An optimal control variance reduction method for density estimation

Ahmed Kebaier*& Arturo Kohatsu-Higa †

December 7, 2007

We study the problem of density estimation of a non-degenerate diffusions using kernel functions. Thanks to Malliavin calculus techniques, we obtain an expansion of the discretization error. Then, we introduce a new control variate method in order to reduce the variance in the density estimation. We prove a stable law convergence theorem of the type obtained in Jacod-Kurtz-Protter, for the first Malliavin derivative of the error process, which leads us to get a CLT for the new variance reduction algorithm. This CLT gives us a precise description of the optimal parameters of the method.

1 Introduction

Let $(X_t)_{0\leq t\leq T}$ be d-dimensional diffusion such that X_T has a smooth density, denoted by $p(x)$. The goal of the present article is to discuss in theoretical terms a control variate method to reduce the variance in the Monte Carlo estimation of $p(x)$.

To introduce the problem, first note that $p(x) = \mathbb{E}\delta_x(X_T)$ where δ_x denotes the Dirac delta distribution function. In order to use the Monte Carlo method we have first to approximate the Dirac delta function .

Consider an integrable continuous function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} \phi(x)dx = 1$ and define the kernel functions

$$\phi_{h,x}(y) = \frac{1}{h}\phi(\frac{y-x}{h}),\ \ h > 0 \quad \text{et} \quad x \in \mathbb{R}.$$

Note that $\phi_{h,x} \to \delta_x$ as $h \to 0$, in a weak sense, according to the assumptions on the function ϕ. The idea is then to approximate the density $p(x) = \mathbb{E}\delta_x(X_T)$ by $\mathbb{E}\phi_{h,x}(X^n_T)$ where $h = n^{-\alpha}$, $\alpha > 0$ and X^n denotes an approximation of X that can be simulated. At this level, a first problem arises. That is, the problem of evaluating the weak error given by $\mathbb{E}\phi_{h,x}(X^n_T) - p(x)$.

Once this problem is solved one fixes the desired error level and from the weak error estimate one obtains a restriction for the value of α. Nevertheless when one carries out the Monte Carlo simulations, one finds that usually the variance of the estimators is relatively high and therefore variance reduction methods have to be studied in order to achieve a prescribed accuracy with less number of calculations.

The present work is framed in this setting. In particular we study a control variate method introduced in the regular case (that is, in the case of the approximation of $\mathbb{E}f(X_T)$ for a smooth function f) in Kebaier (2005).

*Université Paris 13, Institut Galilée, Mathématiques, 99, av. JB Clément, F-99430 Villetaneuse, France. e-mail: kebaier@math.univ-paris13.fr

†Osaka University Graduate School of Engineering Sciences. Machikaneyama cho 1-3. Osaka 560-8531, Japan. email: kohatsu@sigmath.es.osaka-u.ac.jp

Mathematics subject classification: 60H07; 60H35; 60J60; 62G07; 65C05; 60F05

Key words or phrases: kernel density estimation, stochastic differential equations, variance reduction, weak approximation, central limit theorem, Malliavin calculus

To be more precise, suppose that X has smooth coefficients and satisfies the Hörmander condition. Then X_T has a smooth density $p(x)$. Let X^n denote the Euler-Maruyama scheme of time step T/n. Under some extra conditions (see Bally and Talay (1996)) one obtains the following expansion for the density diffusion

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

where $p_n(x) = \mathbb{E}\phi_{h,x}(X_T^n)$ is a regularized density of the Euler scheme X^n. This regularization is needed because under the present conditions X^n may not have a density.

In Kohatsu-Higa and Pettersson (2002), the above result is obtained under weaker conditions on x and X_0 but the expression on the expansion on the error is less explicit. Then a Monte Carlo simulation study of $\mathbb{E}\phi_{h,x}(X_T^n)$ is presented. After obtaining the variance of this estimator the authors propose a variance reduction method using a localization function. The procedure used can be described as follows.

Using the integration by parts formula of Malliavin calculus, Kohatsu-Higa and Pettersson (2002) obtain that

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

where $\psi_{h,x}$ is the primitive function of $\phi_{h,x}$ and H_n is the weight given by the Malliavin calculus. Using this idea, Kohatsu-Higa and Pettersson (2002) prove that

$$\mathbb{E}\left(\phi_{h,x}(X_T^n)\varphi(X_T^n - x)\right) = \mathbb{E}\Big(\psi_{h,x}(X_T^n)(\varphi(X_T^n - x)H_n + \varphi'(X_T^n - x)G_n\Big),$$

for a smooth function φ with $\varphi(0) = 1$. Explicit expressions for the random variables H_n and G_n are obtained and the asymptotic variance is minimized with respect to φ obtaining as result that the optimal φ is of exponential type. These results are later verified trough simulations. The disadvantage of this method is that the computation time of their algorithm is higher than that of the method using kernel density functions.

On the other hand, in Kebaier (2005), the author considered the approximation of $Ef(X_T)$ by a Monte Carlo algorithm where f is a given regular function and X_T is a diffusion process. In particular, the so-called statistical Romberg method is introduced and analyzed. This control variate method gives a variance reduction if parameters are chosen appropriately. The optimal parameters are obtained after a careful study of a central limit theorem for the error process.

In this paper, we generalize these results to the case of density approximations. That is when f is a Dirac delta function under the Hörmander condition.

The method uses two Euler schemes X_T^n and X_T^m with $m << n$ as follows.

Suppose for the moment that $\mathbb{E}\phi_{h,x}(X_T^m)$ can be computed explicitly. Then the classical control variate method can be applied as follows:

$$\frac{1}{N_{n,m}} \sum_{i=1}^{N_{n,m}} \{\phi_{h,x}(X_{T,i}^n) - \phi_{h,x}(X_{T,i}^m)\} + \mathbb{E}\phi_{h,x}(X_T^m),$$

where the index $i = 1, ..., N_{n,m}$ indicates independent simulations of the corresponding random variable. As the last quantity above is in fact not known we will use an additional N_m independent simulations to estimate this quantity. Therefore the final calculation scheme is given by

$$\frac{1}{N_{n,m}} \sum_{i=1}^{N_{n,m}} \{\phi_{h,x}(X_{T,i}^n) - \phi_{h,x}(X_{T,i}^m)\} + \frac{1}{N_m}\sum_{i=1}^{N_m} \phi_{h,x}(\hat{X}_{T,i}^m).$$

Now in order not to increase the number of simulations we simulate a large number, N_m, of sample paths with a coarse time discretization step T/m and few additional sample paths of size $N_{n,m}$ with the fine time discretization step T/n.

In order to choose the parameters h, n, m, N_m and $N_{n,m}$ to achieve a certain desired error level, one has to study the weak error of the above expression together with the variance behavior. This brings us to study a central limit theorem for the error process.

A similar study in the regular case (with one less parameter, h) is carried out in Kebaier (2005). If one choses the same parameters as in the regular case with h given by the kernel density method then there is explosion of variances. Even more, it is one of the conclusions of this article that there is no variance reduction that one can achieve with this method if one uses kernels as have been defined previously.

In fact, one has to use the concept of super-kernel of order s with $s > 2(d+1)$ in order to achieve some variance reduction (see Definition 3.1 and Theorem 6.1).

As these kernels do not correspond exactly with the results in Bally and Talay (1996) or Kohatsu-Higa and Pettersson (2002), we start by finding the expansion of the weak error (see Theorem 3.1).

Our final aim is to find the optimal parameters leading to an optimal complexity of the algorithm. In order to obtain these optimal parameters we extend a result of Jacod and Protter (1998) for the asymptotic behavior of the law of the first Malliavin derivative of the error in the Euler scheme. Using this extension we prove a CLT, for our algorithm, giving us a precise description of the choice of the optimal parameters m, N_m and $N_{n,m}$.

The usual version of the integration by parts formula of Malliavin Calculus in dimension d, see Nualart (1995,2006) (p.103, 2006 edition) is based on using d times the integration by parts formula. Although it is feasible (although long) to prove the stable convergence of the high order weights, we propose instead to use a new integration by parts formula introduced by Malliavin and Thalmaier (2006) which significantly simplifies the proof in the general multi-dimension context.

The optimal parameters given by the CLT lead to an optimal complexity of the algorithm of order $n^{\frac{5}{2}+(d+1)\alpha}$ which is less than the optimal complexity of the Monte Carlo method for the kernel density method which is of order $n^{3+\alpha d}$, where α is the parameter tuning the window size h which depends on the order of the superkernel. Finally, d is the dimension of the problem.

The gain obtained here is of order $n^{\frac{1}{2}-\alpha}$. Consequently, we have an exact mathematical estimate of when and how much variance reduction can be achieved. Whereas, there is less reduction than in the regular case due to the explosion of the variance of our estimators (see section 6 for more details).

The remainder of the paper is organized as follows. In the following section, we introduce some basics of the Malliavin Calculus. In section 3, we study the weak discretization error. Section 4 is devoted to prove the CLT for the classical Monte Carlo method. In section 5 we prove a stable convergence theorem for the first Malliavin derivative of the error in the Euler scheme. In the last section we prove a CLT for the statistical Romberg algorithm and we give the optimal parameters leading to an optimal complexity of the method.

In the Appendices we give the proofs of technical lemmas used throughout the proofs.

2 Malliavin Calculus

2.1 Main definitions and properties

We follow the notations, definitions and results of Nualart (1995,2006). Let $(W_t)_{0\leq t\leq T}$ be a q dimensional standard Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ where $(\mathcal{F}_t)_{0\leq t\leq T}$ denotes the standard filtration. D denotes the Malliavin derivative which takes values in $H := L^2([0,T];\mathbb{R}^q)$. The k-th order derivative of F for a multi-index $k \in \{1,...,q\}^l$, $l \in \mathbb{N}$ of length $|k| = l$ is denoted by $D^k F$, it takes values in $H^{\otimes l}$ and is given by

$$D^k_{t_1,\dots,t_l} F = D^{k_1}_{t_1} \dots D^{k_l}_{t_l} F$$

where $k = (k_1, ..., k_l)$.

Note that the operator D^k is closable for any $k \in \{1,...,q\}^l$. For $p \geq 1$ and $l \in \mathbb{N}$, we denote $\mathbb{D}^{l,p}(W)$ the closure of the space of smooth random variables with respect to the norm $\|\cdot\|_{l,p}$.

We denote $\mathbb{D}^\infty(W) = \bigcap_{p\geq 1}\bigcap_{l\geq 1}\mathbb{D}^{l,p}(W)$. For $F = (F^1, \dots, F^d) \in (\mathbb{D}^\infty(W))^d$, we introduce γ_F the Malliavin covariance matrix of F given by

$$\gamma_F^{ij} = \langle DF^i, DF^j\rangle_H, \quad 1 \leq i, j \leq d$$

2.2 Duality and integration by parts formulas

Let δ denote the adjoint operator of D, which is also called *Skorokhod integral*. The operator δ is unbounded, we denote by $Dom(\delta)$ its domain (see for example Definition 1.3.1 of Nualart (1995,2006)). Note that if $u \in L^2\big([0,T]\times\Omega;\mathbb{R}^q\big)$ is an adapted process, then (see Proposition 1.3.4 in Nualart (1995,2006)) $u \in Dom(\delta)$ and $\delta(u)$ coincides with the Itô integral.

If $F \in \mathbb{D}^{1,2}$ and $u \in Dom(\delta)$ then $Fu \in Dom(\delta)$ and we have

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_H.$$

In such a case we have the following duality formula

$$\mathbb{E}\Big[\langle u, DF\rangle_H\Big] = \mathbb{E}\big[F\delta(u)\big]. \tag{1}$$

In the following we give the definition of a non-degenerate random vector.

Definition 2.1. *A random vector $F = (F^1,\dots,F^d) \in (\mathbb{D}^\infty(W))^d$ is said to be non-degenerate if the Malliavin covariance matrix of F is invertible a.s. and*

$$(\det\gamma_F)^{-1} \in \bigcap_{p\geq 1} L^p(\mathbb{P}^W).$$

For a nondegenerate random vector, the following integration by parts formula plays a key role. (For a proof of the following proposition see Nualart (1998)).

Proposition 2.1. *Let $F \in \big(\mathbb{D}^\infty(W)\big)^d$ be a non-degenerate random vector. Let $f \in C_p^\infty(\mathbb{R}^n)$, and let $G \in \mathbb{D}^\infty(W)$. Fix $k \geq 1$. Then for any multi-index $m = (m_1,\dots,m_k) \in \{1,\dots,d\}^k$ we have*

$$\mathbb{E}\big[\partial_m f(F)G\big] = \mathbb{E}\big[f(F)\,\mathbf{H}_m(F,G)\big],$$

where $\partial_m = \partial_{m_1}\dots\partial_{m_k}$ and the random variable $\mathbf{H}_m(F,G)$ is defined inductively as follows

$$\begin{aligned} \mathbf{H}_{(i)}(F,G) &= \textstyle\sum_{j=1}^d \delta\Big(DF^jG(\gamma_F^{-1})^{ij}\Big) \\ \mathbf{H}_m(F,G) &= \mathbf{H}_{(m_k)}\big(F, \mathbf{H}_{(m_1,\dots,m_{k-1})}(F,G)\big). \end{aligned}$$

2.3 An extension of the integration by parts formula

In the following work we will deal with a d-dimensional diffusion $X = (X^1,\dots,X^d)$ driven by a q-dimensional Brownian motion $W = (W^1,\dots,W^q)$. In order to regularize the Euler scheme associated to the diffusion X, we will employ d additional noises, corresponding to $X^1,\dots,X^d$. In order to do that, we consider a d-dimensional Brownian motion $\bar{W} = (W^{q+1},\dots,W^{q+d})$, independent of $W = (W^1,\dots,W^q)$, and we set

$$\tilde{W} = (W,\bar{W}) = (W^1,\dots,W^q,W^{q+1},\dots,W^{q+d}).$$

Therefore our random vectors are defined on the Wiener space of dimension $r = q+d$, but we should distinguish between the two Brownian motions W et $\bar{W}$ which play different roles in our calculation: W drive the diffusion whereas $\bar{W}$ is an additional noise used for the regularization. Hence, by using again the notations of the preceding subsection we obtain

$$\tilde{D} = (D,\bar{D}) = (D^1,\dots,D^q,D^{q+1},\dots,D^{q+d})$$

and for $\tilde{u} = (u,\bar{u}) = (u^1,\dots,u^q,u^{q+1},\dots,u^{q+d})$ we have

$$\tilde{\delta}(\tilde{u}) = \delta(u) + \bar{\delta}(\bar{u}).$$

The norms $\|F\|_{k,p}$ are norms defined on $\mathbb{D}^{k,p}(\tilde{W})$, thus it involves all the derivatives $\tilde{D} = (D, \bar{D})$. Similarly, the Malliavin covariance matrix of the random vector F is given by

$$\tilde{\gamma}_F = \langle \tilde{D}F, \tilde{D}F \rangle.$$

The auxiliary noise, that we will use, is given by the random vector

$$Z_{n,\theta} := \frac{\bar{W}_T}{n^{\frac{1}{2}+\theta}}, \quad \theta \geq 0. \tag{2}$$

In the following, we introduce the random vector $F = (F_1, \dots, F_d)$ which depends only on $W = (W^1, \dots, W^q)$ and the random variable G which depends only on $\tilde{W} = (W, \bar{W})$. The proposition below, is a natural extension of Proposition 2.1, gives us an explicit expression of $\tilde{H}_i$ which appears in the integration by parts formula.

Proposition 2.2. *Let $F \in \big(\mathbb{D}^\infty(W)\big)^d$ be a non-degenerate random vector. Let $f \in C_p^\infty(\mathbb{R}^d)$, and let $G \in \mathbb{D}^{k,2}(\tilde{W})$. Fix $k \geq 1$. Then for any multi-index $m = (m_1, \dots, m_k) \in \{1, \dots, d\}^k$ we have*

$$\mathbb{E}\big[\partial_m f(F + Z_{n,\theta})G\big] = \mathbb{E}\big[f(F + Z_{n,\theta})\, \tilde{\mathbf{H}}_m(F, G)\big], \tag{3}$$

where the random variable $\tilde{\mathbf{H}}_m(F, G)$ is given by

$$\begin{aligned}
\tilde{\mathbf{H}}_{(i)}(F, G) &= \sum_{j=1}^d \tilde{\delta}\Big(\tilde{D}(F + Z_{n,\theta})^j G(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{ij}\Big) \\
&= \sum_{j=1}^d \delta\Big(G(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{ij} DF^j\Big) + \frac{1}{n^{\frac{1}{2}+\theta}} \sum_{j=1}^d \bar{\delta}\Big(G(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{ij} \bar{D}\bar{W}_T^j\Big), \\
\tilde{\mathbf{H}}_m(F, G) &= \tilde{\mathbf{H}}_{(m_k)}\big(F, \tilde{\mathbf{H}}_{(m_1, \dots, m_{k-1})}(F, G)\big),
\end{aligned}$$

with $\bar{\delta}$ and $\tilde{\delta}$ are respectively the adjoint operators of $\bar{D}$ and $\tilde{D}$.

