsequent approach will make use of data observed at negative times on $[t_0, 0]$ (for some $t_0 < 0$), just to be able to provide accurate estimation of $(\sigma_t\sigma_t^\intercal)^{1/2}$ for t close to 0. Alternatively, one could assume directly that $(\sigma_0\sigma_0^\intercal)^{1/2}$ is known (exactly or with some error) and then combine this a priori knowledge with observation on $[0, T]$. However, we believe that this would be quite artificial and that a full data-driven algorithm is preferable and more realistic for practical applications. Therefore for this section we suppose that the process S is defined on $[t_0, T]$ (with some $t_0 < 0$) and we start the discretization algorithm at time 0 (as in Sections 2.1-2.2-2.3).

Now we define the kernel used for the estimation. Let $K : \mathbb{R} \mapsto \mathbb{R}^+$ be a non-negative bounded function satisfying

$$\int_{-\infty}^0 K(u)\mathrm{d}u = 1.$$

We assume that the kernel K verifies the following hypotheses:

$(\mathbf{H}_K^{supp})$: K has a compact support included in $[-\kappa, 0]$ for some $\kappa > 0$.

$(\mathbf{H}_K^{Lip})$: K is Lipschitz continuous on the interval $[-\kappa, 0]$ with a Lipschitz constant L_K.

Note that K may be discontinuous at $-\kappa$ and 0. For any $\gamma > 0$ denote

$$K_\gamma(t) = \frac{1}{\gamma}K\left(\frac{t}{\gamma}\right).$$

As mentioned before, the process S is defined on $[t_0, T]$ for some $t_0 < 0$. We implicitly suppose that Assumptions 1 and 4 are modified for the case where $t \in [t_0, T]$. However the

optimality results for discretization strategies are still considered on $[0,T]$. We are aware that doing so, we should redefine the probabilistic model on $[t_0,T]$ as a difference with previous sections. However, the reader can easily check that it would not modify the results but would complicate the presentation (which we prefer to avoid for the sake of clarity and conciseness).

Let Assumptions 1 hold. Fix some $(\varepsilon_n)_{n\geq 0}, \rho_N$ and δ satisfying Assumptions 2. Further we will use a particular sequence of estimators σ_t^n that as will be shown also verifies Assumptions 4.

We choose a sequence $\{\mathcal{T}^n\}_{n\geq 0}$ of estimation grids that is $\mathbb{P}$-admissible for some $(\varepsilon'_n)_{n\geq 0}$ with $\varepsilon'_n \to 0$ and $\rho'_N \geq 1$. The sequence of estimation grids may differ from the sequence of the grids representing the optimal discretization strategy. The latter one is admissible for $(\varepsilon_n)_{n\geq 0}$ and ρ_N. Remark that in this section τ_i^n will denote the stopping times of the estimations grid $\mathcal{T}^n = \{t_0 = \tau_0^n < \cdots < \tau_i^n < \cdots < \tau_{N_T^n}^n = T\}$.

Choose a positive sequence $(\gamma_n)_{n\geq 0}$ such that $\gamma_n \to 0$; we can assume without loss of generality that $t_0 < -\kappa\gamma_n$ for any $n \geq 0$. Define the rolling empirical covariance matrix, where the kernel K represents a weight function, as follows

$$\Sigma_t^n = \sum_{\tau_i^n < t} K_{\gamma_n}(\tau_{i-1}^n - t)\Delta S_{\tau_i^n} \Delta S_{\tau_i^n}^{\mathsf{T}}, \tag{2.4.1}$$

where, for any process U, we set $\Delta U_t := U_t - U_{\tau_{i-1}^n}$ for $t \in (\tau_{i-1}^n, \tau_i^n]$.

Note that for each $n \geq 0$ the process Σ^n is adapted (as τ_i^n's are adapted). Moreover it is continuous between the pairs of stopping times τ_{i-1}^n and τ_i^n and takes the left limit at τ_i^n's. Thus the process Σ_t^n is adapted cáglád for each n.

The main result of this section is the following theorem.

Theorem 2.4.1. *Let $(\alpha_n)_{n\geq 0}$ be a sequence of real positive numbers such that*

$$\varepsilon_n^{-\delta/2}\alpha_n \underset{n\to+\infty}{\to} 0.$$

Define the sequence of estimators $\{\sigma_t^n : 0 \leq t \leq T\}_{n\geq 0}$ by setting

$$\sigma_t^n = (\Sigma_t^n + \alpha_n \operatorname{Id}_d)^{1/2},$$

where Σ_t^n is the symmetric matrix given by (2.4.1). Suppose that $\varepsilon'_n, \varepsilon_n$ and γ_n satisfy the assumptions

(i) $\varepsilon'_n \to 0$, $\varepsilon_n \to 0, \gamma_n \to 0$;

(ii) $\varepsilon_n^{-\delta/2}\gamma_n^{\theta_\sigma/2} \to 0$, *as* $n \to \infty$;

(iii) there exists $\rho \in (0,2)$ *such that* $\dfrac{(\varepsilon'_n)^{1-\rho/2}}{\varepsilon_n^{\delta/2}\gamma_n} \to 0$, *as* $n \to \infty$.

Then the estimators σ^n fulfill Assumptions 4 and the sequence of discretization strategies (2.2.12) based on the estimators σ^n is optimal in the sense of Theorem 2.2.5.

Remark 2.4.2. *Clearly the set of hypotheses in Theorem 2.4.1 is non-contradictory since we can first choose γ_n that converges to 0 fast enough to satisfy (ii) and then choose ε'_n converging to 0 fast enough to satisfy (iii).*

The rest of the section is devoted to the proof. The above result states optimality for convergence in probability, optimality in almost sure sense would be similar by requiring square summability of the sequences ε_n and $\frac{\varepsilon'_n}{\varepsilon_n^{\delta/2}\gamma_n}$, the detailed analysis is left to the reader.

We now prove Theorem 2.4.1. The following lemma is a technical property of the kernel K with the stochastic grid $\mathcal{T}^n$.

Lemma 2.4.3. *Consider a sequence of grids, admissible (in the a.s. sense of Definition 2.2.1) for ε'_n and ρ'_N, with $\sum_{n\geq 0}(\varepsilon'_n)^2 < +\infty$. Denote $\phi(s)$ the latest stopping time of a estimation grid strictly before s, where we omit the dependence on n and on the grid. Then, for any $\rho > 0$, we have a.s. for any $n \geq 0$ and any $t \in [t_0, T]$*

$$\int_{-\infty}^{t} |K_{\gamma_n}(s-t) - K_{\gamma_n}(\phi(s)-t)|\mathrm{d}s \leq \frac{C_0(\varepsilon'_n)^{2-\rho}}{\gamma_n}$$

(note that the a.s. finite random variable C_0 may depend on ρ).

Proof. We recall the a.s. control of time step for the admissible estimation grid (see Lemma 1.3.2), i.e.

$$\sup_{n>0}(\varepsilon'_n)^{\rho-2} \sup_{1\leq i\leq N_T^n} \Delta\tau_i^n < +\infty \quad a.s. \tag{2.4.2}$$

So we obtain

$$\begin{aligned}
&\int_{-\infty}^{t} |K_{\gamma_n}(s-t) - K_{\gamma_n}(\phi(s)-t)|\mathrm{d}s \leq \int_{t-\kappa\gamma_n\leq\phi(s)<s\leq t} |K_{\gamma_n}(s-t) - K_{\gamma_n}(\phi(s)-t)|\mathrm{d}s \\
&+ \int_{\phi(s)<t-\kappa\gamma_n\leq s\leq t} |K_{\gamma_n}(s-t) - K_{\gamma_n}(\phi(s)-t)|\mathrm{d}s \\
&\leq \frac{L_K}{\gamma_n^2}\kappa\gamma_n \sup_{1\leq i\leq N_T^n} \Delta\tau_i^n + \frac{2}{\gamma_n} \sup_{u\in[-\kappa,0]} K(u) \sup_{1\leq i\leq N_T^n} \Delta\tau_i^n \leq \frac{C_0(\varepsilon'_n)^{2-\rho}}{\gamma_n},
\end{aligned}$$

which implies the result. □

Now we show a preliminary result on the convergence of the estimators Σ_t^n.

Proposition 2.4.4. *Suppose that $\varepsilon'_n, \varepsilon_n$ and γ_n satisfy the assumptions (i)-(ii)-(iii) of Theorem 2.4.1. Then we have that*

$$\varepsilon_n^{-\delta/2} \sup_{t\in[0,T]} |\Sigma_t^n - \Sigma_t| \xrightarrow[n\to+\infty]{\mathbb{P}} 0 \tag{2.4.3}$$

where $\Sigma_t = \sigma_t\sigma_t^\mathsf{T}$.

Proof. Note that (i) and (iii) imply $\dfrac{\varepsilon_n'}{\varepsilon_n^{\delta/2}\gamma_n} \to 0$. First suppose that

$$\sum_{n\geq 0}\left(\frac{\varepsilon_n'}{\varepsilon_n^{\delta/2}\gamma_n}\right)^2 < +\infty.$$

In particular, this implies that $\sum_{n\geq 0}(\varepsilon_n')^2 < +\infty$. For this case we will prove the convergence (2.4.3) in the a.s. sense and then, as in the proof of Theorem 2.2.5, use the subsequence principle from Lemma 2.2.2 to pass to the general case. Write

$$\varepsilon_n^{-\delta/2}|\Sigma_t^n - \Sigma_t| \leq \varepsilon_n^{-\delta/2}\left|\int_{-\infty}^t K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s - \Sigma_t\right| \tag{2.4.4}$$

$$+\,\varepsilon_n^{-\delta/2}\left|\int_{-\infty}^t K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s - \sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\Delta S_{\tau_i^n}\Delta S_{\tau_i^n}^{\mathsf{T}}\right|. \tag{2.4.5}$$

Let us first show the convergence

$$\varepsilon_n^{-\delta/2}\sup_{t\in[0,T]}\left|\int_{-\infty}^t K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s - \Sigma_t\right| \overset{a.s.}{\underset{n\to+\infty}{\to}} 0.$$

Using the Hölder property of σ_t (and by consequence of Σ_t), the assumption 2.4 and that $\int_{-\infty}^t K_{\gamma_n}(s-t)\mathrm{d}s = 1$, we have

$$\begin{aligned}\varepsilon_n^{-\delta/2}\left|\int_{-\infty}^t K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s - \Sigma_t\right| &\leq \varepsilon_n^{-\delta/2}\int_{-\infty}^t K_{\gamma_n}(s-t)|\Sigma_s - \Sigma_t|\mathrm{d}s\\ \leq C_0\varepsilon_n^{-\delta/2}\gamma_n^{\theta_\sigma/2}\int_{-\infty}^t K_{\gamma_n}(s-t)\mathrm{d}s &= C_0\varepsilon_n^{-\delta/2}\gamma_n^{\theta_\sigma/2}\overset{a.s.}{\underset{n\to+\infty}{\to}} 0\end{aligned}$$

uniformly on $[0,T]$, in view of the assumption (ii) of the proposition.

Now consider the second term of the decomposition (2.4.4). Had the grid $\mathcal{T}^n$ been deterministic (i.e., the usual case in the literature), the analysis would have been quite standard, using direct martingale arguments. Here the stochasticity of $\mathcal{T}^n$ and of the factors $K_{\gamma_n}(\tau_{i-1}^n - t)$ complicate significantly the analysis. Using the Itô formula we write

$$\varepsilon_n^{-\delta/2}\left(\int_{-\infty}^t K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s - \sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\Delta S_{\tau_i^n}\Delta S_{\tau_i^n}^{\mathsf{T}}\right) \tag{2.4.6}$$

$$=\varepsilon_n^{-\delta/2}\left(\int_{-\infty}^t K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s - \sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\int_{\tau_{i-1}^n}^{\tau_i^n}\Sigma_s \mathrm{d}s\right) \tag{2.4.7}$$

$$-\,2\varepsilon_n^{-\delta/2}\sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\int_{\tau_{i-1}^n}^{\tau_i^n}\Delta S_s \mathrm{d}M_s^{\mathsf{T}} \tag{2.4.8}$$

$$-\,2\varepsilon_n^{-\delta/2}\sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\int_{\tau_{i-1}^n}^{\tau_i^n}\Delta S_s \mathrm{d}A_s^{\mathsf{T}}. \tag{2.4.9}$$

a) For the first summand in the decomposition (2.4.7) we have

$$\begin{aligned}
&\varepsilon_n^{-\delta/2}\left|\int_{-\infty}^{t} K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s - \sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\int_{\tau_{i-1}^n}^{\tau_i^n}\Sigma_s \mathrm{d}s\right| \\
&\le \varepsilon_n^{-\delta/2}\int_{-\infty}^{\phi(t)} |K_{\gamma_n}(s-t) - K_{\gamma_n}(\phi(s)-t)|\cdot|\Sigma_s|\mathrm{d}s + \varepsilon_n^{-\delta/2}\left|\int_{\phi(t)}^{t} K_{\gamma_n}(s-t)\Sigma_s \mathrm{d}s\right| \\
&\le C_0 \sup_{t\in[t_0,T]}|\Sigma_t|\frac{(\varepsilon_n')^{2-\rho}}{\gamma_n\varepsilon_n^{\delta/2}} + C_0\sup_{t\in[t_0,T]}|\Sigma_t|\sup_u K(u)\frac{(\varepsilon_n')^{2-\rho}}{\gamma_n\varepsilon_n^{\delta/2}},
\end{aligned}$$

where for the last inequality we used Lemma 2.4.3 for ρ from (iii), and (2.4.2). Further

$$\frac{(\varepsilon_n')^{2-\rho}}{\gamma_n\varepsilon_n^{\delta/2}} = \varepsilon_n^{\delta/2}\gamma_n\frac{(\varepsilon_n')^{2-\rho}}{\gamma_n^2\varepsilon_n^{\delta}} \to 0$$

in view of (iii). Thus the first summand in (2.4.7) tends to 0, uniformly in $t\in[0,T]$.

b) Now let us handle the second term in (2.4.7). Define a martingale process on $[t_0,T]$, valued in $\mathbb{R}^d\otimes\mathbb{R}^d$, as follows:

$$P_t^n = \int_{t_0}^{t}\Delta S_s \mathrm{d}M_s^{\mathsf{T}}.$$

The Abel transformation yields

$$\begin{aligned}
\sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\int_{\tau_{i-1}^n}^{\tau_i^n}\Delta S_s \mathrm{d}M_s^{\mathsf{T}} &= \sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)(P_{\tau_i^n}^n - P_{\tau_{i-1}^n}^n) &(2.4.10)\\
&= \sum_{\tau_i^n<t} P_{\tau_i^n}^n(K_{\gamma_n}(\tau_{i-1}^n - t) - K_{\gamma_n}(\tau_i^n - t)) + P_{\phi(t)}^n K_{\gamma_n}(\phi(t)-t). &(2.4.11)
\end{aligned}$$

Consequently we obtain

$$\begin{aligned}
&\left|\varepsilon_n^{-\delta/2}\sum_{\tau_i^n<t} K_{\gamma_n}(\tau_{i-1}^n - t)\int_{\tau_{i-1}^n}^{\tau_i^n}\Delta S_s \mathrm{d}M_s^{\mathsf{T}}\right| \\
&\le \varepsilon_n^{-\delta/2}\sup_{t_0\le s\le T}|P_s^n|\left(\sum_{\tau_i^n<t}|K_{\gamma_n}(\tau_{i-1}^n - t) - K_{\gamma_n}(\tau_i^n - t)| + K_{\gamma_n}(\phi(t)-t)\right) \\
&\le \varepsilon_n^{-\delta/2}\sup_{t_0\le s\le T}|P_s^n|\left(\gamma_n^{-2}L_K\sum_{t-\kappa\gamma_n\le\tau_{i-1}^n<\tau_i^n<t}(\tau_i^n - \tau_{i-1}^n) + 2\gamma_n^{-1}\sup_u K(u)\right) \\
&\le \varepsilon_n^{-\delta/2}\sup_{t_0\le s\le T}|P_s^n|\left(\gamma_n^{-2}L_K\kappa\gamma_n + 2\gamma_n^{-1}\sup_u K(u)\right) \le C_0\frac{1}{\varepsilon_n^{\delta/2}\gamma_n}\sup_{t_0\le s\le T}|P_s^n|.
\end{aligned}$$

The quadratic variation of the (k,l)-element of the matrix-valued martingale $\dfrac{1}{\varepsilon_n^{\delta/2}\gamma_n}P^n$ is

equal to

$$\frac{1}{\varepsilon_n^{\delta}\gamma_n^2}\int_{t_0}^{T}|\Delta S_s^k|^2\mathrm{d}\langle M^l\rangle_s \le C_0\frac{(\varepsilon_n')^2}{\varepsilon_n^{\delta}\gamma_n^2}.$$

Hence $\sum_{n\ge 0}\langle \varepsilon_n^{-\delta/2}\gamma_n^{-1}P_{k,l}^n\rangle_T < +\infty$ a.s. using the temporary assumption $\sum_n\left(\frac{\varepsilon_n'}{\varepsilon_n^{\delta/2}\gamma_n}\right)^2 < +\infty$ made at the beginning of the proof. From this using Lemma 2.A.2 for a process defined on $[t_0,T]$ (see Remark 2.A.3) we deduce that

$$\frac{1}{\varepsilon_n^{\delta/2}\gamma_n}\sup_{t_0\le s\le T}|P_s^n| \xrightarrow[n\to+\infty]{a.s.} 0,$$

Thus the second term in the decomposition (2.4.7) a.s. converges to 0 uniformly in $t\in[0,T]$.
c) Denote $t\to|A^l|_t$ the total variation process of the l-th component A. For the (k,l)-element of the third (matrix) summand in (2.4.7) we get

$$\begin{aligned}\varepsilon_n^{-\delta/2}\left|\sum_{\tau_i^n<t}K_{\gamma_n}(\tau_{i-1}^n-t)\int_{\tau_{i-1}^n}^{\tau_i^n}\Delta S_s^k\mathrm{d}A_s^l\right| &\le C_0\varepsilon_n^{-\delta/2}\frac{\sup_u K(u)}{\gamma_n}\varepsilon_n'\sum_{\tau_i^n<t}\int_{\tau_{i-1}^n}^{\tau_i^n}\mathrm{d}|A^l|_s\\ &\le C_0\varepsilon_n^{-\delta/2}\frac{\sup_u K(u)}{\gamma_n}\varepsilon_n'(|A|_T-|A|_{t_0}).\end{aligned}$$

In view of (iii) we have $\dfrac{\varepsilon_n'}{\gamma_n\varepsilon_n^{\delta/2}}\to 0$, which implies the a.s. convergence to 0 in the sup-norm. Hence the a.s. convergence

$$\varepsilon_n^{-\delta/2}\sup_{t\in[0,T]}|\Sigma_t^n-\Sigma_t| \xrightarrow[n\to+\infty]{a.s.} 0$$

is proved under the assumption of square summability of $(\frac{\varepsilon_n'}{\varepsilon_n^{\delta/2}\gamma_n})_{n\ge0}$.

To prove (2.4.3) in the general case under (i)-(ii)-(iii) we use the subsequence principle. For any subsequence $(\iota(n))_{n\ge0}$ of positive integers there exists another subsequence $(\iota\circ\iota'(n))_{n\ge0}$ for which $\sum_{n\ge0}\left(\dfrac{\varepsilon'_{\iota\circ\iota'(n)}}{\varepsilon_{\iota\circ\iota'(n)}^{\delta/2}\gamma_{\iota\circ\iota'(n)}}\right)^2 < +\infty$. Thus as shown earlier in the proof, we have $\varepsilon_{\iota\circ\iota'(n)}^{-\delta/2}\sup_{t\in[0,T]}|\Sigma_t^{\iota\circ\iota'(n)}-\Sigma_t| \xrightarrow[n\to+\infty]{a.s.} 0$. Using Lemma 2.2.2 we obtain the desired convergence (2.4.3). □

Now we are ready to finish the proof of the main result of this section.

Proof of Theorem 2.4.1. Using Lemma 2.3.5 and the Frobenius norm $|.|_F$ on matrices, write

$$|\sigma_t^n-(\sigma_t\sigma_t^{\mathsf{T}})^{1/2}|_F = |(\Sigma_t^n+\alpha_n\,\mathrm{Id}_d)^{1/2}-(\Sigma_t)^{1/2}|_F \tag{2.4.12}$$

$$\le \frac{1}{2\sqrt{\min(\lambda_{\min}(\Sigma_t^n)+\alpha_n,\lambda_{\min}(\Sigma_t))}}(|\Sigma_t^n-\Sigma_t|_F+\alpha_n\sqrt{d}). \tag{2.4.13}$$

From this it is easy to deduce via Lemma 2.2.2 the convergence

$$\varepsilon_n^{-\delta/2} \sup_{t\in[0,T]} |\sigma_t^n - (\sigma_t\sigma_t^\mathsf{T})^{1/2}| \xrightarrow[n\to+\infty]{\mathbb{P}} 0,$$

using that $\varepsilon_n^{-\delta/2}\alpha_n \to 0$ and the convergence

$$\varepsilon_n^{-\delta/2} \sup_{t\in[0,T]} |\Sigma_t^n - \Sigma_t| \xrightarrow[n\to+\infty]{\mathbb{P}} 0$$

given by Proposition 2.4.4. Indeed for some subsequence $(\iota(n))_{n\geq 0}$ we have

$$\varepsilon_{\iota(n)}^{-\delta/2} \sup_{t\in[0,T]} |\Sigma_t^{\iota(n)} - \Sigma_t| \xrightarrow[n\to+\infty]{a.s.} 0, \tag{2.4.14}$$

therefore a.s.

$$\inf_{n\geq 0} \inf_{t\in[0,T]} \left(\lambda_{\min}(\Sigma_t^{\iota(n)}) + \alpha_{\iota(n)}\right) > 0,$$

and thus, in view of (2.4.13) and (2.4.14), we obtain

$$\varepsilon_{\iota(n)}^{-\delta/2} \sup_{t\in[0,T]} |\sigma_t^{\iota(n)} - (\sigma_t\sigma_t^\mathsf{T})^{1/2}| \xrightarrow[n\to+\infty]{a.s.} 0.$$

In the above, the strict positivity uniformly in n stems from the convergence of $\Sigma^{\iota(n)}$, the fact that $\inf_{t\in[0,T]} \lambda_{\min}(\Sigma_t) > 0$ and the strict positivity of $\lambda_{\min}(\Sigma_t^{\iota(n)}) + \alpha_{\iota(n)}$ for any $n \geq 0$. Hence finally using Lemma 2.2.2 we deduce that

$$\varepsilon_n^{-\delta/2} |\sigma_t^n - (\sigma_t\sigma_t^\mathsf{T})^{1/2}| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Note that from the construction of the sequence $\{\sigma_t^n : t \in [0,T]\}_{n\geq 0}$ it satisfies Assumptions 4 by taking subsequences as above. So the assumptions of Theorem 2.2.5 are verified and its application shows that the sequence of discretization strategies based on $\{\sigma_t^n : t \in [0,T]\}_{n\geq 0}$ is optimal. The proof is finished. □

2.5 A numerical example

We consider a 2-dimensional diffusion process S on $[t_0, T], t_0 = -0.1, T = 1$ with a deterministic diffusion coefficient, so that S is a Gaussian process that can be simulated exactly. Namely we define $S_{t_0} = 0$ and

$$\mathrm{d}S_t = (\Sigma_t)^{1/2}\mathrm{d}B_t \quad \text{for } t \in [-0.1, 1],$$

where B is a standard 2-dimensional Brownian motion and

$$\Sigma_t = \begin{pmatrix} \sigma_1(t)^2 & \rho\sigma_1(t)\sigma_2 \\ \rho\sigma_1(t)\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Here $\sigma_1(t) = 1 + k\sin(mt)$ with $k = 0.3, m = 10$. The other parameters are set as follows:

$$\sigma_2 = 0.1, \qquad \rho = -0.2.$$

Now we take the following function v, it does not depend on t:

$$v(t, S) = \begin{pmatrix} S^2 \\ 60S^2 \end{pmatrix}.$$

Note that an efficient discretization of $\int_0^1 v(t, S_t) \cdot \mathrm{d}S_t$ must take into account the difference of the sensitivities of v with respect to S^1 and S^2.

We simulate the process S exactly on $[-0.1, 1]$, on a regular grid with a step 0.00002 used for the estimation. In this example we will not perform asymptotic tests, so we will not define the whole sequence $(\varepsilon_n)_{n\geq 0}$. Instead we consider a fixed n and we may directly specify small ε_n that we take.

In this test we consider 4 different discretization methods:

1. asymptotically optimal discretization method, given in Theorem 1.5.2, using the exact knowledge of Σ_t;

2. asymptotically optimal discretization method based on estimation of the diffusion coefficient, given by Theorem 2.4.1 that uses a kernel $K(u) = \frac{1}{\kappa}\mathbb{1}_{-\kappa\leq u\leq 0}$;

3. discretization method based on hitting times of equal circles, i.e. with the quadratic form process Λ, generating the ellipsoids in (2.2.12), equal to $\lambda\,\mathrm{Id}$;

4. regular deterministic discretization grid.

The method based on the hitting of equally sized circles may be seen as a choice by default when one does not want to estimate the model.

We fix ε_n to be 0.025. To get the grids of approximately equal sizes, we take $\lambda = 2$ in the construction of the circle hitting times grid. This value of ε_n is empirically chosen to get nearly 1000 discretization points on $[0, 1]$, further for the deterministic discretization we set the grid size equal exactly 1000. The lookback window for the estimation (i.e. $\kappa\gamma_n$) is 0.002 with 100 estimation points.

We simulate 25 trajectories of the process S. For each trajectory we compute the optimal lower bound equal to $\left(\int_0^1 \mathrm{Tr}(X_t)\mathrm{d}t\right)^2$. Further for each discretization method we calculate the grid size N_1^n, the quadratic variation of the discretization error $\langle Z^n\rangle_1$, using the finest grid with time step 0.00002.

The quantity of interest for each method is $\beta_{n,1} = \dfrac{N_1^n\langle Z^n\rangle_1}{\left(\int_0^1 \mathrm{Tr}(X_t)\mathrm{d}t\right)^2}$. Note that the theoretical results imply that the first two methods are asymptotically optimal and thus must have $\beta_{n,1}$ close to 1.

On Figure 2.1 we display the values of $\beta_{n,1}$ for the 4 discretization methods in each of the 25 simulations.

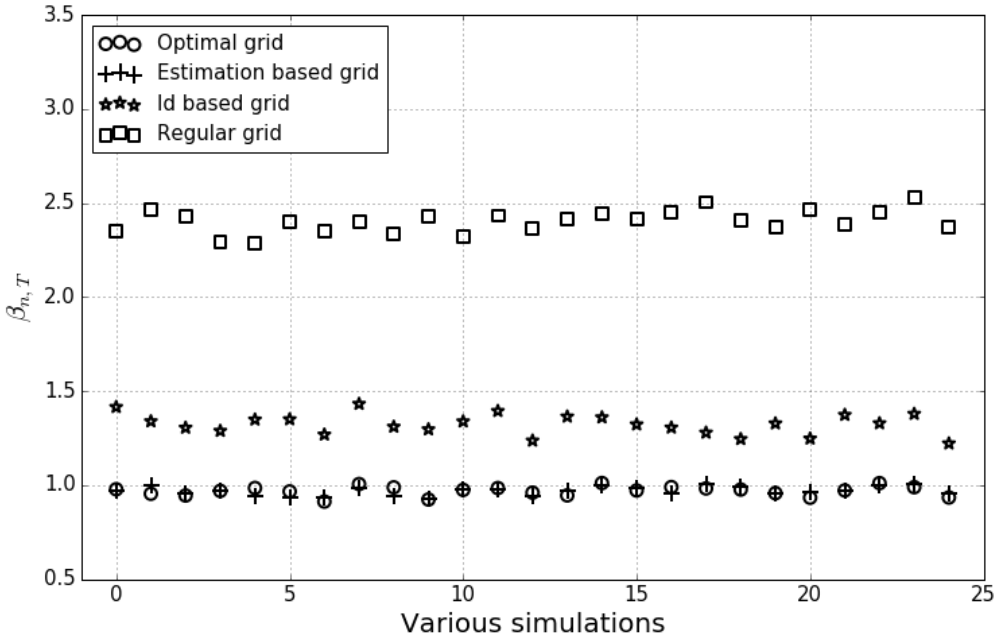

Figure 2.1: The values of the ratio $\beta_{n,1}$ of the renormalized discretization error and the optimal limit lower bound for 4 discretization methods, 25 different simulations.

The numerical test strongly confirms the theoretical results about the optimality of the estimation based algorithms. We see that for both standard and adaptive versions of the optimal discretization, the values of $\beta_{n,1}$ are very close to 1, thus the renormalized error nearly attains the optimal lower bound. For the circle hitting times method, the strategies are less accurate, we lose nearly 40 percent in efficiency. Further the first three methods based on stopping times largely outperform the regular discretization with approximately the same grid size. Remark that this is true in exactly each of the 25 simulations which is in line with the a.s. nature of the theoretical results. To conclude, the tests confirm that the use of adaptive estimation in the case of no prior information about the model does not impact the quality of the discretization method and the model-adaptive version of the optimal discretization works as well as the initial one.

2.A Appendix

2.A.1 Convergence of discrete quadratic variation

We reformulate Proposition 1.3.9 in our setting.

Proposition 2.A.1. *Let S be a semimartingale of the form (2.1.1)-(2.1.2) satisfying Assumptions 1. Let $\mathcal{T} = \{\mathcal{T}^n : n \geq 0\}$ be a sequence of strategies satisfying 2.2.1 with*

$\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *For any continuous adapted* $d\times d$*-matrix process* $(H_t)_{0\leq t\leq T}$*, we have*

$$\sum_{\tau_{i-1}^n<T} \Delta S_{\tau_i^n}^{\mathsf{T}} H_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \overset{a.s.}{\underset{n\to+\infty}{\to}} \int_0^T \mathrm{Tr}\left(H_t \mathrm{d}\langle M\rangle_t\right).$$

Further we recall the following result from [GL14a] (see Corollary 2.1) about the a.s. convergence to zero of sequences of martingales.

Lemma 2.A.2. *Let* $p>0$ *and let* $\{M_t^n : 0\leq t\leq T\}_{n\geq 0}$ *be a sequence of scalar continuous local martingales vanishing at zero. Then*

$$\sum_{n\geq 0}\langle M^n\rangle_T^{p/2} < +\infty \quad a.s. \iff \sum_{n\geq 0}\sup_{0\leq t\leq T}|M_t^n|^p < +\infty \quad a.s..$$

In particular the left hand side implies that $\sup_{0\leq t\leq T}|M_t^n| \overset{a.s.}{\underset{n\to+\infty}{\to}} 0$.

Remark 2.A.3. *Note that from a simple time-change argument the statement of Lemma 2.A.2 holds for martingales defined on* $[t_0,T], t_0<T$*, and vanishing at* t_0*, where the supremum is be taken over* $[t_0,T]$*.*

2.A.2 Proof of Lemma 2.3.1

Here we closely follow the proof of [GL14a, Proposition 2.4]. First note that the convergence (2.3.1) implies

$$\sup_{n\geq 0}\left(\sup_{t\in[0,T]}\left|[H_t^n]^{\varepsilon_n^\delta}\right|\right) \leq \sup_{n\geq 0}\left(\sup_{t\in[0,T]}|H_t^n| + \varepsilon_n^\delta\right) < +\infty.$$

From this we deduce (similarly as in [GL14a, Proposition 2.4]) that the sequence of stopping times $\mathcal{T}^n$ is a.s. finite for all $n\geq 0$.

We now have to verify the two assumptions 2.2.1 and 2.2.1 in Definition 2.2.1 about admissibility. Let us begin with 2.2.1:

$$\sup_{t\in[0,T]}|\Delta S_t|^2 \leq \frac{1}{\inf_{t\in[0,T]}\lambda_{\min}([H_t^n]^{\varepsilon_n^\delta})}\sup_{1\leq i\leq N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}(\Delta S_t^{\mathsf{T}}[H_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta}\Delta S_t) \leq \frac{2}{\varepsilon_n^\delta}\varepsilon_n^{2+\delta} = 2\varepsilon_n^2$$

using (2.1.9).

Now let us establish 2.2.1. For n large enough so that $\varepsilon_n\leq 1$, we have

$$\begin{aligned}
\varepsilon_n^{2\rho_N} N_T^n &\leq \varepsilon_n^{2+\delta} N_T^n = \varepsilon_n^{2+\delta} + \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}}[H_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta}\Delta S_{\tau_i^n} \\
&\leq \varepsilon_n^{2+\delta} + \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}} H_{\tau_{i-1}^n}\Delta S_{\tau_i^n} + \sum_{\tau_i^n<T}|\Delta S_{\tau_i^n}|^2\left|[H_{\tau_{i-1}^n}^n]^{\varepsilon_n^\delta} - H_{\tau_{i-1}^n}\right| \\
&\leq \varepsilon_n^{2+\delta} + \sum_{\tau_i^n<T}\Delta S_{\tau_i^n}^{\mathsf{T}} H_{\tau_{i-1}^n}\Delta S_{\tau_i^n} + \sup_{n\geq 0}\left(\varepsilon_n^\delta + \sup_{t\in[0,T]}|H_t^n - H_t|\right)\sum_{\tau_i^n<T}|\Delta S_{\tau_i^n}|^2.
\end{aligned}$$

Using only 2.2.1 and the continuity of H, Proposition 2.A.1 gives us that

$$\sum_{\tau_i^n<T} \Delta S_{\tau_i^n}^{\mathsf{T}} H_{\tau_{i-1}^n} \Delta S_{\tau_i^n} \overset{a.s.}{\underset{n\to+\infty}{\to}} \int_0^T \mathrm{Tr}(H_t \mathrm{d}\langle M\rangle_t),$$
$$\sum_{\tau_i^n<T} |\Delta S_{\tau_i^n}|^2 \overset{a.s.}{\underset{n\to+\infty}{\to}} \mathrm{Tr}(\langle M\rangle_T).$$

Thus the sequence $(\varepsilon_n^{2\rho_N} N_T^n)_n$ is a.s. finite and the assumption 2.2.1 is proved. The sequence $\{\mathcal{T}^n\}_{n\geq 0}$ is admissible. □

2.A.3 Proof of Lemma 2.3.2

Using that σ_t is a.s. continuous and non-degenerate we get that

$$\inf_{t\in[0,T]} \lambda_{\min}((\sigma_t\sigma_t^{\mathsf{T}})^{1/2}) > 0.$$

The above and the convergence (2.3.3) readily implies that

$$\inf_{n\geq 0}\left(\inf_{t\in[0,T]} \lambda_{\min}(\sigma_t^n)\right) > 0.$$

The fact that it holds for any $n \geq 0$ (and not only asymptotically) is made possible owing to Assumptions 3. Assumptions 1 (in particular (1) and (1)) imply that C_t is non-zero for any $t \in [0,T]$ a.s.. Thus from the continuity of C_t we deduce that

$$\inf_{t\in[0,T]} \mathrm{Tr}(C_t C_t^{\mathsf{T}}) > 0. \tag{2.A.1}$$

Writing

$$|C_t C_t^{\mathsf{T}} - C_t^n (C_t^n)^{\mathsf{T}}| = |\sigma_t^{\mathsf{T}} (D_x v_t)^{\mathsf{T}} \sigma_t \sigma_t^{\mathsf{T}} D_x v_t \sigma_t - (\sigma_t^n)^{\mathsf{T}} (D_x v_t)^{\mathsf{T}} \sigma_t^n (\sigma_t^n)^{\mathsf{T}} D_x v_t \sigma_t^n|$$

and using that $\sup_{n\geq 0}(|\sigma_t^n|)$ is a.s. bounded, we easily deduce from (2.3.3) that

$$\sup_{t\in[0,T]} |C_t C_t^{\mathsf{T}} - C_t^n (C_t^n)^{\mathsf{T}}| \overset{a.s.}{\underset{n\to+\infty}{\to}} 0,$$

which in particular implies that

$$\inf_{n\geq 0}\left(\inf_{t\in[0,T]} \mathrm{Tr}(C_t^n (C_t^n)^{\mathsf{T}})\right) > 0. \tag{2.A.2}$$

Consider now the matrix equation (2.2.1)

$$2\,\mathrm{Tr}(x)x + 4x^2 = cc^{\mathsf{T}}. \tag{2.A.3}$$

For $x \in \mathcal{S}_+^d$, we have the easy inequalities

$$\mathrm{Tr}(x^2) \leq \mathrm{Tr}(x)^2 \leq d\,\mathrm{Tr}(x^2).$$

Then taking the trace in (2.A.3) we get

$$(4+2d)\,\mathrm{Tr}(x^2) \geq 2\,\mathrm{Tr}(x)^2 + 4\,\mathrm{Tr}(x^2) = \mathrm{Tr}(cc^\mathsf{T}),$$

from which we deduce

$$\mathrm{Tr}(x)^2 \geq \mathrm{Tr}(x^2) \geq \frac{1}{4+2d}\,\mathrm{Tr}(cc^\mathsf{T}).$$

Combining this with the properties (2.A.1) and (2.A.2) leads to

$$\inf_{t\in[0,T]} \mathrm{Tr}(X_t) > 0, \qquad \inf_{n\geq 0}\left(\inf_{t\in[0,T]} \mathrm{Tr}(X_t^n)\right) > 0,$$

which finishes the proof. □

2.A.4 Proof of the counter-example 2.1.1

We have

$$D_x v = \begin{pmatrix} \sqrt{14} & 0 \\ 0 & \sqrt{104} \end{pmatrix}.$$

We use a misspecified model with $\tilde{\sigma} = \mathrm{Id}_2$. Let us follow the construction of the discretization strategy in [GL14a, Theorem 3.3] (here all the processes are constant, we mark by tilde those corresponding to the misspecified model). Thus we obtain

$$\tilde{C} = \tilde{\sigma}^\mathsf{T}(D_x v)^\mathsf{T}\tilde{\sigma} = \begin{pmatrix} \sqrt{14} & 0 \\ 0 & \sqrt{104} \end{pmatrix}.$$

One easily checks that $\tilde{X} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ satisfies the matrix equation

$$2\,\mathrm{Tr}(\tilde{X})\tilde{X} + 4\tilde{X}^2 = \tilde{C}\tilde{C}^\mathsf{T} = \begin{pmatrix} 14 & 0 \\ 0 & 104 \end{pmatrix}.$$

Thus we obtain

$$\tilde{\Lambda} = (\tilde{\sigma}^{-1})^\mathsf{T}\tilde{X}\tilde{\sigma}^{-1} = \tilde{X},$$

which is the matrix used by the optimizer to construct the discretization times as the ellipsoid hitting times. For the given sequence $(\varepsilon_n)_{n\geq 0}$ with $\varepsilon_n \to 0$ we define $\{\mathcal{T}^n\}_{n\geq 0}$, where $\mathcal{T}^n = \{\tau_i^n\}_{i\geq 0}$, to be a sequence of discretization strategies such that: for any $n \in \mathbb{N}$

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T}\tilde{\Lambda}(S_t - S_{\tau_{i-1}^n}) \geq \varepsilon_n^2\} \wedge T. \end{cases} \tag{2.A.4}$$

Using that the true diffusion matrix is $\sigma = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$, so that $\sigma^\mathsf{T}\tilde{\Lambda}\sigma = \mathrm{Id}_2$, the above definition (2.A.4) is equivalent to

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (B_t - B_{\tau_{i-1}^n})^\mathsf{T}\,\mathrm{Id}_2(B_t - B_{\tau_{i-1}^n}) \geq \varepsilon_n^2\} \wedge T. \end{cases} \tag{2.A.5}$$

Now consider the renormalized discretization error for $\{\mathcal{T}^n\}_{n\geq 0}$,

$$\begin{aligned} \tilde{N}_T^n \langle \tilde{Z}^n \rangle_T &= \tilde{N}_T^n \left\langle \sqrt{14} \int_0^\cdot \Delta S_t^1 \mathrm{d}S_t^1 + \sqrt{104} \int_0^\cdot \Delta S_t^2 \mathrm{d}S_t^2 \right\rangle_T \\ &= \tilde{N}_T^n \left(14 \int_0^T (\Delta B_t^1)^2 \mathrm{d}t + \frac{104}{2^4} \int_0^T (\Delta B_t^2)^2 \mathrm{d}t \right). \end{aligned}$$

For a moment, assume that the following holds for $j = 1, 2$:

$$\tilde{N}_T^n \int_0^T (\Delta B_t^j)^2 \mathrm{d}t \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} \frac{T^2}{4} \tag{2.A.6}$$

(which proof is given at the end of the section). Thus we obtain the convergence

$$\tilde{N}_T^n \langle \tilde{Z}^n \rangle_T \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} \frac{41}{8} T^2.$$

At the same time for the optimal discretization algorithm (which attains the optimal lower bound) the $\mathbb{P}$-limit is equal to $\mathrm{Tr}(X)^2 T^2$, where the matrix X is the solution of the equation $2\,\mathrm{Tr}(X)X + 4X^2 = CC^\mathsf{T}$ (see [GL14a, Theorem 3.3] combined with Lemma 2.2.2). Here

$$CC^\mathsf{T} = (\sigma^\mathsf{T}(D_x v)^\mathsf{T}\sigma)^2 = \begin{pmatrix} 14 & 0 \\ 0 & \frac{13}{2} \end{pmatrix}.$$

For the case of diagonal CC^T the solution X can be calculated analytically. Following the proof of [GL14a, Lemma 3.1], $y = \mathrm{Tr}(X)$ is the unique solution to the equation

$$6y = \sqrt{y^2 + 4 \cdot 14} + \sqrt{y^2 + 4 \cdot \frac{13}{2}}.$$

After simple calculations and simplifications, y must solve the bi-quadratic equation

$$1152y^4 - 5904y^2 + 900 = 0,$$

whose solutions for y^2 are $\frac{369 \pm 9\sqrt{1481}}{144}$. Using the inequality $\mathrm{Tr}(X)^2 \geq \dfrac{\mathrm{Tr}(CC^\mathsf{T})}{4 + 2d} = \frac{41}{16}$ from Section 2.A.3, we derive that $\mathrm{Tr}(X)^2 = \frac{369 + 9\sqrt{1481}}{144}$. Therefore we obtain the desired result (2.1.7) with

$$\beta = \frac{41}{8} - \frac{369 + 9\sqrt{1481}}{144} \approx 0.157. \tag{2.A.7}$$

Proof of (2.A.6). [GL14a, Theorem 3.3] applied with $S_t = B_t$ and $D_x v = \mathrm{Id}_2$ (here we have $X_t = X = \Lambda = \frac{1}{2\sqrt{2}}\mathrm{Id}_2$) and combined with Lemma 2.2.2 yields

$$\tilde{N}_T^n \int_0^T ((\Delta B_t^1)^2 + (\Delta B_t^2)^2)\mathrm{d}t \xrightarrow[n\to+\infty]{\mathbb{P}} \left(\int_0^T \mathrm{Tr}(X_t)\mathrm{d}t\right)^2 = \frac{T^2}{2}, \tag{2.A.8}$$

here we suppose that the discretization rule from [GL14a, Theorem 3.3] uses a sequence $(\tilde{\varepsilon}_n)_{n\geq 0}$ satisfying $\tilde{\varepsilon}_n^2 = \frac{\varepsilon_n^2}{2\sqrt{2}}$ so that it is coherent with the one defined in (2.A.5). We now prove that the left hand side of (2.A.6) has a limit in probability, which is a constant. For $j = 1, 2$ write

$$\tilde{N}_T^n \int_0^T (\Delta B_t^j)^2\mathrm{d}t = (\tilde{N}_T^n \tilde{\varepsilon}_n^2)^2 \frac{1}{\tilde{N}_T^n} \sum_{i=1}^{\tilde{N}_T^n - 1} \tilde{\varepsilon}_n^{-4} \int_{\tau_{i-1}^n}^{\tau_i^n} (\Delta B_t^j)^2\mathrm{d}t + o_n(1),$$

where $o_n(1)$ comes for the last summand. From [GL14a, Proof of Theorem 3.2] we obtain

$$\tilde{N}_T^n \tilde{\varepsilon}_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^T \mathrm{Tr}(X_t)\mathrm{d}t = \frac{T}{\sqrt{2}}.$$

Let $(W_t)_{t\geq 0}$ be a 2-dimensional Brownian motion independent of B. Define a sequence of stopping times $(\theta_i)_{i\geq 0}$ by $\theta_0 = 0$ and

$$\theta_{i+1} = \inf\{t > \theta_i : |W_t - W_{\theta_i}|^2 \geq 1\}.$$

Let $\check{N}_T^n$ be the number of θ_i satisfying $\theta_i \leq T\varepsilon_n^{-2}$, it is easy to justify that $\check{N}_T^n \xrightarrow{a.s.} +\infty$. Define a sequence of random variables

$$Q_i = \int_{\theta_i}^{\theta_{i+1}} (W_s^1 - W_{\theta_i}^1)^2\mathrm{d}s.$$

Using the scaling property of B we get the following equality of distributions

$$\frac{1}{\tilde{N}_T^n} \sum_{i=1}^{\tilde{N}_T^n - 1} \varepsilon_n^{-4} \int_{\tau_{i-1}^n}^{\tau_i^n} (\Delta B_t^j)^2\mathrm{d}t \overset{d}{=} \frac{1}{\check{N}_T^n} \sum_{i=1}^{\check{N}_T^n - 1} Q_i.$$

Hence using that $\check{N}_T^n \xrightarrow{a.s.} +\infty$ the Law of Large Numbers implies

$$\frac{1}{\tilde{N}_T^n} \sum_{i=1}^{\tilde{N}_T^n - 1} \tilde{\varepsilon}_n^{-4} \int_{\tau_{i-1}^n}^{\tau_i^n} (\Delta B_t^j)^2\mathrm{d}t \xrightarrow[n\to+\infty]{\mathbb{P}} (2\sqrt{2})^2 \mathbb{E}(Q)$$

and so

$$\tilde{N}_T^n \int_0^T (\Delta B_t^j)^2\mathrm{d}t \xrightarrow[n\to+\infty]{\mathbb{P}} 4T^2\mathbb{E}(Q).$$

As the limit is the same for $j = 1, 2$ in view of the symmetry we deduce $\mathbb{E}(Q) = \frac{1}{16}$ from (2.A.8). Thus, (2.A.6) is proved. □

Chapter 3

Central limit theorem for discretization errors based on stopping time sampling

Contents

3.1 Introduction

Statement of the problem and motivation. In this work we consider the discretization of a multidimensional Itô process S at random stopping times $\tau_0^n = 0 < \tau_1^n < \cdots < \tau_{N_T^n}^n = T$. The number of discretization times N_T^n may be random as well. Our goal is to establish a functional Central Limit Theorem (CLT) for the renormalized discretization error process $(\sqrt{N_t^n}\mathcal{E}_t^n)_{0\le t\le T}$, where $\mathcal{E}_t^n$ is $\mathbb{R}^m$-valued and has the form $\mathcal{E}_t^n = \mathcal{E}_t^{n,1} + \mathcal{E}_t^{n,2}$ with

$$\mathcal{E}_t^{n,1} = \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} \mathcal{M}_{\tau_{i-1}^n}(S_s - S_{\tau_{i-1}^n})\mathrm{d}s, \quad \mathcal{E}_t^{n,2} = \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} (S_s - S_{\tau_{i-1}^n})^{\mathsf{T}} \mathcal{A}_{\tau_{i-1}^n} \mathrm{d}B_s. \tag{3.1.1}$$

Here, B is a d-dimensional Brownian motion, $\mathcal{M}$ and $\mathcal{A}$ are arbitrary adapted continuous processes with values in $\mathrm{Mat}_{m,d}$ and $\mathrm{Mat}_{d,d} \otimes \mathbb{R}^m$ respectively (so that $\mathcal{A}_t$ maps bilinearly $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ to $x^{\mathsf{T}}\mathcal{A}_t y \in \mathbb{R}^m$; see the notation at the end of this section and Section 3.2.3 for more details).

Analysis of the discretization errors based on deterministic discretization grids is a well studied subject developed in works such as [Roo80, JP98, HM05, GT09, GT01, MZ06] among others, see also [JP12] and references therein. However, in practice the discretization times may be random, which makes the analysis much more complicated. The setting of random discretization grids has gained a lot of attention due to applications in high frequency finance (see e.g. [Fuk10, FR12, RR10, RR12]). The importance of the subject was, in particular, emphasized in [DGM$^+$01, Section 1.1] and [ASJ14, Chapter 9]. See also [GW02] for empirical evidence about the connection of volatility and inter-transaction duration in finance, [Fuk10] for modeling bid or ask quotation data and tick time sampling. Many works remark the non-negligible impact of the randomness of the discretization times with respect to the classical deterministic case when dealing with convergence results. For example, in [ASM03] it is observed a considerable effect of random sampling on the estimators in the setting of parametric inference for diffusions. In [LZZ13] the authors note that taking into account the endogenous randomness of the observation grids, when it exists, may substantially improve the performance of the integrated volatility estimator. Certain works (such as e.g. [LR13, ZS16]) consider the case of random but, so called, strongly predictable discretization times. Though important, this case is more basic compared to stopping times. While the theory of stopping time discretization grids has recently experienced substantial progress (see the literature discussion below), the existing results possess a number of drawbacks. First, many of them only cover the 1-dimensional case and typically consider particular examples of the error term, e.g. related to integrated variance estimation (e.g. [Fuk10, Fuk11b, FR12]). Second, in terms of the discretization grids under study, most general works (such as [Fuk11b]) consider an abstract setting with assumptions that are difficult to verify and thus do not explicitly describe the class of those grids. Generalization to the multidimensional case is highly non-trivial both regarding the extension of the central limit theorem (in particular, characterization of the limit distribution) given the abstract assumptions on the moments, and even more regarding the determination of the class of random grids verifying the assumptions. For example, natural candidates for endogenously generated

discretization times are first exit times from random domains, whose analysis is much more complicated in multidimensional setting (while in dimension 1 such a domain is given by the two boundary points). Less abstract works (such as e.g. [FR12, RR12, LZZ13, LMR+14]) study only specific classes of grids and are restricted to either the case of grids given by hitting times or random times driven by a noise independent on the process S, while a combination of these two types of grids has not been addressed.

In our our we propose a unified treatment and we aim at closing these gaps in the existing literature on the subject. Our goal is to prove a functional CLT for the sequence of the renormalized discretization error processes $(\sqrt{N_t}\mathcal{E}_t^n)_{0\leq t\leq T}$ in the multidimensional case for a general error term of the form (3.1.1) for a sufficiently general concrete class of random discretization grids (i.e. specified directly by its definition and not by abstract assumptions) with explicit characterization of the limit distribution. In particular, the model for the process S_t allows quite general (non-Markov) Itô processes verifying mild regularity assumptions, and therefore it includes most of the models relevant in practice. The class of random discretization grids allows a combination of the endogenous randomness generated by S and independent noise, and includes the exit times from general random domains – a framework that was not previously studied in the literature. As seen in (3.2.6) , it will take the form of

$$\tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin \varepsilon_n D^n_{\tau_{i-1}^n}\} \wedge (\tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}) \wedge T,$$

for some parameter $\varepsilon_n \to 0$, some stochastic domain $\mathrm{D}^n_\cdot$ indexed by time, some independent random variables $(U_{n,i})_{i,n}$, some negligible terms $\Delta_{n,i}$. More general forms are even allowed in Section 3.3

In particular, the form (3.1.1) of the error term covers such important applications as error analysis for integrated variance estimation, optimal tracking strategies and parametric estimation for processes. In these applications a discretization error process can be typically decomposed into a linear part of the form (3.1.1) and the rest, that gives negligible contribution. To illustrate consider a process of the form $S_t := S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}B_s$ and let $\Delta S_t := S_t - S_{\tau_i^n}$ with τ_i^n the largest discretization time before t.

1. *Integrated variance estimation.* Here the goal is to estimate $\int_0^t \mathrm{Tr}(\sigma_s\sigma_s^\mathsf{T})\mathrm{d}s$ using the random process observations (see, e.g., [BNS05, MZ06, RR12, LZZ13, LMR+14]). Using the Itô formula we write the error process

$$\sum_{\tau_{i-1}^n<t} |\Delta S_{\tau_i^n\wedge t}|^2 - \int_0^t \mathrm{Tr}(\sigma_s\sigma_s^\mathsf{T})\mathrm{d}s = 2\int_0^t \Delta S_s^\mathsf{T}\sigma_s \mathrm{d}B_s + 2\int_0^t b_s^\mathsf{T}\Delta S \mathrm{d}s.$$

2. *Optimal tracking strategies.* This is related to the minimization of the tracking error of a continuous-times strategy, which, for some function $v : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$, may be written in the form

$$\int_0^t v(s, S_s)\mathrm{d}S_s - \sum_{\tau_{i-1}^n<t} v(\tau_{i-1}^n, S_{\tau_{i-1}^n})\Delta S_{\tau_i^n\wedge s} \approx \sum_{\tau_{i-1}^n<t} \int_{\tau_{i-1}^n}^{\tau_i^n\wedge t} \nabla_S v(\tau_{i-1}^n, S_{\tau_{i-1}^n})\Delta S_s \mathrm{d}S_s,$$

which is a particular case of (3.1.1). See, e.g., [Fuk11b, Fuk11a, GL14a] and Chapters 1-2.

3. *Parametric estimation for processes.* In the problem of the parametric inference for a diffusion process, depending on a parameter $\xi \in \Xi$, and observed at discrete times, the proof of the CLT for the renormalized estimation error sequence $(\sqrt{N_T^n}(\xi^n - \xi^\star))_{n\geq 0}$ often requires the CLT for discretization errors of the form (3.1.1). This application is developped in Chapter 4, where the question of optimal contrast estimator is investigated together with optimal observation grids, see also [GCJ93, GCJ94, ASM04].

There also exists a number of works related to the asymptotic analysis of the Euler scheme error, see [FO15] and references therein. We remark, however, that our setting is quite different since we deal we discretization of the true process trajectories, and thus we do not consider this problem here.

Background results. A number of works deal with the case of strongly predictable (possibly up to conditioning on some independent noise) grids. This case is studied in [ASM03, ASM04] in the setting of parametric inference for diffusions. [DG04] investigates the inference problem for Markov processes observed at random times and provides, in particular, their asymptotic normality. Sampling scheme intervals are made of exponential times with intensity that may depend on the underlying Markov process and on the parameter vector to be estimated (compare with our analysis in Section 3.2.2, where we do not use any Markovian assumptions). [BNS05] studies the estimation of general power variations for time-changed regular schemes. In [BNGJ+06] a CLT is proved for realized power and bipower variations of continuous semimartingales for even functions, [KP08] provide an extension of this result for more general functions. Among other works we also mention [Roo80, GT01, HM05, GT09] where the asymptotic properties of the discretization error for stochastic integrals are studied in various settings. In [GCJ93, GCJ94] an asymptotic normality property is given for an estimator of a parameter of a diffusion process based on discrete observations. Though in [GCJ94] random observation times are allowed, the framework is quite different since those times are not stopping times and are chosen by the user given anticipative observations.

Some more recent papers study the case of endogenously triggered discretization times. In [RR10, RR12] the authors construct a financial high-frequency price model which combines microstructure noise, including rounding noise, and sampling at transaction times on the basis of suitably defined hitting times, and then estimate the integrated volatility. They also provide the asymptotic analysis of their estimator. A different approach with endogenous random sampling times can be found in [LZZ13, LMR+14], the authors prove a CLT for the realized volatility in a general endogenous time setting. Random discretization schemes given by hitting times appear naturally in the problem of optimizing the tracking error which is interpreted as a discretization error of stochastic integral (see [Fuk11b, Fuk11a, GL14a] and Chapters 1-2).

In the above mentioned works the CLT typically holds with a limit having zero correlation

with the initial Brownian motion, if the time grids satisfy the centering property

$$\mathbb{E}[(B^j_{\tau^n_{i+1}} - B^j_{\tau^n_i})^{2p+1}|\mathcal{F}_{\tau^n_i}] = 0 \tag{3.1.2}$$

for the odd moments of the increments of the j-th component of the Brownian motion B. When the observation times are deterministic, (3.1.2) can be proved thanks to Gaussian centering property. When $(\tau^n_i)_{i\leq N^n_T}$ are hitting times of symmetric intervals (thus the property (3.1.2) still holds), [Fuk11a] shows also a CLT with an independent Brownian motion at the limit in the case of dimension 1. When (3.1.2) holds asymptotically, that is the third moment of the increments oscillates around 0 fast enough ($\varepsilon_n^{-1}\mathbb{E}[(B^j_{\tau^n_{i+1}} - B^j_{\tau^n_i})^3|\mathcal{F}_{\tau^n_i}] \to^{\mathbb{P}} 0$), as for instance in the case of random grids made of hitting times of regularly spaced meshes, the authors of [FR12] prove also no bias in the limit. However, many important cases require treatment of irregular grids without the centering property (3.1.2) and the analysis of asymptotic bias. For example, in [LZZ13] a CLT with biased limit distribution is given in the case of integrated volatility estimation on randomly spaced observations in dimension 1.

The closest to our setting are the works by Fukasawa *et al.* In particular, [Fuk10] proves the CLT for hitting times of a regular grid in dimension 1. In [Fuk11b] the CLT for renormalized discretization error sequence is established in a very general (allowing, in particular, the asymptotic bias) setting in dimension 1. The result, though quite important, is given in terms of abstract assumptions on the moments of the process increments, which determine, in particular, the limit distribution. Verification of these assumptions is hard beyond simple cases (and would be particularly complicated if the result were generalized to the multidimensional case). Hence, an application of this result to a concrete example requires a substantial amount of work, including the assumption checking and finding the limit distribution.

Our contribution.

- To our knowledge, this is the first attempt to study the convergence in distribution of discretization errors for a concrete general class of Itô processes and random discretization grids given by stopping times. In particular, our models for the process and the discretization times are specified directly, in simple terms and without abstract assumptions, so that verification for a specific example is quite straightforward. In addition, we provide explicitly the limit distribution (the asymptotic bias and covariance matrix) in a tractable form in terms of the underlying model. We consider both multidimensional process and multidimensional error term.

- Our class of random discretization grids includes, in particular, hitting times of general random multidimensional domains (under quite mild assumptions). To our knowledge, this is the first work that studies such discretization grids.

- Our class of random grids allows a combination of endogenous (e.g. given by hitting times) and exogenous noise (given by independent random variables, e.g. Poisson process) in the definition of discretization times, while a majority of previous works is restricted to only one of these cases.

- We consider a general error term $\mathcal{E}_t^n = \mathcal{E}_t^{n,1} + \mathcal{E}_t^{n,2}$ given by (3.1.1) which covers simultaneously most of the applications of interest.

- We do not impose any Markovian assumptions either on the process or on discretization times.

- An important and direct application of our results is when time grids are made of hitting times of random ellipsoids. Such grids naturally appear in [GL14a] and Chapters 1-2 as optimal discretization strategies regarding the minimization of quadratic variation criterion for multidimensional models and play important role in the problem of hedging error optimization in finance (see [Fuk11a]).

- Furthermore, to derive the above CLT for general grids, we have proved several important results on the sensitivity of the exit times of Brownian semimartingales from bounded domains with respect to the model and domain perturbations. They are of their own interest and may be useful in other problems.

- Though our framework is multidimensional, we consider only discretization times that are synchronized for different components of the process. This may be not true in various applications. See, for example, [HY08] for the treatment of asynchronous observations in high-frequency finance. Nevertheless, our work is an important intermediary step for the study of more general asynchronous discretization schemes and provides useful machinery to tackle this problem. Generalization of our results to, for example, bipower variations (see [BNS05, BNGJ$^+$06, KP08]) is left for future research.

Organization of the chapter. In Section 3.2 we introduce the stochastic model for the semimartingale S and describe the class of random discretization grids under study. Further we state the main theorem of this work and provide various examples and applications of our result. Section 3.3 is devoted to the proof of the main theorem, which contains two important blocks: a general abstract CLT for discretization errors based on random grids (Section 3.3.1) and certain important properties of the semimartingale exit times from general domains (Section 3.3.2). The completion of the proof is given in Section 3.3.3. In Section 3.4 we continue with the proof of the general abstract CLT, while Section 3.5 is devoted to the proof of the semimartingale exit time properties. Supplementary material and technical results are given in Appendix.

Notation used throughout this work.

- $v \cdot w$ denotes the scalar product in $\mathbb{R}^d$.

- $\mathrm{Mat}_{m,d}$ denotes the set of $m \times d$ real matrices. $\mathrm{Tr}(.)$ and $^\intercal$ stand respectively for the trace and transpose operators.

- We write $(M)_{ij}$ for the components of a matrix M, $M_{i:}$ (resp. $M_{:i}$) its i-th row (resp. i-th column), and a^k for the components of the vector a.

- $\mathcal{S}_d$, $\mathcal{S}_d^+$ and $\mathcal{S}_d^{++}$ denote respectively the set of symmetric, positive semidefinite symmetric and positive definite symmetric real $d \times d$ matrices.
- For $M \in \mathrm{Mat}_{m,d}$ we denote by $\|M\| := \sqrt{\mathrm{Tr}(MM^\intercal)}$ its Frobenius norm. For $M \in \mathrm{Mat}_{d,d}$, we recall the easy inequality $|\mathrm{Tr}(M)| \leq \sqrt{d}\|M\|$.
- For $M \in \mathcal{S}_d$ we denote $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ the smallest and the largest eigenvalue of M.
- We denote by: $\overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}}$ - a.s. convergence uniform on $[0,T]$, $\overset{u.c.p.}{\underset{n\to+\infty}{\longrightarrow}}$ - convergence in probability uniform on $[0,T]$, $\underset{[0,T]}{\Longrightarrow}$ - convergence in distribution on $[0,T]$ in the sense of processes w.r.t. the uniform topology..
- $B_d(x_0, R)$ denotes a d-dimensional closed ball with radius R and center x_0.
- $\mathcal{U}(0,1)$ stands for the distribution of a uniform random variable on $[0,1]$.
- $\mathcal{C}_{\sup}([0,T])$ denotes the normed vector space of continuous processes on $[0,T]$ with the sup-norm.
- If $f : \mathbb{R}^d \mapsto \mathbb{R}$ is a smooth function, then ∇f (resp. $\nabla^2 f$) stands for the gradient (resp. the Hessian) of f, as a row vector (resp. as a square matrix).
- A $f : \mathbb{R}^d \mapsto \mathbb{R}$ is an α-homogeneous function (for some $\alpha \in \mathbb{N}$) if $f(cx) = c^\alpha f(x)$ for all $c \geq 0, x \in \mathbb{R}^d$.
- All the further asymptotic convergences are stated through a positive deterministic sequence $(\varepsilon_n)_{n\geq 0}$ with $\varepsilon_n \to 0$. Without loss of generality and for the sake of simplicity, from now on we assume $\varepsilon_n \leq 1$ for any n.
- For any subinterval $I \subset [0,T]$ denote $N^n(I) := \#\{\tau_i^n \in I\}$ for the number of grid times in I. Let $|I|$ denote the length of I.
- In what follows, we may consider the conditional expectation of scalar random variables $\mathcal{X}$ that are non necessarily integrable. We adopt the following convention. When $\mathcal{X}$ is non-negative, $\mathbb{E}_t(\mathcal{X})$ can be properly defined as a random variable valued in $\mathbb{R}_+ \cup \{+\infty\}$. In the case of $\mathbb{E}_t(|\mathcal{X}|) < +\infty$ a.s. we define $\mathbb{E}_t(\mathcal{X}) := \mathbb{E}_t(\mathcal{X}_+) - \mathbb{E}_t(\mathcal{X}_-)$ where $\mathcal{X}_+$ and $\mathcal{X}_-$ are the positive and the negative parts of $\mathcal{X}$.

3.2 Stochastic model, random grids, main result

3.2.1 Probabilistic model

Let $T > 0$ and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ be a filtered probability space supporting a d-dimensional Brownian motion $(B_t)_{0\leq t\leq T}$. We assume that the filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ satisfies the usual assumptions of being right-continuous and $\mathbb{P}$-complete. Let $(S_t)_{0\leq t\leq T}$ be a d-dimensional continuous $\mathcal{F}$-adapted semimartingale.

Our first CLT (Theorem 3.2.7) and the computation of explicit limits in Section 3.2.4 will be derived under the following assumptions and for stopping times of the form (3.2.6). A slightly more general version of CLT is established in Section 3.3.1, for abstract stopping times satisfying some structure conditions ($\mathbf{H}_R$)-($\mathbf{H}_\mathcal{B}$).

($\mathbf{H}_S$)**:** The process S is of the form

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}B_s, \quad t \in [0,T], \tag{3.2.1}$$

where

- the starting point S_0 is an $\mathcal{F}_0$-measurable random variable;
- $(b_t)_{0\le t\le T}$ is a $\mathcal{F}$-adapted d-dimensional stochastic process;
- $(\sigma_t)_{0\le t\le T}$ is a continuous $\mathcal{F}$-adapted $\mathrm{Mat}_{d,d}$-valued process, such that σ_t is invertible a.s. for all $t \in [0,T]$ and σ_0, σ_0^{-1} are bounded;
- for some a.s. finite random variable $C_\sigma > 0$ satisfying $\mathbb{E}\left(C_\sigma^4 | \mathcal{F}_0\right) < +\infty$ and a parameter $\eta_\sigma \in (0,1]$, we have

$$|\sigma_t - \sigma_s| \le C_\sigma |t-s|^{\eta_\sigma/2} \qquad \forall s,t \in [0,T] \quad \text{a.s.}$$

We remark that the boundedness of σ_0 and σ_0^{-1} above is needed mainly to guarantee that certain processes are integrable at 0 in the proof of Proposition 3.4.1 in Section 3.A.2, which is an important step of our main proof. Later similar boundedness condition is assumed for some other processes for the same reason.

($\mathbf{H}_\Delta$)**:** There exist positive $\mathcal{F}$-adapted processes $(v_t)_{0\le t\le T}$ and $(\delta_t)_{0\le t\le T}$, such that v_t is a.s. bounded and δ_t is a.s. continuous, and for which we have a.s. for all $t \in [0,T]$

$$v_t^{-1} \le \inf_{t\le s\le \psi(t)} \lambda_{\min}(\sigma_s \sigma_s^\mathsf{T}) \le \sup_{t\le s\le \psi(t)} \|\sigma_s \sigma_s^\mathsf{T}\| \le v_t, \qquad \sup_{t\le s\le \psi(t)} |b_s| \le v_t,$$

where

$$\psi(t) := \inf\{s \ge t : |S_s - S_t| \ge \delta_t\} \wedge T, \quad t \in [0,T].$$

In ($\mathbf{H}_\Delta$) the key assumption is that v_t is $\mathcal{F}$-adapted, so that it allows $\mathcal{F}_t$-measurable control on $[t, \psi(t)]$ for $t \in [0,T]$.

