Electronic Journal of Statistics
Vol. 14 (2020) 2098–2122
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1708

Parametric inference for diffusions observed at stopping times

Emmanuel Gobet* **and Uladzislau Stazhynski**

CMAP, Ecole Polytechnique and CNRS, Route de Saclay, 91128 Palaiseau cedex, France.
e-mail: emmanuel.gobet@polytechnique.edu
e-mail: uladzislau.stazhynski@polytechnique.edu

Abstract: In this paper we study the problem of parametric inference for multidimensional diffusions based on observations at random stopping times. We work in the asymptotic framework of high frequency data over a fixed horizon. Previous works on the subject (such as [10, 17, 19, 5] among others) consider only deterministic, strongly predictable or random, independent of the process, observation times, and do not cover our setting. Under mild assumptions we construct a consistent sequence of estimators, for a large class of stopping time observation grids (studied in [20, 23]). Further we carry out the asymptotic analysis of the estimation error and establish a Central Limit Theorem (CLT) with a mixed Gaussian limit. In addition, in the case of a 1-dimensional parameter, for any sequence of estimators verifying CLT conditions without bias, we prove a uniform a.s. lower bound on the asymptotic variance, and show that this bound is sharp.

MSC 2010 subject classifications: 62Mxx, 62Fxx, 60F05, 60G40, 60Gxx, 62F12.
Keywords and phrases: Diffusion coefficient estimation, observation at stopping times, consistent sequence of estimators, local asymptotic mixed normality, asymptotic variance, optimal lower bound.

Received October 2018.

1. Introduction

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

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma(s, S_s, \xi)\mathrm{d}B_s, \qquad t \in [0,T], \quad S_0 \in \mathbb{R}^d, \tag{1.1}$$

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 (1.1) 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

*This author research is part of the Chaire Risque Financiers of the Fondation du Risque and of the Finance for Energy Market Research Centre (FiME lab) of Institut Europlace de Finance.

important in practice (see the examples in [25, 14] for instance) and it has a large impact on inference procedure, as it is argued in [4].

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 [25] for empirical evidence about the connection of volatility and inter-transaction duration in finance, and [14] 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 [21] for additional motivation and discussion. To the best of our knowledge this setting has not yet been studied in the literature, except in [30] 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 [21] in a concrete setting (i.e. for explicitly defined class of grids, and not given by abstract assumptions, as a difference with [30]), in several dimensions (as a difference with above references) and with a tractable limit characterization. Note that in [15], 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 MLE 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 [21] 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 [31, 13] and the lecture notes [27].

The nonparametric estimation of the diffusion coefficient $\sigma(.)$ is investigated in [12] for equidistant observations times on a fixed time interval. In [17] 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. [8, Chapter 5] for the definition: this LAMN property is established in [10] for one-dimensional S, and in [19] 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, [1, 2, 3, 9]. Note that these works consider only the case of deterministic observation grids.

In [18] 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 [18] 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 [4, 11] 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 [26], and it can take the form of *strongly predictable* times (τ_i^n is known at time τ_{i-1}^n). In [5], the time increments are simply independent and identically distributed. In [14, 16], 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, [30] 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 [23, Remark 1], contains most of the examples, interesting in practice.
- Using our asymptotic analysis and applying the results of [21] 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 [15, 20, 23].

Outline of the paper. In Section 2 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 3 is devoted to the statements of the main results of the paper. We continue with the proofs in Section 4. Several technical points are postponed to Section A.

2. The model

Let $(B_t)_{0\leq t\leq T}$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ 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 $(S_t)_{0\le t\le T}$ of the form

$$S_t = S_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma(s, S_s, \xi^\star)\mathrm{d}B_s, \qquad t\in[0,T], \quad S_0 \in \mathbb{R}^d, \qquad (2.1)$$

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. $(b_t)_{0\le t\le T}$ 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 $0\le s\le t\le T$ we have $|b_t - b_s| \le C|t-s|^{\eta_b}$.

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 $[0,T]$.

2.1. Random observation grids

We consider a sequence of random observation grids

$$\{(\tau_0^n := 0 < \tau_1^n < \cdots < \tau_i^n < \cdots < \tau_{N_T^n}^n := T) : n \ge 0\}$$

on the interval $[0,T]$ and suppose that for each n, only the values $(\tau_i^n, S_{\tau_i^n})_{0\le i\le N_T^n}$ are available for the parameter estimation: these are the observation data. For each n, $(\tau_i^n : 0\le i\le N_T^n)$ 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 [10, 19]) given the sequence of observations $\{(\tau_i^n, S_{\tau_i^n})_{0\le i\le N_T^n} : n\ge 0\}$ is so far out of reach, and we leave these problems for further investigation. However we establish a partial optimality result in Section 3.4.

Our statistics analysis is based on the asymptotic techniques, developed recently in [20, 23, 24], 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 $(\varepsilon_n)_{n\ge0}$ with $\varepsilon_n\to0$ and a parameter $\rho_N\ge1$ (compare to [23, Definition 1]):

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

$$\sup_{n\ge0}\left(\varepsilon_n^{-2}\sup_{1\le i\le N_T^n}\sup_{t\in(\tau_{i-1}^n,\tau_i^n]}|S_t - S_{\tau_{i-1}^n}|^2\right) < +\infty. \qquad (2.2)$$

$(\mathbf{A}_N)$**:** For some $\rho_N \in [1, (1+2\eta_b) \wedge 4/3)$ the following non-negative random variable is a.s. finite:

$$\sup_{n \geq 0}(\varepsilon_n^{2\rho_N} N_T^n) < +\infty. \tag{2.3}$$

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

$$(\varepsilon_{\iota(n)}, \{\tau_0^{\iota(n)} := 0 < \tau_1^{\iota(n)} < \cdots < \tau_i^{\iota(n)} < \cdots < \tau_{N_T^{\iota(n)}}^{\iota(n)} := T\})_{n\geq 0}$$

of

$$(\varepsilon_n, \{\tau_0^n := 0 < \tau_1^n < \cdots < \tau_i^n < \cdots < \tau_{N_T^n}^n := T\})_{n\geq 0}$$

there exists another subsequence

$$(\varepsilon_{\iota\circ\iota'(n)}, \{\tau_0^{\iota\circ\iota'(n)} := 0 < \tau_1^{\iota\circ\iota'(n)} < \cdots < \tau_i^{\iota\circ\iota'(n)} < \cdots < \tau_{N_T^{\iota\circ\iota'(n)}}^{\iota\circ\iota'(n)} := T\})_{n\geq 0}$$

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 [23, Remark 1].
- The sequences of grids based on exit times from general random domains and, possibly, extra independent noise. Namely let $\{(D_t^n)_{0\leq t\leq T} : n \geq 0\}$ 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 [21, Section 2.2]) to an adapted continuous domain-valued process $(D_t)_{0\leq t\leq T}$. 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 $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ with $\mathcal{T}^n = \{\tau_i^n, i = 1,\ldots,N_T^n\}$ given by

$$\begin{cases} \tau_0^n & := 0, \\ \tau_i^n & := \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n}) \notin \varepsilon_n D_{\tau_{i-1}^n}^n\} \\ & \qquad \wedge(\tau_{i-1}^n + \varepsilon_n^2 G_{\tau_{i-1}^n}(U_{n,i}) + \Delta_{n,i}) \wedge T, \end{cases} \tag{2.4}$$

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 [21, Section 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 [21, Section 2.2.3]). In addition, we can rely on a

CLT for a general discretization error term based on such grids (see [21, Theorem 2.4]). The optimal observation grid in Section 3.4 is of the above form, taking some ellipsoid for D^n and $G(\cdot) = +\infty$, $\Delta_{n,i} = 0$.

