Discretization and simulation of Zakai equation

Emmanuel GOBET[*] Gilles PAGÈS[†] Huyên PHAM[‡]
Jacques PRINTEMS[§]

September 2005
this revised version : June 2006

Abstract

This paper is concerned with numerical approximations for stochastic partial differential Zakai equation of nonlinear filtering problem. The approximation scheme is based on the representation of the solutions as weighted conditional distributions. We first accurately analyse the error caused by an Euler type scheme of time discretization. Sharp error bounds are calculated: we show that the rate of convergence is in general of order $\sqrt{\delta}$ (δ is the time step), but in the case when there is no correlation between the signal and the observation for the Zakai equation, the order of convergence becomes δ. This result is obtained by carefully employing techniques of Malliavin calculus. In a second step, we propose a simulation of the time discretization Euler scheme by a quantization approach. This formally consists in an approximation of the weighted conditional distribution by a conditional discrete distribution on finite supports. We provide error bounds and rate of convergence in terms of the number N of the grids of this support. These errors are minimal at some optimal grids which are computed by a recursive method based on Monte Carlo simulations. Finally, we illustrate our results with some numerical experiments arising from correlated Kalman-Bucy filter.

Key words: Stochastic partial differential equations, nonlinear filtering, Zakai equation, Euler scheme, quantization, Malliavin calculus.

MSC Classification (2000): 60H35, 60H15, 60G35, 60H07, 65C20

[*]LMC, ENSIMAG-INP Grenoble, emmanuel.gobet@imag.fr
[†]LPMA, Université Paris 6-Paris 7, gpa@ccr.jussieu.fr
[‡]LPMA, Université Paris 6-Paris 7, pham@math.jussieu.fr
[§]LAMA, Université Paris 12, printems@univ-paris12.fr

1 Introduction

We are interested in numerical approximation for the measure-valued process V governed by the following stochastic partial differential equations (SPDE) written in weak form: for all test functions $f \in C_b^2(\mathbb{R}^d)$,

$$
\begin{aligned}
< V_t, f > &= < \mu_0, f > + \int_0^t < V_s, Lf > ds \\
&\quad + \int_0^t < V_s, hf + \gamma^\intercal \nabla f > .dW_s,
\end{aligned} \tag{1.1}
$$

where μ_0 is an initial probability measure. We denote by $\mathcal{M}(\mathbb{R}^d)$ the set of finite signed measures on $\mathbb{R}^d$. Here L is the second-order differential operator:

$$
Lf(x) = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \partial^2_{x_i x_j} f(x) + \sum_{i=1}^{d} b_i(x) \partial_{x_i} f(x),
$$

W is a q-dimensional Brownian motion, $a = (a_{ij})$ is a $d \times d$ matrix-valued, $\gamma = (\gamma_{il})$ is a $d \times q$ matrix-valued, $b = (b_i)$ is a $\mathbb{R}^d$-vector valued, and $h = (h_l)$ is a $\mathbb{R}^q$-vector valued function defined on $\mathbb{R}^d$, in the form:

$$
\begin{aligned}
a &= \sigma\sigma^\intercal + \gamma\gamma^\intercal, \\
b &= \beta + \gamma h,
\end{aligned}
$$

for some $d \times d$ matrix-valued function $\sigma = (\sigma_{ij})$ and $\mathbb{R}^d$-vector valued function $\beta = (\beta_i)$ on $\mathbb{R}^d$. The transpose and the scalar product are respectively denoted by $^\intercal$ and a dot. The Euclidean norm of a vector is denoted $|.|$ and one uses the norm $|\sigma| = \sqrt{\mathrm{Tr}(\sigma\sigma^\intercal)}$ for a matrix σ.

When the distribution V_t admits a density $v(t,x)$, one may usually rewrite (1.1) in the form:

$$
\begin{aligned}
dv(t,x) &= \left(\frac{1}{2} \sum_{i,j=1}^{d} \partial^2_{x_i x_j} [a_{ij}(x) v(t,x)] - \sum_{i=1}^{d} \partial_{x_i} [b_i(x) v(t,x)] \right) dt \\
&\quad + \left(h^\intercal(x) v(t,x) - \nabla[\gamma(x) v(t,x)] \right) dW_t.
\end{aligned} \tag{1.2}
$$

Under appropriate conditions, it is proved in [20], that the solution V to (1.1) can be characterized through the following system of diffusions:

$$
X_t = X_0 + \int_0^t \beta(X_s) ds + \int_0^t \sigma(X_s) dB_s + \int_0^t \gamma(X_s) dW_s, \tag{1.3}
$$

$$
X_0 \rightsquigarrow \mu_0,
$$

$$
\xi_t = \exp(Z_t) = \exp\left(\int_0^t h(X_s).dW_s - \frac{1}{2} \int_0^t |h(X_s)|^2 ds \right), \tag{1.4}
$$

$$
< V_t, f > = E_{W} \left[f(X_t) \xi_t \right], \tag{1.5}
$$

where B is a $\mathbb{R}^d$-Brownian motion independent of W, and E_{W} denotes the conditional expectation given W. We also denote P_{W} the corresponding conditional probability.

Actually, equation (1.1) is the so-called Zakai equation arising from nonlinear filtering problem : here, X given in (1.3) is a d-dimensional signal, and W is a q-dimensional observation process (with correlated noise when $\gamma \neq 0$) given by :

$$W_t = \int_0^t h(X_s)ds + U_t,$$

on a probability space $(\Omega, \mathcal{F}, P)$ equipped with filtration $(\mathcal{F}_t)$ under which B and U are independent Brownian motions. The nonlinear filtering problem consists in estimating the conditional distribution of X given W, i.e. we want to compute the measure-valued process π_t characterized by:

$$< \pi_t, f > = E^P[f(X_t)|\mathcal{F}_t^W],$$

where $\mathcal{F}_t^W$ is the filtration generated by the whole observation of W until t. Under suitable conditions, there exists a reference probability measure Q, such that:

$$\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t} = \xi_t = \exp\left(\int_0^t h(X_s).dW_s - \frac{1}{2}\int_0^t |h(X_s)|^2 ds\right),$$

and (B, W) are two independent Brownian motions under Q. By the Kallianpur-Striebel formula, we have

$$< \pi_t, f > = \frac{< V_t, f >}{< V_t, 1 >},$$

where

$$< V_t, f > = E_W^Q[f(X_t)\xi_t],$$

satisfies Zakai equation (1.1). From now on the symbol E will denote the expectation with respect to the probability Q.

1.1 A short discussion of related literature

Numerical approximations of Zakai equation and more generally of SPDEs have been extensively studied in the literature. We cite the survey paper [17] and the references therein. Roughly summarizing, one may classify between the following approaches:

– Approximations based on the analytic expression (1.2) vary from finite difference of finite elements methods, splitting up methods or Galerkin's approximation. We cite for instance for the finite difference method the papers of [33], [15], [16] for Zakai equation and SPDE, and for finite element methods the recent paper [35]. For the splitting up method of Zakai equation and SPDE, see [4], [11], [23], [18]. See also [34] for a time discretization analysis of θ-schemes of parabolic type SPDEs driven by a(n infinite dimensional) Wiener process.

– A first algorithm based on some uniform quantization grids of the state process is mentioned in [22].

– Another point of view, developed and studied in [24] and [5], is based on the Wiener chaos decomposition of the solution to the Zakai equation. We mention also Wong-Zakai type approximations considered in [19].

– The third approach is based on the probabilistic representation (1.5) of the solution as a weighted (or unnormalized) conditional distribution. For the Zakai Equation of nonlinear filtering problem, papers [21] and [10] develop approximation methods by replacing the signal process by a finite state Markov chain on an uniform grid prescribed *a priori.* This method is somewhat equivalent to the finite difference method.

– The so-called particular Monte Carlo method is based on a particle approximation of the conditional distribution. It has recently given raise to extensive studies, see for instance [8], [6], [7] for the nonlinear filtering problem. We will compare some of our results to those obtained in [7] (in which the diffusion X does no depend on the observation process *i.e.* $\gamma = 0$).

1.2 Contribution and organization of the paper

The first contribution of our work consists in accurately estimating the error due to time discretization on the conditional expectation (1.5). Without conditioning, classical results yield an error at most linear w.r.t. the time step δ (see for instance [1]). Here, the situation is unusual because of the conditional expectation and our analysis makes clear the role of the correlation factor between the underlying process X and the observation process W. As concerns the proof, we use Malliavin calculus techniques, but the fact that we work conditionally to W induces some specific technicalities.

In a second part, we propose a simulation algorithm for the SPDE (1.1) based on an optimal quantization approach. Basically, this means a spatial discretization of the dynamics of the Euler time-discretization (X_k, V_k) of (1.3)-(1.5) optimally fitted to its probabilistic features. To be more specific, we first recall some short background on optimal quantization of a random vector. Let $X : (\Omega, \mathcal{F}, P) \to \mathbb{R}^d$ be a random vector and let $\Gamma = \{x^1, \dots, x^N\}$ be a subset (or *grid*) of $\mathbb{R}^d$ having N elements. We approximate X by one of its Borel closest neighbour projection $\widehat{X}^\Gamma := \mathrm{Proj}_\Gamma(X)$ on Γ. Such a projection is canonically associated to a Voronoi tessellation $(C_i(\Gamma))_{1 \le i \le N}$ that is a Borel partition of $\mathbb{R}^d$ satisfying for any $i = 1, \dots, N$:

$$C_i(\Gamma) \quad \subset \quad \left\{ \xi \in \mathbb{R}^d : |\xi - x^i| = \min_j |\xi - x^j| \right\}.$$

Hence

$$\widehat{X}^\Gamma \quad = \quad \mathrm{Proj}_\Gamma(X) \ := \ \sum_{i=1}^{N} x^i \mathbf{1}_{\{X \in C_i(\Gamma)\}}.$$

As soon as $X \in L^p(\Omega, P, \mathbb{R}^d)$ the induced L^p-*quantization error* is given by

$$\|X - \widehat{X}^\Gamma\|_p = \left(E \min_{1 \le i \le N} |X - x^i|^p \right)^{\frac{1}{p}} < \infty.$$

The L^p-optimal N-quantization problem for X consists in finding a grid Γ^* which achieves the lowest L^p-quantization error among all grids of size at most N. Such an optimal grid does exist (see [14]), its size is exactly N if the support of X is infinite; it is generally not unique (except in 1-dimension where uniqueness holds when the distribution P_X of X has a log-concave density). The rate of convergence of the lowest L^p-quantization error as $N \to +\infty$ is ruled by the so-called Zador theorem (see [14]). For historical reasons, this theorem is usually stated with the p^{th} power of the L^p-quantization error, known as the *L^p-distortion.*

Theorem 1.1 *Assume that $X \in L^{p+\eta}(\Omega, P, \mathbb{R}^d)$ for some $\eta > 0$. Let f denote the probability density of the absolutely continuous part of its distribution P_X (f is possibly 0). Then,*

$$\lim_N \left(N^{\frac{p}{d}} \min_{|\Gamma| \le N} \|X - \widehat{X}^\Gamma\|_p^p \right) = J_{p,d} \|f\|_{\frac{d}{d+p}}.$$

The constant $J_{p,d}$ corresponds to the uniform distribution over $[0,1]^d$ and in that case the above $\lim_N$ also holds as an infimum.

The constant $J_{p,d}$ is unknown as soon as $d \ge 3$ although one knows that $J_{p,d} \sim (d/(2\pi e))^{\frac{p}{2}}$ as $d \to \infty$. This theorem says that the lowest L^p-quantization error goes to 0 at a $N^{-\frac{1}{d}}$-rate when $N \to \infty$. For more details about these results, we refer to [14] and the references therein.

From a computational viewpoint, no closed form is available for optimal quantization grids Γ^* except for some very specific 1-dimensional distributions like the uniform one. Several algorithms can be implemented to compute these optimal (or at least some efficient locally optimal) grids. Several of them rely on the differentiability of the L^p-distortion function as a function of the grid (viewed as a N-tuple of $(\mathbb{R}^d)^N$): if P_X is continuous, it is differentiable at any grid of size N and its gradient admits an integral representation with respect to the distribution of X. Consequently one may search for optimal grids by implementing a Newton-Raphson procedure (in 1-dimension) or a stochastic gradient descent (in d-dimension). These numerical aspects have been extensively investigated in [31] with a special attention to the d-dim normal distribution. Efficient grids for these distributions are now available for many sizes in dimensions $d = 1$ up to 10 (can be downloaded at `www.quantification.finance-mathematique.com`); the extension to the quantization of Markov chains, including its numerical aspects, has already been discussed in several papers for various fields of applications like American option pricing, nonlinear filtering, or stochastic control (see *e.g.* [2], [28], [30] or [29]).

We now briefly explain in this introduction how to apply vector quantization method to the Zakai SPDE (1.1). The process (X_k) is simply a time-discretization of a diffusion independent of V. In particular, given an observation W, (X_k) can be easily simulated and the idea is to quantize optimally at each time step k, the random vector X_k by a finite distribution $\hat{X}_k$. This provides in turn an approximation of (V_k) as the conditional distribution of $\hat{X}_k$, weighted by a Girsanov like term.

Let us mention that this approach can be applied to a wider family of stochastic SPDEs *e.g.* when the functions h and γ (and possibly β and σ in the diffusion process) depend upon V_t. This is the case of the stochastic McKean-Vlasov equation where $h \equiv 0$ and $\gamma(x,V) = \int \bar{\gamma}(x,v)V(dv)$ (V positive measure). We refer to [13] for some theoretical and numerical developments on this equation.

Our main results concerning the rate of convergence can be summed up as follows. First we prove under some regularity assumptions that the error induced by a time discretization with step δ is in general of order $\sqrt{\delta}$ although in the case $\gamma = 0$, the order of convergence is improved to δ. As concerns spatial discretization error, we obtain a $n^{\frac{3}{2}}/\bar{N}^{\frac{1}{d}}$ (where $\delta = T/n$ and $\bar{N} = N/n$ denotes the (average) size of the quantization grids used at every time step). Finally (when $\gamma \neq 0$) our global error term has the form

$$\frac{1}{\sqrt{n}} + \frac{n^{\frac{3}{2}}}{\bar{N}^{\frac{1}{d}}}.$$

Numerical experiments carried out in Section 4 suggest that a significantly better space order holds true like (when $d = 1$) $\frac{c_1 + c_2 n + o(n)}{\bar{N}}$ where $c_2 \ll c_1$.

The finite element method applied to (1.2) would provide the same kind of rate (in [35] the Wiener process W is infinite dimensional which induces worst rates for time and space discretization). However, these methods require an implicit time integration in order to be stable. This requires to invert a $N^d \times N^d$ linear system (even if it is sparse) at each time step which becomes very expensive as the dimension d grows (say $d \geq 3$ or 4).

As concerns Monte Carlo methods based on interacting particles procedures like [8] or [6], the main difference of our approach in terms of complexity is that most part of our computations (the quantization of the d-dimensional process X) can be made off-line. This compensates the dependency in d of its theoretical rate of convergence at least in medium dimensions. Since the algorithm proposed here is similar to the quantized nonlinear filters developed in [28] from a computational point of view, we refer to the detailed discussion carried out in it.

The paper is organized as follows. Section 2 is devoted to the time discretization error of the SPDE (1.1). The above result is established using Malliavin calculus techniques . We describe precisely in Section 3 the optimal quantization algorithm for the Zakai equation and we analyse the resulting error. Finally, we illustrate our results in Section 4 with several simulations concerning the Zakai equation in the linear case.