Example 3.2.1. *On $(\Omega, \mathcal{F}, \mathbb{P})$ consider a Brownian motion $(B_t)_{0\le t\le T}$ and a continuous-time Markov chain $(P_t)_{0\le t\le T}$ taking values in $\mathbb{N}_R := \{1,\dots,R\}$, that is aimed at modeling a regime-switching behavior (see [Nor98, Chapter 2]). The label $r \in \mathbb{N}_R$ stands for indexing the different regimes. The transition from state r to state r' in two successive times is given by a Frobenius matrix M_F and the distributions of time interval between two jumps are exponential distributions, with a parameter depending on M_F. Define the $\mathbb{P}$-augmented right-continuous extension $(\mathcal{F}_t)_{0\le t\le T}$ of the filtration generated by (B,P). Consider the processes*

$$\sigma_t = \sigma\left(t, (S_{s\wedge t})_{0\le s\le T}\right), \qquad b_t = b\left(P_t, t, (S_{s\wedge t})_{0\le s\le T}\right)$$

for functions $\sigma : [0,T] \times \mathcal{C}_{\sup}([0,T]) \to \mathrm{Mat}_{d,d}$ *such that* σ_t^{-1} *exists for all* $t \in [0,T]$ a.s. *and* $b : \mathbb{N}_R \times [0,T] \times \mathcal{C}_{\sup}([0,T]) \to \mathbb{R}^d$. *Suppose that* $\sigma(\cdot,\cdot)$ *is continuous and that* $b(r,\cdot,\cdot)$ *is continuous for all* $r \in \mathbb{N}_R$. *Thus for a given continuous positive process* v_t, *since* σ_t *is invertible, we may choose* δ_t *(continuous in* t*) small enough, such that if the trajectory* $(S_{s\wedge\psi(t)})_{0\le s\le T}$ *is at distance at most* δ_t *from* $(S_{s\wedge t})_{0\le s\le T}$ *we may upper and lower bound the eigenvalues of* $\sigma\left(u,(S_{s\wedge u})_{0\le s\le T}\right), u \in [t,\psi(t)]$, *using* v_t. *Similar reasoning yields the condition on* b_t *in* ($\mathbf{H}_\Delta$). *We remark that this model is path-dependent (thus non-Markovian) and non-only driven by Brownian motions (which justifies the use of general filtration). It also includes the diffusion model* $\sigma_t = \sigma(t,S_t)$ *as a particular case.*

3.2.2 Class of random discretization grids

In this section we discuss the class of random discretization grids for which we study the discretization error, in particular, for which we establish the functional CLT with explicit limit characterization.

- This class is quite large and includes the hitting times of general random domains. Notably, it allows almost arbitrary random domain processes under some mild regularity assumptions. We claim that this is the most general concrete framework (i.e. with explicit description and without any abstract assumption) for endogenously generated discretization schemes for multidimensional processes considered in the literature.

- In addition we allow to incorporate additional independent noise of quite general form while constructing the discretization times.

In particular, examples include random grids given by a combination of the hitting times of random domains with the times generated by a Poisson process having general random path-dependent intensity and independent source of randomness.

We recall that $(\varepsilon_n)_{n\ge 0}$ is a deterministic sequence with $\varepsilon_n \in (0,1]$ and $\varepsilon_n \to 0$.

A set of regular bounded domains

We recall that a domain is a non-empty open connected set, see [GT83, p.10]. Let $\widetilde{\mathcal{D}}$ be the set of bounded domains D in $\mathbb{R}^d$ which contains 0, and let $\mathcal{D}$ be the subset of $\widetilde{\mathcal{D}}$ which element D has a boundary ∂D of class $\mathcal{C}^2$. For any $D \in \widetilde{\mathcal{D}}$, define the signed distance $\delta_{\partial D} : \mathbb{R}^d \to \mathbb{R}$ to its boundary by

$$\delta_{\partial D}(x) := (\mathbb{1}_{x\in D} - \mathbb{1}_{x\notin D}) \inf\{|x-y| : y \in \partial D\}. \tag{3.2.2}$$

We recall that without any regularity on ∂D, $\delta_{\partial D}$ is a Lipschitz function with Lipschitz constant smaller than 1 (see [GT83, Section 14.6, p. 354]). For any $D^1, D^2 \in \widetilde{\mathcal{D}}$ define

$$\mu(D^1,D^2) := \sup_{x\in\partial D^1} |\delta_{\partial D^2}(x)| + \sup_{x\in\partial D^2} |\delta_{\partial D^1}(x)|.$$

Lemma 3.2.2. $\mu(.,.)$ *is a distance on the set* $\widetilde{\mathcal{D}}$ *of domains of* $\mathbb{R}^d$ *containing 0.*

Proof. It is obviously non-negative and symmetric.
Assume that $\mu(D^1, D^2) = 0$ for $D^1, D^2 \in \widetilde{\mathcal{D}}$ and let us show that $D^1 = D^2$. We have $\sup_{x\in\partial D^1} |\delta_{\partial D^2}(x)| = 0$: for any $x \in \partial D^1$, since the boundary ∂D^2 is compact, there exists a $y(x) \in \partial D^2$ such that $0 = \delta_{\partial D^2}(x) = |x - y(x)|$ which shows that $\partial D^1 \subset \partial D^2$. Using $\sup_{x\in\partial D^2} |\delta_{\partial D^1}(x)| = 0$, we get the converse inclusion and therefore, $\partial D^1 = \partial D^2$. But since D^1 and D^2 are open connected sets containing 0, we must have $D^1 = D^2$.
It remains to prove that μ satisfies to the triangular inequality: this is an easy verification that we leave to the reader. The proof is complete. □

To allow greater generality and deal with intersection of J smooth domains (to encompass domains with corners like polyhedrons) we introduce appropriate notations. For any integer $J > 0$, let

$$\mathcal{D}^J := \{(D_1, \dots, D_J) : D_j \in \mathcal{D}\}, \qquad \mathcal{D}^J_\cap := \Big\{ \bigcap_{j=1}^J D_j : D_j \in \mathcal{D}\Big\}. \tag{3.2.3}$$

An element of $\mathcal{D}^J$ is a sequence of J domains, while an element of $\mathcal{D}^J_\cap$ is a domain of $\mathbb{R}^d$. We generalize $\mu(\cdot,\cdot)$ to $\mu^J(\cdot,\cdot)$ on $\mathcal{D}^J$ (resp. $\mathcal{D}^J_\cap$) by setting, for any D^1, D^2 in $\mathcal{D}^J$ (resp. $\mathcal{D}^J_\cap$),

$$\mu^J(D^1, D^2) := \sum_{j=1}^J \mu(D^1_j, D^2_j),$$

with obvious definitions of D^i_j. Since μ is a distance on $\widetilde{\mathcal{D}}$, μ^J defines also a distance on $\mathcal{D}^J$ (resp. $\mathcal{D}^J_\cap$). In what follows the continuity for a $\mathcal{D}^J$ or $\mathcal{D}^J_\cap$-valued process is meant with respect to $\mu^J(\cdot,\cdot)$.

For a domain $D \in \mathcal{D}^J_\cap$, the notation εD stands naturally as $\varepsilon D := \{y \in \mathbb{R}^d : y/\varepsilon \in D\}$ and similarly for $D \in \mathcal{D}^J$.

Class of random discretization grids

Fix some integer $J > 0$. We consider a $\mathcal{D}^J_\cap$-valued continuous $\mathcal{F}$-adapted process $(D_t)_{0\le t\le T}$ and a sequence of $\mathcal{D}^J_\cap$-valued continuous $\mathcal{F}$-adapted processes $\{(D^n_t)_{0\le t\le T} : n \ge 0\}$. All these domains of $\mathcal{D}^J_\cap$ are under the form

$$D^n_t := \bigcap_{j=1}^J D^n_{j,t}, \qquad D_t := \bigcap_{j=1}^J D_{j,t}.$$

Suppose that for some positive constants $r_0, \bar{r}_0$ the initial domain D_0 verifies

$$B_d(0, r_0) \subset D_0 \subset B_d(0, \bar{r}_0) \ \text{ a.s.} \tag{3.2.4}$$

We will assume the following approximation and continuity properties.

($\mathbf{H}_D^1$): There exists a constant $\eta_{\mathcal{D}} > 0$ such that

$$\sup_{n\geq 0}\left(\varepsilon_n^{-\eta_{\mathcal{D}}}\sup_{0\leq t\leq T}\mu^J(D_t^n, D_t)\right) < +\infty. \tag{3.2.5}$$

($\mathbf{H}_D^2$): There exists a continuous $\mathcal{F}$-adapted positive process $(L_t)_{0\leq t\leq T}$ such that L_0^{-1} is a bounded random variable and the following holds a.s. for all $t \in [0,T]$ and any $D \in \{D_{j,t}^n, D_{j,t},\ n \geq 0, j = 1,\dots,J\}$

1. the signed distance $\delta_{\partial D}(\cdot)$ is $\mathcal{C}^2$ on the set $\{x \in \mathbb{R}^d : |\delta_{\partial D}(x)| \leq L_t\}$;
2. we have $\sup_{x\in D}|x| \leq L_t^{-1}$ and

$$\inf_{x:|\delta_{\partial D}(x)|\leq L_t}|\nabla\delta_{\partial D}(x)| \geq \frac{1}{2}, \qquad \sup_{x:|\delta_{\partial D}(x)|\leq L_t}(|\nabla\delta_{\partial D}(x)| + \|\nabla^2\delta_{\partial D}(x)\|) \leq L_t^{-1}.$$

Remark 3.2.3. *Assumption ($\mathbf{H}_D^2$) is quite mild. Indeed, following [GT83, Lemma 14.16] for any $D \in \mathcal{D}$ there exists $L_D > 0$ such that the distance function (3.2.2) is $\mathcal{C}^2$ on the set $\{x \in \mathbb{R}^d : |\delta_{\partial D}(x)| \leq L_D\}$. Further, using that $\nabla\delta_{\partial D}(\cdot)$ restricted to ∂D is the inward unit vector at the boundary, the boundedness of D and ∂D, we get the existence of $L_D > 0$ such that, in addition, $\sup_{x\in D}|x| \leq L_D^{-1}$ and*

$$\inf_{x:|\delta_{\partial D}(x)|\leq L_D}|\nabla\delta_{\partial D}(x)| \geq \frac{1}{2}, \qquad \sup_{x:|\delta_{\partial D}(x)|\leq L_D}(|\nabla\delta_{\partial D}(x)| + \|\nabla^2\delta_{\partial D}(x)\|) \leq L_D^{-1}.$$

Therefore ($\mathbf{H}_D^2$) only requires some continuity and uniformity properties of L_D for the random domain-valued processes $D_{j,t}^n, D_{j,t},\ n \geq 0, j = 1,\dots,J$.

Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ supports an i.i.d. family of random variables $U := \{U_{n,i} : i, n \in \mathbb{N}\}$ with $U_{n,i} \sim \mathcal{U}(0,1)$, that are independent of $\mathcal{F}_T$. Define the filtration $\mathcal{F}_t^U := \mathcal{F}_t \vee \sigma(U)$. Let $G : (t,\omega,u) \in [0,T]\times\Omega\times[0,1] \mapsto \mathbb{R}^+ \cup \{+\infty\}$ be a $\mathcal{P}\otimes\mathcal{B}([0,1])$-measurable mapping, where $\mathcal{P}$ denotes the σ-field of predictable sets of $[0,T]\times\Omega$. In what follows, we will simply write $G_t(u)$.

Now we present the class of random discretization grids that constitutes the principal object of our analysis. Define a sequence of discretization grids $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ with $\mathcal{T}^n = \{\tau_i^n, i = 0,\dots,N_T^n\}$ given by

$$\begin{cases}\tau_0^n := 0,\\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin \varepsilon_n D_{\tau_{i-1}^n}^n\} \wedge (\tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}) \wedge T,\end{cases} \tag{3.2.6}$$

where $(\Delta_{n,i})_{n,i\in\mathbb{N}}$ is a family of random variables such that τ_i^n's are $\mathcal{F}^U$-stopping times and $\Delta_{n,i}$ is independent of $U_{m,j}$ for $m \neq n$ or $j > i$. The variables $\Delta_{n,i}$ play the role of error terms, we make an additional assumption on it later. We define the counting process $N_t^n := \#\{i \geq 1 : \tau_i^n \leq t\}$ for any $t \in [0,T]$, this is a càdlàg $\mathcal{F}^U$-adapted process.

Remark 3.2.4. *Note that $G_t(\cdot)$ may take the value of $+\infty$. However τ_i^n is always well defined since we take the minimum with the exit time in (3.2.6). In particular, if $G_t(\cdot) = +\infty$*

for all $t \in [0,T]$ we simply get a sequence of random grids given by exit times without exogenous source of randomness.

Define the normed vector space

$$\mathcal{H} := \left\{ u = (u_n, n \in \mathbb{N}) : u_n \in \mathbb{R},\ \|u\|_{\mathcal{H}} := \sum_{n\in\mathbb{N}} \frac{|u_n|}{2^n} < +\infty \right\},$$

and consider the $\mathcal{H}$-valued $\mathcal{F}^U$-adapted càdlàg process $Z_t := (Z_{n,t}, n \in \mathbb{N})$ on $[0,T]$ defined by

$$Z_{n,t} := \frac{N_t^n}{N_t^n + 1}, \quad n \in \mathbb{N}.$$

Let $(\bar{\mathcal{F}}_t)_{0\le t\le T}$ be the right-continuous extension of the filtration $(\mathcal{F}_t \vee \sigma(Z_r, r \le t))_{0\le t\le T}$. Since Z_t is $\mathcal{F}^U$-adapted and $\mathcal{F}^U$ is right-continuous, we naturally have

$$\mathcal{F}_t \subset \bar{\mathcal{F}}_t \subset \mathcal{F}_t^U. \tag{3.2.7}$$

Thus the filtration $\bar{\mathcal{F}}$ verifies the usual conditions. We also remark that the definition of Z_t implies that the $\mathcal{F}^U$-stopping times τ_i^n given by (3.2.6) are $\bar{\mathcal{F}}$-stopping times.

Suppose the following condition:

($\mathbf{H}_G$): 1. With probability 1, for all $u \in [0,1]$ the process $(G_t(u))_{0\le t\le T}$ is continuous on $\mathbb{R}^+ \cup \{+\infty\}$. Moreover there exists an $\mathcal{F}_T \otimes \mathcal{B}([0,1])$-measurable mapping $G_* : \Omega \times [0,1] \to \mathbb{R}^+$ not a.e. equal to zero, such that a.s. for all $n \ge 0$ and $1 \le i \le N_T^n$ we have

$$G_{\tau_{i-1}^n}(U_{n,i}) + \varepsilon_n^{-2}\Delta_{n,i} \ge G_*(U_{n,i}).$$

2. For some constant $\eta > 0$ and an $\bar{\mathcal{F}}$-adapted bounded process $(p_t)_{0\le t\le T}$ we have a.s. for all $n \ge 0$ and $1 \le i \le N_T^n$

$$\mathbb{E}\left(|\Delta_{n,i}||\bar{\mathcal{F}}_{\tau_{i-1}^n}\right) \le p_{\tau_{i-1}^n}\varepsilon_n^{2+\eta}. \tag{3.2.8}$$

The following lemma states certain important properties of the filtration $\bar{\mathcal{F}}$.

Lemma 3.2.5. *The following properties hold.*

(i) The $\mathcal{F}$-Brownian motion $(B_t)_{0\le t\le T}$ is also a $\bar{\mathcal{F}}$-Brownian motion. Moreover any $\mathcal{F}$-adapted continuous semimartingale has the same characteristics (finite variation part, local martingale part and quadratic variation) w.r.t. $\bar{\mathcal{F}}$.

(ii) For any $\bar{\mathcal{F}}_{\tau_{i-1}^n} \otimes \mathcal{B}([0,1])$-measurable mapping $f : \Omega \times [0,1] \to \mathbb{R}^+$ we have

$$\mathbb{E}(f(\omega,U_{n,i})|\bar{\mathcal{F}}_{\tau_{i-1}^n}) = \int_0^1 f(\omega,x)\mathrm{d}x.$$

Proof. Item (i). Observe that [Pro04, Theorem 2, Chap. VI] ensures that any $\mathcal{F}$-semimartingale remains a $\mathcal{F}^U$-semimartingale with the same characteristics. Now we extend this property

to the filtration $\bar{\mathcal{F}}$. For this, consider a square-integrable continuous $\mathcal{F}$-martingale M: using that it is a $\mathcal{F}^U$-martingale as recalled before, M is also a $\bar{\mathcal{F}}$-martingale in view of (3.2.7) and of the equality

$$\mathbb{E}(M_t|\bar{\mathcal{F}}_s) = \mathbb{E}(\mathbb{E}(M_t|\mathcal{F}_s^U)|\bar{\mathcal{F}}_s) = \mathbb{E}(M_s|\bar{\mathcal{F}}_s) = M_s.$$

In addition, M has the same quadratic variation $\langle M\rangle$ w.r.t. $\bar{\mathcal{F}}$ since it is characterized by the fact that $M^2 - \langle M\rangle$ is a martingale. The same conclusion can be extended to the case of local martingales since the localization times may be chosen as $\nu^k = \inf\{t \in [0,T] : \langle M\rangle_t \geq k\}$, which are $\bar{\mathcal{F}}$-stopping times, and thus by the previous argument each process $M_{\cdot\wedge\nu^k}$ is a $\bar{\mathcal{F}}$-martingale. Finally the property of having finite variation is independent of the filtration. *Item* (ii). It is sufficient to show that $U_{n,i}$ is independent of $\bar{\mathcal{F}}_{\tau_{i-1}^n}$. Indeed, $U_{n,i}$ is independent of $\mathcal{F}_T$ and of $(Z_{m,t})_{0\leq t\leq T}$ for $m \neq n$. Moreover, $N_{n,\cdot}$ is a counting process, thus its natural filtration (or equivalently that of $Z_{n,\cdot}$) is right-continuous (see [Pro04, Theorem 25, Chap. I]). So, it is enough to show that $U_{n,i}$ is independent of Z_{n,τ_{i-1}^n}. This follows from the construction (3.2.6) of the times τ_i^n and the properties of $\Delta_{n,i}$, in particular, since $U_{n,i}$ is completely unused up to the time τ_{i-1}^n, and no information about it is available at τ_{i-1}^n. □

In what follows by adapted process we mean $\bar{\mathcal{F}}$-adapted, for $\mathcal{F}$-adapted processes we will specify it explicitly if this property is needed. We also denote $\mathbb{E}_t(\cdot) := \mathbb{E}(\cdot|\bar{\mathcal{F}}_t)$.

Example: combination of hitting times and Poisson point process with general stochastic intensity

In this section we present the example of Poisson random times having general random path-dependent intensity and based on independent source of randomness (see [Str10] for an introduction to Poisson point processes), for which ($\mathbf{H}_G$) holds.

Let $(\lambda_t)_{0\leq t\leq T}$ be a strictly positive $\mathcal{F}$-adapted continuous stochastic process, playing the role of a stochastic intensity, and suppose that the following assumption holds.

($\mathbf{H}_\lambda$)**:** For some constant $\eta_\lambda \in (0,1]$ we have

$$|\lambda_t - \lambda_s| \leq C_\lambda |t-s|^{\eta_\lambda}, \quad 0 \leq s \leq t \leq T, \quad \text{a.s.}$$

and, in addition, $\mathbb{E}(C_\lambda \lambda_*^{-(2+\eta_\lambda)}) < +\infty$ where $\lambda_* := \inf_{0\leq t\leq T} \lambda_t$.

For a given trajectory of $(\lambda_t)_{0\leq t\leq T}$ define a sequence of independent Poisson point processes $(\mathcal{P}^n)_{n\geq 0}$, where for each $n \geq 0$ the process $\mathcal{P}^n$ has the intensity $\{\varepsilon_n^{-2}\lambda_t, t \in [0,T]\}$ and is based on the random noise $(U_{n,i})_{i\in\mathbb{N}}$ (see (3.2.11) below for a precise definition). Define a sequence of random discretization grids $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ with $\mathcal{T}^n = \{\tau_i^n, i = 0, \ldots, N_T^n\}$ as follows

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin \varepsilon_n D_{\tau_{i-1}^n}^n \text{ or } t \in \mathcal{P}^n\} \wedge T. \end{cases} \tag{3.2.9}$$

Then our claim is that $\mathcal{T}$ belongs to the class of grids described in Section 3.2.2, of the form

(3.2.6), and it satisfies to ($\mathbf{H}_G$). Indeed, let

$$G_t(u) := -\frac{1}{\lambda_t}\log(1-u), \tag{3.2.10}$$

which is the inverse c.d.f. of the exponential distribution with parameter λ_t. The next Poisson time $\tilde{\tau}_i^n$ after τ_{i-1}^n is defined by the equation

$$\varepsilon_n^{-2}\int_{\tau_{i-1}^n}^{\tilde{\tau}_i^n}\lambda_s \mathrm{d}s = -\log(1-U_{n,i}), \tag{3.2.11}$$

so that $\Delta_{n,i}$ is such that (in view of (3.2.6))

$$\tilde{\tau}_i^n = \tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}. \tag{3.2.12}$$

It readily follows that

$$G_{\tau_{i-1}^n}(U_{n,i})+\varepsilon_n^{-2}\Delta_{n,i} = \varepsilon_n^{-2}(\tilde{\tau}_i^n-\tau_{i-1}^n) \geq (\sup_{0\leq t\leq T}\lambda_t)^{-1}\varepsilon_n^{-2}\int_{\tau_{i-1}^n}^{\tilde{\tau}_i^n}\lambda_s \mathrm{d}s = (\sup_{0\leq t\leq T}\lambda_t)^{-1}|\log(1-U_{n,i})|.$$

We have completed the proof of ($\mathbf{H}_G$)-1.
Now, let us establish (3.2.8). Combining (3.2.10)-(3.2.11)-(3.2.12) and invoking Assumption ($\mathbf{H}_\lambda$), we obtain

$$|\Delta_{n,i}| = \left|\tilde{\tau}_i^n - \tau_{i-1}^n - \lambda_{\tau_{i-1}^n}^{-1}\int_{\tau_{i-1}^n}^{\tilde{\tau}_i^n}\lambda_s \mathrm{d}s\right| \leq \lambda_{\tau_{i-1}^n}^{-1}\left|\int_{\tau_{i-1}^n}^{\tilde{\tau}_i^n}|\lambda_s-\lambda_{\tau_{i-1}^n}|\mathrm{d}s\right| \leq \lambda_*^{-1}C_\lambda(\tilde{\tau}_i^n-\tau_{i-1}^n)^{1+\eta_\lambda}.$$

Further (3.2.11) yields

$$\tilde{\tau}_i^n - \tau_{i-1}^n \leq \lambda_*^{-1}\int_{\tau_{i-1}^n}^{\tilde{\tau}_i^n}\lambda_s \mathrm{d}s = \lambda_*^{-1}|\log(1-U_{n,i})|\varepsilon_n^2,$$

which finally implies

$$|\Delta_{n,i}| \leq C_\lambda\lambda_*^{-(2+\eta_\lambda)}|\log(1-U_{n,i})|^{1+\eta_\lambda}\varepsilon_n^{2+2\eta_\lambda}.$$

Using Lemma 3.2.5-(ii), we deduce that

$$\mathbb{E}_{\tau_{i-1}^n}(|\Delta_{n,i}|) \leq \left(\int_0^1|\log(1-x)|^{1+\eta_\lambda}\mathrm{d}x\right)\mathbb{E}_{\tau_{i-1}^n}\left(C_\lambda\lambda_*^{-(2+\eta_\lambda)}\right)\varepsilon_n^{2+2\eta_\lambda}.$$

The process $\mathbb{E}_t\left(C_\lambda\lambda_*^{-(2+\eta_\lambda)}\right) < +\infty$ is a martingale due to ($\mathbf{H}_\lambda$) and thus has a cádlág version, hence it is a.s. bounded. We have proved ($\mathbf{H}_G$)-2. All in all, ($\mathbf{H}_G$) holds in this general framework of Poisson point process with stochastic intensity. □

3.2.3 Main result

We are now in a position to state a functional CLT for a general multidimensional discretization error in the setting presented in the previous subsections. The CLT limit is defined in terms of the solution to the following matrix-valued quadratic equation.

Lemma 3.2.6 ([GL14a, Lemma 3.1]). *Let c be a $d \times d$-matrix symmetric non-negative real matrix. Then the equation*

$$2\,\mathrm{Tr}(x)x + 4x^2 = c \tag{3.2.13}$$

admits exactly one solution $x(c) \in \mathcal{S}_d^+$. Moreover, the mapping $c \mapsto x(c)$ is continuous.

Proof. We remark that in [GL14a, Lemma 3.1], the input matrix on the right hand side of (3.2.13) is $\tilde{c}^2$ instead of c here. Of course, it does not modify the existence and uniqueness properties in the form we state them here. Only the continuity property is questionable: in [GL14a, Lemma 3.1] the continuity of $\tilde{c} \mapsto x(\tilde{c}^2) = x(c)$ is proved. However one may easily deduce the continuity of $c \mapsto x(c)$ from their proof as well: indeed, this is a direct consequence of the representation [GL14a, eq. (A.7)] and of the fact that y_λ is continuous in $(\lambda_i^2)_{i=1}^d$ (in the notation of [GL14a, Section A.4]). □

Fix a random grid sequence $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ of the form (3.2.6). Define

$$\begin{aligned} &\varphi(t) := \max\{\tau \in \mathcal{T}^n : \tau \leq t\}, \qquad \bar{\varphi}(t) := \min\{\tau \in \mathcal{T}^n : \tau > t\},\ \bar{\varphi}(T) := T, \\ &\Delta X_t := X_t - X_{\varphi(t)}, \end{aligned} \tag{3.2.14}$$

where the dependence on n is omitted for the sake of simplicity.

Let $(\mathcal{M}_t)_{0\leq t\leq T}$ and $(\mathcal{A}_t)_{0\leq t\leq T}$ be adapted continuous processes with values in $\mathrm{Mat}_{m,d}$ and $\mathrm{Mat}_{d,d} \otimes \mathbb{R}^m$ respectively (recall that an element $\mathcal{A}_t \in \mathrm{Mat}_{d,d} \otimes \mathbb{R}^m$ is given by m real $d \times d$ matrices as $[\mathcal{A}_{1,t}, \ldots, \mathcal{A}_{m,t}]^\mathsf{T}$ for which we write $x^\mathsf{T}\mathcal{A}_t y := [x^\mathsf{T}\mathcal{A}_{1,t}y, \ldots, x^\mathsf{T}\mathcal{A}_{m,t}y]^\mathsf{T} \in \mathbb{R}^m$). Consider an $\mathbb{R}^m$-valued discretization error process given by

$$\mathcal{E}_t^n := \mathcal{E}_t^{n,1} + \mathcal{E}_t^{n,2}, \quad t \in [0,T],$$

with $\mathcal{E}_t^{n,1}$ and $\mathcal{E}_t^{n,2}$ of the form

$$\mathcal{E}_t^{n,1} := \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} \mathcal{M}_{\tau_{i-1}^n} \Delta S_s \mathrm{d}s, \qquad \mathcal{E}_t^{n,2} := \sum_{\tau_{i-1}^n < t} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} \Delta S_s^\mathsf{T} \mathcal{A}_{\tau_{i-1}^n} \mathrm{d}B_s. \tag{3.2.15}$$

Note that this is the most general form of an error term which is linear (or bi-linear) in terms of ΔS_s and $\mathrm{d}B_s$.

Now we introduce some processes that are involved in the explicit characterization of the limit distribution. Let W be a standard Brownian motion with $W_0 = 0$ and $U \sim \mathcal{U}(0,1)$ be independent of W, both independent of $\bar{\mathcal{F}}_T$. Set

$$\tau(t) := \inf\{s \geq 0 : \sigma_t W_s \notin D_t\} \wedge G_t(U),\ t \in [0,T].$$

In addition, for any measurable $f : \mathbb{R}^d \to \mathbb{R}$ define

$$\mathcal{B}_t[f(\cdot)] := \mathbb{E}_t\left(f(\sigma_t W_{\tau(t)})\right),\ t \in [0,T], \tag{3.2.16}$$

and

$$m_t := \mathbb{E}_t(\tau(t)),\ t \in [0,T]. \tag{3.2.17}$$

Define an $\mathbb{R}^d$-valued adapted continuous process $(Q_t)_{0\le t\le T}$ by

$$Q_t := \frac{1}{3} m_t^{-1} \begin{pmatrix} (\sigma_t\sigma_t^\mathsf{T})^{-1}_{11}\mathcal{B}_t[f(x) := (x^1)^3] \\ \vdots \\ (\sigma_t\sigma_t^\mathsf{T})^{-1}_{dd}\mathcal{B}_t[f(x) := (x^d)^3] \end{pmatrix}. \tag{3.2.18}$$

Denote $\mathcal{A}_t^\mathsf{T} := [\mathcal{A}_{1,t}^\mathsf{T}, \dots, \mathcal{A}_{m,t}^\mathsf{T}]^\mathsf{T}$ and $\mathcal{A}_t^{ij} := \frac{1}{2}(\mathcal{A}_{i,t}\mathcal{A}_{j,t}^\mathsf{T} + \mathcal{A}_{i,t}^\mathsf{T}\mathcal{A}_{j,t})$. Since $\mathcal{A}_t^{ij}$ is symmetric, by Lemma 3.B.1 we may write $\mathcal{A}_t^{ij} = \mathcal{A}_t^{ij+} - \mathcal{A}_t^{ij-}$, where $\mathcal{A}_t^{ij+}$ and $\mathcal{A}_t^{ij-}$ are continuous symmetric non-negative definite matrices. Define a $\mathrm{Mat}_{m,m}$-valued process $(\mathcal{K}_t)_{0\le t\le T}$ by

$$\mathcal{K}_t^{ij} := m_t^{-1}\mathcal{B}_t\left[f(x) := ((\sigma_t^{-1}x)^\mathsf{T} X_t^{ij+}(\sigma_t^{-1}x))^2 - ((\sigma_t^{-1}x)^\mathsf{T} X_t^{ij-}(\sigma_t^{-1}x))^2\right] - Q_t^\mathsf{T}\mathcal{A}_t^{ij}Q_t, \tag{3.2.19}$$

for all $1 \le i,j \le m$, where X_t^{ij+} (resp. X_t^{ij-}) is the solution of the matrix equation (3.2.13) for $c = \sigma_t^\mathsf{T}\mathcal{A}_t^{ij+}\sigma_t$ (resp. $\sigma_t^\mathsf{T}\mathcal{A}_t^{ij-}\sigma_t$).

Here is the main result of this chapter which provides the $\bar{\mathcal{F}}$-stable functional convergence of $(\sqrt{N_t^n}\mathcal{E}_t^n)_{0\le t\le T}$ in distribution as $n \to \infty$. For stable convergence, see [JS02, p. 512]-[JP12, Section 2.2.1.] for definition and properties.

Theorem 3.2.7. *Assume that S satisfies $(\mathbf{H}_S)$, $(\mathbf{H}_\Delta)$ and $\mathcal{T}$ is given by (3.2.6) and satisfies $(\mathbf{H}_D^1)$, $(\mathbf{H}_D^2)$ and $(\mathbf{H}_G)$. Assume that $\mathcal{M}_0$ and $\mathcal{A}_0$ are bounded random variables. Then the processes Q and $\mathcal{K}$ are adapted continuous and $\mathcal{K}_t \in \mathcal{S}_m^+$ a.s. for all $t \in [0,T]$. Denote $\mathcal{K}_t^{1/2}$ the matrix principal square root of $\mathcal{K}_t$. Then there exists an m-dimensional Brownian motion W defined on an extended probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and independent of B such that the following functional $\bar{\mathcal{F}}$-stable convergence in distribution holds:*

$$\sqrt{N_t^n}\mathcal{E}_t^n \underset{[0,T]}{\overset{d}{\Longrightarrow}} \sqrt{\int_0^t m_s^{-1}\mathrm{d}s}\left(\int_0^t \mathcal{M}_s Q_s \mathrm{d}s + \int_0^t Q_s^\mathsf{T}\mathcal{A}_s \mathrm{d}B_s + \int_0^t \mathcal{K}_s^{1/2}\mathrm{d}W_s\right). \tag{3.2.20}$$

3.2.4 Examples

Below we discuss several examples where the characteristics $m, Q, \mathcal{K}$ of the limit distribution (3.2.20) may be explicit or easily computable using only some basic numerical calculations. We consider a general process $(S_t)_{0\le t\le T}$ verifying $(\mathbf{H}_S)$, $(\mathbf{H}_\Delta)$ and sequence of domain-valued processes $(D_t^n)_{0\le t\le T}, n \ge 0$ verifying $(\mathbf{H}_D^1)$, $(\mathbf{H}_D^2)$, while we only specify explicitly the process $(D_t)_{0\le t\le T}$.

Case $d = 1$, hitting times of stochastic time-dependent barriers. First consider the case $d = 1$, $G_t(\cdot) \equiv +\infty$ and the domain-valued process $D_t := (-\alpha_t, \beta_t) \subset \mathbb{R}$ for some adapted continuous a.s. positive processes $(\alpha_t)_{0\le t\le T}$ and $(\beta_t)_{0\le t\le T}$. Recall that

$$\tau(t) := \inf\{r > 0 : \sigma_t W_r \notin (-\alpha_t, \beta_t)\}, \qquad \mathcal{B}_t[f(\cdot)] := \mathbb{E}_t\Big(f(\sigma_t W_{\tau(t)})\Big).$$

In this case the distribution of $\sigma_t W_{\tau(t)}$ is explicitly known: $\mathbb{P}_t(\sigma_t W_{\tau(t)} = -\alpha_t) = \frac{\beta_t}{\alpha_t+\beta_t}$ and $\mathbb{P}_t(\sigma_t W_{\tau(t)} = \beta_t) = \frac{\alpha_t}{\alpha_t+\beta_t}$, so that $\mathcal{B}_t[f(x) := x^k] = \frac{\alpha_t\beta_t^k + (-1)^k\beta_t\alpha_t^k}{\alpha_t+\beta_t}$. In particular, an easy calculation from (3.2.16) and (3.2.17) yields

$$m_t = \mathbb{E}_t(\tau(t)) = \mathbb{E}_t((W_{\tau(t)})^2) = \alpha_t\beta_t\sigma_t^{-2}, \qquad Q_t = \frac{1}{3}m_t^{-1}\sigma_t^{-2}\mathcal{B}_t[f(x) := x^3] = \frac{1}{3}(\beta_t - \alpha_t).$$

To calculate $\mathcal{K}_t$ we remark that $\mathcal{A}_t^{11+} = (\mathcal{A}_t)^2$, $\mathcal{A}_t^{11-} = 0$ and thus $(X_t^{11+})^2 = \frac{1}{6}\sigma_t^2(\mathcal{A}_t)^2$. This further implies

$$\mathcal{K}_t = m_t^{-1}\frac{1}{6}\sigma_t^2(\mathcal{A}_t)^2\sigma_t^{-4}\mathcal{B}_t[f(x) := x^4] - Q_t^2(\mathcal{A}_t)^2 = \frac{(\mathcal{A}_t)^2}{18}(\alpha_t^2 + \beta_t^2 + \alpha_t\beta_t).$$

So finally we get

$$\begin{aligned}\sqrt{N_t^n}\mathcal{E}_t^n \underset{[0,T]}{\overset{d}{\Longrightarrow}} \frac{1}{3}\sqrt{\int_0^t \frac{\sigma_s^2}{\alpha_s\beta_s}\mathrm{d}s}\Big(\int_0^t \mathcal{M}_s(\beta_s - \alpha_s)\mathrm{d}s + \int_0^t (\beta_s - \alpha_s)\mathcal{A}_s\mathrm{d}B_s \\ + \frac{1}{\sqrt{2}}\int_0^t \mathcal{A}_s\sqrt{\alpha_s^2 + \beta_s^2 + \alpha_s\beta_s}\mathrm{d}W_s\Big).\end{aligned} \tag{3.2.21}$$

From (3.2.21) we can easily deduce the result of [Fuk10, Theorem 3.1] (for $\varphi(x) = x$; the general case may be easily deduce by applying $\varphi^{-1}(\cdot)$ to S_t) which studies a particular case of $\alpha_t = \beta_t = 1$ and considers the estimation of integrated variance (see Section 3.1), so that $\mathcal{A}_t = 2\sigma_t$. In this case, invoking Theorem 3.3.1 yields

$$\varepsilon_n^{-1}\mathcal{E}_t^n \underset{[0,T]}{\overset{d}{\Longrightarrow}} \int_0^t \mathcal{K}_s^{1/2}\mathrm{d}W_s$$

where $\mathcal{K}_t = \frac{2\sigma_t^2}{3}$, and Theorem 3.4.3 justifies that

$$\varepsilon_n^{-2}\sum_{\tau_{i-1}^n < T} |\Delta S_{\tau_i^n}|^4 \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \int_0^T \sigma_t^2\mathrm{d}t,$$

which, all in all, coincide with the results in [Fuk10, Theorem 3.1]. Theorem 3.2.7 uses the normalization $\sqrt{N_t^n}$, which is somewhat more natural for a CLT, and it writes

$$\sqrt{N_t^n}\mathcal{E}_t^n \underset{[0,T]}{\overset{d}{\Longrightarrow}} \sqrt{\frac{2}{3}\int_0^t \sigma_s^2\mathrm{d}s}\int_0^t \sigma_s\mathrm{d}W_s.$$

Note that our work provides tractable limit distribution characterization in a more general setting than [Fuk10] in terms of the discretization times, the shape of the error terms; furthermore it covers the multidimensional case.

Now suppose that $G_t(\cdot)$ is not always $+\infty$. Let T_0 be deterministic and τ be the first exit time of σW from an interval $[-\alpha, \beta]$. Thus the distribution of $W_{\tau \wedge T_0}$ is equal to

$$\mathbb{P}(\tau \leq T_0, \sigma W_\tau = -\alpha)\delta_{-\alpha}(\mathrm{d}x) + k(x)\mathbb{1}_{[-\alpha,\beta]}(x)\mathrm{d}x + \mathbb{P}(\tau \leq T_0, \sigma W_\tau = \beta)\delta_\beta(\mathrm{d}x),$$

where, following [RY99, p.111, Exercise 3.15], $k(x)$ equals

$$\frac{1}{\sqrt{2\pi T_0}\sigma} \sum_{k=-\infty}^{+\infty} \left\{ \exp\left(-\frac{1}{2T_0\sigma^2}(x + 2k(\alpha+\beta))^2 \right) - \exp\left(-\frac{1}{2T_0\sigma^2}(x - 2\beta + 2k(\alpha+\beta))^2 \right) \right\},$$

and, from [BS02, p.212, formulas 3.0.6],

$$\mathbb{P}(\tau \leq T_0, \sigma W_\tau = -\alpha) = \int_0^{\sigma^2 T_0} \mathrm{ss}_s(\beta, \alpha+\beta)\mathrm{d}s,$$

$$\mathbb{P}(\tau \leq T_0, \sigma W_\tau = \beta) = \int_0^{\sigma^2 T_0} \mathrm{ss}_s(\alpha, \alpha+\beta)\mathrm{d}s$$

for $\mathrm{ss}_t(\cdot,\cdot)$ given under an explicit form in [BS02, p.641].

Let $\mathcal{N}(\alpha, \beta, \mu, \sigma^2, p) := \int_{-\alpha}^{\beta} x^p p_{\mu,\sigma}(x)\mathrm{d}x$, where $p_{\mu,\sigma}(x) := (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Note that the explicit value of $\mathcal{N}(\alpha, \beta, \mu, \sigma^2, p)$ in terms of the standard Gaussian c.d.f. maybe easily deduced (recursively in p) via integration by parts. Further define

$$M_p(\alpha, \beta, \sigma, T_0) := \sum_{k=-\infty}^{+\infty} \left\{ \mathcal{N}(\alpha, \beta, -2k(\alpha+\beta), T_0\sigma^2, p) - \mathcal{N}(\alpha, \beta, 2\beta - 2k(\alpha+\beta), T_0\sigma^2, p) \right\}.$$

Note that in practice $M_p(\alpha, \beta, \sigma, T_0)$ is well approximated by a finite sum due to the fast decay of e^{-x^2}. Now a simple calculation yields that $\mathcal{B}_t[f(x) := x^p]$ equals

$$\int_0^1 \sigma_t^p \left(M_p(\alpha_t, \beta_t, \sigma_t, G_t(u)) + \int_0^{\sigma_t^2 G_t(u)} ((-\alpha_t)^p \,\mathrm{ss}_s(\beta_t, \alpha_t+\beta_t) + \beta_t^p \,\mathrm{ss}_s(\alpha_t, \alpha_t+\beta_t))\mathrm{d}s \right) \mathrm{d}u,$$

which allows to easily deduce the explicit form of the limit distribution in (3.2.20) through the computations of $m, Q, \mathcal{K}$ (at least, using a numerical integration routine).

Case $d > 1$, hitting times of symmetric domains, ellipsoid based grids. Suppose that for all $t \in [0,T]$ the domain D_t is symmetric (i.e. $D_t = -D_t$), denote $\tau(t) = \inf\{r > 0 : \sigma_t W_r \notin D_t\} \wedge G_t(U)$. Let us prove that $Q_t = 0$. Indeed, in view of (3.2.18), this follows from

$$\mathbb{E}_t((W^i_{\tau(D_t)\wedge T})^3) = \mathbb{E}_t((-W^i_{\tau(-D_t)\wedge T})^3) = \mathbb{E}_t((-W^i_{\tau(D_t)\wedge T})^3) = -\mathbb{E}_t((W^i_{\tau(D_t)\wedge T})^3),$$

where we denote $\tau(D)$ the first exist time of $\sigma_t W$ from a domain D, and $T > 0$ is fixed.

We suppose again that $G_t(\cdot) \equiv +\infty$. Consider the case $d > 1$. For an $\mathcal{S}_d^{++}$-valued process $(\Sigma_t)_{0\le t\le T}$ we take $D_t = \{x \in \mathbb{R}^d : x^\mathsf{T}\Sigma_t x \le 1\}$. Hence

$$\tau(t) = \inf\{r > 0 : W_r^\mathsf{T}(\sigma_t^\mathsf{T}\Sigma_t\sigma_t)W_r \ge 1\}.$$

Let $\sigma_t^\mathsf{T}\Sigma_t\sigma_t = U_t^\mathsf{T}\Lambda_t U_t$ where U_t is orthogonal and Λ_t is diagonal. Then $\tau(t)$ is equal in distribution to $\inf\{r > 0 : W_r^\mathsf{T}\Lambda_t W_r \ge 1\}$. To characterize explicitly the limit distribution (conditionally on σ_t) in (3.2.20), it is enough to calculate $\mathcal{K}_t$ (since $Q_t = 0$), which requires only the calculation of $\mathbb{E}_t(\tau(t))$ and $\mathbb{E}_t\left(\prod_{i=1}^d (W^i_{\tau(t)})^{k_i}\right)$ for $k_1 + \cdots + k_d = 4, k_i \ge 0$.

In the case $d = 2$ we need only to calculate numerically the following 3 functions

$$f_1(\lambda) := \mathbb{E}((W^1_{\tau(\lambda)})^4), \quad f_2(\lambda) := \mathbb{E}((W^1_{\tau(\lambda)}W^2_{\tau(\lambda)})^2), \quad f_3(\lambda) := \mathbb{E}((W^1_{\tau(\lambda)})^3 W^2_{\tau(\lambda)}),$$

where $\tau(\lambda) := \inf\{r > 0 : (W^1_r)^2 + \lambda(W^2_r)^2 \ge 1\}$ for $\lambda > 0$ (other calculations follow from setting $\lambda \mapsto \frac{1}{\lambda}$ and using basic scaling properties). To treat the case with general $G_t(\cdot)$ it is enough to numerically calculate the following 3 functions in 2 parameters

$$\begin{aligned} f_1(\lambda, T_0) &:= \mathbb{E}((W^1_{\tau(\lambda)\wedge T_0})^4), \qquad f_2(\lambda, T_0) := \mathbb{E}((W^1_{\tau(\lambda)\wedge T_0}W^2_{\tau(\lambda)\wedge T_0})^2), \\ f_3(\lambda, T_0) &:= \mathbb{E}((W^1_{\tau(\lambda)\wedge T_0})^3 W^2_{\tau(\lambda)\wedge T_0}). \end{aligned}$$

To the best of our knowledge, explicit formulas for these functions are not available and we have to resort to numerical methods like Monte Carlo methods. For related efficient schemes, see the boundary shifting scheme of [GM10], the walk on moving spheres algorithm of [DH13].

3.3 Proof of the main result (Theorem 3.2.7)

This is based on two general results: first, a CLT (Section 3.3.1) for discretization errors in an abstract setting; second, general properties of exit times from intersection of regular domains (Section 3.3.2). The proof of Theorem 3.2.7 is then completed in Section 3.3.3.

3.3.1 A general CLT

The result of this section is the key ingredient of the proof of Theorem 3.2.7 and constitutes itself a stand-alone contribution. In particular, it generalizes the result of [Fuk11b] in our framework of multidimensional process and general multidimensional error term, with explicit limit coefficients (as opposed to the non-explicit Condition 2.3 of [Fuk11b]).

Within Section 3.3.1 (and Section 3.4 for the proofs) we are working in a slightly more abstract framework regarding S than in Section 3.2. Let $(\Omega, \mathcal{F}, (\bar{\mathcal{F}}_t)_{0\le t\le T}, \mathbb{P})$ be a filtered probability space (with $(\bar{\mathcal{F}}_t)_{0\le t\le T}$ satisfying the usual conditions) and consider a more general semimartingale S satisfying the following extended assumption.

($\mathbf{H}_S^{gen.}$)**:** The process S on $[0, T]$ is given by

$$S_t = A_t + \int_0^t \sigma_s \mathrm{d}B_s, \quad t \in [0, T],$$

where

- the process A is continuous, adapted and of finite variation, and satisfies

$$|A_t - A_s| \leq C_A |t-s|^{\eta_A} \qquad \forall s,t \in [0,T] \quad \text{a.s.}, \tag{3.3.1}$$

  for a random variable C_A, a.s. finite, and a parameter $\eta_A \in (1/2, 1]$;

- $(\sigma_t)_{0\leq t\leq T}$ is a continuous adapted $\mathrm{Mat}_{d,d}$-valued process, such that σ_t is invertible a.s. for all $t \in [0,T]$ and σ_0, σ_0^{-1} are bounded random variables;

- for some a.s. finite random variable $C_\sigma > 0$ and a parameter $\eta_\sigma \in (0,1]$, we have

$$|\sigma_t - \sigma_s| \leq C_\sigma |t-s|^{\eta_\sigma/2} \qquad \forall s,t \in [0,T] \quad \text{a.s.}$$

Let $\mathcal{T} = \{\mathcal{T}^n : n \geq 0\}$ be a sequence of discretization grids made of stopping times, where $\mathcal{T}^n = \{\tau_i^n, i = 0, \ldots, N_T^n\}$. We introduce two assumptions, whose formulation depends on the choice of a particular sequence $(\varepsilon_n)_{n\geq 0}$. For the subsequent CLT, we consider $\varepsilon_n \to 0$; with loss of generality, we assume $\varepsilon_n \leq 1$ for any n.

($\mathbf{H}_R$): 1. There exists an adapted continuous non-decreasing process $(C_t^{(3.3.2)})_{0\leq t\leq T}$ with bounded $C_0^{(3.3.2)}$, such that for $\alpha \in \{2,3,4\}$ and for all $n \geq 0$ and $1 \leq i \leq N_T^n$

$$\sup_{\tau_{i-1}^n < t \leq T} \left(\mathbb{E}_t(|S_{\tau_i^n} - S_{\tau_{i-1}^n}|^\alpha) + |S_{t\wedge\tau_i^n} - S_{\tau_{i-1}^n}|^\alpha \right) \leq C_{\tau_{i-1}^n}^{(3.3.2)} \varepsilon_n^\alpha \tag{3.3.2}$$

   where $\mathbb{E}_t(.) := \mathbb{E}(. \mid \bar{\mathcal{F}}_t)$.

2. The following non-negative random variable is a.s. finite:

$$C_{(3.3.3)} := \sup_{n\geq 0} \left(\varepsilon_n^2 N_T^n \right) < +\infty. \tag{3.3.3}$$

Observe that it is enough to verify (3.3.2) with $\alpha = 4$, by invoking the non-expansion property of (conditional) L^p-norms.

For $\alpha \in \mathbb{N}$ we denote by $\mathcal{P}^\alpha$ the vector space spanned by α-homogeneous polynomial functions $f : \mathbb{R}^d \to \mathbb{R}$. The next set of assumptions is related to the mapping $\mathcal{B}_t[\cdot]$ arising in (3.2.16) in our applications. Since we deal here with a more general setting, we state a more general assumption.

($\mathbf{H}_\mathcal{B}$): 1. There is a linear operator $\mathcal{B}_.[.]$ from the vector space spanned by $\mathcal{P}^\alpha, \alpha = 2,3,4$, into scalar adapted continuous process $(\mathcal{B}_t[f(\cdot)])_{0\leq t\leq T}$, such that the random variable $\mathcal{B}_0[f]$ is bounded for any such f.

2. The $\mathbb{R}$-valued process $m_t := \dfrac{\mathcal{B}_t[f(x) := |x|^2]}{\mathrm{Tr}(\sigma_t \sigma_t^\mathsf{T})}$ is strictly positive and such that m_0^{-1} is bounded.

3. There exists a function $g : [0,1] \to \mathbb{R}_+$ with $\lim_{\varepsilon\to 0}(g(\varepsilon) + \varepsilon^{2(1-\rho)} g(\varepsilon)^{-1}) = 0$ for some $\rho \in (0,1)$, such that for any $f \in \mathcal{P}^\alpha$ with $\alpha \in \{2,3,4\}$ we have, for some a.s. finite

random variable $C_{(3.3.4)}$ and a parameter $\eta \in (0,1]$, that

$$\sup_{\tau^n_{i-1}<(T-g(\varepsilon_n))_+} \left|\varepsilon_n^{-\alpha}\mathbb{E}_{\tau^n_{i-1}}(f(S_{\tau^n_i}-S_{\tau^n_{i-1}}))-\mathcal{B}_{\tau^n_{i-1}}[f(\cdot)]\right| \leq C_{(3.3.4)}\varepsilon_n^{\eta} \tag{3.3.4}$$

for all $n \geq 0$ a.s.

4. We have $\varepsilon_n^{-2}\#\{\tau^n_i : (T-g(\varepsilon_n))_+ \leq \tau^n_i \leq T\} \xrightarrow[n\to+\infty]{\text{a.s.}} 0$.

The assumption $(\mathbf{H}_{\mathcal{B}})$ imposes consistency on the distribution of the discretization grids for various n and specifies a "scaling" property for the grid sequence as $n \to +\infty$. At first sight it looks like similar to [Fuk11b, Condition 2.3], but as we see in Section 3.3.3, it is quite tractable. Moreover, we remark that [Fuk11b, Condition 2.3] involves higher moments (up to 12, as opposed to 4 in our work) and is stated for moment ratios which makes the generalization to the multidimensional case and the practical verification of this condition much harder.

We adopt some of the notation from Section 3.2.3 but with the general notion of $\mathcal{B}_t[f(\cdot)]$ and m_t in $(\mathbf{H}_{\mathcal{B}})$ instead of (3.2.16) and (3.2.17), and for a general sequence of discretization grids $\mathcal{T}$. In particular, we similarly denote $\varphi(t), \bar{\varphi}(t)$ and ΔX_t (for any process X_t) as in (3.2.14).

We consider an $\mathbb{R}^m$-valued discretization error process $\mathcal{E}^n_t := \mathcal{E}^{n,1}_t + \mathcal{E}^{n,2}_t$ with $\mathcal{E}^{n,1}_t$ and $\mathcal{E}^{n,2}_t$ given by (3.2.15). The processes $(Q_t)_{0\leq t\leq T}$ and $(\mathcal{K}_t)_{0\leq t\leq T}$ are derived from m_t and $\mathcal{B}_t[f(\cdot)]$ in the same way as in (3.2.18) and (3.2.19). Here is a general result which provides the $\bar{\mathcal{F}}$-stable functional convergence of $(\sqrt{N^n_t}\mathcal{E}^n_t)_{0\leq t\leq T}$ in distribution.

Theorem 3.3.1. *Assume that S satisfies $(\mathbf{H}^{gen.}_S)$ and consider a sequence of discretization grids $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ with $\mathcal{T}^n = \{\tau^n_i, i = 0,\dots,N^n_T\}$. Assume that S and $\mathcal{T}$ are such that, there is a positive sequence ε_n with $\varepsilon_n \to 0$, such that for any subsequence $(\varepsilon_{\iota(n)})_{n\geq 0}$ there exists another subsequence $(\varepsilon_{\iota'\circ\iota(n)})_{n\geq 0}$ for which $(\mathbf{H}_R)$ and $(\mathbf{H}_{\mathcal{B}})$ hold (for this sub-subsequence). Suppose that $\mathcal{M}_0$ and $\mathcal{A}_0$ are bounded random variables.*
Then there exists an m-dimensional Brownian motion W defined on an extended probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and independent of $\bar{\mathcal{F}}_T$ such that the following convergences hold:

1. *the functional $\bar{\mathcal{F}}$-stable convergence in distribution*

$$\varepsilon_n^{-1}\mathcal{E}^n_t \xrightarrow[[0,T]]{d} \left(\int_0^t \mathcal{M}_s Q_s \mathrm{d}s + \int_0^t Q_s^{\intercal}\mathcal{A}_s \mathrm{d}B_s + \int_0^t \mathcal{K}_s^{1/2}\mathrm{d}W_s\right);$$

2. *the uniform convergence in probability*

$$\varepsilon_n^2 N^n_t \xrightarrow[n\to+\infty]{u.c.p.} \int_0^t m_s^{-1}\mathrm{d}s. \tag{3.3.5}$$

As a consequence, this justifies the convergence in distribution for $(\sqrt{N^n_t}\mathcal{E}^n_t : 0 \leq t \leq T)$ in the functional sense (see [JP12, p.45]). The proof will be given in Section 3.4.

3.3.2 Properties of exit times from domain

Let B be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, (\bar{\mathcal{F}}_t)_{t\geq 0}, \mathbb{P})$, with filtration satisfying the usual assumptions of being right-continuous and $\mathbb{P}$-complete. In this section we present some general properties of domain exit times for d-dimensional continuous Itô semimartingales $(S_t)_{0\leq t\leq T}$ and $(\bar{S}_t)_{0\leq t\leq T}$ of the form

$$S_t = \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}B_s, \quad \bar{S}_t = \sigma_0 B_t, \quad t \geq 0, \tag{3.3.6}$$

where $(b_t)_{t\geq 0}$ and $(\sigma_t)_{t\geq 0}$ are respectively $\mathbb{R}^d$-valued and $\mathrm{Mat}_{d,d}$-valued $\bar{\mathcal{F}}$-adapted stochastic processes, satisfying some assumptions presented below. Here the starting point is $S_0 = 0$, for the sake of simplicity; actually, this is enough for our analysis, since the stopping times under study are essentially defined regarding the increments of S, extensions to $S_0 \neq 0$ would be straightforward. The subsequent results (Lemma 3.3.2, Propositions 3.3.4 and 3.3.5) play a key role in the proof of the CLT (Theorem 3.2.7, which proof is provided in Section 3.3.3).

($\mathbf{H}_{loc}^{D,\sigma}$)**:** The following assumptions hold.

i) Let $J \geq 1$ and $D \in \mathcal{D}_\cap^J$ (i.e. $D = \cap_{j=1}^J D_j$ for some $D_j \in \mathcal{D}$). Define the functions $\delta_{\partial D_j} : \mathbb{R}^d \to \mathbb{R}$ which are the signed distances to ∂D_j (defined in (3.2.2)). Set $L_D > 0$ such that for all j we have $\delta_{\partial D_j}(\cdot) \in \mathcal{C}^2$ on $\{x : |\delta_{\partial D_j}(x)| \leq L_D\}$ and

$$\begin{aligned} &\sup_{x\in D_j} |x| \leq L_D^{-1}, \qquad \inf_{x:|\delta_{\partial D_j}(x)|\leq L_D} |\nabla \delta_{\partial D_j}(x)| \geq \frac{1}{2}, \\ &\sup_{x:|\delta_{\partial D_j}(x)|\leq L_D} (|\nabla \delta_{\partial D_j}(x)| + \|\nabla^2 \delta_{\partial D_j}(x)\|) \leq L_D^{-1}. \end{aligned} \tag{3.3.7}$$

ii) The $\mathrm{Mat}_{d,d}$-valued process $(\sigma_t)_{0\leq t\leq T}$ is adapted continuous, such that for all $t \geq 0$ the matrix σ_t is invertible and

$$|\sigma_t - \sigma_0| \leq C_\sigma t^{\eta_\sigma/2}, \quad \forall t \in [0, T] \quad \text{a.s.}$$

for some $\eta_\sigma > 0$ and some random variable $C_\sigma > 0$ satisfying $m_\sigma := \mathbb{E}\left(C_\sigma^4\right) < +\infty$. In addition, there exist strictly positive and finite constants $\Lambda_{\min}^\sigma, \Lambda_{\max}^\sigma, b_{\max}$ such that

$$\Lambda_{\min}^\sigma \leq \inf_{t\in[0,\tau_0]} \lambda_{\min}(\sigma_t \sigma_t^\mathsf{T}) \leq \sup_{t\in[0,\tau_0]} \|\sigma_t \sigma_t^\mathsf{T}\| \leq \Lambda_{\max}^\sigma, \qquad \sup_{t\in[0,\tau_0]} |b_t| \leq b_{\max}, \tag{3.3.8}$$

where we denote $\tau_0 := \inf\{t \geq 0 : S_t \notin D\}$.

Let $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ be an α-homogeneous function for some $\alpha \geq 2$. It is easy to check that for some constant C_f we have for all $x \in \mathbb{R}^d$

$$|f(x)| \leq C_f |x|^\alpha, \qquad |\nabla f(x)| \leq C_f |x|^{\alpha-1}, \qquad \|\nabla^2 f(x)\| \leq C_f |x|^{\alpha-2}. \tag{3.3.9}$$

In what follows, we fix the parameters $L_D, \eta_\sigma, m_\sigma, \Lambda_{\min}^\sigma, \Lambda_{\max}^\sigma, b_{\max}, C_f$ that are specified by the model. The following notation is quite convenient for the subsequent analysis, it will be

repeatedly used.

Notation 1. *Let $\mathfrak{S}$ be a set of variables. We denote by $\mathcal{C}(\mathfrak{S})$ the set of strictly positive and continuous functions of the variables of $\mathfrak{S}$.*

Remark that such a set $\mathcal{C}(\mathfrak{S})$ is closed under addition, multiplication and all usual operations we may perform in the following analysis.

Let us fix $\mathfrak{S} := \{L_D, \eta_\sigma, m_\sigma, \Lambda^\sigma_{\min}, \Lambda^\sigma_{\max}, b_{\max}, C_f\}$. For the elements of $\mathcal{C}(\mathfrak{S})$ we will omit the dependence on the arguments, the value of a function in $\mathcal{C}(\mathfrak{S})$ is by default assumed to be equal to the value on the parameters fixed above.

Now we state the main results of this section (proofs postponed to Section 3.5). The next lemma is a simple technical result.

Lemma 3.3.2. *Assume $(\mathbf{H}^{D,\sigma}_{loc})$. For any $\varepsilon \in (0,1]$ any stopping times $\nu_1, \nu_2 \in [0,\tau]$, with $\tau := \inf\{t \geq 0 : S_t \notin \varepsilon D\}$, we have*

$$|\mathbb{E}(f(S_{\nu_1}) - f(S_{\nu_2}))| \leq C_f(b_{\max} L_D^{-(\alpha-1)} + \frac{1}{2}\sqrt{d}\Lambda^\sigma_{\max} L_D^{-(\alpha-2)})\varepsilon^{\alpha-2}\mathbb{E}\left(|\nu_1 - \nu_2|\right).$$

Proof. Using the Itô formula, the inequality $|\operatorname{Tr}(M)| \leq \sqrt{d}\|M\|$ for any $M \in \mathrm{Mat}_{m,d}$, the sub-multiplicativity of the Frobenius norm, and since $\varepsilon \leq 1$, we obtain

$$\begin{aligned}|\mathbb{E}\left(f(S_{\nu_1}) - f(S_{\nu_2})\right)| &\leq \left|\mathbb{E}\left(\int_{\nu_1}^{\nu_2} [\nabla f(S_t) b_t + \frac{1}{2}\operatorname{Tr}(\sigma_t^\intercal \nabla^2 f(S_t)\sigma_t)]\,\mathrm{d}t\right)\right| \\ &\leq C_f(b_{\max} L_D^{-(\alpha-1)} + \frac{1}{2}\sqrt{d}\Lambda^\sigma_{\max} L_D^{-(\alpha-2)})\varepsilon^{\alpha-2}\mathbb{E}\left(|\nu_1 - \nu_2|\right).\end{aligned}$$

□

The next results state some important properties of domain exit times, their proofs are postponed to Section 3.5.2. These results are interesting on their own.

Lemma 3.3.3. *Assume $(\mathbf{H}^{D,\sigma}_{loc})$ with $D \in \mathcal{D}$ $(J = 1)$. There exists $R_D \in \mathcal{C}(\mathfrak{S})$ such that, for any $\varepsilon \in (0,1]$, $\tau = \inf\{t \geq 0 : S_t \notin \varepsilon D\}$ and any stopping time ν, the following holds:*

i) for any $p \in \mathbb{N}^*$, a.s. *on the event $\{\nu \leq \tau\}$ we have $\mathbb{E}_\nu\left((\tau - \nu)^p\right) \leq p!\ (R_D\varepsilon^2)^p$;*

ii) for any $c \geq 0$, a.s. *on the event $\{\nu \leq \tau\}$ we have* a.s. $\mathbb{P}_\nu(\tau - \nu \geq \varepsilon^2 c) \leq 2e^{-\frac{c}{2R_D}}$.

The next proposition estimates the weak error between the exit values for S and $\bar{S}$.

Proposition 3.3.4. *Assume $(\mathbf{H}^{D,\sigma}_{loc})$ and let $f \in \mathcal{C}(\mathbb{R}^d, \mathbb{R})$ be an α-homogeneous function with $\alpha \in \{2,3,4\}$. There exists $K \in \mathcal{C}(\mathfrak{S})$ such that for any $\varepsilon \in (0,1]$, the stopping times*

$$\tau = \inf\{t \geq 0 : S_t \notin \varepsilon D\} \quad \text{and} \quad \bar{\tau} = \inf\{t \geq 0 : \bar{S}_t \notin \varepsilon D\}$$

satisfy, for any $T > 0$,

$$\varepsilon^{-\alpha}\left|\mathbb{E}(f(S_{\tau\wedge T}) - f(\bar{S}_{\bar{\tau}\wedge T}))\right| \leq K\varepsilon^{\eta_\sigma}. \tag{3.3.10}$$

The next result gives the estimation of the weak error between the exit values of S from two domains that are close to each other.

Proposition 3.3.5. *Assume* $(\mathbf{H}_{loc}^{D,\sigma})$ *and let* $f \in \mathcal{C}(\mathbb{R}^d, \mathbb{R})$ *be an* α*-homogeneous function with* $\alpha \in \{2,3,4\}$*. There exists* $K \in \mathcal{C}(\mathfrak{S} \cup \{K'\})$ *such that for any* $\varepsilon \in (0,1]$*, any strictly positive constants* K', η' *and any* $D' \in \mathcal{D}_\cap^J$ *such that* $\mu^J(D,D') \le K'\varepsilon^{\eta'}$*, and for which* (3.3.7) *and* $(\mathbf{H}_{loc}^{D,\sigma})$ *hold for* D' *instead of* D *with the same constants* $L_D, \Lambda^\sigma_{\min}, \Lambda^\sigma_{\max}, b_{\max}$*, we have*

$$\varepsilon^{-\alpha} \left|\mathbb{E}(f(S_{\tau\wedge T}) - f(S_{\tau'\wedge T}))\right| \le K\varepsilon^{\eta'},$$

for all $T > 0$*, where*

$$\tau = \inf\{t \ge 0 : S_t \notin \varepsilon D\}, \quad \tau' = \inf\{t \ge 0 : S_t \notin \varepsilon D'\}.$$

3.3.3 Completion of the proof of Theorem 3.2.7

We come back to the setting of Section 3.2.3. Our strategy is to apply the general CLT stated in Theorem 3.3.1. In particular, we aim at checking $(\mathbf{H}_R)$ and $(\mathbf{H}_\mathcal{B})$ for the $\mathcal{B}_t[\cdot]$ given by (3.2.16) for any ε_n satisfying $\sum_{n\ge0}\varepsilon_n^2 < +\infty$. For a general sequence $\varepsilon_n \to 0$ the result will follow in view of the subsequence formulation of Theorem 3.3.1: it is enough to verify the assumptions for some subsequence $\varepsilon_{\iota'\circ\iota(n)}$ (that may be chosen square summable) of arbitrary subsequence $\varepsilon_{\iota(n)}$ of ε_n.

Let us prove $(\mathbf{H}_R)$-1. Recall that we denote $\mathbb{E}_t(\cdot) := \mathbb{E}(\cdot|\bar{\mathcal{F}}_t)$. From the definition of $\mathcal{T}$ in (3.2.6), we have by $(\mathbf{H}_D^2)$ that for all $n \ge 0$ and $1 \le i \le N_T^n$

$$\sup_{\tau_{i-1}^n < t \le T} \left(\mathbb{E}_t(|S_{\tau_i^n} - S_{\tau_{i-1}^n}|^\alpha) + |S_{t\wedge\tau_i^n} - S_{\tau_{i-1}^n}|^\alpha\right) \le 2\left(\sup_{0\le s\le\tau_{i-1}^n, \alpha\in\{2,3,4\}} L_s^{-\alpha}\right)\varepsilon_n^\alpha,$$

which shows $(\mathbf{H}_R)$-1 with $C_t^{(3.3.2)} := 2\left(\sup_{0\le s\le t,\alpha\in\{2,3,4\}} L_s^{-\alpha}\right)$, so that by $(\mathbf{H}_D^2)$ the process $C^{(3.3.2)}$ is continuous and $C_0^{(3.3.2)}$ is bounded.