Subsequence formulation of the assumption ($\mathbf{H}_{\mathcal{T}}$) is motivated by the following subsequence principle:

Lemma 2.1 ([7, Theorem 20.5]). *Consider real-valued random variables. Then* $\mathcal{X}_n \underset{n\to+\infty}{\overset{\mathbb{P}}{\to}} \mathcal{X}$ *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* $\mathcal{X}_{\iota\circ\iota'(n)} \underset{n\to+\infty}{\overset{a.s.}{\to}} \mathcal{X}$.

It allows to first prove a.s. results for the sequences of observation grids verifying ($\mathbf{A}_S^{\mathrm{osc.}}$)-($\mathbf{A}_N$) and $\sum_{n\geq 0}\varepsilon_n^2 < +\infty$, and then pass to the equivalent results in probability in the general case.

2.2. Sequence of estimators

Suppose that $\mathcal{T} := \{\mathcal{T}^n : n \geq 0\}$ 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)

$$\varphi(t) := \max\{\tau \in \mathcal{T}^n : \tau \leq t\}, \quad \Delta H_t := H_t - H_{\varphi(t)}. \tag{2.5}$$

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 [17]. The optimality of the Gaussian based likelihood approximations in the case of regular observation times has been proved in [10, 19]. 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

$$D_{\mathrm{KL}}(\Sigma_1,\Sigma_2) := \int_{\mathbb{R}^d} p_{\Sigma_1}(x)\log\frac{p_{\Sigma_1}(x)}{p_{\Sigma_2}(x)}\mathrm{d}x. \tag{2.6}$$

For some continuous weight function $\omega : [0,T]\times\mathbb{R}^d \to]0,+\infty[$ set $\omega_t := \omega(t, S_t)$; the process $(\omega_t)_{0\leq t\leq T}$ 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 $\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi))\omega_t dt$ 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

$$D_{\mathrm{KL}}(\Sigma_1, \Sigma_2) = \frac{1}{2}\int_{\mathbb{R}^d} \left(\log(\det \Sigma_2) - \log(\det \Sigma_1) + x^\top \Sigma_2^{-1} x - x^\top \Sigma_1^{-1} x\right) p_{\Sigma_1}(x)\mathrm{d}x,$$

and thus

$$\int_0^T D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi))\omega_t \mathrm{d}t = \frac{1}{2} U^\star(\xi) + C_0, \tag{2.7}$$

where C_0 is independent of ξ and

$$\begin{aligned} U^\star(\xi) &:= \int_0^T \int_{\mathbb{R}^d} \left(\log(\det c_t(\xi)) + x^\top c_t^{-1}(\xi) x\right) p_{c_t(\xi^\star)}(x)\omega_t \mathrm{d}x\mathrm{d}t \\ &= \int_0^T \left(\log(\det c_t(\xi)) + \mathrm{Tr}(\sigma_t(\xi^\star)^\top c_t^{-1}(\xi)\sigma_t(\xi^\star))\right)\omega_t \mathrm{d}t. \end{aligned} \tag{2.8}$$

Remark that $\int_0^T \mathrm{Tr}(\sigma_t(\xi^\star)^\top c_t^{-1}(\xi)\sigma_t(\xi^\star))\omega_t \mathrm{d}t$ represents a quadratic variation. Thus we define the following discretized version of $U^\star(\cdot)$, that uses only $(\tau_i^n, S_{\tau_i^n} : 0 \le i \le N_T^n)$,

$$\begin{aligned} U^n(\xi) := &\sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \log\left(\det c_{\tau_{i-1}^n}(\xi)\right)(\tau_i^n - \tau_{i-1}^n) \\ &+ \sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \Delta S_{\tau_i^n}^\top c_{\tau_{i-1}^n}^{-1}(\xi) \Delta S_{\tau_i^n}. \end{aligned} \tag{2.9}$$

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 (2.9) coincides with [17, eq. (3)].

Define the sequence of estimators $(\xi^n)_{n\ge 0}$ as follows:

$$\xi^n := \mathrm{Argmin}_{\xi \in \Xi}\, U^n(\xi) \tag{2.10}$$

(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 $(\tau_i^n, S_{\tau_i^n}, 0 \le i \le N_T^n)$ 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.

3. Main results

For the subsequent convergences, we adopt the following natural notations. By $O_n^{\mathrm{a.s.}}(1)$ (resp. $o_n^{\mathrm{a.s.}}(1)$) 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.

3.1. 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 4.1.

Theorem 3.1. *Assume ($\mathbf{H}_{\mathcal{S}}$), ($\mathbf{H}_{\xi}$) and ($\mathbf{H}_{\mathcal{T}}$). Then for the sequence estimators $(\xi^n)_{n\geq 0}$ given by (2.10) we have the following convergence in probability*

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

3.2. 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 (2.6) 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

$$\mathcal{H}_T := 2\int_0^T \left(\nabla_\xi^2 D_{\mathrm{KL}}(c_t(\xi^\star), c_t(\xi)) \mid_{\xi=\xi^\star}\right) \omega_t \mathrm{d}t = \nabla_\xi^2 U^\star(\xi^\star) \qquad (3.1)$$

is positive definite, and where the second equality follows from (2.7) (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 $\mathcal{A} =$