2.4 Malliavin Thalmaier integration by parts formula

Recently Malliavin and Thalmaier (2006) introduced a new idea of integration by parts based on the Riesz transform. Essentially this amounts to replace the representation of the Dirac function δ_0 by

$$\delta_0 = \Delta Q,$$

where $\Delta = \sum_{i=1}^d \partial_i^2$ is the Laplace operator and Q_d is the fundamental solution of the Poisson equation in the following sense. If f denotes some function, then the solution of the equation $\Delta u = f$ is given by the convolution $Q_d * f$. The explicit expressions for Q_d are $Q_1(x) = x_+$, $Q_2(x) = a_2 \ln|x|$ and $Q_d(x) = a_d|x|^{-(d-2)}$ for $d > 2$ and suitable constants a_d, $d \geq 2$. Then we have a new integration by parts formula

Proposition 2.3. *Let $F \in \big(\mathbb{D}^\infty(W)\big)^d$ be a non-degenerate random vector. Let $G \in \mathbb{D}^{k,2}(\tilde{W})$ and $x \in \mathbb{R}_d$. Then*

$$\mathbb{E}[\delta_0(F - x)G] = \mathbb{E}[\Delta Q_d(F - x)G] = \sum_{r=1}^d \mathbb{E}[\partial_r Q_d(F - x) H_{(r)}(F, G)],$$

where $H_{(r)}(F, G)$ are the weights of the classical integration by parts formula (see Proposition 2.1)

Note that $\partial_r Q_d(F - x)$ is integrable but not bounded. Consequently, the advantage of this new approach is that one has to make just one integration by parts, because we need to remove only one derivative, while in the classical integration by parts formula we have to make d integration by parts in order to remove the d derivatives in $\delta_0(y - x) = \partial^{(1,\dots,d)} \mathbf{1}_{\{y_i \geq x_i;\ i=1,\dots,d\}}$.

3 Weak convergence of the approximate density

Let $(X_t)_{0\leq t\leq T}$ be a $\mathbb{R}^d$-valued diffusion process which is the solution of the following stochastic differential equation

$$dX_t = f(X_t)dY_t, \quad X_0 = x \in \mathbb{R}^d, \tag{4}$$

where $Y_t = (t, W_t^1, \dots, W_t^q)^T$, with $W = (W^1, \dots, W^q)$ a q-dimensional Brownian motion defined on the filtered probability space $\mathcal{B} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, where $(\mathcal{F}_t)_{t\geq 0}$ denotes a filtration satisfying the usual conditions. The function $f : \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times(q+1)}$ is of class $\mathscr{C}_b^{d+3}$. In order to distinguish clearly the drift from the diffusion term we will use indices as follows $f = (f_{ij})_{i=1,\dots,d; j=0,1,\dots,q}$. So that $j = 0$ corresponds to the drift coefficient.

The Euler scheme, denoted by X^n, associated to the diffusion X and with discretization step $\delta = T/n$ is defined as:

$$dX_t^n = f(X_{\eta_n(t)}^n)dY_t, \quad \eta_n(t) = [t/\delta]\delta.$$

The next result gives bounds on the error of the Euler scheme in the sense of $\| \|_{l,p}$-norms. For a proof of this result see Kusuoka and Stroock (1984) and **?**.

Proposition 3.1. *With the previous notation, the following two properties are valid:*

$\mathbf{P}_1)$ $\forall t > 0, \quad X_t^n,\ X_t \in (\mathbb{D}^\infty(W))^d$

$\mathbf{P}_2)$ $\forall p > 1,\ \forall l \in \mathbb{N}^*, \exists K > 0$ *such that:*

$$\sup_{t\in[0,T]} \|X_t\|_{l,p} + \sup_{t\in[0,T]} \|X_t^n\|_{l,p} \leq K(1 + \|x\|) \tag{5}$$

and

$$\sup_{t\in[0,T]} \|X_t^n - X_t\|_{l,p} \leq \frac{K}{\sqrt{n}}. \tag{6}$$

Furthermore $D^k_{t_1,\dots,t_l}F(t)$ *is* $L^p(\Omega)$*-continuous in* $(t, t_1, ..., t_l)$ *for* $t_i \leq t$, $i = 1, ..., l$, $F = X,\ X^n$ *and any* $p > 1$ *and any multi-index* k *such that* $|k| = l$.

Notation:
For a function $V : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, we denote by $\mathbf{D}V$ the Jacobian matrix of V and by $\mathbf{D^2}V$, its Hessian matrix. We suppose that the d-dimensional diffusion process $(X_t)_{0\leq t\leq T}$, which is the solution of (4) has a coefficient f, which satisfies the Hörmander condition (see Section 2.3.2 of Nualart (1995,2006)).

Therefore X admits a smooth density $p_T(x_0, x)$ (see Kusuoka and Stroock (1985)) and in order to simplify the notation, we denote

$$p_T(x_0, x) := p(x).$$

We note here that the Hörmander condition is not enough to guarantee that the Malliavin covariance matrix associated to the Euler scheme X^n, is invertible (this would be true under an ellipticity condition).

To deal with this problem we will regularize the Euler scheme using $X^n + Z_{n,\theta}$ instead of X^n, $Z_{n,\theta}$ denotes a independent random variable defined in Section 2.3 through the relation

$$Z_{n,\theta} = \frac{\bar{W}_T}{n^{\frac{1}{2}+\theta}}$$

where $\tilde{W}$ is a d-dimensional Brownian motion independent of W. Then we have the following result.

Proposition 3.2. *For $\lambda \in [0,1]$ we introduce*

$$X_T^{n,\lambda} = X_T + \lambda(X_T^n - X_T).$$

Then for all $p \geq 1$ there exists a constant $K_T > 0$ and parameters $p', p'' \geq 1$ such that

$$\sup_n \Big\| \left(\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} \right)^{-1} \Big\|_p \leq K_T \Big\| (\det \gamma_{X_T})^{-1} \Big\|_{p'}^{p''} < \infty.$$

Proof. We have that $\mathbb{E}\left(\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} \right)^{-p} = A_n + B_n$ with

$$A_n := \mathbb{E}\Big\{ \left(\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} \right)^{-p} \mathbf{1}_{\left|\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} - \det \gamma_{X_T}\right| < \frac{1}{2}\det \gamma_{X_T}} \Big\}$$

and

$$B_n := \mathbb{E}\Big\{ \left(\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} \right)^{-p} \mathbf{1}_{\left|\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} - \det \gamma_{X_T}\right| \geq \frac{1}{2}\det \gamma_{X_T}} \Big\}$$

As the diffusion X is non-degenerated in the sense of definition 2.1, we deduce that

$$\sup_n A_n \leq 2^p \mathbb{E}\left(\det \gamma_{X_T} \right)^{-p} < +\infty.$$

On the other hand, we have that

$$\tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} = \gamma_{X_T^{n,\lambda}} + \frac{T}{n^{1+2\theta}} I_d.$$

As $\gamma_{X_T^{n,\lambda}}$ is a positive definite matrix we deduce that

$$\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} \geq \left(\frac{T}{n^{1+2\theta}} \right)^d.$$

Therefore, one obtains that

$$B_n \leq \left(\frac{T}{n^{1+2\theta}} \right)^{-dp} \mathbb{P}\Big(\left|\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} - \det \gamma_{X_T}\right| \geq \frac{1}{2}\det \gamma_{X_T} \Big).$$

Therefore using the Markov inequality, we have that

$$B_n \leq 2^k \left(\frac{T}{n^{1+2\theta}} \right)^{-dp} \mathbb{E}\Big\{ (\det \gamma_{X_T})^{-1} \left|\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} - \det \gamma_{X_T}\right| \Big\}^k$$

$$\leq 2^k \left(\frac{T}{n^{1+2\theta}} \right)^{-dp} \Big\| \left|\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} - \det \gamma_{X_T}\right|^k \Big\|_2 \big\| (\det \gamma_{X_T})^{-k} \big\|_2$$

Therefore from the inequalities (5) and (6), we obtain that

$$\Big\| \left|\det \tilde{\gamma}_{X_T^{n,\lambda}+Z_{n,\theta}} - \det \gamma_{X_T}\right|^k \Big\|_2 \leq \frac{C_k}{n^{\frac{k}{2}}}$$

where C_k is a given constant. Finally, if we take $k = 2dp(1+2\theta)$ we obtain that

$$\sup_n B_n < \infty.$$

$\square$

In what follows we are interested in considering the approximation of the marginal density $p(x)$ of the diffusion X using kernel density estimation methods.

Definition 3.1. *Let $\phi \in C_b^\infty(\mathbb{R};\mathbb{R})$, we say that ϕ is a super-kernel of order $s > 2$ if*

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

In what follows, we suppose that ϕ satisfies the following properties:

a) $\displaystyle\int_{\mathbb{R}} |x|^{s+1} |\phi(x)|\,dx < \infty,$ where s denotes the order of the kernel,

b) $\displaystyle\int_{\mathbb{R}} |\phi'(x)|^2\,dx < \infty, \ \int_{\mathbb{R}} |\phi(x)|^l dx < \infty,$ for $l = 1,\ 2,\ 3.$

For $h > 0$, we define

$$\phi_{h,x}(y) = \frac{1}{h}\phi\Big(\frac{y-x}{h}\Big).$$

The parameter h is called *the window size of the kernel.*

We extend the previous concepts to $\mathbb{R}^d$ as follows: Let $\phi_i : \mathbb{R} \mapsto \mathbb{R}$, for $i = 1, \dots, d$ be one dimensional super kernels. We set

$$\phi(u_1, \dots, u_d) = \phi_1(u_1) \times \cdots \times \phi_d(u_d).$$

We say that ϕ is a super kernel of order s if the functions $\phi_i,\ i = 1, \dots, d$ are one dimensional super kernels of order s. Furthermore, we define

$$\phi_{h,x}(y) = \frac{1}{h^d}\phi\Big(\frac{y-x}{h}\Big) = \prod_{i=1}^{d} \phi_{i,h,x}(y_i)$$

In the calculations to follow, we will also use other kernels that stem from ϕ. So, we define for $l = 1, 2, 3$

$$\phi^l_{h,x}(y) = \frac{1}{h^d \phi^l}\Big|\phi\Big(\frac{y-x}{h}\Big)\Big|^l, \qquad \phi^l = \int_{\mathbb{R}^d} |\phi(u)|^l\,du.$$

These positive functions are integrable and integrate to one. Additionally, we define

$$\phi_{(2),h,x}(y) = \frac{1}{h\phi_{(2)}}\Big[\phi'\Big(\frac{y-x}{h}\Big)\Big]^2, \quad \phi_{(2)} = \int |\phi'(x)|^2\,dx$$

Remark 1. One can construct super kernels of infinite order in the following way. We take a symmetric function $\psi \in \mathscr{S}$ (where $\mathscr{S}$ denotes the class of smooth functions rapidly decreasing to zero at infinity) so that $\psi(x) = 1$ in a neighborhood of zero. Next, we define ϕ as the inverse Fourier transform of ψ. That is,

$$\phi(x) := \frac{1}{2\pi}\int_{\mathbb{R}} e^{ix\xi}\psi(\xi)\,d\xi, \quad x \in \mathbb{R}.$$

Due to the symmetric property of ψ, ϕ is a real valued function.

Then the Fourier transform of ϕ is ψ given by

$$\psi(\xi) = \int_{\mathbb{R}} e^{-ix\xi}\phi(x)\,dx, \quad \xi \in \mathbb{R}.$$

As $\psi^{(k)}(0) = 0$, for all $k \in \mathbb{N}$ we conclude also that $\int_{\mathbb{R}} x^k \phi(x)\,dx = 0$ for all $k \in \mathbb{N}$ and as $\psi(0) = 1$ we have that $\int_{\mathbb{R}} \phi(x)\,dx = 1$. The inverse Fourier transform sends the functions $\mathscr{S}$ into $\mathscr{S}$. Therefore $\phi \in \mathscr{S}$ and consequently, it verifies the conditions a) and b) above.

Also, one can easily construct polynomials on compacts which lead to super kernels of order s which are not of order $s+1$. That is, the coefficients of the high order polynomial are determined by equations requiring the smoothness properties and the moment conditions required to the polynomial.

For a multi-index $\alpha \in \{1, ..., d\}^l$, $l \in \mathbb{N}$, we define

$$c_\alpha(\phi) = \int_{\mathbb{R}^d} \prod_{i=1}^{l} u_{\alpha_i} \phi(u)\, du = \prod_{i=1}^{l} \int_{\mathbb{R}^d} u_{\alpha_i}^{p_i} \phi_i(u_i)\, du_i,$$

with $p_i = card\{j : \alpha_j = i\}$. Note in particular, that $c_\alpha = 0$ for $|\alpha| = l < s$. The property that will interest us in the calculations to follow is that the super kernel of order s approximate the Dirac delta function up to the order $s+1$. More precisely, we have the following result.

Lemma 3.1. *1. Let ϕ be a d-dimensional super kernel of order s. Let $f \in \mathcal{C}_b^{s+1}(\mathbb{R}^d; \mathbb{R})$. Then*

$$\left| f(x) - \int_{\mathbb{R}^d} f(y) \phi_{h,x}(y)\, dy - \frac{h^s}{s!} \sum_{|\alpha|=s} c_\alpha(\phi) \partial^\alpha f(x) \right| \le C\, h^{s+1},$$

where $\partial^\alpha f$ denotes the derivative of f corresponding to the multi-index $\alpha = (\alpha_1, \ldots, \alpha_l)$, of length $|\alpha| = l$. Whereas the constant C is given by

$$C = c_s \|f^{(s+1)}\|_\infty \int_{\mathbb{R}^d} \|u\|^{s+1} |\phi(u)|\, du$$

where c_s is a universal constant depending on s and $\|f^{(s+1)}\|_\infty$ is the sup norm of derivatives of order $s+1$ of f.

2. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a positive integrable and bounded function. Suppose that $\int_{\mathbb{R}^d} \varphi(x) dx = 1$. Let $\varphi_{h,x}(y) = \frac{1}{h^d} \varphi\left(\frac{y-x}{h}\right)$, then for every continuous and bounded function f we have

$$\lim_{h \to 0} \int_{\mathbb{R}^d} f(y) \varphi_{h,x}(y)\, dy = f(x).$$

Proof. We have that

$$\begin{aligned} \int_{\mathbb{R}^d} f(y) \phi_{h,x}(y)\, dy - f(x) &= \int_{\mathbb{R}^d} \phi_{h,x}(y)(f(y) - f(x))\, dy \\ &= \int_{\mathbb{R}^d} \phi(u)(f(x+uh) - f(x))\, du \end{aligned}$$

Using a Taylor series expansion of order s for f we obtain

$$\begin{aligned} \int_{\mathbb{R}^d} f(y) \phi_{h,x}(y)\, dy - f(x) &= \sum_{k=1}^{s} \frac{h^k}{k!} \sum_{|\alpha|=k} c_\alpha(\phi) \partial^\alpha f(x) \\ &+ \frac{h^{s+1}}{s!} \sum_{|\alpha|=s+1} \int_{\mathbb{R}^d} \int_0^1 (1-\lambda)^s \partial^\alpha f(x + \lambda u h) \prod_{i=1}^{s+1} u_{\alpha_i} \phi(u)\, d\lambda\, du. \end{aligned}$$

Since $(\phi_j)_{j=1,\ldots,d}$ are super kernels of order s, we conclude that $c_\alpha(\phi) = 0$, for $|\alpha| < s$. Consequently,

$$\begin{aligned} \int_{\mathbb{R}^d} f(y) \phi_{h,x}(y)\, dy - f(x) &= \frac{h^s}{s!} \sum_{|\alpha|=s} c_\alpha(\phi) \partial^\alpha f(x) \\ &+ \frac{h^{s+1}}{s!} \sum_{|\alpha|=s+1} \int_{\mathbb{R}^d} \int_0^1 (1-\lambda)^s \partial^\alpha f(x + \lambda u h) \prod_{i=1}^{s+1} u_{\alpha_i} \phi(u)\, d\lambda\, du. \end{aligned}$$

In the following we evaluate the remainder term.

$$\Big|\int_{\mathbb{R}^d}\int_0^1 (1-\lambda)^s \partial^\alpha f(x+\lambda uh)\prod_{i=1}^{s+1} u_{\alpha_i}\phi(u)\,d\lambda\,du\Big| \le \|f^{(s+1)}\|_\infty \int_{\mathbb{R}^d} \|u\|^{s+1}|\phi(u)|\,du.$$

According to property a) of Definition 3.1, the right side of the inequality is finite and therefore the result follows. The proof of the second assertion follows from the Lebesgue theorem. □

The main theorem of this section gives us an expansion of order 1 of the weak error in the approximation of the density of the hypoelliptic diffusion X.

Before this we study the error process in a form that will also be useful when studying the stable convergence problem.