The verification of the assumptions $(\mathbf{H}_R)$-2 and $(\mathbf{H}_\mathcal{B})$-4 is technical, and it relies on the next Lemma, which is proved in Appendix 3.A.1. The result below gives a quantitative comparison between the empirical measure related to the grid times and the Lebesgue measure.

Lemma 3.3.6. *Assume the conditions of Theorem 3.2.7 and* $\sum_{n\ge0}\varepsilon_n^2 < +\infty$*. Then, for any sequence of non-empty deterministic intervals* $I_n \subset [0,T]$*, such that for some* $\rho \in (0,1)$

$$\varepsilon_n^{-(2-2\rho)}|I_n| \to +\infty, \tag{3.3.11}$$

there exists an a.s. finite random variable C *such that*

$$N^n(I_n) \le C\varepsilon_n^{-2}|I_n|, \quad \forall n \ge 0, \text{ a.s.} \tag{3.3.12}$$

The condition $(\mathbf{H}_R)$-2 follows from Lemma 3.3.6 (with $I_n = [0,T]$ and any $\rho \in (0,1)$),

while the condition ($\mathbf{H}_{\mathcal{B}}$)-4 follows from Lemma 3.3.6 with $I_n := [(T - g(\varepsilon_n))_+, T]$ and the choice $g(\varepsilon) = \varepsilon$, $\rho = 1/3$.

We now prove that the statements 1-2-3 of ($\mathbf{H}_{\mathcal{B}}$) hold with $\mathcal{B}[f]$ and m defined in (3.2.16)-(3.2.17). For a Brownian motion W starting at 0 and $U \sim \mathcal{U}(0,1)$ independent of W (both independent of $\bar{\mathcal{F}}_T$) let

$$\tau(t) := \inf\{s \geq 0 : \sigma_t W_s \notin D_t\} \wedge G_t(U), \tag{3.3.13}$$

$$\tau^n(t) := \inf\{s \geq 0 : \sigma_t W_s \notin D_t\} \wedge G_t(U) \wedge \varepsilon_n^{-2}(T-t). \tag{3.3.14}$$

Since a.s. D_t is a bounded domain and σ_t is invertible, $\tau(t)$ and $\tau^n(t)$ are a.s. finite random variables. Moreover $(W_{s\wedge\tau(t)} : s \geq 0)$ is a bounded martingale (with a $\mathcal{F}_t$-measurable bound depending on $\sigma_t, \sigma_t^{-1}, L_t^{-1}$), thus

$$\mathcal{B}_t[f(\cdot)] := \mathbb{E}_t\left(f(\sigma_t W_{\tau(t)})\right)$$

(given in (3.2.16)) is well defined for any function $f \in \mathcal{P}^\alpha, \alpha \in \{2,3,4\}$. It obviously defines a linear operator from the vector space spanned by $\mathcal{P}^\alpha, \alpha = 2,3,4$, into scalar adapted processes. Note that $\mathcal{B}_0[f]$ is bounded owing to the boundedness of $\sigma_0, \sigma_0^{-1}, L_0^{-1}$.
The aforementioned boundedness on $W_{\cdot\wedge\tau(t)}$ implies $\mathbb{E}_t(W^i_{\tau(t)} W^j_{\tau(t)}) = 0$ for $0 \leq i < j \leq d$ and $\mathbb{E}_t((W^i_{\tau(t)})^2) = \mathbb{E}_t(\tau(t))$: to see these, apply the optional sampling theorem at the stopping time $\tau(t) \wedge k$ and take the limit as $k \uparrow +\infty$, each right hand side converges using the dominated convergence theorem, each left hand side using the monotone convergence theorem. As a consequence and using easy manipulations, we obtain the identity

$$\frac{\mathcal{B}_t[f(x) := |x|^2]}{\mathrm{Tr}(\sigma_t \sigma_t^{\mathsf{T}})} = \mathbb{E}_t(\tau(t)) \underset{(3.2.17)}{=} m_t.$$

Since D_t contains $0 \in \mathbb{R}^d$, $\tau(t) > 0$ a.s. and therefore $m_t > 0$ a.s.; in addition from (3.2.4), we get the boundedness of m_0^{-1} and $\mathcal{B}_0[f(\cdot)]$. We are done with the proof of ($\mathbf{H}_{\mathcal{B}}$)-2.

Observe that to get ($\mathbf{H}_{\mathcal{B}}$)-1, it remains only to justify the continuity of $\mathcal{B}_t[f(\cdot)]$. Using that $\cup_{0\leq t\leq T} D_t$ is a.s. bounded and the local Lipschitz condition of f, we have for some a.s. finite C_T and all $0 \leq s \leq t \leq T$ that

$$\begin{aligned} |\mathcal{B}_t[f(\cdot)] - \mathcal{B}_s[f(\cdot)]| &= \left|\mathbb{E}_t(f(\sigma_t W_{\tau(t)})) - \mathbb{E}_s(f(\sigma_s W_{\tau(s)}))\right| \\ &\leq C_T\left(|\sigma_t - \sigma_s| + \mathbb{E}_T(|W_{\tau(t)} - W_{\tau(s)}|)\right). \end{aligned}$$

The first term on the right hand side is clearly continuous under our assumptions on σ. For the second, write

$$\mathbb{E}_T(|W_{\tau(t)} - W_{\tau(s)}|) \leq \mathbb{E}_T(|W_{\tau(t)} - W_{\tau(s)}|^2)^{1/2} = \mathbb{E}_T(|\tau(t) - \tau(s)|)^{1/2}.$$

Let us fix t, assume $s \to t$ and let us prove that $\mathbb{E}_T(|\tau(t) - \tau(s)|) \to 0$. Define the domains $\tilde{D}_t := \sigma_t^{-1} D_t, \tilde{D}_s := \sigma_s^{-1} D_s$ (where $\sigma^{-1} D = \{\sigma^{-1} x : x \in D\}$), and set

$$\tilde{\tau}(s,t) := \inf\{r \geq 0 : W_r \notin \tilde{D}_s\} \wedge G_t(U),$$

so that

$$\mathbb{E}_T(|\tau(t)-\tau(s)|) \le \mathbb{E}_T(|\tau(t)-\tilde{\tau}(s,t)|) + \mathbb{E}_T(|\tilde{\tau}(s,t)-\tau(s)|). \tag{3.3.15}$$

From the continuity of σ_t and D_t (w.r.t. $\mu^J(\cdot,\cdot)$) one may check that $\mu^J(\tilde{D}_s,\tilde{D}_t)\to 0$: thus, the convergence to 0 of the first term in (3.3.15) readily follows by invoking Corollary 3.5.5 with D and D' equal to the components of $\tilde{D}_s$ and $\tilde{D}_t$ respectively (see (3.2.3)), with $S=W$, and making $K'\to 0$ (in the notation of Corollary 3.5.5).
The second term in (3.3.15) is bounded by $\mathbb{E}_T(|\tilde{\tau}\wedge G_t(U)-\tilde{\tau}\wedge G_s(U)|)$ (where $\tilde{\tau}$ denotes the first exit time of W from $\cup_t \tilde{D}_t$), which converges to zero by the dominated convergence theorem in view of ($\mathbf{H}_G$)-1. The proof of ($\mathbf{H}_{\mathcal{B}}$)-1 is now complete.

It remains to show the condition ($\mathbf{H}_{\mathcal{B}}$)-3 with the choice $g(\varepsilon)=\varepsilon$ made at the beginning. Fix n and i, let $f:\mathbb{R}^d\to\mathbb{R}$ be any α-homogeneous polynomial function of degree $\alpha=2,3,4$. Let

$$\tilde{\tau}_i^n := \inf\{t>\tau_{i-1}^n : S_t - S_{\tau_{i-1}^n}\notin \varepsilon_n D^n_{\tau_{i-1}^n}\}\wedge(\tau_{i-1}^n+\varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}))\wedge T,$$

$$\hat{\tau}_i^n := \inf\{t>\tau_{i-1}^n : S_t - S_{\tau_{i-1}^n}\notin \varepsilon_n D_{\tau_{i-1}^n}\}\wedge(\tau_{i-1}^n+\varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}))\wedge T$$

($\hat{\tau}_i^n$ differs from $\tilde{\tau}_i^n$ by the use of $D_{\tau_{i-1}^n}$ instead of $D^n_{\tau_{i-1}^n}$ in the definition, and τ_i^n differs from $\tilde{\tau}_i^n$ by the use of $\Delta_{n,i}$ in (3.2.6)). Recall that by ($\mathbf{H}_D^2$) $\sup_{n\ge0}\sup_{x\in D_t\cup D_t^n}|x|\le L_t^{-1}$. Define a sequence of events $\Omega_n:=\{\varepsilon_n L_t^{-1}\le\delta_t\ \forall t\in[0,T]\}, n\ge0$, where δ_t is given by ($\mathbf{H}_\Delta$). For any $\tau_{i-1}^n<(T-\varepsilon_n)_+$ (since we consider $g(\varepsilon)=\varepsilon$) and in view of (3.2.16), write

$$\begin{aligned}
&\mathbb{1}_{\Omega_n}\left|\varepsilon_n^{-\alpha}\mathbb{E}_{\tau_{i-1}^n}(f(S_{\tau_i^n}-S_{\tau_{i-1}^n}))-\mathcal{B}_{\tau_{i-1}^n}[f(\cdot)]\right| \\
&\qquad\le \mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(S_{\tau_i^n}-S_{\tau_{i-1}^n}))-\mathbb{E}_{\tau_{i-1}^n}(f(S_{\tilde{\tau}_i^n}-S_{\tau_{i-1}^n}))\right| \\
&\qquad+\mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(S_{\tilde{\tau}_i^n}-S_{\tau_{i-1}^n}))-\mathbb{E}_{\tau_{i-1}^n}(f(S_{\hat{\tau}_i^n}-S_{\tau_{i-1}^n}))\right| \\
&\qquad+\mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(S_{\hat{\tau}_i^n}-S_{\tau_{i-1}^n}))-\mathbb{E}_{\tau_{i-1}^n}(f(\varepsilon_n\sigma_{\tau_{i-1}^n}W_{\tau^n(\tau_{i-1}^n)}))\right| \\
&\qquad+\mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(\varepsilon_n\sigma_{\tau_{i-1}^n}W_{\tau^n(\tau_{i-1}^n)}))-\mathbb{E}_{\tau_{i-1}^n}(f(\varepsilon_n\sigma_{\tau_{i-1}^n}W_{\tau(\tau_{i-1}^n)}))\right|.
\end{aligned} \tag{3.3.16}$$

Remark that the assumption ($\mathbf{H}_{loc}^{D,\sigma}$) is verified on Ω_n for $D_{\tau_{i-1}^n}$ and $D^n_{\tau_{i-1}^n}$ due to ($\mathbf{H}_\Delta$) with η_σ given by ($\mathbf{H}_S$), $m_\sigma=\mathbb{E}_{\tau_{i-1}^n}(C_\sigma^4)$, $\Lambda^\sigma_{\min}=v^{-1}_{\tau_{i-1}^n}$ and $\Lambda^\sigma_{\max}=b_{\max}=v_{\tau_{i-1}^n}$. In addition we may take $L_D=L_{\tau_{i-1}^n}$.
For the first term of the right-hand side of (3.3.16), by applying Lemma 3.3.2 and using that $|\tilde{\tau}_i^n-\tau_i^n|\le|\Delta_{n,i}|$ together with ($\mathbf{H}_G$)-2 we have for some $\bar{\mathcal{F}}_{\tau_{i-1}^n}$-measurable K and for some constant $\eta>0$

$$\mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(S_{\tau_i^n}-S_{\tau_{i-1}^n}))-\mathbb{E}_{\tau_{i-1}^n}(f(S_{\tilde{\tau}_i^n}-S_{\tau_{i-1}^n}))\right|\le p_{\tau_{i-1}^n}K\varepsilon_n^\eta.$$

For the second term we apply Propositions 3.3.5 with $D=D_{\tau_{i-1}^n}$ and $D=D^n_{\tau_{i-1}^n}$ conditionally on $U_{n,i}$ and taking $T:=\varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i})\wedge(T-\tau_{i-1}^n)$. Note that the necessary conditions are verified due to ($\mathbf{H}_D^1$). Since in Proposition 3.3.5 the variable K is independent of T, we may further take (in view of Lemma 3.2.5-(ii)) expectation w.r.t. $U_{n,i}$. Thus we get for some

$\bar{\mathcal{F}}_{\tau_{i-1}^n}$-measurable K and the constant $\eta_{\mathcal{D}} > 0$

$$\mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(S_{\hat{\tau}_i^n} - S_{\tau_{i-1}^n})) - \mathbb{E}_{\tau_{i-1}^n}(f(S_{\hat{\tau}_i^n} - S_{\tau_{i-1}^n}))\right| \le K\varepsilon_n^{\eta_{\mathcal{D}}}.$$

For the third term we similarly apply Propositions 3.3.4 with $D = D_{\tau_{i-1}^n}$ and $D = D^n_{\tau_{i-1}^n}$ conditionally on the coupling $U_{n,i} = U$ and taking $T := \varepsilon_n^2 G_{\tau_{i-1}^n}(U) \wedge (T - \tau_{i-1}^n)$. Again for some $\bar{\mathcal{F}}_{\tau_{i-1}^n}$-measurable $K > 0$ (integrating with respect to $U_{n,i} = U \sim \mathcal{U}(0,1)$ since K is independent of T in Proposition 3.3.4, and in view of Lemma 3.2.5-(ii)) we get

$$\mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(S_{\hat{\tau}_i^n} - S_{\tau_{i-1}^n})) - \mathbb{E}_{\tau_{i-1}^n}(f(\varepsilon_n\sigma_{\tau_{i-1}^n}W_{\tau^n(\tau_{i-1}^n)}))\right| \le K\varepsilon_n^{\eta_\sigma}.$$

Finally for the last term we write using Lemma 3.3.3, $\tau_{i-1}^n < (T - \varepsilon_n)_+$ and (3.3.9), that

$$\begin{aligned}
&\mathbb{1}_{\Omega_n}\varepsilon_n^{-\alpha}\left|\mathbb{E}_{\tau_{i-1}^n}(f(\varepsilon_n\sigma_{\tau_{i-1}^n}W_{\tau^n(\tau_{i-1}^n)})) - \mathbb{E}_{\tau_{i-1}^n}(f(\varepsilon_n\sigma_{\tau_{i-1}^n}W_{\tau(\tau_{i-1}^n)}))\right| \\
&\quad\le 2\mathbb{1}_{\Omega_n}C_f L_{\tau_{i-1}^n}^{-\alpha}\,\mathbb{P}_{\tau_{i-1}^n}(\tau(\tau_{i-1}^n) > (T - \tau_{i-1}^n)\varepsilon_n^{-2}) \\
&\quad\le KC\exp(-C\varepsilon_n^{-1}) \le K\varepsilon_n \sup_{x\ge 0}\left(xe^{-x}\right)
\end{aligned}$$

for some a.s. finite K (independent of T and τ_{i-1}^n) and an $\bar{\mathcal{F}}_{\tau_{i-1}^n}$-measurable C.
In addition from ($\mathbf{H}_G$)-2, Lemmas 3.3.2, 3.3.3 and Propositions 3.3.4, 3.3.5 we also deduce that $\bar{\mathcal{F}}_{\tau_{i-1}^n}$-measurable K in the four latter bounds may be expressed as continuous positive simple expressions of $\eta_\sigma, \mathbb{E}_{\tau_{i-1}^n}(C_\sigma^4), v_{\tau_{i-1}^n}$ and $L_{\tau_{i-1}^n}$. This implies that, due to boundedness of the processes v_t, L_t, $\mathbb{E}_t(C_\sigma^4)$ (since it is a martingale and thus has a càdlàg version) and also p_t, we may choose $K > 0$ uniformly in $n \ge 0$ and $i = 1, \dots, N_T^n$ so that for all $n \ge 0$,

$$\mathbb{1}_{\Omega_n}\sup_{\tau_{i-1}^n < (T-\varepsilon_n)_+}\left|\varepsilon_n^{-\alpha}\mathbb{E}_{\tau_{i-1}^n}(f(S_{\tau_i^n} - S_{\tau_{i-1}^n})) - \mathcal{B}_{\tau_{i-1}^n}[f(\cdot)]\right| \le K\varepsilon_n^{\eta\wedge\eta_{\mathcal{D}}\wedge\eta_\sigma\wedge 1}.$$

Finally, $\mathbb{1}_{\Omega_n} = 1$ except for a finite number of n a.s., hence we easily derive the inequality (3.3.4). Thus, ($\mathbf{H}_{\mathcal{B}}$)-3 is verified. The proof of Theorem 3.2.7 is finished. □

3.4 Proof of the general CLT (Theorem 3.3.1)

We adopt the framework of Section 3.3.1. The overall strategy of proof is standard and consists in proving that the drift and the quadratic variation/covariation of the error $\mathcal{E}_n$ converge in probability to some limits (see details in Subsection 3.4.2). The trick is to switch from convergence in probability to a.s. convergence by using the subsequence principle in Lemma 2.2.2. In our framework, the flexibility in choosing another subsequence ι' is that it can be made to guarantee $\sum_{n\ge 0}\varepsilon^2_{\iota\circ\iota'(n)} < +\infty$ and to make ($\mathbf{H}_R$)-($\mathbf{H}_{\mathcal{B}}$) valid along this sequence $\tilde{\varepsilon}_n = \varepsilon_{\iota\circ\iota'(n)}$. In doing so, we define a new sequence of discretization grids $\tilde{\mathcal{T}} := \{\mathcal{T}^{\iota\circ\iota'(n)} : n \ge 0\}$. Because the new sequence $(\tilde{\varepsilon}_n : n \ge 0)$ is square summable and ($\mathbf{H}_R$)-($\mathbf{H}_{\mathcal{B}}$) hold for $(\tilde{\varepsilon}_n : n \ge 0)$, we are back to the framework of admissible sequences of discretization grids studied in [GL14a] and Chapters 1-2 with a parameter $\rho_N = 1$. This latter framework

is quite interesting since some a.s. results for discretization errors are already available.

The careful reader will have observed that the above references study these convergence results for admissible grid sequences in the context of a Brownian filtration $\mathcal{F}^B$ (this choice of filtration was motivated by the application at hand). However, the reader can check easily that the results of [GL14a] and Chapters 1-2 hold true even if the filtration satisfies the usual assumptions of being only right continuous and $\mathbb{P}$-complete, as for $\bar{\mathcal{F}}$ in particular, because the proofs of the above references mostly use the Itô formula for the continuous semimartingale S of the form $(\mathbf{H}_S^{gen.})$ and the BDG inequalities for the Brownian integral (as in the decomposition of S), both being available when the filtration satisfies the usual assumptions.

3.4.1 Part I: Preliminary almost sure convergence results

We now provide some auxiliary almost sure convergence results that are necessary for the proof of Theorem 3.3.1. These results are, however, of their own interest and hence we put them in a separate section. In view of the above subsequence principle, these results will have to be established for a sub-subsequence $(\tilde{\varepsilon}_n : n \geq 0)$ instead of $(\varepsilon_n : n \geq 0)$. But to maintain simple notation, we keep writing ε_n (instead of $\tilde{\varepsilon}_n$), and therefore, we will have to assume that $(\varepsilon_n : n \geq 0)$ is square summable and $(\mathbf{H}_R)$-$(\mathbf{H}_{\mathcal{B}})$ hold for $(\varepsilon_n : n \geq 0)$.

The next lemma allows to replace locally the values of homogeneous functions of the process increments by their conditional expectations.

Proposition 3.4.1. *Assume the hypotheses $(\mathbf{H}_S^{gen.})$ and $(\mathbf{H}_R)$ for the sequence $(\varepsilon_n)_{n\geq 0}$ with $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. Let $\alpha \in \{2,3,4\}$. For any adapted continuous $\mathcal{P}^\alpha$-valued process $(f_t)_{0\leq t\leq T}$ with bounded f_0 (i.e. given by $f_t = \sum_{\text{finitely many } k} f_t^k P_k$ where P_k are monomials of degree α and f_t^k are adapted continuous scalar process with bounded random variables f_0^k), and for any adapted continuous scalar process $(H_t)_{0\leq t\leq T}$ with bounded H_0, we have*

$$\varepsilon_n^{2-\alpha} \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \left(f_{\tau_{i-1}^n}(\Delta S_{\tau_i^n \wedge t}) - \mathbb{E}_{\tau_{i-1}^n}(f_{\tau_{i-1}^n}(\Delta S_{\tau_i^n})) \right) \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} 0.$$

Similar convergence-in-probability results are typically deduced using the Lenglart inequality (see e.g. [Fuk11b, Proof of Lemma A.2]). However, here, since we need a.s. results to leverage the setting of admissible grid sequences, and due to lack of suitable references we provide our own proof in Section 3.A.2.

Next, we reformulate the above convergence in a form ready to be used in combination with $(\mathbf{H}_{\mathcal{B}})$.

Proposition 3.4.2. *Assume $(\mathbf{H}_S^{gen.})$, $(\mathbf{H}_R)$ and $(\mathbf{H}_{\mathcal{B}})$ for the sequence $(\varepsilon_n)_{n\geq 0}$ with $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. Let $(f_t)_{0\leq t\leq T}$ be adapted continuous $\mathcal{P}^\alpha$-valued process for $\alpha \in \{2,3,4\}$ with bounded f_0 (see the definition in Proposition 3.4.1). Then*

(i) the process $(\mathcal{B}_t[f_t(\cdot)])_{0\leq t\leq T}$ is adapted continuous;

(ii) for some random variable $C_{(3.4.1)}$ a.s. finite and independent of n, we have a.s. for all $n \geq 0$

$$\sup_{\tau_{i-1}^n < (T-g(\varepsilon_n))_+} \left| \varepsilon_n^{-\alpha} \mathbb{E}_{\tau_{i-1}^n}(f_{\tau_{i-1}^n}(S_{\tau_i^n} - S_{\tau_{i-1}^n})) - \mathcal{B}_{\tau_{i-1}^n}[f_{\tau_{i-1}^n}(\cdot)] \right| \leq C_{(3.4.1)} \varepsilon_n^{\eta}; \tag{3.4.1}$$

(iii) for any adapted continuous scalar process $(H_t)_{0 \leq t \leq T}$ we have

$$\varepsilon_n^2 \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \left(\varepsilon_n^{-\alpha} \mathbb{E}_{\tau_{i-1}^n}(f_{\tau_{i-1}^n}(S_{\tau_i^n} - S_{\tau_{i-1}^n})) - \mathcal{B}_{\tau_{i-1}^n}[f_{\tau_{i-1}^n}(\cdot)] \right) \overset{u.c.a.s.}{\underset{n \to +\infty}{\longrightarrow}} 0. \tag{3.4.2}$$

Proof. Statements (i) and (ii) are obvious to check from ($\mathbf{H}_{\mathcal{B}}$)-1 and ($\mathbf{H}_{\mathcal{B}}$)-3.

Let us now prove (iii). Decomposing the sum in (3.4.2) into the contributions of the intervals $[0, t \wedge (T - g(\varepsilon_n))_+)$ and $[t \wedge (T - g(\varepsilon_n))_+, t]$, we write using (3.4.1)

$$\begin{aligned}
&\varepsilon_n^2 \left| \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \left(\varepsilon_n^{-\alpha} \mathbb{E}_{\tau_{i-1}^n}(f_{\tau_{i-1}^n}(S_{\tau_i^n} - S_{\tau_{i-1}^n})) - \mathcal{B}_{\tau_{i-1}^n}[f_{\tau_{i-1}^n}(\cdot)] \right) \right| \\
&\leq C_{(3.4.1)} \varepsilon_n^2 N_t^n \sup_{0 \leq s \leq t} |H_s| \varepsilon_n^{\eta} \\
&+ \left(\sum_{(T-g(\varepsilon_n))_+ \leq \tau_{i-1}^n < T} \varepsilon_n^2 \right) \sup_{0 \leq s \leq t} \left(|H_{\varphi(s)}| \left(|\varepsilon_n^{-\alpha} \mathbb{E}_{\varphi(s)}(f_{\varphi(s)}(S_{\bar{\varphi}(s)} - S_{\varphi(s)}))| + |\mathcal{B}_{\varphi(s)}[f_{\varphi(s)}(\cdot)]| \right) \right) \\
&\overset{u.c.a.s.}{\underset{n \to +\infty}{\longrightarrow}} 0
\end{aligned}$$

where for the first term we used that $\varepsilon_n^2 N_t^n$ is a.s. bounded owing to ($\mathbf{H}_R$)-2, and for the second term the convergence is proved by (i), ($\mathbf{H}_R$)-1 and using that by ($\mathbf{H}_{\mathcal{B}}$)-4 we have $\sum_{(T-g(\varepsilon_n))_+ \leq \tau_{i-1}^n < T} \varepsilon_n^2 \overset{\text{a.s.}}{\underset{n \to +\infty}{\longrightarrow}} 0$. □

The next theorem states the convergence of the renormalized sum of process values at the discretization grid points.

Theorem 3.4.3. *Assume ($\mathbf{H}_S^{gen.}$), ($\mathbf{H}_R$) and ($\mathbf{H}_{\mathcal{B}}$) for the sequence $(\varepsilon_n)_{n \geq 0}$ such that $\sum_{n \geq 0} \varepsilon_n^2 < +\infty$. Let $(m_t)_{0 \leq t \leq T}$ be given by ($\mathbf{H}_{\mathcal{B}}$)-2. Let $(H_t)_{0 \leq t \leq T}$ be an adapted continuous scalar process with bounded H_0. Let $\alpha \in \{2, 3, 4\}$ and $(f_t)_{0 \leq t \leq T}$ be an adapted continuous $\mathcal{P}^\alpha$-valued process with bounded f_0. Then the following uniform convergences hold on $[0, T]$:*

$$\varepsilon_n^2 \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \overset{u.c.a.s.}{\underset{n \to +\infty}{\longrightarrow}} \int_0^t H_s m_s^{-1} \mathrm{d}s, \tag{3.4.3}$$

$$\varepsilon_n^{2-\alpha} \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} f_{\tau_{i-1}^n}(S_{\tau_i^n \wedge t} - S_{\tau_{i-1}^n}) \overset{u.c.a.s.}{\underset{n \to +\infty}{\longrightarrow}} \int_0^t H_s m_s^{-1} \mathcal{B}_s[f_s(\cdot)] \mathrm{d}s. \tag{3.4.4}$$

Proof. Let us first prove (3.4.3). The assumption ($\mathbf{H}_{\mathcal{B}}$)-2 reads $\mathcal{B}_t[f(x) := |x|^2] = m_t \operatorname{Tr}(\sigma_t \sigma_t^\intercal)$,

where the above right-hand side is positive continuous. Let

$$\xi_t := m_t^{-1} \operatorname{Tr}(\sigma_t \sigma_t^{\mathsf{T}})^{-1}, \quad t \in [0,T]; \tag{3.4.5}$$

note that ξ is adapted continuous, ξ_0 is bounded in view of ($\mathbf{H}_{\mathcal{B}}$)-2 and ($\mathbf{H}_S^{gen.}$), and we have

$$\xi_t \mathcal{B}_t[f(x) := |x|^2] = 1, \quad t \in [0,T]. \tag{3.4.6}$$

Now leverage the above equality to write

$$\begin{aligned}
\varepsilon_n^2 \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} &= \varepsilon_n^2 \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \xi_{\tau_{i-1}^n} \mathcal{B}_{\tau_{i-1}^n}[f(x) := |x|^2] \\
&= \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \xi_{\tau_{i-1}^n} |\Delta S_{\tau_i^n \wedge t}|^2 \\
&\quad + \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \xi_{\tau_{i-1}^n} \left(\mathbb{E}_{\tau_{i-1}^n}(|\Delta S_{\tau_i^n}|^2) - |\Delta S_{\tau_i^n \wedge t}|^2 \right) \\
&\quad + \varepsilon_n^2 \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \xi_{\tau_{i-1}^n} \left(\mathcal{B}_{\tau_{i-1}^n}[f(x) = |x|^2] - \varepsilon_n^{-2} \mathbb{E}_{\tau_{i-1}^n} |\Delta S_{\tau_i^n}|^2 \right).
\end{aligned}$$

Applying (3.4.2) from Proposition 3.4.2 with $f_t(x) = |x|^2$, $\alpha = 2$, we justify that the third term above converges uniformly a.s. to 0. Further using Proposition 3.4.1 with $f_t(x) = |x|^2$, $\alpha = 2$, the second term above also converges uniformly a.s. to 0. Finally by Proposition 1.3.9 (it easy to check in the proof that the convergence there holds in the sup-norm) we obtain

$$\sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \xi_{\tau_{i-1}^n} |\Delta S_{\tau_i^n \wedge t}|^2 \xrightarrow[n \to +\infty]{u.c.a.s.} \int_0^t H_s \xi_s \operatorname{Tr}(\sigma_s \sigma_s^{\mathsf{T}}) \mathrm{d}s = \int_0^t H_s m_s^{-1} \mathrm{d}s,$$

where for the last equality we recast the definition of ξ. The proof of (3.4.3) is finished.

Regarding (3.4.4), write

$$\begin{aligned}
\varepsilon_n^{2-\alpha} \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} f_{\tau_{i-1}^n}(S_{\tau_i^n \wedge t} - S_{\tau_{i-1}^n}) &= \varepsilon_n^2 \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \mathcal{B}_{\tau_{i-1}^n}[f_{\tau_{i-1}^n}(\cdot)] \\
&\quad + \varepsilon_n^2 \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \left(\varepsilon_n^{-\alpha} \mathbb{E}_{\tau_{i-1}^n} \left(f_{\tau_{i-1}^n}(S_{\tau_i^n} - S_{\tau_{i-1}^n}) \right) - \mathcal{B}_{\tau_{i-1}^n}[f_{\tau_{i-1}^n}(\cdot)] \right) \\
&\quad + \varepsilon_n^{2-\alpha} \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \left(f_{\tau_{i-1}^n}(S_{\tau_i^n \wedge t} - S_{\tau_{i-1}^n}) - \mathbb{E}_{\tau_{i-1}^n} \left(f_{\tau_{i-1}^n}(S_{\tau_i^n} - S_{\tau_{i-1}^n}) \right) \right).
\end{aligned} \tag{3.4.7}$$

Proposition 3.4.2 and Proposition 3.4.1 imply respectively that the second and the third terms in the above right-hand side converge uniformly a.s. to 0. Last, apply (3.4.3) to the process $(H_t \mathcal{B}_t[f_t(\cdot)])_{0 \le t \le T}$ (which is adapted continuous by Proposition 3.4.2): this shows that the first term of the right-hand side of (3.4.7) converges uniformly a.s. to $\int_0^t H_s m_s^{-1} \mathcal{B}_s[f_s(\cdot)] \mathrm{d}s$. We are done. □

The next lemma gives the limit of integral of weighted increments of S.

Lemma 3.4.4. *Assume* ($\mathbf{H}_S^{gen.}$), ($\mathbf{H}_R$) *and* ($\mathbf{H}_\mathcal{B}$) *for the sequence* $(\varepsilon_n)_{n\geq 0}$ *with* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Let* $(\mathcal{M}_t)_{0\leq t\leq T}$ *be a* $\mathrm{Mat}_{m,d}$*-valued adapted continuous process with bounded* $\mathcal{M}_0$, *and recall the definition* (3.2.18) *of the* $\mathbb{R}^d$*-valued adapted continuous process* $(Q_t)_{0\leq t\leq T}$*:*

$$Q_t := \frac{1}{3}m_t^{-1}\begin{pmatrix}(\sigma_t\sigma_t^\mathsf{T})_{11}^{-1}\mathcal{B}_t[f(x):=(x^1)^3]\\ \vdots\\ (\sigma_t\sigma_t^\mathsf{T})_{dd}^{-1}\mathcal{B}_t[f(x):=(x^d)^3]\end{pmatrix}. \tag{3.4.8}$$

Then

$$\varepsilon_n^{-1}\int_0^t \mathcal{M}_{\varphi(s)}\Delta S_s \mathrm{d}s \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} \int_0^t \mathcal{M}_s Q_s \mathrm{d}s.$$

Proof. For any adapted continuous scalar process $(H_t)_{0\leq t\leq T}$ with bounded H_0 and any coordinate $k\in\{1,\dots,d\}$, the Itô formula yields that

$$\varepsilon_n^{-1}\int_0^t H_{\varphi(s)}\Delta S_s^k \mathrm{d}s = \varepsilon_n^{-1}\sum_{\tau_{i-1}^n<t} H_{\tau_{i-1}^n}(\sigma_{\tau_{i-1}^n}\sigma_{\tau_{i-1}^n}^\mathsf{T})_{kk}^{-1}\times$$
$$\times\left(\frac{1}{3}(\Delta S^k_{\tau_i^n\wedge t})^3 - \int_{\tau_{i-1}^n}^{\tau_i^n\wedge t}(\Delta S_s^k)^2\mathrm{d}S_s^k - \int_{\tau_{i-1}^n}^{\tau_i^n\wedge t}\Delta S_s^k\Delta(\sigma_s\sigma_s^\mathsf{T})_{kk}\mathrm{d}s\right).$$

First, by Theorem 3.4.3 applied with $f_t(x)=(x^k)^3$ we obtain

$$\varepsilon_n^{-1}\sum_{\tau_{i-1}^n<t} H_{\tau_{i-1}^n}(\sigma_{\tau_{i-1}^n}\sigma_{\tau_{i-1}^n}^\mathsf{T})_{kk}^{-1}(\Delta S^k_{\tau_i^n\wedge t})^3 \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} \int_0^t H_s m_s^{-1}(\sigma_s\sigma_s^\mathsf{T})_{kk}^{-1}\mathcal{B}_s[f(x):=(x^k)^3]\mathrm{d}s.$$

Second, apply Lemma 3.B.3 with $\alpha=2$ to get

$$\varepsilon_n^{-1}\sum_{\tau_{i-1}^n<t} H_{\tau_{i-1}^n}(\sigma_{\tau_{i-1}^n}\sigma_{\tau_{i-1}^n}^\mathsf{T})_{kk}^{-1}\int_{\tau_{i-1}^n}^{\tau_i^n\wedge t}(\Delta S_s^k)^2\mathrm{d}S_s^k \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} 0.$$

Finally, in view of (3.3.2) in ($\mathbf{H}_R$) and using the Hölder continuity of σ in ($\mathbf{H}_S^{gen.}$), it readily follows that

$$\left|\varepsilon_n^{-1}\sum_{\tau_{i-1}^n<t} H_{\tau_{i-1}^n}(\sigma_{\tau_{i-1}^n}\sigma_{\tau_{i-1}^n}^\mathsf{T})_{kk}^{-1}\int_{\tau_{i-1}^n}^{\tau_i^n\wedge t}\Delta S_s^k\Delta(\sigma_s\sigma_s^\mathsf{T})_{kk}\mathrm{d}s\right|$$
$$\leq \varepsilon_n^{-1}\sup_{0\leq s\leq t}|H_s(\sigma_s\sigma_s^\mathsf{T})_{kk}^{-1}|\sup_{0\leq s\leq t}|\Delta S_s^k|\sup_{0\leq s\leq t}|\Delta(\sigma_s\sigma_s^\mathsf{T})_{kk}|\,t \leq C\Big(\sup_{1\leq i\leq N_T^n}\Delta\tau_i^n\Big)^{\eta_\sigma/2}$$

for some finite random variable C. The above time step goes almost surely to 0, this is a consequence of ($\mathbf{H}_S^{gen.}$)-($\mathbf{H}_R$), see Theorem 1.3.4 and Lemma 1.3.2. All in all, this implies

$$\varepsilon_n^{-1}\int_0^t H_{\varphi(s)}\Delta S_s^k\mathrm{d}s \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} \frac{1}{3}\int_0^t H_s m_s^{-1}(\sigma_s\sigma_s^\mathsf{T})_{kk}^{-1}\mathcal{B}_s[f(x):=(x^k)^3]. \tag{3.4.9}$$

Now, apply the above for each component $\int_0^t \mathcal{M}^{lk}_{\varphi(s)}\Delta S_s^k\mathrm{d}s$ arising in the product matrix-

vector $\int_0^t \mathcal{M}_{\varphi(s)}\Delta S_s \mathrm{d}s$, we get the announced convergence. □

The next lemma handles the convergence of integral of weighted squared increments of S.

Lemma 3.4.5. *Assume* $(\mathbf{H}_S^{gen.})$, $(\mathbf{H}_R)$ *and* $(\mathbf{H}_\mathcal{B})$ *for the sequence* $(\varepsilon_n)_{n\geq 0}$ *with* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Let* $(H_t)_{0\leq t\leq T}$ *be an adapted continuous* $\mathcal{S}_d^+$*-valued process with bounded* H_0. *Then*

$$\varepsilon_n^{-2}\int_0^t \Delta S_s^\mathsf{T} H_{\varphi(s)}\Delta S_s \mathrm{d}s \xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t m_s^{-1}\mathcal{B}_s[f(x) := ((\sigma_s^{-1}x)^\mathsf{T} X_s(\sigma_s^{-1}x))^2]\mathrm{d}s,$$

where X_s *the solution of the matrix equation* (3.2.13) *for* $c = \sigma_s^\mathsf{T} H_s \sigma_s$ *(remark that* $\sigma_s^\mathsf{T} H_s \sigma_s$ *is in* $\mathcal{S}_d^+$*).*

Proof. Set $\Lambda_s := (\sigma_s^{-1})^\mathsf{T} X_s \sigma_s^{-1}$. First observe that, owing to the properties of Lemma 3.2.6, X and Λ are adapted continuous processes. Moreover, multiply (3.2.13) (with $c = \sigma_s^\mathsf{T} H_s \sigma_s$) by $(\sigma_s^{-1})^\top$ on the left and σ_s^{-1} on the right: this gives the identity

$$2\Lambda_s \operatorname{Tr}(\sigma_s\sigma_s^\top\Lambda_s) + 4\Lambda_s\sigma_s\sigma_s^\top\Lambda_s = H_s. \tag{3.4.10}$$

Besides, for $\tau_{i-1}^n < t$, the Itô formula gives

$$\begin{aligned}(\Delta S_{\tau_i^n\wedge t}^\mathsf{T}\Lambda_{\tau_{i-1}^n}\Delta S_{\tau_i^n\wedge t})^2 &= 4\int_{\tau_{i-1}^n}^{\tau_i^n\wedge t}\Delta S_s^\top\Lambda_{\varphi(s)}\Delta S_s\Delta S_s^\top\Lambda_{\varphi(s)}\mathrm{d}S_s \\ &\quad + \int_{\tau_{i-1}^n}^{\tau_i^n\wedge t}\Delta S_s^\top\left[2\Lambda_{\varphi(s)}\operatorname{Tr}(\sigma_s\sigma_s^\top\Lambda_{\varphi(s)}) + 4\Lambda_{\varphi(s)}\sigma_s\sigma_s^\top\Lambda_{\varphi(s)}^\top\right]\Delta S_s\mathrm{d}s.\end{aligned}$$

Therefore, summing over i for $\tau_{i-1}^n < t$ and using the idendity (3.4.10), we get

$$\begin{aligned}\varepsilon_n^{-2}\sum_{\tau_{i-1}^n<t}(\Delta S_{\tau_i^n\wedge t}^\mathsf{T}\Lambda_{\tau_{i-1}^n}\Delta S_{\tau_i^n\wedge t})^2 &= 4\varepsilon_n^{-2}\int_0^t\Delta S_s^\top\Lambda_{\varphi(s)}\Delta S_s\Delta S_s^\top\Lambda_{\varphi(s)}\mathrm{d}S_s \\ &\quad + \varepsilon_n^{-2}\int_0^t\Delta S_s^\top\left[2\Lambda_{\varphi(s)}\operatorname{Tr}(\Delta(\sigma_s\sigma_s^\top)\Lambda_{\varphi(s)}) + 4\Lambda_{\varphi(s)}\Delta(\sigma_s\sigma_s^\top)\Lambda_{\varphi(s)}^\top\right]\Delta S_s\mathrm{d}s \\ &\quad + \varepsilon_n^{-2}\int_0^t\Delta S_s^\top H_{\varphi(s)}\Delta S_s\mathrm{d}s.\end{aligned} \tag{3.4.11}$$

Lemma 3.B.3 with $\alpha = 3$ implies that $\varepsilon_n^{-2}\int_0^t\Delta S_s^\top\Lambda_{\varphi(s)}\Delta S_s\Delta S_s^\top\Lambda_{\varphi(s)}\mathrm{d}S_s \xrightarrow[n\to+\infty]{u.c.a.s.} 0$. Moreover, the Hölder continuity of σ in $(\mathbf{H}_S^{gen.})$ and the bound (3.3.2) of $(\mathbf{H}_R)$ ensure the existence of a a.s. finite random variable $C > 0$ such that

$$\sup_{t\leq T}\left|\varepsilon_n^{-2}\int_0^t\Delta S_s^\top\left[2\Lambda_{\varphi(s)}\operatorname{Tr}(\Delta(\sigma_s\sigma_s^\top)\Lambda_{\varphi(s)}) + 4\Lambda_{\varphi(s)}\Delta(\sigma_s\sigma_s^\top)\Lambda_{\varphi(s)}^\top\right]\Delta S_s\mathrm{d}s\right| \leq C\Big(\sup_{1\leq i\leq N_T^n}\Delta\tau_i^n\Big)^{\eta_\sigma/2}.$$

The latter bound converges to 0, see the arguments in the proof of Lemma 3.4.4. Therefore,

from (3.4.11), we obtain

$$\varepsilon_n^{-2}\int_0^t \Delta S_s^{\mathsf{T}} H_{\varphi(s)} \Delta S_s \mathrm{d}s - \varepsilon_n^{-2} \sum_{\tau_{i-1}^n < t} (\Delta S_{\tau_i^n \wedge t}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n \wedge t})^2 \xrightarrow[n\to+\infty]{u.c.a.s.} 0.$$

Observe that due to the boundedness of σ_0, σ_0^{-1} and H_0, and the properties of the solution of (3.2.13), the coefficients of X_0 and Λ_0 are bounded random variables. Thus, we can apply Theorem 3.4.3 with $\alpha = 4$ and $f_s(x) := (x^{\mathsf{T}}\Lambda_s x)^2$, to obtain

$$\begin{aligned}\varepsilon_n^{-2} \sum_{\tau_{i-1}^n < t} (\Delta S_{\tau_i^n}^{\mathsf{T}} \Lambda_{\tau_{i-1}^n} \Delta S_{\tau_i^n})^2 &\xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t m_s^{-1} \mathcal{B}_s[f(x) := (x^{\mathsf{T}}\Lambda_s x)^2]\mathrm{d}s \\ &= \int_0^t m_s^{-1} \mathcal{B}_s[f(x) := ((\sigma_s^{-1}x)^{\mathsf{T}} X_s (\sigma_s^{-1}x))^2]\mathrm{d}s.\end{aligned}$$

The proof is complete. □

3.4.2 Part II: Conclusion of the proof

Now we are in a position to finish the proof of Theorem 3.3.1. It boils down to combine previous preliminary results with the application of an abstract CLT for semimartingale sequences. The reference result on this subject is [JS02, Chapter IX, Theorem 7.3]. Next we state a theorem that essentially follows from a simplified version of this general result given in [Fuk11b, Theorem A.1]. For notions of stable convergence in distribution, [JS02, p. 512]-[JP12, Section 2.2.1.].

Theorem 3.4.6. *Let $(\Omega, \mathcal{F}, (\bar{\mathcal{F}}_t)_{0\le t\le T}, \mathbb{P})$ be a filtered probability space supporting a $\bar{\mathcal{F}}$-adapted d-dimensional Brownian motion $(B_t)_{0\le t\le T}$. Let $(S^n)_{n\ge 0}$ be a sequence of adapted continuous semimartingales of the form*

$$S^n = A^n + M^n,$$

where M^n are $\mathbb{R}^m$-valued $\bar{\mathcal{F}}$-local martingales of the form $M^n = \int_0^{\cdot} \alpha_s^n \mathrm{d}B_s$, and A^n are $\mathbb{R}^m$-valued adapted continuous processes with finite variation (note that m and d are not necessarily equal). Suppose that:

a) $\langle M^n\rangle_t \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^t \mathcal{K}_s \mathrm{d}s$ for all $t\in[0,T]$ and $(\mathcal{K}_t)_{0\le t\le T}$ is a $\mathcal{S}_m^+$-valued adapted process;

b) $\langle M^n, B\rangle_t \xrightarrow[n\to+\infty]{\mathbb{P}} 0$ for all $t\in[0,T]$;

c) there exists an adapted continuous $\mathbb{R}^m$-valued process A such that $\sup_{0\le t\le T}|A_t^n - A_t| \xrightarrow[n\to+\infty]{\mathbb{P}} 0$.

We denote by $\mathcal{K}_t^{1/2}$ the principal square root of the symmetric non-negative definite matrix $\mathcal{K}_t$. Let W be a m-dimensional Brownian motion independent of $\bar{\mathcal{F}}_T$ defined on an extended

probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. *Then, we have the following functional* $\bar{\mathcal{F}}$*-stable convergence in distribution*

$$S_t^n \xRightarrow[[0,T]]{d} A_t + \int_0^t \mathcal{K}_s^{1/2} \mathrm{d}W_s.$$

Proof. First we apply [Fuk11b, Theorem A.1] to the martingale sequence M^n. The conditions of [Fuk11b, Theorem A.1] follow from (a)-(b) and the fact that $M^n = \int_0^{\cdot} \alpha_s^n \mathrm{d}B_s$ is orthogonal to all martingales that are orthogonal to B. Note that this result in [Fuk11b] can be easily extended to our multidimensional setting using the standard Cramér-Wold argument. Finally the convergence of S^n follows from (c) and the $\bar{\mathcal{F}}$-stability in [Fuk11b, Theorem A.1]. □

We now proceed to the proof of Theorem 3.3.1. We come back to the setting of Theorem 3.3.1 with general sequence $\varepsilon_n \to 0$. Take any subsequence $(\varepsilon_{\iota(n)})_{n\geq 0}$. Then there exists another subsequence $(\varepsilon_{\iota'\circ\iota(n)})_{n\geq 0}$ which is square summable and for which the assumptions $(\mathbf{H}_R)$ and $(\mathbf{H}_{\mathcal{B}})$ are verified. To simplify the notation we write simply ε_n instead of $\varepsilon_{\iota'\circ\iota(n)}$ until the final part of the proof.

Recall (see definitions (3.2.15)) that

$$\mathcal{E}_t^n = \mathcal{E}_t^{n,1} + \mathcal{E}_t^{n,2},$$

with $\mathcal{E}_t^{n,1}$ and $\mathcal{E}_t^{n,2}$ given by

$$\mathcal{E}_t^{n,1} = \int_0^t \mathcal{M}_{\varphi(s)} \Delta S_s \mathrm{d}s, \qquad \mathcal{E}_t^{n,2} = \int_0^t \Delta S_s^{\mathsf{T}} \mathcal{A}_{\varphi(s)} \mathrm{d}B_s.$$

For two continuous semimartingales $(a_t)_{0\leq t\leq T}$ and $(b_t)_{0\leq t\leq T}$ with values in $\mathbb{R}^m$ and $\mathbb{R}^d$ respectively we denote by $(\langle a; b\rangle_t)_{0\leq t\leq T}$ their $\mathrm{Mat}_{m,d}$-valued quadratic covariation process. Recall that $\mathcal{A}_t = (\mathcal{A}_{1,t}, \ldots, \mathcal{A}_{m,t})^{\mathsf{T}}$ and set

$$\mathcal{A}_t^{ij} := \frac{1}{2}(\mathcal{A}_{i,t}\mathcal{A}_{j,t}^{\mathsf{T}} + \mathcal{A}_{i,t}^{\mathsf{T}}\mathcal{A}_{j,t}).$$

Using Lemma 3.4.4 we obtain for any $l = 1, \ldots, m$ and $(Q_t)_{0\leq t\leq T}$ given by (3.4.8)

$$\left\langle \varepsilon_n^{-1} \int_0^{\cdot} \Delta S_s^{\mathsf{T}} \mathcal{A}_{l,\varphi(s)} \mathrm{d}B_s; B \right\rangle_t = \varepsilon_n^{-1} \int_0^t \Delta S_s^{\mathsf{T}} \mathcal{A}_{l,\varphi(s)} \mathrm{d}s \xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t Q_s^{\mathsf{T}} \mathcal{A}_{l,s} \mathrm{d}s. \tag{3.4.12}$$

Hence

$$\left\langle \varepsilon_n^{-1} \mathcal{E}^{n,2}; B \right\rangle_t = \left\langle \varepsilon_n^{-1} \int_0^{\cdot} \Delta S_s^{\mathsf{T}} \mathcal{A}_{\varphi(s)} \mathrm{d}B_s; B \right\rangle_t \xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t Q_s^{\mathsf{T}} \mathcal{A}_s \mathrm{d}s. \tag{3.4.13}$$

Further we have

$$\left\langle \int_0^t Q_s^{\mathsf{T}} \mathcal{A}_s \mathrm{d}B_s; B \right\rangle_t = \int_0^t Q_s^{\mathsf{T}} \mathcal{A}_s \mathrm{d}s, \tag{3.4.14}$$

which in view of (3.4.13) yields

$$\left\langle \varepsilon_n^{-1} \mathcal{E}^{n,2} - \int_0^{\cdot} Q_s^{\mathsf{T}} \mathcal{A}_s \mathrm{d}B_s; B \right\rangle_t \xrightarrow[n\to+\infty]{u.c.a.s.} 0. \tag{3.4.15}$$

We decompose the quadratic covariation matrix of $\varepsilon_n^{-1}\mathcal{E}^{n,2} - \int_0^\cdot Q_s^\mathsf{T}\mathcal{A}_s \mathrm{d}B_s$ at time t as follows: for any $1 \le i, j \le m$, we have

$$\left\langle \varepsilon_n^{-1}\mathcal{E}^{n,2} - \int_0^\cdot Q_s^\mathsf{T}\mathcal{A}_s \mathrm{d}B_s \right\rangle_t^{ij} = \varepsilon_n^{-2}\int_0^t \Delta S_s^\mathsf{T}\mathcal{A}_{i,\varphi(s)}\mathcal{A}_{j,\varphi(s)}^\mathsf{T}\Delta S_s \mathrm{d}s + \int_0^t Q_s^\mathsf{T}\mathcal{A}_{i,s}\mathcal{A}_{j,s}^\mathsf{T}Q_s \mathrm{d}s - \varepsilon_n^{-1}\int_0^t Q_s^\mathsf{T}(\mathcal{A}_{j,s}\mathcal{A}_{i,\varphi(s)}^\mathsf{T} + \mathcal{A}_{i,s}\mathcal{A}_{j,\varphi(s)}^\mathsf{T})\Delta S_s \mathrm{d}s.$$

By symmetry of the matrix $\left\langle \varepsilon_n^{-1}\mathcal{E}^{n,2} - \int_0^\cdot Q_s^\mathsf{T}\mathcal{A}_s \mathrm{d}B_s \right\rangle_t$, we deduce

$$\left\langle \varepsilon_n^{-1}\mathcal{E}^{n,2} - \int_0^\cdot Q_s^\mathsf{T}\mathcal{A}_s \mathrm{d}B_s \right\rangle_t^{ij} = \varepsilon_n^{-2}\int_0^t \Delta S_s^\mathsf{T}\mathcal{A}_{\varphi(s)}^{ij}\Delta S_s \mathrm{d}s + \int_0^t Q_s^\mathsf{T}\mathcal{A}_s^{ij}Q_s \mathrm{d}s - \varepsilon_n^{-1}\int_0^t Q_s^\mathsf{T}(\mathcal{A}_{j,s}\mathcal{A}_{i,\varphi(s)}^\mathsf{T} + \mathcal{A}_{i,s}\mathcal{A}_{j,\varphi(s)}^\mathsf{T})\Delta S_s \mathrm{d}s.$$

First, apply the dominated convergence theorem by invoking the a.s. continuity of $\mathcal{A}_\cdot$ and $Q_\cdot$ on $[0,T]$, ($\mathbf{H}_R$) and the convergence to 0 of the mesh size of $\mathcal{T}^n$ (see the proof of Lemma 3.4.4), it gives

$$\varepsilon_n^{-1}\int_0^t \left|2Q_{\varphi(s)}^\mathsf{T}\mathcal{A}_{\varphi(s)}^{ij} - Q_s^\mathsf{T}(\mathcal{A}_{j,s}\mathcal{A}_{i,\varphi(s)}^\mathsf{T} + \mathcal{A}_{i,s}\mathcal{A}_{j,\varphi(s)}^\mathsf{T})\right| \Delta S_s \mathrm{d}s \xrightarrow[n\to+\infty]{u.c.a.s.} 0. \tag{3.4.16}$$

Second, from Lemma 3.4.4 we obtain

$$\varepsilon_n^{-1}\int_0^t Q_{\varphi(s)}^\mathsf{T}\mathcal{A}_{\varphi(s)}^{ij}\Delta S_s \mathrm{d}s \xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t Q_s^\mathsf{T}\mathcal{A}_s^{ij}Q_s \mathrm{d}s. \tag{3.4.17}$$

Last, we write $\mathcal{A}_s^{ij} = \mathcal{A}_s^{ij+} - \mathcal{A}_s^{ij-}$ (see Lemma 3.B.1), where $\mathcal{A}_s^{ij+}$ and $\mathcal{A}_s^{ij-}$ are adapted continuous symmetric non-negative definite matrices. Owing to Lemma 3.4.5 we get

$$\varepsilon_n^{-2}\int_0^t \Delta S_s^\mathsf{T}(\mathcal{A}_{\varphi(s)}^{ij})^+\Delta S_s \mathrm{d}s \xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t m_s^{-1}\mathcal{B}_s[f(x) := ((\sigma_s^{-1}x)^\mathsf{T}X_s^{ij+}(\sigma_s^{-1}x))^2]\mathrm{d}s,$$
$$\varepsilon_n^{-2}\int_0^t \Delta S_s^\mathsf{T}(\mathcal{A}_{\varphi(s)}^{ij})^-\Delta S_s \mathrm{d}s \xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t m_s^{-1}\mathcal{B}_s[f(x) := ((\sigma_s^{-1}x)^\mathsf{T}X_s^{ij-}(\sigma_s^{-1}x))^2]\mathrm{d}s,$$

where X_s^{ij+} (resp. X_s^{ij-}) is the solution of the matrix equation (3.2.13) for $c = \sigma_s^\mathsf{T}\mathcal{A}_s^{ij+}\sigma_s$ (resp. $\sigma_s^\mathsf{T}\mathcal{A}_s^{ij-}\sigma_s$). Hence, using that $\mathcal{B}_s[\cdot]$ is linear, we obtain

$$\varepsilon_n^{-2}\int_0^t \Delta S_s^\mathsf{T}\mathcal{A}_{\varphi(s)}^{ij}\Delta S_s \mathrm{d}s \xrightarrow[n\to+\infty]{u.c.a.s.} \int_0^t m_s^{-1}\mathcal{B}_s[f(x) := ((\sigma_s^{-1}x)^\mathsf{T}X_s^{ij+}(\sigma_s^{-1}x))^2 - ((\sigma_s^{-1}x)^\mathsf{T}X_s^{ij-}(\sigma_s^{-1}x))^2]\mathrm{d}s. \tag{3.4.18}$$

Recall the definition (3.2.19), i.e.

$$\mathcal{K}_t^{ij} = m_t^{-1}\mathcal{B}_t\left[f(x) := ((\sigma_t^{-1}x)^\mathsf{T}X_t^{ij+}(\sigma_t^{-1}x))^2 - ((\sigma_t^{-1}x)^\mathsf{T}X_t^{ij-}(\sigma_t^{-1}x))^2)\right] - Q_t^\mathsf{T}\mathcal{A}_t^{ij}Q_t.$$

Thus from (3.4.16), (3.4.17) and (3.4.18) we get the convergence

$$\left\langle \varepsilon_n^{-1}\mathcal{E}^{n,2} - \int_0^{\cdot} Q_s^{\mathsf{T}}\mathcal{A}_s \mathrm{d}B_s \right\rangle_t \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} \int_0^t \mathcal{K}_s \mathrm{d}s. \tag{3.4.19}$$

Note that $\mathcal{K}_s$ is a symmetric non-negative definite matrix since it is the a.s. limit of covariation matrices.

Further we compute the limit for the finite variation part $\mathcal{E}_t^{n,1}$. Owing to Lemma 3.4.4 we directly have

$$\varepsilon_n^{-1}\mathcal{E}_t^{n,1} = \varepsilon_n^{-1}\int_0^t \mathcal{M}_{\varphi(s)}\Delta S_s \mathrm{d}s \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} \int_0^t \mathcal{M}_s Q_s \mathrm{d}s. \tag{3.4.20}$$

For the convergence of $\varepsilon_n^2 N_t^n$ we take advantage of Theorem 3.4.3 to write

$$\varepsilon_n^2 N_t^n = \sum_{\tau_{i-1}^n < t} \varepsilon_n^2 \overset{u.c.a.s.}{\underset{n\to+\infty}{\longrightarrow}} \int_0^t m_s^{-1}\mathrm{d}s. \tag{3.4.21}$$

Now we come back to the initial notation $\varepsilon_{\iota'\circ\iota}$ for the subsequence. Having proved the a.s. convergences (3.4.15), (3.4.19), (3.4.20) and (3.4.21) for $\varepsilon_n = \varepsilon_{\iota'\circ\iota}$, we use the arbitrary choice of $\iota(n)$ and the subsequence principle from Lemma 2.2.2 to get the same convergences in probability along the initial sequence $(\varepsilon_n : n \geq 0)$. So, in particular, we can apply Theorem 3.4.6 with

$$M_t^n = \varepsilon_n^{-1}\mathcal{E}_t^{n,2} - \int_0^t Q_s^{\mathsf{T}}\mathcal{A}_s \mathrm{d}B_s \quad \text{and} \quad A_t^n = \varepsilon_n^{-1}\mathcal{E}_t^{n,1}$$

and after easy manipulations, we obtain the following functional $\bar{\mathcal{F}}$-stable convergence in distribution:

$$\varepsilon_n^{-1}\mathcal{E}_t^n \overset{d}{\underset{[0,T]}{\Longrightarrow}} \left(\int_0^t \mathcal{M}_s Q_s \mathrm{d}s + \int_0^t Q_s^{\mathsf{T}}\mathcal{A}_s \mathrm{d}B_s + \int_0^t \mathcal{K}_s^{1/2}\mathrm{d}W_s \right).$$

The uniform convergence in probability (3.3.5) follows similarly from the $\mathbb{P}$-version of the convergence (3.4.21). The proof of Theorem 3.3.1 is now complete. □

3.5 Proofs of domain exit time properties (Lemma 3.3.3, Propositions 3.3.4 and 3.3.5)

We assume the notation of Section 3.3.2. In particular L_D denotes the constant given by (3.3.7).

3.5.1 Proof of Lemma 3.3.3

We begin by justifying *i)* with $p = 1$. For this we assume without loss of generality that the process S has its coefficients such that

$$\Lambda_{\min}^{\sigma} \leq \inf_{t\geq 0} \lambda_{\min}(\sigma_t\sigma_t^{\mathsf{T}}) \leq \sup_{t\geq 0} \|\sigma_t\sigma_t^{\mathsf{T}}\| \leq \Lambda_{\max}^{\sigma}, \qquad \sup_{t\geq 0} |b_t| \leq b_{\max}. \tag{3.5.1}$$

Indeed, we can still define new $\bar{\mathcal{F}}$-adapted coefficients $\tilde{b}_t = b_t \mathbf{1}_{\tau<t}$ and $\tilde{\sigma}_t = \sigma_t \mathbf{1}_{\tau<t} + \mathbf{1}_{\tau\geq t}\sqrt{\Lambda^\sigma_{\max}/d}$: they satisfy to the above bounds, they coincide with those of S before τ, and therefore the process with new coefficients has the same exit time τ. For the proof of the above lemma, this is enough to consider such a modified process instead of the initial S, or equivalently to assume (3.5.1) for S.
Now, we invoke the rough bound $\tau \leq \bar{\tau} = \inf\{t \geq 0 : |S_{1,t}| \geq \varepsilon L_D^{-1}\}$ which holds since D is included in a ball centered at 0 with radius L_D^{-1}. We now derive two bounds, one for any $\varepsilon \leq \varepsilon_0 \leq 1$, the other for small ε.

1. Take λ as the unique positive solution to $-\lambda b_{\max} + \frac{1}{2}\lambda^2 \Lambda^\sigma_{\min} = 1$: clearly $\lambda \in \mathcal{C}(\mathfrak{S})$; then apply the Itô formula in expectation to get

$$e^{\lambda L_D^{-1}} \geq \mathbb{E}_\nu\left(e^{\lambda S_{1,\bar{\tau}}}\right) = e^{\lambda S_{1,\nu}} + \mathbb{E}_\nu\left(\int_\nu^{\bar{\tau}} e^{\lambda S_{1,s}}(\lambda b^1_s + \frac{1}{2}\lambda^2 |\sigma_{1:,s}|^2)\mathrm{d}s\right)$$
$$\geq \mathbb{E}_\nu\left(\int_\nu^{\bar{\tau}} e^{\lambda S_{1,s}}\mathrm{d}s\right) \geq e^{-\lambda L_D^{-1}}\mathbb{E}_\nu(\bar{\tau} - \nu).$$

   This holds for any $\varepsilon \leq 1$.

2. Now, for $\varepsilon \leq \min(1, \Lambda^\sigma_{\min} L_D/(4b_{\max})) := \bar{\varepsilon}_0 \in \mathcal{C}(\mathfrak{S})$ so that $-2\varepsilon L_D^{-1} b_{\max} + \Lambda^\sigma_{\min} \geq \Lambda^\sigma_{\min}/2$, we have with similar arguments

$$\varepsilon^2 L_D^{-2} \geq \mathbb{E}_\nu\left(S^2_{1,\bar{\tau}}\right) = S^2_{1,\nu} + \mathbb{E}_\nu\left(\int_\nu^{\bar{\tau}} (2S_{1,s}b_{1,s} + |\sigma_{1:,s}|^2)\mathrm{d}s\right) \geq \mathbb{E}_\nu(\bar{\tau} - \nu)\ \Lambda^\sigma_{\min}/2.$$

To summarize, we have justified that for any stopping time ν, a.s. on $\{\nu \leq \tau\}$ we have

$$\mathbb{E}_\nu(\tau - \nu) \leq \mathbb{E}_\nu(\bar{\tau} - \nu) \leq e^{2\lambda/L_D}\mathbf{1}_{\varepsilon>\bar{\varepsilon}_0} + 2\varepsilon^2/(L_D^2\Lambda^\sigma_{\min})\mathbf{1}_{\varepsilon\leq\bar{\varepsilon}_0} \leq \max\left(e^{2\lambda/L_D}/\bar{\varepsilon}_0^2, 2/(L_D^2\Lambda^\sigma_{\min})\right)\varepsilon^2$$
$$=: R_D\varepsilon^2$$

with $R_D \in \mathcal{C}(\mathfrak{S})$.

We now establish *i)* for $p \geq 2$ by induction. Assume that *i)* holds for some $p \geq 1$ and for any stopping time ν: then, on $\{\nu \leq \tau\}$,

$$\mathbb{E}_\nu\Big((\tau-\nu)^{p+1}\Big) = \int_0^\infty (p+1)\mathbb{E}_\nu\Big((\tau-\nu-t)^p\mathbf{1}_{\tau-\nu\geq t}\Big)\mathrm{d}t$$
$$= \int_0^\infty (p+1)\mathbb{E}_\nu\Big(\mathbb{E}_{\nu+t}((\tau-\nu-t)^p)\mathbf{1}_{\tau-\nu\geq t}\Big)\mathrm{d}t$$
$$\leq \int_0^\infty (p+1)\mathbb{E}_\nu\Big(p!(R_D\varepsilon^2)^p\mathbf{1}_{\tau-\nu\geq t}\Big)\mathrm{d}t$$
$$= (p+1)!(R_D\varepsilon^2)^p\mathbb{E}_\nu(\tau-\nu) \leq (p+1)!(R_D\varepsilon^2)^{p+1}$$

using twice the induction assumption (first for the stopping time $\nu+t$ on the event $\{\nu+t \leq \tau\}$, second for ν on the event $\{\nu \leq \tau\}$).

Last we derive *ii)*. On $\{\nu \leq \tau\}$, use the exponential Markov inequality and the estimates

i) to get

$$\mathbb{P}_\nu(\tau-\nu\geq\varepsilon^2 c)\leq\mathbb{E}_\nu\left(e^{\frac{1}{2R_D\varepsilon^2}(\tau-\nu-\varepsilon^2 c)}\right)\leq e^{-\frac{c}{2R_D}}\sum_{p\geq 0}\frac{1}{p!2^p}\mathbb{E}_\nu\left((\frac{\tau-\nu}{R_D\varepsilon^2})^p\right)\leq 2e^{-\frac{c}{2R_D}}.$$

We are done. □

3.5.2 Preparing the proof of Propositions 3.3.4 and 3.3.5

This section is devoted to some preliminary results. Only within this section we assume that

$$\boxed{D\in\mathcal{D}}$$

(we pass to the general case $D\in\mathcal{D}^J_\cap$ in Section 3.5.3). For simplicity we write $\delta(\cdot)$ instead of $\delta_{\partial D}(\cdot)$ since $D\in\mathcal{D}$ is fixed and no confusion is possible. For $\varepsilon>0$ denote $\delta_\varepsilon(x):=\varepsilon\delta(\varepsilon^{-1}x)$.