$[\mathcal{A}^1, \cdots, \mathcal{A}^m]^\intercal$ and $x, y \in \mathbb{R}^d$ we denote $x^\intercal \mathcal{A} y := [x^\intercal \mathcal{A}^1 y, \cdots, x^\intercal \mathcal{A}^m y]^\intercal \in \mathbb{R}^m$. By $x^\intercal \mathcal{A}$ we denote the linear operator in $\mathrm{Mat}_{m,d}$ corresponding to $y \mapsto x^\intercal \mathcal{A} y$ (similarly for $x \mapsto x^\intercal \mathcal{M} y$). 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:

$$\mathcal{M}_t := 2\omega_t b_t^\intercal \nabla_\xi c_t^{-1}(\xi^\star) + \bar{\mathcal{M}}_t, \quad \mathcal{A}_t := 2\omega_t \nabla_\xi c_t^{-1}(\xi^\star) \sigma_t(\xi^\star), \quad t \in [0, T] \tag{3.2}$$

where for $1 \le i \le m, 1 \le j \le d$ we define

$$\bar{\mathcal{M}}_t^{ij} := 2\omega_t \operatorname{Tr}(\sigma_t(\xi^\star)^\intercal \nabla_{\xi_i} c_t^{-1}(\xi^\star) \nabla_{x_j} \sigma_t(\xi^\star)). \tag{3.3}$$

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 3.2. *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*

$$\varepsilon_n^{-\rho_N}(\xi^n - \xi^\star) = -(\mathcal{H}_T^{-1} + o_n^{\mathbb{P}}(1)) \varepsilon_n^{-\rho_N} Z_T^n + o_n^{\mathbb{P}}(1), \tag{3.4}$$

where

$$Z_s^n := \int_0^s \Delta S_t^\intercal \mathcal{A}_{\varphi(t)} \mathrm{d}B_t + \int_0^s \mathcal{M}_{\varphi(t)} \Delta S_t \mathrm{d}t := M_s^n + A_s^n \tag{3.5}$$

for $\mathcal{M}_t$ *and* $\mathcal{A}_t$ *defined in* (3.2).

The proof is done in Section 4.2.

3.3. CLT in the case of ellipoid exit times

We start with the following lemma, that plays an important role in the sequel:

Lemma 3.3 ([20, Lemma 3.1]). *Let* y *be a* $d \times d$*-matrix symmetric non-negative real matrix. Then the equation*

$$2 \operatorname{Tr}(x) x + 4x^2 = y^2 \tag{3.6}$$

admits exactly one solution $x(y) \in \mathcal{S}_d^+$.

Theorem 3.2 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 (3.4) we get

$$\sqrt{N_T^n}(\xi^n - \xi^\star) = -(\mathcal{H}_T^{-1} + o_n^{\mathbb{P}}(1)) \sqrt{N_T^n} Z_T^n + o_n^{\mathbb{P}}(\sqrt{N_T^n} \varepsilon_n^{\rho_N}) \tag{3.7}$$

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 2.1). This makes possible the direct application of general results on CLT for discretization errors of the form (3.5); we refer to [21] 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 [21, Theorem 2.4] where the CLT for discretization errors of the form (3.5) 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 (3.8)). This example is, in particular, used in Section 3.4.

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 [21, Example 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

$$\tau_0^n = 0, \quad \tau_i^n = \inf\{t > \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T} \Sigma_{\tau_{i-1}^n}^n (S_t - S_{\tau_{i-1}^n}) \ge \varepsilon_n^2\} \wedge T. \quad (3.8)$$

Such a sequence verifies $(\mathbf{H}_{\mathcal{T}})$ with $\rho_N = 1$ (which follows from [21, Theorem 2.4], see the proof of Theorem 3.4).

To simplify we note $\sigma_t := \sigma_t(\xi^\star)$ till the end of this section. We consider the setting of [21, Section 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 [21], define, for any $t \in [0,T]$ and any measurable function $f : \mathbb{R}^d \mapsto \mathbb{R}$,

$$\tau(t) := \inf\{s \ge 0 : \sigma_t W_s \notin D_t\}, \qquad \mathcal{B}_t[f(\cdot)] := \mathbb{E}_t\left(f(\sigma_t W_{\tau(t)})\right), \quad (3.9)$$

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 [21, Lemma 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

$$\mathcal{K}_t^{ij} := m_t^{-1}\mathcal{B}_t\left[f(x) := ((\sigma_t^{-1}x)^\mathsf{T} X_t^{ij+}(\sigma_t^{-1}x))^2 - ((\sigma_t^{-1}x)^\mathsf{T} X_t^{ij-}(\sigma_t^{-1}x))^2\right], \quad (3.10)$$

for all $1 \leq i, j \leq m$, where X_t^{ij+} (resp. X_t^{ij-}) is the solution of the matrix equation (3.6) for $c = \sigma_t^\intercal \mathcal{A}_t^{ij+}\sigma_t$ (resp. $\sigma_t^\intercal \mathcal{A}_t^{ij-}\sigma_t$). Remark that the process $(Q_t)_{0\leq t\leq T}$ defined in [21, eq. 2.18] is equal to 0 in our case since the domains D_t and D_t^n are symmetric, see [21, Section 2.4]. Also note that the matrix equation (3.6) may be easily solved numerically, see the details in [20, 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. [22]. The following result is an application of [21, Theorem 2.4 and its proof].

Theorem 3.4. *The process $(\mathcal{K}_t)_{0\leq t\leq T}$ 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 $(\xi^n)_{n\geq 0}$ given by (2.10) we have*

$$\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{\mathcal{L}} \mathcal{H}_T^{-1}\sqrt{\int_0^T m_t^{-1}\mathrm{d}t}\int_0^T \mathcal{K}_t^{1/2}\mathrm{d}\widetilde{W}_t, \tag{3.11}$$

where $\mathcal{H}_T$ is defined in (3.1). More specifically, for Z^n, M^n, A^n defined in (3.5), we have the convergences

$$\begin{aligned} &\varepsilon_n^{-2}\langle Z^n\rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^s \mathcal{K}_t\mathrm{d}t, \text{ for all } s \in [0, T], \\ &\varepsilon_n^{-1}\langle Z^n, B\rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \text{ for all } s \in [0, T], \\ &\varepsilon_n^{-1}\sup_{s\in[0,T]}|A_s^n| \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \\ &\varepsilon_n^2 N_T^n \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^T m_t^{-1}\mathrm{d}t. \end{aligned} \tag{3.12}$$

Proof. Our goal is to check the assumptions of [21, Theorem 2.4]. First note that all random variables σ_0, σ_0^{-1}, $\mathcal{M}_0$ and $\mathcal{A}_0$ are bounded under our setting. Condition [21, $(\mathbf{H}_S)$] follows from $(\mathbf{H}_S)$ and $(\mathbf{H}_\Sigma)$-4. Further [21, $(\mathbf{H}_\Delta)$] follows from $(\mathbf{H}_\Sigma)$-2.