The error process $U^n = (U^n_t)_{0\le t\le T}$, defined by

$$U^n_t = X_t - X^n_t,$$

satisfies the equation

$$dU^n_t = \sum_{j=0}^q (\dot{f}^n_{t,j}).(X_t - X^n_{\eta_n(t)})\,dY^j_t,$$

where

$$\dot{f}^n_{t,j} = \int_0^1 \nabla f_j\Big(X^n_{\eta_n(t)} + \lambda(X_t - X^n_{\eta_n(t)})\Big)\,d\lambda.$$

Therefore the equation satisfied by U^n can be written as:

$$U^n_t = \int_0^t \sum_{j=0}^q \dot{f}^n_{s,j}\,dY^j_s.U^n_s + G^n_t, \tag{7}$$

with

$$G^n_t = \int_0^t \sum_{j=0}^q \dot{f}^n_{s,j}.(X^n_s - X^n_{\eta_n(s)})\,dY^j_s. \tag{8}$$

Note that

$$X^n_s - X^n_{\eta_n(s)} = \sum_{j=0}^q \bar{f}^n_{s,j}(Y^j_s - Y^j_{\eta_n(s)}), \tag{9}$$

with $\bar{f}^n_{s,j} = f_j(X^n_{\eta_n(s)})$. In the following let $(Z^n_t)_{0\le t\le T}$ be the $\mathbb{R}^{d\times d}$ valued solution of

$$Z^n_t = I_d + \int_0^t \sum_{j=0}^q \dot{f}^n_{s,j}\,dY^j_s.Z^n_s\,.$$

From Theorem 56 p.271 in Protter (1990) we obtain that there exists $(Z^n_s)^{-1}$ for all $s\le T$ which satisfies

$$(Z^n_t)^{-1} = I_d - \int_0^t (Z^n_s)^{-1}\sum_{j=1}^q (\dot{f}^n_{s,j})^2 ds - \int_0^t (Z^n_s)^{-1}\sum_{j=0}^q \dot{f}^n_{s,j}dY^j_s$$

and that

$$U^n_t = Z^n_t\Big\{\int_0^t (Z^n_s)^{-1}dG^n_s - \int_0^t (Z^n_s)^{-1}\sum_{j=1}^q (\dot{f}^n_{s,j})^2(X^n_s - X^n_{\eta_n(s)})\,ds\Big\}.$$

We define $Z_t = D_x X_t$ and therefore we have that it satisfies

$$Z_t = I_d + \int_0^t \sum_{j=0}^q \dot{f}_{s,j}\,dY^j_s.Z_s\,.$$

with $\dot{f}_{t,j} = \nabla f_j(X_t)$.

Furthermore Z_t^{-1} exists and satisfies the following explicit linear stochastic differential equation

$$(Z_t)^{-1} = I_d - \int_0^t (Z_s)^{-1} \sum_{j=1}^q (\dot{f}_{s,j})^2 ds - \int_0^t (Z_s)^{-1} \sum_{j=0}^q \dot{f}_{s,j} dY_s^j.$$

Then using the same techniques as in the proof of existence and uniqueness for stochastic differential equations with Lipschitz coefficients (i.e. Gronwall inequality) and its Malliavin derivatives (see e.g. Section 2.2.2 in Nualart (1995,2006)), we obtain that

Lemma 3.2. *For any $t \in [0,T]$ $Z_t^n,\ Z_t,\ (Z_t^n)^{-1},\ (Z_t)^{-1} \in (\mathbb{D}^\infty(W))^{d\times d}$*

$$\forall p \geq 1,\ l \geq 0 \qquad \lim_{n\to\infty} \sup_{0\leq t\leq T} \|Z_t^n - Z_t\|_{l,p} = 0,$$

and

$$\forall p \geq 1\ l \geq 0 \qquad \lim_{n\to\infty} \sup_{0\leq t\leq T} \Big\|(Z_t^n)^{-1} - (Z_t)^{-1}\Big\|_{l,p} = 0.$$

Furthermore $D^k_{t_1,\ldots,t_l} F(t)$ is $L^p(\Omega)$-continuous in $(t, t_1, \ldots., t_l)$ for $t_i \leq t$, $i = 1, \ldots, l$, $p > 1$ and any multi-index k with $|k| = l$ and for $F = Z,\ Z^n,\ (Z)^{-1},\ (Z^n)^{-1}$.

Now we are ready to give the main theorem in this section.

Theorem 3.1. *Let $(X_t)_{0\leq t\leq T}$ be a d-dimensional process solution of (4) satisfying the Hörmander condition and with density function p. We denote by X^n the Euler scheme associated to X and $Z_{n,\theta}$ the auxiliary noise introduced in (2).*

1. *Let ϕ be a super-kernel of order $s > 2$ satisfying the properties a) and b) of Definition 3.1. Then, there exists a constant $C^s_{\phi,x} > 0$ depending on ϕ, $p(x)$ and s such that*

$$\mathbb{E}\Big[\phi_{h,x}\big(X_T^n + Z_{n,\theta}\big)\Big] - p(x) = \frac{C^s_{\phi,x}}{n} + o\left(\frac{1}{n}\right), \quad \textit{with} \quad h = n^{-\alpha},\ \alpha \geq 1/s$$

2. *let $\varphi \in \mathcal{C}_b^\infty(\mathbb{R}^d;\mathbb{R})$ be a positive bounded and integrable function with bounded derivatives. Suppose that $\int_{\mathbb{R}^d} \varphi(x)dx = 1$. Let*

$$\varphi_{h,x}(y) = \frac{1}{h^d}\varphi\Big(\frac{y-x}{h}\Big), \quad h = n^{-\alpha} \quad \textit{with} \quad \alpha > 0,$$

then we have

$$\lim_{n\to 0} \mathbb{E}\,\varphi_{h,x}(X_T^n + Z_{n,\theta}) = p(x).$$

Proof. First we give the proof of the first assertion. We write the weak approximation error as follows

$$\begin{aligned}\mathbb{E}\Big[\phi_{h,x}\big(X_T^n + Z_{n,\theta}\big)\Big] - p(x) =& \mathbb{E}\Big[\phi_{h,x}\big(X_T^n + Z_{n,\theta}\big)\Big] - \mathbb{E}\Big[\phi_{h,x}\big(X_T + Z_{n,\theta}\big)\Big]\\ &+ \mathbb{E}\Big[\phi_{h,x}\big(X_T + Z_{n,\theta}\big)\Big] - \mathbb{E}\Big[\phi_{h,x}\big(X_T\big)\Big]\\ &+ \mathbb{E}\Big[\phi_{h,x}\big(X_T\big)\Big] - p(x).\end{aligned}$$

• Step 1:

We study the last term given by: $\mathbb{E}\Big[\phi_{h,x}\big(X_T\big)\Big] - p(x)$. Using Lemma 3.1 for the function p (we recall that under our hypothesis this is a $\mathcal{C}_b^\infty$ function) we obtain

$$\mathbb{E}\Big[\phi_{h,x}\big(X_T\big)\Big] - p(x) = \frac{h^s}{s!} \sum_{|\beta|=s} c_\beta(\phi)\partial^\beta p(x) + o(h^s),$$

where $\partial^\beta p$ is the partial derivative of p corresponding to the multi-index β. Note that for $h = n^{-\alpha}$, $\alpha \geq 1/s$ we have $o(h^s) = o(1/n)$.

• Step 2:

The second term is given by: $\mathbb{E}\Big[\phi_{h,x}\big(X_T + Z_{n,\theta}\big)\Big] - \mathbb{E}\Big[\phi_{h,x}\big(X_T\big)\Big]$. Using Taylor's expansion, we have

$$\mathbb{E}\Big[\phi_{h,x}\big(X_T + Z_{n,\theta}\big)\Big] - \mathbb{E}\Big[\phi_{h,x}\big(X_T\big)\Big] = \frac{1}{2}\,\mathbb{E}\int_0^1 (1-\lambda)\big(Z_{n,\theta}.\nabla\big)^2\phi_{h,x}\big(X_T + \lambda Z_{n,\theta}\big)\,d\lambda.$$

Since $Z_{n,\theta}$ and X are independent we obtain, after applying the integration by parts formula $d+2$ times, that

$$\begin{aligned}\mathbb{E}\Big[\big(Z_{n,\theta}.\nabla\big)^2\phi_{h,x}\big(X_T + \lambda Z_{n,\theta}\big)\Big] &= \mathbb{E}\int_{\mathbb{R}^d}\big(Z_{n,\theta}.\nabla\big)^2\phi_{h,x}\big(y + \lambda Z_{n,\theta}\big)p(y)\,dy \\ &= \mathbb{E}\int_{\mathbb{R}^d}\psi_{h,x}\big(y + \lambda Z_{n,\theta}\big)\big(Z_{n,\theta}.\nabla\big)^2\partial^{(1,\ldots,d)}p(y)\,dy\end{aligned}$$

where $\psi_{h,x}(y) := \int_{\prod_{i=1}^d(-\infty,y_i)}\phi_{h,x}(t)\,dt$. Since $\psi_{h,x}$ is bounded we obtain that

$$\Big|\mathbb{E}\int_{\mathbb{R}^d}\phi_{h,x}\big(y + \lambda Z_{n,\theta}\big)\big(Z_{n,\theta}.\nabla\big)^2 p(y)\,dy\Big| \leq \frac{c}{n^{1+2\theta}}.$$

The last inequality is immediate using the definition of $Z_{n,\theta}$ and that $\nabla^2\partial^{(1,\ldots,d)}p$ is integrable, since p decreases exponentially fast (see Corollary 3.25 in Kusuoka and Stroock (1985)). The result follows.

• Step 3:

Now we deal with the first term given by

$$A_n = \mathbb{E}\Big[\phi_{h,x}\big(X_T^n + Z_{n,\theta}\big)\Big] - \mathbb{E}\Big[\phi_{h,x}\big(X_T + Z_{n,\theta}\big)\Big].$$

In fact, we have

$$A_n = \int_0^1 \mathbb{E}\Big(\nabla\phi_{h,x}\big(X_T^{n,\lambda} + Z_{n,\theta}\big).U_T^n\Big)d\lambda,$$

where $X_T^{n,\lambda} = X_T + \lambda(X_T^n - X_T)$. In what follows we use the ideas contained in Clement et al. (2006). Recalling equations (7), (8) and (9) we have that

$$A_n = \sum_{j,k=0}^{q}\int_0^1 E\Big(\nabla\phi_{h,x}\big(X_T^{n,\lambda} + Z_{n,\theta}\big)Z_T^n\int_0^T (Z_s^n)^{-1}F_{jk}^n(s)(Y_s^j - Y_{\eta_n(s)}^j)dY_s^k\Big)d\lambda$$

where $F_{jk}^n(s) = \dot{f}_{s,j}^n\bar{f}_{s,k}^n$. If we define $D^0 = I$ (the identity operator) then using the duality formula (1) two times, one obtains

$$\begin{aligned}A_n &= \sum_{j,k=0}^{q}\int_0^1 E\Big(\int_0^T D_s^k\{\nabla\phi_{h,x}\big(X_T^{n,\lambda} + Z_{n,\theta}\big)Z_T^n\}(Z_s^n)^{-1}F_{jk}^n(s)\int_{\eta_n(s)}^s dY_u^j ds\Big)d\lambda \\ &= \sum_{j,k=0}^{q}\int_0^1 E\Big(\int_0^T\int_{\eta_n(s)}^s D_u^j\Big\{D_s^k\{\nabla\phi_{h,x}\big(X_T^{n,\lambda} + Z_{n,\theta}\big)Z_T^n\}(Z_s^n)^{-1}F_{jk}^n(s)\Big\}duds\Big)d\lambda.\end{aligned}$$

Next, if we apply the stochastic derivative operators one obtains that the above is a sum of terms of the type

$$E\Big(\int_0^1\int_0^T\int_{\eta_n(s)}^s \partial^r\phi_{h,x}\big(X_T^{n,\lambda} + Z_{n,\theta}\big)G_{u,s}^{n,r,j,k}dudsd\lambda\Big), \tag{10}$$

where $j,\ k=0,...,q$ and r is a multi-index of order 1 up to order 3. The random variables $G^{n,r,j,k}_{u,s}$ are given by

$$\begin{aligned}
&\left(D^j_u\left\{D^k_s\{Z^n_T\}\right\}(Z^n_s)^{-1}F^n_{jk}(s)+D^k_s\{Z^n_T\}D^j_u\left\{(Z^n_s)^{-1}F^n_{jk}(s)\right\}\right)^a \text{ if } r=(a)\\
&(D^k_s\{X^{n,\lambda}_T\})^a\left(D^j_u\{Z^n_T\}(Z^n_s)^{-1}F^n_{jk}(s)\right)^b+(D^k_s\{X^{n,\lambda}_T\})^a\left(Z^n_T D^j_u\{(Z^n_s)^{-1}F^n_{jk}(s)\}\right)^b+\\
&(D^j_u\{X^{n,\lambda}_T\})^a\left(D^k_s\{Z^n_T\}(Z^n_s)^{-1}F^n_{jk}(s)\right)^b \text{ if } r=(a,b)\\
&(D^j_u\{X^{n,\lambda}_T\})^a(D^k_s\{X^{n,\lambda}_T\})^b(Z^n_T(Z^n_s)^{-1}F^n_{jk}(s))^c \text{ if } r=(a,b,c).
\end{aligned}$$

Here $a,\ b,\ c\in\{1,...,d\}$ denote the component of the corresponding vector. Next for each term one applies the integration by parts formula (3) to obtain that each term of the type (10) can be written as

$$B_n(r,j,k):=E\Big(\int_0^1\int_0^T\int_{\eta_n(s)}^s\psi_{h,x}\big(X^{n,\lambda}_T\big)\tilde{H}_{r+}(X^{n,\lambda}_T+Z_{n,\theta},G^{n,r,j,k}_{u,s})dudsd\lambda\Big),$$

where $r+=(r,1,...,d)$ and $\psi_{h,x}(y):=\int_{\prod_{i=1}^d(-\infty,y_i)}\phi_{h,x}(t)\,dt$.

The proof of $A_n=C/n+o(1/n)$ follows using the following two lemmas. The proof of the first one is just a straightforward analysis exercise and the second lemma is proved in the appendix.

Lemma 3.3. *Let $g,g_n:\{(u,s)\in[0,T]^2;\ u\le s\}\to\mathbb{R},\ \ n\in\mathbb{N}$. Suppose that*

i) g is continuous on the compact set $\{(u,s)\in[0,T]^2;\ u\le s\}$.

ii) $\sup\limits_{0\le u\le s\le T}|g_n(u,s)-g(u,s)|\underset{n\to\infty}{\longrightarrow}0.$

Then

$$\int_0^T\int_{\eta_n(s)}^s g_n(u,s)\,du\,ds=\frac{1}{2n}\int_0^T g(s,s)\,ds+o(1/n).$$

Lemma 3.4. *Under the previous notations we obtain*

$$B_n(r,j,k)=\frac{1}{2n}\int_0^T\mathbb{E}\Big(\mathbf{1}_{\{X_T>x\}}\,\mathbf{H}_{r+}\big(X_T,G^{r,j,k}_{s,s}\big)\Big)\,ds+o\left(\frac{1}{n}\right),\tag{11}$$

with $G^{r,j,k}_{u,s}$ is the limit process given by (here $F_{jk}(s)=\dot{f}_{s,j}f_k(X_t)$)

$$\begin{aligned}
&\left(D^j_u\left\{D^k_s\{Z_T\}\right\}(Z_s)^{-1}F_{jk}(s)+D^k_s\{Z_T\}D^j_u\left\{(Z_s)^{-1}F_{jk}(s)\right\}\right)^a \textit{ if } r=(a)\\
&(D^k_s\{X_T\})^a\left(D^j_u\{Z_T\}(Z_s)^{-1}F_{jk}(s)\right)^b+(D^k_s\{X_T\})^a\left(Z_T D^j_u\{(Z_s)^{-1}F_{jk}(s)\}\right)^b+\\
&(D^j_u\{X_T\})^a\left(D^k_s\{Z_T\}(Z_s)^{-1}F_{jk}(s)\right)^b \textit{ if } r=(a,b)\\
&(D^j_u\{X_T\})^a(D^k_s\{X_T\})^b(Z_T(Z_s)^{-1}F_{jk}(s))^c \textit{ if } r=(a,b,c).
\end{aligned}$$

The proof of the second assertion of the Theorem follows as the first assertion with the exception that the rate is not $1/n$ but $1/n^{2\alpha}$ if $\alpha<1/2$. We mention here that in the proof of the third step above we only need the integrability of ϕ and that $\int_{\mathbb{R}^d}\phi(x)dx=1$. Consequently, the results obtained in this step remain valid in the context of the second assertion of the theorem. □

4 Approximations of non-degenerated diffusions through the Monte Carlo method

Let X be a hypoelliptic diffusion, solution of the stochastic differential equation (4). The goal of this section is to study an approximation of the density $p(x)$ of $X(T)$ using a Monte Carlo method together with a kernel density estimate. That is, in order to evaluate $p(x)$:

• One discretizes the diffusion X through an Euler scheme X^n of step T/n regularized as $X^n + Z_{n,\theta}$ where $Z_{n,\theta}$ is an independent Gaussian random variable of mean zero and standard deviation $n^{-1/2-\theta}$.
• one approximates the distribution $y \mapsto \delta_x(y)$ by the super-kernel $\phi_{h,x}(y)$ of order s , where h denotes the window size.
• then finally one estimates $\mathbb{E}\,\phi_{h,x}(X_T^n + Z_{n,\theta})$ using the Monte Carlo method. This procedure gives the classical kernel estimator given by

$$S^{n,N} := \frac{1}{N}\sum_{i=1}^{N} \phi_{h,x}(X_{T,i}^n + Z_{n,\theta}^i)$$

where $(X_{T,i}^n)_{1\le i\le N}$ and $(Z_{n,\theta}^i)_{1\le i\le N}$ are i.i.d. copies of X_T^n and $Z_{n,\theta}$. In what follows, we prove a central limit theorem analogue to a similar result proved by Duffie and Glynn (1995) which gives a precise choice for the sample size N for the Monte Carlo method. Their choice depends on the step size parameter n from the Euler scheme and is valid for the case where instead of $\phi_{h,x}$ one has a smooth function uniformly in h. Here we extend this result to the degenerate case. That is $\phi_{h,x}$ tends to the delta distribution function as $h \to 0$. This problem is somewhat more complex as we have to decide the optimal values of N and h in function of n.