Lemma 3.5.1. *Assume* $(\mathbf{H}^{D,\sigma}_{loc})$ *with* $D\in\mathcal{D}$ *($J=1$). Let* $\tau_0:=\inf\{t\geq 0: S_t\notin D)\}$. *There exists* $L_\sigma\in\mathcal{C}(\mathfrak{S})$ *such that* $L_\sigma\leq L_D$ *and for any* $t\in[0,\tau_0]$ *we have a.s.*

$$\inf_{0\leq\delta(x)\leq L_\sigma}\mathrm{Tr}(\sigma_t^\mathsf{T}(\nabla\delta^\mathsf{T}\nabla\delta+\delta\nabla^2\delta)(x)\sigma_t)\geq\frac{1}{8}\Lambda^\sigma_{\min}.$$

Proof. Remind of the convention on $\nabla\delta$ as a row vector. By (3.3.7) on the set $|\delta(x)|\leq L_D$ the function $\delta(\cdot)$ is $\mathcal{C}^2$ and $\inf_{0\leq\delta(x)\leq L_D}|\nabla\delta(x)|^2\geq\frac{1}{4}$. For any $x\in D$ such that $|\delta(x)|\leq L_D$ we have

$$\mathrm{Tr}(\sigma_t^\mathsf{T}(\nabla\delta^\mathsf{T}\nabla\delta)(x)\sigma_t)=\nabla\delta(x)^\mathsf{T}\cdot\sigma_t\sigma_t^\mathsf{T}\nabla\delta(x)^\mathsf{T}\geq\frac{1}{4}\Lambda^\sigma_{\min}. \tag{3.5.2}$$

Further using $|\mathrm{Tr}(M)|\sqrt{d}\|M\|$ for $M\in\mathrm{Mat}_{d,d}(\mathbb{R})$ and the sub-multiplicative property of the Frobenius norm, for any $0\leq L\leq L_D$ and $x\in D$ with $|\delta(x)|\leq L$, we have

$$|\mathrm{Tr}(\sigma_t^\mathsf{T}(\delta\nabla^2\delta)(x)\sigma_t)|\leq\sqrt{d}\|\sigma_t\sigma_t^\mathsf{T}(\delta\nabla^2\delta)(x)\|\leq\sqrt{d}\|\sigma_t\sigma_t^\mathsf{T}\|\times\|(\delta\nabla^2\delta)(x)\|\leq\sqrt{d}LL_D^{-1}\Lambda^\sigma_{\max}. \tag{3.5.3}$$

We set $L_\sigma:=L_D\min\left(1,\frac{\Lambda^\sigma_{\min}}{8\sqrt{d}\Lambda^\sigma_{\max}}\right)$, which is a continuous function of L_D, $\Lambda^\sigma_{\min}$ and $\Lambda^\sigma_{\max}$, so that $\frac{1}{4}\Lambda^\sigma_{\min}-\sqrt{d}L_\sigma L_D^{-1}\Lambda^\sigma_{\max}\geq\frac{1}{8}\Lambda^\sigma_{\min}$, which together with (3.5.2) and (3.5.3) implies the announced result. □

Lemma 3.5.2. *Assume* $(\mathbf{H}^{D,\sigma}_{loc})$ *with* $D\in\mathcal{D}$ *($J=1$). There exists* $K\in\mathcal{C}(\mathfrak{S})$ *such that for any* $\varepsilon\in(0,1]$ *and the stopping time*

$$\tau=\inf\{t\geq 0: S_t\notin\varepsilon D\}$$

and any stopping time ν *such that* $\nu\leq\tau$ *a.s. we have*

$$\mathbb{E}(\tau-\nu)\leq K\varepsilon^2\mathbb{E}(\delta(\varepsilon^{-1}S_\nu)). \tag{3.5.4}$$

Proof. Take $\varepsilon \in (0,1]$. Let $L_\sigma \in (0, L_D]$ be given by Lemma 3.5.1 (L_D is defined in (3.3.7)), $l \in (0, L_\sigma]$. We have

$$\mathbb{E}(\tau - \nu) = \mathbb{E}((\tau-\nu)\mathbb{1}_{\delta(\varepsilon^{-1}S_\nu)>l}) + \mathbb{E}((\tau-\nu)\mathbb{1}_{\delta(\varepsilon^{-1}S_\nu)\le l}). \tag{3.5.5}$$

Using Lemma 3.3.3 we get

$$\mathbb{E}((\tau-\nu)\mathbb{1}_{\delta(\varepsilon^{-1}S_\nu)>l}) = \mathbb{E}(\mathbb{1}_{\delta(\varepsilon^{-1}S_\nu)>l}\mathbb{E}_\nu(\tau-\nu)) \le R_D\varepsilon^2\mathbb{P}(\delta(\varepsilon^{-1}S_\nu) > l). \tag{3.5.6}$$

The rest of the proof consists in estimating $\mathbb{1}_{\delta(\varepsilon^{-1}S_\nu)\le l}\mathbb{E}_\nu(\tau-\nu)$. For simplicity we omit the indicator in the calculations, so that *we are working on the event* $\{\delta(\varepsilon^{-1}S_\nu) \le l\}$. Denote $\tau_l := \inf\{t > \nu : \delta_\varepsilon(S_t) \ge l\varepsilon\}$. Note that $\delta(\cdot)$ is $\mathcal{C}^2$ on the set $|\delta(x)| \le l$ since $l \le L_\sigma \le L_D$. Let us write the Itô formula for $\delta_\varepsilon^2(S_t)$ on $[\nu, \tau\wedge\tau_l]$:

$$\begin{aligned}\delta_\varepsilon^2(S_{\tau\wedge\tau_l}) = \delta_\varepsilon^2(S_\nu) &+ 2\int_\nu^{\tau\wedge\tau_l} (\delta_\varepsilon\nabla\delta_\varepsilon)(S_s)\,\mathrm{d}S_s \\ &+ \int_\nu^{\tau\wedge\tau_l} \mathrm{Tr}(\sigma_s^\mathsf{T}(\nabla\delta_\varepsilon^\mathsf{T}\nabla\delta_\varepsilon + \delta_\varepsilon\nabla^2\delta_\varepsilon)(S_s)\sigma_s)\mathrm{d}s.\end{aligned} \tag{3.5.7}$$

Note that by Lemma 3.5.1, $s \le \tau \le \tau_0$, $0 \le \delta(\varepsilon^{-1}S_s) \le l \le L_\sigma$, $(\nabla\delta_\varepsilon^\mathsf{T}\nabla\delta_\varepsilon + \delta_\varepsilon\nabla^2\delta_\varepsilon)(x) = (\nabla\delta^\mathsf{T}\nabla\delta + \delta\nabla^2\delta)(\varepsilon^{-1}x)$ we have for all $s \in [\nu, \tau\wedge\tau_l]$ a.s.

$$\mathrm{Tr}(\sigma_s^\mathsf{T}(\nabla\delta_\varepsilon^\mathsf{T}\nabla\delta_\varepsilon + \delta_\varepsilon\nabla^2\delta_\varepsilon)(S_s)\sigma_s) \ge \frac{1}{8}\Lambda^\sigma_{\min}.$$

So we obtain

$$\int_\nu^{\tau\wedge\tau_l} \mathrm{Tr}(\sigma_s^\mathsf{T}(\nabla\delta_\varepsilon^\mathsf{T}\nabla\delta_\varepsilon + \delta_\varepsilon\nabla^2\delta_\varepsilon)(S_s)\sigma_s)\mathrm{d}s \ge \frac{1}{8}\Lambda^\sigma_{\min}(\tau\wedge\tau_l - \nu). \tag{3.5.8}$$

Further

$$\begin{aligned}\left|\mathbb{E}_\nu\left(\int_\nu^{\tau\wedge\tau_l} (\delta_\varepsilon\nabla\delta_\varepsilon)(S_s)\,\mathrm{d}S_s\right)\right| &= \left|\mathbb{E}_\nu\left(\int_\nu^{\tau\wedge\tau_l} (\delta_\varepsilon\nabla\delta_\varepsilon)(S_s)\,b_s\mathrm{d}s\right)\right| \\ &\le lL_D^{-1}b_{\max}\mathbb{E}_\nu(\tau\wedge\tau_l - \nu).\end{aligned} \tag{3.5.9}$$

Thus from (3.5.7), applying $\mathbb{E}_\nu(\cdot)$, using (3.5.8), (3.5.9) and simply that $\delta_\varepsilon^2(S_\nu) \ge 0$ we get

$$C_1\mathbb{E}_\nu(\tau\wedge\tau_l - \nu) \le \mathbb{E}_\nu(\delta_\varepsilon^2(S_{\tau\wedge\tau_l})),$$

where $C_1 = \frac{1}{16}\Lambda^\sigma_{\min} \in \mathcal{C}(\mathfrak{S})$, for any l satisfying

$$0 < l \le L_\sigma \wedge (L_D\Lambda^\sigma_{\min}b^{-1}_{\max}/16). \tag{3.5.10}$$

We continue with l satisfying (3.5.10). Now using that $\delta_\varepsilon(S_\tau) = 0$ and from the definition of τ_l we get $\mathbb{E}_\nu(\delta_\varepsilon^2(S_{\tau\wedge\tau_l})) = \mathbb{E}_\nu(\delta_\varepsilon^2(S_{\tau_l})\mathbb{1}_{\tau>\tau_l}) = l^2\varepsilon^2\mathbb{P}_\nu(\tau > \tau_l)$, and consequently

$$C_1\mathbb{E}_\nu(\tau\wedge\tau_l - \nu) \le l^2\varepsilon^2\mathbb{P}_\nu(\tau > \tau_l). \tag{3.5.11}$$

Further we write

$$\mathbb{E}_\nu(\tau\wedge\tau_l-\nu)=\mathbb{E}_\nu((\tau-\nu)\mathbb{1}_{\tau<\tau_l})+\mathbb{E}_\nu((\tau_l-\nu)\mathbb{1}_{\tau>\tau_l})=\mathbb{E}_\nu(\tau-\nu)-\mathbb{E}_\nu((\tau-\tau_l)\mathbb{1}_{\tau>\tau_l}). \quad (3.5.12)$$

Using Lemma 3.3.3 (with $R_D\in\mathcal{C}(\mathfrak{S})$) we obtain

$$\mathbb{E}_\nu((\tau-\tau_l)\mathbb{1}_{\tau>\tau_l})=\mathbb{E}_\nu(\mathbb{1}_{\tau>\tau_l}\mathbb{E}_{\tau_l}(\tau-\tau_l))\le R_D\varepsilon^2\mathbb{P}_\nu(\tau>\tau_l). \quad (3.5.13)$$

Hence plugging (3.5.11) and (3.5.13) into (3.5.12) yields

$$\mathbb{E}_\nu(\tau-\nu)\le(R_D+C_1^{-1}l^2)\varepsilon^2\mathbb{P}_\nu(\tau>\tau_l). \quad (3.5.14)$$

Now, we aim at upper bounding the above probability. By taking the conditional expectation $\mathbb{E}_\nu(\cdot)$ of the Itô formula for $\delta_\varepsilon(S_t)$ on $[\nu,\tau\wedge\tau_l]$, we get

$$\begin{aligned} l\varepsilon\mathbb{P}_\nu(\tau>\tau_l)=\mathbb{E}_\nu(\delta_\varepsilon(S_{\tau\wedge\tau_l}))&=\delta_\varepsilon(S_\nu)+\mathbb{E}_\nu\left(\int_\nu^{\tau\wedge\tau_l}\nabla\delta_\varepsilon(S_s)\ b_s\mathrm{d}s\right)\\ &\quad+\frac{1}{2}\mathbb{E}_\nu\left(\int_\nu^{\tau\wedge\tau_l}\mathrm{Tr}(\sigma_s^\mathsf{T}\nabla^2\delta_\varepsilon(S_s)\sigma_s)\mathrm{d}s\right). \end{aligned} \quad (3.5.15)$$

The first expectation in the right-hand side of (3.5.15) is bounded by $L_D^{-1}b_{\max}\mathbb{E}_\nu(\tau\wedge\tau_l-\nu)$, while the second expectation, in view of (3.3.7) and ($\mathbf{H}_{loc}^{D,\sigma}$), is bounded by $\varepsilon^{-1}\sqrt{d}L_D^{-1}\Lambda^\sigma_{\max}\mathbb{E}_\nu(\tau\wedge\tau_l-\nu)$. Therefore, plugging the above into (3.5.15) and using then (3.5.11), we readily obtain

$$\begin{aligned} l\varepsilon^2\mathbb{P}_\nu(\tau>\tau_l)&\le\varepsilon\delta_\varepsilon(S_\nu)+(\frac{1}{2}\sqrt{d}L_D^{-1}\Lambda^\sigma_{\max}+L_D^{-1}b_{\max})\mathbb{E}_\nu(\tau\wedge\tau_l-\nu)\\ &\le\varepsilon\delta_\varepsilon(S_\nu)+\varepsilon^2C_2l^2\mathbb{P}_\nu(\tau>\tau_l), \end{aligned}$$

where $C_2:=(\frac{1}{2}\sqrt{d}L_D^{-1}\Lambda^\sigma_{\max}+L_D^{-1}b_{\max})C_1^{-1}$, so that $C_2\in\mathcal{C}(\mathfrak{S})$. Note that all the previous analysis is valid for any l verifying (3.5.10) and the elements of $\mathcal{C}(\mathfrak{S})$ do not depend on l, so we may now fix $l=l_0:=\min(C_2^{-1}/2,L_\sigma,(L_D\Lambda^\sigma_{\min}b_{\max}^{-1})/16)$ which implies $C_3:=l_0-C_2l_0^2\ge\frac{l_0}{2}>0$. Observe that $l_0,C_3\in\mathcal{C}(\mathfrak{S})$. Hence we obtain

$$\mathbb{P}_\nu(\tau>\tau_{l_0})\le C_3^{-1}\delta(\varepsilon^{-1}S_\nu). \quad (3.5.16)$$

Combining (3.5.14) and (3.5.16) and setting $K:=(R_D+C_1^{-1}l_0^2)C_3^{-1}\in\mathcal{C}(\mathfrak{S})$, we get

$$\mathbb{E}_\nu(\tau-\nu)\le K\varepsilon^2\delta(\varepsilon^{-1}S_\nu). \quad (3.5.17)$$

Remember that this result is obtained on the event $\{\delta(\varepsilon^{-1}S_\nu)\le l_0\}$. Going back to the general notation we have $\mathbb{1}_{\delta(\varepsilon^{-1}S_\nu)\le l_0}\mathbb{E}_\nu(\tau-\nu)\le K\varepsilon^2\mathbb{1}_{\delta(\varepsilon^{-1}S_\nu)\le l_0}\delta(\varepsilon^{-1}S_\nu)$, and then by taking expectation and combining this with (3.5.6) and (3.5.5), we finally obtain

$$\mathbb{E}(\tau-\nu)\le K\varepsilon^2\mathbb{E}(\delta(\varepsilon^{-1}S_\nu))+R_D\varepsilon^2\mathbb{P}(\delta(\varepsilon^{-1}S_\nu)>l_0)\le(K+R_Dl_0^{-1})\varepsilon^2\mathbb{E}(\delta(\varepsilon^{-1}S_\nu))$$

where we have applied the Markov inequality at the last inequality. We are done. □

Lemma 3.5.3. *Assume* $(\mathbf{H}_{loc}^{D,\sigma})$ *with* $D \in \mathcal{D}$ $(J = 1)$*, and let* $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ *be an* α*-homogeneous function with* $\alpha \in \{2, 3, 4\}$*. There exists* $K \in \mathcal{C}(\mathfrak{S})$ *such that for any* $\varepsilon \in (0, 1]$*, for the stopping times*

$$\tau = \inf\{t \geq 0 : S_t \notin \varepsilon D\}, \quad \bar{\tau} = \inf\{t \geq 0 : \bar{S}_t \notin \varepsilon D\}$$

and any stopping time ν *such that* $\nu \leq \tau \wedge \bar{\tau}$ a.s.*, we have*

$$\varepsilon^{-2}\mathbb{E}(|S_\nu - \bar{S}_\nu|^2) + \varepsilon^{-\alpha}|\mathbb{E}(f(S_\nu) - f(\bar{S}_\nu))| \leq K\varepsilon^{\eta_\sigma}. \tag{3.5.18}$$

Proof. We start with a bound on $\mathbb{E}(|S_\nu - \bar{S}_\nu|^2)$:

$$\begin{aligned}\mathbb{E}\left(|S_\nu - \bar{S}_\nu|^2\right) &\leq \mathbb{E}\left(\sum_{k=1}^{+\infty} \mathbb{1}_{\nu/\varepsilon^2 \in [k-1,k)} \sup_{t \leq k\varepsilon^2} |\bar{S}_t - S_t|^2\right) \\ &\leq \sum_{k=1}^{+\infty} \mathbb{P}(\nu/\varepsilon^2 \in [k-1,k))^{1/2} \left[\mathbb{E}\left(\sup_{t \leq k\varepsilon^2} |\bar{S}_t - S_t|^4\right)\right]^{1/2}.\end{aligned}$$

a) Estimate for $\mathbb{E}\left(\sup_{t \leq k\varepsilon^2} |\bar{S}_t - S_t|^4\right)$*:* Denote

$$\bar{M}_t := \int_0^t (\sigma_s - \sigma_0)\mathrm{d}W_s$$

so that $S_t - \bar{S}_t = \int_0^t b_s \mathrm{d}s + \bar{M}_t$. Using the BDG inequalities and $(\mathbf{H}_{loc}^{D,\sigma})$ we obtain

$$\begin{aligned}\mathbb{E}\left(\sup_{t \leq k\varepsilon^2} |S_t - \bar{S}_t|^4\right) &\leq 8\left(b_{\max}^4 (k\varepsilon^2)^4 + \mathbb{E}\left(\sup_{t \leq k\varepsilon^2} |\bar{M}_t|^4\right)\right) \\ &\leq C\left(b_{\max}^4 (k\varepsilon^2)^4 + \mathbb{E}\left(\langle \bar{M}\rangle_{k\varepsilon^2}^2\right)\right),\end{aligned}$$

where C is some universal constant. For the quadratic variation part we get

$$\mathbb{E}\left(\langle \bar{M}\rangle_{k\varepsilon^2}^2\right) = \mathbb{E}\left(\left(\int_0^{k\varepsilon^2} |\sigma_t - \sigma_0|^2 \mathrm{d}t\right)^2\right) \leq \mathbb{E}\left(C_\sigma^4\right)\left(\int_0^{k\varepsilon^2} t^{\eta_\sigma}\mathrm{d}t\right)^2 = C_0(k\varepsilon^2)^{2(\eta_\sigma+1)},$$

with $C_0 := \frac{m_\sigma}{(\eta_\sigma+1)^2}$. So we conclude, using that $k \geq 1, \varepsilon \leq 1$,

$$\mathbb{E}\left(\sup_{t \leq k\varepsilon^2} |S_t - \bar{S}_t|^4\right) \leq C_1 k^4 \varepsilon^{2(2+2\eta_\sigma)}, \tag{3.5.19}$$

where $C_1 := C(b_{\max}^4 + C_0) \in \mathcal{C}(\mathfrak{S})$.

b) Estimate for $\mathbb{P}(\nu/\varepsilon^2 \in [k-1,k))^{1/2}$*:* Lemma 3.3.3-ii) directly yields

$$\mathbb{P}(\nu/\varepsilon^2 \in [k-1,k)) \leq \mathbb{P}(\nu \geq \varepsilon^2(k-1)) \leq R_D e^{-R_D(k-1)}$$

for some $R_D \in \mathcal{C}(\mathfrak{S})$ Hence combining this with (3.5.19) we get

$$\sum_{k=1}^{+\infty} \mathbb{P}(\nu/\varepsilon^2 \in [k-1,k))^{1/2} \left[\mathbb{E}\left(\sup_{t\le k\varepsilon^2} |S_t - \bar{S}_t|^4\right)\right]^{1/2} \le \sqrt{R_D C_1}\left(\sum_{k=1}^{+\infty} e^{-R_D(k-1)/2} k^2\right)\varepsilon^{2+2\eta_\sigma}.$$

Thus for $K = \sqrt{R_D C_1}\left(\sum_{k=1}^{+\infty} e^{-R_D(k-1)/2}k^2\right)$ (so that $K \in \mathcal{C}(\mathfrak{S})$), we get

$$\mathbb{E}_\nu(|S_\nu - \bar{S}_\nu|^2) \le K\varepsilon^{2+2\eta_\sigma}. \tag{3.5.20}$$

Now we proceed with the proof of (3.5.18) regarding f. Recall that the function f verifies (3.3.9). We have

$$\begin{aligned}|\mathbb{E}\left(f(S_\nu) - f(\bar{S}_\nu)\right)| &\le \mathbb{E}\left(|S_\nu - \bar{S}_\nu| \int_0^1 |\nabla f(\lambda S_\nu + (1-\lambda)\bar{S}_\nu)|\mathrm{d}\lambda\right)\\ &\le \left[\mathbb{E}\left(\left(\int_0^1 |\nabla f(\lambda S_\nu + (1-\lambda)\bar{S}_\nu)|\mathrm{d}\lambda\right)^2\right)\right]^{1/2}\left[\mathbb{E}\left(|S_\nu - \bar{S}_\nu|^2\right)\right]^{1/2}.\end{aligned}$$

Using that $\nu \le \tau \wedge \bar{\tau}$ we obtain $|S_\nu| \le \varepsilon L_D^{-1}$ and $|\bar{S}_\nu| \le \varepsilon L_D^{-1}$ so that

$$\left[\mathbb{E}\left(\left(\int_0^1 |\nabla f(\lambda S_\nu + (1-\lambda)\bar{S}_\nu)|\mathrm{d}\lambda\right)^2\right)\right]^{1/2} \le C_f\ L_D^{-(\alpha-1)}\varepsilon^{\alpha-1}. \tag{3.5.21}$$

Now combine (3.5.20) and (3.5.21) to get (up to changing $K \in \mathcal{C}(\mathfrak{S})$) the announced estimate. □

Corollary 3.5.4. *Assume* ($\mathbf{H}_{loc}^{D,\sigma}$) *with* $D \in \mathcal{D}$ $(J = 1)$. *There exists* $K \in \mathcal{C}(\mathfrak{S})$ *such that for any* $\varepsilon \in (0,1]$, *the stopping times*

$$\tau = \inf\{t \ge 0 : S_t \notin \varepsilon D\}, \quad \bar{\tau} = \inf\{t \ge 0 : \bar{S}_t \notin \varepsilon D\}$$

satisfy

$$\mathbb{E}(|\tau - \bar{\tau}|) \le K\varepsilon^{2+\eta_\sigma}, \tag{3.5.22}$$

Proof. Let $\nu := \tau \wedge \bar{\tau}$. Applying Lemma 3.5.2, we get for some $K \in \mathcal{C}(\mathfrak{S})$

$$\mathbb{E}\left(\tau - \nu\right) \le K\varepsilon^2 \mathbb{E}\left(\delta(\varepsilon^{-1}S_\nu)\right). \tag{3.5.23}$$

Using that $\mathbb{1}_{\nu<\tau}\delta(\varepsilon^{-1}\bar{S}_\nu) = 0$ and $\mathbb{1}_{\nu=\tau}\delta(\varepsilon^{-1}S_\nu) = 0$ we write

$$\mathbb{E}\left(\delta(\varepsilon^{-1}S_\nu)\right) = \mathbb{E}\left(\mathbb{1}_{\nu<\tau}(\delta(\varepsilon^{-1}S_\nu) - \delta(\varepsilon^{-1}\bar{S}_\nu))\right) \le L_D^{-1}\varepsilon^{-1}\mathbb{E}(|S_\nu - \bar{S}_\nu|^2)^{1/2}.$$

Using (3.5.18) from Lemma 3.5.3 we get $\varepsilon^2\mathbb{E}(\delta(\varepsilon^{-1}S_\nu)) \le L_D^{-1}K^{1/2}\varepsilon^{2+\eta_\sigma/2}$. In view of (3.5.23), we have proved (up to redefining $K \in \mathcal{C}(\mathfrak{S})$)

$$\mathbb{E}(\mathbb{1}_{\tau>\bar{\tau}}(\tau - \bar{\tau})) = \mathbb{E}(\tau - \nu) \le K\varepsilon^{2+\eta_\sigma}.$$

A similar bound holds for $\mathbb{E}(\mathbb{1}_{\tau<\bar\tau}(\bar\tau-\tau))$: this is justified in the same way, applying Lemma 3.5.2 to $\bar S$ and Lemma 3.5.3. Consequently, the proof of the bound for $\mathbb{E}(|\tau-\bar\tau|)$ is complete. □

Corollary 3.5.5. *Assume* $(\mathbf{H}^{D,\sigma}_{loc})$ *with* $D\in\mathcal{D}$ *(*$J=1$*). There exists* $K\in\mathcal{C}(\mathfrak{S}\cup\{K'\})$ *such that for any* $\varepsilon\in(0,1]$*, any strictly positive constants* K',η' *and for* $D'\in\mathcal{D}$ *such that* $\mu(D,D')\le K'\varepsilon^{\eta'}$*, and for which (3.3.7) and* $(\mathbf{H}^{D,\sigma}_{loc})$ *hold for* D' *instead of* D *with the same constants* $L_D,\Lambda^\sigma_{\min},\Lambda^\sigma_{\max},b_{\max}$*, we have*

$$\mathbb{E}(|\tau-\tau'|)\le K\varepsilon^{2+\eta'} \tag{3.5.24}$$

where

$$\tau=\inf\{t\ge0: S_t\notin\varepsilon D\},\quad \tau'=\inf\{t\ge0: S_t\notin\varepsilon D'\}.$$

In particular, K *is a multiple of* K'*, so that* $K\to0$ *as* $K'\to0$*.*

Proof. Let $\nu:=\tau\wedge\tau'$ and denote by $\delta(\cdot)$ the distance $\delta_{\partial D}(\cdot)$. Using Lemma 3.5.2, we obtain for some $K\in\mathcal{C}(\mathfrak{S})$

$$\mathbb{E}(\tau-\nu)\le K\varepsilon^2\mathbb{E}(\delta(\varepsilon^{-1}S_\nu)).$$

Observe that $\delta(\varepsilon^{-1}S_\nu)\le\mu(D,D')\le K'\varepsilon^{\eta'}$, which gives

$$\mathbb{E}(\mathbb{1}_{\tau\ge\tau'}(\tau-\tau'))=\mathbb{E}(\tau-\nu)\le KK'\varepsilon^{2+\eta'}.$$

A similar bound on $\mathbb{E}(\mathbb{1}_{\tau'\ge\tau}(\tau'-\tau))$ follows from the symmetry between D and D'. □

3.5.3 Proofs of Propositions 3.3.4 and 3.3.5

Now we pass to the general case of $D\in\mathcal{D}^J_\cap$, i.e. of the form $D=\cap_{j=1}^J D_j$. Note that the results of Section 3.5.2 are valid for each $D_j, j=1,\dots,J$.

Proof of Proposition 3.3.4. Let $\nu:=\tau\wedge\bar\tau$. Denote for $j=1,\dots,J$

$$\tau_j=\inf\{t\ge0: S_t\notin\varepsilon D_j\},\quad \bar\tau_j=\inf\{t\ge0:\bar S_t\notin\varepsilon D_j\},$$

so that $\tau=\min(\tau_1,\dots,\tau_J)$ and $\bar\tau=\min(\bar\tau_1,\dots,\bar\tau_J)$. Write

$$\begin{aligned}|\mathbb{E}(f(S_{\tau\wedge T})-f(\bar S_{\bar\tau\wedge T}))|\le{}&|\mathbb{E}(f(S_{\nu\wedge T})-f(\bar S_{\nu\wedge T}))|+|\mathbb{E}(f(S_{\tau\wedge T})-f(S_{\nu\wedge T}))|\\&+|\mathbb{E}(f(\bar S_{\bar\tau\wedge T})-f(\bar S_{\nu\wedge T}))|.\end{aligned}$$

By Lemma 3.5.3 (applied for any j to the domain D_j and the stopping time $\nu\wedge T\le\tau_j\wedge\bar\tau_j$) we have for some $K\in\mathcal{C}(\mathfrak{S})$

$$\varepsilon^{-\alpha}|\mathbb{E}(f(S_{\nu\wedge T})-f(\bar S_{\nu\wedge T}))|\le K\varepsilon^{\eta_\sigma}. \tag{3.5.25}$$

For the next term we have (using that $\varepsilon \leq 1$)

$$|\mathbb{E}(f(S_{\tau\wedge T}) - f(S_{\nu\wedge T}))| \leq \mathbb{E}\left(\left|\int_{\nu\wedge T}^{\tau\wedge T} [\nabla f(S_t)b_t + \frac{1}{2}\operatorname{Tr}(\sigma_t^{\mathsf{T}}\nabla^2 f(S_t)\sigma_t)]\mathrm{d}t\right|\right)$$
$$\leq C_f(b_{\max}L_D^{-(\alpha-1)} + \frac{1}{2}\sqrt{d}\Lambda_{\max}^{\sigma}L_D^{-(\alpha-2)})\varepsilon^{\alpha-2}\mathbb{E}(|\tau\wedge T - \nu\wedge T|)$$
$$\leq C_f(b_{\max}L_D^{-(\alpha-1)} + \frac{1}{2}\sqrt{d}\Lambda_{\max}^{\sigma}L_D^{-(\alpha-2)})\varepsilon^{\alpha-2}\sum_{j=1}^{J}\mathbb{E}(|\tau_j - \bar{\tau}_j|)$$

(since the min function is Lipschitz)

$$\leq C_f(b_{\max}L_D^{-(\alpha-1)} + \frac{1}{2}\sqrt{d}\Lambda_{\max}^{\sigma}L_D^{-(\alpha-2)})\varepsilon^{\alpha-2}J\ K\ \varepsilon^{2+\eta_\sigma}$$

where we have applied Corollary 3.5.4 at the last inequality. We can show a similar bound for $\bar{S}$ and at the end, we obtain the advertised inequality (3.3.10). □

Proof of Proposition 3.3.5. The proof is quite similar to that of Proposition 3.3.4, at the end we invoke Corollary 3.5.5 instead of Corollary 3.5.4. □

Appendix

3.A Technical proofs

3.A.1 Proof of Lemma 3.3.6

We start with some preliminary analysis. Let $(U_i, i \geq 0)$ be i.i.d. random variables uniformly distributed on $[0,1]$ and independent of $\bar{\mathcal{F}}_T$. We keep the same notation for the extended probability space supporting these extra random variables and we simply write $\mathbb{P}_T(\cdot)$ (resp. $\mathbb{E}_T(\cdot)$) for the probability (resp. expectation) conditionally on $\bar{\mathcal{F}}_T$.
Set $V_j = G_*(U_j)$, $j \geq 0$ where $G_*(\cdot)$ is given by ($\mathbf{H}_G$): conditionally on $\bar{\mathcal{F}}_T$, these random variables are i.i.d. Let Y be the random variable given by

$$Y := \inf\{m \geq 1 : \sum_{j=1}^{m} V_j \geq T\}.$$

In view of ($\mathbf{H}_G$) there exists an a.s. finite $\bar{\mathcal{F}}_T$-measurable random integer m_0 such that a.s. we have

$$\gamma := \mathbb{P}_T\left((V_1 + \cdots + V_{m_0}) < T\right) < 1.$$

Our goal is to show that Y has finite (conditional) moments. We write for all $p > 0$

$$\begin{aligned}\mathbb{E}_T(|Y|^p) &\leq \sum_{k\geq 1} k^p \mathbb{P}_T\left(\sum_{j=1}^{k-1} V_j < T\right) \leq \sum_{k\geq 1} k^p \mathbb{P}_T\left((V_1+\cdots+V_{m_0}) < T\right)^{\lfloor (k-1)/m_0\rfloor} \\ &= \sum_{k\geq 1} k^p \gamma^{\lfloor (k-1)/m_0\rfloor} < +\infty.\end{aligned} \tag{3.A.1}$$

We now come back to the main point about proving (3.3.12). For any $n \geq 0$ the grid $\mathcal{T}^n$ may be represented as a union $\mathcal{T}^{n,1} \cup \mathcal{T}^{n,2}$ (possibly non-disjoint), where $\mathcal{T}^{n,1}$ is the grid points with $\tau_i^n = \tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}$ and $\mathcal{T}^{n,2}$ contains the points where exit times occur first (see (3.2.6)). We have $N^n(I_n) \leq N^{n,1}(I_n) + N^{n,2}(I_n)$ with respect to the decomposition $\mathcal{T}^n = \mathcal{T}^{n,1} \cup \mathcal{T}^{n,2}$.

▷ *Upper bound on* $N^{n,1}(I_n)$. Note that from (3.A.1) we get $\mathbb{E}_T(Y) < +\infty$ a.s. Set $C := 1 + 2\mathbb{E}_T(Y)$, let $(Y_i)_{i\geq 0}$ be i.i.d. copies of Y conditionally on $\bar{\mathcal{F}}_T$, and put $m_n := \lceil \varepsilon_n^{-2}|I_n|/T\rceil \to +\infty$. Let $\preceq$ denote the relation of first-order stochastic domination (conditionally on $\bar{\mathcal{F}}_T$). Then using ($\mathbf{H}_G$)-1 and the subadditivity property of counting processes we obtain

$$N^{n,1}(I_n) \preceq \inf\{m \geq 0 : \sum_{j=1}^{m} V_j \geq \varepsilon_n^{-2}|I_n|\} \preceq \sum_{i=1}^{m_n} Y_i.$$

Remark that the latter relation of domination turns into equality in distribution in the particular case of V_j having an exponential distribution due to the additivity of Poisson variables.

Let $p := 2/\rho \geq 2$ for ρ in (3.3.11). Note that from (3.3.11) we have $\varepsilon_n^2/|I_n| \leq C_0 \varepsilon_n^{2\rho}$ so that $\sum_{n\geq 0}(\varepsilon_n^2/|I_n|)^{p/2} < +\infty$. Applying the Markov inequality, the Burkholder inequality (see e.g. [HH80, Theorem 2.10]) and the Minkowsky inequality we obtain (for n large enough so that $m_n \geq 2$)

$$\begin{aligned}&\mathbb{P}_T\Big(T|I_n|^{-1}\varepsilon_n^2 N^{n,1}(I_n) \geq C\Big) \\ &\quad\leq \mathbb{P}_T\left(\frac{\sum_{i=1}^{m_n} Y_i}{m_n - 1} \geq C\right) \leq \mathbb{P}_T\left(\frac{\sum_{i=1}^{m_n}(Y_i - \mathbb{E}_T(Y))}{m_n - 1} \geq 1\right) \\ &\quad\leq \mathbb{E}_T\left(\left|\frac{\sum_{i=1}^{m_n}(Y_i - \mathbb{E}_T(Y))}{m_n - 1}\right|^p\right) \leq C_{\text{Burk.}}\, m_n^{-p}\mathbb{E}_T\left(\left|\sum_{i=1}^{m_n}(Y_i - \mathbb{E}_T(Y))^2\right|^{p/2}\right) \\ &\quad\leq C_{\text{Burk.}} m_n^{-p}\left(\sum_{i=1}^{m_n}\mathbb{E}_T(|Y_i - \mathbb{E}_T(Y)|^p)^{2/p}\right)^{p/2} = C_{\text{Burk.}} m_n^{-p/2}\mathbb{E}_T(|Y - \mathbb{E}_T(Y)|^p)^{2/p} \\ &\quad\leq C_{\text{Burk.}}\mathbb{E}_T(|Y - \mathbb{E}_T(Y)|^p)^{2/p}\left(\frac{T\varepsilon_n^2}{|I_n|}\right)^{p/2}.\end{aligned}$$

So we get

$$\sum_{n\geq 0}\mathbb{P}_T(T|I_n|^{-1}\varepsilon_n^2 N^{n,1}(I_n) \geq C) < +\infty \text{ a.s.}$$

and thus, by the Borel-Cantelli lemma, the event $\{T|I_n|^{-1}\varepsilon_n^2 N^{n,1}(I_n) \geq C\}$ occurs finitely many times conditionally on $\bar{\mathcal{F}}_T$ a.s. This proves $\sup_{n\geq 0}\varepsilon_n^2|I_n|^{-1}N^{n,1}(I_n) \leq C_1$ a.s. for some a.s. finite C_1.

▷ *Upper bound on* $N^{n,2}(I_n)$. Denote $r_* := \inf_{0\leq t\leq T}\sup\{r \geq 0 : B_d(0,r) \subset \cap_{n\geq 0} D_t^n\}$. Let us show that $r_* > 0$ a.s. Indeed for any $n \geq 0$, we have $\inf_{0\leq t\leq T}\sup\{r \geq 0 : B_d(0,r) \subset D_t^n\} > 0$ since each D_t^n contains 0 and in view of the time-continuity of D_t^n w.r.t. the distance $\mu^J(\cdot,\cdot)$. The same holds for $(D_t)_{0\leq t\leq T}$. Now the positivity of r_* follows from the convergence of D_t^n to D_t w.r.t. $\mu^J(\cdot,\cdot)$ uniformly in $t \in [0,T]$ by $(\mathbf{H}_D^1)$. For $N^{n,2}(I_n)$, we write

$$\begin{aligned} N^{n,2}(I_n)\varepsilon_n^2 &\leq \varepsilon_n^2 + r_*^{-2}\sum_{\tau_i^n\in\mathcal{T}^{n,2}\cap I_n,\tau_{i-1}^n\in I_n}\left|S_{\tau_i^n} - S_{\tau_{i-1}^n}\right|^2 \\ &\leq \varepsilon_n^2 + r_*^{-2}\sum_{\tau_i^n\in\mathcal{T}^{n}\cap I_n,\tau_{i-1}^n\in I_n}\left|S_{\tau_i^n} - S_{\tau_{i-1}^n}\right|^2. \end{aligned}$$

We have

$$\sum_{\tau_i^n\in\mathcal{T}^{n}\cap I_n,\tau_{i-1}^n\in I_n}|\Delta S_{\tau_i^n}|^2 - \int_{I_n}\mathrm{Tr}(\sigma_t\sigma_t^\mathsf{T})\mathrm{d}t = 2\int_{I_n}\Delta S_t^\mathsf{T}\sigma_t\mathrm{d}B_t + 2\int_{I_n}\Delta S_t^\mathsf{T}b_t\mathrm{d}t.$$

Further, using $(\mathbf{H}_R)$-1, we obtain that there exists an a.s. finite random variable C such that

$$\left|\int_{I_n}\Delta S_t^\mathsf{T}b_t\mathrm{d}t\right| \leq C\varepsilon_n|I_n| \quad \text{and} \quad \left|\int_{I_n}\Delta S_t^\mathsf{T}\sigma_t\mathrm{d}B_t\right| \leq C\varepsilon_n^{1-\rho}\sqrt{|I_n|},$$

where for the last inequality we apply [GL14a, Corollary 2.1] for the sequence of martingales

$$M_t^n := \frac{\varepsilon_n^{\rho-1}}{\sqrt{|I_n|}}\int_0^t \mathbb{1}_{I_n}(s)\Delta S_s^\mathsf{T}\sigma_s\mathrm{d}B_s,$$

for the parameter $p := 2/\rho$ with ρ given by (3.3.11), in view of the quadratic variation bound

$$\langle M^n\rangle_T = \frac{\varepsilon_n^{2\rho-2}}{|I_n|}\left|\int_{I_n}\Delta S_t^\mathsf{T}\sigma_t\sigma_t^\mathsf{T}\Delta S_t\mathrm{d}t\right| \leq C\varepsilon_n^{4/p}, \qquad \sum_{n\geq 0}\langle M^n\rangle_T^{p/2} < +\infty \quad \text{a.s.}.$$

Using that $\varepsilon_n^{1-\rho}/\sqrt{|I_n|} \to 0$ by (3.3.11), this finally implies

$$N^{n,2}(I_n) \leq 1 + r_*^{-2}\varepsilon_n^{-2}\left(|I_n|\sup_{0\leq t\leq T}\mathrm{Tr}(\sigma_t\sigma_t^\mathsf{T}) + o_n^{\text{a.s.}}(|I_n|)\right),$$

which finishes the proof. □

3.A.2 Proof of Proposition 3.4.1

First let us prove the statement for $f_t = f$, for any $t \in [0,T]$, where $f : \mathbb{R}^d \to \mathbb{R}$ is a continuous α-homogeneous deterministic function. Let $C_f := \sup_{|x|=1}|f(x)|$ and $C^{(3.3.2)}$ be given by $(\mathbf{H}_R)$. First note that from $(\mathbf{H}_R)$ and the homogeneity of f we have for all $n \geq 0$

and for all $t \in [0,T]$ a.s.

$$|f(\Delta S_t)| + |\mathbb{E}_t(f(\Delta S_{\bar{\varphi}(t)}))| \leq C_f C^{(3.3.2)}_{\varphi(t)} \varepsilon_n^\alpha. \tag{3.A.2}$$

Fix $n \geq 0$. Consider the adapted process

$$Z^n_t := \sum_{\tau^n_{i-1}<t} H_{\tau^n_{i-1}} \left(\mathbb{E}_t(f(\Delta S_{\tau^n_i})) - \mathbb{E}_{\tau^n_{i-1}}(f(\Delta S_{\tau^n_i})) \right)$$

(note that the conditional expectations are well defined, see our conventions at the end of the introduction). Define the process

$$\bar{V}_t := C^{(3.3.2)}_t + \sup_{0\leq s\leq t} |H_s| + \left(2C_f C^{(3.3.2)}_t \sup_{0\leq s\leq t} |H_s| + 1 \right)^2 \left(1 + \sum_{n\geq 0} \varepsilon_n^4 N^n_t \right).$$

Note that $\bar{V}_t$ takes finite values due to ($\mathbf{H}_R$)-2 and is adapted càdlàg and non-decreasing. Define

$$\nu_k := \inf\{t \geq 0 : \bar{V}_t \geq k\} \tag{3.A.3}$$

(with the convention $\nu_k = +\infty$ a.s. if $k > \bar{V}_T$). Due to boundedness of H_0 and $C^{(3.3.2)}_0$ we have that

$$\bar{V}_0 = C^{(3.3.2)}_0 + |H_0| + \left(2C_f C^{(3.3.2)}_0 |H_0| + 1 \right)^2 \leq C_{\bar{V}_0}$$

for some deterministic constant $C_{\bar{V}_0}$. Now observe that (since the jumps of N^n_t are of size 1)

$$\bar{V}_{\nu_k} \leq \left(1 + \sum_{n\geq 0} \varepsilon_n^4 \right) (k \vee C_{\bar{V}_0}) =: q(k). \tag{3.A.4}$$

In order to justify the manipulations with the conditional expectations below we remark the following properties

$$\mathbb{1}_{\tau^n_{i-1}\leq \nu^k} |\Delta S_{\tau^n_i}|^\alpha \leq C^{(3.3.2)}_{\nu^k} \varepsilon_n^\alpha \leq q(k)\varepsilon_n^\alpha, \qquad \mathbb{1}_{\tau^n_{i-1}\leq \nu^k} |H_{\tau^n_{i-1}}| \leq q(k) \qquad \text{a.s. .} \tag{3.A.5}$$

It implies that for any stopping time θ and any continuous function Φ we have the equality

$$\mathbb{1}_{\tau^n_{i-1}\leq \nu^k\wedge\theta} \mathbb{E}_{\nu^k\wedge\theta} \Big(\mathbb{1}_{\tau^n_{i-1}\leq \nu^k\wedge\theta} \Phi(\Delta S_{\tau^n_i}) \Big) = \mathbb{1}_{\tau^n_{i-1}\leq \nu^k\wedge\theta} \mathbb{E}_{\nu^k\wedge\theta} \Big(\Phi(\Delta S_{\tau^n_i}) \Big). \tag{3.A.6}$$

Owing to (3.A.5), the random variable inside the conditional expectation on the left hand side is bounded, and therefore its conditional expectation is well-defined (and in any L^p). The random variable inside the conditional expectation on the right hand-side is not necessarily integrable (essentially controlled thanks to ($\mathbf{H}_R$)), but actually, in the next computations, it will be still localised on a set of the form $\{\tau^n_{i-1} \leq \nu^k \wedge \theta\}$, on which we have the equality (3.A.6). Therefore, in what follows, writing $\mathbb{1}_{\tau^n_{i-1}\leq \nu^k\wedge\theta} \mathbb{E}_{\nu^k\wedge\theta} \Big(\Phi(\Delta S_{\tau^n_i}) \Big)$ or $\mathbb{1}_{\tau^n_{i-1}\leq \nu^k\wedge\theta} \mathbb{E}_{\nu^k\wedge\theta} \Big(\mathbb{1}_{\tau^n_{i-1}\leq \nu^k\wedge\theta} \Phi(\Delta S_{\tau^n_i}) \Big)$ is the same and gives random variables that are bounded: for the sake of brevity, we use the notation on the left hand side of (3.A.6).

For $\tau^n_{i-1} < t \wedge \nu_k$ we obtain

$$|H_{\tau^n_{i-1}}| \left(|\mathbb{E}_{t\wedge\nu_k}(f(\Delta S_{\tau^n_i}))| + |\mathbb{E}_{\tau^n_{i-1}}(f(\Delta S_{\tau^n_i}))| \right) \leq \left(\sup_{0\leq s\leq \nu^k} |H_s| \right) 2C_f C^{(3.3.2)}_{\nu^k} \varepsilon^\alpha_n \leq \sqrt{q(k)}\varepsilon^\alpha_n. \tag{3.A.7}$$

Using in addition that $\varepsilon^4_n N^n_{t\wedge\nu_k} \leq q(k)$, we obtain a.s.

$$\begin{aligned} |Z^n_{t\wedge\nu_k}| &\leq \sum_{\tau^n_{i-1}<t\wedge\nu_k} |H_{\tau^n_{i-1}}| \left(|\mathbb{E}_{t\wedge\nu_k}(f(\Delta S_{\tau^n_i}))| + |\mathbb{E}_{\tau^n_{i-1}}(f(\Delta S_{\tau^n_i}))| \right) \leq N^n_{t\wedge\nu_k} \sqrt{q(k)}\varepsilon^\alpha_n \\ &\leq q(k)^{3/2}\varepsilon^{\alpha-4}_n. \end{aligned} \tag{3.A.8}$$

Hence, we get that $\mathbb{E}(|Z^n_{t\wedge\nu_k}|^p) < +\infty$ a.s. for all $p \geq 1$ with an L^p-norm bound independent of $t \in [0,T]$. Using (3.A.5)-(3.A.6) to deal with the conditional expectations and (3.A.8) to be able to interchange the sum and the conditional expectation below, we verify that for any $0 \leq s \leq t \leq T$ we have a.s.

$$\begin{aligned} \mathbb{E}_s(Z^n_{t\wedge\nu^k}) = &\sum_{\tau^n_{i-1}<s\wedge\nu^k} H_{\tau^n_{i-1}} \left(\mathbb{E}_{s\wedge\nu^k}(f(\Delta S_{\tau^n_i})) - \mathbb{E}_{\tau^n_{i-1}}(f(\Delta S_{\tau^n_i})) \right) \\ &+ \mathbb{E}_s \left(\sum_{s\wedge\nu^k\leq\tau^n_{i-1}<t\wedge\nu^k} H_{\tau^n_{i-1}} \left(\mathbb{E}_{t\wedge\nu^k}(f(\Delta S_{\tau^n_i})) - \mathbb{E}_{\tau^n_{i-1}}(f(\Delta S_{\tau^n_i})) \right) \right) = Z^n_{s\wedge\nu^k}. \end{aligned}$$

Hence the process $(Z^n_{t\wedge\nu^k})_{0\leq t\leq T}$ is a martingale, and, in particular, it has a càdlàg modification. Using that $\nu^k = +\infty$ for $k > \bar{V}_T$ we deduce that the process $(Z^n_t)_{0\leq t\leq T}$ is cádlág.

In view of (3.A.2), the final result will follow from the convergence $\varepsilon^{2-\alpha}_n Z^n_t \xrightarrow[n\to+\infty]{u.c.a.s.} 0$. We prove it by leveraging Lemma 3.B.2. Define

$$U^n_t := \varepsilon^{4-2\alpha}_n \sup_{0\leq s\leq t} |Z^n_s|^2, \qquad V^n_t := \varepsilon^4_n N^n_t \left(2C_f C^{(3.3.2)}_t \sup_{0\leq s\leq t} |H_s| \right)^2.$$

Since N^n and Z^n are cádlág, it readily follows that U^n and V^n are càdlàg adapted processes, non-decreasing, vanishing at 0. Note that

$$\sum_{n\geq 0} V^n_t \leq \left(2C_f C^{(3.3.2)}_t \sup_{0\leq s\leq t} |H_s| \right)^2 \sum_{n\geq 0} \varepsilon^4_n N^n_t \leq \bar{V}_t. \tag{3.A.9}$$

Let us check the hypotheses (i)-(ii)-(iii) of Lemma 3.B.2. The assumptions (i)-(ii) follow from (3.A.9). We have already proved (iv) in (3.A.4). Now, we check the relation of domination (iii). We need to show that for some (deterministic) constant $C_0 > 0$ we have, uniformly in k and n,

$$\mathbb{E}(U^n_{t\wedge\nu_k}) \leq C_0 \mathbb{E}(V^n_{t\wedge\nu_k}). \tag{3.A.10}$$

We proceed with the following estimate of $\mathbb{E}(|Z^n_{t\wedge\nu_k}|^2)$ using Fubini's theorem

$$\begin{aligned}\mathbb{E}\left(|Z^n_{t\wedge\nu_k}|^2\right) &= \mathbb{E}\left(\sum_{\tau^n_{i-1}<t\wedge\nu_k} H^2_{\tau^n_{i-1}}\left(\mathbb{E}_{t\wedge\nu_k}(f(\Delta S_{\tau^n_i})) - \mathbb{E}_{\tau^n_{i-1}}(f(\Delta S_{\tau^n_i}))\right)^2\right)\\ &\quad + 2\sum_{1\le i<j<+\infty}\mathbb{E}\Big(1_{\tau^n_{j-1}<t\wedge\nu_k} H_{\tau^n_{i-1}}\left(\mathbb{E}_{t\wedge\nu_k}(f(\Delta S_{\tau^n_i})) - \mathbb{E}_{\tau^n_{i-1}}(f(\Delta S_{\tau^n_i}))\right)\\ &\qquad\qquad \times H_{\tau^n_{j-1}}\mathbb{E}_{\tau^n_{j-1}}\left(\mathbb{E}_{t\wedge\nu_k}(f(\Delta S_{\tau^n_j})) - \mathbb{E}_{\tau^n_{j-1}}(f(\Delta S_{\tau^n_j}))\right)\Big)\\ &\le \mathbb{E}\left(\left(2C_f C^{(3.3.2)}_{t\wedge\nu^k}\sup_{0\le s\le t\wedge\nu^k}|H_s|\right)^2 N^n_{t\wedge\nu^k}\varepsilon_n^{2\alpha}\right) = \varepsilon_n^{2\alpha-4}\mathbb{E}\left(V^n_{t\wedge\nu^k}\right),\end{aligned}$$

where we used (3.A.8) to interchange the sum and the expectation, and (3.A.6)-(3.A.7) to justify that the expectations of the cross-products are well defined and equal 0. In particular, since the process in the right-hand side of the last inequality is non-decreasing, we obtain

$$\varepsilon_n^{4-2\alpha}\sup_{0\le s\le t}\mathbb{E}(|Z^n_{s\wedge\nu_k}|^2) \le \mathbb{E}(V^n_{t\wedge\nu_k}). \tag{3.A.11}$$

Applying Doob's L^2-inequality (see [RY99, Theorem II.1.7]) to the càdlàg martingale $(Z_{t\wedge\nu^k})_{0\le t\le T}$, we obtain

$$\mathbb{E}\left(\sup_{0\le s\le t}|Z^n_{s\wedge\nu_k}|^2\right) \le 4\sup_{0\le s\le t}\mathbb{E}\left(|Z^n_{s\wedge\nu_k}|^2\right).$$

Combining this estimate with (3.A.11) and from the definition of U^n_t we get

$$\mathbb{E}(U^n_{t\wedge\nu_k}) = \varepsilon_n^{4-2\alpha}\mathbb{E}\left(\sup_{0\le s\le t}|Z^n_{s\wedge\nu_k}|^2\right) \le 4\mathbb{E}(V^n_{t\wedge\nu_k}).$$

The convergence $\varepsilon_n^{2-\alpha}Z^n_t \xrightarrow[n\to+\infty]{u.c.a.s.} 0$ now follows from Lemma 3.B.2.

To complete the proof in the general case $f_t = \sum_{\text{finitely many } k} f^k_t P_k$, simply apply the above result to $H_t f^k_t$ and P_k for each k. □

3.B Supplementary material

3.B.1 Decomposition of symmetric matrix into non-negative and non-positive parts

Lemma 3.B.1. *Let $(M_t)_{0\le t\le T}$ be an $\mathcal{S}_d$-valued continuous adapted process on some filtered probability space. Then we can decompose $M_t = M^+_t - M^-_t$ where M^+_t and M^-_t are $\mathcal{S}^+_d$-valued continuous adapted processes; this decomposition, however, is not unique.*

Proof. Let $\lambda_t := \max(\lambda_{\max}(M_t), 0)$. By Hoffman and Wielandt's theorem [HJ90, p. 368], $(\lambda_t)_{0\le t\le T}$ is continuous and we may take $M^+_t := \lambda_t\,\mathrm{Id}, \quad M^-_t := \lambda_t\,\mathrm{Id} - M_t$. □

3.B.2 Fundamental lemma on the a.s. convergence of processes

The following lemma is inspired from [GL14a, Lemma 2.1], but its assumptions better fit our setting.

Lemma 3.B.2. *Let $(U^n)_{n\geq 0}$ and $(V^n)_{n\geq 0}$ be two sequences of non-negative measurable processes. Assume that:*

(i) the series $\sum_{n\geq 0} V^n_t$ converges for all $t \in [0,T]$ a.s.;

(ii) the above limit is upper bounded by a non-decreasing adapted càdlàg process $\bar{V}$;

(iii) there is a constant $c_{(3.B.1)} \geq 0$ such that, for every $n \in \mathbb{N}$, $k \in \mathbb{N}$ and $t \in [0,T]$, we have

$$\mathbb{E}[U^n_{t\wedge\nu^k}] \leq c_{(3.B.1)}\ \mathbb{E}[V^n_{t\wedge\nu^k}] \tag{3.B.1}$$

with the stopping time $\nu^k := \inf\{s \in [0,T] : \bar{V}_s \geq k\}$ (with the usual convention that $\inf\emptyset = +\infty$);

(iv) there is a deterministic function $q : \mathbb{N} \to \mathbb{R}^+$ such that $q(k) \geq k$ and $\bar{V}_{\nu^k} \leq q(k)$ for any k a.s.

Then for any $t \in [0,T]$, the series $\sum_{n\geq 0} U^n_t$ converges almost surely. As a consequence, $U^n_t \overset{\text{a.s.}}{\to} 0$.

Proof. Let $t \in [0,T]$ be fixed. Denote by $\mathcal{N}_V$ the subset of Ω on which the series $(\sum_{n\geq 0} V^n_t)_{0\leq t\leq T}$ do not converge, on which $\bar{V}$ and then $(\nu^k)_{k\geq 0}$ are not defined and on which the inequalities of (iv) are not fulfilled; note that $\mathcal{N}_V$ is built as a countable union of negligible sets, thus it is $\mathbb{P}$-negligible.
For $\omega \notin \mathcal{N}_V$, we have $\bar{V}_{t\wedge\nu^k}(\omega) \leq q(k)$ for any $k \in \mathbb{N}$. Set $\bar{V}^p := \sum_{n=0}^p V^n$: we have $\bar{V}^p \leq \bar{V}$ on $\mathcal{N}_V^c$; thus, the localization of $\bar{V}$ entails that of $\bar{V}^p$ and we have $\bar{V}^p_{t\wedge\nu^k} \leq q(k)$ for any k, p (on $\mathcal{N}_V^c$). Furthermore the relation of domination (iii) writes

$$\mathbb{E}\left[\sum_{n=0}^p U^n_{t\wedge\nu^k}\right] \leq c_{(3.B.1)}\ \mathbb{E}\left[\sum_{n=0}^p V^n_{t\wedge\nu^k}\right] = c_{(3.B.1)}\ \mathbb{E}[\bar{V}^p_{t\wedge\nu^k}] \leq c_{(3.B.1)}\ q(k) \tag{3.B.2}$$

for any k, p (on $\mathcal{N}_V^c$). From Fatou's lemma we get $\mathbb{E}[\sum_{n\geq 0} U^n_{t\wedge\nu^k}] < +\infty$ for any k, therefore the series $\sum_{n\geq 0} U^n_{t\wedge\nu^k}(\omega)$ converges for all ω outside of a $\mathbb{P}$-negligible set $\mathcal{N}_{k,t}$. The set $\mathcal{N}_t := \bigcup_{k\in\mathbb{N}} \mathcal{N}_{k,t} \bigcup \mathcal{N}_V$ is $\mathbb{P}$-negligible, and it follows that for $\omega \notin \mathcal{N}_t$, the series $\sum_{n\geq 0} U^n_{t\wedge\nu^k}(\omega)$ converges for all $k \in \mathbb{N}$. For $\omega \notin \mathcal{N}_t$, we have $\nu^k(\omega) = +\infty$ as soon as $k > \bar{V}_T(\omega)$; thus by taking such k, we complete the convergence of $\sum_{n\geq 0} U^n_t$ on $\mathcal{N}_t^c$. □

3.B.3 Control of martingale processes

Lemma 3.B.3. *Assume that a process S and a sequence of discretization grids $\mathcal{T}$ verify ($\mathbf{H}_S^{gen.}$) and ($\mathbf{H}_R$)-1 with a sequence $(\varepsilon_n : n \geq 0)$ such that $\sum_{n\geq 0} \varepsilon_n^2 < +\infty$. Let $(H_t)_{0\leq t\leq T}$*

be an adapted continuous scalar process and let $f : \mathbb{R}^d \to \mathbb{R}$ *be a* α*-homogeneous function with* $\alpha > 0$*. Then for any* $k = 1, \dots, d$ *we have*

$$\varepsilon_n^{-\alpha+1} \sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} f(\Delta S_s) \mathrm{d}S_s^k \xrightarrow[n\to+\infty]{u.c.a.s.} 0.$$

Proof. Using the decomposition $S = A + M$, we write

$$\sum_{\tau_{i-1}^n < t} H_{\tau_{i-1}^n} \int_{\tau_{i-1}^n}^{\tau_i^n \wedge t} f(\Delta S_s) \mathrm{d}S_s^k = \int_0^t H_{\varphi(s)} f(\Delta S_s) \mathrm{d}A_s^k + \int_0^t H_{\varphi(s)} f(\Delta S_s) \mathrm{d}M_s^k.$$

First, the assumption ($\mathbf{H}_R$)-1 and the inequality $|f(x)| \leq C_f |x|^\alpha$ yield

$$\varepsilon_n^{-\alpha+1} \left| \int_0^t H_{\varphi(s)} f(\Delta S_s) \mathrm{d}A_s^k \right| \leq \varepsilon_n C_f \sup_{0\leq t\leq T} |H_t| \int_0^t (\varepsilon_n^{-1} |\Delta S_s|)^\alpha \mathrm{d}|A^k|_s \xrightarrow[n\to+\infty]{u.c.a.s.} 0.$$

Second, the quadratic variation of the Brownian stochastic integral is

$$\left\langle \varepsilon_n^{-\alpha+1} \int_0^\cdot H_{\varphi(s)} f(\Delta S_s) \mathrm{d}M_s^k \right\rangle_T \leq C_f^2 \sup_{0\leq t\leq T} |H_t|^2 \sup_{0\leq t\leq T} |(\sigma_t \sigma_t^\mathsf{T})_{kk}| \varepsilon_n^{-2\alpha+2} \int_0^T |\Delta S_s|^{2\alpha} \mathrm{d}s \leq C \varepsilon_n^2$$

for some a.s. finite random variable $C > 0$ (using again ($\mathbf{H}_R$)-1). Thus using that $\sum_{n\geq 0} \varepsilon_n^2 < +\infty$ and applying [GL14a, Corollary 2.1] we get

$$\varepsilon_n^{-\alpha+1} \int_0^t H_{\varphi(s)} f(\Delta S_s) \mathrm{d}M_s^k \xrightarrow[n\to+\infty]{u.c.a.s.} 0,$$

which implies the result. □

Chapter 4

Parametric inference for diffusions observed at stopping times

Contents

Statement of the problem. In this work we study the problem of parametric inference for a d-dimensional Brownian semimartingale $(S_t)_{0\le t\le T}$ of the form

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma(s, S_s, \xi)\mathrm{d}B_s, \qquad t \in [0,T], \quad S_0 \in \mathbb{R}^d, \tag{4.0.1}$$

based on a finite random number of observations of S at stopping times. The time horizon $T > 0$ and S_0 are fixed. We assume that the observations are the values of a single trajectory of $(S_t : 0 \le t \le T)$ sampled from the model (4.0.1) with an unknown parameter $\xi = \xi^\star \in \Xi$. Our goal is to estimate $\xi^\star$ using these discrete observations and study the asymptotic properties of the estimator sequence as the number of observations goes to infinity; we work in the high-frequency fixed horizon setting. Handling data at random observation times is important in practice (see the examples in [GW02, Fuk10] for instance) and it has a large impact on inference procedure, as it is argued in [ASM03].

A large number of works (see the references below) are devoted to the inference of diffusion models in the case of deterministic, random independent or strongly predictable observation time grids. In most cases they are based on the approximations of the transition

probability density of the diffusion process, resulting in so called approximate maximum likelihood estimators (AMLEs). However, in practice, the observation times may be random and, moreover, the randomness may be (at least partly) endogenous, i.e. depending on the sampled process itself: see [GW02] for empirical evidence about the connection of volatility and inter-transaction duration in finance, and [Fuk10] for modeling bid or ask quotation data and tick time sampling. In other words, as motivated by those examples, the observation grid may be given by a sequence of general stopping times with respect to a general filtration; see the introduction of Chapter 3 for additional motivation and discussion. To the best of our knowledge this setting has not yet been studied in the literature, except in [LMR+14] where a Central Limit Theorem (CLT) for estimating the integrated volatility in dimension 1 is established assuming the convergence in probability of renormalized quarticity and tricity (however, the authors do not characterize the stopping times for which these convergences hold). One reason for this lack of studies in the literature is essentially that the necessary tools for the analysis of the stopping time discretization grids for multidimensional processes were not available until recently. In particular, the study of the asymptotic normality for a sequence of estimators requires a general central limit theorem for discretization errors based on such grids. Such a result has been very recently obtained in Chapter 3 in a concrete setting (i.e. for explicitly defined class of grids, and not given by abstract assumptions, as a difference with [LMR+14]), in several dimensions (as a difference with above references) and with a tractable limit characterization. Note that in [Fuk11b], the derivation of CLT is achieved in the context of general stopping times, but the limit depends on implicit conditions that are hardly tractable except in certain situations (notably in dimension 1). Another issue is that it is delicate to design an appropriate AMLE method in this stopping times setting: in general, approximation of the increment distribution seems hardly possible in this case, since the expression for the distribution of (S_τ, τ), where τ is a stopping time, is out of reach in multiple dimension even in the simplest cases.

In this work we aim at constructing a consistent sequence of estimators $(\xi^n)_{n\geq 0}$ of the true parameter $\xi^\star$ in the case of random observation grids given by general stopping times. We provide an asymptotic analysis that allows to directly apply the existing results of Chapter 3 on CLTs for discretization errors and show the convergence in distribution of the renormalized error $\sqrt{N_T^n}(\xi^n - \xi^\star)$ (where N_T^n is the number of observation times) to an explicitly defined mixture of normal variables.

Literature background. A number of works study the problem of inference for diffusions. For general references, see the books [Sør04, Fuc13] and the lecture notes [Jac07].

The nonparametric estimation of the diffusion coefficient $\sigma(.)$ is investigated in [FZ93] for equidistant observations times on a fixed time interval. In [GCJ93] the authors consider the problem of the parametric estimation of a multidimensional diffusion under regular deterministic observation grids. They construct consistent sequences of estimators of the unknown parameter based on the minimization of certain contrasts and prove the weak convergence of the error renormalized at the rate $\sqrt{n}$ to a mixed Gaussian variable, where n is the number of observations. The problem of achieving minimal variance estimator is investigated using the local asymptotic mixed normality (LAMN) property, see e.g. [CY90, Chapter 5] for

the definition: this LAMN property is established in [Doh87] for one-dimensional S, and in [Gob01] for higher dimensions using Malliavin calculus techniques, when the n observation times are equidistant on a fixed interval. These latter results show the optimality of Gaussian AMLEs that achieve consistency with minimal variance.

If the time step between the observations is not small, one can use more advanced techniques based on the expansions of transition densities in order to approximate the likelihood of the observations. See, for instance, [AS99, AS02, AS08, CC11]. Note that these works consider only the case of deterministic observation grids.

In [GCJ94] the authors study the case where each new observation time may be chosen by the user depending on the previous observations (so that the times depend on the trajectory of S). The authors exhibit a sequence of sampling schemes with an asymptotic conditional variance achieving the optimal (over all such schemes with random times) bound for LAMN property for all the parameter values simultaneously. We remark that though in [GCJ94] the observation times are random, they are not stopping times, and the perspective is quite different from ours: the authors assume that observations at all times are, in principle, available, and aim at choosing adaptively a finite number of them to optimize the asymptotic variance of the estimator. In our setting observations are stopping times and are not chosen by the user in an anticipative way.

Several works are dedicated to the inference problem with observations at stopping times, but under quite restrictive assumptions on those times as a difference with our general setting. More precisely, in [ASM03, DG04] the authors assume that the time increment $\tau_i^n - \tau_{i-1}^n$ depends only on the information up to τ_{i-1}^n and on extra independent noise. A similar condition is considered in [HJY11], and it can take the form of *strongly predictable* times (τ_i^n is known at time τ_{i-1}^n). In [ASM04], the time increments are simply independent and identically distributed. In [Fuk10, FR12], the authors consider the observation times as exit times of S from an interval in dimension 1: because such one-dimensional exit time can be explicitly approximated, they are able to establish some CLT results for the realized variance. For potentially more general stopping times, but still in dimension 1, [LMR+14] provides CLT results under the extra condition of convergence of the quarticity and tricity. To summarize, all the above results consider stopping times with significant restrictions and, in any case, in one-dimensional setting for S. In the current study, we aim at overcoming these restrictions.

Our contributions.