Conditions [21, $(\mathbf{H}_D^1)$, $(\mathbf{H}_D^2)$] 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 [21, Section 2.2.1]) that $\mu(D_t, D_t^n) \leq 2|\Sigma_t^{-1/2} - (\Sigma_t^n)^{-1/2}|$. 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 [21, ($\mathbf{H}_G$)] is trivial in this case since the function $G(\cdot)$ equals $+\infty$ and $\Delta_{n,i} = 0$ (in the notation of [21]). Other assumptions of [21, Theorem 2.4] follow from ($\mathbf{H}_\Sigma$)-3. Last, the sign in front of $\mathcal{H}_T^{-1}$ in (3.7) does not change the distribution-limit which is symmetric (as that of $\widetilde{W}$). □

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

$$V_T := \left(\int_0^T m_t^{-1}\mathrm{d}t\right)\mathcal{H}_T^{-1}\left(\int_0^T \mathcal{K}_t\mathrm{d}t\right)\mathcal{H}_T^{-1}.$$

3.4. 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 (3.6) 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 (3.1) define

$$V_T^{opt.} := \mathcal{H}_T^{-2}\left(\int_0^T 2\omega_t \,\mathrm{Tr}(X_t(\xi^\star))\mathrm{d}t\right)^2, \tag{3.13}$$

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 [17]; or the hitting times of symmetric boundaries, see [21, Section 2.4] and Theorem 3.4) we prove that $V_T^{opt.}$ is a uniform lower bound on the asymptotic variance of the sequence of estimators (2.10). 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 1. *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 $\mathcal{T}_\kappa$,*

$$\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_T^\kappa),$$

and the limit variance V_T^κ associated with $\mathcal{T}_\kappa$ verifies the condition

$$0 \leq V_T^\kappa - V_T^{opt.} \leq C_0\kappa, \quad \forall \kappa \in (0, \kappa_0].$$

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 $\mathbf{1}_{(-\infty,1/2]} \leq \chi(.) \leq \mathbf{1}_{(-\infty,1]}$, and denote $\chi_\kappa(x) := \chi(x/\kappa)$. 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 3.2 we have the decomposition (3.4), with Z^n given by (3.5). In view of (3.4), 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 1.

Theorem 3.5. *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* (2.10). *For some* $\rho \in [1, \rho_N]$ *suppose that the semimartingale decomposition* $Z_t^n := M_t^n + A_t^n$ *in* (3.5) *verifies*

$$\begin{aligned}
&\varepsilon_n^{-2\rho}\langle M^n\rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} \int_0^s \mathcal{K}_t \mathrm{d}t, \ \textit{for all } s \in [0,T],\\
&\varepsilon_n^{-\rho}\langle M^n, B\rangle_s \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \ \textit{for all } s \in [0,T],\\
&\varepsilon_n^{-\rho} \sup_{0\leq t\leq T} |A_t^n| \xrightarrow[n\to+\infty]{\mathbb{P}} 0
\end{aligned} \tag{3.14}$$

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) $\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_T)$ *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* (3.13).

(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*

$$\begin{cases}\tau_0^n := 0,\\ \tau_i^n = \inf\left\{t \geq \tau_{i-1}^n : (S_t - S_{\tau_{i-1}^n})^\mathsf{T}\Lambda_{\tau_{i-1}^n}^\kappa(\xi^\star)(S_t - S_{\tau_{i-1}^n}) > \varepsilon_n^2\right\} \wedge T\end{cases} \tag{3.15}$$

is κ*-optimal for* $\kappa_0 = 1$ *in the sense of Definition 1.*

We remark that the class of discretization grids over which the universal variance lower bound is obtained in Theorem 3.5 includes most of the examples for which a CLT has been established, since the conditions of the type (3.14) are quite commonly required (see [29, 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 [21, Section 2.2] we have $\rho = \rho_N = 1$. See also the discussion in Section 2.1 and [23, Remark 1].

As we may notice the κ-optimal sequence of discretization grids in (3.15) depends on the unknown parameter $\xi^\star$. Besides, concerning the optimal variance $V_T^{opt.}$ in (3.13), it also involves $\xi^\star$, as well as ω_t: we argue in Section 2.2 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 similar methodology has been designed and analyzed in [24]. 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 [24, Section 3.1].

4. Proofs of the main results

The next lemma provides some important properties of the process $\sigma_t(\cdot)$.

Lemma 4.1. *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 3.2 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 [6, 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

$$\begin{aligned}\sigma_t(\xi^\star) - \sigma_{\varphi(t)}(\xi^\star) &= \sigma(t, S_t, \xi^\star) - \sigma(\varphi(t), S_{\varphi(t)}, \xi^\star)\\ &= \sigma(\varphi(t), S_t, \xi^\star) - \sigma(\varphi(t), S_{\varphi(t)}, \xi^\star) + O_n^{\text{a.s.}}(|\Delta t|)\end{aligned}$$

$$= \sum_{i=1}^{d} \nabla_{x_i} \sigma(\varphi(t), S_{\varphi(t)}, \xi^\star) \Delta S_t^i + O_n^{\text{a.s.}}(|\Delta t| + |\Delta S_t|^2).$$

From ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) and [23, Lemma 3.2] we get $\sup_{t\in[0,T]} \left|O_n^{\text{a.s.}}(|\Delta t|+|\Delta S_t|^2)\right| \leq C_\rho \varepsilon_n^{2-\rho}$ 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 4.2. *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*

$$\sup_{\xi\in\Xi} |U^n(\xi) - U^\star(\xi)| \xrightarrow[n\to+\infty]{\text{a.s.}} 0, \tag{4.1}$$

$$\begin{aligned} &\sup_{\xi\in\Xi} |\nabla_\xi U^n(\xi) - \nabla_\xi U^\star(\xi)| \xrightarrow[n\to+\infty]{\text{a.s.}} 0, \\ &|\nabla_\xi^2 U^n(\xi) - \nabla_\xi^2 U^\star(\xi)| \xrightarrow[n\to+\infty]{\text{a.s.}} 0, \quad \forall \xi \in \Xi. \end{aligned} \tag{4.2}$$

Proof. Using (2.8) and Lemma A.1 we deduce the following expressions for $\nabla_{\xi_k} U^\star(\xi)$ and $\nabla^2_{\xi_k\xi_l} U^\star(\xi)$ ($1 \leq k, l \leq m$):

$$\nabla_{\xi_k} U^\star(\xi) = \int_0^T \operatorname{Tr}\left(\nabla_{\xi_k} c_t(\xi) c_t(\xi)^{-1} + \sigma_t(\xi^\star)^\mathsf{T} \nabla_{\xi_k} c_t^{-1}(\xi) \sigma_t(\xi^\star)\right) \omega_t \mathrm{d}t, \tag{4.3}$$

$$\begin{aligned} \nabla^2_{\xi_k\xi_l} U^\star(\xi) = \int_0^T \operatorname{Tr}\Big(&\nabla^2_{\xi_k\xi_l} c_t(\xi) c_t^{-1}(\xi) + \nabla_{\xi_k} c_t(\xi) \nabla_{\xi_l} c_t^{-1}(\xi) \\ &+ \sigma_t(\xi^\star)^\mathsf{T} \nabla^2_{\xi_k\xi_l} c_t^{-1}(\xi) \sigma_t(\xi^\star)\Big) \omega_t \mathrm{d}t. \end{aligned} \tag{4.4}$$