In what follows we let $N = n^\gamma$, $h = n^{-\alpha}$ where $\gamma > 0$ and $\alpha \ge 1/s$

Theorem 4.1. *With the previous definitions and if we let $\gamma = 2 + \alpha d$ then*

$$n(S^{n,N} - p(x)) \Rightarrow \sigma G + C_{\phi,x}^s$$

with $\sigma^2 = \phi^2\, p(x)$, G is a standard Gaussian random variable and $C_{\phi,x}^s$ is the constant in the error expansion given in Theorem 3.1 and $\phi^2 = \int_{\mathbb{R}^d} |\phi(u)|^2\, du$.

Proof. We have that

$$n(S^{n,N} - p(x)) = \frac{1}{n^{\gamma-1}}\sum_{i=1}^{n^\gamma}\Big\{\phi_{h,x}(X_{T,i}^n + Z_{n,\theta}^i) - \mathbb{E}\big[\phi_{h,x}(X_T^n + Z_{n,\theta})\big]\Big\}$$
$$+\, n\Big[\mathbb{E}\big[\phi_{h,x}(X_T^n + Z_{n,\theta})\big] - p(x)\Big].$$

From Theorem 3.1, we have that

$$n\Big[\mathbb{E}\big[\phi_{h,x}(X_T^n + Z_{n,\theta})\big] - p(x)\Big] \underset{n\to\infty}{\longrightarrow} C_{\phi,x}^s$$

Therefore it remains to prove a central limit theorem for $\frac{1}{n^{\gamma-1}}\sum_{i=1}^{n^\gamma}\zeta_i^n$ where

$$\zeta_i^{n,h} := \Big\{\phi_{h,x}(X_{T,i}^n + Z_{n,\theta}^i) - \mathbb{E}\big[\phi_{h,x}(X_T^n + Z_{n,\theta})\big]\Big\}.$$

We start considering the characteristic function of the previous sum

$$\mathbb{E}\left[\exp\Big(\frac{iu}{n^{\gamma-1}}\sum_{k=1}^{n^\gamma}\zeta_k^n\Big)\right] = \left[\mathbb{E}\exp\Big(\frac{iu\zeta^n}{n^{\gamma-1}}\Big)\right]^{n^\gamma}$$
$$= \left[1 + \frac{1}{n^\gamma}\Big(\frac{-u^2}{2n^{\gamma-2}}\mathbb{E}\,|\zeta^n|^2 + \mathbb{E}\,C_n(\omega)\Big)\right]^{n^\gamma}.$$

Here

$$|\mathbb{E}\,C_n(\omega)| \le \frac{u^3}{6n^{2\gamma-3}}\mathbb{E}\,|\zeta^n|^3.$$

To study the above terms recall from definition 3.1, the definition of $\phi^l_{h,x}$ for $l=2,3$.

Therefore from the second assertion of Theorem 3.1 we have

$$\mathbb{E}\Big[\phi^i_{h,x}(X^n_T+Z_{n,\theta})\Big]=p(x)+\varepsilon_{i,n}(x),\quad i=2,3.$$

with $\lim_n \varepsilon_{i,n}(x)=0$ for $i=2,3$.

Let us start studying the term given by $\mathbb{E}\,|\zeta^{^n}|^2$. We have that

$$\begin{aligned}\mathbb{E}\,|\zeta^{^n}|^2&=\mathbb{E}\left[\phi_{h,x}(X^n_T+Z_{n,\theta})^2\right]-\Big\{\mathbb{E}\left[\phi_{h,x}(X^n_T+Z_{n,\theta})\right]\Big\}^2\\&=\frac{\phi^2}{h^d}\mathbb{E}\left[\phi^2_{h,x}(X^n_T+Z_{n,\theta})\right]-\Big\{\mathbb{E}\left[\phi_{h,x}(X^n_T+Z_{n,\theta})\right]\Big\}^2.\end{aligned}$$

Therefore,

$$\mathbb{E}\,|\zeta^{^n}|^2=\frac{\phi^2}{h^d}\varepsilon_2(x)+\frac{\phi^2}{h^d}p(x)+\Big\{\frac{C^s_{\phi,x}}{n}+o\Big(\frac{1}{n}\Big)+p(x)\Big\}^2$$

where $C^s_{\phi,x}$ is the constant in the error expansion given in Theorem 3.1. Therefore, for $h=n^{-\alpha}$, $\gamma=2+\alpha d$ and $\alpha\geq 1/s$ we have

$$\frac{1}{n^{\gamma-2}}\mathbb{E}\,|\zeta^{^n}|^2\underset{n\to\infty}{\longrightarrow}\phi^2\,p(x).$$

On the other hand, we have that

$$\begin{aligned}\mathbb{E}\,|\zeta^{^n}|^3&=\mathbb{E}\,\Big|\phi_{h,x}(X^n_T+Z_{n,\theta})-\mathbb{E}\big[\phi_{h,x}(X^n_T+Z_{n,\theta})\big]\Big|^3\\&\leq\mathbb{E}\,\big|\phi_{h,x}(X^n_T+Z_{n,\theta})\big|^3+3\mathbb{E}\,\big|\phi_{h,x}(X^n_T+Z_{n,\theta})\big|^2\big|\mathbb{E}\,\phi_{h,x}(X^n_T+Z_{n,\theta})\big|\\&+4\big|\mathbb{E}\,\phi_{h,x}(X^n_T+Z_{n,\theta})\big|^3.\end{aligned}$$

Therefore, as before, we obtain that

$$\begin{aligned}\mathbb{E}\,|\zeta^{^n}|^3\leq\,&h^{-2d}\phi^3\mathbb{E}\,\phi^3_{h,x}(X^n_T+Z_{n,\theta})+3h^{-d}\phi^2\mathbb{E}\,\phi^2_{h,x}(X^n_T+Z_{n,\theta})\big|\mathbb{E}\,\phi_{h,x}(X^n_T+Z_{n,\theta})\big|\\&+4\big|\mathbb{E}\,\phi_{h,x}(X^n_T+Z_{n,\theta})\big|^3\end{aligned}$$

Finally,

$$\begin{aligned}\mathbb{E}\,|\zeta^{^n}|^3\leq\,&h^{-2d}\phi^3(p(x)+\varepsilon_{3,n}(x))+3h^{-d}\phi^2\Big|p(x)+\varepsilon_{2,n}(x)\Big|\Big|\frac{C^s_{\phi,x}}{n}+o\Big(\frac{1}{n}\Big)+p(x)\Big|\\&+4\Big|\frac{C^s_{\phi,x}}{n}+o\Big(\frac{1}{n}\Big)+p(x)\Big|^3.\end{aligned}$$

for $h=n^{-\alpha}$, $\gamma=2+\alpha d$ and $\alpha\geq 1/s$. This leads to

$$\frac{1}{n^{2\gamma-3}}\mathbb{E}\,|\zeta^{^n}|^3\underset{n\to\infty}{\longrightarrow}0.$$

which finishes the proof. □

The interpretation of the above result leads to the previously announced result. That is, in order to approximate the density $p(x)$ through a Monte Carlo method with a tolerance error of order $1/n$, the optimal asymptotic choice of parameters are $h=n^{-\alpha}$ and $N=n^{2+\alpha d}$ with $\alpha\geq 1/s$ where s

denotes the order of the super kernel used for the estimation. This leads to the following algorithmic complexity (that is, number of calculations) of

$$C_{MC} = C \times nN = C \times n^{3+\alpha d},$$

for a given $C > 0$ (here the unit of calculation is one simulation of a random variable). Therefore the optimal complexity of this algorithm is given by

$$C^{\star}_{MC} = C \times n^{3+\frac{d}{s}}.$$

Therefore we conclude that if the order s of the kernel is bigger then the complexity is smaller.

Nevertheless, one should keep in mind that the constant $C^{s}_{\phi,p(x)}$ depends on s and the implementation of this algorithm for high order kernels carries some problems, such as the possibility of non-positive values for $S^{n,N}$ and big constants in the error expansions. Therefore the practical choice of super kernel remains an open problem from the practical point of view.

5 Asymptotic behavior of the Malliavin derivative of the normalized error

5.1 Malliavin derivative of the error process

In the following we denote $\check{W}^n$ the $d \times d$-dimensional process defined by

$$\check{W}^{n,ij}_t = \sqrt{\frac{2}{T}} \int_0^t (W^i_s - W^i_{\eta_n(s)})\, dW^j_s.$$

According to the Theorem 3.2 of Jacod and Protter (1998), the process $\sqrt{n}\check{W}^n$ converge stably in law to a $d \times d$-dimensional Brownian motion $\check{W}$ independent from W and the couple $\sqrt{n}(\check{W}^n, U^n)$ converge stably in law to the couple $(\check{W}, U)$ where the $\mathbb{R}^{d\times d}$-valued process U is solution to

$$U_t = \sum_{j=0}^{q} \int_0^t \dot{f}_{s,j}.U_s\, dY^j_s + \sqrt{\frac{T}{2}} \sum_{i,j=1}^{q} \int_0^t \dot{f}_{s,j}.f_i(X_s)\, d\check{W}^{ij}_s. \tag{12}$$

In order to obtain the equation satisfied by the Malliavin derivative of the error process with respect to W^i, $i = 1, ..., q$, we derive the equation (7):

$$D^i_s U^n_t = \sum_{j=0}^{q} \int_s^t D^i_s(\dot{f}^n_{j,v}.U^n_v)\, dY^j_v + \dot{f}^n_{s,i} U^n_s \mathbf{1}_{\{s\le t\}} + D^i_s G^n_t. \tag{13}$$

Note that the above derivative exists due to the regularity properties of the coefficients of the equation for X. Furthermore, using (8) and (9), we have that

$$D^i_s G^n_t = \dot{f}^n_{s,i}.(X^n_s - X^n_{\eta_n(s)})\mathbf{1}_{\{s\le t\}} + \sum_{j=0}^{q} \int_s^t D^i_s\Big[\dot{f}^n_{u,j}.(X^n_u - X^n_{\eta_n(u)})\Big]\, dY^j_u$$

and

$$D^i_s\Big[\dot{f}^n_{u,j}.(X^n_u - X^n_{\eta_n(u)})\Big] = \sum_{k=0}^{q} D^i_s(\dot{f}^n_{u,j}.\bar{f}^n_{u,k})(Y^k_u - Y^k_{\eta_n(u)}) + \dot{f}^n_{u,j}.\bar{f}^n_{u,i}\mathbf{1}_{\{\eta_n(u)\le s\le u\}}$$

As $D_s Z = 0$ for Z, which is $\mathcal{F}_u$-measurable ($u < s$), the relation (13) becomes for $s \le t$,

$$D^i_s U^n_t = \dot{f}^n_{s,i}(U^n_s + X^n_s - X^n_{\eta_n(s)}) + \sum_{j=0}^{q} \int_s^t \dot{f}^n_{v,j} D^i_s U^n_v\, dW^j_v + \tilde{G}^{n,i}_{s,t}, \tag{14}$$

with

$$\tilde{G}^{n,i}_{s,t} = \sum_{j=0}^{q} \int_s^t D^i_s \dot{f}^n_{v,j}\, U^n_v \, dY^j_v + \sum_{j,k=0}^{q} \int_s^t D^i_s(\dot{f}^n_{u,j}\, \bar{f}^n_{u,k})(Y^k_u - Y^k_{\eta_n(u)})\, dY^j_u$$
$$+ \sum_{j=0}^{q} \int_s^t \dot{f}^n_{u,j}\, \bar{f}^n_{u,i} \mathbf{1}_{\{\eta_n(u) \le s \le u\}}\, dY^j_u. \quad (15)$$

From Theorem 56 p.271 in Protter (1990), it follows that (14) becomes for $t \geq s$,

$$D^i_s U^n_t = Z^n_t (Z^n_s)^{-1} \dot{f}^n_{s,i}(U^n_s + X^n_s - X^n_{\eta_n(s)})$$
$$+ Z^n_t \Big\{ \int_s^t (Z^n_u)^{-1}\, d\tilde{G}^{n,i}_{s,u} - \sum_{j=0}^{q} \int_s^t (Z^n_u)^{-1} \dot{f}^n_{u,j} d\langle \tilde{G}^{n,i}_{s,.}, Y^j \rangle_u \Big\}. \quad (16)$$

5.2 Convergence in law for the normalized Malliavin derivative of the error

The Malliavin derivative of U^n_T is a random vector taking values in the Hilbert space $H = L^2([0,T])$. The aim of this section is to establish the convergence in law for the sequence $\sqrt{n} D U^n_T$. Note that the process U, limit of $\sqrt{n} U^n$, is an adapted process with respect to the filtration of W and $\check{W}$. Using (12), we can compute the derivatives DU_t with respect to the Wiener processes W and obtain that DU_t satisfies for $0 \leq s \leq t \leq T$,

$$D^i_s U_t = \dot{f}_{s,i} U_s + \sum_{j=0}^{q} \int_s^t \dot{f}_{v,j} D^i_s U_v \, dY^j_v + \sum_{j=0}^{q} \int_s^t D^i_s \dot{f}_{v,i} U_v \, dY^j_v$$
$$+ \sqrt{\frac{T}{2}} \sum_{j,k=1}^{q} \int_s^t D^i_s(\dot{f}_{v,j} f_{v,k}) d\check{W}^{kj}_v, \quad (17)$$

or using again Theorem 56 p.271 in Protter (1990), we obtain for $0 \leq s \leq t \leq T$ that,

$$D_s U_t = Z_t (Z_s)^{-1} \dot{f}_{s,i} U_s + Z_t \Big\{ \int_s^t (Z_u)^{-1} dG^i_{s,u} - \sum_{j=0}^{q} \int_s^t (Z_u)^{-1} \dot{f}_{u,j}\, d\langle G^i_{s,.}, Y^j \rangle_u \Big\}, \quad (18)$$

with

$$G^i_{s,t} = \sum_{j=0}^{q} \int_s^t D^i_s \dot{f}_{v,j} U_v \, dY^j_v + \sqrt{\frac{T}{2}} \sum_{j,k=1}^{q} \int_s^t D^i_s(\dot{f}_{v,j} f_{v,i})\, d\check{W}^{kj}_v. \quad (19)$$

Theorem 5.1. *Let $(H^i_t)_{0 \leq t \leq T}$ be a continuous sequence of $\mathbb{R}$-valued process (possibly non adapted). The random vector $(\sqrt{n} U^n_T, \sqrt{n} \int_0^T H^i_s D^i_s U^n_T \, ds)$ converges stably in law to $(U_T, \int_0^T H^i_s D^i_s U_T \, ds)$ where $D^i U_T$ is the Malliavin derivative of U with respect to W^i and solution of (18).*

In order to prove this theorem, we use the two technical lemmas below. The proofs of these lemmas are given in the Appendix 9 (See Jacod and Protter (1998) for related results).