- To the best of our knowledge, this is the first work that analyzes the problem of parametric inference for multidimensional diffusions based on observations at general stopping times.
- Under mild assumptions we construct a sequence of estimators and prove its consistency for a large class of observations grids, which, following Remark 1.2.2, contains most of the examples, interesting in practice.
- Using our asymptotic analysis and applying the results of Chapter 3 we prove the weak convergence of the renormalized error to a mixture of normal variables, for a

quite general class of random observations, which includes exit times from general random domains, and allows combination of endogenous and independent sources of randomness. In addition, we explicitly compute the limit distribution. The asymptotic limit is, in general, biased, and we characterize both asymptotic bias and variance. Such a bias has not been previously observed in parametric inference problems due to centering property of Gaussian increments for strongly predictable grids.

- We provide a uniform lower bound on the limit variance in the case of a 1-dimensional parameter $\xi \in \Xi$, and for the set of observation grids for which the weak convergence to a mixture of normal variables without bias holds. We also prove that this bound is sharp in this class of grids. To the best our knowledge, this result for parametric inference for diffusions is new, and it allows for discussing optimal sampling procedure for instance.

Last, for other applications and results of stopping times in high-frequency regime, see [Fuk11b, GL14a] and Chapter 1.

Outline of the chapter. In Section 4.1 we present the model for the observed process S, the random observation grids, and construct a sequence of estimators $(\xi^n)_{n\geq 0}$ based on the discretized version of the integrated Kullback-Leibler divergence in the Gaussian case. Section 4.2 is devoted to the statements of the main results of the chapter. We continue with the proofs in Section 4.3. Several technical points are postponed to Section 4.A.

4.1 The model

Let $(B_t)_{0\leq t\leq T}$ be a d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ with $(\mathcal{F}_t)_{0\leq t\leq T}$ verifying the usual conditions of being right-continuous and complete. By $|\cdot|$ we denote the Euclidean norm on matrix and tensor vector spaces. Let $\mathrm{Mat}_{m,n}$ be the space of real $m \times n$ matrices, denote by $\mathcal{S}_m^{++}$ (resp. $\mathcal{S}_m^{+}$) the set of positive (resp. non-negative) definite symmetric real $m \times m$ matrices.

Let $\Xi \subset \mathbb{R}^q, q \geq 1$, be a convex compact set, with non-empty interior to avoid degenerate cases. We fix a parameter $\xi^\star \in \Xi \setminus \partial\Xi$ (where $\partial\Xi$ is the boundary of Ξ). The process serving for the observation is a d-dimensional Brownian semimartingale $(S_t)_{0\leq t\leq T}$ of the form

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma(s, S_s, \xi^\star)\mathrm{d}B_s, \qquad t \in [0,T], \quad S_0 \in \mathbb{R}^d, \tag{4.1.1}$$

verifying the following:

($\mathbf{H}_S$):
1. $\sigma : [0,T] \times \mathbb{R}^d \times \Xi \to \mathrm{Mat}_{d,d}$ is a $\mathcal{C}^{1,2,2}$ function;
2. the matrix $\sigma(t, S_t, \xi)$ is invertible for all $\xi \in \Xi$ and $t \in [0,T]$ a.s.;
3. $(b_t)_{0\leq t\leq T}$ is a continuous adapted $\mathbb{R}^d$-valued process such that for some $\eta_b > 0$, for some a.s. finite C and for any $0 \leq s \leq t \leq T$ we have $|b_t - b_s| \leq C|t-s|^{\eta_b}$.

In what follows we denote for simplicity $\sigma_t(\xi) := \sigma(t, S_t, \xi)$. Let $c_t(\cdot) := \sigma_t(\cdot)\sigma_t(\cdot)^\mathsf{T}$. We suppose, in addition, the following parameter identifiability assumption.

($\mathbf{H}_\xi$): For any $\xi \in \Xi \setminus \{\xi^\star\}$ we have a.s. that the continuous trajectories $t \mapsto c_t(\xi^\star)$ and $t \mapsto c_t(\xi)$ are not almost everywhere (w.r.t. the Lebesgue measure) equal on $[0,T]$.

4.1.1 Random observation grids

We consider a sequence of random observation grids

$$\{(\tau_0^n := 0 < \tau_1^n < \cdots < \tau_i^n < \cdots < \tau_{N_T^n}^n := T) : n \geq 0\}$$

on the interval $[0,T]$ and suppose that for each n, only the values $(\tau_i^n, S_{\tau_i^n})_{0\leq i\leq N_T^n}$ are available for the parameter estimation: these are the observation data. For each n, $(\tau_i^n : 0 \leq i \leq N_T^n)$ is a sequence of $\mathcal{F}$-stopping times and N_T^n is a.s. a finite random variable. Here we do not assume further information on the structure of these stopping times (e.g. they are hitting times for S of such or such boundary and so on): we are aware that having this structural information would presumably be beneficial for the inference problem, by making the estimation more accurate. Proving optimality results (like in [Doh87, Gob01]) given the sequence of observations $\{(\tau_i^n, S_{\tau_i^n})_{0\leq i\leq N_T^n} : n \geq 0\}$ is so far out of reach, and we leave these problems for further investigation. However we establish a partial optimality result in Section 4.2.4.

Our statistics analysis is based on the asymptotic techniques, developed recently in [GL14a] and Chapters 1-2, for admissible random discretization grids in the setting of quadratic variation minimization. In this work we adapt these techniques to the problem of parametric estimation.

We introduce the following assumptions that depend on the choice of a positive sequence $(\varepsilon_n)_{n\geq 0}$ with $\varepsilon_n \to 0$ and a parameter $\rho_N \geq 1$ (compare to the definition in Section 1.2.2):

($\mathbf{A}_S^{\text{osc.}}$): The following non-negative random variable is a.s. finite:

$$\sup_{n\geq 0}\left(\varepsilon_n^{-2}\sup_{1\leq i\leq N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}|S_t - S_{\tau_{i-1}^n}|^2\right) < +\infty. \tag{4.1.2}$$

($\mathbf{A}_N$): For some $\rho_N \in [1, (1+2\eta_b)\wedge 4/3)$ the following non-negative random variable is a.s. finite:

$$\sup_{n\geq 0}(\varepsilon_n^{2\rho_N} N_T^n) < +\infty. \tag{4.1.3}$$

Let us now fix $(\varepsilon_n)_{n\geq 0}$ with $\varepsilon_n \to 0$ and a sequence of discretization grids $\mathcal{T}$. We assume for some $\rho_N \in [1, (1+2\eta_b)\wedge 4/3)$ the following hypothesis:

($\mathbf{H}_\mathcal{T}$): For any subsequence $(\varepsilon_{\iota(n)})_{n\geq 0}$ of $(\varepsilon_n)_{n\geq 0}$ there exists another subsequence $(\varepsilon_{\iota'\circ\iota(n)})_{n\geq 0}$ for which the assumptions ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) (with the given ρ_N) are verified.

Remark that the class of grids verifying ($\mathbf{H}_\mathcal{T}$) is very general and covers most of the settings considered in the previous works on inference for diffusions. At the same time, it allows new types of grids that were not studied before. In particular, it includes:

- The sequences of deterministic or strongly predictable discretization grids for which the time steps are controlled from below and from above and for which the step size

tends to zero. Here $\rho_N > 1$, see Remark 1.2.2.

- The sequences of grids based on exit times from general random domains and, possibly, extra independent noise. Namely let $\{(D_t^n)_{0\le t\le T} : n \ge 0\}$ be a sequence of general random adapted processes with values in the set of domains in $\mathbb{R}^d$, that are continuous and converging (in a suitable sense, see the details in Section 3.2.2) to an adapted continuous domain-valued process $(D_t)_{0\le t\le T}$. Consider also an i.i.d. family of random variables $(U_{i,n})_{n,i\in\mathbb{N}}$ uniform on $[0,1]$ and an arbitrary $\mathcal{P}\otimes\mathcal{B}([0,1])$-measurable ($\mathcal{P}$ is the σ-field of predictable sets of $[0,T]\times\Omega$) mapping $G : (t,\omega,u) \in [0,T]\times\Omega\times[0,1] \mapsto \mathbb{R}^+\cup\{+\infty\}$ (to simplify we write $G_t(u)$). Then the discretization grids of the form $\mathcal{T} := \{\mathcal{T}^n : n \ge 0\}$ with $\mathcal{T}^n = \{\tau_i^n, i = 1,\dots,N_T^n\}$ given by

$$\begin{cases} \tau_0^n := 0, \\ \tau_i^n := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin \varepsilon_n D^n_{\tau_{i-1}^n}\} \wedge (\tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}) \wedge T, \end{cases} \tag{4.1.4}$$

where $(\Delta_{n,i})_{n,i\in\mathbb{N}}$ represents some negligible contribution, verify the assumption $(\mathbf{H}_{\mathcal{T}})$ with $\rho_N = 1$ (see the proof in Section 3.3.3). This class of discretization grids allows a coupling of endogenous noise generated by hitting times and extra independent noise given, for example, by a Poisson process with stochastic intensity (see Section 3.2.2). In addition, we can rely on a CLT for a general discretization error term based on such grids (see Theorem 3.2.7). The optimal observation grid in Section 4.2.4 is of the above form, taking some ellipsoid for D^n and $G(\cdot) = +\infty$, $\Delta_{n,i} = 0$.

The subsequence formulation of the assumption $(\mathbf{H}_{\mathcal{T}})$ is motivated by the subsequence principle in Lemma 2.2.2. It allows to first prove a.s. results for the sequences of observation grids verifying $(\mathbf{A}_S^{\text{osc.}})$-$(\mathbf{A}_N)$ and $\sum_{n\ge0}\varepsilon_n^2 < +\infty$, and then pass to the equivalent results in probability in the general case.

4.1.2 Sequence of estimators

Suppose that $\mathcal{T} := \{\mathcal{T}^n : n \ge 0\}$ is a sequence of random grids verifying $(\mathbf{H}_{\mathcal{T}})$ for some $\varepsilon_n \to 0$, and $\rho_N \in [1, (1+2\eta_b)\wedge 4/3)$. Denote for any process H (where we omit the dependence on n)

$$\varphi(t) := \max\{\tau\in\mathcal{T}^n : \tau \le t\}, \quad \Delta H_t := H_t - H_{\varphi(t)}. \tag{4.1.5}$$

Parametric inference for a discretely observed process typically requires a discrete approximation of some criterion, whose optimization yields the true parameter $\xi^\star$. A standard approach is to approximate the likelihood of $S_{\tau_0^n},\dots,S_{\tau_i^n}$, or equivalently of the distribution of $\Delta S_{\tau_i^n}$ conditionally on $S_{\tau_0^n},\dots,S_{\tau_{i-1}^n}$. Gaussian approximations are often used when the distance between observation times is small, see, for instance [GCJ93]. The optimality of the Gaussian based likelihood approximations in the case of regular observation times has been proved in [Doh87, Gob01]. Although the distribution of S_τ as τ is a stopping time may be quite different from Gaussian, we are inspired by the same approach, because of the flexibility and tractability of the subsequent contrast estimator with respect to the choice

of observation times τ_i^n; however, below we present a slightly different interpretation of the same minimization criteria, since in the stopping time case the distribution of process increments is not necessarily close to Gaussian. We also generalize the criteria to account for non-equidistant distribution of the discretization points over $[0,T]$.

Denote $p_\Sigma(x) := (2\pi)^{-d/2}(\det\Sigma)^{-1/2}\exp\left(-\frac{1}{2}x^\mathsf{T}\Sigma^{-1}x\right)$ the density of a centered d-dimensional Gaussian variable $\mathcal{N}_d(0,\Sigma)$ with the covariance matrix Σ (assumed to be non-degenerate). Denote the Kullback-Leibler (KL) divergence between the variables $\mathcal{N}_d(0,\Sigma_1)$ and $\mathcal{N}_d(0,\Sigma_2)$ by

$$D_{\mathrm{KL}}(\Sigma_1,\Sigma_2) := \int_{\mathbb{R}^d} p_{\Sigma_1}(x)\log\frac{p_{\Sigma_1}(x)}{p_{\Sigma_2}(x)}\mathrm{d}x. \tag{4.1.6}$$

For some continuous weight function $\omega : [0,T]\times\mathbb{R}^d \to]0,+\infty[$ set $\omega_t := \omega(t,S_t)$; the process $(\omega_t)_{0\le t\le T}$ is continuous adapted positive. Recall that $D_{\mathrm{KL}}(\Sigma_1,\Sigma_2)$ is always non-negative and equals 0 if and only if $\Sigma_1=\Sigma_2$. Thus, in view of ($\mathbf{H}_\xi$), the minimization of $\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star),c_t(\xi))\omega_t\mathrm{d}t$ naturally yields the true parameter $\xi^\star$. Our goal is to construct a discretized version of this criterion based on the observations of S. We write

$$D_{\mathrm{KL}}(\Sigma_1,\Sigma_2) = \frac{1}{2}\int_{\mathbb{R}^d}\left(\log(\det\Sigma_2)-\log(\det\Sigma_1)+x^\mathsf{T}\Sigma_2^{-1}x - x^\mathsf{T}\Sigma_1^{-1}x\right)p_{\Sigma_1}(x)\mathrm{d}x,$$

and thus

$$\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star),c_t(\xi))\omega_t\mathrm{d}t = \frac{1}{2}U^\star(\xi)+C_0, \tag{4.1.7}$$

where C_0 is independent of ξ and

$$\begin{aligned} U^\star(\xi) &:= \int_0^T\int_{\mathbb{R}^d}\left(\log(\det c_t(\xi)) + x^\mathsf{T}c_t^{-1}(\xi)x\right)p_{c_t(\xi^\star)}(x)\omega_t\mathrm{d}x\mathrm{d}t \\ &= \int_0^T\left(\log(\det c_t(\xi)) + \mathrm{Tr}(\sigma_t(\xi^\star)^\mathsf{T}c_t^{-1}(\xi)\sigma_t(\xi^\star))\right)\omega_t\mathrm{d}t. \end{aligned} \tag{4.1.8}$$

Remark that $\int_0^T \mathrm{Tr}(\sigma_t(\xi^\star)^\mathsf{T}c_t^{-1}(\xi)\sigma_t(\xi^\star))\omega_t\mathrm{d}t$ represents a quadratic variation. Thus we define the following discretized version of $U^\star(\cdot)$, that uses only $(\tau_i^n, S_{\tau_i^n} : 0\le i\le N_T^n)$,

$$U^n(\xi) := \sum_{\tau_{i-1}^n<T}\omega_{\tau_{i-1}^n}\log\left(\det c_{\tau_{i-1}^n}(\xi)\right)(\tau_i^n-\tau_{i-1}^n) + \sum_{\tau_{i-1}^n<T}\omega_{\tau_{i-1}^n}\Delta S_{\tau_i^n}^\mathsf{T}c_{\tau_{i-1}^n}^{-1}(\xi)\Delta S_{\tau_i^n}. \tag{4.1.9}$$

The random function $U^n(.)$ plays the role of a contrast function: it is asymptotically equal to $U^\star(.)$, which minimum is achieved at $\xi^\star$. In the case of regular grids and $\omega_t=1$ the contrast (4.1.9) coincides with [GCJ93, eq. (3)].

Define the sequence of estimators $(\xi^n)_{n\ge0}$ as follows:

$$\xi^n := \mathrm{Argmin}_{\xi\in\Xi}\, U^n(\xi) \tag{4.1.10}$$

(if the minimizing set of $U^n(\cdot)$ is not a single point we take any of its elements). We expect that the minimizer of $U^n(\cdot)$ will asymptotically attain the minimizer of $\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star),c_t(\xi))\omega_t\mathrm{d}t$,

i.e. $\xi^\star$.

Note that the user is free to choose the form of the process ω_t. While the rigorous optimization of the choice of ω_t given only the observations $(\tau_i^n, S_{\tau_i^n}, 0 \leq i \leq N_T^n)$ is complicated, it seems reasonable to increase ω_t on the time intervals where the observation frequency is higher. We have not investigated furthermore in this direction.

4.2 Main results

For the subsequent convergences, we adopt the following natural notations. By $O_n^{\text{a.s.}}(1)$ (resp. $o_n^{\text{a.s.}}(1)$) we denote any a.s. bounded (resp. a.s. converging to 0) sequence of random variables; in addition, denote $O_n^{\text{a.s.}}(x) = xO_n^{\text{a.s.}}(1), o_n^{\text{a.s.}}(x) = xo_n^{\text{a.s.}}(1)$. Similarly we write $o_n^{\mathbb{P}}(1)$ for sequences converging to 0 in probability.

Besides, we introduce a convenient and short notation for denoting random vectors written as a mixture of Gaussian random variables. Given a (possibly stochastic) matrix $V \in \mathcal{S}_m^+$, we denote by $\mathcal{N}(0, V)$ a random variable which is equal in distribution to $V^{1/2}G$ where G is a centered Gaussian m-dimensional vector with covariance matrix Id_m, where $V^{1/2}$ is the principal square root of V, and where G is independent from everything else.

4.2.1 Consistency

The following result states the convergence of the estimators $(\xi^n)_{n\geq 0}$ in probability to $\xi^\star$ for any sequence of random observation grids verifying $(\mathbf{H}_\mathcal{T})$. Its proof is postponed to Section 4.3.1.

Theorem 4.2.1. *Assume $(\mathbf{H}_S)$, $(\mathbf{H}_\xi)$ and $(\mathbf{H}_\mathcal{T})$. Then for the sequence estimators $(\xi^n)_{n\geq 0}$ given by (4.1.10) we have the following convergence in probability*

$$\xi^n \xrightarrow[n\to+\infty]{\mathbb{P}} \xi^\star.$$

4.2.2 Asymptotic error analysis

We now proceed with the asymptotic analysis of the error sequence $(\xi^n - \xi^\star)_{n\geq 0}$. Recall that $D_{\mathrm{KL}}(\Sigma_1, \Sigma_2)$ given in (4.1.6) is always non-negative and equals to 0 if and only if $\Sigma_1 = \Sigma_2$. Thus for any $t \in [0, T]$ the point $\xi^\star \in \Xi \setminus \partial\Xi$ is a minimum of $D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\cdot))$ which implies that $\nabla_\xi^2 D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi)) \mid_{\xi=\xi^\star}$ is positive semidefinite a.s. for all $t \in [0, T]$. We introduce the following assumption:

$(\mathbf{H}_\mathcal{H})$: There exists a subset $\mathcal{I} \subset [0, T]$ of positive Lebesgue measure such that

$$\nabla_\xi^2 D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi)) \mid_{\xi=\xi^\star} \text{ is positive definite for all } t \in \mathcal{I}.$$

Note that in practice, since $\xi^\star$ is not known, the verification of $(\mathbf{H}_\mathcal{H})$ is typically required for all possible values of $\xi^\star \in \Xi \setminus \partial\Xi$. Assumption $(\mathbf{H}_\mathcal{H})$ in particular implies that

$$\mathcal{H}_T := 2\int_0^T \left(\nabla_\xi^2 D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi)) \mid_{\xi=\xi^\star}\right) \omega_t \mathrm{d}t = \nabla_\xi^2 U^\star(\xi^\star) \tag{4.2.1}$$

is positive definite, and where the second equality follows from (4.1.7) (note that we can interchange differentiation and integration via the dominated convergence theorem).

In what follows we assume the following conventions. The gradient of an $\mathbb{R}$-valued function is assumed to be a column vector. For a $\mathrm{Mat}_{d,d}$-valued function $c = c(x)$, $x \in \mathbb{R}^m$, the gradient $\nabla_x c(\cdot)$ is a element of $\mathbb{R}^m \otimes \mathrm{Mat}_{d,d}$. For an element $x \otimes y \in \mathbb{R}^m \otimes \mathrm{Mat}_{d,d}$ we denote $\mathrm{Tr}(x \otimes y) := x\,\mathrm{Tr}(y)$, which extends linearly on the entire space $\mathbb{R}^m \otimes \mathrm{Mat}_{d,d}$. For $\mathcal{A} \in \mathbb{R}^m \otimes \mathrm{Mat}_{d,d}$ so that $\mathcal{A} = [\mathcal{A}^1, \cdots, \mathcal{A}^m]^\mathsf{T}$ and $x, y \in \mathbb{R}^d$ we denote $x^\mathsf{T}\mathcal{A}y := [x^\mathsf{T}\mathcal{A}^1 y, \cdots, x^\mathsf{T}\mathcal{A}^m y]^\mathsf{T} \in \mathbb{R}^m$. By $x^\mathsf{T}\mathcal{A}$ we denote the linear operator in $\mathrm{Mat}_{m,d}$ corresponding to $y \mapsto x^\mathsf{T}\mathcal{A}y$ (similarly for $x \mapsto x^\mathsf{T}\mathcal{M}y$). Finally, partial derivatives of a $\mathrm{Mat}_{d,d}$-valued function are obtained by differentiating each matrix component and take values in $\mathrm{Mat}_{d,d}$.

For $i = 1, \ldots, d$ we denote $\nabla_{x_i}\sigma(\xi) := \nabla_{x_i}\sigma(t, S_t, \xi)$, where $\sigma = \sigma(t, x, \xi)$ is given by ($\mathbf{H}_S$)-1. Define the processes $(\mathcal{M}_t)_{0\le t\le T}$ and $(\mathcal{A}_t)_{0\le t\le T}$ with values in $\mathrm{Mat}_{m,d}$ and $\mathbb{R}^m \otimes \mathrm{Mat}_{d,d}$ respectively as follows:

$$\mathcal{M}_t := 2\omega_t b_t^\mathsf{T} \nabla_\xi c_t^{-1}(\xi^\star) + \bar{\mathcal{M}}_t, \qquad \mathcal{A}_t := 2\omega_t \nabla_\xi c_t^{-1}(\xi^\star)\sigma_t(\xi^\star), \qquad t \in [0, T] \tag{4.2.2}$$

where for $1 \le i \le m, 1 \le j \le d$ we define

$$\bar{\mathcal{M}}_t^{ij} := 2\omega_t\,\mathrm{Tr}(\sigma_t(\xi^\star)^\mathsf{T}\nabla_{\xi_i} c_t^{-1}(\xi^\star)\nabla_{x_j}\sigma_t(\xi^\star)). \tag{4.2.3}$$

Here comes the main result of this section. This is a universal decomposition of the estimation error, available for any stopping time grids, as in ($\mathbf{H}_\mathcal{T}$), which will be the starting point for showing a CLT later.

Theorem 4.2.2. *Assume* ($\mathbf{H}_S$), ($\mathbf{H}_\xi$), ($\mathbf{H}_\mathcal{T}$) *and* ($\mathbf{H}_\mathcal{H}$). *Then, for* ρ_N *as in* ($\mathbf{A}_N$), *we have*

$$\varepsilon_n^{-\rho_N}(\xi^n - \xi^\star) = (\mathcal{H}_T^{-1} + o_n^\mathbb{P}(1))\varepsilon_n^{-\rho_N} Z_T^n + o_n^\mathbb{P}(1), \tag{4.2.4}$$

where

$$Z_s^n := \int_0^s \Delta S_t^\mathsf{T}\mathcal{A}_{\varphi(t)}\mathrm{d}B_t + \int_0^s \mathcal{M}_{\varphi(t)}\Delta S_t \mathrm{d}t := M_s^n + A_s^n \tag{4.2.5}$$

for $\mathcal{M}_t$ *and* $\mathcal{A}_t$ *defined in* (4.2.2).

The proof is done in Section 4.3.2.

4.2.3 CLT in the case of ellipoid exit times

We start with the following lemma, that plays an important role in the sequel:

Lemma 4.2.3 ([GL14a, Lemma 3.1])**.** *Let* y *be a* $d \times d$*-matrix symmetric non-negative real matrix. Then the equation*

$$2\,\mathrm{Tr}(x)x + 4x^2 = y^2 \tag{4.2.6}$$

admits exactly one solution $x(y) \in \mathcal{S}_d^+$.

Theorem 4.2.2 shows that it is enough to study the convergence in distribution of $\sqrt{N_T^n}Z_T^n$ to obtain such a convergence for $\sqrt{N_T^n}(\xi^n - \xi^\star)$. Indeed, from (4.2.4) we get

$$\sqrt{N_T^n}(\xi^n - \xi^\star) = (\mathcal{H}_T^{-1} + o_n^\mathbb{P}(1))\sqrt{N_T^n}Z_T^n + o_n^\mathbb{P}(\sqrt{N_T^n}\varepsilon_n^{\rho_N})$$

where $o_n^{\mathbb{P}}(\sqrt{N_T^n}\varepsilon_n^{\rho_N}) = o_n^{\mathbb{P}}(1)$ from ($\mathbf{H}_{\mathcal{T}}$) and the subsequence principle (Lemma 2.2.2). This makes possible the direct application of general results on CLT for discretization errors of the form (4.2.5); we refer to Chapter 3 for discussion and references on the subject.

Since we are particularly interesting in the case of stopping time discretization grids in the multidimensional case, we use Theorem 3.2.7 where the CLT for discretization errors of the form (4.2.5) with general $\mathcal{M}_t$ and $\mathcal{A}_t$ has been proved in a quite general setting. We state a particular case of this setting, namely the exit times from random ellipsoids (as defined in (4.2.7)). This example is, in particular, used in Section 4.2.4.

Let $(\Sigma_t)_{0\le t\le T}$ and $(\Sigma_t^n)_{0\le t\le T}, n \ge 0$, be adapted continuous $\mathcal{S}_d^{++}$-valued processes, characterizing the ellipsoids. Assume the following:

($\mathbf{H}_\Sigma$): 1. For some $\eta > 0$ and a.s. finite C we have that $\sup_{0\le t\le T} |\Sigma_t - \Sigma_t^n| \le C\varepsilon_n^\eta$ a.s.;

2. There exist positive continuous $\mathcal{F}$-adapted processes $(v_t)_{0\le t\le T}$ and $(\delta_t)_{0\le t\le T}$, such that we have a.s. for all $t \in [0,T]$ that $\sup_{t\le s\le \psi(t)} |b_s| \le v_t$, where $\psi(t) := \inf\{s \ge t : |S_s - S_t| \ge \delta_t\} \wedge T, \quad t \in [0,T]$ (this condition is quite mild, see Example 3.2.1).

3. The random variables b_0 and Σ_0 are bounded.

4. For some $\eta_\sigma > 0$ we have $|\sigma_t - \sigma_s| \le C_\sigma |t-s|^{\eta_\sigma/2}$ for all $0 \le s < t \le T$ and the variable C_σ verifies $\mathbb{E}(C_\sigma^4) < +\infty$ (this condition, in particular, holds for a diffusion process with bounded coefficients b and σ such that their derivatives are also bounded).

Define the sequence of discretization grids $\mathcal{T} = \{\mathcal{T}^n : n \ge 0\}$ by

$$\tau_0^n = 0, \quad \tau_i^n = \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Sigma_{\tau_{i-1}^n}^n (S_t - S_{\tau_{i-1}^n}) \ge \varepsilon_n^2\} \wedge T. \tag{4.2.7}$$

Such a sequence verifies ($\mathbf{H}_{\mathcal{T}}$) with $\rho_N = 1$ (which follows from Theorem 3.2.7, see the proof of Theorem 4.2.4).

To simplify we note $\sigma_t := \sigma_t(\xi^\star)$ till the end of this section. We consider the setting of Section 3.2.2 with $D_t = \{x :\in \mathbb{R}^d : \ x^\mathsf{T}\Sigma_t x = 1\}$ and $D_t^n = \{x :\in \mathbb{R}^d : \ x^\mathsf{T}\Sigma_t^n x = 1\}$. Define the process $m_t := \left[\mathrm{Tr}(\sigma_t^\mathsf{T}\Sigma_t\sigma_t)\right]^{-1}$. Following Chapter 3, define, for any $t \in [0,T]$ and any measurable function $f : \mathbb{R}^d \mapsto \mathbb{R}$,

$$\tau(t) := \inf\{s \ge 0 : \sigma_t W_s \notin D_t\}, \qquad \mathcal{B}_t[f(\cdot)] := \mathbb{E}_t\left(f(\sigma_t W_{\tau(t)})\right), \tag{4.2.8}$$

where W is an extra d-dimensional Brownian motion, independent from everything else. Denote $\mathcal{A}_t^\mathsf{T} := [\mathcal{A}_{1,t}^\mathsf{T}, \dots, \mathcal{A}_{m,t}^\mathsf{T}]^\mathsf{T}$ and $\mathcal{A}_t^{ij} := \frac{1}{2}(\mathcal{A}_{i,t}\mathcal{A}_{j,t}^\mathsf{T} + \mathcal{A}_{i,t}^\mathsf{T}\mathcal{A}_{j,t})$. Since $\mathcal{A}_t^{ij}$ is symmetric, by Lemma 3.B.1 we may write $\mathcal{A}_t^{ij} = \mathcal{A}_t^{ij+} - \mathcal{A}_t^{ij-}$, where $\mathcal{A}_t^{ij+}$ and $\mathcal{A}_t^{ij-}$ are continuous symmetric non-negative definite matrices. Define a $\mathrm{Mat}_{m,m}$-valued process $(\mathcal{K}_t)_{0\le t\le T}$ by

$$\mathcal{K}_t^{ij} := m_t^{-1}\mathcal{B}_t\left[f(x) := ((\sigma_t^{-1}x)^\mathsf{T} X_t^{ij+}(\sigma_t^{-1}x))^2 - ((\sigma_t^{-1}x)^\mathsf{T} X_t^{ij-}(\sigma_t^{-1}x))^2\right], \tag{4.2.9}$$

for all $1 \le i,j \le m$, where X_t^{ij+} (resp. X_t^{ij-}) is the solution of the matrix equation (4.2.6) for $c = \sigma_t^\mathsf{T}\mathcal{A}_t^{ij+}\sigma_t$ (resp. $\sigma_t^\mathsf{T}\mathcal{A}_t^{ij-}\sigma_t$). Remark that the process $(Q_t)_{0\le t\le T}$ defined in (3.2.18) is equal to 0 in our case since the domains D_t and D_t^n are symmetric, see Section 3.2.4.

Also note that the matrix equation (4.2.6) may be easily solved numerically, see the details in [GL14a, Section A.4]. However, analytic solution is only available in dimension 1. In general (especially in multi-dimensional case), the computation of $\mathcal{K}$ is hardly explicit, and requires some numerical methods, like Monte-Carlo schemes suitable for statistics of stopped processes, see e.g. [GM10]. The following result is an application of Theorem 3.2.7 and its proof.

Theorem 4.2.4. *The process $(\mathcal{K}_t)_{0\le t\le T}$ is continuous and $\mathcal{K}_t \in \mathcal{S}_m^+$ a.s. for all $t \in [0,T]$. Denote $\mathcal{K}_t^{1/2}$ the matrix principal square root of $\mathcal{K}_t$. Then there exists an m-dimensional Brownian motion $\widetilde{W}$ defined on an extended probability space $(\tilde\Omega, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and independent of B such that for the sequence estimators $(\xi^n)_{n\ge0}$ given by (4.1.10) we have*

$$\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{d} \mathcal{H}_T^{-1}\sqrt{\int_0^T m_t^{-1}\mathrm{d}t}\int_0^T \mathcal{K}_t^{1/2}\mathrm{d}\widetilde{W}_t, \tag{4.2.10}$$

where $\mathcal{H}_T$ is defined in (4.2.1). More specifically, for Z^n, M^n, A^n defined in (4.2.5), we have the convergences

$$\begin{aligned}
&\varepsilon_n^{-2}\langle Z^n\rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^s \mathcal{K}_t\mathrm{d}t, \text{ for all } s\in[0,T],\\
&\varepsilon_n^{-1}\langle Z^n, B\rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \text{ for all } s\in[0,T],\\
&\varepsilon_n^{-1}\sup_{s\in[0,T]}|A_s^n| \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \qquad \varepsilon_n^2 N_T^n \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^T m_t^{-1}\mathrm{d}t.
\end{aligned} \tag{4.2.11}$$

Proof. Our goal is to check the assumptions of Theorem 3.2.7. First note that all random variables σ_0, σ_0^{-1}, $\mathcal{M}_0$ and $\mathcal{A}_0$ are bounded under our setting. Condition ($\mathbf{H}_S$) follows from ($\mathbf{H}_S$) and ($\mathbf{H}_\Sigma$)-4. Further ($\mathbf{H}_\Delta$) follows from ($\mathbf{H}_\Sigma$)-2.

Conditions ($\mathbf{H}_D^1$)-($\mathbf{H}_D^2$) are straightforward from the definition of D_t and D_t^n, and ($\mathbf{H}_\Sigma$)-1. Namely, for $B_d(0,1)$ the unit ball in $\mathbb{R}^d$ centered at 0, we write

$$D_t = \{\Sigma_t^{-1/2}x : x\in B_d(0,1)\} \quad \text{and} \quad D_t^n = \{(\Sigma_t^n)^{-1/2}x : x\in B_d(0,1)\}$$

from which one may easily get (for the distance $\mu(\cdot,\cdot)$ for domains, as defined in Section 3.2.2) that $\mu(D_t, D_t^n) \le 2|\Sigma_t^{-1/2} - (\Sigma_t^n)^{-1/2}|$. The latter bound can be controlled uniformly in t and n in view of the continuity and the non-degeneracy of Σ_t, Σ_t^n and the condition ($\mathbf{H}_\Sigma$)-1.

Finally ($\mathbf{H}_G$) is trivial in this case since the function $G(\cdot)$ equals $+\infty$ and $\Delta_{n,i} = 0$ (in the notation of Chapter 3). Other assumptions of Theorem 3.2.7 follow from ($\mathbf{H}_\Sigma$)-3. □

Note that the drift b does not enter in the parameters of the CLT, this is due to the symmetry of the domain defining the observation times.

Because $\widetilde{W}$ is independent of everything else, we have the identity

$$\mathcal{H}_T^{-1}\sqrt{\int_0^T m_t^{-1}\mathrm{d}t}\int_0^T \mathcal{K}_t^{1/2}\mathrm{d}\widetilde{W}_t \overset{d}{=} \mathcal{H}_T^{-1}\sqrt{\int_0^T m_t^{-1}\mathrm{d}t}\left(\int_0^T \mathcal{K}_t\mathrm{d}t\right)^{1/2}\mathcal{N}(0,\mathrm{Id}_m)$$

with an extra independent m-dimensional Gaussian random variable $\mathcal{N}(0, \mathrm{Id}_m)$. In other words, the (random) covariance limit of $\sqrt{N_T^n}(\xi^n - \xi^\star)$ is

$$V_T := \left(\int_0^T m_t^{-1} \mathrm{d}t\right) \mathcal{H}_T^{-1} \left(\int_0^T \mathcal{K}_t \mathrm{d}t\right) \mathcal{H}_T^{-1}.$$

4.2.4 Optimal uniform lower bound on the limit variance

In this section we assume $q = 1$, so that $\Xi \subset \mathbb{R}$. Our aim is to seek the optimal observation times (among ellipsoid based stopping times) achieving the lowest possible limit variance.

Let $X_t(\xi)$ be the solution of the matrix equation (4.2.6) with

$$y^2 = \sigma_t(\xi)^\mathsf{T} \nabla_\xi c_t^{-1}(\xi) \sigma_t(\xi) \sigma_t(\xi)^\mathsf{T} \nabla_\xi c_t^{-1}(\xi) \sigma_t(\xi)$$

(note that it is an element of $\mathrm{Mat}_{d,d}(\mathbb{R})$ for a scalar ξ). For $\mathcal{H}_T$ given in (4.2.1) define

$$V_T^{opt.} := \mathcal{H}_T^{-2} \left(\int_0^T 2\omega_t \,\mathrm{Tr}(X_t(\xi^\star)) \mathrm{d}t\right)^2, \tag{4.2.12}$$

which is fixed from now on. In the case where the weak convergence of the renormalized error to a mixture of normal variables holds without bias (e.g. the case of deterministic grids, see [GCJ93]; or the hitting times of symmetric boundaries, see Section 3.2.4 and Theorem 4.2.4) we prove that $V_T^{opt.}$ is a uniform lower bound on the asymptotic variance of the sequence of estimators (4.1.10). In addition, this lower bound is tight in the sense that one can find a sequence of observation times achieving as close as possible this lower bound. This is formalized in the following definition.

Definition 4.2.5. *Let $\kappa_0 > 0$. A parametric family of discretization grid sequences $\{\mathcal{T}_\kappa : \kappa \in (0, \kappa_0]\}$ is κ-optimal if there exists an a.s. finite random variable C_0 independent of κ such that $\sqrt{N_T^n}(\xi^n - \xi^\star)$ converges in distribution to a mixture of centered normal variables for all $\mathcal{T}_\kappa$,*

$$\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{d} \mathcal{N}(0, V_T^\kappa),$$

and the limit variance V_T^κ associated with $\mathcal{T}_\kappa$ verifies the condition

$$0 \le V_T^\kappa - V_T^{opt.} \le C_0 \kappa, \quad \forall \kappa \in (0, \kappa_0].$$

The subsequent κ-optimal observation times are related to some random ellipsoid hitting times, which are built as follows. Let $\chi(.)$ be a smooth function such that $\mathbb{1}_{(-\infty,1/2]} \le \chi(.) \le \mathbb{1}_{(-\infty,1]}$, let $\chi_\kappa(x) := \chi(x/\kappa)$. Let $\Lambda_t(\xi) := 2\omega_t \sigma_t^{-1}(\xi)^\mathsf{T} X_t(\xi) \sigma_t^{-1}(\xi)$, define

$$\Lambda_t^\kappa(\xi) := \Lambda_t(\xi) + \kappa \chi_\kappa(\lambda_{\min}(\Lambda_t(\xi)))\, \mathrm{Id}_d,$$

where $\lambda_{\min}(M)$ stands for the smallest eigenvalue of $M \in \mathcal{S}_d^+$. Hence, $\Lambda_t^\kappa(\xi) \in \mathcal{S}_d^{++}$ as soon as $\kappa > 0$. Recall that under the general assumptions of Theorem 4.2.2 we have the decomposition (4.2.4), with Z^n given by (4.2.5). In view of (4.2.4), to study the weak

convergence of $\sqrt{N_T^n}(\xi^n - \xi^\star)$ we essentially need to consider $\sqrt{N_T^n} Z_T^n$. The result below states that under standard conditions implying the CLT for $\sqrt{N_T^n} Z_T^n$ (and hence for $\sqrt{N_T^n}(\xi^n - \xi^\star)$) there exists a uniform lower bound on the limit variance. We also show the tightness of this bound in the sense of Definition 4.2.5.

Theorem 4.2.6. *Assume* $(\mathbf{H}_S)$, $(\mathbf{H}_\xi)$, $(\mathbf{H}_\mathcal{T})$ *and* $(\mathbf{H}_\mathcal{H})$. *Let* $(\xi^n)_{n\geq 0}$ *be defined by* (4.1.10). *For some* $\rho \in [1, \rho_N]$ *suppose that the semimartingale decomposition* $Z_t^n := M_t^n + A_t^n$ *in* (4.2.5) *verifies*

$$\begin{aligned} &\varepsilon_n^{-2\rho} \langle M^n \rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^s \mathcal{K}_t \mathrm{d}t, \ \textit{for all } s \in [0,T], \\ &\varepsilon_n^{-\rho} \langle M^n, B \rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \ \textit{for all } s \in [0,T], \\ &\varepsilon_n^{-\rho} \sup_{0\leq t\leq T} |A_t^n| \xrightarrow[n\to+\infty]{\mathbb{P}} 0 \end{aligned} \tag{4.2.13}$$

for some adapted non-negative continuous process $(\mathcal{K}_t)_{0\leq t\leq T}$. *Assume also that* $N_T^n \langle Z^n \rangle_T$ *converges in probability to an a.s. finite random variable. Then, the following holds:*

(i) $\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{d} \mathcal{N}(0, V_T)$ *for some non-negative random variable* V_T *(asymptotic variance).*

(ii) The asymptotic variance V_T *satisfies the following uniform lower bound:* $V_T \geq V_T^{opt.}$ a.s. *for* $V_T^{opt.}$ *defined in* (4.2.12).

(iii) Assuming, in addition, $(\mathbf{H}_\Sigma)$*-2,3,4, the lower bound* $V_T^{opt.}$ *is tight in the following sense: the parametric family of discretization grid sequences* $\{\mathcal{T}_\kappa : \kappa \in (0,1]\}$ *given for any* $\varepsilon_n \to 0$ *by* $\mathcal{T}_\kappa = \{\mathcal{T}_\kappa^n : n \geq 0\}$ *with* $\mathcal{T}_\kappa^n = (\tau_i^n)_{0\leq i\leq N_T^n}$ *written as*

$$\tau_0^n := 0, \quad \tau_i^n = \inf\left\{ t \geq \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Lambda_{\tau_{i-1}^n}^\kappa (\xi^\star)(S_t - S_{\tau_{i-1}^n}) > \varepsilon_n^2 \right\} \wedge T \tag{4.2.14}$$

is κ*-optimal for* $\kappa_0 = 1$ *in the sense of Definition 4.2.5.*

We remark that the class of discretization grids over which the universal variance lower bound is obtained in Theorem 4.2.6 includes most of the examples for which a CLT has been established, since the conditions of the type (4.2.13) are quite commonly required (see [JS02, Chapter IX, Theorem 7.3] for a classical result). Typically for deterministic or strongly predictable grids the conditions will hold with $\rho = \rho_N > 1$, while in the setting of Section 3.2.2 we have $\rho = \rho_N = 1$. See also the discussion in Section 4.1.1 and Remark 1.2.2.

As we may notice the κ-optimal sequence of discretization grids in (4.2.14) depends on the unknown parameter $\xi^\star$. Besides, concerning the optimal variance $V_T^{opt.}$ in (4.2.12), it also involves $\xi^\star$, as well as ω_t: we argue in Section 4.1.2 that the rigorous optimization of ω_t (to minimize $V_T^{opt.}$) is out of reach because $\xi^\star$ is unknown. However, for all these extra optimization steps, a heuristic approach might be used. Namely in practice, one may pre-estimate $\xi^\star$ on some initial interval $[0, T_1]$ using any reasonable consistent estimator and then proceed with the estimation that achieves the limit variance close to the optimum on $[T_1, T]$ using this pre-estimator instead of $\xi^\star$. A thorough analysis of the limit variance in our case

would be possible, although quite technical; we naturally expect that such a method would constitute a κ-optimal family of strategies for $T_1 = \kappa^2 T$ in view of the robustness results for the optimal sequence of discretization grids produced in Section 2.3.1.

4.3 Proofs of the main results

The next lemma provides some important properties of the process $\sigma_t(\cdot)$.

Lemma 4.3.1. *Assume* ($\mathbf{H}_S$)-*1. Let $\mathcal{T}$ be any sequence of observation grids verifying* ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) *with* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Then the following holds:*

(i) For any $\eta_\sigma \in (0,1)$ we have that for some a.s. finite random variable C_0

$$|\sigma_t(\xi^\star) - \sigma_s(\xi^\star)| \leq C_0|t-s|^{\eta_\sigma/2} \quad \forall s,t\in[0,T] \quad \text{a.s.}$$

(ii) For $(\nabla_x\sigma_t(\xi^\star))_{0\leq t\leq T}$ defined in Section 4.2.2 and any $\rho > 0$ we have

$$\varepsilon_n^{-(2-\rho)} \sup_{0\leq t\leq T}\left|\sigma_t(\xi^\star) - \sigma_{\varphi(t)}(\xi^\star) - \sum_{i=1}^d \nabla_{x_i}\sigma_{\varphi(t)}(\xi^\star)\Delta S_t^i\right| \xrightarrow[n\to+\infty]{\text{a.s.}} 0.$$

Proof. To prove (i) remark that $(S_t)_{0\leq t\leq T}$ is Hölder continuous with any exponent smaller than $1/2$ by [BY82b, Theorem 5.1]. We conclude by using that $\sigma = \sigma(t,x,\xi^\star)$ is locally Lipschitz in t and x due to the continuous differentiability, and that $(S_t)_{0\leq t\leq T}$ is a.s. bounded on $[0,T]$.

To prove (ii) we use the differentiability of $\sigma(t,x,\xi^\star)$ in t and x by ($\mathbf{H}_S$)-1. We write

$$\begin{aligned}
\sigma_t(\xi^\star) - \sigma_{\varphi(t)}(\xi^\star) &= \sigma(t,S_t,\xi^\star) - \sigma(\varphi(t),S_{\varphi(t)},\xi^\star)\\
&= \sigma(\varphi(t),S_t,\xi^\star) - \sigma(\varphi(t),S_{\varphi(t)},\xi^\star) + O_n^{\text{a.s.}}(|\Delta t|)\\
&= \sum_{i=1}^d \nabla_{x_i}\sigma(\varphi(t),S_{\varphi(t)},\xi^\star)\Delta S_t^i + O_n^{\text{a.s.}}(|\Delta t| + |\Delta S_t|^2).
\end{aligned}$$

From ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) and Lemma 1.3.2 we get $\sup_{t\in[0,T]}|O_n^{\text{a.s.}}(|\Delta t|+|\Delta S_t|^2)| \leq C_\rho\varepsilon_n^{2-\rho}$ for any $\rho > 0$ and some a.s. finite C_ρ, which finishes the proof. □

The next lemma states the a.s. convergence of $U^n(\cdot)$ to $U^\star(\cdot)$, as well as the corresponding results for the derivatives $\nabla_\xi U^n(\cdot)$ and $\nabla_\xi^2 U^n(\cdot)$.

Lemma 4.3.2. *Assume* ($\mathbf{H}_S$)-*1,2. Let $\mathcal{T}$ be any sequence of observation grids verifying* ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) *with* $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$. *Then the following convergences hold*

$$\sup_{\xi\in\Xi}|U^n(\xi) - U^\star(\xi)| \xrightarrow[n\to+\infty]{\text{a.s.}} 0, \tag{4.3.1}$$

$$\sup_{\xi\in\Xi}|\nabla_\xi U^n(\xi) - \nabla_\xi U^\star(\xi)| \xrightarrow[n\to+\infty]{\text{a.s.}} 0, \qquad |\nabla_\xi^2 U^n(\xi) - \nabla_\xi^2 U^\star(\xi)| \xrightarrow[n\to+\infty]{\text{a.s.}} 0, \quad \forall\xi\in\Xi. \tag{4.3.2}$$

Proof. Using (4.1.8) and Lemma 4.A.1 we deduce the following expressions for $\nabla_{\xi_k} U^\star(\xi)$ and $\nabla^2_{\xi_k\xi_l} U^\star(\xi)$ $(1 \le k, l \le m)$:

$$\nabla_{\xi_k} U^\star(\xi) = \int_0^T \mathrm{Tr}\left(\nabla_{\xi_k} c_t(\xi) c_t(\xi)^{-1} + \sigma_t(\xi^\star)^\mathsf{T} \nabla_{\xi_k} c_t^{-1}(\xi) \sigma_t(\xi^\star)\right) \omega_t \mathrm{d}t, \tag{4.3.3}$$

$$\nabla^2_{\xi_k\xi_l} U^\star(\xi) = \int_0^T \mathrm{Tr}\left(\nabla^2_{\xi_k\xi_l} c_t(\xi) c_t^{-1}(\xi) + \nabla_{\xi_k} c_t(\xi) \nabla_{\xi_l} c_t^{-1}(\xi) + \sigma_t(\xi^\star)^\mathsf{T} \nabla^2_{\xi_k\xi_l} c_t^{-1}(\xi) \sigma_t(\xi^\star)\right) \omega_t \mathrm{d}t. \tag{4.3.4}$$

Recall that

$$U^n(\xi) = \sum_{\tau_{i-1}<T} \omega_{\tau^n_{i-1}} \log(\det c_{\tau^n_{i-1}}(\xi))(\tau^n_i - \tau^n_{i-1}) + \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^\mathsf{T}_{\tau^n_i} c^{-1}_{\tau^n_{i-1}}(\xi) \Delta S_{\tau^n_i}. \tag{4.3.5}$$

Let us first prove that for any $\xi \in \Xi$

$$U^n(\xi) \xrightarrow[n\to+\infty]{\text{a.s.}} U^\star(\xi). \tag{4.3.6}$$

The convergence of the first term in the right-hand side of (4.3.5) follows from the standard Riemann integral approximation, using that $\sup \Delta\tau^n_i \xrightarrow[n\to+\infty]{\text{a.s.}} 0$ by Lemma 1.3.2, so we get

$$\sum_{\tau_{i-1}<T} \omega_{\tau^n_{i-1}} \log(\det c_{\tau^n_{i-1}}(\xi))(\tau^n_i - \tau^n_{i-1}) \xrightarrow[n\to+\infty]{\text{a.s.}} \int_0^T \log(\det c_t(\xi)) \omega_t \mathrm{d}t. \tag{4.3.7}$$

For the second term we have by Proposition 1.3.9

$$\sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^\mathsf{T}_{\tau^n_i} c^{-1}_{\tau^n_{i-1}}(\xi) \Delta S_{\tau^n_i} \xrightarrow[n\to+\infty]{\text{a.s.}} \int_0^T \mathrm{Tr}\left(\sigma_t(\xi^\star)^\mathsf{T} c_t^{-1}(\xi) \sigma_t(\xi^\star)\right) \omega_t \mathrm{d}t. \tag{4.3.8}$$

Hence the convergence (4.3.6) follows now from taking the sum of (4.3.7) and (4.3.8). Further using Lemma 4.A.1 we obtain

$$\begin{aligned}\nabla_{\xi_k} U^n(\xi) = &\sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \mathrm{Tr}\left(\nabla_{\xi_k} c_{\tau^n_{i-1}}(\xi) c^{-1}_{\tau^n_{i-1}}(\xi)\right)(\tau^n_i - \tau^n_{i-1}) \\ &+ \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^\mathsf{T}_{\tau^n_i} (\nabla_{\xi_k} c^{-1}_{\tau^n_{i-1}}(\xi)) \Delta S_{\tau^n_i},\end{aligned} \tag{4.3.9}$$

$$\begin{aligned}\nabla^2_{\xi_k\xi_l} U^n(\xi) = &\sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \mathrm{Tr}\left(\nabla^2_{\xi_k\xi_l} c_{\tau^n_{i-1}}(\xi) c^{-1}_{\tau^n_{i-1}}(\xi) + \nabla_{\xi_k} c_{\tau^n_{i-1}}(\xi) \nabla_{\xi_l} c^{-1}_{\tau^n_{i-1}}(\xi)\right)(\tau^n_i - \tau^n_{i-1}) \\ &+ \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^\mathsf{T}_{\tau^n_i} \nabla^2_{\xi_k\xi_l} c^{-1}_{\tau^n_{i-1}}(\xi) \Delta S_{\tau^n_i}.\end{aligned} \tag{4.3.10}$$

Using (4.3.3), (4.3.4) and applying the same reasoning as for the proof of (4.3.6) we also

show the following convergences for any $\xi \in \Xi$

$$\nabla_\xi U^n(\xi) \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla_\xi U^\star(\xi), \qquad \nabla^2_\xi U^n(\xi) \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla^2_\xi U^\star(\xi). \tag{4.3.11}$$

Further from (4.3.9) and (4.3.10), using ($\mathbf{H}_S$)-1,2, the compactness of Ξ, the continuity of ω_t and the convergence $\sum_{\tau_i^n \le T} |\Delta S_{\tau_i^n}|^2 \xrightarrow[n\to+\infty]{\text{a.s.}} \mathrm{Tr}(\langle S\rangle_T)$ by Proposition 1.3.9, we have a.s.

$$\begin{aligned}
\sup_{n\ge 0}\left(\sup_{\xi\in\Xi}|\nabla_\xi U^n(\xi)|\right) &\le C \sup_{0\le t\le T}\left[\omega_t\left(\sup_{\xi\in\Xi}|\nabla_\xi c_t(\xi)c_t^{-1}(\xi)| + \sup_{\xi\in\Xi}|\nabla_\xi c_t^{-1}(\xi)|\right)\right] < +\infty,\\
\sup_{n\ge 0}\left(\sup_{\xi\in\Xi}|\nabla^2_\xi U^n(\xi)|\right) &\le C \sup_{0\le t\le T}\Big[\omega_t\Big(\sup_{\xi\in\Xi}(|\nabla^2_\xi c_t(\xi)||c_t^{-1}(\xi)|) + \sup_{\xi\in\Xi}|\nabla_\xi c_t(\xi)|^2\\
&\qquad + \sup_{\xi\in\Xi}|\nabla^2_\xi c_t^{-1}(\xi)|\Big)\Big] < +\infty,
\end{aligned}$$

for some a.s. finite $C > 0$. This implies that the sequences $(U^n(\cdot))_{n\ge0}, (\nabla_\xi U^n(\cdot))_{n\ge0}$ are equicontinuous and hence the convergences in (4.3.6) and (4.3.11) are uniform in $\xi \in \Xi$. We are done. □

4.3.1 Proof of Theorem 4.2.1

First suppose that $\sum_{n\ge0}\varepsilon_n^2 < +\infty$ and that the grid sequence $\mathcal{T}$ verifies ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$).

Recall that $D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi)) \ge 0$ and the equality holds if and only if $c_t(\xi^\star) = c_t(\xi)$. From ($\mathbf{H}_\xi$) we have that for any $\xi \ne \xi^\star$ the processes $c_t(\xi^\star)$ and $c_t(\xi)$ are not almost everywhere equal on $[0,T]$. Hence $\xi^\star$ is the unique minimum of $\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi))\omega_t \mathrm{d}t$, and in view of (4.1.7) we have that a.s.

$$\xi^\star = \mathrm{Argmin}_{\xi\in\Xi} U^\star(\xi).$$

Further, Lemma 4.3.2 implies that $U^n(\xi) \xrightarrow[n\to+\infty]{\text{a.s.}} U^\star(\xi)$ uniformly in $\xi \in \Xi$, from which we deduce that $\xi^n \xrightarrow[n\to+\infty]{\text{a.s.}} \xi^\star$ since $\xi^n = \mathrm{Argmin}_{\xi\in\Xi} U^n(\xi)$.

Finally the convergence $\xi^n \xrightarrow[n\to+\infty]{\mathbb{P}} \xi^\star$ for $\mathcal{T}$ verifying ($\mathbf{H}_\mathcal{T}$) with general $\varepsilon_n \to 0$ follows from the subsequence principle in Lemma 2.2.2. □

4.3.2 Proof of Theorem 4.2.2

First suppose that $\sum_{n\ge0}\varepsilon_n^2 < +\infty$ and the grid sequence $\mathcal{T}$ verifies ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$).

Step 1. We start by showing the convergence

$$\int_0^1 \nabla^2_\xi U^n(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla^2_\xi U^\star(\xi^\star) =: \mathcal{H}_T. \tag{4.3.12}$$

Let $1 \leq k, l \leq m$. In view of the convergence $\nabla^2_\xi U^n(\xi^\star) \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla^2_\xi U^\star(\xi^\star)$ from Lemma 4.3.2 it is enough verify that

$$\int_0^1 \nabla^2_{\xi_k\xi_l} U^n(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u - \nabla^2_{\xi_k\xi_l} U^n(\xi^\star) \xrightarrow[n\to+\infty]{\text{a.s.}} 0. \tag{4.3.13}$$

Denote $\gamma_t(\xi) := \mathrm{Tr}(\nabla^2_{\xi_k\xi_l} c_t(\xi) c_t^{-1}(\xi) + \nabla_{\xi_k} c_t(\xi) \nabla_{\xi_l} c_t^{-1}(\xi))$. Using the representation (4.3.10) for $\nabla^2_{\xi_k\xi_l} U^n(\cdot)$, we get that the left-hand side in (4.3.13) is equal to

$$\begin{aligned}
&\sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \left(\int_0^1 \gamma_{\tau^n_{i-1}}(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u - \gamma_{\tau^n_{i-1}}(\xi^\star) \right) (\tau^n_i - \tau^n_{i-1}) \\
&\quad + \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^{\mathsf{T}}_{\tau^n_i} \left(\int_0^1 \nabla^2_{\xi_k\xi_l} c^{-1}_{\tau^n_{i-1}}(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u - \nabla^2_{\xi_k\xi_l} c^{-1}_{\tau^n_{i-1}}(\xi^\star) \right) \Delta S_{\tau^n_i}.
\end{aligned}$$

Now (4.3.13) follows from the convergence $\xi^n \xrightarrow[n\to+\infty]{\text{a.s.}} \xi^\star$ for $\mathcal{T}$ verifying $(\mathbf{A}^{\text{osc.}}_S)$-$(\mathbf{A}_N)$ (see the proof of Theorem 4.2.1) and the dominated convergence theorem (in view of the differentiability and invertibility properties of σ from $(\mathbf{H}_S)$-1,2 and the compactness of Ξ).

Step 2: linearization. Our strategy is to analyse $\xi^n - \xi^\star$ using the second order Taylor decomposition of $U^n_T(\cdot)$ near $\xi^\star$ and invoking Theorem 4.2.1. From $(\mathbf{H}_\mathcal{H})$ the matrix $\mathcal{H}_T = \nabla^2_\xi U^\star(\xi^\star)$ is positive definite. Define the following sequence of events

$$\Omega^n := \{\xi^n \in \Xi \setminus \partial\Xi\} \cap \left\{ \int_0^1 \nabla^2_\xi U^n(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u \in \mathcal{S}^{++}_q \right\}.$$

From the convergences (4.3.12) and $\xi^n \xrightarrow[n\to+\infty]{\text{a.s.}} \xi^\star$, and since $\xi^\star \notin \partial\Xi$ we obtain $\mathbb{1}_{\Omega^n} \xrightarrow[n\to+\infty]{\text{a.s.}} 1$. On Ω^n we have $\nabla_\xi U^n(\xi^n) = 0$, which implies

$$\mathbb{1}_{\Omega^n}(\xi^n - \xi^\star) = \mathbb{1}_{\Omega^n} \left(\int_0^1 \nabla^2_\xi U^n(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u \right)^{-1} \nabla_\xi U^n(\xi^\star)$$

by the Taylor formula. This implies, in view of (4.3.12) and since $\mathbb{1}_{\Omega\setminus\Omega^n} = 0$ for n large enough, that

$$\varepsilon_n^{-\rho_N}(\xi^n - \xi^\star) = \left(\mathcal{H}_T^{-1} + o^{\text{a.s.}}_n(1) \right) \varepsilon_n^{-\rho_N} \nabla_\xi U^n(\xi^\star) + o^{\text{a.s.}}_n(1). \tag{4.3.14}$$

Step 3: expansion of $\nabla_\xi U^n(\xi^\star)$. Now let us analyze the term $\nabla_\xi U^n(\xi^\star)$. Using the expression (4.3.9) of $\nabla_\xi U^n(\cdot)$ and applying the Itô formula, we obtain

$$\begin{aligned}
\nabla_\xi U^n(\xi^\star) &= \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \mathrm{Tr}\left(\nabla_\xi c_{\tau^n_{i-1}}(\xi^\star) c^{-1}_{\tau^n_{i-1}}(\xi^\star) \right) (\tau^n_i - \tau^n_{i-1}) + \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^{\mathsf{T}}_{\tau^n_i} \nabla_\xi c^{-1}_{\tau^n_{i-1}}(\xi^\star) \Delta S_{\tau^n_i} \\
&= \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \mathrm{Tr}\left(\nabla_\xi c_{\tau^n_{i-1}}(\xi^\star) c^{-1}_{\tau^n_{i-1}}(\xi^\star) + \sigma_{\tau^n_{i-1}}(\xi^\star)^{\mathsf{T}} \nabla_\xi c^{-1}_{\tau^n_{i-1}}(\xi^\star) \sigma_{\tau^n_{i-1}}(\xi^\star) \right) (\tau^n_i - \tau^n_{i-1})
\end{aligned}$$

$$+\int_0^T \omega_{\varphi(t)} \operatorname{Tr}\left((\sigma_t(\xi^\star)+\sigma_{\varphi(t)}(\xi^\star))^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)(\sigma_t(\xi^\star)-\sigma_{\varphi(t)}(\xi^\star))\right)\mathrm{d}t$$
$$+2\int_0^T \omega_{\varphi(t)}\Delta S_t^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)b_t\mathrm{d}t+2\int_0^T \omega_{\varphi(t)}\Delta S_t^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)\sigma_t(\xi^\star)\mathrm{d}B_t. \quad (4.3.15)$$

Consider the four terms on the right-hand side of (4.3.15). The first term is equal to 0 since, using that $\nabla_\xi c^{-1}_{\tau^n_{i-1}}(\xi^\star) = -c^{-1}_{\tau^n_{i-1}}(\xi^\star)\nabla_\xi c_{\tau^n_{i-1}}(\xi^\star)c^{-1}_{\tau^n_{i-1}}(\xi^\star)$, we have

$$\operatorname{Tr}\left(\sigma_{\tau^n_{i-1}}(\xi^\star)^\mathsf{T}\nabla_\xi c^{-1}_{\tau^n_{i-1}}(\xi^\star)\sigma_{\tau^n_{i-1}}(\xi^\star)\right) = -\operatorname{Tr}\left(\nabla_\xi c_{\tau^n_{i-1}}(\xi^\star)c^{-1}_{\tau^n_{i-1}}(\xi^\star)\right).$$

For the second term, using Lemma 4.3.1 and the properties $(\mathbf{A}_S^{\text{osc.}})$-$(\mathbf{A}_N)$ we deduce that

$$\begin{aligned}\int_0^T \omega_{\varphi(t)} \operatorname{Tr}\left((\sigma_t(\xi^\star)+\sigma_{\varphi(t)}(\xi^\star))^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)(\sigma_t(\xi^\star)-\sigma_{\varphi(t)}(\xi^\star))\right)\mathrm{d}t&\\ = 2\int_0^T \operatorname{Tr}\left(\sigma_{\varphi(t)}(\xi^\star)^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)\sum_{i=1}^d \nabla_{x_i}\sigma_{\varphi(t)}(\xi^\star)\Delta S_t^i\right)\omega_{\varphi(t)}\mathrm{d}t \;+\; e_T^n&\\ = \int_0^T \bar{\mathcal{M}}_{\varphi(t)}\Delta S_t\mathrm{d}t \;+\; e^n_{T,2},&\end{aligned}$$

where for any $\rho>0$ and any $\eta_\sigma\in(0,1)$, using Lemma 1.3.2 and Lemma 4.3.1-(i) we have

$$|e^n_{T,2}| \le C_0(\varepsilon_n^{2-\rho}+\varepsilon_n \sup_t|t-\varphi(t)|^{\eta_\sigma/2}) \le C_0\varepsilon_n^{1+(2-\rho)\eta_\sigma/2}.$$

Here, C_0 is a notation standing for any a.s. finite random variable (independent on n), which values may change throughout the computations. Note that $\varepsilon_n^{-\rho_N}|e^n_{T,2}| \le C_0\varepsilon_n^{-\rho_N+1+(2-\rho)\eta_\sigma/2} \xrightarrow[n\to+\infty]{\text{a.s.}} 0$ for ρ small enough, since $\rho_N<4/3$ by $(\mathbf{A}_N)$. Also remark that the process $(\bar{\mathcal{M}}_t)_{0\le t\le T}$ is the same as defined in (4.2.3), Section 4.2.2.

The third term of (4.3.15) may be written as

$$2\int_0^T \omega_{\varphi(t)}\Delta S_t^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)b_{\varphi(t)}\mathrm{d}t + e^n_{T,3},$$

where, in view Lemma 1.3.2 and Lemma 4.3.1-(i), we have

$$|e^n_{T,3}| \le C_0\varepsilon_n \sup_t|t-\varphi(t)|^{\eta_b} \le C_0\varepsilon_n^{1+(2-\rho)\eta_b}.$$

Again $(\mathbf{A}_N)$ implies that $\varepsilon_n^{-\rho_N}|e^n_{T,3}| \xrightarrow[n\to+\infty]{\text{a.s.}} 0$ for ρ small enough.

Finally, the last term of (4.3.15) equals

$$2\int_0^T \omega_{\varphi(t)}\Delta S_t^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)\sigma_{\varphi(t)}(\xi^\star)\mathrm{d}B_t + e^n_{T,4},$$

where, $(e^n_{t,4})_{0\le t\le T} : n\ge 0)$ is a sequence of continuous local martingales verifying for some

a.s. finite C_0, C_1

$$\langle e^n_{\cdot,4}\rangle_T \le C_0 \sup_{0\le t\le T}\Big(|\Delta S_t|^2|\sigma_t(\xi^\star)-\sigma_{\varphi(t)}(\xi^\star)|^2\Big) \le C_1\varepsilon_n^{2+2\eta_\sigma}$$

for any $\eta_\sigma \in (0,1)$ using $(\mathbf{A}^{\text{osc.}}_S)$, Lemma 4.3.1-(i) and Lemma 1.3.2. This implies $\varepsilon_n^{-\rho_N}|e^n_{T,4}| \xrightarrow[n\to+\infty]{\text{a.s.}} 0$ via an application of [GL14a, Corollary 2.1, p large enough] to the sequence $\varepsilon_n^{-\rho_N} e^n_{\cdot,4}$.

Hence, we deduce that $\nabla_\xi U^n(\xi^\star)$ is equal, up to some negligible contribution, to Z^n_T given in (4.2.5). So finally this implies

$$\varepsilon_n^{-\rho_N}(\xi^n-\xi^\star) = \Big(\mathcal{H}_T^{-1}+o_n^{\text{a.s.}}(1)\Big)\varepsilon_n^{-\rho_N} Z^n_T + o_n^{\text{a.s.}}(1).$$

Step 4: convergence in probability. For a general $\mathcal{T}$ satisfying $(\mathbf{H}_\mathcal{T})$ with $\varepsilon_n \to 0$ the result is obtained via the subsequence principle (Lemma 2.2.2). □

4.3.3 Proof of Theorem 4.2.6

Recall that $\Lambda_t(\xi) = 2\omega_t\sigma_t^{-1}(\xi)^\mathsf{T} X_t(\xi)\sigma_t^{-1}(\xi)$, where $X_t(\xi)$ is the solution of the matrix equation (4.2.6) with $y^2 = \sigma_t(\xi)^\mathsf{T}\nabla_\xi c_t^{-1}(\xi)\sigma_t(\xi)\sigma_t(\xi)^\mathsf{T}\nabla_\xi c_t^{-1}(\xi)\sigma_t(\xi)$.

Central Limit Theorem. All the conditions for applying the CLT of Theorem 3.2.7 are fulfilled, and we get

$$\varepsilon_n^{-\rho} Z^n_T \xrightarrow{d} \int_0^T \mathcal{K}_t^{1/2}\mathrm{d}\widetilde{W}_t,$$

with an independent Brownian motion $\widetilde{W}$. Moreover, the above convergence is $\mathcal{F}$-stable (see [JP12, Section 2.2.1] for related definition and properties). Therefore, together with the convergence of $\varepsilon_n^{2\rho}N^n_T$, we deduce the announced result in (i).