Recall that

$$\begin{aligned} U^n(\xi) = &\sum_{\tau_{i-1}<T} \omega_{\tau_{i-1}^n} \log(\det c_{\tau_{i-1}^n}(\xi))(\tau_i^n - \tau_{i-1}^n) \\ &+ \sum_{\tau_{i-1}^n<T} \omega_{\tau_{i-1}^n} \Delta S_{\tau_i^n}^\mathsf{T} c_{\tau_{i-1}^n}^{-1}(\xi) \Delta S_{\tau_i^n}. \end{aligned} \tag{4.5}$$

Let us first prove that for any $\xi \in \Xi$

$$U^n(\xi) \xrightarrow[n\to+\infty]{\text{a.s.}} U^\star(\xi). \tag{4.6}$$

The convergence of the first term in the right-hand side of (4.5) follows from the standard Riemann integral approximation, using that $\sup \Delta\tau_i^n \xrightarrow[n\to+\infty]{\text{a.s.}} 0$ by [23, Lemma 3.2], so we get

$$\sum_{\tau_{i-1}<T} \omega_{\tau_{i-1}^n} \log(\det c_{\tau_{i-1}^n}(\xi))(\tau_i^n - \tau_{i-1}^n) \xrightarrow[n\to+\infty]{\text{a.s.}} \int_0^T \log(\det c_t(\xi)) \omega_t \mathrm{d}t. \tag{4.7}$$

For the second term we have by [23, Proposition 3.7]

$$\sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^{\mathsf{T}}_{\tau^n_i} c^{-1}_{\tau^n_{i-1}}(\xi) \Delta S_{\tau^n_i} \xrightarrow[n\to+\infty]{\text{a.s.}} \int_0^T \mathrm{Tr}\left(\sigma_t(\xi^\star)^{\mathsf{T}} c_t^{-1}(\xi)\sigma_t(\xi^\star)\right)\omega_t \mathrm{d}t. \quad (4.8)$$

Hence the convergence (4.6) follows now from taking the sum of (4.7) and (4.8). Further using Lemma A.1 we obtain

$$\begin{aligned}\nabla_{\xi_k} U^n(\xi) = &\sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \mathrm{Tr}\left(\nabla_{\xi_k} c_{\tau^n_{i-1}}(\xi) c^{-1}_{\tau^n_{i-1}}(\xi)\right)(\tau^n_i - \tau^n_{i-1}) \\ &+ \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^{\mathsf{T}}_{\tau^n_i} (\nabla_{\xi_k} c^{-1}_{\tau^n_{i-1}}(\xi)) \Delta S_{\tau^n_i},\end{aligned} \quad (4.9)$$

$$\begin{aligned}\nabla^2_{\xi_k\xi_l} U^n(\xi) = &\sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \mathrm{Tr}\Big(\nabla^2_{\xi_k\xi_l} c_{\tau^n_{i-1}}(\xi) c^{-1}_{\tau^n_{i-1}}(\xi) \\ &\qquad + \nabla_{\xi_k} c_{\tau^n_{i-1}}(\xi) \nabla_{\xi_l} c^{-1}_{\tau^n_{i-1}}(\xi)\Big)(\tau^n_i - \tau^n_{i-1}) \\ &+ \sum_{\tau^n_{i-1}<T} \omega_{\tau^n_{i-1}} \Delta S^{\mathsf{T}}_{\tau^n_i} \nabla^2_{\xi_k\xi_l} c^{-1}_{\tau^n_{i-1}}(\xi) \Delta S_{\tau^n_i}.\end{aligned} \quad (4.10)$$

Using (4.3), (4.4) and applying the same reasoning as for the proof of (4.6) we also show the following convergences for any $\xi \in \Xi$

$$\nabla_\xi U^n(\xi) \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla_\xi U^\star(\xi), \qquad \nabla^2_\xi U^n(\xi) \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla^2_\xi U^\star(\xi). \quad (4.11)$$

Further from (4.9) and (4.10), using ($\mathbf{H}_S$)-1,2, the compactness of Ξ, the continuity of ω_t and the convergence $\sum_{\tau^n_i \le T} |\Delta S_{\tau^n_i}|^2 \xrightarrow[n\to+\infty]{\text{a.s.}} \mathrm{Tr}(\langle S\rangle_T)$ by [23, Proposition 3.7], we have a.s.

$$\begin{aligned}\sup_{n\ge 0}\left(\sup_{\xi\in\Xi} |\nabla_\xi U^n(\xi)|\right) &\le C \sup_{0\le t\le T}\left[\omega_t\left(\sup_{\xi\in\Xi} |\nabla_\xi c_t(\xi) c_t^{-1}(\xi)| + \sup_{\xi\in\Xi} |\nabla_\xi c_t^{-1}(\xi)|\right)\right] \\ &< +\infty,\end{aligned}$$

$$\begin{aligned}\sup_{n\ge 0}\left(\sup_{\xi\in\Xi} |\nabla^2_\xi U^n(\xi)|\right) \le C \sup_{0\le t\le T}\Big[\omega_t\Big(&\sup_{\xi\in\Xi}(|\nabla^2_\xi c_t(\xi)||c_t^{-1}(\xi)|) + \sup_{\xi\in\Xi} |\nabla_\xi c_t(\xi)|^2 \\ &+ \sup_{\xi\in\Xi} |\nabla^2_\xi c_t^{-1}(\xi)|\Big)\Big] < +\infty,\end{aligned}$$

for some a.s. finite $C>0$. This implies that the sequences $(U^n(\cdot))_{n\ge 0}, (\nabla_\xi U^n(\cdot))_{n\ge 0}$ are equicontinuous and hence the convergences in (4.6) and (4.11) are uniform in $\xi \in \Xi$. We are done. □

4.1. Proof of Theorem 3.1

First suppose that $\sum_{n\ge 0} \varepsilon_n^2 < +\infty$ and that the grid sequence $\mathcal{T}$ verifies ($\mathbf{A}^{\text{osc.}}_S$)-($\mathbf{A}_N$).

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 $\xi \neq \xi^\star$ 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 (2.7) we have that a.s.

$$\xi^\star = \mathrm{Argmin}_{\xi\in\Xi}\, U^\star(\xi).$$

Further, Lemma 4.2 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 2.1. □

4.2. Proof of Theorem 3.2

First suppose that $\sum_{n\geq 0} \varepsilon_n^2 < +\infty$ and the grid sequence $\mathcal{T}$ verifies ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$).