Lemma 5.1. *Let $(H^n_t = (H^{1,n}_t, ..., H^{d,n}_t))_{0 \leq t \leq T}$ be a tight sequence of continuous processes (possibly non adapted) taking values in $\mathbb{R}^d$. The sequence of random vectors $(\sqrt{n} \int_0^T H^{i,n}_s (Y^j_s - Y^j_{\eta_n(s)})\, ds; i \in \{1, ..., d\}, j \in \{0, ..., q\})_{n \in \mathbb{N}}$ converge in probability to 0.*

Lemma 5.2. *Let $(H_t)_{0 \leq t \leq T}$ be a continuous $\mathbb{R}$-valued process (possibly non adapted) and let $(K^n_u)_{0 \leq u \leq T}$ be a sequence of adapted and continuous processes taking values in $\mathbb{R}^d$ and such that $\sup_n \mathbb{E} \int_0^T \|K^n_u\|^2 \, du < \infty$. Then the sequence $(\sqrt{n} \int_0^T H_s (\int_0^T \mathbf{1}_{\{\eta_n(u) \leq s \leq u\}} K^n_u \, dY^j_u) ds)_{n \in \mathbb{N}}$ converge in probability to 0.*

Lemma 5.3. *Let H^i, K^i, L^i be three real processes with continuous trajectories on $[0,T]$ and let $(\xi^{ij}_{s,u})_{0\le s\le u\le T}$, $(\zeta^{ijk}_{s,u})_{0\le s\le u\le T}$ be two adapted processes (wrt u), taking values in $\mathbb{R}^{d\times d}$, with continuous trajectories and such that*

$$\mathbb{E}\int_0^T\int_0^u\Big(\max_j\|\xi^{ij}_{s,u}\|^p+\max_{j,k}\|\zeta^{ijk}_{s,u}\|^p\Big)dsdu<\infty \qquad \textit{for} \quad p>2,\ i=1,...,q.$$

Then

$$\sqrt{n}\Big(U^n_T,\int_0^T H^i_sU^n_s\,ds,\int_0^T K^i_s\Big(\sum_{j=1}^q\int_s^T\xi^{ij}_{s,u}U^n_u\,dW^j_u\Big)ds,$$
$$\int_0^T L^i_s\Big(\sum_{j,k=1}^q\int_s^T\zeta^{ijk}_{s,u}\,d\check{W}^{n,kj}_u\Big)ds;i=1,...,q\Big)$$

stably converge in law to

$$\Big(U_T,\int_0^T H^i_sU_s\,ds,\int_0^T K^i_s\Big(\sum_{j=1}^q\int_s^T\xi^{ij}_{s,u}U_u\,dW^j_u\Big)ds,$$
$$\int_0^T L^i_s\Big(\sum_{j,k=1}^q\int_s^T\zeta^{ijk}_{s,u}\,d\check{W}^{kj}_u\Big)ds,i=1,...,q\Big)$$

Proof of Theorem 5.1. Using the relation (16), we have

$$\int_0^T H^i_sD^i_sU^n_T\,ds=Z^n_T\int_0^T H^i_s(Z^n_s)^{-1}\dot f^n_{s,i}(U^n_s+X^n_s-X^n_{\eta_n(s)})\,ds+Z^n_T\,I^{n,i}, \tag{20}$$

where

$$I^{n,i}=\int_0^T H^i_s\Big(\int_s^T(Z^n_u)^{-1}d\tilde G^{n,i}_{s,u}-\sum_{j=0}^q\int_s^T(Z^n_u)^{-1}\dot f^n_{u,j}\,d\langle\tilde G^{n,i}_{s,.},Y^j\rangle_u\Big)\,ds.$$

Using (18)

$$\int_0^T H^i_sD^i_sU_T\,ds=Z_T\int_0^T H^i_s(Z_s)^{-1}\dot f_{s,i}U_s\,ds+Z_TI^i$$

with

$$I^i=\int_0^T H^i_s\Big(\int_s^T(Z_u)^{-1}dG^i_{s,u}-\sum_{j=0}^q\int_s^T(Z_u)^{-1}\dot f_{u,j}\,d\langle G^i_{s,.},Y^j\rangle_u\Big)\,ds.$$

Now, let us prove that

$$\int_0^T H^i_s(Z^n_s)^{-1}\dot f^n_{s,i}(U^n_s+X^n_s-X^n_{\eta_n(s)})\,ds=\int_0^T H^i_s(Z_s)^{-1}\dot f_{s,i}U^n_s\,ds+\xi^{n,i} \tag{21}$$

with $\mathbb{P}\lim_{n\to\infty}\big(\sqrt{n}\xi^{n,i}\big)=0$, where we use the notation $\mathbb{P}\lim$ for convergence in probability.

In fact, the tightness of $\sqrt{n}U^n$ (see Theorem 3.2 of Jacod and Protter (1998)) and

$$\mathbb{P}\lim_{n\to\infty}\sup_{0\le s\le T}|(Z^n_s)^{-1}\dot f^n_{s,j}-(Z_s)^{-1}\dot f_{s,j}|=0$$

give that

$$\mathbb{P}\lim_{n\to\infty}\sqrt{n}\int_0^T H^i_s\big[(Z^n_s)^{-1}\dot f^n_{s,i}-(Z_s)^{-1}\dot f_{s,i}\big]U^n_s\,ds=0.$$

In the other hand, we can write

$$\sqrt{n}\int_0^T H^i_s(Z^n_s)^{-1}\dot f^n_{s,i}(X^n_s-X^n_{\eta_n(s)})\,ds=\sqrt{n}\sum_{j=0}^q\int_0^T H^i_s(Z^n_s)^{-1}\dot f^n_{s,i}\bar f^n_{s,j}(Y^j_s-Y^j_{\eta_n(s)})\,ds.$$

Then as $H^i_\cdot(Z^n_\cdot)^{-1}\dot f^n_{\cdot,i}\bar f^n_{\cdot,j}$ is a tight sequence of continuous process we obtain from Lemma 5.1 that the above integral converges to zero in probability and therefore $\mathbb{P}\lim_{n\to\infty}(\sqrt{n}\xi^{n,i})=0$.

Let's study now the sequence (I^n). First, note that using (15) we obtain that

$$I^{n,i}=\sum_{j=0}^{q}\int_0^T H^i_s\int_s^T(Z^n_u)^{-1}\Big\{A^{n,i,j}_{s,u}U^n_u+\sum_{k=0}^{q}B^{n,i,j,k}_{s,u}(Y^k_u-Y^k_{\eta(u)})+C^{n,i,j}_u\mathbf{1}_{\{\eta_n(u)\le s\le u\}}\Big\}dY^j_u ds \quad (22)$$

where

$$\begin{aligned}
A^{n,i,j}_{s,u} &= D^i_s\dot f^n_{u,j}-\mathbf{1}_{\{j=0\}}\sum_{l=1}^{q}\dot f^n_{u,l}D^i_s\dot f^n_{u,l}\\
B^{n,i,j,k}_{s,u} &= D^i_s(\dot f^n_{u,j}\bar f^n_{u,k})-\mathbf{1}_{\{k=0\}}\sum_{l=1}^{q}\dot f^n_{u,l}D^i_s(\dot f^n_{u,l}\bar f^n_{u,k})\\
C^{n,i,j}_u &= \dot f^n_{u,j}\bar f^n_{u,i}-\mathbf{1}_{\{j=0\}}\sum_{l=1}^{q}\dot f^n_{u,l}\dot f^n_{u,l}\bar f^n_{u,i}.
\end{aligned}$$

Now we study each of the three terms in (22). First, the third term in (22) satisfies that

$$\sum_{j=0}^{q}\sqrt{n}\int_0^T H^i_s\int_s^T(Z^n_u)^{-1}C^{n,i,j}_u\mathbf{1}_{\{\eta_n(u)\le s\le u\}}dY^j_u ds$$

with $\sup_n\mathbb{E}\int_0^T\|(Z^n_u)^{-1}C^{n,i,j}_u\|^2du<\infty$. Therefore this term tends to zero due to Lemma 5.2. Now, consider the second term with $jk=0$. Note that

$$\sum_{j,k=0;jk=0}^{q}\sqrt{n}\int_0^T H^i_s\int_s^T(Z^n_u)^{-1}B^{n,i,j,k}_{s,u}(Y^k_u-Y^k_{\eta(u)})dY^j_u ds$$

tends to zero as $jk=0$. Next, if we define $B^{i,j,k}_{s,u}=D^i_s(\dot f_{u,j}f_{u,k})$

$$\sum_{j,k=1}^{q}\sqrt{n}\int_0^T H^i_s\int_s^T((Z^n_u)^{-1}B^{n,i,j,k}_{s,u}-(Z_u)^{-1}B^{i,j,k}_{s,u})(Y^k_u-Y^k_{\eta(u)})dY^j_u ds$$

tends to zero in $L^1(\Omega)$ as $\sqrt{n}\check W^{n,kj}$ is bounded uniformly in $L^p(\Omega)$ and $(Z^n_u)^{-1}B^{n,i,j,k}_{s,u}-(Z_u)^{-1}B^{i,j,k}_{s,u}$ converges to zero in $L^p(\Omega\times[0,T]^2)$, therefore this term also converges to zero. Then for the remaining

$$\sum_{j,k=1}^{q}\sqrt{\frac{nT}{2}}\int_0^T H^i_s\int_s^T(Z_u)^{-1}B^{i,j,k}_{s,u}d\check W^{n,kj}_u ds,$$

we will apply Lemma 5.3 at the end together with the analysis for the first term of (22). For that first term, consider as previously

$$\sum_{j=0}^{q}\sqrt{n}\int_0^T H^i_s\int_s^T((Z^n_u)^{-1}A^{n,i,j}_{s,u}-(Z_u)^{-1}A^{i,j}_{s,u})U^n_u dY^j_u ds,$$

where $A^{i,j}_{s,u}=D^i_s\dot f_{u,j}-\mathbf{1}_{\{j=0\}}\sum_{l=1}^{q}\dot f_{u,l}D^i_s\dot f_{u,l}$. Again this term goes to zero in $L^1(\Omega)$ as the sequence $\sqrt{n}U^n$ is bounded uniformly in $L^p(\Omega)$ and $(Z^n_u)^{-1}A^{n,i,j}_{s,u}-(Z_u)^{-1}A^{i,j}_{s,u}$ converges to zero in $L^p(\Omega\times[0,T]^2)$. Now the only terms that have been left are

$$\sum_{j,k=1}^{q}\sqrt{\frac{nT}{2}}\int_0^T H^i_s\int_s^T(Z_u)^{-1}B^{i,j,k}_{s,u}d\check W^{n,kj}_u ds+\sum_{j=1}^{q}\sqrt{n}\int_0^T H^i_s\int_s^T(Z_u)^{-1}A^{i,j}_{s,u}U^n_u dY^j_u ds.$$

For this term is clear how to define $\xi^{ij}_{s,u}$ and $\zeta^{ijk}_{s,u}$ which satisfy the required conditions in Lemma 5.3. Note that one has to apply Lemma 5.3 to the term above and the first term on the right of (21). Therefore the proof is finished. □

6 An optimal control variate method for density estimation

The aim of this section, is to analyze the statistical Romberg method as a control variate introduced in Kebaier (2005) in the case of density estimation. In order to reduce variance in the density estimation of a non-degenerate d-dimensional diffusion $(X_t)_{0\leq t\leq T}$, we will use another estimation of the same density using less steps and slightly more simulation paths.

That is, we discretize the diffusion by two Euler schemes with time steps T/n and T/m ($m << n$). Under the Hörmander condition, the statistical Romberg method approximates the density $p(x)$ of the diffusion $(X_t)_{0\leq t\leq T}$ by

$$\frac{1}{N_m}\sum_{i=1}^{N_m}\phi_{h,x}(\hat{X}^m_{T,i}+\hat{Z}^i_{m,\theta})+\frac{1}{N_{n,m}}\sum_{i=1}^{N_{n,m}}\left\{\phi_{h,x}(X^n_{T,i}+Z^i_{n,\theta})-\phi_{h,x}(X^m_{T,i}+Z^i_{m,\theta})\right\},$$

where $\hat{X}^m_T$ is a second Euler scheme with step T/m and such that the Brownian paths W, used for X^n_T and X^m_T have to be independent of the Brownian paths (denoted by $\hat{W}$) used in order to simulate $\hat{X}^m_T$. Furthermore

$$Z_{n,\theta}=\frac{\bar{W}_T}{n^{\frac{1}{2}+\theta}},\quad \hat{Z}_{m,\theta}=\frac{\breve{W}_T}{m^{\frac{1}{2}+\theta}},\quad \theta\geq 0,$$

where $\bar{W}$ and $\breve{W}$ are two independent d-dimensional Brownian motions independent of W and $\hat{W}$.

In order to run the statistical Romberg algorithm, we have to optimize the parameters in the method. In the same manner as in Kebaier (2005), we establish a central limit theorem which will lead to a precise description of how to choose the parameters N_m, $N_{n,m}$, m and h as functions of n. The essential difference with the problem studied in Kebaier (2005) is that the variance of the estimators explode. This issue will be resolved through an appropriate renormalization procedure and an appropriate decomposition of the derivatives of the kernel function.

In the following, we suppose that for a given $0<\beta<2/3$ we have

$$m=n^{\beta},\ N_m=n^{\gamma_1},\ \ N_{n,m}=n^{\gamma_2},\ \ h=n^{-\alpha},$$

where $\gamma_1,\gamma_2>0$, and $\alpha\geq 1/s$ (the parameter s denotes the order of the super-kernel ϕ). We set

$$V_n:=\frac{1}{n^{\gamma_1}}\sum_{i=1}^{n^{\gamma_1}}\phi_{h,x}(\hat{X}^{n^\beta}_{T,i}+\hat{Z}^i_{n^\beta,\theta})+\frac{1}{n^{\gamma_2}}\sum_{i=1}^{n^{\gamma_2}}\Big\{\phi_{h,x}(X^n_{T,i}+Z^i_{n,\theta})-\phi_{h,x}(X^{n^\beta}_{T,i}+Z^i_{n^\beta,\theta})\Big\}.$$

Note that the first derivatives of the kernel function ϕ have the following decomposition

$$\frac{\partial\phi}{\partial x_i}(x)=\phi^1_i(x)-\phi^2_i(x)$$

where, $\phi^1_i(x):=\big(\frac{\partial\phi}{\partial x_i}(x)\big)_+$ and $\phi^2_i(x):=\big(\frac{\partial\phi}{\partial x_i}(x)\big)_-$. We prefer this notation because it will be easier to handle in the coming calculations. The condition $\int_{\mathbb{R}^d}|\nabla\phi(x)|^2dx<\infty$ (see Definition 3.1) implies that $\int_{\mathbb{R}^d}|\phi^j_i(x)|^2dx<+\infty$, for $i=1,...,d$ and $j=1,2$. In the following we define the constant

$$C^{j\,j'}_{i\,i'}=\int_{\mathbb{R}^d}\phi^j_i(x)\phi^{j'}_{i'}(x)\,dx,\quad j,j'\in\{1,2\},\ i,i'\in\{1,...,d\}.$$

Theorem 6.1. *Let*

$$\tilde{\sigma}^2:=\sum_{j,j'=1}^{2}\sum_{i,i'=1}^{d}(-1)^{j+j'}C^{j\,j'}_{i\,i'}\Big\{\mathbb{E}\big[\delta_x(X_T)U^i_TU^{i'}_T\big]+T\delta_{ii'}p(x)\mathbf{1}_{\{\theta=0\}}\Big\},$$

where $\delta_x(.)$ stands for the Dirac delta function and $\delta_{ii'}$ is the Kroeneker delta function. Assume that $h=n^{-\alpha}$, $\gamma_1=2+\alpha d$, $\gamma_2=(d+2)\alpha+2-\beta$ and $1/s\leq\alpha<\beta/(d+2)$ with $0<\beta<2/3$.

Then

$$n\big(V_n-p(x)\big)\Rightarrow\tilde{\sigma}G+C^s_{\phi,x}$$

where G is a standard Gaussian r.v. and $C^s_{\phi,x}$ is the discretization constant of Theorem 3.1.

Before proving the theorem we introduce an essential result about the rate of explosion of the variances of the estimators. In what follows we extend the previous notation to

$$\phi^1_{i,h,x}(y) = \frac{1}{h^d}\phi^1_i(\frac{y-x}{h}), \qquad \phi^2_{i,h,x}(y) = \frac{1}{h^d}\phi^2_i(\frac{y-x}{h}).$$

Lemma 6.1. *Under the notation and assumptions of the above theorem, we have*

1. $n^{\beta-\alpha(2+d)}\mathbb{E}\left[\phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta}) - \phi_{h,x}(X_T)\right]^2 \underset{n\to\infty}{\longrightarrow} \tilde{\sigma}^2.$

2. $n^{\beta-\alpha(2+d)}\mathbb{E}\left[\phi_{h,x}(X_T^{n} + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right]^2 \underset{n\to\infty}{\longrightarrow} \tilde{\sigma}^2.$

We remark here that the assertion 1 above is satisfied also for $\beta \geq 2/3$.