2 Time discretization error

In this section, we study the error caused by a time discretization of the system (1.3)-(1.4)-(1.5) characterizing the solution to the SPDE (1.1) on a finite time interval $[0,T]$. We consider regular discretization times $t_k = k\delta$, $k = 0,\dots,n$, where $\delta = T/n$ is the time step, and we denote $\phi(t) = \sup\{t_k : t_k \leq t\}$. We then use an Euler scheme as follows:

$$\begin{aligned} X_t^\delta &= X_0 + \int_0^t \beta(X_{\phi(s)}^\delta)ds + \int_0^t \sigma(X_{\phi(s)}^\delta)dB_s + \int_0^t \gamma(X_{\phi(s)}^\delta)dW_s, \\ Z_t^\delta &= \int_0^t h(X_{\phi(s)}^\delta).dW_s - \frac{1}{2}\int_0^t |h(X_{\phi(s)}^\delta)|^2 ds, \\ < V_t^\delta, f > &= E_W\left[f(X_t^\delta)\exp(Z_t^\delta)\right]. \end{aligned}$$

By denoting $\bar{X}_k = X^\delta_{t_k}$, $\bar{V}_k = V^\delta_{t_k}$, $\Delta\bar{B}_k = B_{t_k} - B_{t_{k-1}}$, $\Delta\bar{W}_k = W_{t_k} - W_{t_{k-1}}$, the Euler scheme reads at the discretization times t_k, $k = 0, \ldots, n$:

$$\bar{X}_{k+1} = \bar{X}_k + \beta(\bar{X}_k)\delta + \sigma(\bar{X}_k)\Delta\bar{B}_{k+1} + \gamma(\bar{X}_k)\Delta\bar{W}_{k+1}, \tag{2.1}$$

$$\bar{X}_0 = X_0 \rightsquigarrow \mu_0, \tag{2.2}$$

$$< \bar{V}_k, f > = E_W\left[f(\bar{X}_k)\exp\left(\sum_{j=0}^{k-1} g(\bar{X}_j, \Delta\bar{W}_{j+1})\right)\right], \tag{2.3}$$

where

$$g(x, \Delta\bar{W}) = h(x).\Delta\bar{W} - \frac{1}{2}|h(x)|^2\delta.$$

Denote by $\bar{P}_{k,W}(x, dx')$ the conditional probability of $\bar{X}_k$ given W and $\bar{X}_{k-1} = x$. From (2.1), we have:

$$\bar{P}_{k,W}(x, dx') \rightsquigarrow \mathcal{N}\left(x + \beta(x)\delta + \gamma(x)\Delta\bar{W}_k, \delta\sigma(x)\sigma^\intercal(x)\right).$$

As usual, we set for any $f \in \mathcal{B}(\mathbb{R}^d)$, set of bounded measurable functions on $\mathbb{R}^d$:

$$\bar{P}_{k,W}f(x) = E_W\left[f(\bar{X}_k)\middle|\, \bar{X}_{k-1} = x\right] = \int f(x')\bar{P}_{k,W}(x, dx'),$$

for any $x \in \mathbb{R}^d$. Hence, by the distribution of iterated conditional expectations, we have the following inductive formula for $\bar{V}_k$, $k = 0, \ldots, n$:

$$< \bar{V}_{k+1}, f > = < \bar{V}_k, \exp\left(g(., \Delta\bar{W}_{k+1})\right)\bar{P}_{k+1,W}f >, \tag{2.4}$$

$$\bar{V}_0 = \mu_0. \tag{2.5}$$

We denote by $BL_1(\mathbb{R}^d)$ the unit ball of bounded Lipschitz functions on $\mathbb{R}^d$:

$$BL_1(\mathbb{R}^d) = \{f : \mathbb{R}^d \mapsto \mathbb{R} \text{ satisfying } |f(x)| \leq 1 \text{ and } |f(x) - f(y)| \leq |x - y| \text{ for all } x, y\}$$

and we consider the following metric on $\mathcal{M}(\mathbb{R}^d)$:

$$\rho(V_1, V_2) = \sup\left\{|< V_1, f > - < V_2, f >|, \quad f \in BL_1(\mathbb{R}^d)\right\},$$

for any $V_1, V_2 \in \mathcal{M}(\mathbb{R}^d)$.

2.1 Main results

To simplify the following convergence analysis, we assume that coefficients are very smooth and in addition, that they satisfy a uniform ellipticity condition.

(H1) *(i)* The functions β, σ and γ are of class C^∞ with bounded derivatives.
(ii) The function h is of class C^∞, is bounded and its derivatives as well.
(iii) For some $\epsilon_0 > 0$, one has $\sigma\sigma^\intercal(x) \geq \epsilon_0$ Id uniformly in x.

We recall some notation from [12]. We set $X_t^{\delta,\lambda} = X_t^\delta + \lambda(X_t - X_t^\delta)$ and $e^{\bar{Z}_T^\delta} = \int_0^1 e^{Z_T^\delta + \lambda(Z_T - Z_T^\delta)} d\lambda$. In addition, for any smooth function $a : \mathbb{R}^d \mapsto \mathbb{R}^{d'}$ we denote its derivative by a' which is $\mathbb{R}^{d'} \otimes \mathbb{R}^d$-valued. Finally, we repeatedly use the notation $a'(t) = \int_0^1 a'(X_t^{\delta,\lambda}) d\lambda$. Now, consider the unique solution of the linear equation $\mathcal{E}_t = \mathrm{Id} + \int_0^t \beta'(s)\mathcal{E}_s\, ds + \sum_{j=1}^d \int_0^t \sigma_j'(s)\mathcal{E}_s\, dB_s^j + \sum_{j=1}^q \int_0^t \gamma_j'(s)\mathcal{E}_s\, dW_s^j$ (as usual, σ_j and γ_j are the j-th column of the matrix σ and γ). Then, Lemma 4.3 in [12] gives

$$\begin{aligned} X_t - X_t^\delta = \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1} \{ [\beta(X_s^\delta) - \beta(X_{\phi(s)}^\delta)] \\ - \sum_{j=1}^d \sigma_j'(s)[\sigma_j(X_s^\delta) - \sigma_j(X_{\phi(s)}^\delta)] - \sum_{j=1}^q \gamma_j'(s)[\gamma_j(X_s^\delta) - \gamma_j(X_{\phi(s)}^\delta)] \}\, ds \\ + \sum_{j=1}^d \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}[\sigma_j(X_s^\delta) - \sigma_j(X_{\phi(s)}^\delta)]\, dB_s^j + \sum_{j=1}^q \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}[\gamma_j(X_s^\delta) - \gamma_j(X_{\phi(s)}^\delta)]\, dW_s^j. \end{aligned} \tag{2.6}$$

For any $f \in BL_1(\mathbb{R}^d)$, we put $f_\delta(x) = E(f(x + \delta \tilde{B}_T))$ where $\tilde{B}$ is an extra d-dimensional Brownian motion independent on B and W. Clearly, f_δ is of class C_b^∞, $\|f_\delta\|_\infty + \sup_{x \neq y} \frac{|f_\delta(x) - f_\delta(y)|}{|x-y|} \leq C$, $\|f_\delta - f\|_\infty \leq C\delta$, both estimates being uniform in $BL_1(\mathbb{R}^d)$.

The main result of this section is the following.

Theorem 2.1 *Assume* **(H1)**. *For* $f \in BL_1(\mathbb{R}^d)$, *set*

$$\begin{aligned} A_1(f) &= -e^{Z_T^\delta} f_\delta'(T)\mathcal{E}_T \big[\sum_{j=1}^q \int_0^T \big(\mathcal{E}_s^{-1} \int_{\phi(s)}^s \gamma_j'(X_r^\delta)\gamma(X_{\phi(r)}^\delta) dW_r \big) dW_s^j \big], \\ A_2(f) &= -e^{\bar{Z}_T^\delta} f(X_T) \big(\sum_{i=1}^q \int_0^T \big[\int_{\phi(s)}^s h_i'(X_r^\delta)\gamma(X_{\phi(r)}^\delta) dW_r \big] dW_s^i \big), \\ A_3(f) &= -\sum_{i,j=1}^q f(X_T) e^{\bar{Z}_T^\delta} \big(\int_0^T h_i'(s)\mathcal{E}_s \big(\int_0^s \mathcal{E}_r^{-1} \big[\int_{\phi(r)}^r \gamma_j'(X_u^\delta)\gamma(X_{\phi(u)}^\delta) dW_u \big] dW_r^j \big) dW_s^i \big), \\ A_4(f) &= \frac{1}{2} e^{\bar{Z}_T^\delta} f(X_T) \int_0^T \big[(\|h\|^2)'(s)\mathcal{E}_s \big(\sum_{j=1}^q \int_0^s \mathcal{E}_r^{-1} \big(\int_{\phi(r)}^r \gamma_j'(X_u^\delta)\gamma(X_{\phi(u)}^\delta) dW_u \big) dW_r^j \big) \big] ds. \end{aligned}$$

Then, one has

$$\left\| \rho(V_T, V_T^\delta) \right\|_2 \leq C\delta + \sup_{f \in BL_1(\mathbb{R}^d)} \| E_W [A_1(f) + A_2(f) + A_3(f) + A_4(f)] \|_2,$$

with

$$\sup_{f \in BL_1(\mathbb{R}^d)} \| E_W (A_1(f) + A_2(f) + A_3(f) + A_4(f)) \|_2 \leq C\sqrt{\delta},$$

for some constant C.

Remark 2.1 The fact that $\sqrt{\delta}$ is an upper bound for the error is clear, if we use classic L^p-estimates between X and X^δ. But we know that this argument involving pathwise errors is not optimal when errors on laws are considered [1]. The result above makes clear the role of the correlation in the error on conditional expectations.

1. When there is no correlation between signal and observation, i.e. $\gamma = 0$ (which is not really relevant in a filtering problem), the four terms $A_i(f)$, $i = 1, \dots, 4$, vanish and the rate of convergence for the approximation of V_T is of order δ, the time discretization step.

2. For constant function γ, the three contributions $A_1(f), A_3(f), A_4(f)$ vanish and it remains $A_2(f)$ of order $\sqrt{\delta}$ coming from the approximation of e^{Z_T}.

3. In the general case, the error will be inexorably of order $\sqrt{\delta}$. Indeed, main contributions in the error essentially behave like $\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} (W_s - W_{t_i}) dW_s = \frac{1}{2} \sum_{i=0}^{n-1} \big([W_{t_{i+1}} - W_{t_i}]^2 - [t_{i+1} - t_i] \big)$, which L^2-norm equals $C\sqrt{\delta}$.

2.2 Proof of Theorem 2.1

The proof relies on Malliavin calculus techniques: we refer the reader to [26], from which we borrow our notations. For technical reasons, it will be useful to work with the extended Wiener process $\mathcal{W} = \begin{pmatrix} B \\ \tilde{B} \\ W \end{pmatrix}$: all the further Malliavin calculus computations are made relatively to $\mathcal{W}$. Set $H = L^2([0,T], \mathbb{R}^{2d+q})$ and denote $\tilde{X}_t^{\delta,\lambda} = X_t^{\delta,\lambda} + \frac{\delta}{\sqrt{2}} \tilde{B}_t$. For $F \in \mathbb{D}^{1,p}$, we write $\mathcal{D}F = (\mathcal{D}^B F, \mathcal{D}^{\tilde{B}} F, \mathcal{D}^W F)$ for the components relatively to the three Brownian motions B, $\tilde{B}$ and W; the partial Malliavin covariance matrix of F is denoted by $\gamma^F = \int_0^T [\mathcal{D}_t^B F, \mathcal{D}_t^{\tilde{B}} F, 0][\mathcal{D}_t^B F, \mathcal{D}_t^{\tilde{B}} F, 0]^\intercal dt = \int_0^T \mathcal{D}_t^B F [\mathcal{D}_t^B F]^\intercal dt + \int_0^T \mathcal{D}_t^{\tilde{B}} F [\mathcal{D}_t^{\tilde{B}} F]^\intercal dt$ (see Section 2.1 in [26]). Following Section 1.3 in [26], the Skorokhod integral, i.e. the adjoint operator of $\mathcal{D}$, is denoted by $\boldsymbol{\delta}$ (with a boldface symbol to avoid confusion with the time step δ): for a process u in the domain of $\boldsymbol{\delta}$, for its Skorokhod integral we write $\boldsymbol{\delta}(u)$ and $\int_0^T u_t \boldsymbol{\delta} \mathcal{W}_t$ as well.

As in section 4.5.2 of [12], a localization factor $\psi_T^\delta \in [0,1]$ will be needed in the control of residual terms to justify integration by parts formulas: it satisfies the following properties

a) for any integers k and p, $\psi_T^\delta \in \mathbb{D}^{k,p}$ and $\sup_\delta \|\psi_T^\delta\|_{\mathbb{D}^{k,p}} \leq \frac{C}{T^q}$ for some $C, q \geq 0$;

b) for any $k \geq 1$, there are $C, q \geq 0$ such that $P(\psi_T^\delta \neq 1) \leq \frac{C}{T^q} \delta^k$;

c) $\{\psi_T^\delta \neq 0\} \subset \{\forall \lambda \in [0,1] : \det(\gamma^{\tilde{X}_T^{\delta,\lambda}}) \geq \frac{1}{2} \det(\gamma^{X_T})\}$.

We omit the details of its tedious construction and we simply refer to [12] (we mention that the non degeneracy condition **(H1)** ***iii)*** is used to get the above estimates with $1/T^q$, but it could also be replaced by hypo-ellipticity type assumption). To prepare the proof, we now state a series of technical results (justified later), which will help to derive a suitable stochastic analysis conditionally on W.

Lemma 2.1 *In the following, $\Phi(W)$ stands for a functional measurable w.r.t. W, which belongs to $\mathbb{D}^\infty$.*

i) For any r.v. $Y \in L^2$, $E_W(Y)$ is the unique r.v. satisfying the equality $E(Y\Phi(W)) = E(E_W(Y)\Phi(W))$ for any functional $\Phi(W) \in \mathbb{D}^\infty$.

ii) For any $\Phi(W) \in \mathbb{D}^\infty$ and $F \in \mathbb{D}^{1,2}$, one has $\Phi(W)F \in \mathbb{D}^{1,1}$, with $\mathcal{D}^B(\Phi(W)F) = \Phi(W)\mathcal{D}^B F$ and $\mathcal{D}^{\tilde{B}}(\Phi(W)F) = \Phi(W)\mathcal{D}^{\tilde{B}} F$.

iii) For $\Phi(W)$ and G in $\mathbb{D}^\infty$, $g \in C_b^\infty$ and any multi-index α, one has

$$\begin{cases} E\left(\Phi(W)\partial^\alpha g(X_T)G\right) = E\left(\Phi(W)g(X_T)H_\alpha(X_T,G)\right), \\ \|H_\alpha(X_T,G)\|_2 \le C\frac{\|G\|_{\mathbb{D}^{k,p}}}{T^q} \end{cases} \tag{2.7}$$

for some integers k, p, q. Furthermore, if $G = 0$ on $\{\psi_T^\delta = 0\}$, then for any $\lambda \in [0,1]$, one has

$$\begin{cases} E\left(\Phi(W)\partial^\alpha g(\tilde{X}_T^{\delta,\lambda})G\right) = E\left(\Phi(W)g(\tilde{X}_T^{\delta,\lambda})H_\alpha(\tilde{X}_T^{\delta,\lambda},G)\right), \\ \|H_\alpha(\tilde{X}_T^{\delta,\lambda},G)\|_2 \le C\frac{\|G\|_{\mathbb{D}^{k,p}}}{T^q} \end{cases} \tag{2.8}$$

with some constants C, k, p, q uniform in δ and $\lambda \in [0,1]$.

The result below is one of the keys of our error analysis. The estimates of order δ are rather surprising. Indeed, at the first glance, each stochastic integral (for fixed r) at the left hand side of (2.9) is of order $\sqrt{\delta}$, but the mean over r helps in improving this estimate to get δ, provided that the processes g and h satisfy some suitable controls. Its proof is postponed to the end of this section.

Proposition 2.1 *For $g \in \mathbb{D}^\infty(H)$ and $h \in \mathbb{D}^\infty(H)$, one has*

$$\begin{aligned} \int_0^T g_r\Big(\int_{\phi(r)}^r h_u \boldsymbol{\delta}\mathcal{W}_u\Big)dr &= \int_0^T\Big(\int_0^T g_r h_u \mathbf{1}_{\phi(r)\le u\le r} dr\Big)\boldsymbol{\delta}\mathcal{W}_u \\ &\quad + \int_0^T\Big(\int_0^T \mathcal{D}_u g_r \cdot h_u \mathbf{1}_{\phi(r)\le u\le r} dr\Big)du, \end{aligned} \tag{2.9}$$

and the above random variable belongs to $\mathbb{D}^\infty$. Under extra assumptions, both terms in the r.h.s. above are of order δ.

i) Assume that $N_{k,p}(g) = \sum_{j=0}^k \left[E\left(\int_0^T \|\mathcal{D}^j g_r\|^p_{L^p([0,T]^j)} dr\right)\right]^{1/p} < +\infty$ and $N_{k,p}(h) < +\infty$ for any k and p. Then, the first term of r.h.s. of (2.9) is of order δ in $\mathbb{D}^{k,p}$, for any $k \in \mathbb{N}$ and $p > 1$:

$$\Big\|\int_0^T\Big(\int_0^T g_r h_u \mathbf{1}_{\phi(r)\le u\le r} dr\Big)\boldsymbol{\delta}\mathcal{W}_u\Big\|_{\mathbb{D}^{k,p}} \le C\ N_{k+1,q}(g) N_{k+1,q}(h)\ \delta \tag{2.10}$$

for some constants C and q depending only on k and p.

ii) Assume that $M_{k,p}(g) = \sum_{j=1}^k \sup_{0\le r\le T}\left[E\|\mathcal{D}^j g_r\|^p_{L^p([0,T]^j)}\right]^{1/p} < +\infty$ and $N_{k,p}(h) < +\infty$ for any k and p. Then, the second term of r.h.s. of (2.9) is of order δ in $\mathbb{D}^{k,p}$, for any $k \in \mathbb{N}$ and $p \ge 1$:

$$\Big\|\int_0^T\Big(\int_0^T \mathcal{D}_u g_r \cdot h_u \mathbf{1}_{\phi(r)\le u\le r} dr\Big)du\Big\|_{\mathbb{D}^{k,p}} \le C\ M_{k+1,q}(g)\ N_{k,q}(h)\ \delta \tag{2.11}$$

for some constants C and q depending only on k and p.