Step 1. We start by showing the convergence

$$\int_0^1 \nabla_\xi^2 U^n(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla_\xi^2 U^\star(\xi^\star) =: \mathcal{H}_T. \tag{4.12}$$

Let $1 \leq k, l \leq m$. In view of the convergence $\nabla_\xi^2 U^n(\xi^\star) \xrightarrow[n\to+\infty]{\text{a.s.}} \nabla_\xi^2 U^\star(\xi^\star)$ from Lemma 4.2 it is enough verify that

$$\int_0^1 \nabla_{\xi_k\xi_l}^2 U^n(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u - \nabla_{\xi_k\xi_l}^2 U^n(\xi^\star) \xrightarrow[n\to+\infty]{\text{a.s.}} 0. \tag{4.13}$$

Denote $\gamma_t(\xi) := \mathrm{Tr}(\nabla_{\xi_k\xi_l}^2 c_t(\xi) c_t^{-1}(\xi) + \nabla_{\xi_k} c_t(\xi) \nabla_{\xi_l} c_t^{-1}(\xi))$. Using the representation (4.10) for $\nabla_{\xi_k\xi_l}^2 U^n(\cdot)$, we get that the left-hand side in (4.13) is equal to

$$\sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \left(\int_0^1 \gamma_{\tau_{i-1}^n}(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u - \gamma_{\tau_{i-1}^n}(\xi^\star) \right) (\tau_i^n - \tau_{i-1}^n)$$
$$+ \sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \Delta S_{\tau_i^n}^\mathsf{T} \left(\int_0^1 \nabla_{\xi_k\xi_l}^2 c_{\tau_{i-1}^n}^{-1}(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u - \nabla_{\xi_k\xi_l}^2 c_{\tau_{i-1}^n}^{-1}(\xi^\star) \right) \Delta S_{\tau_i^n}.$$

Now (4.13) follows from the convergence $\xi^n \xrightarrow[n\to+\infty]{\text{a.s.}} \xi^\star$ for $\mathcal{T}$ verifying ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) (see the proof of Theorem 3.1) 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 $U_T^n(\cdot)$ near $\xi^\star$ and invoking Theorem 3.1. 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 (4.12) and $\xi^n \xrightarrow[n\to+\infty]{\text{a.s.}} \xi^\star$, and since $\xi^\star \notin \partial\Xi$ we obtain $\mathbf{1}_{\Omega^n} \xrightarrow[n\to+\infty]{\text{a.s.}} 1$. On Ω^n we have $\nabla_\xi U^n(\xi^n) = 0$, which implies

$$\mathbf{1}_{\Omega^n}(\xi^n - \xi^\star) = -\mathbf{1}_{\Omega^n} \left(\int_0^1 \nabla^2_\xi U^n(\xi^\star + u(\xi^n - \xi^\star))\mathrm{d}u \right)^{-1} \nabla_\xi U^n(\xi^\star)$$

by the Taylor formula. This implies, in view of (4.12) and since $\mathbf{1}_{\Omega\setminus\Omega^n} = 0$ for n large enough, that

$$\varepsilon_n^{-\rho_N}(\xi^n - \xi^\star) = -\left(\mathcal{H}_T^{-1} + o_n^{\text{a.s.}}(1)\right)\varepsilon_n^{-\rho_N}\nabla_\xi U^n(\xi^\star) + o_n^{\text{a.s.}}(1). \tag{4.14}$$

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 (4.9) of $\nabla_\xi U^n(\cdot)$ and applying the Itô formula, we obtain

$$\begin{aligned}
&\nabla_\xi U^n(\xi^\star) \\
&= \sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \operatorname{Tr}\left(\nabla_\xi c_{\tau_{i-1}^n}(\xi^\star) c^{-1}_{\tau_{i-1}^n}(\xi^\star)\right)(\tau_i^n - \tau_{i-1}^n) \\
&\quad + \sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \Delta S_{\tau_i^n}^{\mathsf{T}} \nabla_\xi c^{-1}_{\tau_{i-1}^n}(\xi^\star)\Delta S_{\tau_i^n} \\
&= \sum_{\tau_{i-1}^n < T} \omega_{\tau_{i-1}^n} \operatorname{Tr}\Big(\nabla_\xi c_{\tau_{i-1}^n}(\xi^\star) c^{-1}_{\tau_{i-1}^n}(\xi^\star) \\
&\qquad\qquad + \sigma_{\tau_{i-1}^n}(\xi^\star)^{\mathsf{T}} \nabla_\xi c^{-1}_{\tau_{i-1}^n}(\xi^\star)\sigma_{\tau_{i-1}^n}(\xi^\star)\Big)(\tau_i^n - \tau_{i-1}^n) \\
&\quad + \int_0^T \omega_{\varphi(t)} \operatorname{Tr}\left((\sigma_t(\xi^\star) + \sigma_{\varphi(t)}(\xi^\star))^{\mathsf{T}} \nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)(\sigma_t(\xi^\star) - \sigma_{\varphi(t)}(\xi^\star))\right)\mathrm{d}t \\
&\quad + 2\int_0^T \omega_{\varphi(t)} \Delta S_t^{\mathsf{T}} \nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star) b_t \mathrm{d}t + 2\int_0^T \omega_{\varphi(t)} \Delta S_t^{\mathsf{T}} \nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)\sigma_t(\xi^\star)\mathrm{d}B_t.
\end{aligned} \tag{4.15}$$

Consider the four terms on the right-hand side of (4.15). The first term is equal to 0 since, using that $\nabla_\xi c^{-1}_{\tau_{i-1}^n}(\xi^\star) = -c^{-1}_{\tau_{i-1}^n}(\xi^\star)\nabla_\xi c_{\tau_{i-1}^n}(\xi^\star) c^{-1}_{\tau_{i-1}^n}(\xi^\star)$, we have

$$\operatorname{Tr}\left(\sigma_{\tau_{i-1}^n}(\xi^\star)^{\mathsf{T}} \nabla_\xi c^{-1}_{\tau_{i-1}^n}(\xi^\star)\sigma_{\tau_{i-1}^n}(\xi^\star)\right) = -\operatorname{Tr}\left(\nabla_\xi c_{\tau_{i-1}^n}(\xi^\star) c^{-1}_{\tau_{i-1}^n}(\xi^\star)\right).$$