Proof. As θ is constant throughout the proof we sometimes use some notations where θ is not explicitly written. Let's prove the first assertion of the lemma.
• Step 1:
The Taylor formula gives

$$\begin{aligned}\phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta}) - \phi_{h,x}(X_T) &= \sum_{i=1}^{d}\frac{\partial\phi_{h,x}}{\partial x_i}(X_T)\big(U_T^{n^\beta,i} + Z^i_{n^\beta,\theta}\big)\\ &\quad + \frac{1}{2}\sum_{i,i'=1}^{d}\frac{\partial^2\phi_{h,x}}{\partial x_i\partial x_{i'}}(\xi_T^{n,ii'})\big(U_T^{n^\beta,i} + Z^i_{n^\beta,\theta}\big)\big(U_T^{n^\beta,i'} + Z^{i'}_{n^\beta,\theta}\big)\end{aligned}$$

where $U_T^{n^\beta} = X_T^{n^\beta} - X_T$ and $\xi_T^n \in \prod_{i=1}^d [X_T^i, X_T^{n^\beta,i} + Z^i_{n^\beta,\theta}]$. Note that

$$\left\|\frac{\partial^2\phi_{h,x}}{\partial x_i\partial x_{i'}}\right\|_\infty \leq h^{-(d+2)}\|\phi''\|_\infty,$$

where $\|\phi''\|_\infty = \max_{i,i'}\sup_{x\in\mathbb{R}^d}\left|\frac{\partial^2\phi(x)}{\partial x_i\partial x_{i'}}\right|$. Then there exists a constant $C_T > 0$ such that

$$\begin{aligned}n^{\frac{\beta-\alpha(2+d)}{2}}\left\|\frac{\partial^2\phi_{h,x}}{\partial x_i\partial x_{i'}}(\xi_T^{n,ii'})\big(U_T^{n^\beta,i} + Z^i_{n^\beta,\theta}\big)\big(U_T^{n^\beta,i'} + Z^{i'}_{n^\beta,\theta}\big)\right\|_2 &\leq C_T n^{\frac{-\alpha(2+d)-\beta}{2}}h^{-(d+2)}\|\phi''\|_\infty\\ &= C_T n^{\frac{\alpha(2+d)-\beta}{2}}\|\phi''\|_\infty \longrightarrow 0 \quad \text{as} \quad n\to\infty.\end{aligned}$$

Consequently, in order to obtain the first assertion of the lemma it suffices to prove that

$$n^{\frac{\beta-\alpha(2+d)}{2}}\left\|\sum_{i=1}^{d}\frac{\partial\phi_{h,x}}{\partial x_i}(X_T)\big(U_T^{n^\beta,i} + Z^i_{n^\beta,\theta}\big)\right\|_2 \longrightarrow \tilde{\sigma} \quad \text{as} \quad n\to\infty. \tag{23}$$

• **Step 2:** We have

$$\begin{aligned}\left\|\sum_{i=1}^{d}\frac{\partial\phi_{h,x}}{\partial x_i}(X_T)\big(U_T^{n^\beta,i} + Z^i_{n^\beta,\theta}\big)\right\|_2^2 &= \sum_{i,i'=1}^{d}\mathbb{E}\Big\{\frac{\partial\phi_{h,x}}{\partial x_i}(X_T)\frac{\partial\phi_{h,x}}{\partial x_{i'}}(X_T)\big(U_T^{n^\beta,i} + Z^i_{n^\beta,\theta}\big)\big(U_T^{n^\beta,i'} + Z^{i'}_{n^\beta,\theta}\big)\Big\}\\ &= \sum_{j,j'=1}^{2}\sum_{i,i'=1}^{d}\mathbb{E}\Big\{\frac{(-1)^{j+j'}}{h^{2+d}}\phi^j_{i,h,x}(X_T)\phi^{j'}_{i',h,x}(X_T)Y_{ii'}^{n^\beta}\Big\},\end{aligned}$$

where $Y_{ii'}^{n^\beta} \equiv Y_{ii'}^{n^\beta}(\theta) := \big(U_T^{n^\beta,i} + Z^i_{n^\beta,\theta}\big)\big(U_T^{n^\beta,i'} + Z^{i'}_{n^\beta,\theta}\big)$ and

$$\frac{\partial\phi_{h,x}}{\partial x_i}(y) = \frac{1}{h}\big[\phi^1_{i,h,x}(y) - \phi^2_{i,h,x}(y)\big].$$

Then

$$\left\|\sum_{i=1}^{d}\frac{\partial\phi_{h,x}}{\partial x_i}(X_T)\big(U_T^{n^\beta,i}+Z^i_{n^\beta,\theta}\big)\right\|_2^2=\sum_{j,j'=1}^{2}\sum_{i,i'=1}^{d}\frac{(-1)^{j+j'}C^{j\,j'}_{i\,i'}}{h^{2+d}}\mathbb{E}\Big[\varphi^{j\,j'}_{i\,i',h,x}(X_T)Y^{n^\beta}_{ii'}\Big],\qquad(24)$$

where

$$\varphi^{j\,j'}_{i\,i',h,x}(y)=(C^{j\,j'}_{i\,i'})^{-1}h^d\phi^j_{i,h,x}(y)\phi^{j'}_{i',h,x}(y).$$

In order to evaluate the limit of the last quantity, one needs to use an integration by parts formula. We use the Malliavin-Thalmaier integration by parts formula introduced in Proposition 2.3. However, to apply this formula we first make appear a Dirac function inside the expectation in 24.

More precisely, let $(\xi^{j\,j'}_{i\,i',h})_{\{i,i'=1\ldots d,\ j,j'=1,2.\}}$ be random vectors independent of all other random variables, so that their density is given by $\varphi^{j\,j'}_{i\,i',h,x}(.)$. Consequently, we have that

$$\xi^{j\,j'}_{i\,i',h}\to x\ \text{ a.s. }\quad\text{as } h\to 0.$$

Then

$$n^{\beta-\alpha(2+d)}\left\|\sum_{i=1}^{d}\frac{\partial\phi_{h,x}}{\partial x_i}(X_T)\big(U_T^{n^\beta,i}+Z^i_{n^\beta,\theta}\big)\right\|_2^2=n^\beta\sum_{j,j'=1}^{2}\sum_{i,i'=1}^{d}(-1)^{j+j'}C^{j\,j'}_{i\,i'}\ \mathbb{E}\Big[\delta_0\big(X_T-\xi^{j\,j'}_{i\,i',h}\big)Y^{n^\beta}_{ii'}\Big],$$

Applying the integration by parts formula in Proposition 2.3, we obtain

$$\begin{aligned}n^{\beta-\alpha(2+d)}&\left\|\sum_{i=1}^{d}\frac{\partial\phi_{h,x}}{\partial x_i}(X_T)\big(U_T^{n^\beta,i}+Z^i_{n^\beta,\theta}\big)\right\|_2^2\\&=n^\beta\sum_{r=1}^{d}\sum_{j,j'=1}^{2}\sum_{i,i'=1}^{d}(-1)^{j+j'}C^{j\,j'}_{i\,i'}\ \mathbb{E}\Big[\partial_rQ_d\big(X_T-\xi^{j\,j'}_{i\,i',h}\big)\,\mathbf{H}_{(r)}\big(X_T,Y^{n^\beta}_{ii'}\big)\Big],\end{aligned}$$

To deal with the last obtained quantity, we need the following technical lemma which is proved in the Appendix.

Lemma 6.2. *Let F_h be a d-dimensional random vector, independent of X_T such that*

$$F_h\to x\ \text{ a.s. }\quad\text{as } h\to 0.$$

Assume that F_h have a density function $\rho_{h,x}(y)=\frac{1}{h}\rho(\frac{y-x}{h})$ where ρ is a density function. Then, for $r=1,\ldots,d$

1. *$\partial_rQ_d\big(X_T-F_h\big)\to\partial_rQ_d\big(X_T-x\big)$ a.s. as $h\to 0$.*
2. *For any $0<\delta<(d-1)^{-1}$, we have $\sup_{h>0}\mathbb{E}\Big|\partial_rQ_d\big(X_T-F_h\big)\Big|^{1+\delta}<\infty$.*

We only need to study the behavior of

$$n^\beta\mathbb{E}\Big[\partial_rQ_d\big(X_T-x\big)\,\mathbf{H}_{(r)}\big(X_T,Y^{n^\beta}_{ii'}\big)\Big]$$

in order to prove relation (23). This will be the aim of the next step.

• **Step 3:**

We have that

$$\mathbf{H}_{(r)}\big(X_T,Y^{n^\beta}_{ii'}\big)=Y^{n^\beta}_{ii'}\,\mathbf{H}_{(r)}(X_T,1)-\sum_{j,k=1}^{d}(\gamma^{-1}_{X_T})_{jr}\int_0^T D^k_sX^j_TD^k_sY^{n^\beta}_{ii'}\,ds$$

According to Theorem 5.1 we have that

$$n^{\beta/2} Y_{ii'}^{n^\beta} \Rightarrow^{stably} Y_{ii',\theta} := (U_T^i + \bar{W}_T^i \mathbf{1}_{\{\theta=0\}})(U_T^{i'} + \bar{W}_T^{i'} \mathbf{1}_{\{\theta=0\}}).$$

Therefore as $s \mapsto D_s X_T$ is continuous for $s \in [0,T]$, we have that

$$\begin{aligned} n^\beta \, \mathbf{H}_{(r)}\Big(X_T, Y_{ii'}^{n^\beta}\Big) &\Rightarrow^{stably} Y_{ii'} \, \mathbf{H}_{(r)}(X_T, 1) - 2 \sum_{j,k=1}^{d} (\gamma_{X_T}^{-1})_{jr} \int_0^T D_s^k (U_T^i U_T^{i'}) D_s^k X_T^j \, ds \\ &= \mathbf{H}_{(r)}\Big(X_T, Y_{ii'}\Big). \end{aligned}$$

As the diffusion X and the associated Euler scheme satisfies Proposition 3.1 and using Proposition 7.1 we have that

$$n^\beta \sup_n \Big\| \mathbf{H}_{(r)}\big(X_T, Y_{ii'}^{n^\beta}\big) \Big\|_{l,p} < \infty.$$

Therefore, according to the Lemma 6.2, the sequence $n^\beta \partial_r Q_d(X_T - x) \, \mathbf{H}_{(r)}\Big(X_T, Y_{ii'}^{n^\beta}\Big)$ is uniformly integrable and therefore

$$n^\beta \sum_{r=1}^{d} \mathbb{E}\Big\{ \partial_r Q_d(X_T - x) \, \mathbf{H}_{(r)}\Big(X_T, Y_{ii'}^{n^\beta}\Big) \Big\} \underset{n\to\infty}{\longrightarrow} \sum_{r=1}^{d} \mathbb{E}\Bigg\{ \partial_r Q_d(X_T - x) \, \mathbf{H}_{(r)}\Big(X_T, Y_{ii'}\Big) \Bigg\} = \mathbb{E}\Big\{ \delta_x(X_T) Y_{ii'} \Big\}.$$

The last equality follows from an application of the integration by parts formula. As $\bar{W}_T$ is independent from $(W, \check{W})$, we have that

$$\mathbb{E}\Big(\delta_x(X_T) U_T^i \bar{W}_T^j\Big) = 0,$$

$$\mathbb{E}\big(\delta_x(X_T) \bar{W}_T^i \bar{W}_T^j\big) = \mathbb{E}\big(\delta_x(X_T)\big) T \delta_{ij} = T p(x) \delta_{ij}.$$

Therefore

$$\mathbb{E}\Big\{ \delta_x(X_T) Y_{ii'} \Big\} = \mathbb{E}\Big\{ \delta_x(X_T) U_T^i U_T^{i'} \Big\} + T p(x) \mathbf{1}_{\{\theta=0\}} \delta_{ii'}$$

Therefore we finally obtain that

$$n^{\beta - \alpha(2+d)} \Big\| \sum_{i=1}^{d} \frac{\partial \phi_{h,x}}{\partial x_i}(X_T)(U_T^{n^\beta, i} + Z_{n^\beta,\theta}^i) \Big\|_2 \underset{n\to\infty}{\longrightarrow} \tilde{\sigma}^2,$$

from which the first assertion of the Lemma follows.

The second assertion is a consequence of the first. In fact using the triangular inequality, we have that

$$\begin{aligned} n^{\frac{\beta - \alpha(2+d)}{2}} \Bigg| \Big\| \phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta}) \Big\|_2 - \Big\| \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta}) - \phi_{h,x}(X_T) \Big\|_2 \Bigg| \leq \\ n^{\frac{\beta - \alpha(2+d)}{2}} \Big\| \phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T) \Big\|_2. \end{aligned}$$

As the first assertion is also valid for $\beta' \in (\beta, 1)$. We apply this first assertion noting that $\alpha \leq \frac{\beta}{d+2} < \frac{\beta'}{d+2}$ which gives

$$\lim_{n\to\infty} n^{\frac{\beta' - \alpha(2+d)}{2}} \| \phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T) \|_2 = \tilde{\sigma}.$$

From here it follows that

$$n^{\frac{\beta - \alpha(2+d)}{2}} \Big\| \phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T) \Big\|_2 \underset{n\to\infty}{\longrightarrow} 0.$$

From here the proof of the second assertion follows. □

Proof of Theorem 6.1. We have

$$n\big(V_n - p(x)\big) := \frac{1}{n^{\gamma_1-1}}\sum_{i=1}^{n^{\gamma_1}} \zeta_i^n + \frac{1}{n^{\gamma_2-1}}\sum_{i=1}^{n^{\gamma_2}} \tilde{\zeta}_i^n + n\big(\mathbb{E}\phi_{h,x}(X_T^n + Z_{n,\theta}) - p(x)\big)$$

with

$$\zeta^n = \phi_{h,x}(\hat{X}_T^{n^\beta} + \hat{Z}_{n^\beta,\theta}) - \mathbb{E}\,\phi_{h,x}(\hat{X}_T^{n^\beta} + \hat{Z}_{n^\beta,\theta})$$

and

$$\tilde{\zeta}^n = \phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta}) - \mathbb{E}\left\{\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right\}.$$

From Theorem 4.1 and for $\gamma_1 = 2 + \alpha d$ we have that

$$\frac{1}{n^{\gamma_1-1}}\sum_{i=1}^{n^{\gamma_1}} \zeta_i^n \Rightarrow N(0,\sigma^2) \qquad \text{with} \quad \sigma^2 = \phi^2 p(x)$$

where $\phi^2 = \int_{\mathbb{R}^d} \phi^2(u)\,du$. Therefore to finish the proof it is enough to prove a central limit theorem for $\frac{1}{n^{\gamma_2-1}}\sum_{i=1}^{n^{\gamma_2}} \tilde{\zeta}_i^n$, as the random variables ζ^n and $\tilde{\zeta}^n$ are independent. As in the proof of Theorem 4.1 we have that

$$\mathbb{E}\left[\exp\Big(\frac{iu}{n^{\gamma_2-1}}\sum_{k=1}^{n^{\gamma_2}} \tilde{\zeta}_k^n\Big)\right] = \left[1 + \frac{1}{n^{\gamma_2}}\Big(\frac{-u^2}{2n^{\gamma_2-2}}\mathbb{E}\,|\tilde{\zeta}^{n,h}|^2 + \mathbb{E}\,\tilde{C}_n(\omega)\Big)\right]^{n^{\gamma_2}},$$

with

$$|\mathbb{E}\,\tilde{C}_n(\omega)| \le \frac{u^3}{6n^{2\gamma_2-3}}\mathbb{E}\,|\tilde{\zeta}^n|^3.$$

Now we prove that

$$\frac{1}{n^{\gamma_2-2}}\mathbb{E}\,|\tilde{\zeta}^n|^2 \underset{n\to\infty}{\longrightarrow} \tilde{\sigma}^2$$

and

$$\frac{1}{n^{2\gamma_2-3}}\mathbb{E}\,|\tilde{\zeta}^n|^3 \underset{n\to\infty}{\longrightarrow} 0,$$

which will give as in the proof of Theorem 4.1

$$\frac{1}{n^{\gamma_2-1}}\sum_{i=1}^{n^{\gamma_2}} \tilde{\zeta}_i^n \to \tilde{\sigma} G$$

where G is a standard Gaussian random variable.