Let us turn to the proof of Theorem 2.1. It consists in proving

$$
\begin{aligned}
E(\Phi(W)[f(X_T^\delta)e^{Z_T^\delta} - f(X_T)e^{Z_T}]) &= E(\Phi(W)e^{Z_T^\delta}[(f - f_\delta)(X_T^\delta) - (f - f_\delta)(X_T)]) \quad (2.12)\\
&\quad + E(\Phi(W)e^{Z_T^\delta}[f_\delta(X_T^\delta) - f_\delta(X_T)]) \quad (2.13)\\
&\quad + E(\Phi(W)f(X_T)[e^{Z_T^\delta} - e^{Z_T}]) \quad (2.14)\\
&= E(\Phi(W)[A_1(f) + A_2(f) + A_3(f) + A_4(f) + R])
\end{aligned}
$$

for any functional $\Phi(W) \in \mathbb{D}^\infty$, with $\|R\|_2 = O(\delta)$ uniformly w.r.t. $f \in BL_1(\mathbb{R}^d)$. Since $\|f - f_\delta\|_\infty \le C\delta$ for $f \in BL_1(\mathbb{R}^d)$, the term (2.12) can be neglected in our expansion.

In the following computations, we simply write Φ instead of $\Phi(W)$.

2.2.1 Contribution (2.13)

A Taylor's formula combined with (2.6) and Itô's formula between $\phi(s)$ and s gives

$$
\begin{aligned}
E(\Phi e^{Z_T^\delta}[f_\delta(X_T^\delta) - f_\delta(X_T)]) &= E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha^{0,0}(u)du]ds) \quad (2.15)\\
&\quad + E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha^{0,1}(u)dB_u]ds) \quad (2.16)\\
&\quad + E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha^{0,2}(u)dW_u]ds) \quad (2.17)\\
&\quad + E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha^{1,0}(u)du]dB_s) \quad (2.18)\\
&\quad + \sum_{i=1}^d E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha_i^{1,1}(u)dB_u]dB_s^i) \quad (2.19)\\
&\quad + \sum_{i=1}^d E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha_i^{1,2}(u)dW_u]dB_s^i) \quad (2.20)\\
&\quad + E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha^{2,0}(u)du]dW_s) \quad (2.21)\\
&\quad + \sum_{i=1}^q \sum_{j=1}^d E(\Phi e^{Z_T^\delta} f'_\delta(T)\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha_{i,j}^{2,1}(u)dB_u^j]dW_s^i) \quad (2.22)\\
&\quad + E(\Phi A_1(f)), \quad (2.23)
\end{aligned}
$$

where coefficients $\alpha_\cdot^\cdot \in \mathbb{D}^\infty(H)$ with $N_{k,p}(\alpha_\cdot^\cdot) + M_{k,p}(\alpha_\cdot^\cdot) < +\infty$ for any k, p, uniformly w.r.t. δ (actually, this is a consequence of the stronger estimate $\sup_{r\in[0,T]} \|\mathcal{D}_{s_1,\cdots,s_k}^k \alpha_\cdot^\cdot(r)\|_p < \infty$, see e.g. [12]). For instance, one easily check that $\alpha_{i,j}^{2,1}(u) = -\gamma'_i(X_{\phi(u)}^\delta)\sigma_j(X_{\phi(u)}^\delta)$.

Terms in factor of Φ in (2.15)(2.18)(2.21) clearly satisfy $\|R\|_2 = O(\delta)$ (remind that $\|f'\|_\infty \le C$ uniformly in $f \in BL_1(\mathbb{R}^d)$).

The contributions (2.16) and (2.17) give a contribution of order δ in L^p-norm by an application of estimates (2.10-2.11).

Terms (2.19) contain most of the difficulties that we have to face in this error analysis: we may here give detailed arguments ((2.20) is handled in the same way). Note that

$f_\delta(x) = E(f_{\delta/\sqrt{2}}(x + \frac{\delta}{\sqrt{2}}\tilde{B}_T))$ as well for the derivatives: thus, each term of the sum in (2.19) equals

$$\int_0^1 d\lambda E(\Phi\psi_T^\delta e^{Z_T^\delta} f'_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda})\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha_i^{1,1}(u)dB_u]dB_s^i) \tag{2.24}$$

$$+\int_0^1 d\lambda E(\Phi(1-\psi_T^\delta) e^{Z_T^\delta} f'_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda})\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha_i^{1,1}(u)dB_u]dB_s^i). \tag{2.25}$$

Since $P(\psi_T^\delta \neq 1) \leq C\frac{\delta^2}{T^q}$, (2.25) provides a negligible contribution. Besides, if we transform the Itô integral w.r.t. B^i into a Lebesgue integral, using the duality relationship (see Section 1.3 in [26]) and Property ii) of Lemma 2.1, we obtain that (2.24) can be rewritten under the form

$$\begin{aligned}&\int_0^1 d\lambda E(\Phi \int_0^T \mathcal{D}_s^{B^i}[\psi_T^\delta e^{Z_T^\delta} f'_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda})\mathcal{E}_T]\mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha_i^{1,1}(u)dB_u]ds)\\ =& \sum_{\kappa:|\kappa|=1,2} \int_0^1 d\lambda E(\Phi\ \partial_x^\kappa f_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda}) \int_0^T \alpha_{\kappa,i}^{1,1}(s)[\int_{\phi(s)}^s \alpha_i^{1,1}(u)dB_u]ds)\end{aligned}$$

where the summation holds on differentiation multi-indices κ with length equal to 1 and 2. In addition the coefficients $\alpha_{\kappa,i}^{1,1}$ and $\alpha_{\kappa,i}^{1,1}$ satisfy $N_{k,p}(\alpha_{\kappa,i}^{1,1}) + M_{k,p}(\alpha_{\kappa,i}^{1,1}) < +\infty$ for any k and p. If we put $G = \int_0^T \alpha_{\kappa,i}^{1,1}(s)[\int_{\phi(s)}^s \alpha_i^{1,1}(u)dB_u]ds$, we remark that $G \in \mathbb{D}^\infty$, that $G = 0$ if $\psi_T^\delta = 0$ because of the local property of the derivative operator (Proposition 1.3.7 in [26]) and that $\|G\|_{\mathbb{D}^{k,p}} \leq C\delta$ applying Proposition 2.1. Thus, Lemma 2.1 completes the estimate, and the factor of Φ in (2.24) is of order δ in L^2-norm, uniformly w.r.t. $f \in BL_1(\mathbb{R}^d)$.

We now consider (2.22). As for (2.19), we introduce ψ_T^δ: the term with $1-\psi_T^\delta$ can be neglected as before. Using analogous computations as above, it is straightforward to see that we have to control

$$\begin{aligned}&\int_0^1 d\lambda E(\Phi\psi_T^\delta e^{Z_T^\delta} f'_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda})\mathcal{E}_T \int_0^T \mathcal{E}_s^{-1}[\int_{\phi(s)}^s \alpha_{i,j}^{2,1}(u)dB_u^j]dW_s^i)\\ =& \int_0^1 d\lambda \int_0^T\int_0^T E(\mathcal{D}_u^{B^j}[\mathcal{D}_s^{W^i}[\Phi\psi_T^\delta e^{Z_T^\delta} f'_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda})\mathcal{E}_T]\mathcal{E}_s^{-1}]\mathbf{1}_{\phi(s)\leq u\leq s}\alpha_{i,j}^{2,1}(u))du\ ds\\ =& \sum_{\kappa:|\kappa|=1,2} \int_0^1 d\lambda E(\Phi\ \partial_x^\kappa f_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda}) \int_0^T\int_0^T \hat{\alpha}_{i,j}^{\kappa,2,1}(s)\mathbf{1}_{\phi(s)\leq u\leq s}\alpha_{i,j}^{2,1}(u)du\ ds) \quad (2.26)\\ &+ \sum_{\kappa:|\kappa|=1,2} \int_0^1 d\lambda E(\int_0^T \mathcal{D}_s^{W^i}[\Phi\ \partial_x^\kappa f_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda})](\int_0^T \alpha_{i,j}^{\kappa,2,1}(s)\mathbf{1}_{\phi(s)\leq u\leq s}\alpha_{i,j}^{2,1}(u)du)ds). \quad (2.27)\end{aligned}$$

For (2.26), it is enough to apply (2.8) with $G = \int_0^T\int_0^T \hat{\alpha}_{i,j}^{\kappa,2,1}(s)\mathbf{1}_{\phi(s)\leq u\leq s}\alpha_{i,j}^{2,1}(u)du\ ds$ that clearly satisfies $\|G\|_{\mathbb{D}^{k,p}} \leq C\delta$: this proves the expected estimate of order δ. The same conclusion holds for each term in (2.27): indeed, they can be transformed in $\int_0^1 d\lambda E(\Phi\ \partial_x^\kappa f_{\delta/\sqrt{2}}(\tilde{X}_T^{\delta,\lambda})$ $\int_0^T(\int_0^T \alpha_{i,j}^{\kappa,2,1}(s)\mathbf{1}_{\phi(s)\leq u\leq s}\alpha_{i,j}^{2,1}(u)du)\boldsymbol{\delta}W_s^i)$ and we conclude with Lemma 2.1.

2.2.2 Contribution (2.14)

It can be decomposed as $E(\Phi f(X_T)[e^{Z_T^\delta} - e^{Z_T}]) = E(\Phi f(X_T)e^{\bar{Z}_T^\delta}[Z_T^\delta - Z_T])$, that is

$$E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T [h(X_{\phi(s)}^\delta) - h(X_s^\delta)].dW_s)) \tag{2.28}$$

$$+E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T [h(X_s^\delta) - h(X_s)].dW_s)) \tag{2.29}$$

$$-\frac{1}{2}E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T [\|h\|^2(X_{\phi(s)}^\delta) - \|h\|^2(X_s^\delta)]ds)) \tag{2.30}$$

$$-\frac{1}{2}E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T [\|h\|^2(X_s^\delta) - \|h\|^2(X_s)]ds)). \tag{2.31}$$

In the sequel, the main idea is to use Itô's formula and the stochastic expansion (2.6) to expand the differences $h(X_{\phi(s)}^\delta) - h(X_s^\delta)$, $h(X_s^\delta) - h(X_s)$ and so on. It will raise iterated stochastic integrals and as before, the ones for which conditional expectation w.r.t. W is of order $\sqrt{\delta}$ are essentially of type $\int_0^T \cdots (\int_{\phi(s)}^s \cdots dW_u)dW_s$ (and not $\int_0^T \cdots (\int_{\phi(s)}^s \cdots dB_u)dW_s$ or $\int_0^T \cdots (\int_{\phi(s)}^s \cdots dW_u)dB_s$).

We now go into details. Since (2.28) can be rewritten as $E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\sum_{i=1}^q \int_0^T [h_i(X_{\phi(s)}^\delta) - h_i(X_s^\delta)]dW_s^i))$, it equals

$$-E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\sum_{i=1}^q \int_0^T [\int_{\phi(s)}^s h_i'(X_r^\delta)\beta(X_{\phi(r)}^\delta)dr]dW_s^i)) \tag{2.32}$$

$$-E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\sum_{i=1}^q \int_0^T [\int_{\phi(s)}^s h_i'(X_r^\delta)\sigma(X_{\phi(r)}^\delta)dB_r]dW_s^i)) \tag{2.33}$$

$$-E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\sum_{i=1}^q \int_0^T [\int_{\phi(s)}^s h_i'(X_r^\delta)\gamma(X_{\phi(r)}^\delta)dW_r]dW_s^i)). \tag{2.34}$$

The factor of Φ in (2.32) clearly satisfies the required estimate and can be neglected. The term (2.33) can also be discarded from the main part of the error using the same arguments as for (2.22). Finally, the term (2.34) gives $A_2(f)$.

Term (2.29). Owing to (2.6), it writes $\sum_{i=1}^q E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T [h_i(X_s^\delta) - h_i(X_s)]dW_s^i))$, equals

$$-\sum_{i=1}^q \sum_{j=1}^d E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T h_i'(s)\mathcal{E}_s(\int_0^s \mathcal{E}_r^{-1}[\int_{\phi(r)}^r \sigma_j'(X_u^\delta)\sigma(X_{\phi(u)}^\delta)dB_u]dB_r^j)dW_s^i))$$

$$-\sum_{i=1}^q \sum_{j=1}^d E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T h_i'(s)\mathcal{E}_s(\int_0^s \mathcal{E}_r^{-1}[\int_{\phi(r)}^r \sigma_j'(X_u^\delta)\gamma(X_{\phi(u)}^\delta)dW_u]dB_r^j)dW_s^i))$$

$$-\sum_{i,j=1}^q E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T h_i'(s)\mathcal{E}_s(\int_0^s \mathcal{E}_r^{-1}[\int_{\phi(r)}^r \gamma_j'(X_u^\delta)\sigma(X_{\phi(u)}^\delta)dB_u]dW_r^j)dW_s^i)) \tag{2.35}$$

$$-\sum_{i,j=1}^q E(\Phi f(X_T)e^{\bar{Z}_T^\delta}(\int_0^T h_i'(s)\mathcal{E}_s(\int_0^s \mathcal{E}_r^{-1}[\int_{\phi(r)}^r \gamma_j'(X_u^\delta)\gamma(X_{\phi(u)}^\delta)dW_u]dW_r^j)dW_s^i)) \tag{2.36}$$

$$+E(\Phi R)$$

with $\|R\|_2 = O(\delta)$ by estimates (2.10-2.11). The term (2.36) gives $A_3(f)$, while the other contributions can be neglected. To justify this assertion, let us consider for instance (2.35), techniques being the same for the other ones. First, we can replace f by f_δ since $\|f - f_\delta\|_\infty \leq C\delta$. Then, three applications of duality relationship yield:

$$E(\Phi f_\delta(X_T) e^{\bar{Z}_T^\delta}(\int_0^T h_i'(s)\mathcal{E}_s(\int_0^s \mathcal{E}_r^{-1}[\int_{\phi(r)}^r \gamma_j'(X_u^\delta)\sigma(X_{\phi(u)}^\delta)dB_u]dW_r^j)dW_s^i))$$

$$= \int_0^T \int_0^T \int_0^T E(\mathcal{D}_u^B[\mathcal{D}_r^{W^j}[\mathcal{D}_s^{W^i}[\Phi f_\delta(X_T)e^{\bar{Z}_T^\delta}]h_i'(s)\mathcal{E}_s]\mathcal{E}_r^{-1}] \cdot \gamma_j'(X_u^\delta)\sigma(X_{\phi(u)}^\delta)\mathbf{1}_{\phi(r)\leq u\leq r})du\, dr\, ds.$$

The term inside the expectation can be split into a sum involving the derivative of Φ and of f. Presumably, the more difficult term to estimate is of the form

$$\int_0^T \int_0^T \int_0^T E(\mathcal{D}_r^{W^j}[\mathcal{D}_s^{W^i}[\Phi\ \partial_x^\kappa f_\delta(X_T)]]\alpha(u,r,s)\mathbf{1}_{\phi(r)\leq u\leq r})du\, dr\, ds.$$

We omit the details for the other ones which are easier to handle. Two integration by parts with fixed W (see iii) in Lemma 2.1) show that it equals

$$E(\Phi\ \partial_x^\kappa f_\delta(X_T)\int_0^T(\int_0^T(\int_0^T \alpha(u,r,s)\mathbf{1}_{\phi(r)\leq u\leq r}du)\boldsymbol{\delta}W_r^j)\boldsymbol{\delta}W_s^i).$$

Then, we conclude using (2.7) with $\| \int_0^T(\int_0^T(\int_0^T \alpha(u,r,s)\mathbf{1}_{\phi(r)\leq u\leq r}du)\boldsymbol{\delta}W_r^j)\boldsymbol{\delta}W_s^i\|_{\mathbb{D}^{k,p}} \leq C\delta$.

Term (2.30). It yields a contribution of order δ, by an application of Itô's formula and inequalities (2.10-2.11). At last, the term (2.31) is equal to $-\frac{1}{2}\int_0^T E(\Phi f(X_T)e^{\bar{Z}_T^\delta}[\|h\|^2(X_s^\delta) - \|h\|^2(X_s)])ds$: in this form, the analysis is analogous to that of (2.13) and we omit the details. It gives the contribution $A_4(f)$ and some residual terms of order δ.

2.2.3 Proof of Lemma 2.1

The two first statements are straightforward. Statement i) immediately follows from the fact that any $\Phi(W) \in L^2$ can be approximated in L^2 by a sequence of $\mathbb{D}^\infty$-r.v. using the chaos expansion (see Theorem 1.1.1 in [26]). Statement ii) is clear from the definition of $\mathbb{D}^{1,p}$, $\mathcal{D}^B$ and $\mathcal{D}^{\tilde{B}}$.

Statement iii) is an integration by parts formula, that puts the differentiation/integration only on B and $\tilde{B}$, but not on W. Its proof is an easy adaptation of Proposition 3.2.1 in [27]. The estimate (2.7) is standard using in particular $\|[\gamma^{X_T}]^{-1}\|_p \leq \frac{C}{T^q}$ under the non-degeneracy condition **(H1)** ***iii)*** (see Theorem 3.3.1 in [27]). We only prove (2.8) which is less usual because of the localization factor G. Using ii), one obtains the following equalities:

$$[\mathcal{D}^B(\Phi(W)g(\tilde{X}_T^{\delta,\lambda})), \mathcal{D}^{\tilde{B}}(\Phi(W)g(\tilde{X}_T^{\delta,\lambda}))] = \Phi(W)g'(\tilde{X}_T^{\delta,\lambda})[\mathcal{D}^B\tilde{X}_T^{\delta,\lambda}, \mathcal{D}^{\tilde{B}}\tilde{X}_T^{\delta,\lambda}],$$

$$\int_0^T \mathcal{D}_t(\Phi(W)g(\tilde{X}_T^{\delta,\lambda}))[\mathcal{D}_t^B\tilde{X}_T^{\delta,\lambda}, \mathcal{D}_t^{\tilde{B}}\tilde{X}_T^{\delta,\lambda}, 0]^\intercal dt = \Phi(W)g'(\tilde{X}_T^{\delta,\lambda})\gamma^{\tilde{X}_T^{\delta,\lambda}}.$$

Note that $\gamma^{\tilde{X}_T^{\delta,\lambda}} \geq \frac{\delta^2}{2}\mathrm{Id}$ and thus $\gamma^{\tilde{X}_T^{\delta,\lambda}}$ is invertible (it is the purpose of the small perturbation of $X^{\delta,\lambda}$ with $\delta\tilde{B}/\sqrt{2}$). Then, the duality relationship leads to

$$E(\Phi(W)\partial_{x_i}g(\tilde{X}_T^{\delta,\lambda})G)$$
$$= E(\int_0^T \mathcal{D}_t(\Phi(W)g(\tilde{X}_T^{\delta,\lambda}))[Ge^i\cdot[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathcal{D}_t^B\tilde{X}_T^{\delta,\lambda}, Ge^i\cdot[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathcal{D}_t^{\tilde{B}}\tilde{X}_T^{\delta,\lambda}, 0]^{\intercal}dt)$$
$$= E(\Phi(W)g(\tilde{X}_T^{\delta,\lambda})\int_0^T [Ge^i\cdot[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathcal{D}_t^B\tilde{X}_T^{\delta,\lambda}, Ge^i\cdot[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathcal{D}_t^{\tilde{B}}\tilde{X}_T^{\delta,\lambda}, 0]\boldsymbol{\delta}\mathcal{W}_t).$$

For longer multi-index α, we iterate the procedure and construct $H_\alpha(\tilde{X}_T^{\delta,\lambda}, G)$ by the recurrence formula $H_{\alpha'+[e^i]^{\intercal}}(\tilde{X}_T^{\delta,\lambda}, G) = \int_0^T [H_{\alpha'}(\tilde{X}_T^{\delta,\lambda}, G)e^i\cdot[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathcal{D}_t^B\tilde{X}_T^{\delta,\lambda}, H_{\alpha'}(\tilde{X}_T^{\delta,\lambda}, G)e^i\cdot[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathcal{D}_t^{\tilde{B}}\tilde{X}_T^{\delta,\lambda}, 0]\boldsymbol{\delta}\mathcal{W}_t$. Concerning the estimation on $\|H_\alpha(\tilde{X}_T^{\delta,\lambda}, G)\|_2$, note first that since the derivative operator and the Skorokhod integral are local (see Propositions 1.3.6 and 1.3.7 in [26]), one has $H_\alpha(\tilde{X}_T^{\delta,\lambda}, G) = H_\alpha(\tilde{X}_T^{\delta,\lambda}, G)\mathbf{1}_{\psi_T^\delta>0}$ owing to the property on G. Using the standard inequality $\|H_\alpha(\tilde{X}_T^{\delta,\lambda}, G)\mathbf{1}_A\|_p \leq C\|[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathbf{1}_A\|_{q_1}^{p_1}\|\tilde{X}_T^{\delta,\lambda}\|_{k_2,q_2}^{p_2}\|G\|_{\mathbb{D}^{k_3,q_3}}$ (Proposition 2.4 in [1]) combined with $\|[\gamma^{\tilde{X}_T^{\delta,\lambda}}]^{-1}\mathbf{1}_{\psi_T^\delta>0}\|_p \leq \frac{C}{T^q}$ (take into account c) of property on ψ_T^δ), we easily complete the expected estimation.

2.2.4 Proof of Proposition 2.1

To prove (2.9), take $\Psi \in \mathbb{D}^\infty$ and write using twice Fubini's theorem and the duality relationship alternatively:

$$\begin{aligned}
E\big(\Psi\int_0^T g_r(\int_{\phi(r)}^r h_u\boldsymbol{\delta}\mathcal{W}_u)dr\big) &= \int_0^T E\big(\Psi g_r(\int_{\phi(r)}^r h_u\boldsymbol{\delta}\mathcal{W}_u)\big)dr\\
&= \int_0^T\int_0^T E\big(\mathcal{D}_u[\Psi g_r]\mathbf{1}_{\phi(r)\leq u\leq r}\cdot h_u\big)du\ dr\\
&= \int_0^T E\big(\mathcal{D}_u\Psi\cdot\int_0^T g_r h_u\mathbf{1}_{\phi(r)\leq u\leq r}dr\big)du + \int_0^T E\big(\Psi\int_0^T \mathcal{D}_u g_r\cdot h_u\mathbf{1}_{\phi(r)\leq u\leq r}dr\big)du\\
&= E\big(\Psi\int_0^T(\int_0^T g_r h_u\mathbf{1}_{\phi(r)\leq u\leq r}dr)\boldsymbol{\delta}\mathcal{W}_u\big) + E\big(\Psi\int_0^T(\int_0^T \mathcal{D}_u g_r\cdot h_u\mathbf{1}_{\phi(r)\leq u\leq r}dr)du\big).
\end{aligned}$$

It is standard to check that $\int_0^T g_r(\int_{\phi(r)}^r h_u\boldsymbol{\delta}\mathcal{W}_u)dr$ belongs to $\mathbb{D}^\infty$ (see Lemma 1.3.4 in [26]). The original feature of our result is specifically related to (2.10) and (2.11). For this, we use the following general estimates, that we prove at the end.

Lemma 2.2 *For appropriately defined random variables* $(g_{r,s}, h_{u,s}, g_{r,s,u})_{r,s,u}$*, we have*

$$\left[E\left(\int_{[0,T]^j} ds \int_0^T du\Big|\int_0^T g_{r,s}h_{u,s}\mathbf{1}_{\phi(r)\le u\le r}dr\Big|^2\right)^{p/2}\right]^{1/p} \tag{2.37}$$
$$\le C_{p,q}(T)\ \delta \left[E\left(\int_{[0,T]^{j+1}} |h_{u,s}|^q\ duds\right)\right]^{1/q} \left[E\left(\int_{[0,T]^{j+1}} |g_{r,s}|^q\ drds\right)\right]^{1/q},$$

$$\left[E\left(\int_{[0,T]^j} ds \int_0^T du\Big|\int_0^T g_{r,s,u}h_{u,s}\mathbf{1}_{\phi(r)\le u\le r}dr\Big|^2\right)^{p/2}\right]^{1/p} \tag{2.38}$$
$$\le C_{p,q}(T)\ \delta \left[E\left(\int_{[0,T]^{j+1}} |h_{u,s}|^q\ duds\right)\right]^{1/q} \sup_{0\le r\le T} \left[E\left(\int_{[0,T]^{j+1}} |g_{r,s,u}|^q\ dsdu\right)\right]^{1/q},$$

for q *large enough.*

We are now in a position to derive (2.10). Consider first $k = 0$. To control the L^p-norms of the first term in the r.h.s. of (2.9), we invoke the continuity of the Skorokhod integral (Proposition 2.4.3 in [27]) to get

$$\|\int_0^T (\int_0^T g_r h_u \mathbf{1}_{\phi(r)\le u\le r}dr)\boldsymbol{\delta}\mathcal{W}_u\|_p$$
$$\le C\Big(\|\int_0^T g_r h_{.}\mathbf{1}_{\phi(r)\le .\le r}dr\|_{L^p(\Omega,H)} + \|\int_0^T \mathcal{D}(g_r h_{.})\mathbf{1}_{\phi(r)\le .\le r}dr\|_{L^p(\Omega,H^{\otimes 2})}\Big). \tag{2.39}$$

From (2.37), we easily get that the first term above is bounded by $N_{0,q}(h)N_{0,q}(h)\delta$, for q large enough. With analogous computations, the second term in the r.h.s. of (2.39) is bounded by $CN_{1,q}(h)N_{1,q}(h)\delta$. Estimates (2.10) have been proved when $k = 0$. For $k \ge 1$, the successive derivative of the r.h.s of (2.9) are standard to compute and can be expressed in a similar form than before: then, analogous computations can be performed and this proves (2.10) for any k. The derivation of (2.11) is analogous, using in addition (2.38).

Proof of Lemma 2.2 The Cauchy-Schwarz inequality yields

$$\begin{aligned}\int_0^T du\ |\int_0^T g_{r,s,u}h_{u,s}\mathbf{1}_{\phi(r)\le u\le r}dr|^2 &\le \int_0^T du|h_{u,s}|^2\Big(\int_u^{\phi(u)+\delta} |g_{r,s,u}|\ dr\Big)^2\\ &\le [\int_0^T du|h_{u,s}|^4]^{1/2}[\int_0^T du(\int_u^{\phi(u)+\delta} |g_{r,s,u}|\ dr)^4]^{1/2}\\ &\le \delta^{3/2}[\int_0^T du|h_{u,s}|^4]^{1/2}[\int_0^T du\int_u^{\phi(u)+\delta} |g_{r,s,u}|^4\ dr]^{1/2}.\end{aligned}$$

If g does not depend on u, the last term above is bounded by $\delta^{1/2}[\int_0^T |g_{r,s}|^4\ dr]^{1/2}$. Then, the derivation of (2.37) is easy, using Hölder's inequalities. To obtain (2.38), i.e. when g depends on u, the previous computation to get the missing factor $\delta^{1/2}$ does not directly work: before, one has to integrate over s and ω, the other arguments remaining unchanged.

3 Simulation of Zakai equation and quantization error

3.1 The quantization algorithm

In this section, we propose a quantization approach for the numerical implementation of formulae in (2.1), (2.3) and (2.5). Here, those formulae are written as:

$$\begin{aligned} \bar{X}_{k+1} &= \bar{X}_k + \beta(\bar{X}_k)\delta + \sigma(\bar{X}_k)\Delta\bar{B}_{k+1} + \gamma(\bar{X}_k)\Delta\bar{W}_{k+1} \\ &=: F_\delta(\bar{X}_k, \Delta\bar{B}_{k+1}, \Delta\bar{W}_{k+1}), \quad (3.1) \\ < \bar{V}_{k+1}, f > &= < \bar{V}_k, \exp\left(g(., \Delta\bar{W}_{k+1})\right) \bar{P}_{k+1,W} f > \quad (3.2) \end{aligned}$$

for $k = 0, \ldots, n-1$, with

$$g(x, \Delta W) = h(x).\Delta W - \frac{1}{2}|h(x)|^2\delta, \quad (3.3)$$

and $\bar{P}_{k+1,W}(x, dx')$ is a normal distribution with mean $x + \beta(x)\delta + \gamma(x)\Delta\bar{W}_{k+1}$ and variance $\sigma(x)\sigma^\intercal(x)\delta$.

We construct an approximation of $\bar{V}_k$ as follows. At each time t_k, $k = 0, \ldots, n$, we are given a grid $\Gamma_k = \{x_k^1, \ldots, x_k^{N_k}\}$ of N_k points in $\mathbb{R}^d$, associated to Voronoi tessellations $C_i(\Gamma_k)$, $i = 1, \ldots, N_k$:

$$C_i(\Gamma_k) = \left\{ u \in \mathbb{R}^d : |u - x_k^i| = \min_j |u - x_k^j| \right\}.$$

We then approximate the process $(\bar{X}_k)$ by the marginal quantized process $(\hat{X}_k)$ defined as:

$$\hat{X}_k = \mathrm{Proj}_{\Gamma_k}(\bar{X}_k) := \sum_{i=1}^{N_k} x_k^i 1_{\{\bar{X}_k \in C_i(\Gamma_k)\}}.$$

We thus define the conditional probability $\hat{P}_{k,W}$ of $\hat{X}_k$ given $\hat{X}_{k-1}$, and W. In other words, $\hat{P}_{k,W}$ is a (random) probability transition matrix $\{\hat{p}_{k,W}^{ij},\ i = 1, \ldots, N_{k-1}, j = 1, \ldots, N_k\}$ characterized by:

$$\hat{p}_{k,W}^{ij} = P_W\left[\hat{X}_k = x_k^j \middle| \hat{X}_{k-1} = x_{k-1}^i\right].$$

Finally, the random measure-valued process $(\bar{V}_k)$ is approximated by the discrete random measure process $(\hat{V}_k)$ defined by:

$$\begin{aligned} \hat{V}_0 &= \text{law of } \hat{X}_0, \\ < \hat{V}_{k+1}, f > &= < \hat{V}_k, \exp\left(g(., \Delta\bar{W}_{k+1})\right) \hat{P}_{k+1,W} f > . \quad (3.4) \end{aligned}$$

From an algorithmic viewpoint, this reads as:

$$\hat{V}_k = \sum_{i=1}^{N_k} \hat{v}_k^i \delta_{x_k^i}, \quad (\delta_x \text{ is the Dirac mass at } x)$$

for $k = 0, \ldots, n$, where the weights $\hat{v}_k^i$ are computed in a forward induction as:

$$\begin{aligned} \hat{v}_0^i &= \hat{p}_0^i := P[\hat{X}_0 = x_0^i] = P[\bar{X}_0 \in C_i(\Gamma_0)], \quad i = 1, \ldots, N_0, \\ \hat{v}_{k+1}^j &= \sum_{i=1}^{N_k} \hat{v}_k^i \hat{p}_{k+1,W}^{ij} \exp\left(g(x_k^i, \Delta \bar{W}_{k+1})\right), \quad j = 1, \ldots, N_{k+1}. \end{aligned}$$

The implementation of the above method requires optimally for each $k = 0, \ldots, n$:

• a grid Γ_k which minimizes the L^p-quantization error

$$\|\Delta_k\|_p = \|\bar{X}_k - \hat{X}_k\|_p$$

as well as an estimation of this error,

• the weights of the joint distribution $(\hat{X}_{k-1}, \hat{X}_k)$ and marginal distribution $\hat{X}_{k-1}$:

$$\begin{aligned} \hat{r}_{k,W}^{ij} &= P_W\left[\hat{X}_k = x_k^j,\ \hat{X}_{k-1} = x_{k-1}^i\right] = P_W\left[\bar{X}_k \in C_j(\Gamma_k),\ \bar{X}_{k-1} \in C_i(\Gamma_{k-1})\right], \\ \hat{q}_{k-1,W}^i &= P_W\left[\hat{X}_{k-1} = x_k^i\right] = P_W\left[\bar{X}_{k-1} \in C_i(\Gamma_{k-1})\right], \end{aligned}$$

for $i = 1, \ldots, N_{k-1}$, $j = 1, \ldots, N_k$, so that

$$\hat{p}_{k,W}^{ij} = \frac{\hat{r}_{k,W}^{ij}}{\hat{q}_{k-1,W}^i}.$$

This program is achieved as follows:

- Monte-Carlo simulation of M independent copies $(\bar{X}_0^{(m)}, \ldots, \bar{X}_n^{(m)})$, $m = 1, 2, \ldots, M$, distributed according to $(\bar{X}_0, \ldots, \bar{X}_n)$.

- Recursive optimization of the grids $\Gamma_0, \ldots, \Gamma_n$ by a *Competitive Learning Vector Quantization* procedure and computation of the probability weights $\hat{r}_{k,W}^{ij}$ and $\hat{q}_{k-1,W}^i$, $k = 1, \ldots, n$. As a byproduct, we also have an estimation of the L^2 quantization errors $\|\Delta_k\|_2$, $k = 0, \ldots, n$.

3.2 Analysis of quantization error

The next theorem states an error estimation for the approximation of $\bar{V}_n$ under the following condition on the coefficients of the s.d.e X:

(H2) (i) The functions β, σ and γ are Lipschitz.
(ii) The function h is bounded and Lipschitz.

Theorem 3.1 *Under* **(H2)**, *for all* $p \in [1, +\infty)$ *and* $p' > p$, *there exists a positive real constant* $C_{p,p'}$ *such that:*

$$\left\|\rho(\bar{V}_n, \hat{V}_n)\right\|_p \leq C_{p,p'} \frac{1}{\sqrt{\delta}} \sum_{k=0}^{n} \|\Delta_k\|_{p'} \qquad (\textit{with } \delta = T/n).$$

We first need the following classic result about L^p-Lipschitz property of Euler schemes.

Lemma 3.1 *Let G_δ be a functional in the form:*

$$G_\delta(x,\varepsilon) \quad = \quad x + \delta B(x) + \sqrt{\delta}\,\Sigma(x)\varepsilon,$$

where B and Σ are Lipschitz functions on $\mathbb{R}^d$, and ε is a Gaussian white noise. Then, for all $p \in [1,\infty)$, there exists a constant C_p such that for all $x, x' \in \mathbb{R}^d$:

$$\left\|G_\delta(x,\varepsilon) - G_\delta(x',\varepsilon)\right\|_p \quad \leq \quad C_p(1+\delta)|x-x'|.$$

We refer *e.g.* [30] for a detailed proof in a slightly more general setting where ε is only symmetric and lies in L^p.

One defines for every $k = 1,\ldots,n$ the operator $\bar{H}_{k,W}$ by

$$\bar{H}_{k,W}(f)(x) = \exp g(x, \Delta\bar{W}_k)\bar{P}_{k,W}(f)(x), \quad \forall f \in BL_1(\mathbb{R}^d),\ \forall x \in \mathbb{R}^d,$$

where g is defined by (3.3). One defines

$$\bar{H}_{0,W}(f) \ = <\mu_0, f> .$$

One easily checks that (with the former notations)

$$<\bar{V}_k, f> = \ E_{W}(\bar{H}_{k,W}(f)(\bar{X}_{k-1})) \ = <\bar{V}_{k-1}, \bar{H}_{k,W}(f)>$$

so that, for every $k = 0,\ldots,n$,

$$<\bar{V}_k, f> = \ (\bar{H}_{0,W} \circ \bar{H}_{1,W} \circ \cdots \circ \bar{H}_{k,W})(f).$$

This equality can be written either in forward or backward recursive form. The backward form will be an important tool for proofs:

$$\begin{aligned} \bar{U}_{n,W} f &:= f, \\ \bar{U}_{k-1,W} f &:= \bar{H}_{k,W}(\bar{U}_{k,W} f), \qquad k = 1,\ldots n. \end{aligned} \tag{3.5}$$

then, one checks *using the Markov property and the iterated conditional expectation rule* that

$$\bar{U}_{0,W} f \ = <\bar{V}_n, f> .$$

For every $k = 1,\ldots,n$, one approximates the operator $\bar{H}_{k,W}$ by its natural quantized counterpart $\widehat{H}_{k,W}$ defined on the grid $\Gamma_{k-1} = \{x^1_{k-1},\ldots,x^i_{k-1},\ldots,x^{N_{k-1}}_{k-1}\}$ by

$$\widehat{H}_{k,W}(f)(x^i_{k-1}) := \exp g(x^i_{k-1}, \Delta\bar{W}_k) \sum_j f(x^j_k) P_{W}(\widehat{X}_k = x^j_k \,|\, \widehat{X}_{k-1} = x^i_{k-1})$$

so that

$$\widehat{H}_{k,W}(f)(\widehat{X}_{k-1}) = \exp g(\widehat{X}_{k-1}, \Delta\bar{W}_k)\, E_{W}(f(\widehat{X}_k) \,|\, \widehat{X}_{k-1}).$$

Then, one sets

$$\widehat{H}_{0,W}(f) := \sum_j f(x^j_0) P_{W}(\widehat{X}_0 = x^j_0).$$

We then notice that the approximation of $\bar{V}_k$ defined in (3.4) satisfies:

$$< \widehat{V}_k, f > = \ (\widehat{H}_{0,W} \circ \widehat{H}_{1,W} \circ \cdots \circ \widehat{H}_{k,W})(f), \quad k = 1, \ldots, n. \tag{3.6}$$

Once again, this equality can be read in backward form as follows:

$$\begin{aligned} \widehat{U}_{n,W} f(x_n^i) &:= f(x_n^i),\ i = 1, \ldots, N_n, \\ \widehat{U}_{k-1,W} f(x_{k-1}^i) &:= \widehat{H}_{k,W}(\widehat{U}_{k,W} f)(x_{k-1}^i),\ \ i = 1, \ldots, N_{k-1},\ k = 1, \ldots n, \quad (3.7) \\ \text{so that} \quad < \widehat{V}_n, f > &= \widehat{U}_{0,W} f. \quad (3.8) \end{aligned}$$

The proof is designed as follows: we wish to establish a backward induction between the error terms $\|\bar{U}_{k,W} f(\bar{X}_k) - \widehat{U}_{k,W} f(\widehat{X}_k)\|_p$ at successive times k and $k+1$ involving the quantization error $\|\bar{X}_{k+1} - \hat{X}_{k+1}\|_p$ of the Euler scheme. Unfortunately a naive approach makes the final error explode because of successive use of Holder inequality. So we are led to introduce a process $\bar{Y}_k$ starting at $\bar{X}_0$ but produced by a *biased* dynamics $G_{\delta,p}$ (instead of F_δ) which corresponds to a step-by-step discrete Girsanov (implicit) change of probability. Thus we can simultaneously take advantage of the martingale property of the Doléans exponential and of the independence property of the increments $\Delta\bar{W}_k$: it makes possible not to use Hölder Inequality at a crucial step (see (3.15) below) which would cause an explosion of the constants. Finally we use a revert Girsanov change of probability to come back to the quantization error of the original dynamics $(\bar{X}_k)$.

Proof of Theorem 3.1. We will assume for convenience that $\delta = T/n \in (0,1]$ throughout the proof.

STEP 1: BACKWARD INDUCTION ON THE ERROR $\|\bar{U}_{k,W} f(\bar{Y}_k) - \widehat{U}_{k,W} f(\widehat{Y}_k)\|_p$

Set temporarily

$$\begin{aligned} G_{\delta,p}(y,v,w) &:= F_\delta(y, v, w + p\delta h(y)) \\ &= y + \delta(\beta(y) + p\gamma(y)h(y)) + \sigma(y)v + \gamma(y)w, \\ \bar{Y}_k &:= G_{\delta,p}(\bar{Y}_{k-1}, \Delta\bar{B}_k, \Delta\bar{W}_k),\ k \geq 1, \\ \bar{Y}_0 &= X_0, \\ \text{and} \quad \widetilde{Y}_k &:= F_\delta(\bar{Y}_{k-1}, \Delta\bar{B}_k, \Delta\bar{W}_k),\ k \geq 1. \end{aligned}$$

Let $\bar{\mathcal{F}}_k$ denote the σ-field $\sigma(\Delta\bar{B}_\ell, \Delta\bar{W}_\ell, \ell = 1, \ldots, k)$. Set, for every $k = 0, \ldots, n$,

$$\hat{Y}_k := \mathrm{Proj}_{\Gamma_k}(\bar{Y}_k) \quad \text{and} \quad \hat{\widetilde{Y}}_k := \mathrm{Proj}_{\Gamma_k}(\widetilde{Y}_k).$$

With these notations, one checks that for every $f \in BL_1(\mathbb{R}^d)$,

$$\bar{H}_{k,W}(f)(\bar{Y}_{k-1}) = \exp g(\bar{Y}_{k-1}, \Delta\bar{W}_k) E_W(f(\widetilde{Y}_k) \,|\, \bar{Y}_{k-1}) \tag{3.9}$$

and

$$\widehat{H}_{k,W}(f)(\widehat{Y}_{k-1}) = \exp g(\bar{Y}_{k-1}, \Delta\bar{W}_k) E_W(f(\hat{\widetilde{Y}}_k) \,|\, \hat{Y}_{k-1}). \tag{3.10}$$

Consequently

$$\begin{aligned}
\bar U_{k-1,W}f(\bar Y_{k-1}) \;-\; & \widehat U_{k-1,W}f(\widehat Y_{k-1})\\
=\; & \bar H_{k,W}(\bar U_{k,W}f)(\bar Y_{k-1}) - \widehat H_{k,W}(\widehat U_{k,W}f)(\widehat Y_{k-1})\\
=\; & (\bar U_{k-1,W}f)(\bar Y_{k-1}) - E_W\left((\bar U_{k-1,W}f)(\bar Y_{k-1})\,|\,\widehat Y_{k-1}\right)\\
& +E_W\left(\bar H_{k,W}(\bar U_{k,W}f)(\bar Y_{k-1}) - \widehat H_{k,W}(\widehat U_{k,W}f)(\widehat Y_{k-1})\,|\,\widehat Y_{k-1}\right).
\end{aligned}$$

Let us deal with the two above terms successively. The random vector $\widehat Y_{k-1}$ being a function of $\bar Y_{k-1}$ and conditional expectation $E(\,.\,|\,W,\widehat Y_{k-1})$ being an L^p-contraction, one gets

$$\begin{aligned}
\left\|\bar U_{k-1,W}f(\bar Y_{k-1}) - E_W\left((\bar U_{k-1,W}f)(\bar Y_{k-1}) \;\;|\;\; \widehat Y_{k-1}\right)\right\|_p &\le \left\|(\bar U_{k-1,W}f)(\bar Y_{k-1}) - (\bar U_{k-1,W}f)(\widehat Y_{k-1})\right\|_p\\
&+\;\left\|E_W\left((\bar U_{k-1,W}f)(\widehat Y_{k-1}) - (\bar U_{k-1,W}f)(\bar Y_{k-1})\,|\,\widehat Y_{k-1}\right)\right\|_p\\
&\le 2\left\|(\bar U_{k-1,W}f)(\bar Y_{k-1}) - (\bar U_{k-1,W}f)(\widehat Y_{k-1})\right\|_p.
\end{aligned}$$

Consequently, using the expressions (3.9) and (3.10) and once again the contraction property and the $\sigma(\bar Y_{k-1})$-measurability of $\hat Y_{k-1}$ yield

$$\begin{aligned}
\left\|\bar U_{k-1,W}f(\bar Y_{k-1}) \;-\; \right.& \left.\widehat U_{k-1,W}f(\widehat Y_{k-1})\right\|_p\\
\le\;& 2\left\|(\bar U_{k-1,W}f)(\bar Y_{k-1}) - (\bar U_{k-1,W}f)(\widehat Y_{k-1})\;\right\|_p \qquad (3.11)\\
&+\left\|e^{g(\bar Y_{k-1},\Delta\bar W_k)}(\bar U_{k,W}f)(\widetilde Y_k) - e^{g(\widehat Y_{k-1},\Delta\bar W_k)}(\widehat U_{k,W}f)(\widehat{\widetilde Y}_k)\right\|_p
\end{aligned}$$

(when $p=2$, the 2 factor can be deleted). Let us deal now with the second term of the sum in the right hand side. First note that

$$\begin{aligned}
&\left\|e^{g(\bar Y_{k-1},\Delta\bar W_k)}(\bar U_{k,W}f)(\widetilde Y_k) - e^{g(\widehat Y_{k-1},\Delta\bar W_k)}(\widehat U_{k,W}f)(\widehat{\widetilde Y}_k)\right\|_p\\
=\;&\left\|\exp g(\bar Y_{k-1},\Delta\bar W_k)\left(\bar U_{k,W}f(\widetilde Y_k) - \exp\left(g(\widehat Y_{k-1},\Delta\bar W_k) - g(\bar Y_{k-1},\Delta\bar W_k)\right)\widehat U_{k,W}f(\widehat{\widetilde Y}_k)\right)\right\|_p.
\end{aligned}$$

Set $L_p(\delta) := \exp\left((p-1)\|h\|_\infty^2\,\delta/2\right)$. A change of variable "à la Girsanov" yields for every nonnegative Borel function Θ and every $p\in(1,+\infty)$,

$$\begin{aligned}
&\left\|\exp\left(g(\bar Y_{k-1},\Delta\bar W_k)\right)\Theta(\bar Y_{k-1},\Delta\bar B_k,\Delta\bar W_k)\right\|_p^p\\
\le\;&(L_p(\delta))^p E\left(\exp\left(ph(\bar Y_{k-1}).\Delta\bar W_k\right) - p^2|h(\bar Y_{k-1})|^2\delta/2)\Theta^p(\bar Y_{k-1},\Delta\bar B_k,\Delta\bar W_k)\right)\\
\le\;&(L_p(\delta))^p E\left(\Theta^p(\bar Y_{k-1},\Delta\bar B_k,\Delta\bar W_k + p\delta h(\bar Y_{k-1})\right)
\end{aligned}$$

so that

$$\left\|\exp(g(\bar Y_{k-1},\Delta\bar W_k))\Theta(\bar Y_{k-1},\Delta\bar B_k,\Delta\bar W_k)\right\|_p \le L_p(\delta)\|\Theta(\bar Y_{k-1},\Delta\bar B_k,\Delta\bar W_k+p\delta h(\bar Y_{k-1})\|_p. \quad (3.12)$$

Applying the above inequality with $\Theta(y,v,w) = (\bar U_{k,W} f)(G_{\delta,p}(y,v,w))$ leads to

$$
\begin{aligned}
&\left\| e^{g(\bar Y_{k-1},\Delta \bar W_k)}(\bar U_{k,W} f)(\widetilde Y_k) - e^{g(\widehat Y_{k-1},\Delta \bar W_k)}(\widehat U_{k,W} f)(\widehat{\widetilde Y}_k)\right\|_p \\
\le\;& L_p(\delta)\left\|\Big(\bar U_{k,W} f(\bar Y_k) - \exp\Big(g(\widehat Y_{k-1},\Delta\bar W_k + p\delta h(\bar Y_{k-1})) - g(\bar Y_{k-1},\Delta \bar W_k + p\delta h(\bar Y_{k-1}))\Big)\widehat U_{k,W} f(\widehat Y_k)\Big)\right\|_p \\
\le\;& L_p(\delta)\left\|\bar U_{k,W} f(\bar Y_k) - \widehat U_{k,W} f(\widehat Y_k)\right\|_p \\
&+L_p(\delta)\left\|\Big(1-\exp\Big(g(\widehat Y_{k-1},\Delta\bar W_k + p\delta h(\bar Y_{k-1})) - g(\bar Y_{k-1},\Delta \bar W_k + p\delta h(\bar Y_{k-1}))\Big)\Big)\widehat U_{k,W}(f)(\widehat Y_k)\right\|_p \\
\le\;& L_p(\delta)\left\|\bar U_{k,W} f(\bar Y_k) - \widehat U_{k,W} f(\widehat Y_k)\right\|_p \\
&+L_p(\delta)\left\|1-\exp\Big(g(\widehat Y_{k-1},\Delta\bar W_k + p\delta h(\bar Y_{k-1})) - g(\bar Y_{k-1},\Delta \bar W_k + p\delta h(\bar Y_{k-1}))\Big)\right\|_{rp}\left\|\widehat U_{k,W} f(\widehat Y_k)\right\|_{sp} \qquad (3.13)
\end{aligned}
$$

where $r>1$ and $s=\frac{r}{r-1}$ are conjugate Holder exponents. Now

$$
\left\|\widehat U_{k,W} f(\widehat Y_k)\right\|_{sp} \;=\; \left\|\exp g(\widehat Y_k,\Delta\bar W_k)\,\widehat U_{k+1,W} f(\widehat Y_k)\right\|_{sp}.
$$

Applying (3.12) (with sp) yields

$$
\left\|\widehat U_{k,W} f(\widehat Y_k)\right\|_{sp} \le L_{sp}(\delta)\left\|\widehat U_{k+1,W} f(\widehat Y^{(sp)}_{k+1})\right\|_{sp}
$$

for some $\bar{\mathcal F}_{k+1}$-measurable random vector $\widehat Y^{(sp)}_{k+1}$ which we have no need to specify (since f is bounded). One derives by induction that

$$
\left\|\widehat U_{k,W} f(\widehat Y_k)\right\|_{sp} \le (L_{sp}(\delta))^{n-k}\|\widehat U_{n,W} f(\widehat Y^{(sp)}_n)\|_{sp} \le (L_{sp}(\delta))^{n-k}\|f\|_\infty \le C_{p,r,\|h\|_\infty,T}\|f\|_\infty \qquad (3.14)
$$

with $K_{p,r,\|h\|_\infty,T} = \exp\left((sp-1)\|h\|^2_\infty T/2\right)$.

Let us deal now with the L^{rp}-norm of the exponential term. First temporarily set $\widehat\Delta_k(h) := h(\widehat Y_k) - h(\bar Y_k)$. Then, standard computations show that

$$
\begin{aligned}
&\left\|1-\exp\Big(g(\widehat Y_{k-1},\Delta\bar W_k + p\,\delta\, h(\bar Y_{k-1})) - g(\bar Y_{k-1},\Delta \bar W_k + p\delta h(\bar Y_{k-1}))\Big)\right\|_{rp} \\
=\;& \left\|1-\exp\Big((p-1)\delta h(\bar Y_{k-1}).\widehat\Delta_{k-1}(h) + \widehat\Delta_{k-1}(h)\Delta\bar W_k - |\widehat\Delta_{k-1}(h)|^2\delta/2\Big)\right\|_{rp}.
\end{aligned}
$$

Now using the elementary inequality $|e^x-1|\le |x|e^{x_+}$ where $x_+ := \max(x,0)$ and the fact that $x\mapsto x_+$ is non-decreasing yield

$$
\begin{aligned}
\|1 \;-\;& \exp\Big(g(\widehat Y_{k-1},\Delta\bar W_k + p\,\delta\, h(\bar Y_{k-1})) - g(\bar Y_{k-1},\Delta \bar W_k + p\delta h(\bar Y_{k-1}))\Big)\Big\|_{rp} \\
\le\;& \left\||\widehat\Delta_{k-1}(h)|\left|(p-1)\delta h(\bar Y_{k-1}) + \Delta\bar W_k - (\widehat\Delta_{k-1}(h))\delta/2\right|\exp\left(2(p-1)\delta\|h\|^2_\infty + 2\|h\|_\infty|\Delta\bar W_k|\right)\right\|_{rp} \\
\le\;& L_{4p-3}(\delta)\sqrt{\delta}[h]_{\rm Lip}\left\|\,|\bar Y_{k-1}-\widehat Y_{k-1}|\left((p-1)\sqrt{\delta}\|h\|_\infty) + |Z_k| + \|h\|_\infty\sqrt{\delta}\right)\exp\left(2\|h\|_\infty\sqrt{\delta}|Z_k|\right)\right\|_{rp}
\end{aligned}
$$

where $Z_k := \frac{\Delta \bar{W}_k}{\sqrt{\delta}}$ is a $\mathcal{N}(0; I_d)$ random vector *independent* of $\bar{\mathcal{F}}_{k-1}$. Finally

$$\|1 - \exp\left(g(\widehat{Y}_{k-1}, \Delta \bar{W}_k + p\delta h(\bar{Y}_{k-1})) - g(\bar{Y}_{k-1}, \Delta \bar{W}_k + p\delta h(\bar{Y}_{k-1}))\right)\Big\|_{rp}$$
$$\leq C_{p,r,\delta,\|h\|_\infty,T}\sqrt{\delta}[h]_{\mathrm{Lip}}\|\widehat{Y}_{k-1} - \bar{Y}_{k-1}\|_{rp} \tag{3.15}$$

with

$$C_{p,r,\delta,\|h\|_\infty,T} = L_{4p-3}(\delta)\left\|\left((p-1)\sqrt{\delta}\|h\|_\infty) + |Z| + \sqrt{\delta}\|h\|_\infty\right)\exp\left(2\|h\|_\infty\sqrt{\delta}|Z|\right)\right\|_{rp}.$$

(Note that this real constant is increasing as a function of δ.) Plugging the estimates in (3.15) and (3.14) into (3.13) yields for every $k = 1, \dots, n$,

$$\begin{aligned}\left\|e^{g(\bar{Y}_{k-1},\Delta\bar{W}_k)}(\bar{U}_{k,W}f)(\widetilde{Y}_k) - e^{g(\widehat{Y}_{k-1},\Delta\bar{W}_k)}(\widehat{U}_{k,W}f)(\widehat{\widetilde{Y}}_k)\right\|_p &\leq L_p(\delta)\left\|\bar{U}_{k,W}f(\bar{Y}_k) - \widehat{U}_{k,W}f(\widehat{Y}_k)\right\|_p\\ &\quad + B(\delta)\|\bar{Y}_{k-1} - \widehat{Y}_{k-1}\|_{rp} \qquad (3.16)\end{aligned}$$

with $B(\delta) := C_{p,r,\|h\|_\infty,T}\sqrt{\delta}[h]_{Lip}\|f\|_\infty$ (with $C_{p,r,\|h\|_\infty,T} = C_{p,r,1,\|h\|_\infty,T}K_{p,r,\|h\|_\infty,T}L_p(1)$).

Now let us pass to the first term in the right hand side of (3.11). Let $(\bar{Y}^{k,y}_\ell)_{\ell=k,\dots,n}$ be the sequence obtained by iterating $G_{p,\delta}(., \Delta\bar{B}_\ell, \Delta\bar{W}_\ell)$ from y at time $\ell = k$ *i.e.*

$$\forall\, \ell \in \{k+1, \dots, n\}, \qquad \bar{Y}^{k,y}_\ell = G_{p,\delta}(\bar{Y}^{k,y}_{\ell-1}, \Delta\bar{B}_\ell, \Delta\bar{W}_\ell), \quad \bar{Y}^{k,y}_k := y.$$

The same proof as above shows that, for any couple (Z_{k-1}, Z'_{k-1}) of $\bar{\mathcal{F}}_{k-1}$-measurable L^p-integrable random variables

$$\begin{aligned}\left\|(\bar{U}_{k-1,W}f)(Z_{k-1}) - (\bar{U}_{k-1,W}f)(Z'_{k-1})\ \right\|_p &\leq L_p(\delta)\|\bar{U}_{k,W}(\bar{Y}^{k-1,Z_{k-1}}_k) - \bar{U}_{k,W}(\bar{Y}^{k-1,Z'_{k-1}}_k)\|_p\\ &\quad + B(\delta)\|\bar{Y}^{k-1,Z_{k-1}}_{k-1} - \bar{Y}^{k-1,Z'_{k-1}}_{k-1}\|_{rp}\end{aligned}$$

so that by induction,

$$\begin{aligned}\left\|(\bar{U}_{k-1,W}f)(\bar{Y}_{k-1}) - (\bar{U}_{k-1,W}f)(\widehat{Y}_{k-1})\ \right\|_p &\leq B(\delta)\sum_{\ell=k}^{n}(L_p(\delta))^{\ell-k}\|\bar{Y}^{k-1,\bar{Y}_{k-1}}_{\ell-1} - \bar{Y}^{k-1,\widehat{Y}_{k-1}}_{\ell-1}\|_{rp}\\ &\quad + (L_p(\delta))^{n+1-k}[f]_{\mathrm{Lip}}\|\bar{Y}^{k-1,\bar{Y}_{k-1}}_n - \bar{Y}^{k-1,\widehat{Y}_{k-1}}_n\|_{rp}.\end{aligned}$$

Now, Lemma 3.1 (applied to $G_{\delta,p}$) implies the existence of a real constant $C_{rp} > 0$ such that

$$\|\bar{Y}^{k-1,\bar{Y}_{k-1}}_\ell - \bar{Y}^{k-1,\widehat{Y}_{k-1}}_\ell\|_{rp} \leq (1 + C_{rp}\delta)^{\ell+1-k}\|\bar{Y}_{k-1} - \widehat{Y}_{k-1}\|_{rp}.$$

Setting $L'_{p,r}(\delta) = L_p(\delta)(1 + C_{rp}\delta)$ finally yields for every $k = 1, \dots, n$,

$$\begin{aligned}\left\|(\bar{U}_{k-1,W}f)(\bar{Y}_{k-1}) \right.&- \left.(\bar{U}_{k-1,W}f)(\widehat{Y}_{k-1})\ \right\|_p \leq C(\delta)\|\bar{Y}_{k-1} - \widehat{Y}_{k-1}\|_{2p}\\ \text{with} \qquad C(\delta) &= L_p(T)e^{C_{rp}}\left(C_{p,r,\|h\|_\infty,T}\frac{[h]_{\mathrm{Lip}}\|f\|_\infty\sqrt{\delta}}{L'_{p,r}(\delta) - 1} + [f]_{\mathrm{Lip}}\right). \qquad (3.17)\\ &\leq L_p(T)e^{C_{rp}}\left(C'_{p,r,\|h\|_\infty,T}\frac{[h]_{\mathrm{Lip}}\|f\|_\infty}{\sqrt{\delta}} + [f]_{\mathrm{Lip}}\right). \qquad (3.18)\end{aligned}$$

Plugging (3.16) and (3.17) into (3.11) finally yields the induction

$$\begin{aligned}\left\|\bar U_{k-1,W}f(\bar Y_{k-1})-\widehat U_{k-1,W}f(\widehat Y_{k-1})\right\|_p &\le L_p(\delta)\left\|\bar U_{k,W}f(\bar Y_k)-\widehat U_{k,W}f(\widehat Y_k)\right\|_p\\ &\quad +A(\delta)\|\bar Y_{k-1}-\widehat Y_{k-1}\|_{rp}\end{aligned}$$

with

$$\begin{aligned}A(\delta) &= B(\delta)+2\,C(\delta)\le C''_{p,r,\|h\|_\infty,T}\left([h]_{\rm Lip}\|f\|_\infty(\sqrt{\delta}+\frac{1}{\sqrt{\delta}})+[f]_{\rm Lip}\right)\\ &\le \frac{C_{p,r,\|h\|_\infty,[h]_{\rm Lip},\|f\|_\infty,[f]_{\rm Lip},T}}{\sqrt{\delta}}\end{aligned}$$

since $\delta\in(0,1]$. A new induction leads to

$$\begin{aligned}\left\|<\bar V_n,f>-<\hat V_n,f>\right\|_p &= \left\|\bar U_{0,W}f(\bar X_0)-\widehat U_{0,W}f(\widehat X_0)\right\|_p\\ &= \left\|\bar U_{0,W}f(\bar Y_0)-\widehat U_{0,W}f(\widehat Y_0)\right\|_p\\ &\le A(\delta)\sum_{k=0}^n(L_p(\delta))^k\|\bar Y_k-(\widehat U_{n,W}f)(\widehat Y_n)\|_{rp}+(L_p(\delta))^n\|(\bar U_{n,W}f)(\bar Y_n)-\widehat Y_k\|_p\\ &\le \frac{C_{p,r,\|h\|_\infty,[h]_{\rm Lip},\|f\|_\infty,[f]_{\rm Lip},T}}{\sqrt{\delta}}\sum_{k=0}^n\|\bar Y_k-\widehat Y_k\|_{rp}+L_p(T)[f]_{\rm Lip}\|\bar Y_n-\widehat Y_n\|_{rp}.\end{aligned}\quad(3.19)$$

STEP 2 (GLOBAL REVERT GIRSANOV TRANSFORM): Now, we aim to come back to $\bar X_k$ by introducing a revert Girsanov transform:

$$\|\bar Y_k-\widehat Y_k\|_{rp}^{rp}=E\left(Z_k(Z_k)^{-1}|\bar Y_k-\widehat Y_k|^{rp}\right)$$

where

$$Z_k=\exp\left(-\sum_{\ell=1}^k ph(\bar Y_{\ell-1}).\Delta\bar W_\ell-p^2|h(\bar Y_{\ell-1})|^2\delta/2\right).$$

It follows that

$$E\left(Z_k(Z_k)^{-1}|\bar Y_k-\widehat Y_k|^{rp}\right)=E\left(\exp\left(\sum_{\ell=1}^k ph(\bar X_{\ell-1}).\Delta\bar W_\ell-p^2|h(\bar X_{\ell-1})|^2\delta/2\right)|\bar X_k-\widehat X_k|^{rp}\right)$$

so that by the Holder inequality applied with two conjugate exponents r', $s'>1$,

$$\begin{aligned}\|\bar Y_k-\widehat Y_k\|_{rp}^{rp} &\le \left(E\exp\left(\sum_{\ell=1}^k s'ph(\bar X_{\ell-1}).\Delta\bar W_\ell-s'p^2|h(\bar X_{\ell-1})|^2\delta/2\right)\right)^{1/s'}\left(E|\bar X_k-\widehat X_k|^{rr'p}\right)^{1/r'}\\ &\le \exp\left(k(s'-1)p^2\|h\|_\infty^2\,\delta/2\right)\|\bar X_k-\widehat X_k\|_{rr'p}^{rp}.\end{aligned}$$

Finally

$$\|\bar Y_k-\widehat Y_k\|_{rp}\le\exp\left(kp\|h\|_\infty^2\,\delta/4\right)\|\bar X_k-\widehat X_k\|_{4p}\le C_{p,r,r',\|h\|_\infty,T}\|\bar X_k-\widehat X_k\|_{rr'p}.$$

One completes the proof by setting $r=r'=\sqrt{p'/p}>1$ and plugging this last inequality into (3.19). $\diamond$

3.3 Global error

Combining the results established in the former sections, we obtain the following result.

Theorem 3.2 *Assume* **(H1)**-**(H2)**. *Let $p' > 2$ and let $N \geq n \geq 1$. Assume that for every $k \in \{0, \ldots, n\}$, Γ_k is an $L^{p'}$-optimal grid of size $[N/(n+1)]$ for X_k. There exists a real constant C (depending on p' but not n) such that*

$$\|\rho(V_T, \widehat{V}_n)\|_2 \leq C\left(\frac{1}{n^\theta} + \frac{n^{\frac{3}{2}}}{\bar{N}^{\frac{1}{d}}}\right) \tag{3.20}$$

with $\theta = 0$ if $\gamma \equiv 0$ and $\theta = 1/2$ otherwise and $\bar{N} = N/n$.

Proof. Combining results obtained in Theorems 2.1 and 3.1, yields

$$\|\rho(V_T, \widehat{V}_n)\|_2 \leq C\left(\frac{1}{n^\theta} + \sqrt{\delta}\sum_{k=0}^{n}\|\Delta_k\|_{p'}\right)$$

where $\Delta_k = X_k - \widehat{X}_k = X_k - \mathrm{Proj}_{\Gamma_k}(X_k)$. It follows from the non parametric version of Zador's Theorem recently established in [25] that for every $p, \delta > 0$, there exists a universal real constant $C_{p,\delta}$ such that for every $N \geq 1$, and every $\mathbb{R}^d$-valued random vector Y

$$\min_{\Gamma\subset\mathbb{R}^d, |\Gamma|\leq N}\|Y - \widehat{Y}^\Gamma\|_p \leq C_{p,\delta}\|Y\|_{p+\delta}N^{-\frac{1}{d}}.$$

Applying this result to our framework yields (with $\delta = 1$)

$$\begin{aligned}\sum_{k=0}^{n}\|\Delta\|_{p'} &\leq C_{p'}\sup_n\max_{0\leq k\leq n}\|\bar{X}_k\|_{p'+1}(n+1)(N/(n+1))^{\frac{1}{d}}\\ &\leq Cn^{\frac{3}{2}}(N/n)^{-\frac{1}{d}}\end{aligned}$$

where C is a finite real constant since we know that b and σ, γ having at most linear growth the family of Euler schemes $((\bar{X}_k)_{0\leq k\leq n})_{n\geq 1}$ satisfies $\sup_n \max_{0\leq k\leq n}\|\bar{X}_k\|_r < +\infty$ for any $r > 0$. $\diamondsuit$

Remark 3.1 • The $n^{\frac{3}{2}}$ in the spatial error term of (3.20) is most likely not optimal (see Section 4). It probably comes from the specific technicalities induced by quantization. It corresponds *e.g.* to rate obtained for the "quenched error" in [7]. As shown by our numerical experiments the spatial error term behaves most likely as $O(n \times (N/n)^{-\frac{1}{d}})$ or $O((N/n)^{-\frac{1}{d}})$ depending on some stability conditions between n and $\bar{N}$ (see section 4 for a detailed explanation).

• As an example one can compare our error rate with that obtained in [7] (in the $\gamma \equiv 0$ setting) where an error bound is of the form

$$\frac{1}{n} + \sqrt{\frac{n}{M}}$$

where M denotes the number of Monte Carlo trials obtained under some regularity assumptions on the diffusion coefficients, h and f (regardless of the dimension). In this case, M can be compared with our N/n, i.e. the mean value of points per time layers in our algorithm.

4 Numerical simulations and estimation of the rates of convergence

Since the expression of the global error given by (3.20) does not separate clearly the time and space parameters, we will try in this section to investigate separately the rate of convergence in time and in space in the following (linear) case:

$$\begin{aligned} \beta(x) &= (A-\Gamma H)x, \qquad h(x) = Hx, \\ \gamma(x) &= \Gamma, \qquad \sigma(x) = \Sigma, \end{aligned}$$

where A, Γ, Σ and H are constant matrices of appropriate dimensions. We also suppose that μ_0 is a Gaussian law with mean m_0 and covariance matrix R_0. Then it is well-known that the solution to the Zakai equation (1.1) is explicitly given by:

$$< V_t, f > \;=\; \left[\int f(\hat{m}_t + R(t)^{\frac{1}{2}}x)\frac{\exp\left(-\frac{1}{2}|x|^2\right)}{(2\pi)^{\frac{d}{2}}}dx\right] < V_t, 1 >, \tag{4.1}$$

where $R(t)$ is the solution to the Riccati equation:

$$\begin{aligned} \frac{dR}{dt} &= AR + RA^{\intercal} + \Sigma\Sigma^{\intercal} + \Gamma\Gamma^{\intercal} - (RH^{\intercal} + \Gamma)(HR + \Gamma^{\intercal}), \\ R(0) &= R_0, \end{aligned} \tag{4.2}$$

$\hat{m}_t$ is solution of:

$$\begin{aligned} d\hat{m}_t &= A\hat{m}_t dt + (RH^{\intercal} + \Gamma)(dW_t - H\hat{m}_t dt), \\ \hat{m}_0 &= m_0, \end{aligned} \tag{4.3}$$

and

$$< V_t, 1 > \;=\; \exp\left(\int_0^t H\hat{m}_s.dW_s - \frac{1}{2}\int_0^t |H\hat{m}_s|^2 ds\right). \tag{4.4}$$

In other words, the normalized measure π_t defined by

$$< \pi_t, f > \;=\; \frac{< V_t, f >}{< V_t, 1 >},$$

is a Gaussian distribution with mean $\hat{m}_t$ and variance $R(t)$.

We introduce now the quantized normalized filter for a given function $f \in BL_1(\mathbb{R})$ as

$$< \hat{\pi}_k^{\delta}, f >:= \frac{< \hat{V}_k, f >}{< \hat{V}_k, 1 >}, \quad k = 0, \ldots, n,$$

where we have emphasized the dependence of the filter in $\delta = T/n$ by a superscript. The unnormalized filters $\hat{V}_k$ are computed according to algorithm (3.4).

The exact normalized filter is approximated owing to (4.1) using the following way. Since R is an explicitly known function (solution of (4.2)), it is sufficient to approximate $\hat{m}_t$, solution of the SDE (4.3) with a refined Euler scheme of step

$$\delta_{ref} = \frac{T}{1024} \ll \delta.$$

Indeed, for each path of the observation W, (4.3) and (4.4) are discretized as

$$\bar{m}_{l+1} = \bar{m}_l + \delta_{ref} A \bar{m}_l + (R(l\delta_{ref})H^\intercal + \Gamma)(W_{(l+1)\delta_{ref}} - W_{l\delta_{ref}} - H\bar{m}_l\delta_{ref}), \quad (4.5)$$

$$\bar{Z}_{l+1} = \bar{Z}_l + H\bar{m}_l.(W_{(l+1)\delta_{ref}} - W_{l\delta_{ref}}) - \frac{1}{2}|H\bar{m}_l|^2\delta_{ref}, \quad \bar{\xi}_l = \exp(\bar{Z}_l), \quad (4.6)$$

and so a very close approximation of the exact normalized filter, in the sense that it can be considered as the exact solution as long as δ remains considerably larger than δ_{ref}, is

$$< \pi^{\delta_{ref}}_{l\delta_{ref}}, f > := \int f(\bar{m}_l + R(l\delta_{ref})^{\frac{1}{2}}x) \frac{\exp\left(-\frac{1}{2}|x|^2\right)}{(2\pi)^{\frac{d}{2}}} dx,$$

where $R(t)$ is computed owing to an exact quadrature formula.

We now estimate the rate of convergence of the scheme with respect to the spatial and time discretization. In order to smoothen undesirable time oscillations of the error, we focus on the following temporal mean of the quadratic quantization error for the normalized filter, namely

$$\mathrm{Err}(\delta, \bar{N}) = \frac{1}{n} E \sum_{k=0}^{n} \left| < \hat{\pi}^\delta_k, f > - < \pi^{\delta_{ref}}_{t_k}, f > \right|^2, \quad (4.7)$$

where $t_k = k\delta = l(k)\delta_{ref}$ and $\bar{N} = N/n$ denotes the mean number of points per time layers. Then $\mathrm{Err}(\delta, \bar{N})$ is simply an approximation of the squared $L^2([0,T], dt)$-norm of the error.

We test the error for the following test functions:

$$f_0(x) = x, \quad f_1(x) = \exp(-x^2), \quad f_2(x) = \exp(-x). \quad (4.8)$$

The expectation in (4.7) is computed by a Monte Carlo method with $M = 100$ trajectories of the observations W.

The parameters of our simulations are

$$\Sigma = 1, \quad B = -0.5, \quad H = 1, \quad T = 1.$$

Such a choice of parameter is motivated by the fact that it provides not too small values for $R(t)$. Otherwise, there would not be enough points around $m_0 = 0$ in order to be able to “capture” the behaviour of the signal around its mean 0.

We will also change a bit the model and consider the following equations:

$$\begin{cases} dX_t = BX_t\,dt + \Sigma\,dB_t + \Gamma dW_t, \\ dW_t = HX_t dt + \varepsilon dU_t. \end{cases} \quad (4.9)$$

The formulæ above need to be changed as follows $\Gamma \rightsquigarrow \varepsilon\Gamma$ and $H \rightsquigarrow H/\varepsilon$. The reason for introducing this new degree of freedom on the noise level may look paradoxical since small

ε will provide large errors. But precisely, these large errors make it possible to display the rate of convergence more efficiently than with $\varepsilon = 1$ which produces smaller errors. Let us take the example of the spatial order. Indeed, we will see that as the discretization parameters $\bar{N}$ get larger and larger the error $\mathrm{Err}(\delta, \bar{N})$ is decreasing as a function of $\bar{N}$ until some threshold depending *a priori* on δ and in the number M of observations (*i.e.* paths of W). Beyond this threshold, the error becomes more or less constant because the difference with the exact solution will be of the same order of the temporal discretization. Subsequently the sum of the two errors will become indistinguishable from the temporal one. Therefore a small ε will provide bigger errors and so we will have more relevant points before reaching this threshold.

• **Estimation of the spatial discretization rate.** We first estimate the spatial rate of convergence in the case $\Gamma = 0$ (no correlation between the signal process X and the observation process W). For four values of $n = 1/\delta \in \{16, 32, 64, 256\}$, we estimate $\bar{N} \mapsto Err(\delta, \bar{N})$ with $\bar{N} = 2^{-\ell}$, $\ell = 1, \ldots, 7$. As a first step, for each value of n and of $\bar{N}$, we compute an optimal quantization $(\hat{X}_k)_k$ of the Euler scheme $(\bar{X}_k)_k$ of (4.9) (which is a version of (3.1)), according to the algorithm described in subsection 3.1. Then, for each test function f in (4.8) and each observation path of W, we compute recursively $< \hat{V}_k^\delta, f >$ and $< \hat{V}_k^\delta, 1 >$ using (3.4) and then $< \hat{\pi}_k^\delta, f >$. On the other hand, we compute the exact solutions using (4.5) and finally we compute $Err(\delta, \bar{N})$ as defined by(4.7) by summing up over the M trajectories sampled from the observation process W.

Note that since $\Gamma = 0$, the quantization optimization procedure of $(X_k)_k$ is a one shot process which does not depend on the observations W.

The results are summarized in Figures 1 and 2. It seems to have two regimes of convergence when $\bar{N}$ becomes larger. From one hand, on Figure 1 are displayed the error (4.7) for low values of n. It seems that its square root behaves like $O(1/\bar{N})$ for the three values of n before a threshold depending (linearly) on n after that the error remains unchanged.

From other hand, for high values of n (but still below $n_{ref} = 1024$), Figure 2 suggests a slower rate of convergence in $O((\bar{N})^{-1/2})$.

This suggests, having in mind (3.20) and Remark 3.1, a decomposition of the global error of the form

$$\frac{C_1}{n} + \frac{C_2(n)}{\bar{N}}$$

where $C_1 > 0$ and $C_2(n) = C_2 + c_2 n + o(n)$ with $C_2 > 0$, $c_2 > 0$ and $c_2 \ll C_2$.

For low values of n, C_2 remains constant and hence we get obviously

$$\frac{C_1}{n} \leq \frac{C_2}{\bar{N}} \quad \Longleftrightarrow \quad \bar{N} \leq Cn = \bar{N}_1^*(n),$$

and we get an order $O(1/\bar{N})$.

For high values of n, the linear part of C_2 becomes larger and hence we get obviously in the same manner

$$\frac{C_1}{n} \leq \frac{c_2 n}{\bar{N}} \quad \Longleftrightarrow \quad \bar{N} \leq C' n^2 = \bar{N}_2^*(n),$$

and hence we have the order $O((\bar{N})^{-1/2})$.

In fact, this emphasizes that the scheme needs some stability criterion involving n and $\bar{N}$ in order to converge at the true rate $O(1/\bar{N})$.

The quantization step of the algorithm can also be the cause of this rate. Indeed, during the quantization optimization of the signal X, we need to simulate at each time step an Euler increment of X in (4.9). This simulation is used to compute the weights of the "quantization tree" of X (weight of the Voronoi cells and the transition probabilities) and to process the optimization. Here the Euler increment of X, namely $\Sigma\sqrt{\delta}\,\chi$ where χ denotes a real valued normal random variable becomes very small as n grows; and so it is when $n = 256$. This implies that the Euler increment will mainly "hit" the closest cell in the upper time layer (not to speak about the ability of random number generator to simulate the tail of distributions). Consequently the transition probabilities are not computed accurately enough, given the size of the simulation and can explain the downgrading of the rate of convergence in time. One can conclude this experiment by saying that there is a "CFL" involving the mean spatial unit length and the time step parameter and a second "CFL" involving the time discretization parameter and the size of the simulation (this one has been precisely analyzed in [3]).

These results enlighten the Remark 3.1 concerning the improvement of Theorem 3.1.

• **Estimation of the time discretization rate of convergence.** Now we look for the rate of convergence with respect to δ. For that purpose, we use $\bar{N} = 100$ quantization points in each time layer. The rate of convergence in time will be estimated with

$$\Gamma \in \{0, 0.5\}, \quad \varepsilon \in \{0.1, 0.5, 1.0\}, \quad \delta = 2^{-m}, \quad m = 1, \ldots, 8.$$

Let us see now why we used normalized filter instead of non normalized one. In Figure 3 are displayed typical examples of graphs $k \mapsto < \hat{V}_k^\delta, f >$, $t \mapsto < V_t, f >$, $k \mapsto < \hat{\pi}_k^\delta, x >$ and $t \mapsto < \pi_t, x >$ for $\Gamma = 0$, $\varepsilon = 0.1$, $\delta = 1/256$ and $\bar{N} = N_n = 100$. The exact filters are still computed using (4.5) and (4.6). We verify on that example that the normalized filter seems to be better computed than the unnormalized one. It explains why we did not use the unnormalized version of the error. Indeed, for such a level of noise for the observations, ($\varepsilon = 0.1$) the unnormalized filter $< \hat{V}_k^\delta, f >$ has very large values. This is true for all tested functions f and all time discretization $\delta = 1/n$. Furthermore, it is also true on all sampled trajectories of W (not all depicted). Therefore it is difficult for numerical reasons to compute errors based on $< \hat{V}_k^\delta, f >$ for $\varepsilon = 0.1$.

Let us consider first the uncorrelated case ($\Gamma = 0$). Figure 4 shows the error plotted against the time step in a $\log - \log$ scale for f given by (4.8). We can see again that for a fixed ε given, the time error decreases until a threshold and then remains flat. We also see that this threshold grows as the inverse of the noise level ε. Before reaching this threshold, for every ε and every function f, the rate seems to be of order $\delta = 1/n$ as established in Theorem 2.1.

Let emphasize that, once again in this case, the quantization procedure does not depend on the observations. Therefore, it can be carried out *off-line*. This is no longer true in the correlated case. Then (*e.g.* if $\Gamma = 0.5$), we will have to compute $M = 100$ quantizations (one per observation path) of the signal $(X_k)_k$ for every $n \in \{2, 4, 8, 16, 32, 64, 128, 256\}$, i.e. 800 optimal grids. The previous study in the uncorrelated case seems to indicate that we need a small level of noise on the observations in order to display a rate with a significant number of time steps. This is why we have chosen $\varepsilon = 0.1$ for the simulations. Figure 5 shows the errors obtained as a function of n in a log–log scale for the functions (4.8). The rates of convergence are the same in each case. A linear regression seems to indicate a rate of $O(n^{-3/4})$ which is better than $O(n^{-1/2})$ stated in Theorem 2.1. An explanation of this unexpected behavior could be the following one. The constant in factor of the term $n^{-1/2}$ is presumably very small compared to the one associated to n^{-1}: thus, small values of n make an intermediate rate of convergence appear, while the rate $n^{-1/2}$ would be observed for larger n (in the asymptotic regime).

References

[1] V. Bally, D. Talay (1996): The distribution of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function, *Probab. Theory Related Fields*, **104**, 43-60 (1996).

[2] V. Bally, G. Pagès (2003): A quantization algorithm for solving discrete time multidimensional optimal stopping problems, *Bernoulli*, **9**, $n^0$6, pp.1003-1049.

[3] V. Bally, G. Pagès (2003), Error analysis of the optimal quantization algorithm for obstacle problems, *Stoch. Proc. and their Appl.*, **106**, $n^0$6, pp.1-40.

[4] A. Bensoussan, R. Glowinski, R. Rascanu (1989): Approximation of Zakai equation by the splitting-up method, *Stochastic Systems and Optimization*, Lect. Notes in Cont. Inf. Sci., **136**, 257-265, Springer Verlag.

[5] A. Budhiraja, G. Kallianpur (1996): Approximations to the solution of the Zakai equation using multiple Wiener and Stratonovitch integral expansions, *Stochastics and Stochastics Reports*, **56**, 271-315.

[6] D. Crisan, T. Lyons (1999): A particle approximation of the solution of the Kushner-Stratonovitch equation, *Probability Theory and Related Fields*, **115**, 549-578.

[7] D. Crisan, P. Del Moral, T. Lyons (1999): Interacting particle approximations of the e Kushner-Stratonovitch equation, *Adv. Appl. Prob.*, **31**,819-838.

[8] P. Del Moral (1995): Nonlinear filtering using random particles, *Theory of Prob. Appl.*, **40**, 690-701.

[9] L. Devroye (1986): *Non uniform random variate generation*, Springer Verlag.

[10] G. Di Masi, M. Pratelli, W. Runggaldier (1985): An Approximation for the nonlinear filtering problem with error bounds, *Stochastics and Stochastics Reports*, **14**, 247-271.

[11] P. Florchinger, F. Le Gland (1991): Time-discretization of the Zakai equation for diffusion processes observed in correlated noise, *Stochastics and Stochastics Reports*, **35**, 233-256.

[12] E. Gobet, R. Munos (2005): Sensitivity analysis using Itô-Malliavin calculus and martingales. Application to stochastic control problem, *SIAM J. Cont. Optim.*, **43**, 1676-1713.

[13] E. Gobet, G. Pagès, H. Pham, J. Printems (2005): Discretization and simulation for a class of SPDEs with applications to Zakai and McKean Vlasov equations, pre-print LPMA 958, Univ. Paris 6 & 7.

[14] S. Graf, H. Luschgy, *Foundations of quantization for random vectors*, Lecture Notes in Mathematics $n^0$1730, Springer, 2000, 230p.

[15] I. Gyöngy (1998) : Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise. I. *Potential Anal.*, **9**, 1-25.

[16] I. Gyöngy (1999) : Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise. II. *Potential Anal.*, **11**, 1-37.

[17] I. Gyöngy (2002): Approximations of stochastic partial differential equations, *Stochastic Partial Differential Equations*, Trento, Lect. Notes in Pure and Appl. Math., **227**, 287-307, Dekker.

[18] I. Gyöngy, N. Krylov (2003) : On the splitting-up method and stochastic partial differential equations, *Ann. Probab.* **31**, 564–591.

[19] Y. Hu, G. Kallianpur, J. Xiong (2002): An approximation for Zakai equation, *Appl. Math. Optim.*, **45**, 23-44.

[20] T. Kurtz and J. Xiong (1999): Particle representations for a class of nonlinear SPDEs, *Stochastic Processes and their Applications*, **83**, 103-126.

[21] H. Kushner (1979): A robust discrete state approximation of the optimal nonlinear filter for a diffusion, *Stochastics and Stochastics Reports*, **3**, 75-83.

[22] H. Korezlioglu, W.J. Runggaldier (1993): Filtering for non linear systems driven by non white noises: an approximation scheme, *Stochastics and Stochastice reports*, **44**, pp.65-102.

[23] F. Le Gland (1992) : Splitting-up approximation for SPDEs and SDEs with application to nonlinear filtering. *Stochastic partial differential equations and their applications (Charlotte, NC, 1991)*, 177–187, Lecture Notes in Control and Inform. Sci., 176, Springer, Berlin.

[24] S. Lototsky, R. Mikulevicius, B. Rozovskii (1997): Nonlinear filtering revisited: a spectral approach, *SIAM J. Cont. Optim.*, **33**, 1716-1730.

[25] H. Luschgy, G. Pagès, Functional quantization rate and mean pathwise regularity of processes with an application to Lévy processes, pre-pub. LPMA-1048, 2006.

[26] D. Nualart (1995): Malliavin calculus and related topics, Springer Verlag.

[27] D. Nualart (1998): Analysis on Wiener space and anticipating stochastic calculus, Lectures on probability theory and statistics (Saint-Flour, 1995), 123-167, Springer Berlin.

[28] G. Pagès, H. Pham (2005): Optimal quantization methods for non linear filtering with discrete-time observations, *Bernoulli*, **11**, 893-932.

[29] G. Pagès, H. Pham, J. Printems (2004): Optimal quantization methods and applications to numerical problems in finance, *Handbook of numerical methods in finance*, ed. S. Rachev, Birkhauser.

[30] G. Pagès, H. Pham, J. Printems (2004), An Optimal Markovian Quantization Algorithm for Multidimensional Stochastic Control Problems, *Stochastics and Dynamics*, **4**, 501-545.

[31] G. Pagès, J. Printems (2003): Optimal quadratic quantization for numerics: the Gaussian case, *Monte Carlo Methods and Applications*, **9**, $n^0 2$, pp.135-166.

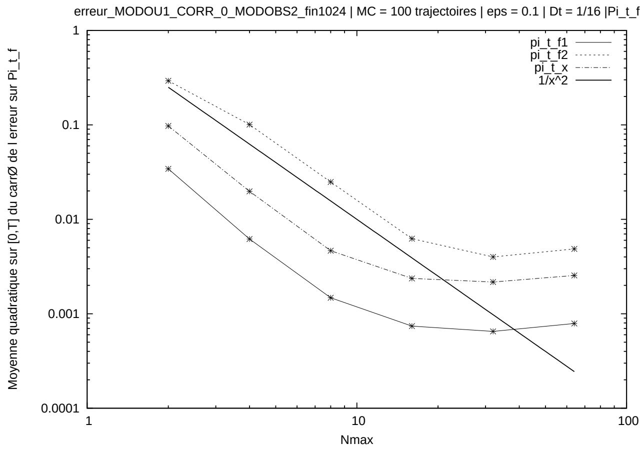
[32] J. Picard (1984): Approximations of nonlinear filtering problems and order of convergence, *Filtering and Control of Random Processes*, Lect. Notes in Cont. Inf. Sci., **61**, 219-236, Springer Verlag.

[33] M. Piccioni (1987): Convergence of implicit discretization schemes for linear differential equations with application to filtering, *Stochastic partial differential equations and applications*, Trento, Lect. Notes in Math., 1236, 208-229, Springer.

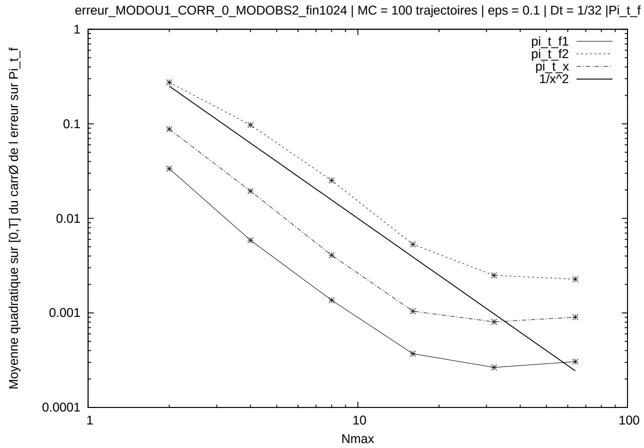
[34] J. Printems (2001): On the discretization in time of parabolic stochastic partial differential equations, *Math. Model. and Numer. Anal.*, **35**(6), 1055-1078.

[35] J.B. Walsh (2005) : Finite element methods for parabolic stochastic PDE's. *Potential Anal.*, **23**, 1–43.

(a) $n = 16$



(b) $n = 32$



(c) $n = 64$

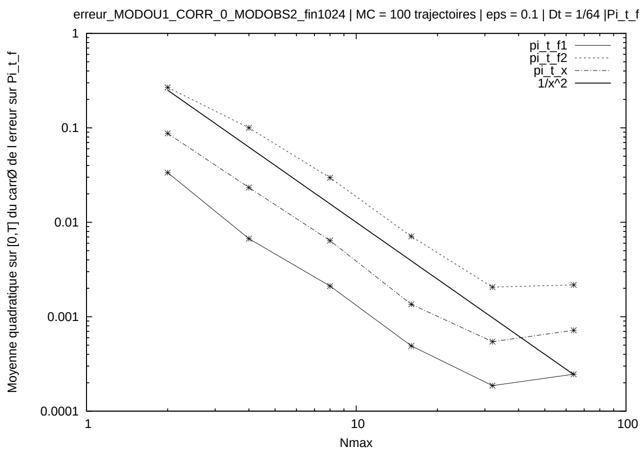


Figure 1: Error $Err(\delta, \bar{N})$ as a function of $\bar{N}$ for several time discretization n. The straight line depicts $\bar{N} \mapsto 1/\bar{N}^2$ and the dash lines denotes the errors computed with the different functions (4.8). Here $\varepsilon = 0.1$.

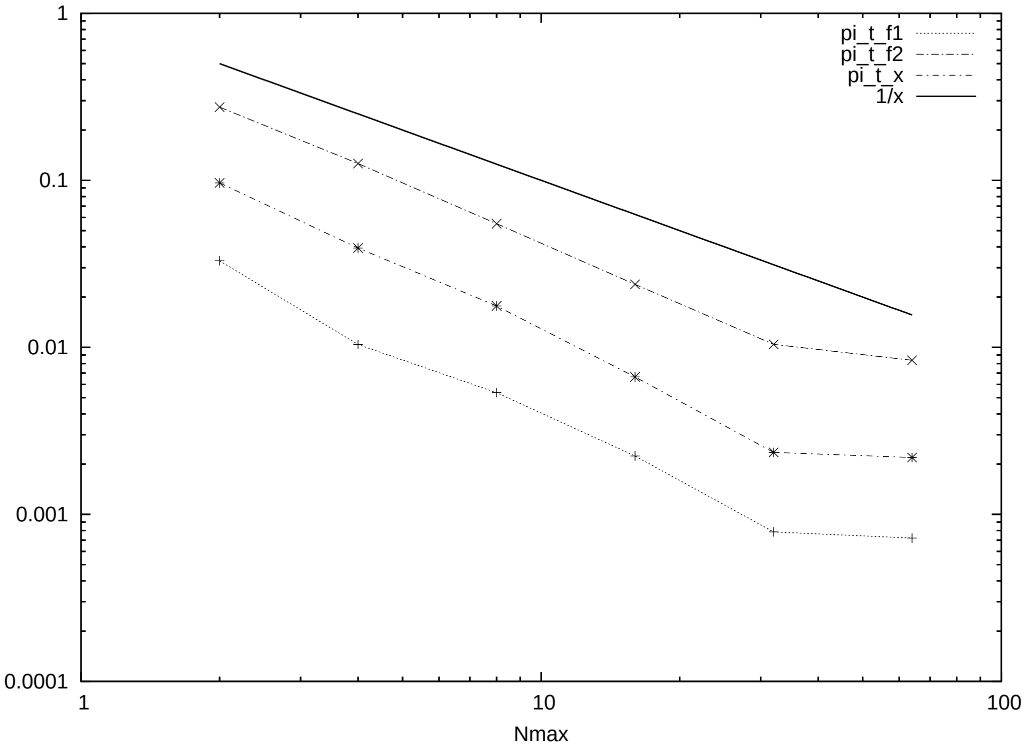
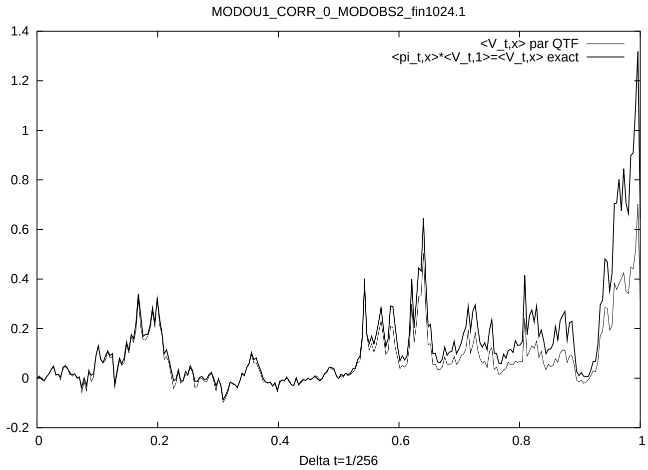
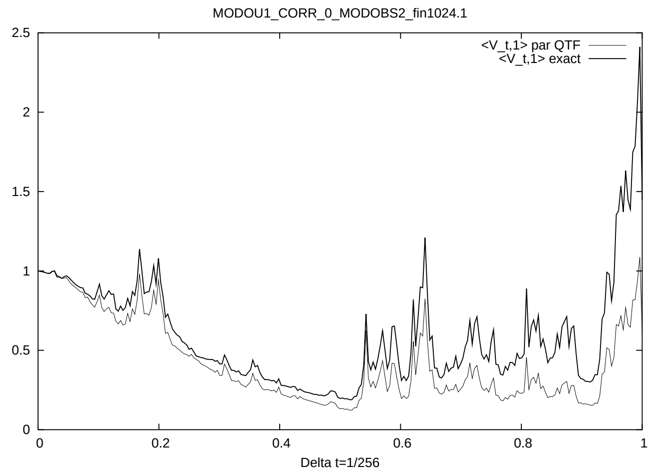


Figure 2: Rate of convergence of (4.7) with $n = 256$. Here again $\varepsilon = 0.1$.

a)



b)



c)

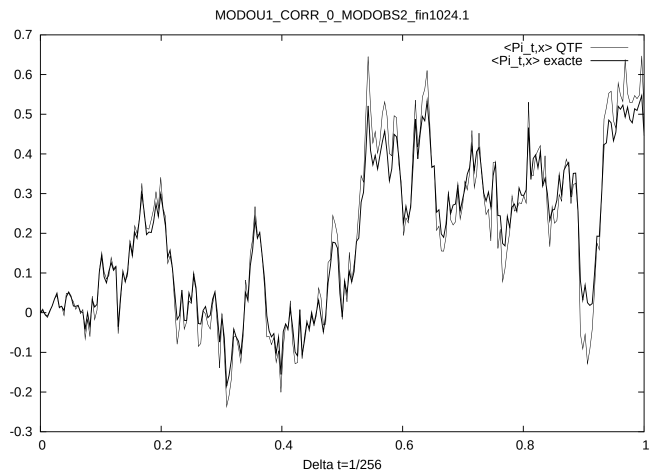
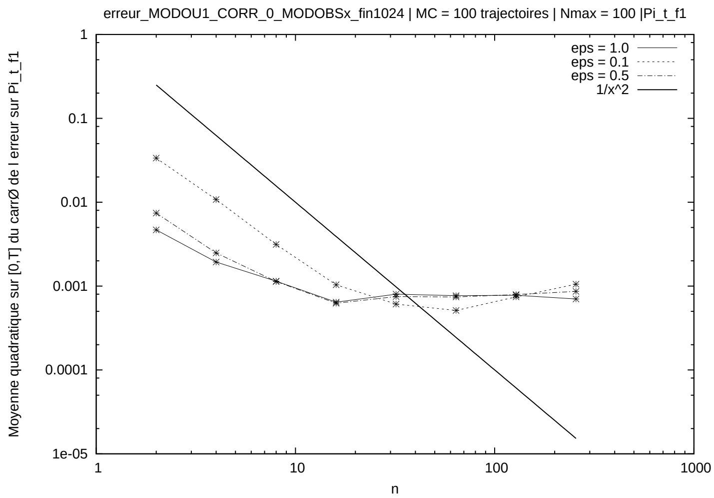
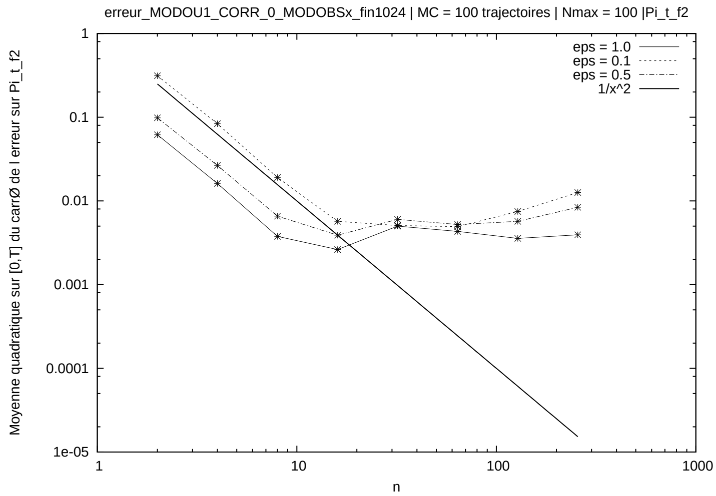


Figure 3: Examples of curves $k \mapsto < \hat{V}_k^\delta, x >$ a), $k \mapsto < \hat{V}_k^\delta, 1 >$ b), $k \mapsto < \hat{\pi}_k^\delta, x >$ c) with $\delta = 1/256$ and $N_n = 100$ computed with the same trajectory of observation. Here $\varepsilon = 0.1$ and $\Gamma = 0$. The thick line depicts the exact filter computed according a time step $\delta_{ref} = 1/1024$ and the thin line the quantized filter.

a)



b)



c)

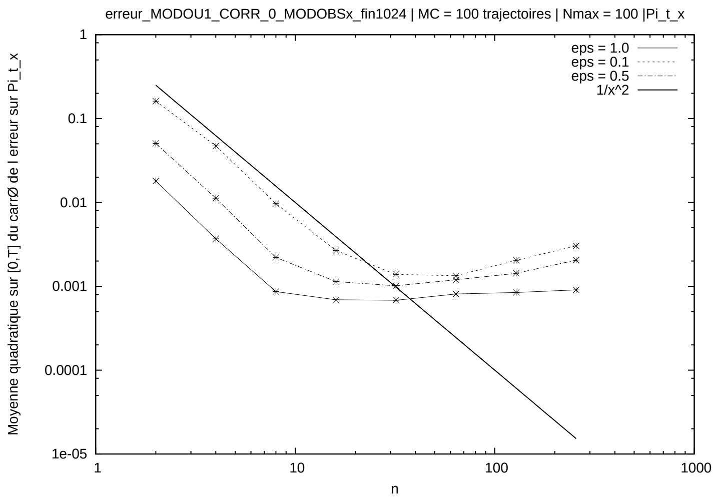


Figure 4: Square of the error (4.7) where $f(x) = \exp(-x^2)$ a), $f(x) = \exp(-x)$ b) and $f(x) = x$ c) as a function of the time step n in a log–log scale. Non correlated case.

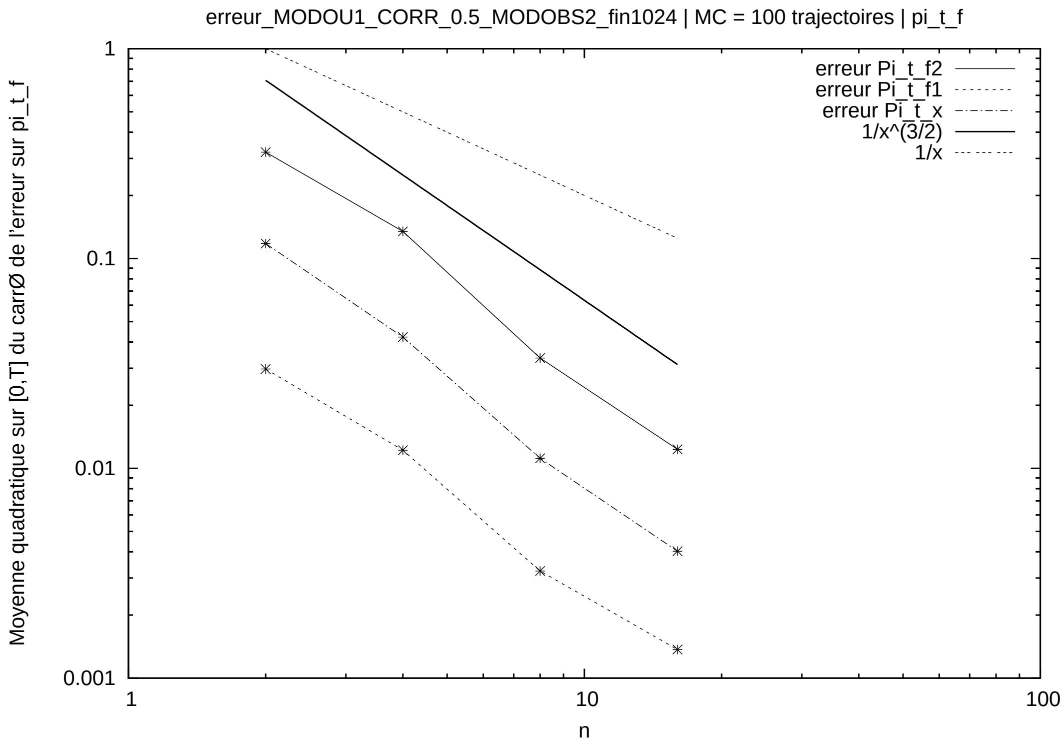


Figure 5: Error (4.7) as a function of the time step n in a log–log scale. Correlated case. The three functions I_d, $f_1(x) = \exp(-x)$ and $f_2(x) = \exp(-x^2))$ are depicted.