For the second term, using Lemma 4.1 and the properties ($\mathbf{A}_S^{\text{osc.}}$)-($\mathbf{A}_N$) we deduce that

$$\int_0^T \omega_{\varphi(t)} \operatorname{Tr}\left((\sigma_t(\xi^\star) + \sigma_{\varphi(t)}(\xi^\star))^{\mathsf{T}} \nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)(\sigma_t(\xi^\star) - \sigma_{\varphi(t)}(\xi^\star))\right)\mathrm{d}t$$

$$= 2\int_0^T \mathrm{Tr}\left(\sigma_{\varphi(t)}(\xi^\star)^\mathsf{T}\nabla_\xi c^{-1}_{\varphi(t)}(\xi^\star)\sum_{i=1}^d \nabla_{x_i}\sigma_{\varphi(t)}(\xi^\star)\Delta S^i_t\right)\omega_{\varphi(t)}\mathrm{d}t \;+\; e^n_{T,2}$$
$$= \int_0^T \bar{\mathcal{M}}_{\varphi(t)}\Delta S_t \mathrm{d}t \;+\; e^n_{T,2},$$

where for any $\rho > 0$ and any $\eta_\sigma \in (0,1)$, we have

$$|e^n_{T,2}| \le C_0(\varepsilon_n^{2-\rho} + \varepsilon_n \sup_t |t-\varphi(t)|^{\eta_\sigma/2}).$$

Here, C_0 is a notation standing for any a.s. finite random variable (independent on n), which values may change throughout the computations. Also remark that the process $(\bar{\mathcal{M}}_t)_{0\le t\le T}$ is the same as defined in (3.3), Section 3.2. Now using [23, Lemma 3.2], we get

$$\varepsilon_n^{-\rho_N}|e^n_{T,2}| \le C_0\left(\varepsilon_n^{-\rho_N+2-\rho} \vee \varepsilon_n^{-\rho_N+1+(2-\rho)\eta_\sigma/2}\right) \xrightarrow[n\to+\infty]{\text{a.s.}} 0$$

by choosing ρ small enough and η_σ close enough to 1 (recall that $\rho_N < 4/3$ by ($\mathbf{A}_N$)).

The third term of (4.15) 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 [23, Lemma 3.2] and ($\mathbf{H}_S$)-3, 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 (4.15) 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

$$\langle e^n_{\cdot,4}\rangle_T \le C_0 \sup_{0\le t\le T}\left(|\Delta S_t|^2|\sigma_t(\xi^\star) - \sigma_{\varphi(t)}(\xi^\star)|^2\right) \le C_1\varepsilon_n^{2+2\eta_\sigma}$$

for any $\eta_\sigma \in (0,1)$ using ($\mathbf{A}_S^{\text{osc.}}$), Lemma 4.1-(i) and [23, Lemma 3.2]. This implies $\varepsilon_n^{-\rho_N}|e^n_{T,4}| \xrightarrow[n\to+\infty]{\text{a.s.}} 0$ via an application of [20, 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 (3.5). So finally this implies

$$\varepsilon_n^{-\rho_N}(\xi^n - \xi^\star) = -\left(\mathcal{H}_T^{-1} + o_n^{\text{a.s.}}(1)\right)\varepsilon_n^{-\rho_N} Z_T^n + o_n^{\text{a.s.}}(1).$$

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 2.1). □

4.3. Proof of Theorem 3.5

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 (3.6) 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 Central Limit Theorem of [21, Theorem 4.7] are fulfilled, and we get

$$\varepsilon_n^{-\rho} Z_T^n \xrightarrow{\mathcal{L}} \int_0^T \mathcal{K}_t^{1/2} \mathrm{d}\widetilde{W}_t,$$

with an independent Brownian motion $\widetilde{W}$. Moreover, the above convergence is $\mathcal{F}$-stable (see [28, Section 2.2.1] for related definition and properties). Therefore, together with the convergence of $\varepsilon_n^{2\rho} N_T^n$, we deduce the announced result in (i).

Lower bound. We have

$$N_T^n \langle Z^n\rangle_T = N_T^n \int_0^T 4\omega_t^2 \Delta S_t^\mathsf{T} \nabla_\xi c_t^{-1}(\xi^\star)\sigma_t(\xi^\star)\sigma_t(\xi^\star)^\mathsf{T}\nabla_\xi c_t^{-1}(\xi^\star)\Delta S_t \mathrm{d}t.$$

Take some subsequence $\iota(n)$ such that $\sum_{n\geq 0}\varepsilon_{\iota(n)}^2 < +\infty$ and such that the convergence of $N_T^{\iota(n)}\langle Z^{\iota(n)}\rangle_T$ holds a.s.. Then $\mathcal{H}_T^{-2} N_T^{\iota(n)}\langle Z^{\iota(n)}\rangle_T \xrightarrow[n\to+\infty]{\text{a.s.}} V_T$ where V_T is the limit variance of $\sqrt{N_T^n}(\xi^n - \xi^\star)$, in view of the above arguments for proving (i). From the proof of [23, Theorem 4.2] we obtain that

$$V_T = \mathcal{H}_T^{-2}\lim_n N_T^{\iota(n)}\langle Z^{\iota(n)}\rangle_T \geq \mathcal{H}_T^{-2}\left(\int_0^T 2\omega_t \operatorname{Tr}(X_t(\xi^\star))\mathrm{d}t\right)^2 =: V_T^{opt.} \quad \text{a.s..}$$

This finishes the proof of (ii).

κ-optimal sequence. We now prove (iii). Let Z^n be defined in Theorem 3.2 based on $\mathcal{T}_\kappa^n$, and $(\xi^n)_{n\geq 0}$ be the corresponding estimator sequence. By Theorem 3.4 we get the convergence $\sqrt{N_T^n}(\xi^n - \xi^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_T^\kappa)$. In addition, by Proposition 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 3.4.

Appendix 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 preliminary lemma provides the expressions for $\nabla_\xi f(G(\xi), x)$ and $\nabla_\xi^2 f(G(\xi), x)$.

Lemma A.1. *We have (for all $1 \leq k, l \leq m$)*

$$\nabla_{\xi_k} \log(\det G(\xi)) = \mathrm{Tr}(\nabla_{\xi_k} G(\xi) G^{-1}(\xi)), \tag{A.1}$$

$$\nabla^2_{\xi_k \xi_l} \log(\det G(\xi)) = \mathrm{Tr}\left(\nabla^2_{\xi_k \xi_l} G(\xi) G^{-1}(\xi) + \nabla_{\xi_k} G(\xi) \nabla_{\xi_l} G^{-1}(\xi)\right), \tag{A.2}$$

and, as a consequence,

$$\nabla_{\xi_k} f(G(\xi), x) = \mathrm{Tr}\left(\nabla_{\xi_k} G(\xi) G^{-1}(\xi)\right) + x^\mathsf{T} \nabla_{\xi_k} G^{-1}(\xi) x, \tag{A.3}$$

$$\begin{aligned}\nabla^2_{\xi_k \xi_l} f(G(\xi), x) = {} & \mathrm{Tr}\left(\nabla^2_{\xi_k \xi_l} G(\xi) G^{-1}(\xi) + \nabla_{\xi_k} G(\xi) \nabla_{\xi_l} G^{-1}(\xi)\right) \\ & + x^\mathsf{T} \nabla^2_{\xi_k \xi_l} G^{-1}(\xi) x.\end{aligned} \tag{A.4}$$

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 (A.1), a second derivation now implies (A.2). The expressions (A.3) and (A.4) now follow from the definition of $f(G, x)$ and (A.1)-(A.2). □

Appendix 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 (3.2). 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 (3.6) 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 [20, Theorem 3.2].