Lower bound. We have

$$N^n_T\langle Z^n\rangle_T = N^n_T\int_0^T 4\omega_t^2\Delta S_t^\mathsf{T}\nabla_\xi c_t^{-1}(\xi^\star)\sigma_t(\xi^\star)\sigma_t(\xi^\star)^\mathsf{T}\nabla_\xi c_t^{-1}(\xi^\star)\Delta S_t\mathrm{d}t.$$

Take some subsequence $\iota(n)$ such that $\sum_{n\ge0}\varepsilon^2_{\iota(n)} < +\infty$ and such that the convergence of $N^{\iota(n)}_T\langle Z^{\iota(n)}\rangle_T$ holds a.s.. Then $\mathcal{H}_T^{-2}N^{\iota(n)}_T\langle Z^{\iota(n)}\rangle_T \xrightarrow[n\to+\infty]{\text{a.s.}} V_T$ where V_T is the limit variance of $\sqrt{N^n_T}(\xi^n-\xi^\star)$, in view of the above arguments for proving (i). From the proof of Theorem 1.4.2 we obtain that

$$V_T = \mathcal{H}_T^{-2}\lim_n N^{\iota(n)}_T\langle Z^{\iota(n)}\rangle_T \ge \mathcal{H}_T^{-2}\left(\int_0^T 2\omega_t\,\mathrm{Tr}(X_t(\xi^\star))\mathrm{d}t\right)^2 =: V_T^{opt.} \quad \text{a.s..}$$

This finishes the proof of (ii).

κ-optimal sequence. We now prove (iii). Let Z^n be defined in Theorem 4.2.2 based on $\mathcal{T}^n_\kappa$, and $(\xi^n)_{n\ge0}$ be the corresponding estimator sequence. By Theorem 4.2.4 we get

the convergence $\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{d} \mathcal{N}(0, V_T^\kappa)$. In addition, by Proposition 4.B.1, since $N_T^n \langle Z^n \rangle_T \xrightarrow[n\to+\infty]{\mathbb{P}} \mathcal{H}_T^2 V_T^\kappa$, we obtain $0 \leq V_T^\kappa - V_T^{opt.} \leq C_0 \kappa$ for some a.s. finite C_0 independent of κ and ε_n. □

Remark that taking $\kappa = 0$ in the definition of $\mathcal{T}^\kappa$ would lead to a grid verifying ($\mathbf{H}_\mathcal{T}$) with $\rho_N > 1$ which is not covered by Theorem 4.2.4.

4.A Technical results

Let $G \in \mathcal{C}^2(\Xi, \mathcal{S}_d^{++})$. Define $f : \mathrm{Mat}_{d,d}(\mathbb{R}) \times \mathbb{R}^d \to \mathbb{R}$ by

$$f(G, x) := \log(\det G) + x^\mathsf{T} G^{-1} x.$$

The following lemma provides the expressions for $\nabla_\xi f(G(\xi), x)$ and $\nabla_\xi^2 f(G(\xi), x)$.

Lemma 4.A.1. *We have (for all $1 \leq k, l \leq m$)*

$$\nabla_{\xi_k} \log(\det G(\xi)) = \mathrm{Tr}(\nabla_{\xi_k} G(\xi) G^{-1}(\xi)), \tag{4.A.1}$$

$$\nabla^2_{\xi_k \xi_l} \log(\det G(\xi)) = \mathrm{Tr}\left(\nabla^2_{\xi_k \xi_l} G(\xi) G^{-1}(\xi) + \nabla_{\xi_k} G(\xi) \nabla_{\xi_l} G^{-1}(\xi)\right), \tag{4.A.2}$$

and, as a consequence,

$$\nabla_{\xi_k} f(G(\xi), x) = \mathrm{Tr}\left(\nabla_{\xi_k} G(\xi) G^{-1}(\xi)\right) + x^\mathsf{T} \nabla_{\xi_k} G^{-1}(\xi) x, \tag{4.A.3}$$

$$\nabla^2_{\xi_k \xi_l} f(G(\xi), x) = \mathrm{Tr}\left(\nabla^2_{\xi_k \xi_l} G(\xi) G^{-1}(\xi) + \nabla_{\xi_k} G(\xi) \nabla_{\xi_l} G^{-1}(\xi)\right) + x^\mathsf{T} \nabla^2_{\xi_k \xi_l} G^{-1}(\xi) x. \tag{4.A.4}$$

Proof. Using the Jacobi formula we get

$$\nabla_{\xi_k} \log(\det G(\xi)) = \frac{\nabla_{\xi_k} \det G(\xi)}{\det G(\xi)} = \mathrm{Tr}\left(\nabla_{\xi_k} G(\xi) G^{-1}(\xi)\right),$$

which gives (4.A.1), a second derivation now implies (4.A.2). The expressions (4.A.3) and (4.A.4) now follow from the definition of $f(G, x)$ and (4.A.1)-(4.A.2). □

4.B κ-optimal discretization strategies

Let $(S_t)_{0 \leq t \leq T}$ verify ($\mathbf{H}_S$). Let $(\mathcal{A}_t)_{0 \leq t \leq T}$ be given by (4.2.2). Fix $i \in \{1, \ldots, m\}$ and let $2\omega_t H_t = \mathcal{A}_t^i$ with $H_t = \nabla_\xi c_t^{-1}(\xi^\star) \sigma_t(\xi^\star)$. Consider the discretization error process of the form

$$Z_s^n := \int_0^s 2\omega_{\varphi(t)} \Delta S_t^\top H_{\varphi(t)} \mathrm{d}B_t.$$

In this section to simplify we write $\sigma_t := \sigma_t(\xi^\star)$. Let X_t be the solution of the matrix equation (4.2.6) with $y^2 = \sigma_t^\mathsf{T} H_t H_t^\mathsf{T} \sigma_t = \sigma_t^\mathsf{T} \nabla_\xi c_t^{-1} \sigma_t \sigma_t^\mathsf{T} \nabla_\xi c_t^{-1} \sigma_t$. The next result essentially follows from [GL14a, Theorem 3.2].

Proposition 4.B.1. *Assume* ($\mathbf{H}_S$), ($\mathbf{H}_\xi$) *and* ($\mathbf{H}_\mathcal{H}$). *Let* $\kappa \in (0,1]$, *for* $t \in [0,T]$ *set* $\Lambda_t := 2\omega_t(\sigma_t^{-1})^\mathsf{T} X_t \sigma_t^{-1}$ *and* $\Lambda_t^\kappa := \Lambda_t + \kappa\chi_\kappa(\lambda_{\min}(\Lambda_t))\,\mathrm{Id}_d$ *(recall the definition of* $\chi_\kappa(\cdot)$ *from Section 4.2.4). For a given* $n \in \mathbb{N}$, *define the discretization grid* $\mathcal{T}_\kappa^n$ *by*

$$\tau_0^n := 0, \quad \tau_i^n = \inf\left\{t \geq \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Lambda^\kappa_{\tau_{i-1}^n} (S_t - S_{\tau_{i-1}^n}) > \varepsilon_n^2\right\} \wedge T. \tag{4.B.1}$$

Then, the sequence of strategies $\mathcal{T}_\kappa = \{\mathcal{T}_\kappa^n : n \geq 0\}$ *verifies* ($\mathbf{H}_\mathcal{T}$), *and it is asymptotically* κ*-optimal in the following sense: we have* $N_T^n\langle Z^n\rangle_T \xrightarrow[n\to+\infty]{\mathbb{P}} V_T^\kappa$ *with* V_T^κ *verifying*

$$0 \leq V_T^\kappa - \left(\int_0^T 2\omega_t \,\mathrm{Tr}(X_t)\mathrm{d}t\right)^2 \leq C_0\kappa \tag{4.B.2}$$

for some a.s. finite random variable C_0 *independent of* $\kappa \in (0,1]$.

Proof. First note that from Theorem 4.2.4 (note that Λ_0^κ is obviously bounded, as needed in ($\mathbf{H}_\Sigma$)-(3)) we get the convergence $N_T^n\langle Z^n\rangle_T \xrightarrow[n\to+\infty]{\mathbb{P}} V_T^\kappa$. Take a subsequence of ε_n for which $\sum_{n\geq 0}\varepsilon^2_{\iota(n)} < +\infty$ and the grid sequence $\mathcal{T}$ verifies ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$). Without loss of generality we assume that for this subsequence the convergence to V_T^κ holds a.s. Let $A_t := \int_0^t b_s \mathrm{d}s$ be the finite variation part and M_t be the martingale part of S_t. Then, using Lemma 1.3.2, we get for any $\rho > 0$ and for some a.s. finite $C > 0$ that $\sup_{t\in[0,T]} |\Delta A_t| \leq |b|_\infty \sup_{t\in[0,T]} |\Delta t| \leq C\varepsilon_n^{2-\rho}$. Hence one may easily check that for $\bar{Z}_s^n := \int_0^s 2\omega_{\varphi(t)} \Delta M_t^\top H_{\varphi(t)} \mathrm{d}B_t$ we have

$$N_T^n\langle Z^n\rangle_T - N_T^n\langle \bar{Z}^n\rangle_T \xrightarrow[n\to+\infty]{\text{a.s.}} 0. \tag{4.B.3}$$

By ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) and Theorem 1.3.4, the sequence of grids $\mathcal{T}_\kappa$ is admissible for the process M_t in the sense of [GL14a]. Thus, for the subsequence $(\varepsilon_{\iota(n)})_{n\geq 0}$, the statement follows from (4.B.3) and [GL14a, Theorem 3.2] applied to $N_T^n\langle \bar{Z}^n\rangle_T$, with

$$C_0 := \left(\sup_{\kappa\in(0,1]} C_\kappa\right)\left(\int_0^T \chi_\kappa(\lambda_{\min}(\Lambda_t))\,\mathrm{Tr}\left(\sigma_t\sigma_t^\mathsf{T}\right)\mathrm{d}t\right)$$

where $C_\kappa := \int_0^T \Big(8\omega_t \,\mathrm{Tr}(X_t) + 3\kappa\chi_\kappa(\lambda_{\min}(\Lambda_t))\,\mathrm{Tr}(c_t)\Big)\mathrm{d}t$. For general case it is enough to note that the limit V_T^κ is the same for any subsequence due to the convergence in probability for the entire sequence $(\varepsilon_n)_{n\geq 0}$. □

Part II

Uncertainty quantification for stochastic approximation limits

Chapter 5

Uncertainty quantification for stochastic approximation limits using chaos expansion

Contents

5.1 Introduction

Since the seminal work of Robbins and Monro [RM51], the method of stochastic approximation (SA for short) has become mainstream for various applications, such as optimization, parameter estimation, signal processing, adaptive control, Monte Carlo optimization of stochastic systems (see [KY97a, BMP90]), stochastic gradient descent methods in machine

learning (see e.g. [BC05, SSS08, BCN17]), adaptive Monte Carlo sampler (see e.g. [HST01, AT08, FMP11, FJLS16, FS00, DVA98]), and efficient tail computations [BFP09], among others.

SA is used to find zeros of an intractable function $h : \mathbb{R}^q \to \mathbb{R}^q$ given in the form of an expectation $h(z) := \mathbb{E}[H(z, V)]$, where V is some random variable and H can be computed explicitly (as opposed to its expectation h). A common application of SA is where h is the gradient of a convex function c given by an expectation, i.e.

$$h(z) = \nabla_z c(z) = \nabla_z \mathbb{E}[C(z, V)].$$

In this case SA corresponds to the minimization of c and is called Stochastic Gradient Descent. Remark that in order to apply SA we need to have $\nabla_z \mathbb{E}[C(z, V)] = \mathbb{E}[\nabla_z C(z, V)]$ and $H := \nabla_z C$ to be known. If only the function C is known one may apply a slightly different Kiefer-Wolfowitz procedure ([KW52]) using finite differences.

In applications, the choice of the model for V is of great importance. Quite often it is chosen from a parametric family of distributions $\{\mu(\theta, \mathrm{d}v) : \theta \in \Theta \subset \mathbb{R}^d\}$, so that the parameter θ must be pre-estimated or set by an expect opinion. Obviously, a perfect specification of θ is rarely possible. In some cases, where we lack information about θ, it is reasonable to assume the model for V to be uncertain. This may be expressed via additional randomness of the parameter θ. Below we present several problems that are solved by SA, and for which the problem of model uncertainty is important:

- Minimization of expected cost (or risk; or utility maximization) under model uncertainty. In this case V models a stochastic system and z corresponds to the parameter determining the strategy of interaction with this system. Further we have $H(z, v) := \nabla_z C(z, v)$ where C is some cost function. The goal is to find a strategy z which minimizes the expected cost $\mathbb{E}[C(z, V)]$. Under suitable assumptions this writes as $\mathbb{E}[\nabla_z C(z, V)] = 0$ and thus may be solved by SA. In this case the model uncertainty problem for V is highly relevant.

- SA may be used to calculate quantiles of a distribution, also known as Value-at-Risk (VaR) in finance (and more generally to calculate a pair of risk measures VaR and CVaR which are widely used, see [BFP09] for details). In financial applications V represents a future random value of some portfolio for which the choice of the distribution is not easy. Often we lack information about it and we need an efficient way to compute the risk measures for a family of models to analyze the model risk. In particular, such analysis is required by financial regulators.

- In some applications the Bayesian approach is used to specify the model for V. Here one considers a parametric family of distributions $\{\mu(\theta, \mathrm{d}v), \theta \in \Theta\}$ with some prior law of θ. After the observation of the data the law of θ is updated to some posterior distribution π. In this case the randomness of θ naturally yields model uncertainty for V.

Motivated by the examples above, we consider the following mathematical framework. We study the uncertainty of the SA procedure in the following forms: *(i)* the distribution

of V depends on an unknown parameter θ in Θ, i.e. $V \sim \mu(\theta, \mathrm{d}v)$, for which only some probability distribution π is available; *(ii)* the function H is modeled through a dependency in the parameter θ. For each θ the solution $z^\star$ of the equation

$$h(z, \theta) := \int_{\mathcal{V}} H(z, V, \theta)\mu(\theta, \mathrm{d}v) = 0 \tag{5.1.1}$$

depends on θ, so that $z^\star = \phi^\star(\theta)$ for some function $\phi^\star(\cdot)$. Our goal is to compute $\phi^\star(\cdot)$ so that we can efficiently quantify the probability distribution of the SA limit $\phi^\star(\theta)$ given the probability distribution π for θ.

Uncertainty quantification. Chaos expansion approach. In the last two decades, UQ has become a huge concern regarding both research and industrial applications. In this work we study the UQ problem for the SA limits, which, to the best of our knowledge, has not been investigated so far.

In UQ applications (see [LK10, Smi14]), the goal is often to quantify the dependence of a solution to an auxiliary problem on some uncertain parameter θ. We denote this dependence by $\theta \mapsto \phi^\star(\theta)$. Quite often $\phi^\star$ solves a Partial Differential Equation (PDE) (see e.g. [LK10]); in our setting, the function $\phi^\star$ is defined for (π-almost) all θ, as the limit of an SA algorithm parameterized by θ.

For the UQ analysis, a first possible approach is based on crude Monte Carlo (MC) methods: they consist of sampling M values θ_m under π, and then compute, for each sample θ_m, an approximation $\widehat{\phi^\star(\theta_m)}$ of $\phi^\star(\theta_m)$. The distribution of the random variable $\{\phi^\star(\theta), \theta \sim \pi\}$ is then approximated by the empirical distribution of $\{\widehat{\phi^\star(\theta_m)} : 1 \leq m \leq M\}$. When $\phi^\star$ solves a PDE, a global error analysis is performed in [BTZ04], accounting for both the sampling error and the PDE discretization error, which are decoupled in some way. In our SA setting, a naive approach would be to compute $\widehat{\phi^\star(\theta_m)}$ as the output of a standard SA algorithm for fixed θ_m (see the discussion in Section 5.2.1).

A second method, developed in [LBM86, KH92], is a perturbative approach taking advantage of a stochastic expansion of $\{\phi^\star(\theta), \theta \sim \pi\}$ that is available when θ has small variations (a restriction that we do not need or want to impose in our case).

A third strategy, which dates back to Wiener [Wie38] and has been developed in the fields of engineering and UQ in the 2000s (see [GS03, LK10] and references therein), is based on chaos expansions. This technique, also known as the spectral method, consists of projecting the unknown function $\phi^\star : \Theta \mapsto \mathbb{R}^q$ on an orthonormal basis $\{\theta \mapsto B_i(\theta), i \geq 0\}$ of the real-valued and squared-integrable (w.r.t. π) functions and computing the $\mathbb{R}^q$-valued coefficients $\{u_i^\star, i \geq 0\}$ of $\phi^\star$ in its decomposition

$$\phi^\star = \sum_{i \geq 0} u_i^\star B_i. \tag{5.1.2}$$

Availability of $u_i^\star$'s (or their approximations) allows to efficient quantify the distribution $\{\phi^\star(\theta), \theta \sim \pi\}$. In the most common case where $B_0 \equiv 1$, the expectation and the variance-

covariance matrix of $\{\phi^\star(\theta), \theta \sim \pi\}$ are related to the coefficients $\{u_i^\star, i \geq 0\}$ through

$$\mathbb{E}_{\theta\sim\pi}[\phi^\star(\theta)] = u_0^\star \text{ and } \mathbb{V}\mathrm{ar}_{\theta\sim\pi}(\phi^\star(\theta)) = \sum_{i\geq 1} u_i^\star (u_i^\star)^\top.$$

Higher order moments are also available, in the case of polynomial basis (see [LK10, Appendix C]). For an approximation of more general statistics, one can and sample i.i.d. variables $\theta \sim \pi$ to efficiently calculate the empirical estimators.

Non-linearity of the problem. SA in infinite-dimensional spaces. The calculation of the chaos expansion coefficients for the SA limit $\phi^\star$ is non-trivial due to the non-linearity of the setting. Indeed, for a given θ the value $\phi^\star(\theta)$ may solve, for example, an optimization problem. Examples of chaos expansion in the linear setting include the works on UQ for linear PDEs, see e.g. [LK10, Section 4.4] and [GS91, Section 3.3]. Here, in certain cases, it is possible to reduce the problem to finite dimension if we are only interested in some truncation of $\phi^\star$. By contrast, in our setting, the projection of the equation (5.1.1) given (for some basis $\{B_i, i \geq 0\}$) by

$$\int_\Theta h\left(\sum_{i=0}^m u_i B_i(\theta), \theta\right) B_i(\theta)\pi(\mathrm{d}\theta) = 0, \qquad i = 0, \dots, m,$$

with the variable set restricted to $u_0, \dots, u_m$, will not yield the coefficients $u_0^\star, \dots, u_m^\star$ in the decomposition (5.1.2) of $\phi^\star$. The analysis of the corresponding error is quite complicated. In [KB09], the authors provide a finite dimensional procedure to approximate the function $\phi^\star$ that minimizes $\phi \mapsto \int_\Theta L(\phi(\theta), \theta)\pi(\mathrm{d}\theta)$, for some explicit function L, and they analyze the error due to such finite dimensional truncation. The error estimation is quite rough and provides little insight on how to fix the truncation level in advance.

Moreover, even when the function $\phi^\star$ is explicitly known, though estimating individual coefficients $u_i^\star$ in (5.1.2) is straightforward by MC simulations, the global convergence of a method where more and more coefficients are computed by Monte Carlo is subject to a nontrivial tuning of the speeds at which the number of coefficients and the number of simulations go to infinity (see [GS14]). In our case, the function $\phi^\star$ is unknown and is given by SA limits which is much more complex.

In Section 5.2.3 we argue that an infinite dimensional procedure is needed to achieve convergence to the true coefficients $\{u_i^\star, i \geq 0\}$ (or even a finite number of them). We design a method based on chaos expansion that is able to overcome the finite dimensional truncation by increasing the dimension. Thus our method belongs to the family of infinite dimensional SA algorithms, however it's fully constructive since at each iteration the dimension is finite.

We argue that previous works on infinite dimensional SA cannot be applied to our setting. There exists a large number of works on such SA methods. In [Wal77], [BS89] and [YZ90], the authors study SA in Hilbert spaces in the case $H(z, V) = \tilde{H}(z) + V$, where z here lives in a Hilbert space. The conditions of convergence are infinite dimensional formal analogues of those in the finite dimensional case (see Remark 5.3.3). Unfortunately, although interesting from a theoretical point of view, these SA algorithms are defined directly in the infinite

dimensional Hilbert space, so that they are not feasible in practice.

An alternative is to iteratively increase the dimension, to ∞ in the limit, while remaining finite at each iteration. There have been several papers in this direction, generally known as the sieve approach. [Gol88] proves almost-sure convergence in the norm topology for a modified Kiefer-Wolfowitz (see [KW52]) procedure in infinite dimensional Hilbert space using a sieve approach. [Nix84] shows asymptotic normality for a modified sieve-type Robbins-Monro procedure. [Yin92] proves almost-sure convergence in the weak topology for a sieve-type Robbins-Monro procedure. The latter three papers treat specific expressions of $H : \mathcal{H} \times \mathcal{V} \to \mathcal{H}$ while [CW02] combines the unrealistic approach (as in [Wal77, BS89, YZ90]) with the sieve approach (as in [Nix84, Gol88, Yin92]), deriving results on the convergence and asymptotic normality for SA with growing dimension in a quite general setting.

However, none of these literatures is adapted for dealing with UQ. Indeed, all of these previous papers in a infinite dimensional Hilbert space $\mathcal{H}$ solve problems of the form

$$\text{find } \phi^\star \in \mathcal{H} : \qquad \int H(\phi^\star, v)\mu(\mathrm{d}v) = 0,$$

so that, first, the distribution of V does not account for the uncertainty and, second, for any v, the computation of the quantity $H(\phi, v)$ may have a prohibitive computational cost depending on how a function ϕ appears in the definition of H.

Our SA algorithm in this chapter combines *(i)* the sieve approach in the special case $\mathcal{H}$ is L_2^π for some probability distribution π, *(ii)* the UQ framework by allowing μ to depend upon $\theta \in \Theta$, $\theta \sim \pi$, and *(iii)* a tractable computational cost of the function H. In the special case where μ does not depend on θ and $\mathcal{H} = L_2^\pi$, it can be compared to one of the methods in [CW02]: see the TRMP algorithm in [CW02, Section II]. But our method is able to address the case when the underlying scalar product of L_2^π is not explicit. Moreover, our proof of convergence - while addressing the more general framework where the scalar product is approximated - relies on weaker assumptions (see Remark 5.3.3, where we show that, under our assumptions for convergence, the assumptions of [CW02] may fail to hold).

We also remark that recently numerous works have been devoted to statistical learning in Hilbert spaces, in particular, reproducing kernel Hilbert spaces (RKHS, see e.g. [DB16] and references therein). However, statistical learning in Hilbert spaces reduces to finite dimensional SA: based on N input/output examples $\{(x_i, y_i), 1 \le i \le N\}$, it consists of solving

$$\mathrm{argmin}_{\varphi \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} \left(L(\varphi(x_i), y_i) + \Omega(\varphi) \right) \tag{5.1.3}$$

where $\mathcal{H}$ is a RKHS associated to a positive-definite real-valued kernel K, L is a non-negative loss function and $\Omega(f)$ is a penalty term. By the Representer Theorem, the solution admits a representation of the form $\phi^\star = \sum_{i=1}^N \omega_i K(\cdot, x_i)$ so that, under regularity conditions on L and Ω, the solution of (5.1.3) can be solved by a SA algorithm in $\mathbb{R}^N$.

Our contributions and summary. To the best of our knowledge this is the first work that studies rigorously the problem of Uncertainty Quantification for Stochastic Approximation limits. We provide the following contributions:

- A fully constructive, easy to implement, algorithm for UQ analysis of SA limits in a chaos expansion setup is obtained. It is dubbed USA (Uncertainty for Stochastic Approximation);

- A convergence proof is provided under easy-to-check hypotheses, in terms of underlying problems corresponding to fixed values of θ, avoiding conditions involving Hilbert space notions that are often hard to check in practice;

- Complexity issues, extensive reports and discussion on numerical tests are provided.

The chapter is outlined as follows. USA is introduced in Section 5.2. Section 5.3 states the almost-sure convergence of USA and its L^p convergence with respect to the underlying Hilbert space norm. The proof is deferred to Section 5.4. Section 5.5 presents the results of numerical experiments, including a detailed discussion of the choice of the design parameters.

Note that beyond model uncertainty, applications of our approach include sensitivity analysis with respect to θ, or quasi-regression of an unknown function (see [AO01]), for instance in the context of outer Monte Carlo computations involving some unknown inner function $\theta \mapsto \phi^\star(\theta)$, which are left for future research.

5.2 Problem Formulations and Algorithmic Solutions

Let $\mathcal{V}$ be a metric space endowed with its Borel σ-field, Θ be a subset of $\mathbb{R}^d$, and $H : \mathbb{R}^q \times \mathcal{V} \times \Theta \to \mathbb{R}^q$. Let π be a probability distribution on Θ and μ be a transition kernel from Θ to $\mathcal{V}$. We define the scalar product induced by π by

$$\langle f; g \rangle_\pi := \int_\Theta f(\theta) g(\theta) \pi(\mathrm{d}\theta), \tag{5.2.1}$$

for any measurable functions $f, g : \Theta \to \mathbb{R}$. By extension, for measurable functions $f = (f_1, \cdots, f_q) : \Theta \to \mathbb{R}^q$ and $g : \Theta \to \mathbb{R}$, we write in vector form

$$\langle f; g \rangle_\pi := \begin{bmatrix} \langle f_1; g \rangle_\pi \\ \cdots \\ \langle f_q; g \rangle_\pi \end{bmatrix}. \tag{5.2.2}$$

We denote by L_2^π the Hilbert space of functions $f : \Theta \to \mathbb{R}^q$ such that the norm $\|f\|_\pi := \sqrt{\sum_{i=1}^q \langle f_i; f_i \rangle_\pi}$ is finite (we omit the dependence on q since it will not lead to confusion).

We consider the following problem:

$$\text{Finding } \phi^\star \text{ in } L_2^\pi \text{ such that } \int_{\mathcal{V}} H(\phi^\star(\theta), v, \theta) \mu(\theta, \mathrm{d}v) = 0, \qquad \pi\text{-a.s.} \tag{5.2.3}$$

We work on a probability space with expectation denoted by $\mathbb{E}$.

5.2.1 SA Approach "θ by θ"

A naive approach for solving (5.2.3) is to calculate $\phi^\star(\theta)$ for each value of θ separately, for example by the following standard (unparameterized) SA scheme (see [BMP90, Duf97, KY97a]): for a fixed θ, given a deterministic sequence $\{\gamma_k, k \geq 1\}$ of positive step sizes and a sequence of independent and identically distributed (i.i.d.) r.v. $\{V_k, k \geq 1\}$ sampled from $\mu(\theta, \mathrm{d}v)$, we may obtain $z^\star = \phi^\star(\theta)$ as the limit of an iterative scheme

$$z^{k+1} = z^k - \gamma_{k+1} H(z^k, V_{k+1}, \theta). \tag{5.2.4}$$

Explicit conditions can be formulated in order to obtain the convergence $z^\star = \lim_k z^k$ (see e.g. [Duf97, Chapter 1]). However, except in the case where Θ is finite with few elements, the estimation of $\phi^\star(\theta)$, separately for each $\theta \in \Theta$, is too demanding computationally.

5.2.2 Chaos Extension Setup and Approach "Coefficient by Coefficient"

Let $\{\theta \mapsto B_i(\theta), i \geq 0\}$ be an orthonormal basis of L_2^π (for the scalar product (5.2.1)). Orthonormal polynomials are natural candidates, but there are other possibilities.

Example 5.2.1 (of orthogonal bases)**.** *See [CHQZ06, Chapter 2] for the four first examples based on orthogonal polynomials in dimension $d = 1$. An orthonormal basis $\{B_i, i \geq 0\}$ can then be obtained by renormalization of the given orthogonal basis.*

(i) If $\pi(\mathrm{d}\theta)$ has the density $1/(\pi\sqrt{1-\theta^2})$ with respect to the Lebesgue measure on $\Theta = [-1, 1]$, then the Chebyshev polynomials of the first kind form an orthogonal basis.

(ii) If $\pi(\mathrm{d}\theta)$ has the density $2\sqrt{1-\theta^2}/\pi$ w.r.t. the Lebesgue measure on $\Theta = [-1, 1]$, then the Chebyshev polynomials of the second kind form an orthogonal basis.

(iii) If $\pi(\mathrm{d}\theta)$ is the uniform distribution on the interval $\Theta = [-1, 1]$, then the Legendre polynomials form an orthogonal basis.

(iv) More generally, if $\pi(\mathrm{d}\theta)$ is the distribution on $[-1, 1]$ with density proportional to $(1-\theta)^\alpha(1+\theta)^\beta$ for some $\alpha, \beta > -1$, then the Jacobi polynomials form an orthogonal basis.

(v) If $\pi(\mathrm{d}\theta)$ is the uniform distribution on the interval $\Theta = [-\pi, \pi]$, then we have the orthogonal Fourier basis $\{1, \cos(i\theta), \sin(i\theta), i \geq 1\}$.

(vi) If $\{B_i, i \geq 0\}$ is an orthogonal basis on $\Theta \subset \mathbb{R}$ with respect to the distribution $\pi(\mathrm{d}\theta) = \pi(\theta)\mathrm{d}\theta$, then, for any continuously differentiable increasing function φ, $\{B_i(\varphi(\cdot)), i \geq 0\}$ is an orthogonal basis on $\varphi^{-1}(\Theta)$ with respect to the distribution $\pi(\varphi(v))\varphi'(v)\mathrm{d}v$.

(vii) For a multidimensional distribution ($d > 1$), with independent components, an orthogonal basis is given by the set of all possible products of basis functions of a single variable (see [CHQZ06, Section 5.8]).

For $x, y \in \mathbb{R}^q$ we denote by $x \cdot y$ and $|x|$ the scalar product and the Euclidean norm in $\mathbb{R}^q$. We denote by l_2 the normed vector space of the $\mathbb{R}^q$-valued sequences $\{u_i, i \geq 0\}$ with

$\|u\|_{l_2}^2 = \sum_{i\geq 0} |u_i|^2 < +\infty$. As is well known, given an orthonormal basis $\{B_i, i \geq 0\}$ in L_2^π, any function $\phi \in L_2^\pi$ is characterized by a sequence $\{u_i, i \geq 0\}$ in l_2 such that $\phi = \sum_{i\geq 0} u_i B_i$. Throughout the chapter, we use the natural isomorphism $\mathbf{Is} : l_2 \to L_2^\pi$ given by

$$\phi = \mathbf{Is}(u) = \sum_{i\geq 0} u_i B_i, \text{ i.e. } u_i = \langle \phi; B_i \rangle_\pi \text{ for each } i \in \mathbb{N}, \tag{5.2.5}$$

and the corresponding isometry $\|\phi\|_\pi = \|u\|_{l_2}$ (see [Mus14, Proposition 10.32]).

Assuming $\phi^\star \in L_2^\pi$, an alternative strategy for solving (5.2.3) consists of the estimation of the $\mathbb{R}^q$-valued coefficients $\{u_i^\star, i \geq 0\}$ of $\phi^\star$, combined with a truncation at a fixed level m of the expansion (5.2.5) and a Monte Carlo approximation of the coefficients $\{u_i^\star, i \leq m\}$, i.e., for $i \leq m$,

$$u_i^\star = \langle \phi^\star; B_i \rangle_\pi \approx \hat{u}_i^\star := \frac{1}{M} \sum_{k=1}^{M} \widehat{\phi^\star(\theta_{k,i})}\, B_i(\theta_{k,i}), \tag{5.2.6}$$

where $\{\theta_{k,i}, k \geq 1, i \leq m\}$ are i.i.d. with distribution π and $\widehat{\phi(\theta_{k,i})}$ is an approximation of $\phi(\theta_{k,i})$. Let us discuss the computational cost of this approach, in the case $q = 1$ for ease of notation (and dimension d of θ). In the case of a Jacobi polynomial basis, the following control on the truncation error of ϕ holds (see [Fun92, Theorem 6.4.2] or [CHQZ06, Chapter 5]):

$$\left\| \sum_{i>m} u_i B_i \right\|_\pi^2 = O\left(m^{-\frac{2(\eta-1)}{d}} \right), \tag{5.2.7}$$

where η is the order of continuous differentiability of ϕ (in some cases the order may be strengthened to $O\left(m^{-\frac{2\eta}{d}}\right)$). Neglecting the error associated with the approximation $\widehat{\phi(\theta_{k,i})} \approx \phi(\theta_{k,i})$, we have

$$\mathbb{E}\left[\left\| \sum_{i=0}^{m} (u_i - \hat{u}_i) B_i \right\|_\pi^2 \right] = O\left(\frac{m}{M} \right). \tag{5.2.8}$$

For balancing the error components (5.2.7) and (5.2.8), we must set $M \propto m^{1+\frac{2(\eta-1)}{d}}$. To reach a precision ϵ, m has to increase as $\epsilon^{-d/(2(\eta-1))}$ and M has to increase as $\epsilon^{-(1+d/(2(\eta-1)))}$. The computational cost in terms of number of Monte Carlo samples to estimate m coefficients is therefore $\epsilon^{-(1+d/(\eta-1))}$. This quantity suffers from the curse of dimensionality, which makes this approach fairly inefficient when combined with a nested procedure for the computation of $\widehat{\phi(\theta_{k,i})}$, e.g. through (5.2.4) if $\phi = \phi^\star$.

Also note that in the simple case where SA is reduced to MC (i.e. for $H(z, v) = z - v$), the approximation $\widehat{\phi^\star(\theta_{k,i})}$ in (5.2.6) will be given by a second MC, so that (5.2.6) results in a two-stage MC procedure which thus converges two times slower and is highly inefficient. The right approach here would be to approximate the average $\int_{\Theta\times\mathcal{V}} v\mu(\theta, \mathrm{d}v) B_i(\theta)\pi(\mathrm{d}\theta)$ directly using i.i.d. simulations from $(\theta, V) \sim \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta)$. This motivates the construction of an algorithm which couples in an efficient way the outer Monte Carlo sampling of θ and the inner Monte Carlo sampling used to feed the SA algorithm in order to optimize the

calculation.

5.2.3 The USA Algorithm

Through the isomorphism (5.2.5), the problem (5.2.3) can be restated on l_2 as

$$\text{Finding } u^\star \text{ in } l_2; \quad \int_{\mathcal{V}} H\left(\sum_{i\geq 0} u_i^\star B_i(\theta), v, \theta\right) \mu(\theta, \mathrm{d}v) = 0, \ \pi\text{-a.s..} \tag{5.2.9}$$

Note that the problem (5.2.3) is equivalent to finding $\phi^\star \in L_2^\pi$ such that

$$\int_\Theta \left(\int_{\mathcal{V}} H\left(\phi^\star(\theta), v, \theta\right) \ \mu(\theta, \mathrm{d}v)\right) B_i(\theta)\pi(\mathrm{d}\theta) = 0_{\mathbb{R}^q}, \qquad \forall i \geq 0. \tag{5.2.10}$$

This observation can be used for devising an original SA scheme for the $u_i^\star$ in (5.2.9).

A first attempt in this direction is to restrict the problem to a set of functions ϕ of the form $\sum_{i=0}^m u_i B_i$, for some fixed $m \geq 1$. If, in addition, $\mu(\theta, \mathrm{d}v) = \mu(\mathrm{d}v)$, and assuming the scalar product $\langle \cdot ; \cdot \rangle_\pi$ (corresponding to the integral in (5.2.11)) computable exactly (possibly at a large computational cost), then an SA algorithm for the computation of $\{u_i^\star, i \leq m\}$ consists of iterating:

$$u_i^{k+1} = u_i^k - \gamma_{k+1} \int_\Theta H\left(\sum_{j=0}^m u_j^k B_j(\theta), V_{k+1}, \theta\right) B_i(\theta)\, \pi(\mathrm{d}\theta) \qquad i = 0, \cdots, m, \tag{5.2.11}$$

where the $\{(V_k), k \geq 0\}$ are i.i.d. with distribution $\mu(\mathrm{d}v)$ and $\{\gamma_k, k \geq 1\}$ is a deterministic stepsize sequence. In the more general case, an SA algorithm for the computation of $\{u_i^\star, i \leq m\}$ is given by

$$u_i^{k+1} = u_i^k - \gamma_{k+1}\ H\left(\sum_{j=0}^m u_j^k B_j(\theta_{k+1}), V_{k+1}, \theta_{k+1}\right) B_i(\theta_{k+1}), \quad i = 0, \cdots, m, \tag{5.2.12}$$

where the $\{(\theta_k, V_k), k \geq 0\}$ are i.i.d. with distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$.

However, in practice, $\phi^\star$ is typically not of the form $\sum_{i=0}^m u_i B_i$ and, even when it is, we may not know for which m. We emphasize that, in the general case $\phi^\star \in L_2^\pi$, as the first argument of H in (5.2.12) is the current truncation $\sum_{i=0}^m u_i^k B_i(\theta_{k+1})$ and not $\phi^\star(\theta_{k+1})$, this algorithm does not converge to the projection of $\phi^\star$ onto the space spanned by $\{B_0, \cdots, B_m\}$. See the numerical evidence reported in Section 5.5.4.

Accordingly, the final version of the algorithm tackles the infinite dimensionality of the problem space L_2^π on which the problem is stated by increasing m, to recover in the limit the full sequence of the coefficients $\{u_i^\star, i \geq 0\}$ defining a solution $\phi^\star = \sum_i u_i^\star B_i$. Toward this aim, we introduce a sequence m_k which specifies the number of coefficients u_i that are updated at the iteration k. The sequence $\{m_k, k \geq 0\}$ is nondecreasing and converges to ∞.

The USA algorithm corresponds to the update of the sequence $\{u_i^k, i \geq 0\}$ through the

following SA scheme, where $\Pi_{\mathcal{A}}$ denotes the projection on a suitable convex subset $\mathcal{A}$ of l_2:

Input: Sequences $\{\gamma_k, k \geq 1\}$, $\{m_k, k \geq 1\}$, $\{M_k, k \geq 1\}$, $K \in \mathbb{N}$, $\{u_i^0, i = 0, \ldots, m_0\}$, a convex set $\mathcal{A} \subseteq l_2$

for $k = 0$ **to** $K - 1$, **do**

  sample $(\theta_{k+1}^s, V_{k+1}^s), s = 1 \ldots, M_{k+1}$, under the distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$;

  for $i > m_{k+1}$ set $u_i^{k+1} = 0$;

  for $i = 0$ **to** m_{k+1}, **do**

$$\hat{u}_i^{k+1} = u_i^k - \gamma_{k+1} M_{k+1}^{-1} \sum_{s=1}^{M_{k+1}} H\left(\sum_{j=0}^{m_k} u_j^k B_j(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) B_i(\theta_{k+1}^s)$$

$$u^{k+1} = \Pi_{\mathcal{A}}(\hat{u}^{k+1})$$

Output: The vector $\{u_i^K, i = 0, \ldots, m_K\}$.

Algorithm 3: The USA algorithm for the coefficients of the basis decomposition of $\phi^\star$.

The inputs of the algorithm are: a positive stepsize sequence $\{\gamma_k, k \geq 1\}$; two integer valued sequences $\{m_k, k \geq 1\}$ and $\{M_k, k \geq 1\}$ corresponding to the number of nonnull coefficients in the approximation of $\phi^\star$ and to the number of Monte Carlo draws of the pair (θ, V) at each iteration k; an initial value $u_0 \in \mathbb{R}^{m_0}$; a total number of iterations K; a convex subset $\mathcal{A}$ of l_2 on which to project each newly updated sequence of coefficients.

The output of the algorithm is a sequence $u^K = \{u_i^K, i \leq m_K\}$ approximating a solution $u^\star$ to the problem (5.2.9). The corresponding approximation ϕ^K of a solution $\phi^\star$ to the problem (5.2.3) is then

$$\phi^K := \sum_{i=0}^{m_K} u_i^K B_i. \tag{5.2.13}$$

Remark 5.2.2. *The motivations for the introduction of the projection set $\mathcal{A}$ and for the averaging over M_k draws at step k are discussed in the respective sections 5.3.2 and 5.5.5.*

5.3 The USA Algorithm Converges

5.3.1 Assumptions

For simplicity of presentation above, we assumed a uniqueness of $\phi^\star$. However, the USA algorithm is proved below to converge even in the case of multiple zeros. Accordingly, Problem (5.2.9) is reformulated as

Finding $u^\star$ in $\mathcal{T}^\star$ where

$$\mathcal{T}^\star := \left\{ u^\star \in l_2; \qquad \int_{\mathcal{V}} H\left(\sum_{i \geq 0} u_i^\star B_i(\theta), v, \theta\right) \mu(\theta, \mathrm{d}v) = 0, \ \pi\text{-a.s.} \right\}. \tag{5.3.1}$$

We do not restrict ourselves to the case of a singleton $\mathcal{T}^\star$. However, we introduce the following assumption on the target set $\mathcal{T}^\star$ in order to guarantee the existence of a (random) limit point ϕ^∞ of the algorithm in this set:

C1. *The set $\mathcal{T}^\star$ is compact and non-empty.*

Remark 5.3.1. *Allowing for multiple limits is quite standard in the SA literature. From the point of view of the application to UQ, it may seem meaningless to quantify the uncertainty of a non-uniquely defined quantity. However, enabling multiple limits appears to be the right setting when some components of the vector-valued function $\phi^\star(\cdot) \in \boldsymbol{Is}(\mathcal{T}^\star)$ are unique and some other are multiple. This encompasses the important case of computing quantiles and average quantiles (cf. [BFP09]) of a (uncertain) distribution: the SA approximation for the quantile component may converge to several limits, while for the average quantile component, the limit is unique.*

C2. *$\{M_k, k \geq 1\}$ and $\{m_k, k \geq 1\}$ are deterministic sequences of positive integers; $\{\gamma_k, k \geq 1\}$ is a deterministic sequence of positive real numbers such that, for some $\kappa > 0$,*

$$\sum_{k\geq 1} \gamma_k = +\infty, \sum_{k\geq 1} \gamma_k^{1+\kappa} < +\infty, \sum_{k\geq 1} \gamma_k^2 \frac{Q_{m_k}}{M_k} < +\infty, \sum_{k\geq 1} \gamma_k^{1-\kappa} q_{m_k} < +\infty, \tag{5.3.2}$$

where the sequences $\{q_m, m \geq 0\}$ and $\{Q_m, m \geq 0\}$ are defined by

$$q_m := \sup_{u^\star \in \mathcal{T}^\star} \sum_{i>m} |u_i^\star|^2, \qquad Q_m := \sup_{\theta\in\Theta} \sum_{i\leq m} |B_i(\theta)|^2. \tag{5.3.3}$$

Remark 5.3.2. *Since $\mathcal{T}^\star$ is compact, we have $\lim_m q_m = 0$ (cf. the proof of Lemma 5.B.1). Assumption **C2** requires, in particular, that $Q_m < +\infty$ for any m. If Θ is bounded, then this is verified for any basis of continuous functions. In the case of polynomial basis, the coefficients Q_m are related to the Christoffel functions [Nev86].*

C3. *For any $z \in \mathbb{R}^q$,*

$$\int_{\Theta\times\mathcal{V}} |H(z, v, \theta)|\, \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta) < \infty;$$

For any $z \in \mathbb{R}^q$ and $\theta \in \Theta$,

$$h(z, \theta) = \int_{\mathcal{V}} H(z, v, \theta)\, \mu(\theta, \mathrm{d}v)$$

exists; For any $\phi \in L_2^\pi$, the mapping $h(\phi(\cdot), \cdot) : \theta \mapsto h(\phi(\theta), \theta)$ is in L_2^π; The mapping $\phi \mapsto h(\phi(\cdot), \cdot)$ from L_2^π into itself is continuous.

C4. *For π-almost every θ, for any $z_\theta, z_\theta^\star \in \mathbb{R}^q$ such that $h(z_\theta, \theta) \neq 0$ and $h(z_\theta^\star, \theta) = 0$,*

$$(z_\theta - z_\theta^\star) \cdot h(z_\theta, \theta) > 0.$$

Remark 5.3.3. *Previous works on SA in a Hilbert space $\mathcal{H}$ typically require an assumption of the type*

$$\int_\Theta (\phi(\theta) - \phi^\star(\theta)) \cdot \hat{h}^m(\phi(\theta), \theta)\pi(\mathrm{d}\theta) > 0, \quad \forall \phi \in L_2^\pi \setminus \boldsymbol{Is}(\mathcal{T}^\star), \phi^\star \in \boldsymbol{Is}(\mathcal{T}^\star),$$

for m large enough, where $\hat{h}^m(\phi(\cdot), \cdot)$ is the approximation of $h(\phi(\cdot), \cdot)$ using the first m elements of a basis of $\mathcal{H}$: See e.g. [CW02, Assumption A3P(2)], which only requires the

above condition for every $\phi \neq \phi^\star$ in the vector space spanned by the first m basis functions B_i. However, even this relaxed assumption does not hold in general in our setting. As a counter-example, one may take any $\phi^\star = \sum_{i\geq 0} u_i^\star B_i$ with non null coefficients $u_i^\star$, $h(z,\theta) = z - \phi^\star(\theta)$, and $\phi = \phi^m$ given, for every m, as the truncation

$$\phi^m := \mathrm{Trunc}_m(\phi^\star) = \sum_{i\leq m} u_i^\star B_i$$

of order m of $\phi^\star$. Then, as $\mathrm{Trunc}_m(\phi^m - \phi^\star) = 0_{L_2^\pi}$ (by definition of ϕ^m), we have

$$\int_\Theta (\phi^m(\theta) - \phi^\star(\theta)) \cdot \hat{h}^m(\phi(\theta),\theta)\pi(\mathrm{d}\theta) = \int (\phi^m(\theta) - \phi^\star(\theta)) \cdot \mathrm{Trunc}_m(\phi^m - \phi^\star)(\theta)\pi(\mathrm{d}\theta) = 0,$$

for every m.

*By contrast, **C4** is the standard assumption for SA with fixed θ.*

C5. *a) There exists a constant C_H such that, for any $z \in \mathbb{R}^q$,*

$$\sup_{\theta\in\Theta} \int_{\mathcal{V}} |H(z,v,\theta)|^2 \mu(\theta,\mathrm{d}v) \leq C_H(1+|z|^2).$$

b) The map from L_2^π into $\mathbb{R}$ defined by $\phi \mapsto \int_{\mathcal{V}\times\Theta} |H(\phi(\theta),v,\theta)|^2 \pi(\mathrm{d}\theta)\mu(\theta,\mathrm{d}v)$ is bounded, i.e. it maps bounded sets into bounded sets.

Note that **C5**-b implies that $\phi \mapsto h(\phi(\cdot),\cdot)$ is a bounded map from L_2^π into itself.

C6. *For any $B > 0$, there exists a constant $C_B > 0$ such that, for any $(\phi,\phi^\star) \in L_2^\pi \times \boldsymbol{Is}(\mathcal{T}^\star)$ with $\|\phi - \phi^\star\|_\pi \leq B$,*

$$\int (\phi - \phi^\star)(\theta) \cdot h(\phi(\theta),\theta)\, \pi(\mathrm{d}\theta) \geq C_B \min_{\bar{\phi}\in \boldsymbol{Is}(\mathcal{T}^\star)} \left\|\phi - \bar{\phi}\right\|_\pi^2.$$

Note that the above minimum exists since $\mathbf{Is}(\mathcal{T}^\star)$ is compact, by **C1**.

5.3.2 Projection Set

We address the convergence of the algorithm 3 for three possible choices regarding the projection set $\mathcal{A}$ (which always includes $\mathcal{T}^\star$).

Case 1. $\mathcal{A} := l_2$.

Case 2. *$\mathcal{A}$ is a closed ball of l_2 containing $\mathcal{T}^\star$.*

Case 3. *$\mathcal{A}$ is a closed convex set of l_2 containing $\mathcal{T}^\star$, with compact intersections with closed balls of l_2.*

Note that the projection set $\mathcal{A}$ is bounded in Case 2 and unbounded in the two other cases (for sure in Case 1 and potentially in 3).

Case 1 is the most convenient from the algorithmic viewpoint since no actual projection is required. However, it requires a stronger condition **C5**-a to ensure the stability and an additional assumption **C6** for the convergence.

The projection on a ball $\{u \in l_2 : \|u\|_{l_2} \leq B\}$ is given simply by

$$u \mapsto \min\left(1, \frac{B}{\|u\|_{l_2}}\right) u. \tag{5.3.4}$$

Hence, the projection required in Case 2 is quite straightforward. The milder assumption **C5**-b is required for the stability but one still needs **C6** for the convergence.

Case 3 requires a potentially nontrivial projection on a closed convex set: see e.g. Example 5.3.4 below. The stronger condition **C5**-a is required for both the stability and the convergence, but **C6** is not needed.

We now give an example of the set $\mathcal{A}$ in Case 3.

Example 5.3.4. *Given a positive sequence $\{a_n, n \geq 0\}$ such that $\sum_{i\geq 0} a_i^2 < \infty$ and an increasing sequence of non-negative integers $\{d_n, n \geq 0\}$, define the closed convex set $\mathcal{A}$:*

$$\mathcal{A} := \left\{ u \in l_2 : \sum_{d_n \leq i < d_{n+1}} |u_i|^2 \leq a_n^2 \quad \forall n \geq 0 \right\}. \tag{5.3.5}$$

When $d_0 = 0$, the set $\mathcal{A}$ is a compact convex subset of l_2 (see Lemma 5.B.1). Otherwise, it is not necessarily compact. However, the set $\mathcal{A} \cap \{u \in l_2 : \sum_{i\geq 0} u_i^2 \leq B\}$ is a compact subset for any $B > 0$ (see Corollary 5.B.2). The orthogonal projection on $\mathcal{A}$ consists of projecting $(u_{d_n}, \ldots, u_{d_{n+1}-1})$ on the ball of radius a_n for all $n \geq 0$.

5.3.3 Main Result

Theorem 5.3.5. *Assume **C1** to **C4** and **C5**-a if $\mathcal{A}$ is unbounded or **C5**-b if $\mathcal{A}$ is bounded. Let there be given i.i.d. random variables $\{(\theta_k^s, V_k^s), 1 \leq s \leq M_k, k \geq 1\}$ with distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$. Let u^K and ϕ^K be the outputs of the USA Algorithm (cf. (5.2.13)).*

Stability. *For any $\phi^\star \in \boldsymbol{Is}(\mathcal{T}^\star)$, $\lim_{k\to+\infty} \left\|\phi^k - \phi^\star\right\|_\pi$ exists, is finite* a.s., *and we have*

$$\sup_{k\geq 0} \mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right] < +\infty. \tag{5.3.6}$$

Convergence. *In addition, in case 3, and in cases 1 and 2 under the additional assumption **C6**, there exists a random variable ϕ^∞ taking values in $\boldsymbol{Is}(\mathcal{T}^\star)$ such that*

$$\lim_{k\to\infty} \left\|\phi^k - \phi^\infty\right\|_\pi = 0 \text{ a.s. } \textit{and, for any } p \in (0,2), \lim_{k\to\infty} \mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^p\right] = 0. \tag{5.3.7}$$

Remark 5.3.6. *The standard assumption ensuring a central limit theorem (CLT) for SA algorithms in a Hilbert space (cf. [CW02, Assumption B3(1)] or [Nix84, Section 3, equation*

3.3]) is not satisfied in our setup: as a counter-example, one can take a function $h(z,\theta)$ such that $\partial_z h(\phi^\star(\theta),\theta) = \theta$ and any polynomial basis, due to the recurrence relations of order two that are intrinsic to such bases. The study of convergence rates and CLT for the USA algorithm is therefore a problem per se, which we leave for future research.

5.4 Proof of Theorem 5.3.5

Throughout the proof, we will use the notation

$$\phi^k := \sum_{i\geq 0} u_i^k B_i = \sum_{i=0}^{m_k} u_i^k B_i$$

(recalling that $u_i^k = 0$ for any $i > m_k$ in the USA Algorithm). For any $z = (z_1, \cdots, z_q) \in \mathbb{R}^q$ and any real-valued sequence $p := \{p_i, i \geq 0\}$ such that $\sum_{i\geq 0} p_i^2 < \infty$ we write

$$z \otimes p := ((z_1 p_0, \cdots, z_q p_0), (z_1 p_1, \cdots, z_q p_1), \cdots) \in l_2.$$

Set $\mathbf{B}^m(\theta) := (B_0(\theta), \ldots, B_m(\theta), 0, 0, \ldots)$. Define the filtration

$$\mathcal{F}_k := \sigma\left(\theta_\ell^s, V_\ell^s, 1 \leq s \leq M_\ell, 1 \leq \ell \leq k\right),\ k \geq 1.$$

We fix $u^\star \in \mathcal{T}^\star$, which exists by C1, and we set $\phi^\star := \mathbf{Is}(u^\star)$.

5.4.1 Stability

The first step is to prove that the algorithm is stable in the sense that

$$\lim_k \left\|u^k - u^\star\right\|_{l_2} \text{ exists a.s. }, \tag{5.4.1}$$

$$\sup_k \mathbb{E}\left[\left\|u^k - u^\star\right\|_{l_2}^2\right] < +\infty, \tag{5.4.2}$$

$$\liminf_{k\to\infty} \int_\Theta (\phi^k(\theta) - \phi^\star(\theta)) \cdot h(\phi^k(\theta),\theta)\pi(\mathrm{d}\theta) = 0, \quad \text{a.s.} \tag{5.4.3}$$

Using the definition of u^{k+1} in the USA algorithm and the property $\Pi_{\mathcal{A}}(u^\star) = u^\star$, we obtain (recalling that, in all cases 1 to 3, $\mathcal{T}^\star \subseteq \mathcal{A}$)

$$\begin{aligned}\left\|\phi^{k+1} - \phi^\star\right\|_\pi^2 = \left\|u^{k+1} - u^\star\right\|_{l_2}^2 = \left\|\Pi_{\mathcal{A}}(\hat{u}^{k+1}) - \Pi_{\mathcal{A}}(u^\star)\right\|_{l_2}^2 \leq \left\|\hat{u}^{k+1} - u^\star\right\|_{l_2}^2 \\ = \left\|u^k - u^\star - \gamma_{k+1}\mathcal{H}^k - \gamma_{k+1}\eta^{k+1}\right\|_{l_2}^2,\end{aligned}$$

where

$$\mathcal{H}^k := \mathbb{E}\left[\frac{1}{M_{k+1}} \sum_{s=1}^{M_{k+1}} H\left(\phi^k(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) \otimes \mathbf{B}^{m_{k+1}}(\theta_{k+1}^s) \middle| \mathcal{F}_k\right]$$

$$= \int_{\Theta\times\mathcal{V}} H\left(\phi^k(\theta), v, \theta\right) \otimes \mathbf{B}^{m_{k+1}}(\theta)\ \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$$
$$= \int_{\Theta} h(\phi^k(\theta), \theta) \otimes \mathbf{B}^{m_{k+1}}(\theta)\ \pi(\mathrm{d}\theta),$$
$$\eta^{k+1} := \frac{1}{M_{k+1}} \sum_{s=1}^{M_{k+1}} H\left(\phi^k(\theta^s_{k+1}), V^s_{k+1}, \theta^s_{k+1}\right) \otimes \mathbf{B}^{m_{k+1}}(\theta^s_{k+1}) - \mathcal{H}^k.$$

For the equivalent definitions of $\mathcal{H}^k$, we used the Fubini theorem and **C3**. Observe that, by definition of $\mathbf{B}^{m_k}$, $\mathcal{H}^k$ and η^{k+1} are sequences in l_2 such that, for all $i > m_{k+1}$,

$$\mathcal{H}^k_i = 0_{\mathbb{R}^q}, \qquad \eta^{k+1}_i = 0_{\mathbb{R}^q}.$$

Define

$$\overline{\mathcal{H}}^k_i := \begin{cases} \mathcal{H}^k_i & i \le m_{k+1}, \\ \int_\Theta h(\phi^k(\theta), \theta) B_i(\theta)\ \pi(\mathrm{d}\theta) & i > m_{k+1}. \end{cases}$$

Recalling that $u^k_i = 0_{\mathbb{R}^q}$ for $i > m_{k+1}$, we obtain

$$\begin{aligned}
\left\|u^{k+1} - u^\star\right\|^2_{l_2} &= \left\|u^k - u^\star\right\|^2_{l_2} - 2\gamma_{k+1} \sum_{i=0}^{m_{k+1}} (u^k_i - u^\star_i)\cdot \mathcal{H}^k_i \\
&\quad - 2\gamma_{k+1} \sum_{i=0}^{m_{k+1}} (u^k_i - u^\star_i)\cdot \eta^{k+1}_i \\
&\quad + 2\gamma^2_{k+1} \sum_{i=0}^{m_{k+1}} \eta^{k+1}_i \cdot \mathcal{H}^k_i + \gamma^2_{k+1} \left\|\eta^{k+1}\right\|^2_{l_2} + \gamma^2_{k+1} \left\|\mathcal{H}^k\right\|^2_{l_2} \\
&= \left\|u^k - u^\star\right\|^2_{l_2} - 2\gamma_{k+1} \sum_{i\ge 0} (u^k_i - u^\star_i)\cdot \overline{\mathcal{H}}^k_i \\
&\quad + 2\gamma^2_{k+1} \sum_{i=0}^{m_{k+1}} \eta^{k+1}_i \cdot \mathcal{H}^k_i - 2\gamma_{k+1} \sum_{i=0}^{m_{k+1}} (u^k_i - u^\star_i)\cdot \eta^{k+1}_i \\
&\quad - 2\gamma_{k+1} \sum_{i>m_{k+1}} u^\star_i \cdot \overline{\mathcal{H}}^k_i + \gamma^2_{k+1} \left\|\eta^{k+1}\right\|^2_{l_2} + \gamma^2_{k+1} \left\|\mathcal{H}^k\right\|^2_{l_2}.
\end{aligned} \tag{5.4.4}$$

C4 implies that, for each θ,

$$\sum_{i\ge 0} \left((u^k_i - u^\star_i)\cdot h\left(\phi^k(\theta), \theta\right)\right) B_i(\theta) = (\phi^k(\theta) - \phi^\star(\theta))\cdot h\left(\phi^k(\theta), \theta\right) \ge 0.$$

Taking expectation with respect to $\theta \sim \pi$ and applying the Fubini theorem (which follows from **C3**), we obtain for all $k \ge 0$

$$R^k := \int_\Theta (\phi^k(\theta) - \phi^\star(\theta))\cdot h\left(\phi^k(\theta), \theta\right) \pi(\mathrm{d}\theta) = \sum_{i\ge 0} (u^k_i - u^\star_i)\cdot \overline{\mathcal{H}}^k_i \ge 0. \tag{5.4.5}$$

Note also that $\sum_{i=0}^{+\infty}(u_i^k - u_i^\star)\cdot\mathcal{H}_i^k \in \mathcal{F}_k$. By definition, $\mathbb{E}\left[\eta_i^{k+1}|\mathcal{F}_k\right] = 0$, so that

$$\mathbb{E}\left[\sum_{i=0}^{m_{k+1}} \eta_i^{k+1}\cdot\mathcal{H}_i^k\Big|\mathcal{F}_k\right] = 0, \qquad \mathbb{E}\left[\sum_{i=0}^{m_{k+1}} (u_i^k - u_i^\star)\cdot\eta_i^{k+1}\Big|\mathcal{F}_k\right] = 0. \tag{5.4.6}$$

Let us consider the term $\left\|\eta^{k+1}\right\|_{l_2}^2$. We write

$$\begin{aligned}
\mathbb{E}\big[\left\|\eta^{k+1}\right\|_{l_2}^2|\mathcal{F}_k\big] &\le \mathbb{E}\left[\left\|\frac{1}{M_{k+1}}\sum_{s=1}^{M_{k+1}} H\left(\phi^k(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right)\otimes\mathbf{B}^{m_{k+1}}(\theta_{k+1}^s) - \mathcal{H}^k\right\|_{l_2}^2|\mathcal{F}_k\right] \\
&\le \frac{1}{M_{k+1}}\int_{\Theta\times\mathcal{V}}\left\|H\left(\phi^k(\theta), v, \theta\right)\otimes\mathbf{B}^{m_{k+1}}(\theta)\right\|_{l_2}^2\pi(\mathrm{d}\theta)\mu(\theta,\mathrm{d}v) \\
&= \frac{1}{M_{k+1}}\int_{\Theta\times\mathcal{V}}\left|H\left(\phi^k(\theta), v, \theta\right)\right|^2\left(\sum_{i=0}^{m_{k+1}} B_i(\theta)^2\right)\pi(\mathrm{d}\theta)\mu(\theta,\mathrm{d}v) \\
&\le \frac{Q_{m_{k+1}}}{M_{k+1}}\int_{\Theta\times\mathcal{V}}\left|H\left(\phi^k(\theta), v, \theta\right)\right|^2\pi(\mathrm{d}\theta)\mu(\theta,\mathrm{d}v).
\end{aligned} \tag{5.4.7}$$

Next we consider the term $2\gamma_{k+1}\sum_{i>m_{k+1}} u_i^\star\cdot\overline{\mathcal{H}}_i^k$. By using $2ab \le a^2 + b^2$ with $a \leftarrow (\gamma_{k+1}^{1-\kappa})^{1/2}|u_i^\star|$ and $b \leftarrow (\gamma_{k+1}^{1+\kappa})^{1/2}|\overline{\mathcal{H}}_i^k|$, we have

$$\begin{aligned}
\left|2\gamma_{k+1}\sum_{i>m_{k+1}} u_i^\star\cdot\overline{\mathcal{H}}_i^k\right| &\le \gamma_{k+1}^{1-\kappa}\left(\sum_{i>m_{k+1}}^{+\infty}|u_i^\star|^2\right) + \gamma_{k+1}^{1+\kappa}\left(\sum_{i>m_{k+1}}^{+\infty}\left|\overline{\mathcal{H}}_i^k\right|^2\right) \\
&\le \gamma_{k+1}^{1-\kappa}q_{m_{k+1}} + \gamma_{k+1}^{1+\kappa}\left\|\overline{\mathcal{H}}^k\right\|_{l_2}^2,
\end{aligned} \tag{5.4.8}$$

where we used **C2** in the last inequality. Note that

$$\begin{aligned}
\left\|\overline{\mathcal{H}}^k\right\|_{l_2}^2 &= \sum_{i=0}^{+\infty}\left|\int_\Theta h(\phi^k(\theta),\theta)B_i(\theta)\ \pi(\mathrm{d}\theta)\right|^2 = \int_\Theta\left|h(\phi^k(\theta),\theta)\right|^2\ \pi(\mathrm{d}\theta) \\
&\le \int_{\Theta\times\mathcal{V}}\left|H\left(\phi^k(\theta), v, \theta\right)\right|^2\pi(\mathrm{d}\theta)\mu(\theta,\mathrm{d}v).
\end{aligned} \tag{5.4.9}$$

Combining (5.4.4), (5.4.5), (5.4.6), (5.4.7), (5.4.8), and (5.4.9), we obtain

$$\begin{aligned}
\mathbb{E}\big[\left\|u^{k+1} - u^\star\right\|_{l_2}^2|\mathcal{F}_k\big] &\le \left\|u^k - u^\star\right\|_{l_2}^2 - 2\gamma_{k+1}R^k + \gamma_{k+1}^{1-\kappa}q_{m_{k+1}} \\
&\quad + \left(\gamma_{k+1}^2 + \gamma_{k+1}^{1+\kappa} + \gamma_{k+1}^2\frac{Q_{m_{k+1}}}{M_{k+1}}\right)\int_{\Theta\times\mathcal{V}}\left|H\left(\phi^k(\theta), v, \theta\right)\right|^2\pi(\mathrm{d}\theta)\mu(\theta,\mathrm{d}v).
\end{aligned} \tag{5.4.10}$$

To control the integral in (5.4.10), we distinguish two cases.

First case: $\mathcal{A}$ is unbounded. Using **C5**-a we write

$$\int_{\Theta\times\mathcal{V}} \left|H\left(\phi^k(\theta), v, \theta\right)\right|^2 \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v) \le C_H \int_\Theta \left(1 + \left|\phi^k(\theta)\right|^2\right) \pi(\mathrm{d}\theta)$$
$$\le C_1 \left(1 + \left\|u^k - u^\star\right\|_{l_2}^2\right),$$

where $C_1 := 2C_H(1 + \sup_{u^\star\in\mathcal{T}^\star} \|u^\star\|_{l_2}^2)$. Note that C_1 is finite by **C1**.

Second case: $\mathcal{A}$ is bounded. Note that, by definition of u^k, there exists a constant B such that a.s. $\sup_{k\ge0} \left\|u^k\right\|_{l_2} \le B$. Assumption **C5**-b implies that, for some finite and positive C_2,

$$\sup_{k\ge0} \int_{\mathcal{V}\times\Theta} |H(\phi^k(\theta), v, \theta)|^2 \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v) \le C_2.$$

In either case, we deduce from (5.4.10) that

$$\mathbb{E}\left[\left\|u^{k+1} - u^\star\right\|_{l_2}^2 |\mathcal{F}_k\right] \le \left\|u^k - u^\star\right\|_{l_2}^2 - 2\gamma_{k+1}R^k + \gamma_{k+1}^{1-\kappa} q_{m_{k+1}} \tag{5.4.11}$$
$$+ \left(\gamma_{k+1}^2 + \gamma_{k+1}^{1+\kappa} + \gamma_{k+1}^2 \frac{Q_{m_{k+1}}}{M_{k+1}}\right) (C_1 \vee C_2) \left(1 + \left\|u^k - u^\star\right\|_{l_2}^2\right).$$

Conclusion. In view of the above controls and of **C2**, the assumptions of the Robbins-Siegmund lemma are verified (see [RS71]). An application of this lemma yields that $\lim_k \left\|u^k - u^\star\right\|_{l_2}^2$ exists and $\sum_{k\ge0} \gamma_{k+1}R^k < +\infty$ a.s.. This concludes the proof of (5.4.1). Taking expectations in (5.4.11) and applying the Robbins-Siegmund lemma to the sequence $\mathbb{E}\left[\left\|u^k - u^\star\right\|_{l_2}^2\right]$ yields (5.4.2). Note also that

$$R := \liminf_{k\to+\infty} R^k = 0, \quad \text{a.s.}. \tag{5.4.12}$$

Indeed, on the event $\{R > 0\}$, there exists a finite random index K such that $R^k > R/2$ holds for any $k \ge K$, which implies that $\sum_{k\ge0} \gamma_{k+1}R^k < +\infty$ (as, by assumption, $\sum_{k\ge1} \gamma_k = +\infty$). Therefore $\{R > 0\} \subseteq \{\sum_{k\ge0} \gamma_{k+1}R^k < +\infty\}$, where we saw above that $\{\sum_{k\ge0} \gamma_{k+1}R^k < +\infty\}$ is a zero probability event. Hence so is $\{R > 0\}$, which proves (5.4.3).

We know from (5.4.1) that $\lim_k \left\|\phi^k - \phi'\right\|_\pi$ exists a.s. for any $\phi' \in \mathbf{Is}(\mathcal{T}^\star)$. For later use we need the existence of this limit simultaneously for all $\phi' \in \mathbf{Is}(\mathcal{T}^\star)$ with probability one. Note that $\lim_k \left\|\phi^k - \phi'\right\|_\pi$ is continuous in ϕ' (by triangle inequality). Using that $\mathbf{Is}(\mathcal{T}^\star)$ is separable as a subset of a separable Hilbert space L_2^π, we deduce that

$$\lim_k \left\|\phi^k - \phi'\right\|_\pi \text{ exists for all } \phi' \in \mathbf{Is}(\mathcal{T}^\star), \quad \text{a.s.} \tag{5.4.13}$$

5.4.2 Proof of the Almost Sure Convergence in (5.3.7)

Proof for Case 1 or Case 2. Under the assumption **C1**, $\mathbf{Is}(\mathcal{T}^\star)$ is bounded so that, by (5.4.1), the random variable $B := \sup_{\phi^\star \in \mathcal{T}^\star} \sup_k \left\|\phi^k - \phi^\star\right\|_\pi$ is finite with probability one. Since by (5.4.12) $\liminf_k R^k = 0$, with probability one, there exists a subsequence $\{\zeta(k), k \geq 1\}$ such that $\lim_k R^{\zeta(k)} = 0$.

From (5.4.5) and by **C6** applied with $\phi \leftarrow \phi^{\zeta(k)}$ and $\phi^\star \leftarrow \mathbf{Is}(u^\star)$, there exists a positive random variable C_B (finite a.s. and independent of k by definition of the r.v. B) such that

$$R^{\zeta(k)} \geq C_B \min_{\bar{\phi} \in \mathbf{Is}(\mathcal{T}^\star)} \left\|\phi^{\zeta(k)} - \bar{\phi}\right\|_\pi^2 .$$

Let $\{\bar{\phi}^k, k \geq 0\}$ be an $\mathbf{Is}(\mathcal{T}^\star)$-valued sequence such that, for all k,

$$\min_{\bar{\phi} \in \mathbf{Is}(\mathcal{T}^\star)} \left\|\phi^{\zeta(k)} - \bar{\phi}\right\|_\pi^2 = \left\|\phi^{\zeta(k)} - \bar{\phi}^k\right\|_\pi^2 .$$

Such a sequence exists since $\mathcal{T}^\star$ is compact by **C1**. Using that $\lim_k R^{\zeta(k)} = 0$ we obtain $\lim_k \left\|\phi^{\zeta(k)} - \bar{\phi}^k\right\|_\pi = 0$ a.s.. Since the sequence $\{\bar{\phi}^k, k \geq 0\}$ is in a compact set $\mathbf{Is}(\mathcal{T}^\star)$ (see **C1**), up to extraction of a subsequence it converges to a random limit $\phi^\infty \in \mathbf{Is}(\mathcal{T}^\star)$. Hence

$$\lim_k \left\|\phi^{\zeta(k)} - \phi^\infty\right\|_\pi = 0 \text{ a.s..}$$

In view of (5.4.13), we deduce

$$\lim_k \left\|\phi^k - \phi^\infty\right\|_\pi = \lim_k \left\|\phi^{\zeta(k)} - \phi^\infty\right\|_\pi = 0 \quad \text{a.s..}$$

This concludes the proof of (5.3.7).

Proof for Case 3. Since by (5.4.12) $\liminf_k R^k = 0$ with probability one, there exists a (random) subsequence $\{\zeta(k), k \geq 1\}$ such that $\lim_k R^{\zeta(k)} = 0$ a.s. Since the sequence $\{u^{\zeta(k)}, k \geq 0\}$ is bounded in l_2 a.s.(as $\lim_k \left\|u^k - u^\star\right\|_{l_2}$ exists a.s.) and belongs to the convex set $\mathcal{A}$ by construction, hence it belongs to a compact set (see Corollary 5.B.2). Therefore we can assume (up to extraction of another subsequence) the existence of $u^\infty \in L_2^\pi$ such that $\lim_k \left\|u^{\zeta(k)} - u^\infty\right\|_{l_2} = 0$ a.s. We now prove that u^∞ is a $\mathcal{T}^\star$-valued random variable (possibly depending on the choice of $u^\star \in \mathcal{T}^\star$). Set $\phi^\infty := \mathbf{Is}(u^\infty)$ and define

$$R^\infty := \int_\Theta \left(\phi^\infty - \phi^\star\right)(\theta) \cdot h(\phi^\infty(\theta), \theta)\, \pi(\mathrm{d}\theta).$$

Then for any $j \geq 1$,

$$\begin{aligned} R^j - R^\infty = &\int_\Theta \left(\phi^j - \phi^\infty\right)(\theta) \cdot h\left(\phi^j(\theta), \theta\right) \pi(\mathrm{d}\theta) \\ &+ \int_\Theta \left(\phi^\infty - \phi^\star\right)(\theta) \cdot \left(h\left(\phi^j(\theta), \theta\right) - h\left(\phi^\infty(\theta), \theta\right)\right) \pi(\mathrm{d}\theta). \end{aligned}$$

By either **C5**-b or **C5**-a (depending on whether $\mathcal{A}$ is bounded or not) and since $\sup_k \left\|u^k\right\|_{l_2} < \infty$ a.s., we have $\sup_k \left\|h\left(\phi^{\zeta(k)}(\cdot),\cdot\right)\right\|_\pi^2 < \infty$ a.s.. Since

$$\lim_k \left\|\phi^{\zeta(k)} - \phi^\infty\right\|_\pi = \lim_k \left\|u^{\zeta(k)} - u^\infty\right\|_{l_2} = 0, \quad \text{a.s.},$$

it follows that

$$\lim_k \int_\Theta \left(\phi^{\zeta(k)} - \phi^\infty\right)(\theta) \cdot h\left(\phi^{\zeta(k)}(\theta),\theta\right) \pi(\mathrm{d}\theta) = 0, \quad \text{a.s.}$$

Furthermore, since, by **C3**, $\phi \mapsto h(\phi(\cdot),\cdot)$ is continuous in L_2^π, we have

$$\lim_k \int_\Theta \left(\phi^\infty - \phi^\star\right)(\theta) \cdot \left(h\left(\phi^{\zeta(k)}(\theta),\theta\right) - h\left(\phi^\infty(\theta),\theta\right)\right) \pi(\mathrm{d}\theta) = 0 \text{ a.s.}$$

Hence $0 = \lim_k R^{\zeta(k)} = R^\infty$ a.s. In view of the definition of R^∞ and of **C4**, we deduce that $u^\infty \in \mathring{\mathcal{T}}^\star$ a.s.. In view of (5.4.13), this implies that $\lim_k \left\|\phi^k - \phi^\infty\right\|_\pi = \lim_k \left\|\phi^{\zeta(k)} - \phi^\infty\right\|_\pi = 0$.

5.4.3 Proof of the L^2-Control (5.3.6) and of the L^p-Convergence in (5.3.7)

The L^2-control

$$\sup_{k\geq 0} \mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^2\right] < +\infty$$

follows directly from (5.4.2) and the boundedness of $\mathcal{T}^\star$ (see **C1**). This proves (5.3.6). Let $C > 0$ and $p \in (0,2)$. We write

$$\mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^p\right] = \mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^p \; U^n_{\{\|\phi^k - \phi^\infty\|_\pi > C\}}\right] + \mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^p \; U^n_{\{\|\phi^k - \phi^\infty\|_\pi \leq C\}}\right].$$

The first term on the right hand side converges to 0 as $C \to +\infty$, uniformly in k: indeed, we have

$$\mathbb{E}\left[\left\|\phi^k - \phi^\infty\right\|_\pi^p U^n_{\{\|\phi^k - \phi^\infty\|_\pi > C\}}\right] \leq \frac{\sup_{l\geq 0} \mathbb{E}\left[\left\|\phi^l - \phi^\infty\right\|_\pi^2\right]}{C^{2-p}}.$$

For any fixed $C > 0$, the second term converges to zero by the dominated convergence theorem. This concludes the proof of Theorem 5.3.5. □

5.5 Numerical Investigations

This section is devoted to the numerical analysis of the convergence of the USA algorithm .

There cannot be any comparison, performance-wise, between the USA algorithm and the naive algorithms of Sections 5.2.1 and 5.2.2. The "θ by θ" algorithm of Section 5.2.1 is of course no option unless a finite set Θ, with reasonable cardinality, is considered. As for the "coefficient by coefficient" algorithm of Section 5.2.2, it requires one (standard, admittedly) SA algorithm for each estimate $\widehat{\phi(\theta_{k,i})}$ of $\phi(\theta_{k,i})$ in (5.2.6): since k indexes Monte Carlo

draws, it means a nested Monte Carlo approach, which can only be achieved, on realistic applications, by resorting to concurrent computing resources. Instead, the USA algorithm is a single SA procedure (in increasing space dimension) for the joint estimation of the coefficients $u_i^\star$.

Hence, the section is purely focused on the USA algorithm. We discuss its parameterization and we test empirically the sensitivity of its performance with respect to the latter.

Notably, the possibility of letting the number m_k of estimated coefficients $u_i^\star$ tend to infinity appears not only as a necessary ingredient for proving the theoretical convergence (see Theorem 5.3.5), but also as an important feature for its numerical performance, regarding, in particular, the estimation of the lower order coefficients $u_i^\star$ and the mitigation of the burn-in phase. We illustrate this assertion numerically, by testing both the genuine USA algorithm with increasing m_k and the fixed dimension version with $m_k = m$ (for different values of m), respectively referred to as the "increasing m_k" and the "fixed m" algorithms henceforth. The speed of the dimension growth turns out to be a determining factor of the practical convergence rate of the algorithm. A correct tuning of this speed allows achieving the right balance between the truncation error, i.e. the error due to the non-estimation of the coefficients beyond the m_k^{th} one, and the estimation error on the "active" coefficients up to m_k. Balancing these two contributions of the error seems to be the way to reach an optimal performance of the algorithm.

5.5.1 Design Parameterization of the USA Algorithm

When running the USA algorithm, the user has to choose some design parameters: given a problem of the form (5.3.1) and the corresponding sequence $\{q_m, m \geq 0\}$ via (5.3.2), the user has to choose the orthogonal basis $\{B_i(\theta), i \geq 0\}$, which fixes in turn the sequence $\{Q_m, m \geq 0\}$. It remains to choose $\{\gamma_k, k \geq 1\}$, $\{m_k, k \geq 0\}$ and $\{M_k, k \geq 1\}$. In this section, we consider sequences of the form

$$\gamma_k = k^{-a}, \quad m_k = \lfloor k^b \rfloor + 1, \quad M_k = \lfloor k^p \rfloor + 1, \tag{5.5.1}$$

for $a, p \geq 0$ and $b > 0$, and we discuss how to choose these constants assuming that

$$q_m = O\left(m^{-\delta}\right), \qquad Q_m = O\left(m^{\Delta}\right), \tag{5.5.2}$$

for some $\delta > 0$ and $\Delta \geq 0$.

An easy calculation shows that **C2** is satisfied ($\kappa > 0$ ensuring **C2** exists) if

$$0 < a \leq 1, \qquad 2 - \delta b < 2a, \qquad b\Delta + 1 < 2a + p. \tag{5.5.3}$$

Given $\delta > 0$ and $\Delta \geq 0$, there always exist a, b, p satisfying these conditions.

Figure 5.1 displays the lines $x \mapsto 1$, $x \mapsto 2(1-x)/\delta$ and $x \mapsto (2x-1)/\Delta$ for different values of the pair (δ, Δ) with $\Delta > 0$. The colored area corresponds to the points (a, b) satisfying the conditions (5.5.3) in the case $p = 0$, i.e. in the case where the numbers of Monte Carlo draws is constant over iterations. Note that this set becomes all the more restrictive as $\delta \to 0$ and $\Delta \to \infty$. Choosing $p > 0$ gives more flexibility, but it also leads

to higher computational cost (since the number of Monte Carlo simulations increases along iterations, see the discussion in Section 5.5.5).

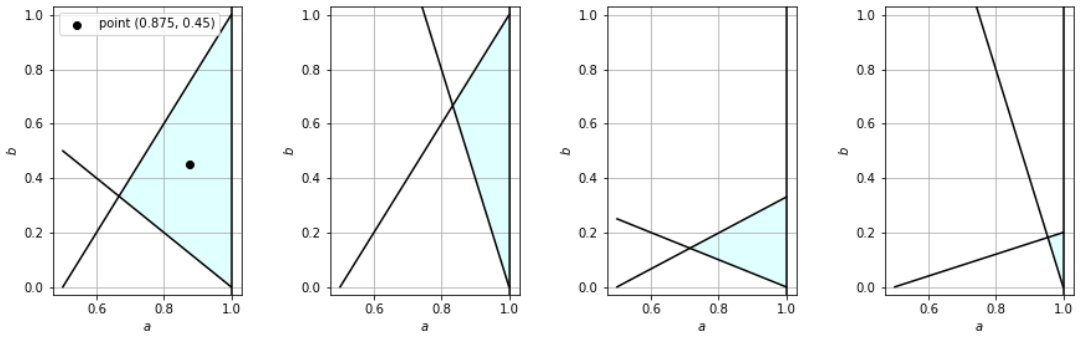

Figure 5.1: For different values of (δ, Δ), in the case $p = 0$, the colored area is the admissible set of points (a, b) satisfying (5.5.3). From left to right: $(\delta, \Delta) = (2, 1), (0.5, 1), (4, 3)$, and $(0.5, 5)$.

5.5.2 Benchmark Problem

We consider the problem (5.3.1) in the case where

$$\Theta = [-\pi, \pi], \qquad \pi(\mathrm{d}\theta) = \frac{1}{2\pi}\mathbf{1}_{[-\pi,\pi]}\,\mathrm{d}\theta. \tag{5.5.4}$$

We perform tests for two different models of function H:

1. $H_1(z, v, \theta) := (z - \phi^\star(\theta))\left(1 + \dfrac{\cos(v)}{2}\cos(z - \phi^\star(\theta))\right) + v,$
2. $H_2(z, v, \theta) := (z - \phi^\star(\theta))\left(1 + \dfrac{\cos(v)}{2}\sin(z - \phi^\star(\theta))\right),$

for a common function $\phi^\star : \Theta \to \mathbb{R}$ given by

$$\phi^\star(\theta) := \left|\frac{4}{5} + \frac{1}{4}\exp(\sin(\theta)) - \cosh(\sin(\theta)^2)\right|(1 + \sin(2\theta)), \tag{5.5.5}$$

and where, for any $\theta \in \Theta$, the conditional distribution $\mu(\theta, \mathrm{d}v)$ is a centered Gaussian distribution with variance θ^2.

The functions h_1, h_2 corresponding to H_1, H_2 (cf. **C3**) are equal to

$$h_1(z, \theta) = (z - \phi^\star(\theta))\left(1 + \frac{\mathbb{E}[\cos(\theta\, Y)]}{2}\cos(z - \phi^\star(\theta))\right), \tag{5.5.6}$$

$$h_2(z, \theta) = (z - \phi^\star(\theta))\left(1 + \frac{\mathbb{E}[\cos(\theta\, Y)]}{2}\sin(z - \phi^\star(\theta))\right), \tag{5.5.7}$$

where $Y \sim \mathcal{N}(0, 1)$. In both cases we have $q = 1$ and $\mathbf{Is}(\mathcal{T}^\star) = \{\phi^\star\}$.

It is easily checked that for any $z \in \mathbb{R}$, $\theta \in \Theta$ and $i = 1, 2$ we have

$$\int_{\mathcal{V}} |H_i(z, v, \theta)|^2 \mu(\theta, \mathrm{d}v) \leq 8|z - \phi^\star(\theta)|^2 + 2\theta^2, \qquad (z - \phi^\star(\theta)) \cdot h_i(z, \theta) \geq \frac{1}{2}(z - \phi^\star(\theta))^2.$$

Hence, the assumptions **C3**, **C4**, **C5**, and **C6** are satisfied for both models.

The two models for H above correspond to two possible behaviours of the martingale increment sequence $\{\eta^k, k \geq 1\}$ (cf. Section 5.4.1): $\mathbb{E}\left[\left\|\eta^k\right\|_{l_2}\right]$ bounded away from 0 (case of Model 1) or $\mathbb{E}\left[\left\|\eta^k\right\|_{l_2}\right] \to 0$ (case of Model 2). While the first case is more general, the second one may also appear in practice and leads to quite different behaviour of the USA algorithm, requiring a different tuning of the parameters.

In real-life applications, the target function $\phi^\star$ is bound to be less challenging than the present one, e.g. monotone and/or convex/concave with respect to θ or some of its components. Moreover, the user may be interested with a few coefficients $u_i^\star$ only, whereas we show numerical results up to $m_K = 250$ below.

The choice of $\mathcal{N}(0, \theta^2)$ for the kernel $\mu(\theta, \mathrm{d}v)$ is purely illustrative. This distribution could be replaced by any other one (simulatable i.i.d.) without expectable impact regarding the qualitative conclusions drawn from the numerical experiments below.

Finally, for the orthonormal basis $\{B_i, i \geq 0\}$, we choose the normalized trigonometric basis on $\Theta = [-\pi, \pi]$ (cf. Example 5.2.1(v)). Therefore, we have $\sup_{i \geq 0} \sup_\Theta |B_i(\theta)| < +\infty$, so that

$$Q_m = O(m),$$

i.e. $\Delta = 1$ in (5.5.1). Since $\phi^\star$ extended by periodicity outside $[-\pi, \pi]$ is piecewise continuously differentiable, its truncation error satisfies (see Lemma 5.A.1)

$$\sum_{i=m+1}^{+\infty} |u_i^\star|^2 = O\left(m^{-2}\right),$$

i.e. we have $\delta = 2$ in (5.5.2). Numerically, one can check that the practical rate of convergence lies somewhere between 2 and 3, i.e. the theoretical value $\delta = 2$ above is reasonably sharp (meaning that our example $\phi^\star$ is close to a "real" $\delta = 2$ example and not much "easier", which also motivated our choice of this particular function $\phi^\star$).

5.5.3 Performance Criteria

In the numerical experiments that follow, we compare the performances of the algorithms with increasing m_k and fixed m, for different choices of (a, b, p). The comparison relies on the root-mean-square errors, where the exact expectation is approximated by the mean value over 50 independent runs of the algorithms. After K iterations, the square of the total error $\mathcal{E}^2$ is decomposed into the mean squared SA error $\mathcal{E}_{sa}^2$, which is the error restricted to the $(m_K + 1)$ estimated coefficients, and the squared truncation error $\mathcal{E}_{tr}^2$, i.e. $\mathcal{E}^2 = \mathcal{E}_{sa}^2 + \mathcal{E}_{tr}^2$

where

$$\mathcal{E}^2 = \mathbb{E}\left[\left\|u^K - u^\star\right\|_{l_2}^2\right], \quad \mathcal{E}_{sa}^2 = \mathbb{E}\left[\sum_{i=0}^{m_K}(u_i^K - u_i^\star)^2\right], \quad \text{and } \mathcal{E}_{tr}^2 = \sum_{i=m_K+1}^{+\infty}(u_i^\star)^2$$

(recalling $u_i^K = 0$ for $i > m_K$). The benchmark values for the coefficients $u_i^\star$ are pre-calculated by high-precision numerical integration. With the exception of Figures 5.3 and 5.5[left], all our graphs are error plots in log-log scale.

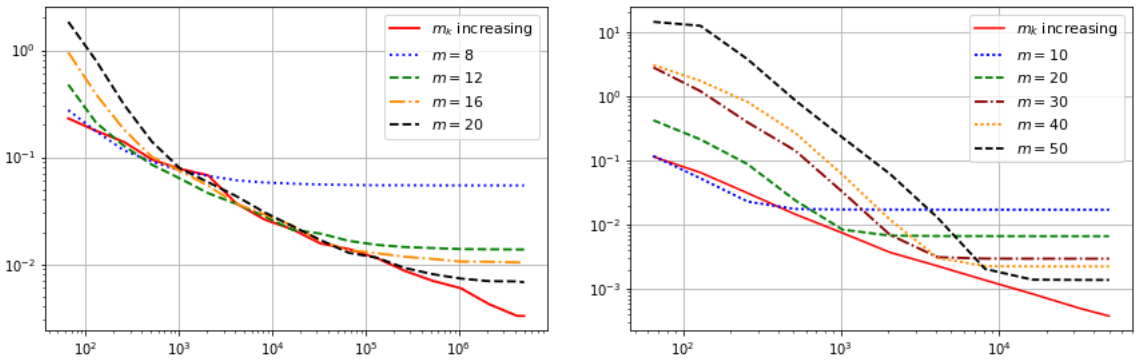

Figure 5.2: The total error $\mathcal{E}$ as a function of the number of iterations, for different choices of the sequence $\{m_k, k \geq 0\}$: m_k increasing (solid line) and: [left] for Model 1 with $m_k = m = 8, 12, 16, 20$ (other lines); [right] for Model 2 with $m_k = m = 10, 20, 30, 40, 50$ (other lines).

5.5.4 Impact of the Increasing Dimension

In this section, we discuss the role of the sequence $\{m_k, k \geq 0\}$. Since $(\delta, \Delta) = (2, 1)$, the set of admissible pairs (a, b) for our benchmark problem is given by the leftmost graph of Figure 5.1.

We take $a = 0.875$, which is in the middle of the corresponding admissible interval. For Model 1 (i.e. $H = H_1$), a heuristic may be applied to choose intelligently the value of b. In finite dimensional SA schemes, the squared L^2-error after k iterations is typically of the order of $\gamma_k = k^a$ (see [Duf97, Chapter 2]). For the USA algorithm we may expect (in the case $p = 0$) a growth of the variance at least proportional to the dimension $m_k \approx k^b$. This suggests a heuristic guess for the SA-error order given by $1/k^{a-b}$. Now, by (5.5.2) (with in our case $\delta = 2$), the truncation error is of order $k^{-b\delta}$. Hence, to optimize the convergence rate, we take b such that $b\delta = a - b$. This approximately corresponds to $b = 3$, which is the value that we use in our tests for Model 1. In the case of Model 2 (i.e. $H = H_2$), this heuristics does not apply, because the variance of the martingale increments goes to 0. In this case we simply take $b = 0.45$, so that $(a, b) = (0.875, 0.45)$ lies in the middle of the admissible set (see Figure 5.1[left]). Also note that the range $[0.25, 0.5]$ for b is reasonable in view of the total number K of iterations that we commonly use in the algorithm and of the

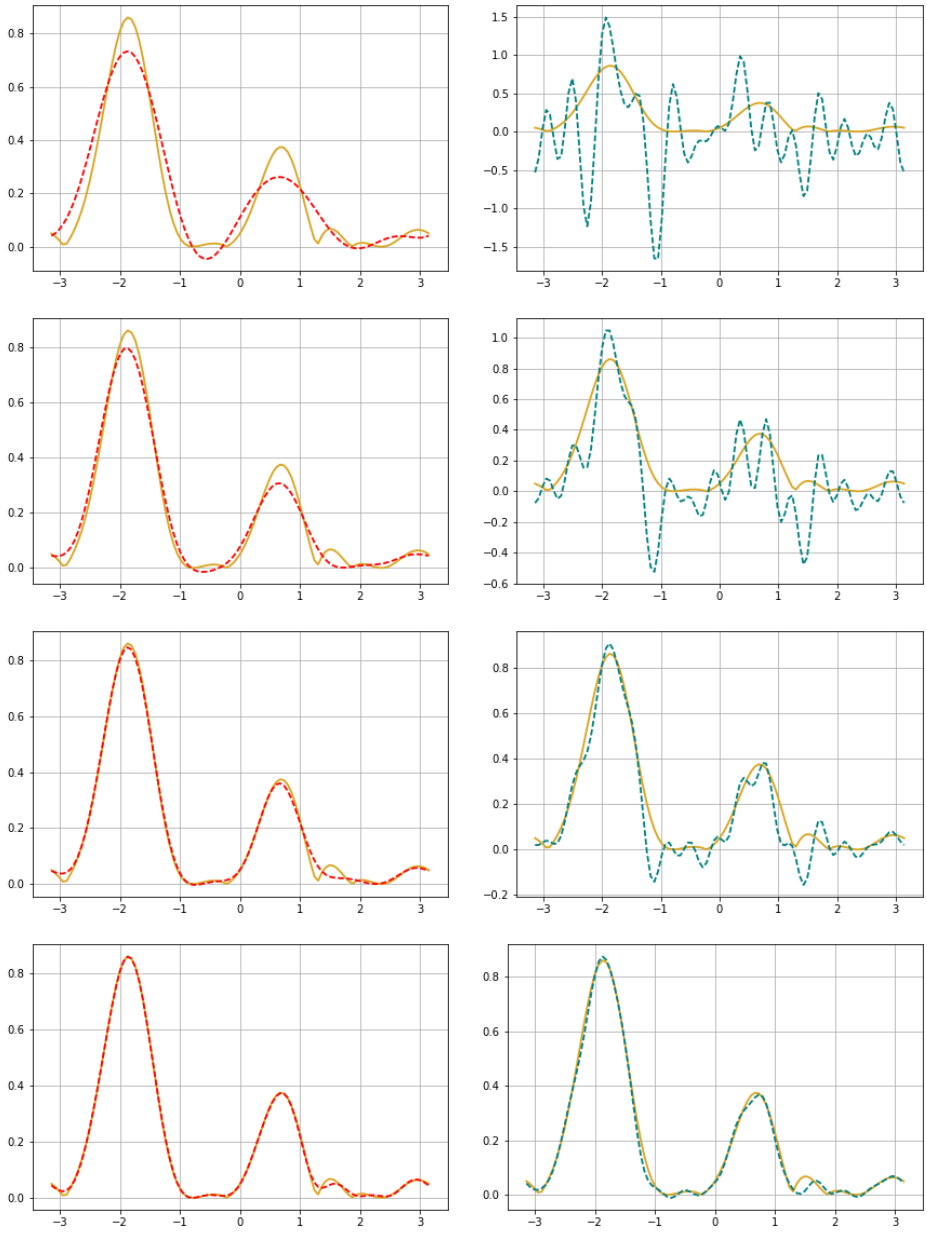

Figure 5.3: The functions $\phi^\star$ and ϕ^K are displayed in respective solid line and dashed lines, as a function of $\theta \in [-\pi, \pi]$. On the left, $\{m_k, k \geq 0\}$ is increasing and on the right, it is constant and equal to $m = 30$. From top to bottom, $K \in \{128, 256, 512, 1024\}$.

number m_K of the coefficients of interest.

Figure 5.2 displays the total error $\mathcal{E}$ for different strategies on the sequence $\{m_k, k \geq 0\}$: the solid line is the case $m_k = \lfloor k^b \rfloor + 1$ (with $b = 0.3$ and 0.45 for Models 1 and 2 respectively), while the other lines correspond to the cases $m_k = m = 8, 12, 16, 20$ for Model

1 and $m_k = m = 10, 20, 30, 40, 50$ for Model 2 (larger values of m are used here since the convergence is expected to be faster for Model 2). The performance of the algorithm with increasing m_k is similar or better throughout the whole iteration path. This holds true in the burn-in phase, which is typically related to disproportion of the first values of γ_k and the magnitude of the solution (estimated coefficients). In fact, with increasing m_k, the dimension gradually grows with k, with larger values of γ_k naturally associated with the estimation of the first, larger coefficients, whereas, when $m_k = m$ is constant, the higher order "small" coefficients are involved from the very beginning along with the larger values of γ_k, leading to a longer burn-in phase. It is also true on the convergence part, where the fixed dimension algorithms $m_k = m$ only converge up to a certain accuracy depending on the value of m. The superior performance of the increasing dimension version is more pronounced for Model 2, while for Model 1 fixed m versions have similar performance within certain ranges of values of K. However, in practice we do not know in advance the length of the burn-in phase or the magnitude of the truncation error for various m. Hence, the genuine USA algorithm with increasing m_k, which provides optimal performance without the need for additional knowledge, is always preferable.

Let us now analyse the weak burn-in phase performance of the fixed m version for Model 2 (cf. Figure 5.2[right]). Figure 5.3 displays the result of a single run of the USA algorithm in this case. In dashed line, the function $\theta \mapsto \phi^K(\theta)$ is displayed for $\theta \in [-\pi, \pi]$. For comparison, the function $\theta \mapsto \phi^\star(\theta)$ is displayed in solid line. To illustrate the advantage of the USA algorithm on the burn-in phase due to gradual dimension growth, we show the estimated function ϕ^K for different values of K (from top to bottom, $K \in \{128, 256, 512, 1024\}$) and for m_k increasing (left panels) versus $m_k = m = 30$ for any k (right panels). The increasing dimension m_k leads to a smoother convergence, with intermediate iterations looking closer to a projection of $\phi^\star$ on the subspace spanned by a smaller number of basis functions. We conclude that in the case where the variance of the martingale increment is small or goes to 0, the progressive dimension growth plays a key role in the USA algorithm performance.

In Figure 5.4, we show that increasing m_k is also key for an accurate determination of the lower order coefficients (e.g. in the case where only the first few coefficients of the expansion of $\phi^\star$ are of interest to the user). In fact, as already mentioned in Section 5.2.3, the algorithm with fixed m does typically not converge to the first $(m+1)$ coefficients of the decomposition of $\phi^\star$. In Figure 5.4[left], the L^2-error on the first 4 coefficients is displayed as a function of the number of iterations K, for two strategies on m_k (case of Model 2): the solid line is the case $m_k = O(k^b)$ with $b = 0.45$ and the dotted line is the case $m = 3$. In Figure 5.4[right], the total error $\mathcal{E}$ and the truncation error $\mathcal{E}_{tr}$ are displayed, resp. in dash-dot line and dashed line in the case m_k is the constant sequence equal to $m = 3$. These figures show that, when $m_k \to +\infty$, USA converges (which is the claim of Theorem 5.3.5), whereas, when $m_k = m$ for any k, it does not: the total error does not reach the truncation error since there is a non vanishing bias on the estimation of the first $(m + 1)$ coefficients (the SA-error $\mathcal{E}_{sa}$ does not vanish when $K \to +\infty$).

For Model 1, similar effects (not reported here) are visible, even if a bit less obvious due to the slower convergence of both versions (fixed and increasing dimension) of the algorithm in this case."

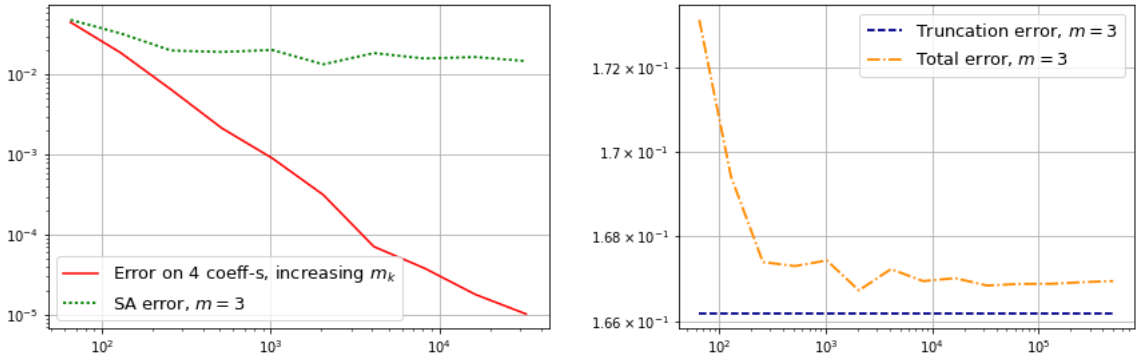

Figure 5.4: Model 2: [left] in the case $m_k \to \infty$ (solid line) and $m_k = m = 3$ (dotted line), the error $\left(\mathbb{E}\sum_{i=0}^{3}(u_i^K - u_i^\star)^2\right)^{1/2}$ as a function of the number of iterations K; [right] in the case $m_k = m = 3$, the truncation error $\mathcal{E}_{tr}$ (dashed line) and the total error $\mathcal{E}$ (dash-dot line) displayed as a function of K.

5.5.5 Impact of the Design Parameters for the Increasing m_k USA Algorithm

In this section, we discuss the impact of the choice of a, b, and p on the performance on the USA algorithm.

Role of b

In this paragraph, we set as before $a = 0.875$, $p = 0$, and we test different values of b. The range of admissible values of b is $(0.125, 0.75)$. We take $b \in \{0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$ for Model 1 and $b \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$ for Model 2 (as we see next, these values of b suffice to explain the behavior of the algorithm for each of the models).

Figure 5.5 displays the evolution of the total error $\mathcal{E}$ as a function of the number of iterations K for different values of b for both models of H.

For Model 1, the variance increases with the dimension, which makes the SA error larger as b increases. At the same time the truncation error has the decrease rate $b\delta$, so that for too small values of b the truncation error dominates the SA error. Hence there is a trade-off between the two errors, with optimal values of b somewhere in the middle. This phenomenon is observed on Figure 5.5[left, center]. For $b = 0.2, 0.25, 0.3$, the total error is dominated by the truncation error, while from $b = 0.35$ the error is dominated by the SA error so that, as b increases further, the convergence becomes slower due to additional variance which augments the SA error.

For Model 2, since the variance of the martingale increments goes to 0, the effect of additional variance due to a larger dimension is not visible. We observe that larger values of b lead to better convergence up to $b = 0.70$. However, as we may see, the gain in the speed of convergence from taking larger b decreases as we approach the border of the admissible interval. In addition, this analysis in terms of the number of iterations K does not take into

account the higher computational cost due to a dimension growing faster and each iteration becoming longer when b is larger. For example, for $b = 0.70$, we made only $K = 2500$ iterations, because the computational effort becomes too large beyond this. To conclude we suggest that in this case optimal values of b (for given a) in terms of both convergence and computational cost lie near the middle of the admissible interval.

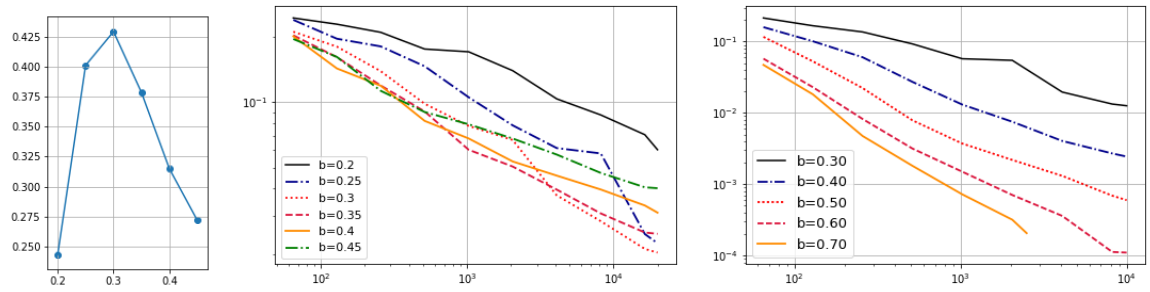

Figure 5.5: [left] Empirical L^2 convergence rate for $b \in \{0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$, Model 1. Total error $\mathcal{E}$ as a function of the number of iterations K, for different values of b; [middle] Model 1, $b \in \{0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$; [right] Model 2, $b \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$.

Role of a

In this paragraph, still for $p = 0$, taking as in Section 5.5.4 $b = 0.3$ for Model 1 and $b = 0.45$ for Model 2, we compare different values of a.

Figure 5.6 displays the total error $\mathcal{E}$ as a function of the number of iterations K for different values of a for both models 1 and 2.

For Model 1, we see that the convergence rate is better for larger values of a. This is in line with classical results for finite dimensional stochastic approximation, whereby the L^2 error is of order $\gamma_k^{1/2} \sim k^{-a/2}$ (see [Duf97, Chapter 2]).

For Model 2, varying a does not produce much effect (except for a slight decline as a approaches 1), because the step-size controls the variance of the corresponding martingale noise, but in the case of Model 2 the variance of the martingale increments goes to 0 anyway.

Role of p

In this section we consider the case $p > 0$, i.e. the number of Monte Carlo samples at each iteration increases along the USA iterations. One may check that all the triples of the parameters (a, b, p) used below lie in the admissible set (cf. (5.5.3)).

In the analysis that follows, we want to keep track of the dependence of the error with respect to a computational cost proxied by the total number of Monte Carlo draws of the pair (θ, v), i.e., after K iterations, $\sum_{k=0}^{K-1} M_k \approx O(K^{p+1})$. As we want to have the same dimension growth speed with respect to the computational cost for different tests, we take $b = \bar{b}(p+1)$, with $\bar{b} = 0.3$ for Model 1 and $\bar{b} = 0.45$ for Model 2.

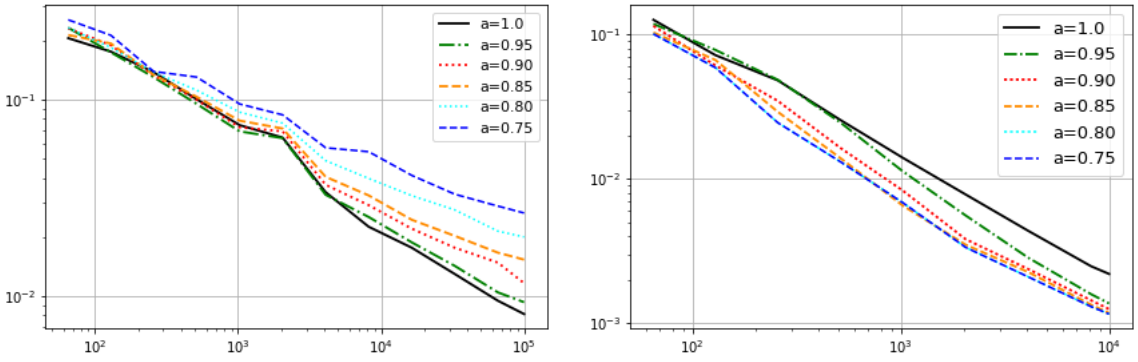

Figure 5.6: The total error $\mathcal{E}$ as a function of the number of iterations, for different values of a in $\{0.75, 0.80, 0.85, 0.9, 0.95, 1.0\}$: [left] Model 1 and [right] Model 2.

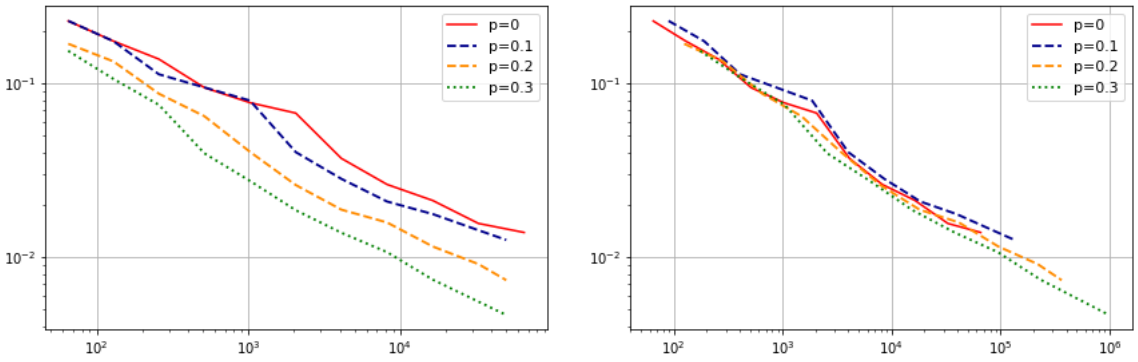

Figure 5.7: Model 1, total error $\mathcal{E}$ of the USA algorithm for different values of $p \in \{0, 0.1, 0.2, 0.3\}$ as a function of the number of iterations [left] and of the total number of Monte Carlo draws [right]. Here $a = 0.875$ and $b = 0.3(p + 1)$.

We first set $a = 0.875$ (as in Sections 5.5.4 and 5.5.5). Figures 5.7 and 5.8 display the total error $\mathcal{E}$ as a function of the number of iterations (left) and as a function of the total number of Monte Carlo draws (right) for triples of the form $(a, \bar{b}(p + 1), p)$ with various p. The results show that, even though larger p yield a better convergence in terms of the number of iterations K, there is no much difference when the computational cost is taken into account (i.e. in terms of the number of Monte Carlo draws).

Taking a larger p allows taking a smaller a (see (5.5.3)), so that γ_k decreases at a lower rate. To see if it is possible to take advantage of this balance in the case of Model 2 (since typically with Model 1 the convergence is slower for smaller a, see Section 5.5.5), we test triples of the form $\big(\bar{a}/(p + 1), \bar{b}(p + 1), p\big)$, with $(\bar{a}, \bar{b}) = (0.875, 0.45)$ and different values of p. Figure 5.9 displays the results. The conclusions are similar as for the previous test. Hence, on our problem, it seems difficult to take advantage of the degree of freedom provided

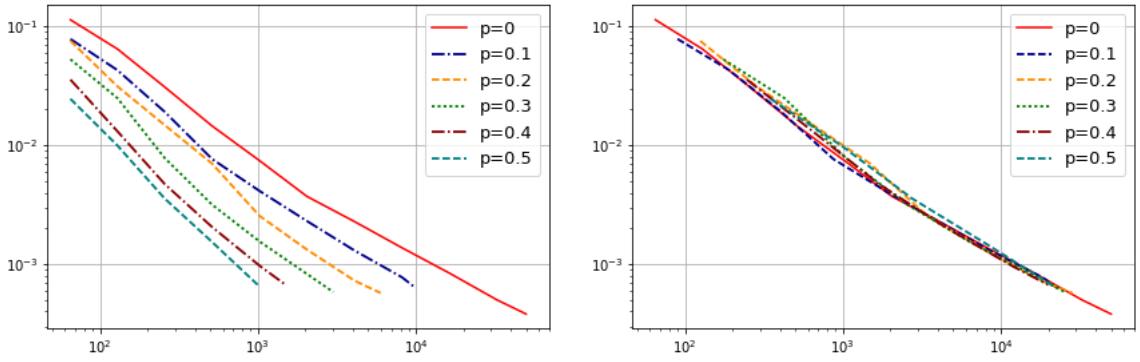

Figure 5.8: Model 2, total error $\mathcal{E}$ of the USA algorithm for different values of $p \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ as a function of the number of iterations [left] and of the total number of Monte Carlo draws [right]. Here $a = 0.875$ and $b = 0.45(p+1)$.

by M_k by going beyond the obvious choice $M_k = M$ for any k. Such a degree of freedom could still be useful to ensure the convergence of $\sum_{k\geq 0} \gamma_k^2 Q_{m_k} M_k^{-1}$ (as required by **C2**) in situations where $\Delta > 1$ (i.e. Q_m grows faster than in our example), and therefore ensure the convergence of the algorithm in such cases, even if this comes at a higher computational cost.

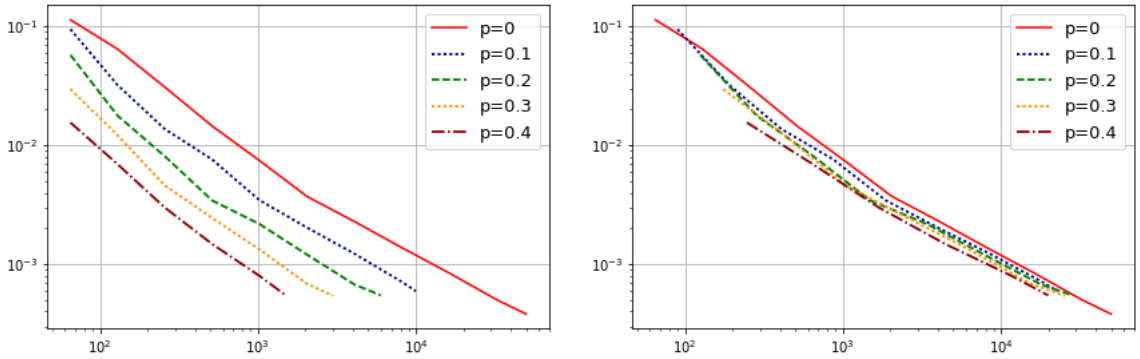

Figure 5.9: Model 2, total error $\mathcal{E}$ of the USA algorithm for different values of $p \in \{0, 0.1, 0.2, 0.3, 0.4\}$ as a function of the number of iterations [left] and of the total number of Monte Carlo draws [right]. Here $a = 0.875/(p+1)$ and $b = 0.45(p+1)$.

To summarize, we observe from our experiments that the growing dimension feature of the USA algorithm is essential for the asymptotic convergence, as well as the lower order coefficients estimation, and also yields a milder burn-in phase. As expected, the convergence is generally faster for Model 2 due to reduced variance effect. In the more general setting (Model 1), the parameter b plays a crucial role in the performance of the algorithm, a good rule of thumb (in the case $p = 0$) being to take b satisfying $b\delta = a - b$. However, in the special

case where the variance of the martingale increments goes to 0, this rule does not work and one should take larger values of b. Empirical convergence rates are naturally bounded by $a/2$ for Model 1 (which follows from corresponding results for finite dimensional SA), while for Model 2 the convergence is much faster. Finally, taking $p > 0$ may potentially be useful for verifying the assumptions of Theorem 5.3.5 in some cases, but it makes no real difference in terms of convergence speed when the latter is assessed with respect to the total number of simulations.

5.A Truncation error for trigonometric basis

Lemma 5.A.1. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be 2π-periodic and piecewise continuously differentiable. Let $\{u_i, i \geq 0\}$ be the coefficients of its decomposition with respect to the normalized trigonometric basis (cf. Example 5.2.1(v)). Then for some $C > 0$*

$$\sum_{i=m+1}^{+\infty} |u_i|^2 \leq C m^{-2}.$$

Proof. Consider the Fourier decomposition of the function ϕ' on $[-\pi, \pi]$:

$$\phi'(x) = v_0 + \sum_{m \geq 1} (v_{2m-1} \sin(mx) + v_{2m} \cos(mx)).$$

As ϕ is 2π-periodic (i.e. $\phi(-\pi) = \phi(\pi)$), integrating by parts yields, for any $m \geq 1$,

$$\begin{aligned} u_{2m-1} &:= \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(x) \sin(mx) \mathrm{d}x = -\frac{1}{2\pi m} \int_{-\pi}^{\pi} \phi(x) \mathrm{d} \cos(mx) \\ &= -\frac{1}{2\pi m} \phi(x) \cos(mx) \Big|_{-\pi}^{\pi} + \frac{1}{2\pi m} \int_{-\pi}^{\pi} \phi'(x) \cos(mx) \mathrm{d}x = \frac{v_{2m}}{m}, \end{aligned}$$

and

$$\begin{aligned} u_{2m} &:= \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(x) \cos(mx) \mathrm{d}x = \frac{1}{2\pi m} \int_{-\pi}^{\pi} \phi(x) \mathrm{d} \sin(mx) \\ &= \frac{1}{2\pi m} \phi(x) \sin(mx) \Big|_{-\pi}^{\pi} - \frac{1}{2\pi m} \int_{-\pi}^{\pi} \phi'(x) \sin(mx) \mathrm{d}x = -\frac{v_{2m-1}}{m}. \end{aligned}$$

Hence,

$$\sum_{i=2m-1}^{+\infty} |u_i|^2 \leq \frac{1}{m^2} \sum_{i=2m-1}^{+\infty} |v_i|^2 \leq \frac{\|\phi'\|_{\pi}^2}{m^2},$$

which implies the result. □

5.B Compact sets in l_2

Lemma 5.B.1. *For a positive sequence $\{a_n, n \geq 0\}$ such that $\sum_{i \geq 0} a_i^2 < \infty$ and an increasing sequence of non-negative integers $\{d_n, n \geq 0\}$ such that $d_0 = 0$, the closed convex set $\mathcal{A}$:*

$$\mathcal{A} := \left\{ u \in l_2 : \sum_{d_n \le i < d_{n+1}} |u_i|^2 \le a_n^2 \quad \forall n \ge 0 \right\} \tag{5.B.1}$$

is compact.

Proof. By [KB09, Theorem 3] a subset $\mathcal{A}$ of l_2 is relatively compact if and only if

$$\sup_{u \in \mathcal{A}} \sum_{i \ge n} |u_i|^2 \text{ is finite for every } n \text{ and converges to 0 as } n \to +\infty.$$

For $\mathcal{A}$ given by (5.B.1) it is clear that for l such that $d_l \le n$ we have

$$\sup_{u \in \mathcal{A}} \sum_{i \ge n} |u_i|^2 \le \sup_{u \in \mathcal{A}} \sum_{i \ge d_l} |u_i|^2 \le \sum_{j \ge l} a_j^2 \to 0$$

as $n, l \to +\infty$. Since $\mathcal{A}$ is also closed we deduce that it is compact. □

Corollary 5.B.2. *Let $\mathcal{A}$ be defined by (5.B.1) (with d_0 not necessarily 0). For any constant $B > 0$ the set $\{u \in \mathcal{A} : \|u\|_{l_2} < B\}$ is convex compact.*

Proof. The result follows directly from Lemma 5.B.1 using that $\sum_{i < d_0} |u_i|^2 \le B^2$ for any $u \in \mathcal{A}$. □

Chapter 6

Uncertainty quantification for stochastic approximation limits: L^2-convergence rate

Contents

6.1 Introduction

In this chapter we continue the study of the Uncertainty Quantification (UQ) problem for Stochastic Approximation (SA) limits. In Chapter 5 we designed a new method, called the USA (Uncertainty for SA) algorithm, to compute the chaos expansion coefficients of the SA limit as a function of the uncertain parameter and proved its a.s. and L^p convergence. Our goal is to analyze the L^2-convergence rate of this algorithm.

Let us briefly recall the setting of Chapter 5. We consider SA that is typically used to find zeros of an intractable function $h : \mathbb{R}^q \to \mathbb{R}^q$ that is only available in the form of an expectation as $h(z) := \mathbb{E}[H(z, V)]$, i.e. for solving equations of the form

$$\mathbb{E}[H(z, V)] = 0, \tag{6.1.1}$$

where V is some random variable. In Chapter 5 the problem (6.1.1) is considered under the presence of uncertainty. Assume for simplicity that (6.1.1) has a unique solution $z^\star$.

Uncertainty in the solution $z^\star$ appears in the situations where the distribution of the random noise V or the function H is not known exactly. This is studied in the form of: *(i)* a parametric dependence $V \sim \mu(\theta, \mathrm{d}v)$ where the distribution of V depends on an unknown parameter θ for which only some probability distribution $\pi(\mathrm{d}\theta)$ is available; *(ii)* through a dependency of H in the uncertain parameter θ. This leads to the equation of the form

$$h(z, \theta) := \int_{\mathcal{V}} H(z, v, \theta)\mu(\theta, \mathrm{d}v) = 0, \qquad \pi\text{-a.s.} \tag{6.1.2}$$

Denote $z^\star := \phi^\star(\theta)$ the solution of (6.1.2) for fixed $\theta \in \Theta$. The USA algorithm developed in Chapter 5 aims at calculating the coefficients of the function $\theta \mapsto \phi^\star(\theta)$ on an orthogonal basis $\{B_i, i \geq 0\}$ of the Hilbert space of square integrable functions with respect to the distribution $\pi(\mathrm{d}\theta)$. It provides recursively a sequence of vectors $(u^k)_{k\geq 0}$ so that each vector u^k is of dimension $m_k + 1$ (for some increasing integer sequence $m_k \to \infty$) and constitutes a current approximation of the first m_k+1 coefficients of the decomposition of $\phi^\star$ on $\{B_i, i \geq 0\}$. An iteration of the USA algorithm (up to some technical formalities) is given as follows: for $i = 0, \ldots, m_{k+1}$

$$u_i^{k+1} = u_i^k - \gamma_{k+1} M_{k+1}^{-1} \sum_{s=1}^{M_{k+1}} H\left(\sum_{j=0}^{m_k} u_j^k B_j(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) B_i(\theta_{k+1}^s), \tag{6.1.3}$$

while for $i > m_{k+1}$ we set $u_i^{k+1} = 0$. Here $\{\gamma_k, k \geq 1\}$ is a positive step-size sequence converging to 0, $\{M_k, k \geq 1\}$ is some (possibly constant) sequence and $\{(\theta_k^s, V_k^s), k \geq 1, s = 1, \ldots, M_k\}$ are i.i.d. simulations from the distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$. Each iteration of the USA algorithm provides an approximation of the function $\phi^\star$ given by $\phi^k := \sum_{i=0}^{m_k} u_i^k B_i$. Further we may use ϕ^k to approximate the distribution of the uncertain SA limit by $\{\phi^k(\theta), \theta \sim \pi\}$ using the i.i.d. simulations of $\theta \sim \pi$.

We refer to Chapter 5 for the motivation of the UQ problem for SA limits and the construction of the iterative procedure (6.1.3) and its comparison with other techniques to tackle the UQ problem for SA limits.

In Chapter 5 we prove the a.s. and L^p ($p < 2$) convergence of the error $\left\|\phi^k - \phi^\star\right\|_\pi$ to 0 as $k \to \infty$ (where $\|\cdot\|_\pi$ is the Hilbert space norm). Remark that, though the USA approximates an infinite-dimensional object, it is fully constructive since the iterates are finite-dimensional. Chapter 5 contains various numerical investigations and, in particular, uses an empirical L^2-error (i.e. an empirical version of $\mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right]$) as a performance criteria. Most of the numerical examples in Chapter 5 confirm the L^2-convergence of the USA algorithm with some positive polynomial rate. More importantly, the performance of the USA depends on a careful tuning of the dimension growth speed (i.e. parameters determining the sequence $m_k \to \infty$). Though some heuristics for its choice are given in Section 5.5.4, a full theoretical study is needed to provide further insights on this issue.

Our contribution. In this chapter we analyze the L^2-convergence rate of the sequence $\{\phi^k, k \geq 0\}$ provided by the USA algorithm. Our main result explicitly provides $\alpha > 0$ such

that for some constant $C_\alpha > 0$ we have for all $k \geq 0$

$$\mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right] \leq C_\alpha \gamma_k^\alpha. \tag{6.1.4}$$

Control of the form γ_k^α is motivated by similar results in the finite dimensional case, where typically the SA squared error is proved to be of order $O(\gamma_k)$, i.e. $\alpha = 1$ (see e.g. [Duf97, Chapter 2]).

We consider the contribution of this work valuable for the following reasons.

Firstly, while in the finite-dimensional results on the SA convergence rate, the convergence speed typically depends only on the step-size sequence γ_k, in our setting, the exponent α in (6.1.4) will depend non-trivially on the model, the regularity of $\phi^\star$, the choice of the basis functions and the design parameters of the USA algorithm. The knowledge of this dependence plays an important role in the correct tuning of the algorithm to guarantee the L^2-convergence with the best possible rate, given the model specification. We illustrate how the obtained results justify the optimality of the heuristic choice of the dimension growth speed used in Chapter 5 (see Section 6.3.2).

Secondly, the iterative procedure (6.1.3) belongs to the class of infinite dimensional SA algorithms. A number of works has been devoted to such SA procedures, see e. g. [Wal77, BS89, YZ90, Nix84, Gol88, Yin92] (we will not give details about all of them here, an extensive discussion can be found in Chapter 5). A particular family of algorithms, which iteratively increase the dimension up to ∞ at the limit while remaining finite at each iteration, is usually referred to as the sieve approach. Such algorithms are fully constructive, in particular, the USA algorithm belongs to this class. There were a few papers studying the convergence rate results for sieve-type SA: [Nix84] shows asymptotic normality for a modified sieve-type Robbins-Monro procedure in the case of independent noise $H(z, V) = \tilde{H}(z) + V$, [CW02] generalizes the previous works on SA in Hilbert spaces and derives results on the convergence, the asymptotic normality and the mean convergence rate for sieve-type SA in a quite general setting. However, none of these works is adapted for dealing with the analysis of the USA algorithm convergence rate. In addition to the arguments in the introduction of Chapter 5, we note that the asymptotic normality result for sieve-type SA in [CW02, Theorem 3.1] (see also [Nix84] for a similar result in a less general setting) assumes a hypothesis (see [CW02, Assumption B3(1)] and [Nix84, eq. (3.3)]) that in our setting would take the form

$$(\mathrm{Id} - \Pi_n) G \Pi_n \to 0 \quad \text{as} \quad n \to +\infty, \tag{6.1.5}$$

where G is an operator on the Hilbert space defined by $G(f)(\cdot) = \nabla_z h(\phi^\star(\cdot), \cdot) f(\cdot)$, and Π_n is the projection on the subspace spanned by the first $n+1$ elements of the basis. Note that in our setting (6.1.5) does not hold even in the most standard cases. For example, an orthogonal polynomial basis $\{B_i, i \geq 0\}$ should satisfy a recurrent relation of the form

$$B_{n+1}(\theta) = (a_n \theta + b_n) B_n(\theta) + c_n B_{n-1}(\theta),$$

so that, taking for example h such that $\nabla_z h(\phi^\star(\theta), \theta) = \theta$, we get

$$(\mathrm{Id} - \Pi_n) G \Pi_n(B_n) = a_n^{-1} B_{n+1},$$

with a_n^{-1} typically not converging to 0 (e.g. $a_n = 2$ for all n for Chebyshev polynomials of the first kind, see [CHQZ06, Chapter 2]). The results on the mean rate convergence (e.g. [CW02, Section V]) also require certain properties on the truncations of $\nabla_z h(\phi^\star(\cdot), \cdot)$ that are generally not verified in our case. See, for example, [CW02, Assumptions A.3P, D.1P]: the counter-example in Remark 5.3.3 is valid for both of them.

Outline of the chapter. In Section 6.2 we present the model under study with various comments and examples. Section 6.3 is devoted to the statement of the main result. In Section 6.3.2 we interpret the assumptions and the main result in the case of polynomial sequences and apply it to optimize the dimension growth speed in the USA algorithm. In Section 6.4 we proceed with the proof of the main theorem. Some technical details are given in Appendix 6.A.

6.2 Model and assumptions

Let $\mathcal{V}$ be a metric space endowed with its Borel σ-field and $\Theta \subset \mathbb{R}^d$. Consider a function $H : \mathbb{R}^q \times \mathcal{V} \times \Theta \to \mathbb{R}^q$. Let π be a probability distribution on Θ and μ be a transition kernel from Θ to $\mathcal{V}$. For any measurable functions $f, g : \Theta \to \mathbb{R}$ we define the scalar product induced by π as

$$\langle f; g \rangle_\pi := \int_\Theta f(\theta) g(\theta) \pi(\mathrm{d}\theta). \tag{6.2.1}$$

By extension, for any measurable functions $f = (f_1, \cdots, f_q) : \Theta \to \mathbb{R}^q$ and $g : \Theta \to \mathbb{R}$, we write in the vector form

$$\langle f; g \rangle_\pi := \begin{bmatrix} \langle f_1; g \rangle_\pi \\ \cdots \\ \langle f_q; g \rangle_\pi \end{bmatrix}. \tag{6.2.2}$$

We denote by L_2^π the Hilbert space of functions $f : \Theta \to \mathbb{R}^q$ such that the norm $\|f\|_\pi := \sqrt{\sum_{i=1}^q \langle f_i; f_i \rangle_\pi}$ is finite. Let us fix an orthonormal basis $\{B_i, i \geq 0\}$ of L_2^π. Assume that we work on a probability space with expectation denoted by $\mathbb{E}$.

The USA algorithm was proposed in Chapter 5 to solve the following problem:

$$\text{Find } \phi^\star \text{ in } L_2^\pi \text{ such that } \int_{\mathcal{V}} H(\phi^\star(\theta), v, \theta) \mu(\theta, \mathrm{d}v) = 0, \qquad \pi\text{-a.s.} \tag{6.2.3}$$

Consider the decomposition of the solution $\phi^\star := \sum_{i \geq 0} u_i^\star B_i$. The USA algorithm aims at calculating the coefficients $\{u_i^\star, i \geq 0\}$. It is given through the update of an approximating sequence $\{u_i^k, i \geq 0\}$ (having only a finite number $m_k + 1$ of non-zero elements for all k) by

the SA scheme presented in Algorithm 4 below.

Input: Sequences $\{\gamma_k, k \geq 1\}$, $\{m_k, k \geq 0\}$, $\{M_k, k \geq 1\}$, $K \in \mathbb{N}$, $\{u_i^0, i = 0, \dots, m_0\}$
for $k = 0$ **to** $K - 1$, **do**
  sample $(\theta_{k+1}^s, V_{k+1}^s), s = 1 \dots, M_{k+1}$, under the distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$;
  for $i > m_{k+1}$ define $u_i^{k+1} = 0$
  for $i = 0$ **to** m_{k+1}, **do**
    $u_i^{k+1} = u_i^k - \gamma_{k+1} M_{k+1}^{-1} \sum_{s=1}^{M_{k+1}} H\left(\sum_{j=0}^{m_k} u_j^k B_j(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) B_i(\theta_{k+1}^s)$
Output: The vector $\{u_i^K, i = 0, \dots, m_K\}$.

Algorithm 4: The USA algorithm for the coefficients of the basis decomposition of $\phi^\star$.

See Section 5.2.3 for the details on the derivation of the USA algorithm. At iteration k the sequence $\{u_i^k, 0 \leq i \leq m_k\}$ approximates the chaos expansion coefficients of a solution $\phi^\star$ to the problem (6.2.3). The corresponding approximation ϕ^k of $\phi^\star$ is then given by

$$\phi^k := \sum_{i=0}^{m_k} u_i^k B_i. \tag{6.2.4}$$

The most general version of the algorithm also allows to project each newly updated approximation on some known convex subset $\mathcal{A}$ of L_2^π and converges under slightly weaker assumptions (see Section 5.3.2). However in this work, the assumptions will be sufficient for the version of the USA algorithm without projection, thus we do not consider such a set $\mathcal{A}$ here.

The USA algorithm approximates the whole function $\phi^\star$ within a single iterative procedure and avoids nested calculations. Theorem 5.3.5 states the a.s. and $L^p, p < 2$, convergence of ϕ^k to $\phi^\star$ with respect to $\|\cdot\|_\pi$. The full discussion about the advantages of this approach, compared to more naive methods, can be found in Section 5.2.

The goal of this work is to derive the L^2-convergence rate for the sequence $\{\phi^k, k \geq 0\}$ produced by Algorithm 4 (possibly under slightly stronger assumptions than those in Chapter 5). Namely, we aim at finding such $\alpha > 0$, that for some constant $C_\alpha > 0$ and for all $k \geq 0$ we would have $\mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right] \leq C_\alpha \gamma_k^\alpha$. Such a rate α is supposed to be defined in terms of the regularity of the solution $\phi^\star$, the chaos expansion basis $\{B_i, i \geq 0\}$ and the input sequences γ_k, m_k and M_k of the USA algorithm.

In Section 6.2.1 we continue with the assumptions on the model (namely on the functions H and h). Various comments on these assumptions are given in Section 6.2.2.

6.2.1 Assumptions

In what follows we denote $|x|$ and $x \cdot y$ the Euclidean norm and scalar product in $\mathbb{R}^q$.

The first assumption states properties that are quite standard for SA setting. We also suppose the uniqueness of the solution to (6.2.3)

H 1. *The following conditions hold:*

1. *For any $z \in \mathbb{R}^q$ and $\theta \in \Theta$ the integral $h(z,\theta) := \int_{\mathcal{V}} H(z,v,\theta)\,\mu(\theta,\mathrm{d}v)$ exists and*
$$\int_{\Theta\times\mathcal{V}} |H(z,v,\theta)|\mu(\theta,\mathrm{d}v)\pi(\mathrm{d}\theta) < +\infty.$$
2. *For any $\phi \in L_2^\pi$, the function $h(\phi(\cdot),\cdot)$ belongs to L_2^π. The mapping $\phi \mapsto h(\phi(\cdot),\cdot)$ from L_2^π into itself is continuous.*
3. *There exists a unique solution $\phi^\star \in L_2^\pi$ to the problem (6.2.3).*
4. *For π-a.a. $\theta \in \Theta$ and any $z \in \mathbb{R}^q$ such that $h(z,\theta) \neq 0$ we have $(z - \phi^\star(\theta))\cdot h(z,\theta) > 0$.*

Denote Mat_q the space of $q \times q$ matrices, and $\|\cdot\|_{\mathbb{R}^q}^{op.}$ the induced operator norm on Mat_q (i.e. $\|M\|_{\mathbb{R}^q}^{op.} = \sup_{|x|=1} |Mx|$). Note that h takes values in R^q and thus $\nabla_z h(z,\theta)$ is a $q \times q$ matrix.

H 2. *The derivative $\nabla_z h(z,\theta)$ exists for any $z \in \mathbb{R}^q$ and π-a.a. $\theta \in \Theta$, in addition:*

1. *For some constant $L_0 > 0$*
$$\sup_{z\in\mathbb{R}^q}\sup_{\theta\in\Theta} \|\nabla_z h(z,\theta)\|_{\mathbb{R}^q}^{op.} \le L_0.$$
2. *There exists a constant $C_H > 0$ such that for any $z \in \mathbb{R}^q$*
$$\sup_{\theta\in\Theta} \int_{\mathcal{V}} |H(z,v,\theta)|^2 \mu(\theta,\mathrm{d}v) \le C_H(1+|z|^2).$$

H 3. *For some constant $A_0 > 0$ one of the following holds:*

(a) for all $z \in \mathbb{R}^q$ and π-a.a. $\theta \in \Theta$ the matrix $\nabla_z h(z,\theta)$ is symmetric positive definite with all eigenvalues greater or equal to A_0,

(b) for all $z \in \mathbb{R}^q$ and π-a.a. $\theta \in \Theta$ the matrix $\nabla_z h(z,\theta)$ is upper triangular and such that for all $x \in \mathbb{R}^q$ we have $x^\mathsf{T}\nabla_z h(z,\theta)x \ge A_0|x|^2$ (in particular, diagonal elements of $\nabla_z h(z,\theta)$ are positive and greater or equal to A_0).

Let $\{u_i^\star, i \ge 0\}$ be the coefficients of $\phi^\star \in L_2^\pi$ in the base $\{B_i, i \ge 0\}$. For the sequence $\{m_k, k \ge 1\}$ controlling the dimension growth of the procedure define $\psi(m) := \inf\{k \ge 0 : m_k \ge m\}$. Also define

$$q_m := \sum_{i>m} |u_i^\star|^2, \qquad Q_m := \sup_{\theta\in\Theta} \sum_{i\le m} |B_i(\theta)|^2. \tag{6.2.5}$$

We note that a bound of the type $q_m = O(m^{-\delta})$ for some $\delta > 0$ may be deduced for various bases depending on the regularity of the solution $\phi^\star(\cdot)$ (see [CHQZ06, Chapter 2] and Section 5.2.2 for examples and discussion).

Define the set $\mathcal{E} \subset \mathbb{R}$ by

$$\mathcal{E} := \left\{ \alpha > 0 : \sup_{n \geq 0} \left(\gamma_n^{-\alpha} q_{m_n} + \gamma_n^{1-\alpha} \frac{Q_{m_n}}{M_n} + \frac{\gamma_{\psi(n)}^{-(1+\alpha/2)} |u_n^\star|}{(\psi(n+1) - \psi(n))} \right) < +\infty \right\}. \tag{6.2.6}$$

The next assumption specifies certain conditions on the input parameter sequences γ_k, m_k and M_k of the USA algorithm.

H 4. *Suppose that $\{M_k, k \geq 1\}$ and $\{m_k, k \geq 1\}$ are deterministic sequences of positive integers; $\{\gamma_k, k \geq 1\}$ is a deterministic decreasing sequence of positive real numbers such that:*

1. *The set $\mathcal{E}$ defined in (6.2.6) is non-empty.*
2. *For some $\kappa > 0$,*

$$\begin{aligned} &\sum_{k \geq 1} \gamma_k = +\infty, \quad \sum_{k \geq 1} \gamma_k^{1+\kappa} < +\infty, \quad \lim_{k \to \infty} \gamma_k^{-1} \log(\gamma_{k-1}/\gamma_k) = 0, \\ &\sum_{k \geq 1} \gamma_k^2 \frac{Q_{m_k}}{M_k} < +\infty, \quad \sum_{k \geq 1} \gamma_k^{1-\kappa} q_{m_k} < +\infty, \quad \sup_{m \geq 0} \frac{\gamma_{\psi(m)}}{\gamma_{\psi(m+1)}} < \infty. \end{aligned} \tag{6.2.7}$$

   where $\{q_m, m \geq 0\}$ and $\{Q_m, m \geq 0\}$ are defined in (6.2.5).

3. *The sequence $\{\psi(m), m \geq 0\}$ is strictly increasing (i.e. m_k grows only by 1).*

Assumption **H4** is fairly technical and may be not easy to understand in the current form. To make it clearer, we later interpret it in the case where the sequences involved have polynomial growth, see Section 6.3.2. We remark that $\mathcal{E}$ will appear to be a set of α for which (6.1.4) holds.

6.2.2 Comments on the assumptions

Here we discuss the assumptions **H1**, **H2** and **H3**. In particular, it will be useful to interpret them on the following example of SA applied to the minimization of a function given in the form of an expectation:

Example 6.2.1. *Consider a function of the form $u(z, \theta) := \mathbb{E}[U(z, V, \theta)]$ that is convex in z for fixed θ. Let $h(z, \theta) := \nabla_z u(z, \theta)$, so that (under certain assumptions) $h(z, \theta) = \mathbb{E}[\nabla_z U(z, V, \theta)]$. In this case, the problem (6.2.3) translates as the minimization of $u(\cdot, \theta)$ for an uncertain parameter $\theta \in \Theta$. The SA procedure in this case is known as Stochastic Gradient Descent.*

Assumption **H1** is mostly needed to verify the conditions of Theorem 5.3.5 that states the a.s. convergence. Properties **H1**-1,2 guarantee that the problem is well defined and satisfies basic requirements. **H1**-4 is also quite standard (see [BMP90, KY97b]). In the setting of Example 6.2.1 it follows from the convexity of the function $u(\cdot, \theta)$ for fixed values of θ.

Let us now discuss **H2** and **H3**. Their main purpose is to guarantee a uniform upper bound on $\nabla_z h$ (in **H2**-1) and a uniform positive lower bound on the eigenvalues of $\nabla_z h$,

i.e. the uniform repulsivity (in **H3**). In the case of Example 6.2.1 such global bounds are available (in the case $u(\cdot,\theta) \in \mathcal{C}^2$) if for all $z \in \mathbb{R}^q, \theta \in \Theta$

$$\lambda_{\min} \leq \text{ Eigenvalues of } \nabla_z^2 u(z,\theta) \leq \lambda_{\max} \tag{6.2.8}$$

for some constants $\lambda_{\min}, \lambda_{\max} > 0$.

Note that the uniform repulsivity of $\nabla_z u(\cdot,\theta)$ in a neighborhood of the solution $\phi^\star(\theta)$ is a quite standard assumption in classical results (see [BMP90, KY97b]) on the convergence rate and asymptotic normality of SA without uncertainty (i.e. for fixed θ).

Assumption **H3** imposes certain qualitative restrictions on $\nabla_z h(z,\theta)$ but nevertheless it covers the following important situations:

- general case in dimension $q = 1$;
- the case of convex function minimization (Example 6.2.1) since here $\nabla_z h(z,\theta)$ equals $\nabla_z^2 u(z,\theta)$, which is symmetric provided $u(\cdot,\theta) \in \mathcal{C}^2$;
- multi-component function h with triangular dependence, i.e. $h = (h_1, \ldots, h_q)$ with h_i depending only on $z_1, \ldots, z_i, \theta$ (this example was originally motivated by the simultaneous calculation of a quantile and the average above the quantile, which are also known as VaR and CVaR in financial applications, see [BFP09]);
- the same results will also hold for $\nabla_z h(z,\theta)$ that is lower triangular or that is built from symmetric and triangular blocks satisfying **H3**.

6.3 L^2-convergence rate of the USA algorithm

6.3.1 The main result

Theorem 6.3.1 below shows that $\mathcal{E}$ defined in (6.2.6) appears to be a set of the exponents α for which an L^2-error control with the speed γ_n^α may be provided. The motivation for such definition of $\mathcal{E}$ will be clear from the proof of Theorem 6.3.1. An explanation of the meaning of the three terms in the definition of $\mathcal{E}$ is also given in Section 6.3.2 for polynomial sequences γ_k, m_k and M_k.

Let $\{\phi^k, k \geq 0\}$ be given by (6.2.4). The following result provides the a.s. convergence $\phi^k \to \phi^\star$ and the non-asymptotic control of the squared L^2-error $\mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right]$.

Theorem 6.3.1. *Assume **H1**, **H2**, **H3**, **H4**. Assume that the set $\mathcal{E}$ defined in (6.2.6) is non-empty. Then for $\{\phi^k, k \geq 0\}$ given by (6.2.4) we have*

$$\left\|\phi^k - \phi^\star\right\|_\pi \overset{a.s.}{\underset{k\to+\infty}{\to}} 0,$$

and for any $\alpha \in \mathcal{E}$ we have for some constant $C_\alpha > 0$ and all $n \geq 0$ that

$$\mathbb{E}\left[\|\phi^n - \phi^\star\|_\pi^2\right] \leq C_\alpha \gamma_n^\alpha.$$

Note that if the set $\mathcal{E}$ has a maximum value then $\alpha = \max \mathcal{E}$ is the best bound on the convergence speed available from Theorem 6.3.1. This, for example, holds for the sequences q_m, Q_m, m_k, M_k and γ_k with polynomial growth, see Section 6.3.2.

6.3.2 Example: case of polynomial sequences

In this section we interpret the assumptions and the result of Theorem 6.3.1 in the case where the sequences involved have polynomial growth. Further we apply this result to optimally choose the speed of the dimension growth in the USA algorithm for the model studied in Section 5.5.

For sequences of the form

$$\gamma_k = k^{-a},\ M_k = \lfloor k^p \rfloor + 1,\ k \geq 1, \quad \text{and} \quad m_k = \lfloor k^b \rfloor + 1,\ k \geq 0, \tag{6.3.1}$$

with $a, b > 0$ and $p \geq 0$, we discuss how to choose these parameters to satisfy **H4** assuming that

$$q_m = O\left(m^{-\delta}\right), \qquad Q_m = O\left(m^{\Delta}\right), \tag{6.3.2}$$

for some $\delta > 0$ and $\Delta \geq 0$. Recall that $\psi(m) := \inf\{k \geq 0 : m_k \geq m\}$.

Let us interpret the condition **H4**: $\gamma_k^{-1} \log(\gamma_{k-1}/\gamma_k)$ means simply that $a < 1$. Condition $\sup_{m \geq 0} \dfrac{\gamma_{\psi(m)}}{\gamma_{\psi(m+1)}} < \infty$ is straightforward for any $a, b > 0$. Thus using in addition **C2** and (5.5.3) (where the other conditions of **H4** are expressed), **H4** writes as

$$0 < a < 1, \qquad 2 - \delta b < 2a, \qquad bR + 1 < 2a + p. \tag{6.3.3}$$

Now we analyze the definition (6.2.6) of the set $\mathcal{E}$. Let $\alpha \in \mathcal{E}$. The condition $\sup_n \gamma_n^{-\alpha} q_{m_n} < \infty$ writes as $\alpha \leq b\delta/a$. Further $\sup_n \gamma_n^{1-\alpha} \dfrac{Q_{m_n}}{M_n} < \infty$ is equivalent to $\alpha \leq 1 + \frac{p - b\Delta}{a}$. Finally consider the condition (to simplify we use exact equalities instead of asymptotics)

$$\sup_{n \geq 0} \frac{\gamma_{\psi(n)}^{-(1+\alpha/2)} |u_n^\star|}{(\psi(n+1) - \psi(n))} < \infty. \tag{6.3.4}$$

Denote $Y_n := \gamma_{\psi(n)}^{-(2+\alpha)} / (\psi(n+1) - \psi(n))^2$. It is not difficult to check that Y_n has polynomial growth of order $(2+\alpha)\frac{a}{b} - \left(\frac{2}{b} - 2\right)$. Assume that for some $\bar{\delta} > 0$ we have $|u_n^\star|^2 = O(n^{-\bar{\delta}})$. From (6.3.4) we deduce that

$$\frac{2a + a\alpha - 2}{b} \leq \bar{\delta} - 2 \Longleftrightarrow \alpha \leq \frac{b(\bar{\delta} - 2)}{a} - 2 + \frac{2}{a}.$$

Hence, the analysis above implies

$$\max \mathcal{E} \geq \alpha^\star := \min\left(\frac{b\delta}{a}, 1 + \frac{p - b\Delta}{a}, \frac{b(\bar{\delta} - 2)}{a} - 2 + \frac{2}{a}\right). \tag{6.3.5}$$

The non-emptiness of $\mathcal{E}$ reads as $\alpha^\star > 0$. In this case we get $\mathbb{E}[\|\phi^n - \phi^\star\|_\pi^2] \leq C_\alpha n^{-r}$ where r is given by

$$r := a\alpha = \min\Big(b\delta,\ a + p - b\Delta,\ b(\bar{\delta} - 2) + 2(1 - a)\Big). \tag{6.3.6}$$

In (6.3.6) the rate r writes as a minimum of 3 terms each of which has a natural interpretations. The term $b\delta$ is related to the speed of the truncation error decrease for $\phi^\star$. The term $a + p - b\Delta$ corresponds to the variance coming from the simulations of (θ_s^k, V_s^k). A typical rate in finite dimensional case is $r = a$ (see [Duf97]) while here an additional term bR corresponds to the increase of variance with dimension, and p reduces the variance since larger p means more simulations made at each iteration. The role of the last term $b(\bar{\delta} - 2) + 2(1 - a)$ is more subtle and related to the deviation of the sieve-type iterative procedure in Algorithm 4 from the "genuine" infinite-dimensional dynamics. It compensates the absence of the property $(\mathrm{Id} - \Pi_n)G\Pi_n \to 0$ discussed in the introduction. Details will be given in the proof on Theorem 6.3.1.

6.3.3 Application: parameter tuning in Algorithm 4.

We remark that in this work we do not carry out numerical tests, since an extensive numerical study has been already done in Chapter 5. Below we apply our result to optimize the choice of the dimension growth speed in the USA algorithm, which is determined by the parameter b.

The formula (6.3.6) for the convergence rate suggests how to optimally choose the speed of the dimension growth, determined by the parameter b. The strategy consists in choosing b that maximizes the convergence rate for given values of the other parameters.

Let us consider the example discussed in Chapter 5 (see Model 1 in Section 5.5.2). In the numerical test presented in Section 5.5.5 we have $a = 0.875, p = 0, \delta = 2, \Delta = 1$. Further, though the value of $\bar{\delta}$ cannot, in general, be improved theoretically beyond the trivial bound $\bar{\delta} \geq \delta$, a good heuristic guess when dealing with polynomial sequences is $\bar{\delta} = \delta + 1$. By (6.3.6) this gives the rate $r = \min(b + 0.25, 2b, 0.875 - b)$, so that we get the optimal value 0.3125. This is in line with the value of b giving optimal performance in Chapter 5. Note however that the values of the rate itself provided by (6.3.6) tend to be rather conservative and underestimate the actual empirically obtained rate (which is not a contradiction to Theorem 6.3.1). This has a natural explanation: first, the actual truncation error decrease may be faster than the theoretical bound; second the analysis of the convergence rate involve certain estimations that are by their nature non-optimal (e.g. control by the maximum of the basis functions, see Step 1 of Section 6.4.2). We expect that the rate may be improved in certain specific cases. However, Theorem 6.3.1 is a valuable contribution since it provides a bound on the convergence rate in the general case, that seems hard to improve. Moreover, it provides important information for the optimization of the design parameters, especially the dimension growth rate b, which allows to enhance the numerical performance of the USA algorithm.

6.4 Proof of Theorem 6.3.1

This section is devoted to the proof of Theorem 6.3.1, we suppose that all of its assumptions are verified.

6.4.1 Proof of the a.s. convergence

Let us check the assumptions of Theorem 5.3.5 for the USA Algorithm 4. Conditions **C1**, **C3**, **C4** follow from **H1**, condition **C2** follows from **H4**. Condition **C5**-(a) follows from **H2**-2. To show **C6** (in the case of a unique solution $\phi^\star$) we use **H3** and that for any $\phi \in L_2^\pi$

$$h(\phi(\theta),\theta) = \left(\int_0^1 \nabla_z h(\phi^\star(\theta) + t(\phi(\theta) - \phi^\star(\theta)),\theta)\mathrm{d}t\right)(\phi(\theta) - \phi^\star(\theta)).$$

So, finally by Theorem 5.3.5 for Case 1 (without projections) we obtain

$$\left\|\phi^k - \phi^\star\right\|_\pi \overset{\text{a.s.}}{\underset{k\to+\infty}{\to}} 0 \quad \text{and} \quad \sup_{k\geq 0} \mathbb{E}\left[\left\|\phi^k - \phi^\star\right\|_\pi^2\right] < +\infty. \tag{6.4.1}$$

6.4.2 Proof of the L^2-convergence rate

Recall that we denote by $\{\phi^k, k \geq 0\}$ the sequence given by $\phi^k := \sum_{i\geq 0} u_i^k B_i = \sum_{i=0}^{m_k} u_i^k B_i$ where $\{u_i^k, i \geq 0\}$ is updated via Algorithm 4, so that $u_i^k = 0$ for all $i > m_k$.

Define the filtration

$$\mathcal{F}_k := \sigma\left(\theta_\ell^s, V_\ell^s, 1 \leq s \leq M_\ell, 1 \leq \ell \leq k\right),\ k \geq 1.$$

For $k \geq 0$ and $i = 0, \ldots, m_{k+1}$ let us denote

$$\mathcal{H}_i^k := \int_\Theta h(\phi^k(\theta),\theta)B_i(\theta)\ \pi(\mathrm{d}\theta),$$
$$\eta_i^{k+1} := \frac{1}{M_{k+1}} \sum_{s=1}^{M_{k+1}} H\left(\phi^k(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) B_i(\theta_{k+1}^s) - \mathcal{H}_i^k,$$

and let $\eta^{k+1} := \sum_{i=0}^{m_{k+1}} \eta_i^{k+1} B_i$. Recall the definition of q_m and Q_m in (6.2.5). The next lemma provides an estimation of $\mathbb{E}\left[\left\|\eta^k\right\|_\pi^2\right]$.

Lemma 6.4.1. *Under the assumptions of Theorem 6.3.1, for some deterministic $C > 0$ we have that for all $k \geq 1$*

$$\mathbb{E}\left[\left\|\eta^k\right\|_\pi^2\right] \leq C\frac{Q_{m_k}}{M_k}.$$

Proof. For $k \geq 1$ and $i = 0, \ldots, m_k$ we decompose $\eta_i^k = A_{k,i} + B_{k,i}$ where

$$A_{k,i} := \frac{1}{M_k} \sum_{s=1}^{M_k} \left(H\left(\phi^{k-1}(\theta_k^s), V_k^s, \theta_k^s\right) - h\left(\phi^{k-1}(\theta_k^s), \theta_k^s\right)\right) B_i(\theta_k^s),$$

$$B_{k,i} := \frac{1}{M_k}\sum_{s=1}^{M_k} h\left(\phi^{k-1}(\theta_k^s), \theta_k^s\right) B_i(\theta_k^s) - \mathcal{H}_i^{k-1}.$$

We write for $i = 0, \dots, m_k$

$$\begin{aligned}\mathbb{E}\left[|\eta_i^k|^2 | \mathcal{F}_{k-1}\right] &= \mathbb{E}\left[\mathbb{E}\left[|A_{k,i}|^2 + 2A_{k,i}\cdot B_{k,i} + |B_{k,i}|^2 \middle| \theta_k^1, \dots, \theta_k^{M_k}, \mathcal{F}_{k-1}\right] \middle| \mathcal{F}_{k-1}\right] \\ &= \mathbb{E}[\mathbb{E}[|A_{k,i}|^2 \mid \theta_k^1, \dots, \theta_k^{M_k}, \mathcal{F}_{k-1}] | \mathcal{F}_{k-1}] + \mathbb{E}[|B_{k,i}|^2 | \mathcal{F}_{k-1}] \\ &= \mathbb{E}[|A_{k,i}|^2 \mid \mathcal{F}_{k-1}] + \mathbb{E}[|B_{k,i}|^2 | \mathcal{F}_{k-1}],\end{aligned} \tag{6.4.2}$$

using that

$$\mathbb{E}\left[A_{k,i}\cdot B_{k,i} \middle| \theta_k^1, \dots, \theta_k^{M_k}, \mathcal{F}_{k-1}\right] = \mathbb{E}\left[A_{k,i} \middle| \theta_k^1, \dots, \theta_k^{M_k}, \mathcal{F}_{k-1}\right] \cdot B_{k,i} = 0.$$

Define $\Gamma : \mathbb{R}^q \times \Theta \to \mathbb{R}$ by

$$\Gamma(z,\theta) := \int_{\mathcal{V}} |H(z,v,\theta) - h(z,\theta)|^2 \mu(\theta, \mathrm{d}v) \tag{6.4.3}$$

(well-defined due to **H2**-2). First

$$\mathbb{E}[|A_{k,i}|^2 \mid \theta_k^1, \dots, \theta_k^{M_k}, \mathcal{F}_{k-1}] = \frac{1}{M_k^2}\sum_{s=1}^{M_k} \Gamma\left(\phi^{k-1}(\theta_k^s), \theta_k^s\right) B_i(\theta_k^s)^2,$$

and hence

$$\mathbb{E}[|A_{k,i}|^2 | \mathcal{F}_{k-1}] = \frac{1}{M_k}\int_\Theta \Gamma(\phi^{k-1}(\theta), \theta) B_i(\theta)^2 \pi(\mathrm{d}\theta). \tag{6.4.4}$$

Second

$$\mathbb{E}[|B_{k,i}|^2 | \mathcal{F}_{k-1}] \le \frac{1}{M_k}\int_\Theta \left|h(\phi^{k-1}(\theta), \theta)\right|^2 B_i(\theta)^2 \pi(\mathrm{d}\theta). \tag{6.4.5}$$

From (6.4.2), using (6.4.4), (6.4.5) and taking the sum over $i = 0, \dots, m_k$, we obtain (since $\eta^k(\cdot) = \sum_{i=0}^{m_k} \eta_i^k B_i(\cdot)$) that

$$\mathbb{E}\left[\left\|\eta^k\right\|_\pi^2 | \mathcal{F}_{k-1}\right] \le \frac{Q_{m_k}}{M_k}\int_\Theta \left(\left|h(\phi^{k-1}(\theta), \theta)\right|^2 + \Gamma(\phi^{k-1}(\theta), \theta)\right) \pi(\mathrm{d}\theta). \tag{6.4.6}$$

Note that **H2**-2 implies that for any $z \in \mathbb{R}^q$, for π-a.a. $\theta \in \Theta$ and some deterministic $C > 0$

$$|h(z,\theta)|^2 + \Gamma(z,\theta) \le C(1 + |z|^2).$$

Thus taking expectation in (6.4.6) implies

$$\mathbb{E}\left[\left\|\eta^k\right\|_\pi^2\right] \le C\frac{Q_{m_k}}{M_k}\left(1 + \mathbb{E}\left[\left\|\phi^{k-1}\right\|_\pi^2\right]\right) \le C\sup_{j\ge 0}\left(1 + \mathbb{E}\left[\left\|\phi^j\right\|_\pi^2\right]\right)\frac{Q_{m_k}}{M_k}, \tag{6.4.7}$$

where $\sup_{j\ge 0}\mathbb{E}\left[\left\|\phi^j\right\|_\pi^2\right]$ is finite by (6.4.1). □

Since the gradient $\nabla_z h(\cdot,\cdot)$ exists on $\Theta \times \mathbb{R}^q$ by **H2**, for π-a.a. $\theta \in \Theta$ we have that

$$h(\phi^k(\theta),\theta) = \left(\int_0^1 \nabla_z h(\phi^\star(\theta) + t(\phi^k(\theta) - \phi^\star(\theta)),\theta)\mathrm{d}t\right)(\phi^k(\theta) - \phi^\star(\theta)). \tag{6.4.8}$$

For any $f : \Theta \to \mathbb{R}$ we denote $\mathrm{EssSup}_{\theta\in\Theta} f(\theta)$ the essential supremum of f with respect to the measure π on Θ. Define $\mathcal{M}^k : \Theta \to \mathrm{Mat}_q$ by

$$\mathcal{M}^k(\cdot) := \int_0^1 \nabla_z h(\phi^\star(\cdot) + t(\phi^k(\cdot) - \phi^\star(\cdot)),\cdot)\mathrm{d}t. \tag{6.4.9}$$

From **H2**-1 we deduce that

$$\sup_{k\geq 0} \mathrm{EssSup}_{\theta\in\Theta} \left\|\mathcal{M}^k(\theta)\right\|^{op.}_{\mathbb{R}^q} \leq L_0. \tag{6.4.10}$$

We proceed with the following lemma, which provides a uniform contraction property of the operators $(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k(\theta)), \theta \in \Theta$.

Lemma 6.4.2. *For any $A_1 \in (0, A_0)$ (A_0 given by **H3**) there exists $t \in (0,1]$ ($t = 1$ in the case **H3**-a) and $k_0 \in \mathbb{N}$ both depending only on L_0, q, A_0, A_1 and the sequence $(\gamma_n)_{n\geq 0}$, such that for all $k \geq k_0$ and the matrix $D_t := \mathrm{Diag}(t, t^2, \ldots, t^q)$ we have*

$$\mathrm{EssSup}_{\theta\in\Theta} \left\|D_t^{-1}\left(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k(\theta)\right)D_t\right\|^{op.}_{\mathbb{R}^q} \leq (1 - \gamma_{k+1}A_1).$$

The proof of Lemma 6.4.2 is given in Section 6.A.1 and essentially follows from **H3**.

Let $\Pi_m : L_2^\pi \to L_2^\pi$ be the projection on the subspace of L_2^π spanned by $\{B_i, i = 0, \ldots, m\}$. From Algorithm 4, (6.4.8), (6.4.9) and the definition of η^k we obtain the following recurrent equation for $\{\phi^k - \phi^\star, k \geq 0\}$

$$\phi^{k+1} - \phi^\star = \phi^k - \phi^\star - \gamma_{k+1}\Pi_{m_{k+1}}(\mathcal{M}^k(\phi^k - \phi^\star)) - \gamma_{k+1}\eta^{k+1}. \tag{6.4.11}$$

First let us pass from (6.4.11) to a procedure with iterates having only a finite number of non-zero coefficients in the basis decomposition. Denote $R_m := (\mathrm{Id} - \Pi_m)\phi^\star$ so that q_m defined in (6.2.5) equals $\|R_m\|_\pi^2$. We have

$$\phi^k - \phi^\star = \Pi_{m_k}(\phi^k - \phi^\star) - R_{m_k}. \tag{6.4.12}$$

Since the term R_{m_k} is deterministic, it remains to analyze $\Pi_{m_k}(\phi^k - \phi^\star)$. Denote

$$t^k := (\Pi_{m_k} - \Pi_{m_{k-1}})\phi^\star. \tag{6.4.13}$$

Applying $\Pi_{m_{k+1}}$ to (6.4.11) we get the following recurrence equation for the sequence $\Pi_{m_k}(\phi^k - \phi^\star)$

$$\Pi_{m_{k+1}}(\phi^{k+1} - \phi^\star) = \Pi_{m_k}(\phi^k - \phi^\star) - \gamma_{k+1}\Pi_{m_{k+1}}(\mathcal{M}^k(\phi^k - \phi^\star)) - \gamma_{k+1}\eta^{k+1} - t^{k+1}. \tag{6.4.14}$$

Our strategy is to decompose $\Pi_{m_k}(\phi^k - \phi^\star)$ into two sequences, so that one of them has the

same recurrence equation as (6.4.14) but without the term t^{k+1} and to put all the rest in the second one. Following this idea, define the sequences $\{\mu^k, k \geq 0\}$ and $\{\rho^k, k \geq 0\}$ in L_2^π by $\mu_0 = 0, \rho^0 = \Pi_{m_0}(\phi^0 - \phi^\star)$ and for all $k \geq 1$

$$\mu^{k+1} = \mu^k - \gamma_{k+1}\Pi_{m_{k+1}}\mathcal{M}^k\mu^k - \gamma_{k+1}\eta^{k+1}, \tag{6.4.15}$$

$$\rho^{k+1} = \rho^k - \gamma_{k+1}\Pi_{m_{k+1}}\mathcal{M}^k\rho^k + \gamma_{k+1}\left(\Pi_{m_{k+1}}\mathcal{M}^k R_{m_k} - \gamma_{k+1}^{-1}t^{k+1}\right). \tag{6.4.16}$$

Taking the sum of (6.4.15) and (6.4.16), and using (6.4.14), (6.4.12), it is easy to show by induction that $\Pi_{m_k}(\phi^k - \phi^\star) = \mu^k + \rho^k$.

Now we analyze the sequences $\{\mu^k, k \geq 0\}$ and $\{\rho^k, k \geq 0\}$ separately. We estimate first $\mathbb{E}\left[\left\|\mu^k\right\|_\pi^2\right]$ using the uniform contraction property of the operators $(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k(\theta))$ given by Lemma 6.4.2. Further we provide a deterministic control on $\left\|\rho^k\right\|_\pi^2$. In particular, we obtain, using **H4**, the speed of convergence of these errors to 0. The definition of the convergence rate set $\mathcal{E}$ will be clear from various stages of the proof.

In what follows, C denotes a deterministic constant that may change from line to line.

Step 1. Estimation of $\mathbb{E}\left[\left\|\mu^k\right\|_\pi^2\right]$. From the equation (6.4.15) we have

$$\mu^{k+1} = \Pi_{m_{k+1}}\left(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k\right)\mu^k - \gamma_{k+1}\eta^{k+1}, \tag{6.4.17}$$

Fix $A_1 \in (0, A_0)$. Consider D_t with t given by Lemma 6.4.2. Multiplying (6.4.17) by D_t^{-1} and using that D_t^{-1} and Π_m commute for any m (since linear operations commute with the truncation Π_m) we get

$$D_t^{-1}\mu^{k+1} = \Pi_{m_{k+1}}(D_t^{-1}\left(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k\right)D_t)D_t^{-1}\mu^k - \gamma_{k+1}D_t^{-1}\eta^{k+1}.$$

This implies

$$\mathbb{E}\left[\left\|D_t^{-1}\mu^{k+1}\right\|_\pi^2\right] = \mathbb{E}\left[\left\|\Pi_{m_{k+1}}(D_t^{-1}\left(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k\right)D_t)D_t^{-1}\mu^k\right\|_\pi^2\right] + \gamma_{k+1}^2\mathbb{E}\left[\left\|\eta^{k+1}\right\|_\pi^2\right],$$

since μ^k and $\mathcal{M}^k$ are $\mathcal{F}_k$-measurable, $\mathbb{E}[\eta^{k+1}|\mathcal{F}_k] = 0$ which yields

$$\mathbb{E}\left[\left\langle\Pi_{m_{k+1}}(D_t^{-1}\left(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k\right)D_t)D_t^{-1}\mu^k; D_t^{-1}\eta^{k+1}\right\rangle_\pi\right] = 0.$$

Applying Lemma 6.4.2, we have for some deterministic $k_0 \in \mathbb{N}$ for any $k \geq k_0$

$$\mathbb{E}\left[\left\|D_t^{-1}\mu^{k+1}\right\|_\pi^2\right] \leq (1 - \gamma_{k+1}A_1)^2\mathbb{E}\left[\left\|D_t^{-1}\mu^k\right\|_\pi^2\right] + \gamma_{k+1}^2\mathbb{E}\left[\left\|D_t^{-1}\eta^{k+1}\right\|_\pi^2\right].$$

We deduce by induction that for some constant $C > 0$ coming from the terms with $k \leq k_0$

$$\mathbb{E}\left[\left\|D_t^{-1}\mu^n\right\|_\pi^2\right] \leq C\sum_{k=1}^n \gamma_k^2 \exp\left(-2A_1\sum_{j=k}^n \gamma_j\right)\mathbb{E}\left[\left\|D_t^{-1}\eta^k\right\|_\pi^2\right]$$

(note that C is finite since for all $k \geq 0$ the term $\mathbb{E}\left[\left\|\eta^{k+1}\right\|_\pi^2\right]$ is finite from Lemma 6.4.1 and $\mathbb{E}\left[\left\|\mu^k\right\|_\pi^2\right]$ is finite from (6.4.17) and (6.4.10)). Let $C_{1,t} := C\left\|D_t^{-1}\right\|_{R^q}^{op.}\|D_t\|_{R^q}^{op.}$. Then

$$\mathbb{E}[\|\mu^n\|_\pi^2] \leq C_{1,t}\sum_{k=1}^n \gamma_k^2 \exp\left(-2A_1\sum_{j=k+1}^n \gamma_j\right)\mathbb{E}\left[\left\|\eta^k\right\|_\pi^2\right]. \tag{6.4.18}$$

Let the set $\mathcal{E}$ be given by (6.2.6) and $\alpha \in \mathcal{E}$. Multiplying (6.4.18) by $\gamma_n^{-\alpha}$ and using (6.4.7) we obtain for a different constant $C > 0$

$$\gamma_n^{-\alpha}\mathbb{E}[\|\mu^n\|_\pi^2] \leq C\gamma_n^{-\alpha}\sum_{k=1}^n \gamma_k^{1+\alpha}\exp\left(-2A_1\sum_{j=k+1}^n \gamma_j\right)\left(\gamma_k^{1-\alpha}\frac{Q_{m_k}}{M_k}\right). \tag{6.4.19}$$

In view of **H4** we may apply Lemma 6.A.2 to get

$$\limsup_{n\to\infty}\gamma_n^{-\alpha}\sum_{k=1}^n \gamma_k^{1+\alpha}\exp\left(-2A_1\sum_{j=k}^n \gamma_j\right)\left(\gamma_k^{1-\alpha}\frac{Q_{m_k}}{M_k}\right) \leq \frac{1}{2A_1}\limsup_{n\to\infty}\left(\gamma_n^{1-\alpha}\frac{Q_{m_n}}{M_n}\right) < +\infty,$$

where the last inequality holds since $\alpha \in \mathcal{E}$ (see (6.2.6)). Using (6.4.19) this implies

$$\sup_{n\geq 0}\gamma_n^{-\alpha}\mathbb{E}[\|\mu^n\|_\pi^2] < +\infty. \tag{6.4.20}$$

Step 2. Estimation of $\left\|\rho^k\right\|_\pi$. From (6.4.16) we have

$$\rho^{k+1} = \Pi_{m_{k+1}}\left(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k\right)\rho^k + \gamma_{k+1}\left(\Pi_{m_{k+1}}\mathcal{M}^k R_{m_k} - \gamma_{k+1}^{-1}t^{k+1}\right). \tag{6.4.21}$$

Consider D_t with t given by Lemma 6.4.2. Multiplying (6.4.21) by D_t^{-1} and, using that D_t^{-1} and Π_m commute for any m, we get

$$D_t^{-1}\rho^{k+1} = \Pi_{m_{k+1}}(D_t^{-1}\left(\mathrm{Id} - \gamma_{k+1}\mathcal{M}^k\right)D_t)D_t^{-1}\rho^k + \gamma_{k+1}D_t^{-1}\left(\Pi_{m_{k+1}}\mathcal{M}^k R_{m_k} - \gamma_{k+1}^{-1}t^{k+1}\right).$$

By (6.4.10) we have for some deterministic $C > 0$

$$\left\|\Pi_{m_{k+1}}\mathcal{M}^k R_{m_k}\right\|_\pi \leq L_0\left\|R_{m_k}\right\|_\pi = Cq_{m_k}^{1/2}.$$

By Lemma 6.4.2 for some deterministic $k_0 \in \mathbb{N}$ we have for any $k \geq k_0$

$$\operatorname*{EssSup}_{\theta\in\Theta} \left\| \Pi_{m_{k+1}} D_t^{-1} \left(\mathrm{Id} - \gamma_{k+1} \mathcal{M}^k(\theta) \right) D_t \right\|_{\mathbb{R}^q}^{op.} \leq (1 - \gamma_{k+1} A_1).$$

Thus for all $k \geq k_0$

$$\left\| D_t^{-1} \rho^{k+1} \right\|_\pi = (1 - \gamma_{k+1} A_1) \left\| D_t^{-1} \rho^k \right\|_\pi + C \gamma_{k+1} \left\| D_t^{-1} \right\|_{\mathbb{R}^q}^{op.} \left(q_{m_k}^{1/2} + \gamma_{k+1}^{-1} \left\| t^{k+1} \right\|_\pi \right).$$

So finally we get for some another constant $C > 0$, coming from the terms with $k \leq k_0$, that

$$\|\rho^n\|_\pi \leq C \sum_{k=1}^n \gamma_k \exp\left(-A_1 \sum_{j=k+1}^n \gamma_j \right) \left(q_{m_k}^{1/2} + \gamma_{k+1}^{-1} \left\| t^{k+1} \right\|_\pi \right) \tag{6.4.22}$$

(C is finite from (6.4.21) and $\left\| t^{k+1} \right\|_\pi \leq \|\phi^\star\|_\pi$)). First, for any $\alpha \in \mathcal{E}$ (so that $\sup_n \gamma_n^{-\alpha} q_{m_n} < +\infty$, see (6.2.6)) we write using Lemma 6.A.2

$$\begin{aligned}
&\limsup_{n\to\infty} \gamma_n^{-\alpha/2} \sum_{k=1}^n \gamma_k \exp\left(-A_1 \sum_{j=k+1}^n \gamma_j \right) q_{m_k}^{1/2} \\
&= \limsup_{n\to\infty} \gamma_n^{-\alpha/2} \sum_{k=1}^n \gamma_k^{1+\alpha/2} \exp\left(-A_1 \sum_{j=k+1}^n \gamma_j \right) (\gamma_k^{-\alpha/2} q_{m_k}^{1/2}) \\
&\leq A_1^{-1} \limsup_{n\to\infty} \gamma_n^{-\alpha/2} q_{m_n}^{1/2} < +\infty
\end{aligned}$$

Second, using the definition (6.4.13) of t^k, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \gamma_n^{-\alpha/2} \sum_{k=1}^n \gamma_k^{1+\alpha/2} \exp\left(-A_1 \sum_{j=k+1}^n \gamma_j \right) \left(\gamma_k^{-1-\alpha/2} \left\| t^k \right\|_\pi \right) \\
&= \limsup_{m\to\infty} \gamma_{\psi(m)}^{-\alpha} \sum_{s=1}^m \gamma_{\psi(s)}^{1+\alpha/2} \exp\left(-A_1 \sum_{j=\psi(s)+1}^{\psi(m)} \gamma_j \right) \gamma_{\psi(s)}^{-1-\alpha/2} |u_s^\star|
\end{aligned} \tag{6.4.23}$$

For any $\psi(s) \leq k < \psi(s+1)$ we have $\exp\left(-A_1 \sum_{j=\psi(s)+1}^{\psi(m)} \gamma_j \right) \leq \exp\left(-A_1 \sum_{j=k+1}^{\psi(m)} \gamma_j \right)$. So, using also that $\sup_{m\geq 0} \dfrac{\gamma_{\psi(m)}}{\gamma_{\psi(m+1)}} < +\infty$ and γ_n is decreasing by **H4**, we deduce that the right-hand side term of (6.4.23) will be finite if

$$\limsup_{n\to\infty} \gamma_n^{-\alpha/2} \sum_{k=1}^n \gamma_k^{1+\alpha/2} \exp\left(-A_1 \sum_{j=k}^n \gamma_j \right) \frac{\gamma_k^{-1-\alpha/2} |u_{m_k}^\star|}{(\psi(m_k+1) - \psi(m_k))} < +\infty. \tag{6.4.24}$$

Now (6.4.24) follows from Lemma 6.A.2 and since for any $\alpha \in \mathcal{E}$ we have

$$\limsup_{n\to\infty} \frac{\gamma_n^{-1-\alpha/2} |u_{m_n}^\star|}{(\psi(m_n+1) - \psi(m_n))} = \limsup_{m\to\infty} \frac{\gamma_{\psi(m)}^{-1-\alpha/2} |u_m^\star|}{(\psi(m+1) - \psi(m))} < +\infty.$$

Coming back to (6.4.22), in view of the analysis above, we get for some deterministic $C > 0$

$$\gamma_n^{-\alpha/2} \|\rho^n\|_\pi \le C, \tag{6.4.25}$$

and, in particular, $\mathbb{E}\left[\|\rho^n\|_\pi^2\right] \le C\gamma_n^\alpha$.

Step 3. Completion of the proof. We proceed with the final estimation. We write

$$\phi^n - \phi^\star = \mu^n + \rho^n - R_{m_n}.$$

Since $\alpha \in \mathcal{E}$ we have $\sup_n(\gamma_n^{-\alpha} q_{m_n}) < \infty$ where $q_m = \|R_m\|_\pi^2$. So, using (6.4.20) and (6.4.25) we conclude that for some deterministic $C_1 > 0$

$$\mathbb{E}[\|\phi^n - \phi^\star\|_\pi^2] \le 3\left(\mathbb{E}\left[\|\mu^n\|_\pi^2\right] + \mathbb{E}\left[\|\rho^n\|_\pi^2\right] + q_{m_n}\right) \le C_1\gamma_n^\alpha.$$

The proof is complete. □

6.A Technical Lemmas and Proofs

6.A.1 Proof of Lemma 6.4.2

Before we prove Lemma 6.4.2 we need the following auxiliary result.

Lemma 6.A.1. *Let $\mathcal{M}$ be a $q \times q$ upper triangular matrix such that all diagonal elements are greater or equal to some $A_0 > 0$ and $\|\mathcal{M}\|_{\mathbb{R}^q}^{op.} < a$ for some $a > 0$. Then for any $A_1 \in (0, A_0)$ there exist $\bar\gamma > 0$ and $t \in (0, 1]$ both depending only on a, q, A_0 and A_1 such that for the matrix $D_t := \mathrm{Diag}(t, t^2, \dots, t^q)$ we have for any $0 < \gamma \le \bar\gamma$*

$$\left\|D_t^{-1}(\mathrm{Id} - \gamma\mathcal{M})D_t\right\|_{\mathbb{R}^q}^{op.} \le (1 - \gamma A_1).$$

Proof. We have

$$D_t^{-1}\mathcal{M}D_t = \begin{bmatrix} \lambda_1 & t\mathcal{M}_{1,2} & t^2\mathcal{M}_{1,3} & \cdots & t^{q-1}\mathcal{M}_{1,q} \\ 0 & \lambda_2 & t\mathcal{M}_{2,3} & \cdots & t^{q-1}\mathcal{M}_{2,q} \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & t\mathcal{M}_{q-1,q} \\ 0 & 0 & 0 & \cdots & \lambda_q \end{bmatrix} = \Lambda + R_t,$$

where $\Lambda := \mathrm{Diag}(\lambda_1, \dots, \lambda_q)$ is the diagonal of $\mathcal{M}$ with $\lambda_i \ge A_0 > 0$.

Denote $\|\cdot\|_F$ the matrix Frobenius norm. Let c_q be such constant that for all M

$$c_q^{-1}\|M\|_{\mathbb{R}^q}^{op.} \le \|M\|_F \le c_q\|M\|_{\mathbb{R}^q}^{op.}.$$

for all $q \times q$ matrices M. We have for any $t \in (0,1]$

$$\|R_t\|_{\mathbb{R}^q}^{op.} \le c_q \|R_t\|_F \le c_q t \left(\sum_{i<j\le q} \mathcal{M}_{i,j}^2 \right)^{1/2} \le c_q t \|\mathcal{M}\|_F \le c_q^2 t a.$$

So we get

$$\|R_t\|_{\mathbb{R}^q}^{op.} \le t c_q^2 \, \|\mathcal{M}\|_{\mathbb{R}^q}^{op.} \le t c_q^2 a, \tag{6.A.1}$$

where c_q depends only on q. Take $\bar\gamma > 0$ such that $\bar\gamma a < 1$. In particular, since λ_i's are the eigenvalues of $\mathcal{M}$, we have $\bar\gamma \lambda_i < 1$ for all i and $\bar\gamma A_0 < 1$. Thus for any $A_1 \in (0, A_0)$ there exists $t > 0$ such that $t c_q^2 a \le (A_0 - A_1)/2$ which we fix from now on. For any $\gamma \in (0, \bar\gamma]$ and $t \in (0,1]$ we get

$$\left\| D_t^{-1}(\mathrm{Id} - \gamma\mathcal{M}) D_t \right\|_{\mathbb{R}^q}^{op.} \le \|\mathrm{Id} - \gamma\Lambda\|_{\mathbb{R}^q}^{op.} + \|\gamma R_t\|_{\mathbb{R}^q}^{op.}.$$

Note that

$$\|\mathrm{Id} - \gamma\Lambda\|_{\mathbb{R}^q}^{op.} \le (1 - \gamma A_0)$$

and thus, using that $t c_q^2 a \le (A_0 - A_1)/2$ and (6.A.1) we get

$$\left\| D_t^{-1}(\mathrm{Id} - \gamma\mathcal{M}) D_t \right\|_{\mathbb{R}^q}^{op.} \le (1 - \gamma A_0) + \gamma \frac{A_0 - A_1}{2} = (1 - \gamma A_1).$$

Remark again that the choice of t and $\bar\gamma$ depends only on a, q, A_0 and A_1. □

Proof of Lemma 6.4.2. For the case **H3**-a we take $t = 1$ and $\bar\gamma > 0$ such that $\bar\gamma \, \|\mathcal{M}\|_{\mathbb{R}^q}^{op.} < 1$. Further we use that for any symmetric positive definite matrix $M = U^* \Lambda U$, with U orthogonal and Λ diagonal we have $\|M\|_{\mathbb{R}^q}^{op.} = \|\Lambda\|_{\mathbb{R}^q}^{op.}$.

For the case **H3**-b, let $\bar\gamma > 0$ and $t \in (0,1]$ be given by Lemma 6.A.1. Let k_0 be such that $\gamma_k \le \bar\gamma$ for all $k \ge k_0$. Now, in view of the bound (6.4.10), the result follows from Lemma 6.A.1. □

6.A.2 An auxiliary lemma

For convenience we state here a simplified version of [For14, Lemma 5.9]

Lemma 6.A.2. *Let $\{\gamma_k, k \ge 0\}$ be a positive sequence such that*

$$\lim_{k\to\infty} \gamma_k = 0, \quad \lim_{k\to\infty} \gamma_k^{-1} \log(\gamma_{k-1}/\gamma_k) = 0, \quad \sum_k \gamma_k = +\infty.$$

Let $e_k, k \ge 0$ be a non-negative sequence. Then for any $V > 0, p \ge 0$

$$\limsup_{n\to\infty} \gamma_n^{-p} \sum_{k=1}^{n} \gamma_k^{p+1} \exp\left(-V \sum_{j=k+1}^{n} \gamma_j \right) e_k \le V^{-1} \limsup_{n\to\infty} e_n.$$

Bibliography

[AGZ09] G.W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge University Press, 2009.

[AO01] J. An and A. Owen. Quasi-regression. *Journal of Complexity*, 17(4):588–607, 2001.

[AS99] Y. Aït-Sahalia. Transition densities for interest rate and other nonlinear diffusions. *J. Finance*, 45:1361–1395, 1999.

[AS02] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70:223–262, 2002.

[AS08] Y. Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.*, 36:906–937, 2008.

[ASJ14] Y. Ait-Sahalia and J. Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

[ASM03] Y. Ait-Sahalia and P.A. Mykland. The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, 71(2):483–549, 2003.

[ASM04] Y. Ait-Sahalia and P.A. Mykland. Estimators of diffusions with randomly spaced discrete observations: A general theory. *Annals of Statistics*, 32:2186–2222, 2004.

[AT08] C. Andrieu and J. Thoms. A tutorial on adaptive mcmc. *Statistics and Computing*, 18:343–373, 2008.

[BC05] L. Bottou and Y. Le Cun. On-line learning for very large data sets. *Applied Stochastic Models in Business and Industry*, 21(2):137–151, 2005.

[BCN17] L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *http://arxiv.org:443/find/all/1/all:+AND+curtis+nocedal/0/1/0/all/0/1*, 2017.

[BFP09] O. Bardou, N. Frikha, and G. Pages. Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009.

[Bil95] P. Billingsley. *Probability and measure*. Wiley-Interscience, third edition, Wiley Series in Probability and Statistics, 1995.

[BMP90] A. Benveniste, M. Metivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics*. Springer-Verlag, 1990.

[BNGJ+06] O. E. Barndorff-Nielsen, S.E. Graversen, J. Jacod, M. Podolskij, and N. Shephard. A central limit theorem for realised power and bipower variation of continuous semimartingales. In *From stochastic calculus to mathematical finance. The Shiryaev Festschrift. Allmost all papers based on the presentation at the second Bachelier colloquium on stochastic calculus and probability, Meatbief, France, January 9–15, 2005.*, pages 33–68. Berlin: Springer, 2006.

[BNS05] O. E. Barndorff-Nielsen and N. Shephard. Power variation and time change. *Theory of Probability and Its Applications*, 50:1–15, 2005.

[BP11] J-Ph. Bouchaud and M. Potters. Financial Applications of Random Matrix Theory: a short review. *The Oxford handbook of random matrix theory, 824–850, Oxford Univ. Press, Oxford*, 2011.

[BS89] N. Berman and A. Shwartz. Abstract stochastic approximations and applications. *Stochastic Processes and their Applications*, 31(1):133–149, 1989.

[BS02] A.N. Borodin and P. Salminen. *Handbook of Brownian motion: facts and formulae*. Springer Basel AG, second edition, 2002.

[BTZ04] I. Babuska, R. Tempone, and G.E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42:800–825, 2004.

[BY82a] M.T. Barlow and M. Yor. Semi-martingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. *Journal of Functional Analysis, 49:198–229*, 1982.

[BY82b] M.T. Barlow and M. Yor. Semi-martingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. *Journal of Functional Analysis*, 49:198–229, 1982.

[CC11] J. Chang and S.X. Chen. On the approximate maximum likelihood estimation for diffusion processes. *Ann. Statist.*, 39:2820–2851, 2011.

[CHQZ06] C. Canuto, M.Y. Hussaini, A. Quarteroni, and T.A. Zang. *Spectral Methods: Fundamentals in Single Domains*. Springer Verlag, 2006.

[CW02] X. Chen and H. White. Asymptotic properties of some projection-based Robbins-Monro procedures in a Hilbert space. *Studies in Nonlinear Dynamics and Econometrics*, 6(1), 2002.

[CY90] L. Le Cam and G.L. Yang. *Asymptotics in statistics*. Springer Verlag, 1990.

[DB16] A. Dieuleveut and F. Bach. Nonparametric stochastic approximation with large step-sizes. *The Annals of Statistics*, 44(4):1363–1399, 2016.

[DG04] D. Duffie and P. Glynn. Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals. *Econometrica*, 72(6):1773–1808, 2004.

[DGM+01] M. Dacoronga, R. Gençay, U. Müller, R. Olsen, and O. Pictet. *Introduction to high-frequency finance*. Academic Press, 2001.

[DH13] M. Deaconu and S. Herrmann. Hitting time for Bessel processes - walk on moving spheres algorithm. *Annals of Applied Probability*, 23(6):2259–2289, 2013.

[Doh87] G. Dohnal. On estimating the diffusion coefficient. *Journal of Applied Probability*, 24:105–114, 1987.

[Duf97] M. Duflo. *Random Iterative Models*, volume 34 of *Applications of Mathematics*. Springer, 1997.

[DVA98] D. Dufresne and F.J. Vázquez-Abad. Accelerated simulation for pricing Asian options. *Proceedings of the 1998 Winter Simulation Conference. Piscataway, NJ, IEEE Press*, pages 1493–1500, 1998.

[FJLS16] G. Fort, B. Jourdain, T. Lelievre, and G. Stoltz. Convergence and efficiency of adaptive importance sampling techniques with partial biasing. *https://hal.archives-ouvertes.fr/hal-01389996*, 2016.

[FMP11] G. Fort, E. Moulines, and P. Priouret. Convergence of adaptive and interacting markov chain monte carlo algorithms. *The Annals of Statistics*, 39:3262–3289, 2011.

[FO15] M. Fukasawa and J. Obloj. Efficient discretisation of stochastic differential equations. *arXiv:1506.05680*, 2015.

[For14] G. Fort. Central limit theorems for stochastic approximations with controlled markov chain dynamics. *ESAIM: Probability and Statistics*, 2014.

[FR12] M. Fukasawa and M. Rosenbaum. Central limit theorems for realized volatility under hitting times of an irregular grid. *Stochastic Processes and their Applications*, 122(12):3901–3920, 2012.

[FS00] M.C. Fu and Y. Su. Optimal importance sampling in securities pricing. *Journal of Computational Finance*, 5(4):27–50, 2000.

[Fuc13] C. Fuchs. *Inference for diffusion processes*. Springer-Verlag Berlin Heidelberg, 2013.

[Fuk10] M. Fukasawa. Central limit theorem for the realized volatility based on tick time sampling. *Finance Stoch.*, 14:209–233, 2010.

[Fuk11a] M. Fukasawa. Asymptotically efficient discrete hedging. *Stochastic analysis with Financial Applications. Progress in Probability*, 65:331–346, 2011.

[Fuk11b] M. Fukasawa. Discretization error of stochastic integrals. *Annals of Applied Probability*, 21(4):1436–1465, 2011.

[Fun92] D. Funaro. *Polynomial Approximation of Differential Equations*, volume 8 of *Lecture Notes in Physics Monographs*. Springer-Verlag Heidelberg, 1992.

[FZ93] D. Florens-Zmirou. On estimating the diffusion coefficient from discrete observations. *Journal of applied probability*, 30(4):790–804, 1993.

[GCJ93] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. Henri Poincaré*, 29:119–151, 1993.

[GCJ94] V. Genon-Catalot and J. Jacod. Estimation of the diffusion coefficient for Diffusion processes: Random Sampling. *Scandinavian Journal of Statistics*, 21:193–221, 1994.

[GG04] C. Geiss and S. Geiss. On approximation of a class of stochastic integrals and interpolation. *Stochastics and Stochastics Reports, 76(4):339–362*, 2004.

[GJY03] A. Göing-Jaeschke and M. Yor. A survey and some generalizations of Bessel processes. *Bernoulli*, 9(2):313–349, 2003.

[GL14a] E. Gobet and N. Landon. Almost sure optimal hedging strategy. *Ann. Appl. Probab.*, 24(4):1652–1690, 2014.

[GL14b] E. Gobet and N. Landon. Optimization of joint p-variations of Brownian semimartingales. *Electronic Journal of Probability*, 19(36), 2014.

[GM10] E. Gobet and S. Menozzi. Stopped diffusion processes: boundary corrections and overshoot. *Stochastic Processes and Their Applications*, 120:130–162, 2010.

[GMIC14] P. Gupta, M.K. Mehlawat, M. Inuiguchi, and S. Chandra. *Portfolio Optimization: An Overview*, volume 316. Springer Berlin Heidelberg, 2014.

[Gob01] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion: a malliavin calculus approach. *Bernoulli*, 7:899–912, 2001.

[Gol88] L. Goldstein. Minimizing noisy functionals in hilbert space: An extension of the kiefer-wolfowitz procedure. *Journal of Theoretical Probability*, 1(2):189–204, 1988.

[GS91] R. Ghanem and P. Spanos. *Stochastic Finite Elements*. Dover publications, 1991.

[GS03] R.G. Ghanem and P.D. Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.

[GS14] E. Gobet and K. Surana. A new sequential algorithm for $l2$-approximation and application to Monte-Carlo integration. *Available at* https://hal.archives-ouvertes.fr/hal-00972016, 2014.

[GT83] D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of second order*. Springer Verlag, second edition, 1983.

[GT01] E. Gobet and E. Teman. Discrete time hedging errors for options with irregular payoffs. *Finance and Stochastics*, 5(3):357–367, 2001.

[GT09] S. Geiss and A. Toivola. Weak convergence of error processes in discretizations of stochastic integrals and Besov spaces. *Bernoulli*, 15(4):925–954, 2009.

[GW02] J. Grammig and M. Wellner. Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics*, 106:369–400, 2002.

[HH80] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc., 1980.

[HHJ15] M. Hairer, M. Hutzenthaler, and A. Jentzen. Loss of regularity for Kolmogorov equations. *Annals of Probability*, 43(2):468–527, 2015.

[HJ90] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, 1990. Reprinted with corrections.

[HJY11] T. Hayashi, J. Jacod, and N. Yoshida. Irregular sampling and central limit theorems for power variations: The continuous case. *Annales de l'Institut Henri Poincaré, Probability and Statistics*, 47(4):1197–1218, 2011.

[HM05] T. Hayashi and P. A. Mykland. Evaluating hedging errors: an asymptotic approach. *Math. Finance*, 15(2):309–343, 2005.

[HMGR01] N. Hoffman, T. Müller-Gronbach, and K. Ritter. The Optimal Discretization of Stochastic Differential Equations. *Journal of Complexity*, 17:117–153, 2001.

[Hof97] M. Hoffmann. Minimax estimation of the diffusion coefficient through irregular samplings. *Statistics and Probability Letters*, 32:11–24, 1997.

[HST01] H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. *Statistics and Computing*, 7:223–242, 2001.

[HY08] T. Hayashi and N. Yoshida. Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Annals of the Institute of Statistical Mathematics*, 60:357–396, 2008.

[Jac07] J. Jacod. Statistics and high frequency data. *Unpublished CIMPA lecture notes,* http://archive.schools.cimpa.info/archivesecoles/20140206104549/2009-j-semstat.pdf, 2007.

[JMY16] A. Jentzen, T. Müller-Gronbach, and L. Yaroslavtseva. On stochastic differential equations with arbitrary slow convergence rates for strong approximation. *Communications in Mathematical Sciences*, 14(6):1477–1500, 2016.

[JP98] J. Jacod and Ph. Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability*, 26(1):267–307, 1998.

[JP12] J. Jacod and Ph. Protter. Discretization of processes. *Stochastic Modelling and Applied Probability, 67. Springer-Verlag*, 2012.

[JS02] J. Jacod and A.N. Shiryaev. *Limit theorems for stochastic processes.* Springer Verlag, Berlin, second edition, 2002.

[KB09] A.A. Kulkarni and V.S. Borkar. Finite dimensional approximation and Newton-based algorithm for stochastic approximation in Hilbert space. *Automatica*, 45(12):2815–2822, 2009.

[KH92] M. Kleiber and T.D. Hien. *The Stochastic Finite Element Method: basic perturbation technique and computer implementation.* John Wiley & Sons Ltd, 1992.

[KP91] T.G. Kurtz and P. Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, 19(3):1035–1070, 1991.

[KP08] S. Kinnebrock and M. Podolskij. A note on the central limit theorem for bipower variation of general functions. *Stochastic Processes and their Applications*, 118(6):1056–1070, 2008.

[KS86] P. Krée and C. Soize. *Mathematics of random phenomena.* D. Reidel Publishing Co., Dordrecht, 1986.

[KS12] P. Kree and C. Soize. Mathematics of random phenomena: random vibrations of mechanical structures. *Springer Science & Business Media*, 2012.

[KW52] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[KY97a] H. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*, volume 35 of *Application of Mathematics.* Springer, 1997.

[KY97b] H. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*, volume 35 of *Application of Mathematics.* Springer, 1997.

[LBM86] W.K. Liu, T. Belytschko, and A. Mani. Random field finite elements. *International journal for numerical methods in engineering*, 23(10):1831–1845, 1986.

[LK10] O. Le Maitre and O. Knio. *Spectral Methods for Uncertainty Quantification. With Applications to Computational Fluid Dynamics, Scientific Computation.* Springer Science & Business Media, 2010.

[LMR+14] Y. Li, P.A. Mykland, E. Renaul, L. Zhang, and X. Zheng. Realized Volatility When Sampling Times are Possibly Endogenous. *Econometric Theory*, 30(3):580–605, 2014.

[LPC11] J. Li, Y.-B. Peng, and J.-B. Chen. Nonlinear stochastic optimal control strategy of hysteretic structures. *Structural Engineering and Mechanics, 38(1):39–63*, 2011.

[LR13] C. Lindberg and H. Rootzén. Error distributions for random grid approximations of multidimensional stochastic integrals. *Ann. Appl. Probab.*, 23:834–857, 2013.

[LZZ13] Y. Li, Z. Zhang, and X. Zheng. Volatility inference in the presence of both endogenous time and microstructure noise. *Stochastic Processes and their Applications*, 123(7):2696–2727, 2013.

[MG02] T. Müller-Gronbach. Strong Approximation of Systems of Stochastic Differential Equations. Habilitation thesis. *Technical University of Darmstadt. Available on https://www.researchgate.net/publication/34202229_Strong_approximation_of_systems_of_stochastic_differential_equations*, 2002.

[MG04] T. Müller-Gronbach. Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. *Annals of Applied Probability*, 14(4):1605–1642, 2004.

[Mus14] J. Muscat. *Functional Analysis. An Introduction to Metric Spaces, Hilbert Spaces, and Banach Algebras*. Springer, 2014.

[MZ06] P.A. Mykland and L. Zhang. ANOVA for diffusions and Itô processes. *Annals of Statistics*, 34:1931–1963, 2006.

[Nev86] P. Nevai. Géza Freud, orthogonal polynomials and Christoffel functions. a case study. *Journal of Approximation Theory*, 48(1):3–167, 1986.

[Nix84] R. Nixdorf. An invariance principle for a finite dimensional stochastic approximation method in a Hilbert space. *Journal of Multivariate Analysis*, 15(2):252–260, 1984.

[Nor98] J.R. Norris. *Markov chains*. Cambridge university press, 1998.

[PJ92] B.T. Polyak and A.B. Juditsky. Acceleration of stochastic approximation by averaging. . *SIAM J. Control and Optimization*, 30:838–855, 1992.

[Pro04] P. Protter. *Stochastic integration and differential equations*. Springer Verlag, second edition, 2004.

[PS04] S. R. Pliska and K. Suzuki. Optimal tracking for asset allocation with fixed and proportional transaction costs. *Quantitative Finance*, 4:233–243, 2004.

[RM51] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[Roo80] H. Rootzen. Limit distributions for the error in approximations of stochastic integrals. *The Annals of Probability*, 1980.

[RR10] C.Y. Robert and M. Rosenbaum. On the Microstructural Hedging Error. *SIAM J. Financial Math.*, 1:427–453, 2010.

[RR12] C.Y. Robert and M. Rosenbaum. Volatility and covariation estimation when microstructure noise and trading times are endogenous. *Mathematical Finance*, 22(1):133–164, 2012.

[RS71] H. Robbins and D. Siegmund. A convergence theorem for nonnegative almost supermartingales and some applications. In J.S. Rustagi, editor, *Optimizing Methods in Statistics*, pages 233–257. Academic Press, New-York, 1971.

[RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion (Third ed.)*. Berlin: Springer. ISBN 3-540-64325-7, 1999.

[Smi14] R.C. Smith. *Uncertainty quantification. Theory, implementation, and applications*, volume 12 of *Computational Science & Engineering*. Springer, Philadelphia, PA, 2014.

[Sør04] H. Sørensen. Parametric inference for diffusion processes observed at discrete points in time: a survey. *Internat. Statist. Rev.*, 72(3):337–354, 2004.

[SSS08] S. Shalev-Shwartz and N. Srebro. SVM optimization: inverse dependence on training set size. In *Proceedings of the 25th international conference on Machine learning*, pages 928–935. ACM, 2008.

[Str10] R.L. Streit. *Poisson Point Processes. Imaging, Tracking, and Sensing*. Springer Science+Business Media, LLC, 2010.

[Tal02] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields*, 8(2):163–198, 2002.

[Wal77] H. Walk. An invariance principle for the Robbins-Monro process in a Hilbert space. *Zeitschrift fur Warscheinlichkeitstheorie und Verwandte Gebiete*, 39(2):135–150, 1977.

[Wie38] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.

[Wih09] T.P. Wihler. On the Hölder continuity of matrix functions for normal matrices. *Journal of Inequalities in Pure and Applied Mathematics*, 10:1–5, 2009.

[Yin92] G. Yin. On H-valued stochastic approximation: Finite dimensional projections. *Stochastic Analysis and Applications*, 10(3):363–377, 1992.

[YZ90] G. Yin and Y.M. Zhu. On H-valued Robbins-Monro processes. *Journal of Multivariate Analysis*, 34(1):116–140, 1990.

[ZS16] L. Zhou and Z. Su. Discretization error of irregular sampling approximations of stochastic integrals. *Applied Mathematics-A Journal of Chinese Universities*, 31:296–306, 2016.

université PARIS-SACLAY | École doctorale de mathématiques Hadamard (EDMH)

Titre : Discrétisation de processus à des temps d'arrêt et Quantification d'incertitudes pour des limites d'approximation stochastique

Mots clés : processus stochastiques, discretisation, optimisation, temps d'arret, quantification d'incertitudes, algorithmes stochastiques

Résumé :
Cette thèse est constituée de deux parties. La première partie est consacrée aux problèmes de discrétisation des semimartingales browniennes le long des temps d'arrêt.
On commence par l'étude de la discrétisation optimale des intégrales stochastiques. Dans ce cadre on établit une borne inférieure trajectorielle pour la variation quadratique renormalisée de l'erreur et on fournit une suite de temps d'arrêt asymptotiquement optimale, sans hypothèse de non dégénérescence. De plus on établit une stratégie de discrétisation optimale qui est complètement adaptative au modèle.
Ensuite on étudie les fluctuations statistiques des erreurs de discrétisation en démontrant des Théorèmes Centraux Limites pour des suites de temps d'arrêt généraux. La classe de grilles de discrétisation est assez large et la distribution limite est décrite explicitement. Les résultats sont démontrés dans le cas multidimensionnel pour le processus et pour l'erreur de discrétisation. On applique ces résultats au problème d'inférence statistique paramétrique pour les processus de diffusion sur la base d'observations à temps d'arrêt généraux.
La deuxième partie est consacrée au problème de quantification d'incertitudes pour les limites d'algorithme stochastique. Dans le cadre d'étude la limite est définie comme le zéro d'une fonction donnée par une espérance et typiquement représente la solution d'un problème d'optimisation stochastique. L'espérance est relative à une variable aléatoire dont la loi dépend d'un paramètre incertain (point de vue Bayésien). Ainsi la limite d'algorithme dépend de ce paramètre et est donc incertaine. On introduit un algorithme appelé USA (Uncertainty for Stochastic Approximation) pour calculer de manière efficace les coefficients de développement en chaos de la limite comme fonction du paramètre incertain. La convergence et le taux de convergence dans L^2 de USA sont analysés.

Title : Discretization of processes at stopping times and Uncertainty quantification for stochastic approximation limits

Keywords : stochastic processes, discretization, optimization, stopping times, uncertainty quantification, stochastic algorithms

Abstract :
This thesis consists of two parts.
The first part is devoted to the problems of Brownian semimartingales discretization based on stopping times.
We start with the study of the optimal discretization for stochastic integrals. In this context we establish an almost sure lower bound on the renormalized quadratic variation of the error and provide a sequence of stopping times which are asymptotically optimal, without the non-degeneracy assumption. Also we establish an optimal discretization strategy which is completely adaptive to the model.
Further we study statistical fluctuations of the discretization errors by showing Central Limit Theorems for general sequences of stopping times. The class of discretization grids is quite large and the limit distribution is given explicitly. The results are proved in the multidimensional case both for the process and for the discretization error. We apply these results to the problem of parametric statistical inference for diffusion processes based on observations at general stopping times.
The second part is devoted to the problem of uncertainty quantification for stochastic approximation limits. In our framework the limit is defined as the zero of a function given by an expectation and typically represents the solution to some stochastic optimization problem. The expectation is related to a random variable for which the distribution is supposed to depend on an uncertain parameter (Bayesian point of view). Thereby the limit of the algorithm depends on this parameter and is also uncertain. We introduce an algorithm called USA (Uncertainty for Stochastic Approximation) to efficiently calculate the chaos expansion coefficients of the limit as a function of the uncertain parameter. The convergence and the L^2-convergence rate of the USA are analysed.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France