Let's start with the term $\mathbb{E}\,|\tilde{\zeta}^n|^2$. We have that

$$\mathbb{E}\,|\tilde{\zeta}^n|^2 = \mathbb{E}\left[\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right]^2 - \left\{\frac{C^s_{\phi,x}}{n} - \frac{C^s_{\phi,x}}{n^\beta} + o\Big(\frac{1}{n^\beta}\Big)\right\}^2,$$

where $C^s_{\phi,x}$ is the constant given in Theorem 3.1 associated to the kernel ϕ. Also from Lemma 6.1 and for $\gamma_2 = (d+2)\alpha + 2 - \beta$ we have that

$$\frac{1}{n^{\gamma_2-2}}\mathbb{E}\,|\tilde{\zeta}^n|^2 \underset{n\to\infty}{\longrightarrow} \tilde{\sigma}^2.$$

On the other hand,

$$\begin{aligned}\mathbb{E}\,|\tilde{\zeta}^n|^3 \le{}& 4\mathbb{E}\left|\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right|^3 + \left|\mathbb{E}\left[\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right]\right|^3 \\ &+ 3\mathbb{E}\left|\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right|^2 \times \left|\mathbb{E}\left[\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right]\right|.\end{aligned}$$

Also using Theorem 3.1 we obtain

$$\mathbb{E}\,|\tilde{\zeta}^n|^3 \le 4\mathbb{E}\left|\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right|^3 + \Big|\frac{C^s_{\phi,x}}{n} - \frac{C^s_{\phi,x}}{n^\beta} + o\Big(\frac{1}{n^\beta}\Big)\Big|^3$$

$$+\, 3\mathbb{E}\left|\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right|^2 \Big|\frac{C^s_{\phi,x}}{n} - \frac{C^s_{\phi,x}}{n^\beta} + o\Big(\frac{1}{n^\beta}\Big)\Big|.$$

Applying again Lemma 6.1 we have that for $\gamma_2 = (d+2)\alpha + 2 - \beta$

$$\frac{1}{n^{2\gamma_2-3}}\mathbb{E}\left|\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right|^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Therefore it remains to prove that

$$\frac{1}{n^{2\gamma_2-3}}\mathbb{E}\left|\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right|^3 \underset{n\to\infty}{\longrightarrow} 0.$$

As $\phi_{h,x}$ is a Lipschitz function with Lipschitz constant of c/h^{d+1} for $c > 0$, we obtain

$$\begin{aligned}\frac{1}{n^{2\gamma_2-3}}\mathbb{E}\left|\phi_{h,x}(X_T^n + Z_{n,\theta}) - \phi_{h,x}(X_T^{n^\beta} + Z_{n^\beta,\theta})\right|^3 &\le \frac{c}{n^{2\gamma_2-3}h^{3(d+1)}}\Big[\mathbb{E}|X_T^n - X^{n^\beta}|^3 + \mathbb{E}|Z_{n,\theta} - Z_{n^\beta,\theta}|^3\Big]\\ &\le \frac{c}{n^{2\gamma_2-3}h^{3(d+1)}} \times \frac{C_T}{n^{\frac{3\beta}{2}}} = \frac{c\,C_T}{n^{-(d-1)\alpha+1-\frac{\beta}{2}}} \to 0.\end{aligned}$$

The last convergence is true if $0 < \alpha < \beta/(d+2)$ and $0 < \beta < 2/3$. This finishes the proof of the Theorem. $\square$

Like in the case of the Monte Carlo method one can interpret the previous result as follows: In order to approach the density $p(x)$ using a control variate method of the Romberg type with a global tolerance error of order $1/n$, the parameters needed to use the algorithm are $h = n^{-\alpha}$, $N_1 = n^{2+\alpha d}$ $N_2 = n^{(d+2)\alpha+2-\beta}$ with $\beta/(d+2) > \alpha \ge 1/s$ where s denotes the order of the superkernel ϕ. Therefore the complexity (number of calculations) needed for this algorithm is

$$\begin{aligned}C_{RS} &= C \times mN_1 + (n+m)N_2\\ &\simeq C \times n^{\beta+\alpha d+2} + n^{(d+2)\alpha-\beta+3}, \quad \text{where} \quad \beta/(d+2) > \alpha \ge 1/s.\end{aligned}$$

For $\beta = \frac{1}{2} + \alpha$ we obtain that the complexity of the Romberg method is given by

$$C^\star_{RS} \simeq C \times n^{\frac{5}{2}+(d+1)\alpha}.$$

Here note that the optimal complexity for the Monte Carlo method is given by

$$C^\star_{MC} \simeq C \times n^{3+\alpha d}.$$

Therefore the Romberg control variate method reduces the complexity by a factor of of order $n^{1/2-\alpha}$. Therefore taking into account that $\beta/(d+2) > \alpha \ge 1/s$ we see that if one uses super-kernels of order $s > 2(d+1)$ we obtain a theoretical asymptotic optimal parameter choice of the method.

7 Appendix 1

In this appendix we prove some estimates that are useful to estimate the norms of the weights in the integration by parts formula. In order to simplify the notation we suppose that c is a positive constant that may change from one line to the next.

Lemma 7.1. *For all $l > 1$, $p > 1$ there exists positive constants k_2, p_1, p_2, γ_1 and γ_2 and a positive constant c independent of n, θ and F such that*

$$\|(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{ij}\|_{l,p} \leq c\,\|(\det \tilde{\gamma}_{F+Z_{n,\theta}})^{-1}\|^{\gamma_1}_{p_1}\|F + Z_{n,\theta}\|^{\gamma_2}_{k_2,p_2},$$

Proof. The proof is done by induction on $l = |k|$. The case $l = 0$ is a direct consequence of the Cramer formula for the inverse of a given matrix.

In general, as $D^r\left\{\tilde{\gamma}_{F+Z_{n,\theta}}\tilde{\gamma}^{-1}_{F+Z_{n,\theta}}\right\} = 0$ for any multi-index r, we have that

$$D^r(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{lm} = \sum_{i,j=1}^{d}\sum_{k\in A(r)-\{r\}}(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{li}D^{r-k}\tilde{\gamma}^{ij}_{F+Z_{n,\theta}}D^k(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{jm}.$$

Here $A(r)$ denotes all the subsets of indices of any order taken from elements of r. Then the result follows from the inductive hypothesis. $\square$

Proposition 7.1. *Let $G \in \mathbb{D}^\infty(\tilde{W})$, then*

1. *Let $F \in (\mathbb{D}^\infty(W))^d$ such that $F + Z_{n,\theta}$ is a non-degenerate random vector. For $p > 1$ and for all multi-index m we have*

$$\|\,\tilde{\mathbf{H}}_m(F,G)\|_p \leq c\|G\|_{r,r'}\|(\det \tilde{\gamma}_{F+Z_{n,\theta}})^{-1}\|^{a'}_{a}\left[\|F\|^{b'}_{b,l} + \frac{1}{n^{(\frac{1}{2}+\theta)l'}}\right]$$

where c is a constant depending on p, m and d, whereas r, r', a, a', b, b', l and l' are parameters depending on m, p and d.

2. *Let $F_1, F_2 \in (\mathbb{D}^\infty(W))^d$ such that $F_1 + Z_{n,\theta}$ and $F_2 + Z_{n,\theta}$ are non-degenerate random vectors. For a fixed multi-index m, any $l \geq 1$ and $p > 1$, there exists c, a positive constant depending on p, m and d, whereas k_i, s_i, β_i, p_i, γ_i, for $i = 1, 2$, k_0, s_0, γ_0, $\bar{k}_0$ and $\bar{s}_0$ are parameters depending on m, p and d such that*

$$\|\,\tilde{\mathbf{H}}_m(F_1,G) - \tilde{\mathbf{H}}_m(F_2,G)\|_{l,p} \leq c\prod_{i=1}^{2}(1 + \|F_i\|^{\gamma_i}_{k_i,s_i})(1 + \|(\det \tilde{\gamma}_{F_i+Z_{n,\theta}})^{-1}\|^{\beta_i}_{p_i})$$
$$\times\,\|F_1 - F_2\|^{\gamma_0}_{k_0,s_0}\|G\|_{\bar{k}_0,\bar{s}_0}$$

3. *Let $F \in (\mathbb{D}^\infty(W))^d$ be a non-degenerate random vector. For a fixed multi-index m, any $l \geq 1$ and $p > 1$, there exists a constant c and parameters r_i, k_i, μ_i, for $i = 1, 2, 3$ depending on p, m and d such that*

$$\|\,\tilde{\mathbf{H}}_m(F,G) - \mathbf{H}_m(F,G)\|_{l,p} \leq \frac{c}{n^{(\frac{1}{2}+\theta)\mu}}\|G\|_{k_1,r_1}(1 + \|F + Z_{n,\theta}\|^{\mu_2}_{k_2,r_2})(1 + \|(\det \tilde{\gamma}_{F+Z_{n,\theta}})^{-1}\|^{\mu_3}_{r_3}).$$

Proof. Again the proof is done by induction on the length of the multi-index m. In fact, using the definition of $\tilde{\mathbf{H}}$ and the continuity of the adjoint operator δ, we have

$$\|\,\tilde{\mathbf{H}}_m(F,G)\|_{l,p} \quad\leq\quad \sum_{r\in m}\sum_{j=1}^{d}\|\tilde{D}(F + Z_{n,\theta})^j\,\tilde{\mathbf{H}}_{m-\{r\}}(F,G)(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{rj}\|_{l+1,p}.$$

Then the proof finishes by using Hölder's inequality, Lemma 7.1 and the inductive hypothesis.

The proof of the second assertion is as the previous one, done by induction on the order of the multi-index m

$$\|\tilde{\mathbf{H}}_m(F_1,G)-\tilde{\mathbf{H}}_m(F_2,G)\|_{l,p}\le\sum_{r\in m}\sum_{j=1}^d\|\tilde{D}(F_1-F_2)^j\,\tilde{\mathbf{H}}_{m-\{r\}}(F_1,G)(\tilde{\gamma}^{-1}_{F_1+Z_{n,\theta}})^{rj}\|_{l+1,p}$$

$$+\sum_{r\in m}\sum_{j=1}^d\|\tilde{D}(F_2+Z_{n,\theta})^j(\tilde{\mathbf{H}}_{m-\{r\}}(F_1,G)-\tilde{\mathbf{H}}_{m-\{r\}}(F_2,G))(\tilde{\gamma}^{-1}_{F_1+Z_{n,\theta}})^{rj}\|_{l+1,p}$$

$$+\sum_{r\in m}\sum_{j=1}^d\|\tilde{D}(F_2+Z_{n,\theta})^j\,\tilde{\mathbf{H}}_{m-\{r\}}(F_2,G)((\tilde{\gamma}^{-1}_{F_1+Z_{n,\theta}})^{rj}-(\tilde{\gamma}^{-1}_{F_2+Z_{n,\theta}})^{rj})\|_{l+1,p}.$$

For the first term one applies the Hölder's inequality, the first assertion and Lemma 7.1. For the second, Hölder's inequality, Lemma 7.1 and the inductive hypothesis. For the third, note that

$$(\tilde{\gamma}^{-1}_{F_1+Z_{n,\theta}})^{rj}-(\tilde{\gamma}^{-1}_{F_2+Z_{n,\theta}})^{rj}=\sum_{k,k'=1}^d(\tilde{\gamma}^{-1}_{F_2+Z_{n,\theta}})^{rk}\Big[(\tilde{\gamma}_{F_2+Z_{n,\theta}})^{kk'}-(\tilde{\gamma}_{F_1+Z_{n,\theta}})^{kk'}\Big](\tilde{\gamma}^{-1}_{F_1+Z_{n,\theta}})^{k'j}\quad(25)$$

Note that $\tilde{\gamma}_{F_1+Z_{n,\theta}}=\gamma_{F_1}+\bar{\gamma}_{Z_{n,\theta}}$ and $\tilde{\gamma}_{F_2+Z_{n,\theta}}=\gamma_{F_2}+\bar{\gamma}_{Z_{n,\theta}}$. Consequently, it follows that

$$(\tilde{\gamma}_{F_2+Z_{n,\theta}})^{kk'}-(\tilde{\gamma}_{F_1+Z_{n,\theta}})^{kk'}=\langle DF_2^k-DF_1^k,DF_2^{k'}\rangle_H+\langle DF_1^k,DF_2^{k'}-DF_1^{k'}\rangle_H.$$

From here the result follows.

In the same way as before, we prove the last relation for an index m. We have

$$\begin{aligned}\tilde{\mathbf{H}}_m(F,G)-\mathbf{H}_m(F,G)&=\sum_{j=1}^m\delta\Big(GDF^j\,[(\tilde{\gamma}^{-1}_{F+Z_{n,\theta}})^{mj}-(\tilde{\gamma}_F^{-1})^{mj}]\Big),\\&+\frac{1}{n^{\frac12+\theta}}\sum_{j=1}^d\bar{\delta}\Big(G(\gamma^{-1}_{F+Z_{n,\theta}})^{ij}\bar{D}\bar{W}_T^j\Big)\end{aligned}$$

Therefore the result follows applying (25) and the same arguments as in the previous proofs of assertions 1 and 2. □

8 Appendix 2

Proof of Lemma 3.4. In order to prove the relation (11) is enough to prove that

$$\sup_{0\le u\le s\le T}|\Delta_n(u,s)|:=\sup_{0\le u\le s\le T}|\Delta_n^1(u,s)+\Delta_n^2(u,s)|\to 0$$

with

$$\Delta_n^1(u,s):=\mathbb{E}\Big[\big(\psi_{h,x}\big(X_T^{n,\lambda}+Z_{n,\theta}\big)-\mathbf{1}_{\{X_T>x\}}\big)\,\mathbf{H}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\Big]$$

and

$$\Delta_n^2(u,s):=\mathbb{E}\Big[\psi_{h,x}\big(X_T^{n,\lambda}+Z_{n,\theta}\big)\big\{\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{n,r,j,k}\big)-\mathbf{H}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\big\}\Big],$$

with $h=n^{-\alpha}$. Since for every $p\ge1$ we have that

$$\psi_{h,x}\big(X_T^{n,\lambda}+Z_{n,\theta}\big)\xrightarrow{L^p}\mathbf{1}_{\{X_T>x\}}$$

and

$$\sup_{0\le u\le s\le T}\big\|\mathbf{H}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\big\|_p<\infty$$

then we deduce that

$$\sup_{0\le u\le s\le T}|\Delta_n^1(u,s)|\to 0.$$

In addition we have

$$\sup_{0\le u\le s\le T}|\Delta_n^2(u,s)|\le c\sup_{0\le u\le s\le T}\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{n,r,j,k}\big)-\mathbf{H}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\Big\|_2$$

$$\begin{aligned}\sup_{0\le u\le s\le T}|\Delta_n^2(u,s)| &\le c\sup_{0\le u\le s\le T}\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{n,r,j,k}\big)-\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{r,j,k}\big)\Big\|_2\\ &+c\sup_{0\le u\le s\le T}\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{r,j,k}\big)-\tilde{\mathbf{H}}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\Big\|_2\\ &+c\sup_{0\le u\le s\le T}\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)-\mathbf{H}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\Big\|_2.\end{aligned}$$

Note that

$$\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{n,r,j,k}\big)-\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{r,j,k}\big)\Big\|_2=\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{n,r,j,k}-G_{u,s}^{r,j,k}\big)\Big\|_2.$$

Since $X_T^{n,\lambda}+Z_{n,\theta}$ is uniformly non-degenerate, we conclude using the first assertion of Proposition 7.1 and Proposition 3.1 as well as Lemma 3.2, that

$$\sup_{0\le u\le s\le T}\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{n,r,j,k}-G_{u,s}^{r,j,k}\big)\Big\|_2\to 0,\ \ (n\to\infty).$$

In the same way, since $X_T^{n,\lambda}+Z_{n,\theta}$ and X_T are non-degenerate, we conclude using the second assertion of Proposition 7.1 and relations (5) and (6), that

$$\sup_{0\le u\le s\le T}\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T^{n,\lambda},G_{u,s}^{r,j,k}\big)-\tilde{\mathbf{H}}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\Big\|_2\to 0,\ \ (n\to\infty).$$

Finally, according to the third assertion of Proposition 7.1 we obtain that

$$\sup_{0\le u\le s\le T}\Big\|\tilde{\mathbf{H}}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)-\mathbf{H}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\Big\|_2\to 0,\ \ (n\to\infty).$$

We conclude that

$$\sup_{0\le u\le s\le T}|\Delta_n^2(u,s)|\to 0,\ \ (n\to\infty).$$

The continuity of

$$g(u,s)=E\Big(\mathbf{1}_{\{X_T>x\}}\,\mathbf{H}_{r^+}\big(X_T,G_{u,s}^{r,j,k}\big)\Big)$$

follows from Proposition 3.1 as well as Lemma 3.2. □

Proof of Lemma 6.2. The case $d=1$ is trivial, so we will assume for the rest of the proof that $d\ge 2$. It is clear that the function $\partial_r Q$ is continuous except at the origin. Since the random vector $F_h\to x$ as $h\to 0$ a.s., the first assertion of the lemma follows. Now we prove the second assertion. For $a>0$, we have that

$$\begin{aligned}\mathbb{E}\big|\partial_r Q_d(X_T-F_h)\big|^{1+\delta} &= E\Big\{\big|\partial_r Q_d(X_T-F_h)\big|^{1+\delta}\mathbf{1}_{\{|X_T-F_h|\le 2a\}}\Big\}\\ &+E\Big\{\big|\partial_r Q_d(X_T-F_h)\big|^{1+\delta}\mathbf{1}_{\{|X_T-F_h|>2a\}}\Big\}.\end{aligned}\tag{26}$$

• **Step 1:** First consider the first term on the right of (26). Then we have

$$E\Big\{\big|\partial_r Q_d(X_T-F_h)\big|^{1+\delta}\mathbf{1}_{\{|X_T-F_h|\le 2a\}}\Big\}=\int_{\mathbb{R}^d}\big|\partial_r Q_d(y)\big|^{1+\delta}\mathbf{1}_{\{|y|\le 2a\}}p_h(y)dy$$

where p_h denotes the density of the random vector $X_T - F_h$. If we have that

$$\sup_h \sup_{|y|\leq 2a} p_h(y) \leq C_x,$$

then it follows immediately that

$$E\Big\{\big|\partial_r Q_d(X_T - F_h)\big|^{1+\delta}\mathbf{1}_{\{|X_T - F_h| \,\leq\, 2a\}}\Big\} \leq C_x \int_{|y|\leq 2a} \big|\partial_r Q_d(y)\big|^{1+\delta}\, dy.$$

As $\big|\partial_r Q_d(y)\big|^{1+\delta} \leq C_d/|y|^{(d-1)(1+\delta)}$. Therefore we obtain

$$\sup_h E\Big\{\big|\partial_r Q_d(X_T - F_h)\big|^{1+\delta}\mathbf{1}_{\{|X_T - F_h| \,\leq\, 2a\}}\Big\} < \infty,$$

for $\delta < (d-1)^{-1}$.