Proposition 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 3.4). For a given $n \in \mathbb{N}$, define the discretization grid $\mathcal{T}_\kappa^n$ by*

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

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*

$$0 \leq V_T^\kappa - \left(\int_0^T 2\omega_t \operatorname{Tr}(X_t) \mathrm{d}t \right)^2 \leq C_0 \kappa \tag{B.2}$$

for some a.s. finite random variable C_0 *independent of* $\kappa \in (0, 1]$*.*

Proof. First note that from Theorem 3.4 (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 $\sum_{n\geq 0} \varepsilon_{\iota(n)}^2 < +\infty$ and the grid sequence $\mathcal{T}$ verifies ($\mathbf{A}_S^{\mathrm{osc.}}$)-($\mathbf{A}_N$). 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 [23, Lemma 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

$$N_T^n \langle Z^n \rangle_T - N_T^n \langle \bar{Z}^n \rangle_T \xrightarrow[n\to+\infty]{\text{a.s.}} 0. \tag{B.3}$$

By ($\mathbf{A}_S^{\mathrm{osc.}}$)-($\mathbf{A}_N$) and [23, Theorem 3.4], the sequence of grids $\mathcal{T}_\kappa$ is admissible for the process M_t in the sense of [20]. Thus, for the subsequence $(\varepsilon_{\iota(n)})_{n\geq 0}$, the statement follows from (B.3) and [20, Theorem 3.2] applied to $N_T^n \langle \bar{Z}^n \rangle_T$, with

$$C_0 := \left(\sup_{\kappa\in(0,1]} C_\kappa \right) \left(\int_0^T \chi_\kappa(\lambda_{\min}(\Lambda_t)) \operatorname{Tr}\left(\sigma_t \sigma_t^\top\right) \mathrm{d}t \right)$$

where $C_\kappa := \int_0^T \Big(8\omega_t \operatorname{Tr}(X_t) + 3\kappa \chi_\kappa(\lambda_{\min}(\Lambda_t)) \operatorname{Tr}(c_t) \Big) \mathrm{d}t$. 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 $(\varepsilon_n)_{n\geq 0}$. ☐

References

[1] Y. Ait-Sahalia. Transition densities for interest rate and other nonlinear diffusions. *J. Finance*, 45:1361–1395, 1999. MR1737724

[2] Y. Ait-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70:223–262, 2002. MR1926260

[3] Y. Ait-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.*, 36:906–937, 2008. MR2396819

[4] Y. Ait-Sahalia and P. Mykland. The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, 71(2):483–549, 2003. MR1958137

[5] Y. Ait-Sahalia and P. Mykland. Estimators of diffusions with randomly spaced discrete observations: A general theory. *Annals of Statistics*, 32:2186–2222, 2004. MR2102508

[6] M. 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. MR0680660

[7] P. Billingsley. *Probability and measure*. Wiley-Interscience, Wiley Series in Probability and Statistics, third edition, 1995. MR1324786

[8] L. L. Cam and G. Yang. *Asymptotics in statistics*. Springer Verlag, 1990. MR1066869

[9] J. Chang and S. Chen. On the approximate maximum likelihood estimation for diffusion processes. *Ann. Statist.*, 39:2820–2851, 2011. MR3012393

[10] G. Dohnal. On estimating the diffusion coefficient. *Journal of Applied Probability*, 24:105–114, 1987. MR0876173

[11] D. Duffie and P. Glynn. Estimation of continuous-time Markov processes sampled at random time intervals. *Econometrica*, 72(6):1773–1808, 2004. MR2095532

[12] D. Florens-Zmirou. On estimating the diffusion coefficient from discrete observations. *Journal of applied probability*, 30(4):790–804, 1993. MR1242012

[13] C. Fuchs. *Inference for diffusion processes*. Springer-Verlag Berlin Heidelberg, 2013. MR3015023

[14] M. Fukasawa. Central limit theorem for the realized volatility based on tick time sampling. *Finance Stoch*, 14:209–233, 2010. MR2607763

[15] M. Fukasawa. Discretization error of stochastic integrals. *Annals of Applied Probability*, 21(4):1436–1465, 2011. MR2857453

[16] M. Fukasawa and M. Rosenbaum. Central limit theorems for realized volatility under hitting times of an irregular grid. *Stochastic Processes Appl.*, 122(12):3901–3920, 2012. MR2971719

[17] 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. MR1204521

[18] V. Genon-Catalot and J. Jacod. Estimation of the diffusion coefficient for diffusion processes: Random sampling. *Scandinavian Journal of Statistics*, 21:193–221, 1994. MR1292636

[19] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7:899–912, 2001. MR1873834

[20] E. Gobet and N. Landon. Almost sure optimal hedging strategy. *Ann. Appl. Probab.*, 24(4):1652–1690, 2014. MR3211007

[21] E. Gobet, N. Landon, and U. Stazhynski. Central Limit Theorem for discretization errors based on stopping time sampling. *Preprint hal-01879287, in revision for Bernoulli*, 2018.

[22] E. Gobet and S. Menozzi. Stopped diffusion processes: boundary corrections and overshoot. *Stochastic Processes and Their Applications*, 120:130–162, 2010. MR2576884

[23] E. Gobet and U. Stazhynski. Optimal discretization of stochastic integrals driven by general Brownian semimartingale. *Annales de l'Institut Henri*

Poincaré, Probability and Statistics, 54(3):1556–1582, 2018. MR3825891

[24] E. Gobet and U. Stazhynski. Model-adaptive optimal discretization of stochastic integrals. *Stochastics*, 91(3):321–351, 2019. MR3925661

[25] J. Grammig and M. Wellner. Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics*, 106:369–400, 2002. MR1885374

[26] 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. MR2884231

[27] J. Jacod. Statistics and high frequency data. *Unpublished CIMPA lecture notes,* *http://archive.schools.cimpa.info/archivesecoles/20140206104549/2009-j-semstat.pdf*, 2007.

[28] J. Jacod and P. Protter. Discretization of processes. *Stochastic Modelling and Applied Probability, 67. Springer-Verlag*, 2012. MR2859096

[29] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes*. Springer Verlag, Berlin, second edition, 2002. MR1943877

[30] Y. Li, P. Mykland, E. Renault, L. Zhang, and X. Zheng. Realized volatility when sampling times are possibly endogenous. *Econometric Theory*, 30(3):580–605, 2014. MR3205607

[31] H. Sørensen. Parametric inference for diffusion processes observed at discrete points in time: a survey. *Internat. Statist. Rev.*, 72(3):337–354, 2004.