• **Step 2:** Now, we prove that

$$\sup_h \sup_{|y|\leq 2a} p_h(y) \leq C_x.$$

Since F_h and X are independent, we have that

$$p_h(y) = \int_{\mathbb{R}^d} \rho_{h,x}(u-y)p(u)\, du,$$

where $\rho_{h,x}$ denotes the density of F_h and p denotes the density of X_T. Then it follows that

$$\begin{aligned} p_h(y) &\leq \sup_{u\in\mathbb{R}^d} p(u) \int_{\mathbb{R}^d} \rho_{h,x}(u-y)\, du \\ &= \sup_{u\in\mathbb{R}^d} p(u). \end{aligned}$$

According to Corollary 3.25 in Kusuoka and Stroock (1985) we have that,

$$p(y) \leq \frac{C}{T^l}\exp(-\frac{|y-x|^2}{CT}), \quad \text{for some} \quad C>0, \;\; \text{and} \;\; l>0.$$

In particular, it follows that p is bounded and therefore the result follows.

• **Step 3:** Now we are able to deal with the second term of equality (26). We have

$$\begin{aligned} &E\Big\{\big|\partial_r Q_d(X_T - F_h)\big|^{1+\delta}\mathbf{1}_{\{|X_T - F_h| \,>\, 2a\}}\Big\} = \\ &\qquad \int_{\mathbb{R}^d\times\mathbb{R}^d} \big|\partial_r Q_d(y-z)\big|^{1+\delta}\mathbf{1}_{\{|y-z| \,>\, 2a\}} p(y)\rho_{h,x}(z) dz\, dy. \end{aligned}$$

Using again that $\big|\partial_r Q_d(y)\big|^{1+\delta} \leq C_d/|y|^{(d-1)(1+\delta)}$ and the Kusuoka-Stroock estimate mentioned previously, we have that

$$\begin{aligned} &\int_{\mathbb{R}^d\times\mathbb{R}^d} \big|\partial_r Q_d(y-z)\big|^{1+\delta}\mathbf{1}_{\{|y-z| \,>\, 2a\}} p(y)\rho_{h,x}(z) dz\, dy \\ &\qquad \leq \frac{C_d C}{(2a)^{(d-1)(1+\delta)}T^l}\int_{\mathbb{R}^d}\exp(-\frac{|y-x|^2}{CT})dy \int_{\mathbb{R}^d}\rho_{h,x}(z)dz < \infty. \end{aligned}$$

This completes the proof of the lemma. □

9 Appendix 3

In the following we prove Lemmas 5.1, 5.2 and 5.3.

Proof of Lemma 5.1. The proof uses the same ideas of Jacod and Protter (1998). Note that for $0 \le t < t' \le T$, the sequence $\Big(\sqrt{n}\int_t^{t'} (W_s^j - W_{\eta_n(s)}^j)\,ds\Big)_{n\in\mathbb{N}}$ tends to 0 in $L^2(\Omega)$. In fact, we have

$$\mathbb{E}\Big(\int_t^{t'} (W_s^j - W_{\eta_n(s)}^j)\,ds\Big)^2 \le \frac{c}{n^2}, \quad c > 0.$$

Therefore we have

$$\int_0^T H_s^{i,n}(W_s^j - W_{\eta_n(s)}^j)\,ds = \int_0^T (H_s^{i,n} - H_{m,s}^{i,n})(W_s^j - W_{\eta_n(s)}^j)\,ds + \int_0^T H_{m,s}^{i,n}(W_s^j - W_{\eta_n(s)}^j)\,ds$$

with

$$H_{m,s}^{i,n} = \sum_{k=1}^m H_{\frac{kT}{m}}^{i,n}\mathbf{1}_{\{\frac{(k-1)T}{m} < s \le \frac{kT}{m}\}}.$$

It follows that

$$\begin{aligned}\Big|\int_0^T H_s^{i,n}(W_s^j - W_{\eta_n(s)}^j)\,ds\Big| &\le \sup_{0<s\le T}|H_s^{i,n} - H_{m,s}^{i,n}|\int_0^T |W_s^j - W_{\eta_n(s)}^j|ds \\ &\quad +\Big|\int_0^T H_{m,s}^{i,n}(W_s^j - W_{\eta_n(s)}^j)\,ds\Big|.\end{aligned}$$

Since the sequence $\sqrt{n}\int_0^T |W_s^j - W_{\eta_n(s)}^j|\,ds$ is tight, after some work we deduce the conclusion of the lemma. □

Proof of Lemma 5.2. Without loss of generality, we assume that $T \le 1$ throughout this proof. We denote by

$$H \diamond K^n = \int_0^T H_s\Big(\int_0^T \mathbf{1}_{\{\eta_n(u)\le s\le u\}}K_u^n\,dY_u^j\Big)ds$$

and suppose in a first time that H is deterministic then

$$\begin{aligned}H \diamond K^n &= \int_0^T K_u^n\Big(\int_0^T \mathbf{1}_{\{\eta_n(u)\le s\le u\}}H_s\,ds\Big)dY_u^j \\ &= \int_0^T K_u^n\Big(\int_{\eta_n(u)}^u H_s\,ds\Big)dY_u^j.\end{aligned}$$

It follows that

$$\begin{aligned}\|H \diamond K^n\|_2^2 &\le \mathbb{E}\int_0^T \|K_u^n\|^2\Big(\int_{\eta_n(u)}^u H_s\,ds\Big)^2 du \\ &\le |H|_\infty^2\mathbb{E}\int_0^T \|K_u^n\|^2(u - \eta_n(u))^2\,du \\ &\le \frac{|H|_\infty^2 T^2}{n^2}\mathbb{E}\int_0^T \|K_u^n\|^2\,du,\end{aligned}$$

and consequently $(\sqrt{n}H\diamond K^n)_{n\in\mathbb{N}}$ tends to 0 in $L^2(\Omega)$. Now let H to be arbitrary. We have that $H \in \mathscr{C}([0,T])$, so there exists a sequence $H^l \in \mathscr{C}([0,T])$ of piecewise functions such that $|H - H^l|_\infty \to 0$ and $|H^l|_\infty \le l$ *a.s.*. We have

$$|(H - H^l)\diamond K^n| \le |H - H^l|_\infty \int_0^T\Big|\int_0^T \mathbf{1}_{\{\eta_n(u)\le s\le u\}}K_u^n\,dY_u^j\Big|ds.$$

It is obvious that the sequence

$$\Big(\sqrt{n}\int_0^T\Big|\int_0^T \mathbf{1}_{\{\eta_n(u)\le s\le u\}} K_u^n\, dY_u^j\Big|ds\Big)_{n\in\mathbb{N}}$$

is tight, (because it is bounded in L^2). Consequently:

$$\begin{aligned}
\mathbb{P}\Big(\sqrt{n}|H\diamond K^n|\ge\varepsilon\Big) &\le \mathbb{P}\Big(\sqrt{n}|(H-H^l)\diamond K^n|\ge\frac{\varepsilon}{2}\Big)+\mathbb{P}\Big(\sqrt{n}|H^l\diamond K^n|\ge\frac{\varepsilon}{2}\Big)\\
&\le \mathbb{P}\Big(|H-H^l|_\infty\ge\frac{\varepsilon}{2\delta}\Big)\\
&+\mathbb{P}\Big(\sqrt{n}\int_0^T\Big|\int_0^T \mathbf{1}_{\{\eta_n(u)\le s\le u\}} K_u^n\, dY_u^j\Big|ds\ge\delta\Big)\\
&+\mathbb{P}\Big(\sqrt{n}|H^l\diamond K^n|\ge\frac{\varepsilon}{2}\Big).
\end{aligned}$$

For a fixed l and for a good choice of δ and n we obtain that for a given $\rho>0$,

$$\limsup_{n\to\infty}\mathbb{P}\Big(\sqrt{n}|H\diamond K^n|\ge\varepsilon\Big)\le\rho+\frac{l^2T^2}{n^2}\mathbb{E}\int_0^T\|K_u^n\|^2du+\mathbb{P}\Big(|H-H^l|_\infty\ge\frac{\varepsilon}{2\delta}\Big).$$

Since ρ is arbitrary and $|H-H^l|_\infty\to 0$ *a.s.*, we conclude that

$$\limsup_{n\to\infty}\mathbb{P}\Big(\sqrt{n}|H\diamond K^n|\ge\varepsilon\Big)=0.$$

Which completes the proof. $\square$

Proof of Lemma 5.3. We split the proof of the lemma into two steps
• **Step 1:** We suppose first H^i, K^i and L^i are deterministic. Then we have:

$$\begin{aligned}
\int_0^T K_s^i\Big(\sum_{j=1}^q\int_s^T \xi_{s,u}^{ij}U_u^n\, dW_u^j\Big)ds &= \int_0^T\Big(\int_0^u K_s^i\xi_{s,u}^{ij}ds\Big)U_u^n dW_u^j\\
&=\sum_{j=1}^q\int_0^T \bar{K}_u^{ij}U_u^n\, dW_u^j
\end{aligned}$$

with $\bar{K}_u^{ij}=\int_0^u K_s^i\xi_{s,u}^{ij}\,ds$. In the same manner:

$$\int_0^T L_s^i\Big(\int_s^T\sum_{j,k=1}^q \zeta_{s,u}^{ijk}\, d\check{W}_u^{n,kj}\Big)ds=\sum_{j,k=1}^q\int_0^T \bar{L}_u^{ijk}\, d\check{W}_u^{n,kj},$$

with $\bar{L}_u^{ijk}=\int_0^u L_s^i\zeta_{s,u}^{ijk}ds$. Therefore we need to prove that

$$\sqrt{n}\Big(U_T^n,\int_0^T H_s^iU_s^n\, ds,\sum_{j=1}^q\int_0^T \bar{K}_u^{ij}U_u^n\, dW_u^j,\sum_{j,k=1}^q\int_0^T \bar{L}_u^{ijk}\, d\check{W}_u^{n,kj}\Big)$$

stably converge in law to

$$\Big(U_T,\int_0^T H_s^iU_s\, ds,\sum_{j=1}^q\int_0^T \bar{K}_u^{ij}U_u\, dW_u^j,\sum_{j,k=1}^q\int_0^T \bar{L}_u^{ijk}\, d\check{W}_u^{kj}\Big)$$

Since the process H^i is deterministic and the processes $\bar{K}^{ij}$ and $\bar{L}^{ijk}$ are continuous and adapted as in Lemma 5.1, we deduce, using an approximation argument, that proving the convergence above can be carried into proving that $\sum_{i=1}^m Z_iV_i^n$ stably converge in law to $\sum_{i=1}^m Z_iV_i$ where $Z_1,\dots,Z_m$

are random matrices and $(V_1^n, \ldots, V_m^n)$ are random vectors converging stably to $(V_1, \ldots, V_m)$. This is a classical property of the stable convergence. In fact, $(Z, Z_1, \ldots, Z_u, V_1^n, \ldots, V_m^n)$ converge to $(Z, Z_1, \ldots, Z_u, V_1, \ldots, V_m)$, it follows that $(Z, \sum_{i=1}^m Z_i V_i^n)$ stably converge in law to $(Z, \sum_{i=1}^m Z_i V_i)$, (see Jacod and Shiryaev (2003) chapter VIII §5.c and Theorems 2.3 and 3.2 in Jacod and Protter (1998)).

• **Step 2:** Suppose now H^i, K^i and L^i are arbitrary. Since the processes H^i, K^i and L^i have continuous trajectories on $[0,T]$, we can approach them by three piecewise functions H_l^i, K_l^i, L_l^i. In the following we introduce the following notations

$$H^i.U^n = \int_0^T H_s^i U_s^n \, ds \qquad K^i \star U^n = \int_0^T K_s^i \Big(\int_s^T \sum_{j=1}^q \xi_{s,u} U_u^n dW_u^j \Big) ds$$

$$(L^i | \check{W}^{n,kj}) = \int_0^T L_s^i \Big(\int_s^T \sum_{j,k=1}^n \zeta_{s,u}^{ijk} \, d\check{W}_u^{n,kj} \Big) ds.$$

We have $\sqrt{n} \| H^i.U^n - H_l^i.U^n \| \leq |H^i - H_l^i|_\infty \int_0^T \sqrt{n} \| \bar{U}_s^n \| ds$ where $| \ |_\infty$ denotes the uniform norm on the space $\mathscr{C}([0,T])$. Similarly, we have

$$\sqrt{n} \| K^i \star U^n - K_l^i \star U^n \| \leq |K^i - K_l^i|_\infty \int_0^T \sqrt{n} \Big\| \sum_{j=1}^q \int_s^T \xi_{s,u}^{ij} U_u^n \, dW_u^j \Big\| ds$$

and

$$\sqrt{n} \| (L^i | \check{W}^{n,kj}) - (L_l^i | \check{W}^{n,kj}) \| \leq |L^i - L_l^i|_\infty \int_0^T \sqrt{n} \Big\| \int_s^T \sum_{j,k=1}^q \zeta_{s,u}^{ijk} \, d\check{W}_u^{n,kj} \Big\| ds.$$

Consequently, in order to prove the statement of the lemma, we have just to prove the tightness of

$$\int_0^T \sqrt{n} \| U_s^n \| \, ds, \qquad P_n^i = \int_0^T \sqrt{n} \Big\| \int_s^T \sum_{j=1}^q \xi_{s,u}^{ij} U_u^n \, dW_u^j \Big\| \, ds$$

$$\text{and} \quad Q_n^i = \int_0^T \sqrt{n} \Big\| \int_s^T \sum_{j,k=1}^q \zeta_{s,u}^{ijk} \, d\check{W}_u^{n,kj} \Big\| ds.$$

The tightness of the sequence $\int_0^T \sqrt{n} \| U_s^n \| \, ds$, follows from the convergence of the law of $\sqrt{n} U^n$. For P_n^i and Q_n^i, this is a consequence of the hypothesis on $\xi_{s,u}^{ij}$, $\zeta_{s,u}^{ijk}$. In fact :

$$\begin{aligned}
\| P_n^i \|_2 &\leq \int_0^T \Big\| \int_s^T \sqrt{n} \sum_{j=1}^q \xi_{s,u}^{ij} U_u^n \, dW_u^j \Big\|_2 ds \\
&= \int_0^T \sqrt{n} \Big\| \Big(\int_s^T \| \sum_{j=1}^q \xi_{s,u}^{ij} U_u^n \|^2 du \Big)^{1/2} \Big\|_2 ds \\
&\leq \sqrt{T} \Big(\mathbb{E} \int_0^T \int_s^T \| \sum_{j=1}^q \xi_{s,u}^{ij} \sqrt{n} U_u^n \|^2 du ds \Big)^{1/2} \\
&\leq q \sqrt{T} \Big(\mathbb{E} \int_0^T \Big[\int_0^u \max_j \| \xi_{s,u}^{ij} \|^2 ds \Big] \| \sqrt{n} U_u^n \|^2 du \Big)^{1/2},
\end{aligned}$$

Using that $\sup_n \mathbb{E} \int_0^T \| \sqrt{n} U_u^n \|^q du < \infty$ for $q \geq 1$ and that

$$\mathbb{E} \int_0^T du \int_0^u ds \Big(\max_j \| \xi_{s,u}^{ij} \|^p \Big) < \infty \qquad \text{for} \quad p > 2.$$

we obtain that

$$\sup_n \|P_n\|_2 < \infty.$$

In the same manner we obtain that $\sup_n \|Q_n\|_2 < \infty$ which completes the proof of the lemma. $\square$

References

Bally, V. and Talay, D. (1996). The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods Appl.* **2** 93–128.

Clement, E., Arturo, K.-H. and Lamberton, D. (2006). A duality approach for the weak approximation of stochastic differential equations. *To appear in Annals of Applied Probability* .

Duffie, D. and Glynn, P. (1995). Efficient Monte Carlo simulation of security prices. *Ann. Appl. Probab.* **5** 897–905.

Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307.

Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*, vol. 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.

Kebaier, A. (2005). Statistical romberg extrapolation: A new variance reduction method and applications to option pricing. *To appear in Annals of Applied Probability* .

Kohatsu-Higa, A. and Pettersson, R. (2002). Variance reduction methods for simulation of densities on Wiener space. *SIAM J. Numer. Anal.* **40** 431–450 (electronic).

Kusuoka, S. and Stroock, D. (1984). Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, vol. 32 of *North-Holland Math. Library*. North-Holland, Amsterdam, 271–306.

Kusuoka, S. and Stroock, D. (1985). Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32** 1–76.

Malliavin, P. and Thalmaier, A. (2006). *Stochastic calculus of variations in mathematical finance*. Springer Finance, Springer-Verlag, Berlin.

Nualart, D. (1995,2006). *The Malliavin calculus and related topics*. Probability and its Applications (New York), Springer-Verlag, New York.

Nualart, D. (1998). Analysis on Wiener space and anticipating stochastic calculus. In *Lectures on probability theory and statistics (Saint-Flour, 1995)*, vol. 1690 of *Lecture Notes in Math.* Springer, Berlin, 123–227.

Protter, P. (1990). *Stochastic integration and differential equations*, vol. 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin.