



# An interpolated Stochastic Algorithm for Quasi-Linear PDEs

François Delarue, Stéphane Menozzi

## ► To cite this version:

François Delarue, Stéphane Menozzi. An interpolated Stochastic Algorithm for Quasi-Linear PDEs. Mathematics of Computation, 2008, 77, p. 125-158. hal-00021967

**HAL Id: hal-00021967**

**<https://hal.science/hal-00021967>**

Submitted on 30 Mar 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# AN INTERPOLATED STOCHASTIC ALGORITHM FOR QUASI-LINEAR PDES

FRANÇOIS DELARUE AND STÉPHANE MENOZZI

ABSTRACT. In this paper, we improve the forward-backward algorithm for quasi-linear PDEs introduced in Delarue and Menozzi [8]. The new discretization scheme takes advantage of the standing regularity properties of the true solution through an interpolation procedure. For the convergence analysis, we also exploit the optimality of the square Gaussian quantization used to approximate the conditional expectations involved.

The resulting bound for the error is closely related to the Hölder exponent of the second order spatial derivatives of the true solution and turns out to be more satisfactory than the one previously established.

## 1. INTRODUCTION

**1.1. Short Overview of Numerical Schemes for BSDEs.** The theory for Backward SDEs (*cf.* Pardoux and Peng [24] for the original background) enjoys nowadays a new development through numerical applications. As the classical theory introduced during the 90's for backward equations does, the numerical counterpart offers a double panorama: each discretization procedure for BSDEs provides a conceivable scheme for a certain class of non-linear PDEs and vice and versa. Both implications make sense (see e.g. Douglas *et al.* [9] for a PDE to BSDE approach), but the trend in the current probabilistic literature now consists in exhibiting purely stochastic algorithms for BSDEs and then in deriving alternative methods to analytical finite-difference or finite-element strategies for non-linear PDEs. Of course, this raises the question of the competitiveness of the standing probabilistic methods and draws the objective for the next years: refine as much as possible the earlier algorithms to decrease at most the underlying approximation error and take advantage of the specific stochastic structure to investigate new fields of application (SPDEs, homogenization...).

In this work, we are concerned with non-linear Cauchy problems on  $[0, T] \times \mathbb{R}^d$  of the following form ( $\nabla_x u$  stands for the  $x$ -gradient of  $u$ , seen as a row vector, and  $H_u$  for the  $x$ -Hessian matrix of  $u$ )

$$(\mathcal{E}) \begin{cases} \partial_t u(t, x) + \nabla_x u(t, x)b(x, u(t, x), v(t, x)) \\ + \frac{1}{2}\text{tr}(a(x, u(t, x))H_u(t, x)) + f(x, u(t, x), v(t, x)) = 0, \\ u(T, x) = H(x), \end{cases}$$

with  $v(t, x) \equiv \nabla_x u(t, x)\sigma(x, u(t, x))$ . The stochastic counterpart of  $(\mathcal{E})$  writes as a “fully coupled” Forward Backward Stochastic Differential Equation (FBSDE in

---

Prepublication March 30, 2006.

2000 *Mathematics Subject Classification.* Primary 65C30; secondary 60H10, 60H35.

short). Namely, for a given starting point  $x_0 \in \mathbb{R}^d$ , we consider a diffusion process  $U$  strongly coupled to the solution  $(V, W)$  of a BSDE by the relation

$$(E) \quad \forall t \in [0, T], \begin{cases} U_t = x_0 + \int_0^t b(U_s, V_s, W_s) ds + \int_0^t \sigma(U_s, V_s) dB_s, \\ V_t = H(U_T) + \int_t^T f(U_s, V_s, W_s) ds - \int_t^T W_s dB_s, \end{cases}$$

where  $B$  is a  $d$ -dimensional Brownian motion on a certain filtered probability space  $(\Omega, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  and  $\sigma(x, y)$  a square root of the diffusion matrix  $a(x, y)$ . For a guided tour of the connection between  $(\mathcal{E})$  and  $(E)$ , we refer the reader to Antonelli [1], Ma, Protter and Yong [18], Ma and Yong [19], Delarue [6] and more recently Delarue and Guatteri [7]. Generally speaking, if  $u$  denotes, under suitable assumptions, the solution of  $(\mathcal{E})$ , the backward component of  $(E)$  writes at time  $t$ :  $(V_t, W_t) = (u(t, U_t), v(t, U_t))$ . Concerning the numerical approximation, we mention the works of Douglas *et al.* [9] and Milstein and Tretyakov [20], [21] and [22].

Probabilistic algorithms for  $(\mathcal{E})$  consist in discretizing the following non-linear form of the dynamic programming principle:  $u(t, U_t) = \mathbb{E}[u(t+h, U_{t+h}) | \mathcal{F}_t] + \mathbb{E}[\int_t^{t+h} f(U_s, V_s, W_s) ds | \mathcal{F}_t]$ .

*Autonomous Case.* In the decoupled case, the forward component can be approximated with a standard Euler scheme so that  $V_t = u(t, U_t)$  can be reached provided a suitable estimation of  $(V, W)$  at time  $t+h$ .

Once the approximation of  $V_t$  is available, the next step to iterate the process consists in updating the approximation of the representation process  $W$ . To this end, one usually uses the so called *martingale increment technique*, see e.g. Bally *et al.* [2] or Bouchard and Touzi [3]. Basically, this amounts to say that  $W_t \approx h^{-1} \mathbb{E}[V_{t+h}(B_{t+h} - B_t) | \mathcal{F}_t]$ .

Monte-Carlo techniques are then well fitted to the effective computations of the underlying conditional expectations. Due to the Markov property for  $U$  and to the relationship  $(V_t, W_t) = (u(t, U_t), v(t, U_t))$ , for  $t \in [0, T]$ , these latter reduce to conditional expectations with respect to  $\sigma$ -fields generated by a random vector. Several regression methods are then conceivable: Bouchard and Touzi [3] refer to Malliavin calculus techniques (this involves a rather large number of simulated paths for the underlying diffusion process), and Lemor, Gobet and Warin [16] and [17] make use of a finite function basis (this allows to use the same paths for the approximations of the forward and backward processes).

*Coupled Case.* All the previous methods require an *a priori* discretized version for the process  $U$  and thus fail in our frame, except by considering a global fixed point strategy for the triple  $(U, V, W)$ : given a first  $U$ , compute the associated  $(V, W)$ , and then plug this  $(V, W)$  to compute a new  $U$  and so on. This is not conceivable from a numerical point of view.

The common strategy in the coupled case relies on spatial grids (see e.g. Delarue and Menozzi [8] and Milstein and Tretyakov [23]). At time  $t$ , the initial condition of the process  $U$  in the dynamic programming principle is chosen as a deterministic node  $x$  of a Cartesian spatial grid. Given, for a small  $h > 0$ , an approximation  $(\bar{u}(t+h, x), \bar{v}(t+h, x))$  of the solution of the PDE and of its gradient at  $(t+h, x)$ , this permits to approximate the transition of the diffusion from time  $t$  to time  $t+h$  and to derive an approximation of  $u(t, x)$ . The martingale increment technique provides an approximation of the gradient. Such a procedure can be iterated along a temporal mesh of step  $h$ . Underlying expectations are then estimated with a

quantization argument, that turns out to be cheaper than a classical Monte-Carlo method.

Anyhow, the approximate transition plugged in the dynamic programming principle is supported by a different set than the grid itself, so that the approximated solution  $\bar{u}$  has to be extended from the spatial grid to the whole space. In [8], the considered extension is piecewise constant and thus discontinuous. We propose here to extend  $\bar{u}$  through a piecewise linear interpolation procedure to take the utmost advantage of the standing regularity for the true solution  $u$  (see Milstein and Tretyakov [20], [21] and [22] for a similar procedure).

**1.2. Contribution and Prospects of the Paper.** The numerical analysis we provide in this paper appears as a new improvement towards competitive probabilistic algorithms for quasi-linear PDEs. As in [20], [21] and [22], the global bound we exhibit below (see Theorem 3.2) just holds for  $b$  independent of  $W$ . However, we feel that it is the first one to apply both to an interpolated stochastic scheme and to a classical solution  $u \in \mathcal{C}^{1+\alpha/2, 2+\alpha}([0, T] \times \mathbb{R}^d, \mathbb{R})$ ,  $\alpha$  being possibly small. By way of example, the solutions are required to be at least twice differentiable in time (and therefore four times in space) in the different papers of Milstein and Tretyakov. Generally speaking, the explanation for these different regularity assumptions follows from the error analysis, *i.e.* from the proofs of the convergence of the underlying algorithm, and not from the very definitions of the algorithms.

Two main conclusions follow from Theorem 3.2. First, as forecasted from purely numerical experiments in [8] and as already proved by Milstein and Tretyakov in the very regular frame, the piecewise linear interpolation procedure reduces the error with respect to the piecewise constant one. We prove here that the gain between both is exactly the one expected. Second, we prove that the algorithm still converges for a low number of points for the underlying quantization of the Brownian increments. In this sense, we recover the results observed in the papers of Milstein and Tretyakov where the Brownian motion is approximated by a simple random walk. We also improve our previous work in which quantization is assumed to be “large” enough to ensure the convergence.

The case where  $b$  depends on  $W$  is crucial for applications: the so-called deterministic KPZ equations, *i.e.* the heat equation forced by the square norm of the gradient of the solution, both appears in statistical mechanics and in finance. We manage to establish the convergence of the algorithm in this setting provided the quantization is large enough (see Section 7). Even if not completely satisfactory, this result is, to the best of our knowledge, new in the probabilistic literature devoted to the subject. We also show that the interpretation of the product  $(\nabla_x u)b(x, u, v)$  as a quadratic second member totally fails from a numerical point of view.

Beyond these remarks, several questions are to be investigated in future contributions. First, the interpolation procedure we consider here is well fitted to our own setting since the Lagrange kernel of order one can be interpreted as a family of probability weights. This fails for higher order kernels so that the interest of an interpolation procedure of order two remains open. Second, the real influence of the quantization in the low regular framework (*i.e.* for  $\alpha$  close to zero) is rather subtle to get in light of numerical experiments and is to be understood. Indeed, for a solution with isolated “singularities” (*i.e.* a solution for which  $H_u$  is smooth except in several points), the error observed on various examples may vary with the number of points for the quantization.

Finally, our analysis of the algorithm provides a possible discretization procedure for the FBSDE (E). For systems driven by Lipschitz continuous coefficients, this discretization turns out to be strongly convergent, as already shown in our previous paper. In the case of space Hölder continuous coefficients, as it may be under our assumptions, the error has to be analyzed in the weak sense. We don't investigate this point in the current work.

**1.3. Organization of the Paper.** In Section 2, we state our working assumptions. In Section 3, we introduce the interpolated algorithm and the associated convergence result. Section 4 is dedicated to numerical illustrations. The different proofs are given in the remaining parts of the paper. Section 7 is specifically devoted to the case  $b = b(x, u, v)$ .

## 2. WORKING ASSUMPTIONS AND ASSOCIATED PROPERTIES

For a given  $d \in \mathbb{N}^*$ , we consider the coefficients  $b : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $f : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $\sigma : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^{d \times d}$ ,  $H : \mathbb{R}^d \rightarrow \mathbb{R}$ .

**Assumption (A)** The functions  $b$ ,  $f$ ,  $H$  and  $\sigma$  are said to satisfy Assumption (A) if they are bounded in space, have at most linear growth in the other variables, and are uniformly  $\alpha$ -Hölder continuous in  $x$ ,  $\alpha > 0$ , and uniformly Lipschitz continuous w.r.t. all the variables, if  $a \equiv \sigma\sigma^*$  is uniformly elliptic and if  $H$  is bounded in  $C^{2+\alpha}(\mathbb{R}^d)$ .

From now on, Assumption (A) is in force. We denote by  $|\cdot|$  the Euclidean norm of  $\mathbb{R}^d$ , and by  $\langle \cdot, \cdot \rangle$  the associated inner product.

**2.1. Forward-Backward SDE.** Consider now a given  $T > 0$  and an initial condition  $x_0 \in \mathbb{R}^d$ . According to Delarue and Guatteri [7], there exists a filtered probability space  $(\Omega, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$  endowed with a  $d$ -dimensional Brownian motion  $(B_t)_{0 \leq t \leq T}$  as well as a progressively measurable triple  $(U, V, W)$ , with values in  $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$ , such that  $\mathbb{E} \sup_{t \in [0, T]} (|U_t|^2 + |V_t|^2) < +\infty$ ,  $\mathbb{E} \int_0^T |W_t|^2 dt < +\infty$ , and which satisfies  $\mathbb{P}$  almost surely the couple of equations (E). The distribution of the four-uple  $(B, U, V, W)$  is unique on the space  $\mathcal{C}([0, T], \mathbb{R}^{2d+1}) \times L^2([0, T], \mathbb{R}^d)$ . In other words, the FBSDE (E) admits a unique weak solution. For  $\alpha = 1$ , existence and uniqueness hold in a strong sense.

**2.2. Quasi-Linear PDE.** According to Ladyzhenskaya et al. [15, Ch. 7, Th 7.1] and to [18] (up to a regularization procedure of the coefficients), we claim that (E) admits a solution  $u \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  satisfying:

**Theorem 2.1.** *There exists a constant  $C_{2,1}$ , only depending on  $T$  and on known parameters appearing in (A), such that  $\forall (t, x) \in [0, T] \times \mathbb{R}^d$ ,*

$$\begin{aligned} & |u(t, x)| + |\nabla_x u(t, x)| + |\nabla_{x,x}^2 u(t, x)| + |\partial_t u(t, x)| \\ & + \sup_{t' \in [0, T], t \neq t'} [|t - t'|^{-(1+\alpha)/2} |\nabla u(t, x) - \nabla u(t', x)|] \\ & + \sup_{x \in \mathbb{R}^d, x \neq x'} [|x - x'|^{-\alpha} |\nabla_{x,x}^2 u(t, x) - \nabla_{x,x}^2 u(t, x')|] \leq C_{2,1}. \end{aligned}$$

Moreover,  $u$  is unique in the class of functions  $\tilde{u} \in \mathcal{C}([0, T] \times \mathbb{R}^d, \mathbb{R}) \cap \mathcal{C}^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  for which  $\sup_{(t,x) \in [0, T] \times \mathbb{R}^d} (|\tilde{u}(t, x)| + |\nabla_x \tilde{u}(t, x)|) < +\infty$ .

The connection between (E) and (E) can be summarized as follows

$$(2.1) \quad (V_t, W_t) = (u, v)(t, U_t), \quad V_t = \mathbb{E}[V_T | \mathcal{F}_t] + \mathbb{E} \left[ \int_t^T f(U_s, V_s, W_s) ds | \mathcal{F}_t \right].$$

### 3. ALGORITHM AND MAIN RESULTS

Following Delarue and Menozzi [8], we now introduce the basic objects for the discretization procedure of (E) and (E), namely a temporal mesh as well as a family of spatial grids and an optimal quantization for the Gaussian law. In addition to these ingredients, we consider a collection of interpolating functions associated to the underlying spatial grids.

**3.1. Construction of the Interpolated Algorithm.** For clarity reasons, we choose to define the approximated solution on a family of infinite spatial grids. This is not realistic from a purely numerical point of view, anyhow the truncation procedure is highly discussed in Delarue and Menozzi [8]. It induces heavy computations for the error analysis and is totally useless for our original purpose. For this reason, we consider  $\mathcal{C}_\infty \equiv \delta \mathbb{Z}^d$ ,  $\delta > 0$ , the infinite Cartesian grid of step  $\delta$ .

**Shape functions.** The algorithm we propose below is based on a piecewise multilinear approximation procedure, obtained by tensorization of piecewise linear interpolation. The involved  $[0, 1]$ -valued shape functions are the following:

$$(3.1) \quad \forall z \in \mathcal{C}_\infty, \forall x \in \mathbb{R}^d, \quad \phi_z(x) = \prod_{i=1}^d \Phi(\delta^{-1}(x_i - z_i)),$$

with  $\Phi(t) = (1 - \text{sgn}(t)t)_+$ ,  $\text{sgn}(t) = \mathbf{I}_{t>0} - \mathbf{I}_{t\leq 0}$ . Obviously, for  $z \in \mathcal{C}_\infty$ ,  $\phi_z$  is nonnegative, matches one in  $z$  and vanishes outside the hypercube of center  $z$  and of radius  $\delta$ . It is plain to see that such a family interpolates exactly polynomials of order less than one:

$$(3.2) \quad \forall x \in \mathbb{R}^d, \quad \sum_{z \in \mathcal{C}_\infty} \phi_z(x) = 1, \quad \sum_{z \in \mathcal{C}_\infty} \phi_z(x) z = x.$$

We refer the reader to the literature devoted to finite elements (see e.g. Brenner and Scott [4]) for more general examples of shape functions. Anyhow, due to the stochastic interpretation of the algorithm we provide below, we are to view the underlying family of shape functions in terms of probability weights. Hence, the method is valid only for nonnegative shape functions with sum equal to 1. This prevents us to introduce Lagrangian kernels of order greater than two since they may take negative values (see the monograph of Milstein and Tretyakov [23, p. 425] for similar remarks). In the sequel, we denote for a given function  $\psi : \mathcal{C}_\infty \rightarrow \mathbb{R}$

$$(3.3) \quad \forall x \in \mathbb{R}^d, \quad \psi_\delta(x) = \sum_{z \in \mathcal{C}_\infty} \phi_z(x) \psi(z),$$

its interpolation associated to the sequence  $(\phi_z)_{z \in \mathcal{C}_\infty}$ . It comes for  $z \in \mathcal{C}_\infty$ ,  $\psi(z) = \psi_\delta(z)$ .

**Time Mesh.** Finally, let us introduce a uniform time mesh of  $[0, T]$  with time step  $h > 0$ ,  $h = T/N$ ,  $N \in \mathbb{N}^*$ , i.e.  $\{(t_i \equiv ih)_{i \in [0, N]}\}$ .

**3.2. Algorithm.** In the spirit of Delarue and Menozzi [8], we define

**Algorithm 3.1.**

$$\begin{aligned} \forall x \in \mathbb{R}^d, \quad \bar{u}(T, x) &\equiv H(x), \quad \bar{v}(T, x) \equiv \nabla_x H(x) \sigma(x, H(x)), \\ \forall k \in \llbracket 0, N-1 \rrbracket, \quad \forall x \in \mathcal{C}_\infty, \\ \mathcal{T}(t_k, x) &\equiv b(x, \bar{u}(t_{k+1}, x), \bar{v}(t_{k+1}, x))h + \sigma(x, \bar{u}(t_{k+1}, x))\gamma^* g(\Delta B^k), \\ \bar{v}(t_k, x) &\equiv h^{-1} \mathbb{E}[\bar{u}_\delta(t_{k+1}, x + \mathcal{T}(t_k, x))g(\Delta B^k)^*]\gamma, \\ \bar{u}(t_k, x) &\equiv \mathbb{E}[\bar{u}_\delta(t_{k+1}, x + \mathcal{T}(t_k, x))] + hf(x, \bar{u}(t_{k+1}, x), \bar{v}(t_k, x)). \end{aligned}$$

In the above algorithm,  $\Delta B^k \equiv B_{t_{k+1}} - B_{t_k}$  where  $((B_t)_{t \geq 0}, (\mathcal{F}_t^B)_{t \geq 0})$  denotes a  $d$ -dimensional Brownian motion and its natural augmented filtration on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Also,  $g(\Delta B_k) \equiv h^{1/2} \Pi_M(h^{-1/2} \Delta B^k)$ ,  $\Pi_M$  being the projection mapping onto a square optimal quantization grid  $\Lambda_M^*$ , with  $M$  points, for the  $d$ -dimensional standard Gaussian vector. In other words, integrals with respect to the Gaussian kernel are replaced by discrete sums that turn out to be numerically computable. The only controls concerning the quantized Brownian increment that will be needed in the sequel are the following:

$$(3.4) \quad \mathbb{E}[|g(\Delta B^k) - \Delta B^k|^2] \leq C_{\text{Quantiz}}(d)hM^{-2/d},$$

$$(3.5) \quad \mathbb{E}[\Delta B^k | g(\Delta B^k)] = g(\Delta B^k).$$

Note that the last property simply expresses that the quantized variable is a projector. It also implies  $\mathbb{E}[g(\Delta B^k)] = 0$ . For details about quantization, we refer to the monograph of Graf and Luschgy [11].

The reader used to stochastic literature may wonder why we do not employ a Monte-Carlo strategy. The reason can be explained as follows: replace for a while  $\bar{u}_\delta(t_{k+1}, \cdot)$  by the true solution  $u(t_{k+1}, \cdot)$  in the above induction. Since the latter function belongs to  $\mathcal{C}^{2+\alpha}(\mathbb{R}^d)$ , the quantization procedure then provides a better approximation for the integral with respect to the Gaussian kernel rather than the Monte-Carlo one, that is known to be well fitted to rough frameworks.

Detail finally the meaning of  $\gamma$  in Algorithm 3.1. Denote the covariance matrix of the quantized  $d$ -dimensional standard Gaussian law by  $K_M \equiv \mathbb{E}[\Pi_M(\mathcal{N}(0, I_d)) \times \Pi_M(\mathcal{N}(0, I_d))^*]$ :  $\gamma$  then stands for the lower triangular matrix of the Cholesky writing of  $K_M^{-1}$ , i.e.  $\gamma\gamma^* = K_M^{-1}$  (provided  $\det(K_M) > 0$ ). The introduction of  $\gamma$  in the algorithm follows from the same trick as in the former paragraph: replace for a while  $\bar{u}(t_{k+1}, \cdot)$  and  $\bar{u}_\delta(t_{k+1}, \cdot)$  by  $u(t_{k+1}, \cdot)$  both in the local transition  $\mathcal{T}(t_k, x)$  and in the definition of  $\bar{v}(t_k, x)$  and focus on the resulting martingale increment  $h^{-1} \mathbb{E}[u(t_{k+1}, x + \mathcal{T}(t_k, x))g(\Delta B^k)^*]\gamma$ . Since the true solution is smooth, the Taylor expansion yields as first approximation:  $h^{-1} \mathbb{E}[u(t_{k+1}, x + \mathcal{T}(t_k, x))g(\Delta B^k)^*]\gamma \sim \nabla_x u(t_{k+1}, x) \sigma(x, u(t_{k+1}, x))\gamma^* K_M \gamma = v(t_{k+1}, x)$ . When corrected by  $\gamma$ , the martingale increment associated to a smooth function is worth, up to negligible terms, the gradient of the underlying function multiplied by the diffusion coefficient of the transition. The result fails in the non-corrected case, i.e. for  $\gamma = I_d$ .

In the sequel, the transition  $\mathcal{T}(t_k, x)$  writes  $\mathcal{T}(t_k, x) \equiv \beta(t_k, x)h + \Sigma(t_k, x)\gamma^* \Delta B^k$ .

### 3.3. Main Results.

**Theorem 3.2.** *Assume that  $b$  does not depend on  $v$ . Then, there exist two constants  $c_{3.2} > 0$  and  $C_{3.2}$ , only depending on  $T$  and on known parameters appearing*

in  $(\mathbf{A})$ , such that for  $h < c_{3.2}$ :

$$\begin{aligned} \sup_{x \in \mathbb{R}^d, k \in \llbracket 0, N \rrbracket} |u(t_k, x) - \bar{u}_\delta(t_k, x)|^2 &\leq C_{3.2} [\mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz})] \\ &\equiv C_{3.2} \mathcal{E}^2(\text{global}), \end{aligned}$$

with  $\mathcal{E}(\text{time}) \equiv h^{\alpha/2}$ ,  $\mathcal{E}(\text{space}) \equiv h^{-1}\delta^2$ ,  $\mathcal{E}(\text{quantiz}) \equiv h^{\alpha/2}M^{-2/d}$ .

**Classification of the Errors.**

*Time Error.* This term results from the combination of the  $1/2$ -Hölder continuity of the Brownian motion in the  $L^2$  sense and of the  $\alpha$ -Hölder  $x$ -continuity of the coefficients of the PDE.

Formally, for  $\alpha = 2$ , we recover the rate announced in the papers of Milstein and Tretyakov.

*Spatial error.* The distance between a smooth function (think to the true solution) and its piecewise linear interpolated version at the nodes of the grid is worth  $\delta^2$ . Due to the propagation of the error along the time mesh, the resulting spatial error matches  $h^{-1}\delta^2$ . Compared to the rough projection in [8], *i.e.* piecewise constant interpolation, we gain one order w.r.t.  $\delta$ . This is the expected improvement with such a procedure and the new bound is in this sense satisfactory.

The reader may wonder why we restrict to an interpolation of order one since  $u$  is  $C^{2+\alpha}$  in space. The reason follows once again from the probabilistic nature of the algorithm that prevents us, at least in terms of convergence analysis, from using negative weights deriving from higher order interpolators.

The structure of the spatial error term induces a non-standard CFL condition:  $\delta$  has to be small against  $h$ . Indeed, the spatial grid has to be fine enough to catch the increments of the Brownian motion.

*Quantization error.* For a bounded smooth function  $F \in C_b^2(\mathbb{R}^d, \mathbb{R})$  (*i.e.* with bounded derivatives of order one and two) and an optimally quantized Gaussian kernel, it comes from (3.4) :  $\mathbb{E}[F(\Delta B)] - \mathbb{E}[F(g(\Delta B))] = O(hM^{-2/d})$ . The term  $M^{-2/d}$  is obtained summing along the mesh. The  $h^{\alpha/2}$  corresponds to the Hölder regularity of  $H_u$  and appears through some rather sharp controls during the error analysis.

**Influence of the Quantization.** In comparison with Delarue and Menozzi [8], this new bound for  $\mathcal{E}(\text{quantiz})$  is rather spectacular. The reason is the following: we take advantage of the optimality of the quantization, see (3.5). On the opposite, the quantization in [8] is not assumed to be optimal since the analysis relies on  $L^p$  bounds for the underlying Gaussian quantization, that are known to fail in the optimal setting (see the recent paper of Graf, Luschgy and Pagès [10]).

The crucial point, in our new frame, is the following: the quantization error is less than the temporal one. In particular, there does not seem to be any interest to choose a large support for the quantized Gaussian kernel: it is sufficient to choose one of the roughest quantization grid satisfying  $\det(K_M) > 0$ . Actually, this phenomenon is not so surprising: in the one-dimensional smooth case investigated by Milstein and Tretyakov [23], that is  $\alpha = 2$  in a formal way, the global error is worth  $h$  provided  $\delta = h$  and  $M = 2$  (in fact, the latter authors do not refer to the quantization theory, but directly approximate the Gaussian law by a centered Bernoulli one).



Paradoxically, we feel that a large quantization may be useful in the nonsmooth setting. Indeed, for  $\alpha$  close to zero, the order of the expected global error cannot exceed  $\alpha/2$ , due to the temporal error  $\mathcal{E}(\text{time})$ , and is thus very low. On the opposite, the terms  $\mathcal{E}(\text{space})$  and  $\mathcal{E}(\text{quantiz})$  can be chosen of order  $1/2$  provided  $\delta = h^{3/4}$  and  $M = h^{-d/4}$  (such values are still reasonable in the three-dimensional setting). The question is then the following: how to diminish  $\mathcal{E}(\text{time})$  ?

As explained above, the bound for  $\mathcal{E}(\text{time})$  follows from the smoothness of the coefficients. To understand exactly what happens, focus on a simple semi-linear case: assume that  $b$  vanishes and that  $\sigma$  reduces to identity. Generally speaking, the underlying strategy of Algorithm 3.1 then consists in approximating  $\mathbb{E} \int_{t_k}^{t_{k+1}} f(B_s, u(s, B_s), v(s, B_s)) ds$ ,  $k \in \llbracket 0, N \rrbracket$ , by  $h \mathbb{E} f(B_{t_k}, \bar{u}_\delta(t_k, B_{t_k}), \bar{v}_\delta(t_k, B_{t_k}))$ . If  $f$  is just Hölder continuous in  $x$ , with respect to a small Hölder exponent, we cannot expect to recover less than  $h^{\alpha/2}$  for the temporal error. However, we could approximate  $\mathbb{E} \int_{t_k}^{t_{k+1}} f(B_s, u(s, B_s), v(s, B_s)) ds$  by  $\mathbb{E} \int_{t_k}^{t_{k+1}} f(B_s, \bar{u}_\delta(t_k, B_{t_k}), \bar{v}_\delta(t_k, B_{t_k})) ds$  with a Monte-Carlo method or a quantization procedure depending on the exact value of  $\alpha$  and the affordable complexity. Even if numerically heavy, the modification of Algorithm 3.1 based on this Monte-Carlo method would provide a convergent scheme in the limit case  $\alpha = 0$ .

Another phenomenon may occur for small values of  $\alpha$ . The coefficients and the solution may count isolated singularities (that is isolated points at which  $\alpha$  is actually tiny) and have, elsewhere, large “pockets” of smoothness. Such a case is very difficult to investigate from a theoretical point of view. Anyhow, we observe in practice (see Section 4 below) that quantization may be, in some cases, more efficient for large values of  $M$ . A possible explanation is the following: there may be a competition between the bounds for the global and local errors. In other words, the error is, away from the singularities, of order one, and, around them, of order  $\alpha/2$ , and the combination of both is sensible to the value of  $M$ .

**Typical Values.** In dimensions two and three (that is in cases considered in Section 4), numerical computations show that  $\det(K_M) > 0$  for  $M \geq d + 1$ . We believe the result to be true for higher values of  $d$ , but the proof remains open. For this choice of  $M$  and for  $\delta = h^{1/2+\alpha/4}$ , the error is at most of order  $\alpha/2$ .

**What About a Gradient Dependence in the Drift ?** Compared to [8], there is no  $\mathcal{E}(\text{gradient})$  term in the writing of  $\mathcal{E}(\text{global})$ . The reason is simple: we just focus in the current setting on the case  $b(x, u)$ , whereas  $\mathcal{E}(\text{gradient})$  appears when considering the more general case  $b(x, u, v)$ .

A first strategy to handle the case  $b(x, u, v)$  is the following: the drift term can always be seen as a part of the second member. Of course, this leads to assume  $f$  to be quadratic in  $\nabla_x u$ . From a theoretical point of view, quadratic (F)BSDEs are investigated with a suitable exponential transform that is highly discussed in Kobylanski [14]. The possible adaptation of this strategy to the discretization procedure is formally open, but turns out to be totally out of interest from a numerical point of view.

The KPZ equation, *i.e.* the heat equation driven by a non-linear term of the form  $|\nabla_x u|^2$ , frequently appears in mathematical finance (think to utility maximization, see e.g. Hu et al. [12]) and provides a very interesting numerical example. The non-linearity  $|\nabla_x u|^2$  can be both interpreted as a quadratic second member, so that the drift reduces to zero, or as a first order term with a non-linear drift given by the

gradient itself, so that the second member vanishes. We show in Section 4 that the numerical counterpart of the first writing is totally unstable and may even explode. On the opposite, the second point of view yields a good approximation of the true solution.

This is the reason why the case  $b(x, u, v)$  is so important to investigate. As easily guessed by the reader, we are not able to prove the convergence of Algorithm 3.1 in this larger setting. A possible solution is discussed at the end of the paper: it consists in introducing an intermediate predictor for the gradient in the drift of the approximate transition  $\mathcal{T}(t_k, x)$ , but induces a new error term, denoted in Delarue and Menozzi [8] by  $\mathcal{E}(\text{gradient})$ . The new bound for the global error remains the same, up to a new constant, provided  $\delta^2 \leq h$  and  $\delta^{-3}h^{3/2}M^{-2/d} \leq 1$ . Of course, this condition is not satisfactory since  $\delta^{-3}h^{3/2}M^{-2/d}$  is worth  $h^{-3\alpha/4}/4$  for the above typical values. To recover the order  $\alpha/2$ ,  $M$  has to be chosen equal to  $h^{-3d\alpha/8}$ . This is the best we can do so far.

**What about a Truncation ?** Similarly to what has been done in [8], one could truncate the grid and obtain, for every  $q \geq 1$ , a truncation error  $\mathcal{E}(\text{trunc}) \leq C_q(R/(R + \rho))^q$  where  $R > 0$  is the radius of the initial grid  $\mathcal{C}_0$  and  $\rho > 0$  a truncation parameter. The constant  $C_q$  blows up when  $q$  increases.

#### 4. NUMERICAL EXAMPLES

In this section, we choose to illustrate various behaviors of the algorithm through the approximation of the multidimensional Burgers equations and the deterministic KPZ equation.

**4.1. Multidimensional Burgers equations.** The Burgers equations are a simplified form of the Navier-Stokes equations. The convective and dissipative parts are the same but the pressure term as well as the incompressibility constraint are neglected. The equations write

$$(4.1) \quad \begin{aligned} \partial_t u - (u \cdot \nabla_x)u + \frac{\varepsilon^2}{2} \Delta u &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}^d, \quad \varepsilon > 0, \\ u(T, x) &= H(x), \quad x \in \mathbb{R}^d, \end{aligned}$$

where  $\forall i \in \llbracket 1, d \rrbracket$ ,  $((u \cdot \nabla_x)u)_i = \nabla_x u_i \times u$ . Even though the convergence results are stated for real valued functions, the same analysis could be carried out for systems of equations. Thus, the controls of Section 3.3 are still valid for the solution of (4.1).

In dimension one, it is well known that the equation (4.1) has an explicit solution obtained through a Cole-Hopf factorization, see e.g. [26]. In the multidimensional setting, the factorization can be done provided the final condition  $H$  derives from a potential, namely  $H = \nabla H_0$  where  $H_0$  is a real-valued function. In this case, the solution explicitly writes:  $\forall (t, x) \in [0, T] \times \mathbb{R}^d$ ,

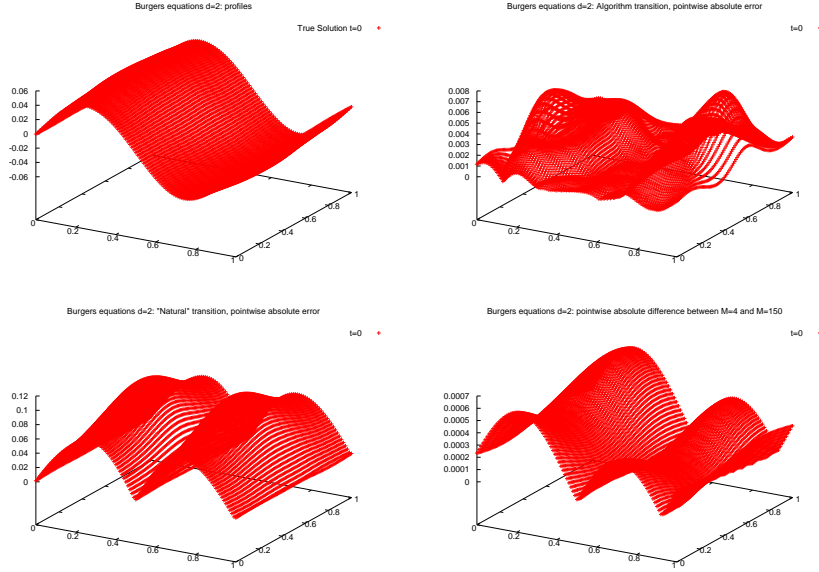
$$u(t, x) = \frac{\mathbb{E}[\nabla H_0(x + \varepsilon B_{T-t}) \exp(-\varepsilon^{-2} H_0(x + \varepsilon B_{T-t}))]}{\mathbb{E}[\exp(-\varepsilon^{-2} H_0(x + \varepsilon B_{T-t}))]}.$$

We always consider the coupled interpretation of the Burgers equations, *i.e.*  $b(x, u) = u$  and  $f = 0$ . This choice turns out to be numerically more robust, see Section 4.2 and also [8], Section 5. In the following, we take  $d = 2$ .

Discuss now the influence of the viscosity parameter. The approximated transitions involved in Algorithm 3.1 are close to  $\varepsilon h^{1/2}$ . To catch them, the spatial grid

has to be fine enough and the spatial step  $\delta$  has to be, at least, less than  $\varepsilon h^{1/2}$ . This empirical condition is confirmed from a numerical point of view. We thus choose the following values for the parameters at hand:  $T = 3/8, h = 2.5 \times 10^{-2}, \delta = .01, \varepsilon^2 = .4$ .

To avoid truncation problems, we choose the periodic initial solution  $H_0(x) = \prod_{i=1,2} \sin^2(\pi x_i)$ . Since the problem is then symmetric, we only present the results obtained for the approximation of the first component of  $u = (u_1, u_2)^*$ . We first plot the profiles of the solution  $u_1$  at  $t = 0$  and the pointwise absolute error between the reference value and the approximation deriving from Algorithm 3.1, both with and without  $\gamma^*$ . We also investigate the influence of the number  $M$  of points used for the quantization (we first choose  $M = 4$  and then  $M = 150$ ).



The reference value is obtained through the explicit representation (4.1) via quantization with 600 points: due to the “large” viscosity, we observe on the left top figure how fast the solution decays.

The left bottom figure suggests, in comparison with the right top one, how the corrector matrix  $\gamma^*$  in the transition is crucial, especially when the number of points in the quantization is small: with  $M = 4$ , the relative error is close to 2 for  $\gamma = I_d$ , but close to .1 for the suitable  $\gamma$  (in this latter case, the relative error is still “large” due to the tiny values for the true solution). Moreover, the last picture confirms that increasing  $M$  does not improve the error.

Let us turn to the case of a smaller viscosity, namely  $\varepsilon^2 = .08$ . For the previous example, with  $M = 4, T = .5$ , let  $\delta$  vary as  $\varepsilon h$ . The reference solution has globally the same shape than in the previous picture and is  $[-.4, .4]$  valued. It comes

$h$	$5 \times 10^{-2}$	$2.5 \times 10^{-2}$	$1.25 \times 10^{-2}$	$6.25 \times 10^{-3}$
$\sup_{t_i, x_j}  (\bar{u}_1 - u_1)(t_i, x_j) $	.121788	.062766	.043893	.035647
$\sup_{t_i, x_j}  (\bar{u}_2 - u_2)(t_i, x_j) $	.12147	.062397	.043695	.035518

These numerical results confirm that in the “smooth” case the error is of order one w.r.t.  $h$ .

**4.2. KPZ Equation.** The KPZ equation can be seen as the primitive of the former Burgers equations. It writes:

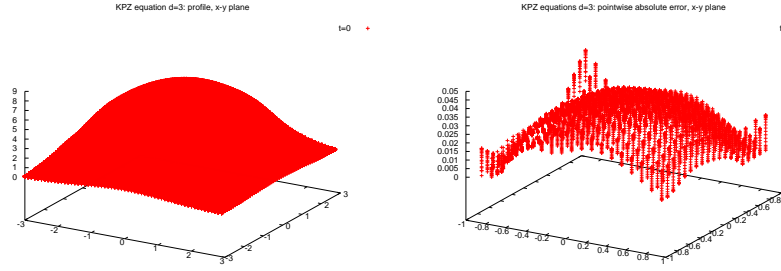
$$(4.2) \quad \begin{aligned} \partial_t u + \frac{1}{2} |\nabla_x u|^2 + \frac{1}{2} \Delta u &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}^d, \\ u(T, x) &= H(x), \quad x \in \mathbb{R}^d. \end{aligned}$$

Equation (4.2) also has an explicit solution obtained through a Cole-Hopf factorization

$$u(t, x) = \log(\mathbb{E}[\exp(H(x + B_{T-t}))]).$$

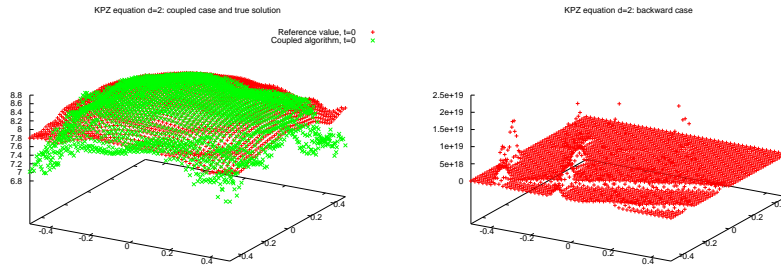
We first concentrate on the 3-dimensional case taking  $H(x) = 10 \exp(-(1/8)(3x_1^2 + 2x_2^2 + x_3^2))$ ,  $T = .5$ ,  $h = .02$ ,  $\delta = .025$ ,  $M = 10$ . We truncate the grid on  $[-3, 3]$ . In the decoupled case, i.e. the underlying process is the Brownian motion, exponential bounds are available for the truncation error through Bernstein like arguments. In the coupled case, since we do not prove the boundedness of the approximated gradient we only have polynomial controls, see [8]. Anyhow, the fast decay of the terminal condition makes the associated error “numerically” reasonable.

The reference value was computed with  $M = 200$ . We plot the difference between the reference value and the coupled algorithm.



In this smooth case, we still observe an error of order one w.r.t.  $h$ .

Let us now turn to an example for which the approximated solution obtained with the decoupled algorithm explodes. Take  $d = 2$ ,  $H(x) = 10 \cos(5|x|^2)$ ,  $T = .1$ ,  $h = .02$ ,  $\delta = .02$ ,  $M = 4$  and truncate the grid for  $|x| \geq 1$ . The reference value is still computed by quantization with  $M = 200$ . One gets



Basically, the numerical integration of a large gradient induces overflows. Hence, the coupled interpretation has to be preferred to the backward one that is quite unstable.

**4.3. A Toy Example in the Low Regular Setting.** To illustrate the behavior of the algorithm for small values of  $\alpha$ , we investigate a linear example. Of course, our algorithm is totally useless in this frame, but reference values can be computed with a Monte-Carlo procedure.

We thus focus on the following one-dimensional example:

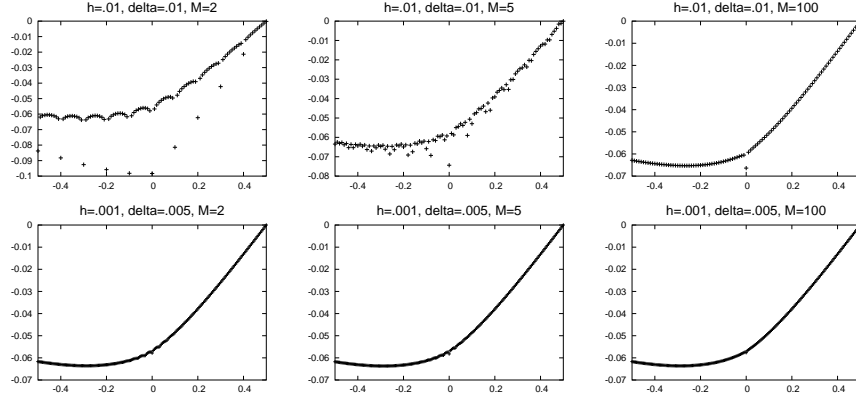
$$\partial_t u(t, x) + \frac{1}{2} \partial_{x,x}^2 u(t, x) + |x|^{1/16} - |x-1|^{1/16} = 0, \quad (t, x) \in [0, 1) \times \mathbb{R},$$

with the null boundary condition at time  $T = 1$ . The second member counts two singularities in zero and one so that  $u$  belongs to  $C^{1+\alpha/2, 2+\alpha}([0, T] \times \mathbb{R})$ ,  $\alpha = 1/16$ . On the opposite, the solution  $u$  has large “pockets” of regularity away from the singular points.

In light of Theorem 3.2, we thus expect the algorithm to behave poorly around the points zero and one and to be closer to the true solution away from them.

Three reference values are computed with  $10^6$  Monte-Carlo simulations. In each case, the empirical standard deviation is less than  $10^{-1}$  so that the underlying error is at most of order  $10^{-4}$ . Here are the results:  $u(-.4, 0) \sim -.062$ ,  $u(-.1, 0) \sim -.060$ ,  $u(0, 0) \sim -.055$ .

We plot below the outcomes of our algorithm on  $[-.5, .5]$  for the following choices: on the first row,  $h = .01$ ,  $\delta = .01$  and  $M = 2, 5, 100$  and on the second row,  $h = .001$ ,  $\delta = .005$  and  $M = 2, 5, 100$ . The Cartesian grid is truncated at the level  $|x| = 5$ .



Comment first what happens in the case  $h = .01$  and  $\delta = .01$ . Each graph exhibits jumps, especially around the singular point zero. They decrease with the size  $M$  of the underlying quantization. Moreover, the mean trends of the curves, that is without taking into account the jumps, are rather close between them, but do not fit exactly the reference values computed with the Monte-Carlo method. In fact, the adequation is less and less satisfactory as getting closer to the singular point zero.

Here is a possible explanation for all these observations: the algorithm is sensitive to the underlying local regularity of the coefficients. On the one hand, each isolated singularity is propagated on a small neighborhood and induces there several local jumps. On the other hand, the local error diminishes on “pockets” of smoothness. In the end, there is a competition between the two phenomena and the combination of both varies with the number  $M$  of quantization points.

Turn now to the second case  $h = .001$  and  $\delta = .005$ . Generally speaking, the results are more satisfactory as expected in light of Theorem 3.2. In particular,

since  $h$ , that induces the dominant term in the global error, is now very small, there are no jumps anymore and quantization doesn't seem to be so influent as in the former case. Anyhow, the fitting is still better away from the singular points due to the local smoothness of the coefficients and of the true solution.

## 5. PROOF, FIRST STEP. DISCRETE FBSDEs

The proof of Theorem 3.2 relies on a discrete version of the FBSDE (E), but we need first to introduce some additional material concerning the interpolation procedure.

**About Constants.** In the following, we keep the same notation  $C, C_\eta, c_\eta$  (or  $C', C'_\eta, c'_\eta$ ) for all finite, non-negative constants which appear in our computations: they may depend on known parameters in **(A)**, on  $T$ , but not on any of the discretization parameters. The index  $\eta$  in the previous notation refers to the numbering of the Proposition, Lemma, Theorem, ... where the constant appears.

**Conditions on Parameters.** The statements of the following Propositions and Lemmas hold for  $h$  small enough.

**5.1. Projection Mappings.** For a given point  $x \in \mathbb{R}^d$ , we want to individuate the cell it belongs to in order to determine the functions  $(\phi_z)_{z \in \mathcal{C}_\infty}$  involved in the interpolation procedure. It is rather obvious that at most  $2^d$  of them are concerned.

We first define the so-called projection to the lowest neighbor:  $\forall x \in \mathbb{R}^d$ ,  $\Pi^{(0)}(x) = (\delta \lfloor \delta^{-1} x_1 \rfloor, \dots, \delta \lfloor \delta^{-1} x_d \rfloor)$ . Following an arbitrary numbering, we denote by  $\Pi^{(i)}(x)$ ,  $i \in \llbracket 1, 2^d \rrbracket$  the remaining projection mappings to the vertices of the cell  $x$  belongs to.

For  $d = 2$ ,  $\Pi^{(0)}, \dots, \Pi^{(3)}$  can be represented as in figure 1 below. With these

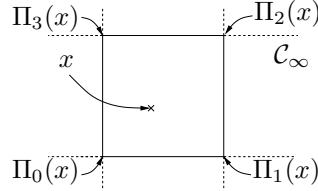


FIGURE 1. Projection Mappings,  $d = 2$

notations, for a function  $\psi : \mathcal{C}_\infty \rightarrow \mathbb{R}$  one also has

$$(5.1) \quad \forall x \in \mathbb{R}^d, \quad \psi_\delta(x) = \sum_{i=0}^{2^d-1} \phi_{\Pi^{(i)}(x)}(x) \psi(\Pi^{(i)}(x)).$$

As explained in Subsection 3.1, to analyze the convergence of Algorithm 3.1, we take advantage of the probabilistic interpretation of the piecewise linear interpolation. Indeed, for  $x \in \mathbb{R}^d$ , the family of nonnegative weights  $(\phi_{\Pi^{(i)}(x)}(x))_{i \in \llbracket 0, 2^d \rrbracket}$  defines a probability measure on the finite set  $\llbracket 0, 2^d \rrbracket$ . We make use in the sequel of the associated cumulative distribution function

$$\forall x \in \mathbb{R}^d, \quad \alpha^{(0)}(x) = 0, \forall i \in \llbracket 1, 2^d \rrbracket, \quad \alpha^{(i)}(x) = \sum_{j=0}^{i-1} \phi_{\Pi^{(j)}(x)}(x).$$

**5.2. Discrete Version of  $(U, V, W)$ .** The key to prove Theorem 3.2 is to associate to Algorithm 3.1 a discrete version of the FBSDE (E). Assume for example that  $(X, Y, Z)$  denotes a possible candidate to mimic  $(U, V, W)$  but along Algorithm 3.1. In order to have a discrete Feynman-Kac formula, we need to define the forward process  $(X_{t_k})_{k \in \llbracket 0, N \rrbracket}$  as a Markov chain with continuous state space. Since the transitions in the algorithm are defined on the grid, we need to specify the transition for  $(X_{t_k})_{k \in \llbracket 0, N \rrbracket}$  according to its spatial position at the current discretization time. The most natural way to proceed consists in randomly choosing, at time  $t_k$ , one of the  $2^d$  possible transitions associated to the cell  $X_{t_k}$  belongs to. This is done by simulating, independently of the Brownian motion, a discrete random variable with weights  $(\phi_k^{(i)} \equiv \phi_{\Pi^{(i)}(X_{t_k})}(X_{t_k}))_{i \in \llbracket 0, 2^d \rrbracket}$ .

**Extension of the Probability Space.** Let  $(U_\ell)_{\ell \in \mathbb{N}^*}$  be a sequence of independent identically distributed random variables, independent of the  $\sigma$ -field  $\bigvee_{t \geq 0} \mathcal{F}_t^B$ , such that  $U_1 \sim \mathcal{U}([0, 1])$ . We set, for  $t \geq 0$ ,  $\mathcal{F}_t \equiv \mathcal{F}_t^B \vee \mathcal{F}_t^U$ , with  $\mathcal{F}_t^U \equiv \sigma(U_1, \dots, U_{k+1})$  with  $k$  such that  $t_k < t \leq t_{k+1}$  ( $\mathcal{F}_0^U \equiv \{\emptyset, \Omega\}$ ). In particular, the  $\sigma$ -field  $\mathcal{F}_{t+}^U$  differs from  $\mathcal{F}_t^U$  for  $t \in \{(t_k)_{k \in \llbracket 0, N \rrbracket}\}$ . For simplicity, we set  $\forall k \in \llbracket 0, N \rrbracket$ ,  $\mathbb{E}_k[\cdot] \equiv \mathbb{E}[\cdot | \mathcal{F}_{t_k}]$ .

The following lemma (whose proof is left to the reader) provides the connection between the variables  $(U_\ell)_{\ell \in \mathbb{N}^*}$  and the projection mappings.

**Lemma 5.1.** *Conditionally to the  $\sigma$ -field  $\mathcal{F}_{t_k}$ ,  $k \in \llbracket 0, N \rrbracket$ ,  $U_{k+1}$  and  $\sigma(B_s - B_{t_k}, t_k \leq s \leq t_{k+1})$  are independent and for every  $\mathbb{R}^d$ -valued and  $\mathcal{F}_{t_k}$ -measurable random variable  $\xi$ :*

$$\forall i \in \llbracket 0, 2^d \rrbracket, \mathbb{E}_k[\mathbf{I}_{U_{k+1} \in [\alpha^{(i)}(\xi), \alpha^{(i+1)}(\xi)]}] = \phi_{\Pi^{(i)}(\xi)}(\xi).$$

**Discrete Representation Processes.** Algorithm 3.1 and Lemma 5.1 motivate, for an initial condition  $x_0 \in \mathbb{R}^d$ , the following definition for the approximating triple  $(X, Y, Z)$ . Set  $X_0 \equiv x_0$  and

$$\begin{aligned} (5.2) \quad X_{t_{k+1}} &\equiv \sum_{i=0}^{2^d-1} \mathbf{I}_{U_{k+1} \in [\alpha^{(i)}(X_{t_k}), \alpha^{(i+1)}(X_{t_k})]} (\Pi^{(i)}(X_{t_k}) + \mathcal{T}(t_k, \Pi^{(i)}(X_{t_k}))) \\ &\equiv \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} X_{t_{k+1}}^{(i)} \text{ for } k \in \llbracket 0, N \rrbracket, \\ (Y_{t_k}, Z_{t_k}) &\equiv (\bar{u}_\delta(t_k, X_{t_k}), \bar{v}_\delta(t_k, X_{t_k})) \text{ for } k \in \llbracket 0, N \rrbracket. \end{aligned}$$

Note carefully that  $X_{t_{k+1}}^{(i)}$  does not stand for the  $i^{\text{th}}$  component of a vector of  $\mathbb{R}^d$ . It is associated to the initial position  $\Pi^{(i)}(X_{t_k})$  and to the transition  $\mathcal{T}(t_k, \Pi^{(i)}(X_{t_k}))$ . Referring to Algorithm 3.1,  $\mathcal{T}(t_k, \Pi^{(i)}(X_{t_k}))$  writes in an obvious manner  $\mathcal{T}(t_k, \Pi^{(i)}(X_{t_k})) \equiv b_k^{(i)} h + \sigma_k^{(i)} \gamma^* g(\Delta B^k)$ .

**Backward Equation.** From the above definition we derive the

**Proposition 5.2** (Discrete Feynman Kac formula). *For all  $k \in \llbracket 0, N \rrbracket$ ,*

$$\begin{aligned} Y_{t_k} &= \mathbb{E}_k \left[ H(X_{t_N}) \right. \\ &\quad \left. + h \sum_{j=k}^{N-1} \sum_{i=0}^{2^d-1} \phi_j^{(i)} f(\Pi^{(i)}(X_{t_j}), \bar{u}(t_{j+1}, \Pi^{(i)}(X_{t_j})), \bar{v}(t_j, \Pi^{(i)}(X_{t_j}))) \right] \\ &\equiv \mathbb{E}_k \left[ H(X_{t_N}) + h \sum_{j=k}^{N-1} \sum_{i=0}^{2^d-1} \phi_j^{(i)} f_j^{(i)} \right] \equiv \mathbb{E}_k \left[ H(X_{t_N}) + h \sum_{j=k}^{N-1} \bar{f}_j(t_j, X_{t_j}) \right]. \end{aligned}$$

*Proof.* Write first, for  $k \in \llbracket 0, N \rrbracket$  (we specify over the symbols “=” the references employed for the computations)

$$\begin{aligned} \mathbb{E}_k[Y_{t_{k+1}}] &\stackrel{(5.2)}{=} \mathbb{E}_k[\bar{u}_\delta(t_{k+1}, X_{t_{k+1}})] \stackrel{(3.3)}{=} \sum_{z \in \mathcal{C}_\infty} \bar{u}(t_{k+1}, z) \mathbb{E}_k[\phi_z(X_{t_{k+1}})] \\ (5.3) \quad &\stackrel{(5.2), \text{Le. 5.1}}{=} \sum_{z \in \mathcal{C}_\infty} \bar{u}(t_{k+1}, z) \sum_{i=0}^{2^d-1} \mathbb{E}_k[\chi_{k+1}^{(i)}] \mathbb{E}_k[\phi_z(X_{t_{k+1}}^{(i)})] \\ &\stackrel{\text{Le. 5.1}}{=} \sum_{z \in \mathcal{C}_\infty} \bar{u}(t_{k+1}, z) \sum_{i=0}^{2^d-1} \phi_k^{(i)} \mathbb{E}_k[\phi_z(X_{t_{k+1}}^{(i)})]. \end{aligned}$$

From Algorithm 3.1, one also gets

$$\begin{aligned} Y_{t_k} &\stackrel{(5.2)}{=} \bar{u}_\delta(t_k, X_{t_k}) \stackrel{(5.1)}{=} \sum_{i=0}^{2^d-1} \phi_k^{(i)} \bar{u}(t_k, \Pi^{(i)}(X_{t_k})) \\ (5.4) \quad &\stackrel{\text{Al. 3.1}}{=} \sum_{i=0}^{2^d-1} \phi_k^{(i)} \{ \mathbb{E}_k[\bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)})] + h f_k^{(i)} \} \\ &\stackrel{(5.2), \text{Le. 5.1}}{=} \sum_{i=0}^{2^d-1} \phi_k^{(i)} \left\{ \sum_{z \in \mathcal{C}_\infty} \bar{u}(t_{k+1}, z) \mathbb{E}_k[\phi_z(X_{t_{k+1}}^{(i)})] + h f_k^{(i)} \right\}. \end{aligned}$$

Equations (5.3) and (5.4) yield  $\mathbb{E}_k[Y_{t_{k+1}}] + h \sum_{i=0}^{2^d-1} \phi_k^{(i)} f_k^{(i)} = Y_{t_k}$ . The proof of the proposition follows by induction.  $\square$

**5.3. Associated *a priori* Estimates.** From (5.2), we derive  $\forall k \in \llbracket 0, N \rrbracket$

$$Y_{t_{k+1}} = \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} \bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)}).$$

For  $k \in \llbracket 0, N \rrbracket$  and  $i \in \llbracket 0, 2^d \rrbracket$ , the random variable  $\bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)})$  is  $\mathcal{F}_{t_k} \vee \sigma(B_t - B_{t_k}, t_k \leq t \leq t_{k+1})$  measurable. Thanks to the Martingale Representation Theorem (see e.g. Theorem III.4.33 in Jacod and Shiryaev [13]), there exists a process  $(\bar{Z}_t^{(i)})_{t_k \leq t \leq t_{k+1}}$  with values in  $\mathbb{R}^d$ , progressively measurable with respect to the previous filtration and with finite moment of order two, such that:

$$(5.5) \quad \bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)}) = \mathbb{E}_k[\bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)})] + \int_{t_k}^{t_{k+1}} \bar{Z}_s^{(i)} dB_s,$$



so that  $Y_{t_{k+1}} = \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} \left\{ \mathbb{E}_k[\bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)})] + \int_{t_k}^{t_{k+1}} \bar{Z}_s^{(i)} dB_s \right\}$ . Define for  $t \in [t_k, t_{k+1}]$ ,

$$(5.6) \quad \bar{Z}_t = \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} \bar{Z}_t^{(i)}.$$

Rewrite now

$$(5.7) \quad \begin{aligned} Y_{t_{k+1}} &= \sum_{i=0}^{2^d-1} \phi_k^{(i)} \mathbb{E}_k[\bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)})] + \int_{t_k}^{t_{k+1}} \bar{Z}_s dB_s + \Delta R_{k+1}(1) \\ &\stackrel{\text{Le. 5.1}}{=} \mathbb{E}_k[Y_{t_{k+1}}] + \int_{t_k}^{t_{k+1}} \bar{Z}_s dB_s + \Delta R_{k+1}(1), \\ \Delta R_{k+1}(1) &\equiv \sum_{i=0}^{2^d-1} \left( \chi_{k+1}^{(i)} - \phi_k^{(i)} \right) \mathbb{E}_k[\bar{u}_\delta(t_{k+1}, X_{t_{k+1}}^{(i)})]. \end{aligned}$$

Note from Lemma 5.1 that  $\mathbb{E}_k[\Delta R_{k+1}(1)] = 0$ . Use eventually Proposition 5.2 to obtain:

$$(5.8) \quad \begin{aligned} Y_{t_{k+1}} &= Y_{t_k} - h \bar{f}_\delta(t_k, X_{t_k}) + \int_{t_k}^{t_{k+1}} \bar{Z}_s dB_s + \Delta R_{k+1}(1), \\ Y_{t_N} + h \sum_{k=1}^N \bar{f}_\delta(t_{k-1}, X_{t_{k-1}}) &= Y_0 + \int_0^T \bar{Z}_s dB_s + \sum_{k=1}^N \Delta R_k(1). \end{aligned}$$

This allows to apply the BSDE machinery already used in [8] provided the following *a priori* estimates.

**Proposition 5.3.** *For  $k \in \llbracket 1, N \rrbracket$ ,  $hZ_{t_{k-1}} = \mathbb{E}_{k-1}[\int_{t_{k-1}}^{t_k} \bar{Z}_s ds]\gamma$  and for  $i \in \llbracket 0, 2^d \rrbracket$ ,  $h\bar{v}(t_{k-1}, \Pi^{(i)}(X_{t_{k-1}})) = \mathbb{E}_{k-1}[\int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} ds]\gamma$ .*

**Proposition 5.4.**  $\exists C_{5.4} \geq 0$  s.t.  $\sup_{(k,x) \in \llbracket 0, N \rrbracket \times \mathbb{R}^d} |\bar{u}_\delta(t_k, x)|^2 \leq C_{5.4}$ .

**Proposition 5.5.**  $\exists C_{5.5} \geq 0$  s.t.  $\mathbb{E}[\int_0^T |\bar{Z}_s|^2 ds] + h \sum_{k=0}^{N-1} \mathbb{E}[|\bar{v}|_\delta^2(t_k, X_{t_k}) + |Z_{t_k}|^2] + h \sup_{(k,x) \in \llbracket 0, N \rrbracket \times \mathbb{R}^d} |\bar{v}_\delta(t_k, x)|^2 \leq C_{5.5}$ .

*Proof of Proposition 5.3.* From (5.8), write for a given  $k \in \llbracket 0, N \rrbracket$ :

$$\begin{aligned} Y_{t_{k+1}} + h \bar{f}_\delta(t_k, X_{t_k}) &= Y_{t_k} + \int_{t_k}^{t_{k+1}} \bar{Z}_s dB_s + \Delta R_{k+1}(1), \\ \mathbb{E}_k \int_{t_k}^{t_{k+1}} \bar{Z}_s ds &= \mathbb{E}_k[Y_{t_{k+1}}(\Delta B^k)^*] \stackrel{\text{Le. 5.1, (3.5)}}{=} \mathbb{E}_k[Y_{t_{k+1}} g(\Delta B^k)^*] \stackrel{\text{Le. 5.1}}{=} hZ_{t_k} \gamma^{-1}. \end{aligned}$$

Similar arguments and (5.5) yield the second statement of the Proposition.  $\square$

*Proof of Proposition 5.4.* We apply the basic strategy of the BSDE theory using a discrete version of Itô's formula, see Shiryaev [25], Chapter VII, Subsection 9 or Lemma 6.8 in [8]. We get:  $|Y_T|^2 = |Y_0|^2 + 2 \sum_{k=1}^N \langle Y_{t_k} - Y_{t_{k-1}}, Y_{t_{k-1}} \rangle + \sum_{k=1}^N |Y_{t_k} - Y_{t_{k-1}}|^2$ , with  $Y_{t_k} - Y_{t_{k-1}} = -h \bar{f}_\delta(t_{k-1}, X_{t_{k-1}}) + \int_{t_{k-1}}^{t_k} \bar{Z}_s dB_s + \Delta R_k(1)$  (cf. (5.8)).

From (5.7),  $\mathbb{E}_{k-1}[\Delta R_k(1)] = 0$ . Similarly,

$$\begin{aligned}
 \mathbb{E}_{k-1} \left[ \int_{t_{k-1}}^{t_k} \bar{Z}_s dB_s \right] &\stackrel{(5.6)}{=} \sum_{i=0}^{2^d-1} \mathbb{E}_{k-1} \left[ \chi_k^{(i)} \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} dB_s \right], \\
 &\stackrel{\text{Le. 5.1}}{=} \sum_{i=0}^{2^d-1} \mathbb{E}_{k-1} [\chi_k^{(i)}] \mathbb{E}_{k-1} \left[ \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} dB_s \right] = 0.
 \end{aligned}
 \tag{5.9}$$

Hence,

$$\begin{aligned}
 \mathbb{E}|Y_T|^2 &= |Y_0|^2 + 2h \sum_{k=1}^N \mathbb{E} \langle -\bar{f}_\delta(t_{k-1}, X_{t_{k-1}}), Y_{t_{k-1}} \rangle \\
 &\quad + \sum_{k=1}^N \mathbb{E} \left[ \left\{ h \bar{f}_\delta(t_{k-1}, X_{t_{k-1}}) - \int_{t_{k-1}}^{t_k} \bar{Z}_s dB_s - \Delta R_k(1) \right\}^2 \right].
 \end{aligned}
 \tag{5.10}$$

As above, the expectations of the double products involving  $\bar{f}_\delta(t_{k-1}, X_{t_{k-1}})$  and  $\int_{t_{k-1}}^{t_k} \bar{Z}_s dB_s$  on the one hand and  $\bar{f}_\delta(t_{k-1}, X_{t_{k-1}})$  and  $\Delta R_k(1)$  on the other hand vanish. Note finally that

$$\begin{aligned}
 \mathbb{E}_{k-1} \left[ \left( \int_{t_{k-1}}^{t_k} \bar{Z}_s dB_s \right) \Delta R_k(1) \right] \\
 &\stackrel{(5.6)}{=} \sum_{i=0}^{2^d-1} \mathbb{E}_{k-1} \left[ \left( \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} dB_s \right) (\chi_k^{(i)} \Delta R_k(1)) \right] \\
 &\stackrel{\text{Le. 5.1}}{=} \sum_{i=0}^{2^d-1} \mathbb{E}_{k-1} \left[ \left( \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} dB_s \right) \right] \mathbb{E}_{k-1} [\chi_k^{(i)} \Delta R_k(1)] = 0.
 \end{aligned}
 \tag{5.11}$$

Plug (5.9) and (5.11) into (5.10):

$$\begin{aligned}
 \mathbb{E}|Y_T|^2 &= |Y_0|^2 + 2h \sum_{k=1}^N \mathbb{E} \langle -\bar{f}_\delta(t_{k-1}, X_{t_{k-1}}), Y_{t_{k-1}} \rangle \\
 &\quad + h^2 \sum_{k=1}^N \mathbb{E} [\bar{f}_\delta^2(t_{k-1}, X_{t_{k-1}})] + \mathbb{E} \int_0^T |\bar{Z}_s|^2 ds + \sum_{k=1}^N \mathbb{E} [\Delta R_k^2(1)].
 \end{aligned}
 \tag{5.12}$$

From Assumption **(A)** and (3.2), there exists a constant  $C$  such that:

$$\begin{aligned}
 |Y_0|^2 + \mathbb{E} \int_0^T |\bar{Z}_s|^2 ds &\leq \mathbb{E}|Y_T|^2 + Ch \sum_{k=1}^N \mathbb{E} \left[ |Y_{t_{k-1}}| \left( 1 + |\bar{u}_\delta(t_k, \cdot)|_\infty \right. \right. \\
 &\quad \left. \left. + \sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} |\bar{v}(t_{k-1}, \Pi_i(X_{t_{k-1}}))| \right) \right].
 \end{aligned}
 \tag{5.13}$$

From Young's inequality and from Jensen's inequality (applied to interpolation weights), we derive for every  $\eta > 0$ :

$$(5.14) \quad \begin{aligned} |Y_0|^2 + \mathbb{E} \int_0^T |\bar{Z}_s|^2 ds &\leq \mathbb{E}|Y_T|^2 + C\eta^{-1}h \sum_{k=1}^N \mathbb{E}[|Y_{t_{k-1}}|^2] \\ &+ C\eta h \sum_{k=1}^N \mathbb{E} \left[ 1 + |\bar{u}_\delta(t_k, \cdot)|_\infty^2 + \sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} |\bar{v}|^2(t_{k-1}, \Pi_i(X_{t_{k-1}})) \right]. \end{aligned}$$

Set  $Q \equiv h \sum_{k=1}^N \sum_{i=0}^{2^d-1} \mathbb{E} \left[ \phi_{k-1}^{(i)} |\bar{v}|^2(t_{k-1}, \Pi_i(X_{t_{k-1}})) \right]$ . Owing to Proposition 5.3:

$$\begin{aligned} Q &= h^{-1} \sum_{k=1}^N \sum_{i=0}^{2^d-1} \mathbb{E} \left[ \phi_{k-1}^{(i)} \left| \mathbb{E}_{k-1} \left[ \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} ds \right] \gamma \right|^2 \right] \\ &\leq C \sum_{k=1}^N \sum_{i=0}^{2^d-1} \mathbb{E} \left[ \phi_{k-1}^{(i)} \mathbb{E}_{k-1} \left[ \int_{t_{k-1}}^{t_k} |\bar{Z}_s^{(i)}|^2 ds \right] \right]. \end{aligned}$$

Write now,

$$(5.15) \quad Q \stackrel{\text{Le. 5.1}}{\leq} C \sum_{k=1}^N \sum_{i=0}^{2^d-1} \mathbb{E} \left[ \chi_k^{(i)} \int_{t_{k-1}}^{t_k} |\bar{Z}_s^{(i)}|^2 ds \right] \stackrel{(5.6)}{=} C \mathbb{E} \left[ \int_0^T |\bar{Z}_s|^2 ds \right].$$

From (5.15) and (5.14), we derive that for  $\eta$  and  $h$  small enough

$$(5.16) \quad |Y_0|^2 + \frac{1}{2} \mathbb{E} \left[ \int_0^T |\bar{Z}_s|^2 ds \right] \leq \mathbb{E}|Y_T|^2 + C + Ch \sum_{k=1}^N |\bar{u}_\delta(t_k, \cdot)|_\infty^2.$$

Recall that  $|Y_T| \leq |H|_\infty$ . Thus  $|\bar{u}_\delta(0, \cdot)|_\infty^2 \leq C + Ch \sum_{k=1}^N |\bar{u}_\delta(t_k, \cdot)|_\infty^2$ . As usual in BSDE theory, we could establish in a similar way that for every  $j \in \llbracket 1, N \rrbracket$ :  $|\bar{u}_\delta(t_j, \cdot)|_\infty^2 \leq C + Ch \sum_{k=j+1}^N |\bar{u}_\delta(t_k, \cdot)|_\infty^2$ . A discrete version of Gronwall's Lemma yields the result.  $\square$

*Proof of Proposition 5.5.* The  $L^2$ -estimate of  $\bar{Z}$  follows from Proposition 5.4 and (5.16). Then, the  $L^2$ -estimate of  $Z$  follows from the  $L^2$ -estimate of  $\bar{Z}$  and Proposition 5.3. The  $L^2$ -estimate of  $|\bar{v}|_\delta$  then comes from the earlier definition of  $Q$ . Finally, as a consequence of Proposition 5.4 and the definition of  $\bar{v}_\delta$ , see Algorithm 3.1, we deduce the estimate of the *supremum* norm of  $\bar{v}_\delta$ .  $\square$

## 6. PROOF, SECOND STEP: STABILITY PROPERTIES

This section focuses on the second step of the proof of Theorem 3.2. As in [8] we rely on the smoothness properties of the solution given in Theorem 2.1. Our strategy follows the uniqueness part of the *Four Step Scheme* of Ma, Protter and Yong [18]. More precisely, we compare the process  $(Y_{t_k})_{k \in \llbracket 0, N \rrbracket}$  with the process  $(u(t_k, X_{t_k}))_{k \in \llbracket 0, N \rrbracket}$ .

**6.1. Stability Theorem.** Applying the usual FBSDE machinery, we establish in Subsection 6.2:

**Theorem 6.1.** *There exists a constant  $C_{6.1}$  such that:*

$$(6.1) \quad \begin{aligned} & |(\bar{u}_\delta - u)(0, x)|^2 + C_{6.1}^{-1} h \sum_{k=1}^N \mathbb{E}[(|\bar{v} - v|^2)_\delta(t_{k-1}, X_{t_{k-1}})] \\ & \leq C_{6.1} \left[ \mathcal{E}^2(\text{global}) + h \sum_{k=1}^N |(\bar{u}_\delta - u)(t_k, \cdot)|_\infty^2 \right]. \end{aligned}$$

We then derive Theorem 3.2 from Theorem 6.1 and Gronwall's Lemma (up to a modification of the initial condition).

## 6.2. Proof of Theorem 6.1.

*Starting Point: Time Continuous Processes.* For the proof, we need to extend the definition of  $X$  to the whole set  $[0, T]$ . Put for all  $k \in \llbracket 0, N \rrbracket$  and  $t \in (t_k, t_{k+1})$ :

$$(6.2) \quad \begin{aligned} X_t & \equiv \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} [\Pi^{(i)}(X_{t_k}) + b_k^{(i)}(t - t_k) + \sigma_k^{(i)} \gamma^*(B_t - B_{t_k})] \\ & \equiv \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} X_t^{(i)}. \end{aligned}$$

Note that  $X_{t_k+}^{(i)} = \Pi^{(i)}(X_{t_k})$ . Hence, the extended process  $(X_t)_{0 \leq t \leq T}$  is discontinuous at times  $(t_k)_{k \in \llbracket 1, N \rrbracket}$ , both at  $t_k-$  and  $t_k+$  (of course, it is also discontinuous at times  $0+$  and  $T-$ ). At a given time  $t_k$ ,  $k \in \llbracket 1, N \rrbracket$ , the size of the jumps performed by the process depends, on the right, on the spatial projection error, and, on the left, on the quantization error. The first one is bounded by  $\delta$  and is of mean zero. The second one is easily controlled by the distortion, cf. (3.4). More precisely, for all  $k \in \llbracket 0, N \rrbracket$

$$(6.3) \quad X_{t_{k+1}} - X_{t_{k+1}-} = \sum_{i=0}^{2^d-1} \left[ \chi_{k+1}^{(i)} \sigma_k^{(i)} \gamma^* [g(\Delta B^k) - \Delta B^k] \right].$$

Moreover, one easily obtains the following

**Lemma 6.2.** *There exists  $C_{6.2} \geq 0$  s.t. for every  $k \in \llbracket 0, N-1 \rrbracket$ :*

$$\forall i \in \llbracket 0, 2^d \rrbracket, \forall t \in (t_k, t_{k+1}), \mathbb{E}_k[|X_t^{(i)} - X_{t_k+}^{(i)}|^2] \leq C_{6.2} h.$$

Referring to the structure of the PDE  $(\mathcal{E})$ , we set

$$(6.4) \quad \begin{aligned} & \forall t \in [0, T], \bar{V}_t \equiv u(t, X_t), \bar{W}_t \equiv \nabla_x u(t, X_t) \sigma(X_t, \bar{V}_t), \\ & \forall t \notin \{(t_k)_{k \in \llbracket 0, N \rrbracket}\}, i \in \llbracket 0, 2^d \rrbracket, \begin{cases} \bar{V}_t^{(i)} \equiv u(t, X_t^{(i)}), \\ \bar{W}_t^{(i)} \equiv \nabla_x u(t, X_t^{(i)}) \sigma(X_t^{(i)}, \bar{V}_t^{(i)}), \\ \quad \equiv \nabla_x u(t, X_t^{(i)}) \bar{\sigma}_t^{(i)}. \end{cases} \end{aligned}$$

Note moreover that the martingale part of  $(\bar{V}_t)_{0 \leq t \leq T}$  is driven, for  $t \in (t_k, t_{k+1})$ ,  $k \in \llbracket 0, N \rrbracket$ , by:

$$(6.5) \quad W_t \gamma^* \equiv \sum_{i=0}^{2^d-1} [\chi_{k+1}^{(i)} \nabla_x u(t, X_t^{(i)}) \sigma_k^{(i)} \gamma^*] \equiv \sum_{i=0}^{2^d-1} [\chi_{k+1}^{(i)} W_t^{(i)} \gamma^*].$$

From Theorem 2.1 and Lemma 6.2, we derive the following *a priori estimates* for  $\bar{V}^{(i)}, \bar{W}^{(i)}$ ,  $i \in \llbracket 0, 2^d \rrbracket$ . For all  $k \in \llbracket 0, N \rrbracket$  and  $s \in (t_k, t_{k+1})$ ,

$$(6.6) \quad \mathbb{E}_k [|\bar{V}_s^{(i)} - \bar{V}_{t_k+}^{(i)}|^2 + |\bar{W}_s^{(i)} - \bar{W}_{t_k+}^{(i)}|^2] \leq Ch.$$

*Step One: Itô's formula for  $\bar{V}$ .* Applying Itô's formula to  $(u(t, X_t^{(i)}))_{t_k < t < t_{k+1}}$ , for all  $i \in \llbracket 0, 2^d \rrbracket$  and  $k \in \llbracket 0, N \rrbracket$ , and using the PDE  $(\mathcal{E})$ , it comes :

$$\begin{aligned} \bar{V}_{t_{k+1}} - \bar{V}_{t_k} &= \bar{V}_{t_{k+1}} - \bar{V}_{t_{k+1}-} + \bar{V}_{t_k+} - \bar{V}_{t_k} \\ &+ \sum_{i=0}^{2^d-1} \left\{ \chi_{k+1}^{(i)} \left[ \int_{t_k}^{t_{k+1}} [F(s, X_s^{(i)}, X_{t_k+}^{(i)}, \bar{u}(t_{i+1}, X_{t_k+}^{(i)}), \sigma_k^{(i)} \gamma^* \gamma (\sigma_k^{(i)})^*) \right. \right. \\ &\quad \left. \left. - F(s, X_s^{(i)}, X_s^{(i)}, \bar{V}_s^{(i)}, \bar{\sigma}_s^{(i)} (\bar{\sigma}_s^{(i)})^*)] ds \right. \right. \\ &\quad \left. \left. - \int_{t_k}^{t_{k+1}} f(X_s^{(i)}, \bar{V}_s^{(i)}, \bar{W}_s^{(i)}) ds \right] \right\} + \int_{t_k}^{t_{k+1}} W_s \gamma^* dB_s, \end{aligned}$$

with  $F(s, x, x', y, A) = (1/2) \text{tr}(AH_u(s, x)) + \nabla_x u(s, x)b(x', y)$ ,  $s \in [0, T]$ ,  $x, x' \in \mathbb{R}^d$ ,  $y \in \mathbb{R}$  and  $A \in \mathbb{R}^{d \times d}$ .

*Step Two: Difference of the Processes.* The strategy is well-known: make the difference between  $\bar{V}$  and  $Y$  and apply the usual BSDE machinery to estimate the difference. From Proposition 5.2 and (5.8), we claim:

$$\begin{aligned} (6.7) \quad & \bar{V}_{t_{k+1}} - Y_{t_{k+1}} - [\bar{V}_{t_k} - Y_{t_k}] \\ &= \{\bar{V}_{t_{k+1}} - \bar{V}_{t_{k+1}-} + \Delta R_{k+1}(3)\} + \{\bar{V}_{t_k+} - \bar{V}_{t_k}\} \\ &+ \left\{ \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} \left[ \int_{t_k}^{t_{k+1}} [F(s, X_s^{(i)}, X_{t_k+}^{(i)}, \bar{u}(t_{i+1}, X_{t_k+}^{(i)}), \sigma_k^{(i)} (\sigma_k^{(i)})^*) \right. \right. \\ &\quad \left. \left. - F(s, X_s^{(i)}, X_s^{(i)}, \bar{V}_s^{(i)}, \bar{\sigma}_s^{(i)} (\bar{\sigma}_s^{(i)})^*)] \right. \right. \\ &\quad \left. \left. - f(X_s^{(i)}, \bar{V}_s^{(i)}, \bar{W}_s^{(i)}) + f_k^{(i)}] ds \right] \right\} \\ &+ \left\{ \int_{t_k}^{t_{k+1}} [W_s \gamma^* - \bar{Z}_s] dB_s + \Delta R_{k+1}(2) \right\} \\ &\equiv \Delta E_{k+1}(1) + \Delta E_{k+1}(2) + \Delta E_{k+1}(3) + \Delta E_{k+1}(4), \end{aligned}$$

denoting

$$\begin{aligned} (6.8) \quad \Delta R_{k+1}(2) &\equiv h \sum_{i=0}^{2^d-1} [(\phi_k^{(i)} - \chi_{k+1}^{(i)}) f_k^{(i)}] - \Delta R_{k+1}(1), \\ \Delta R_{k+1}(3) &\equiv \sum_{i=0}^{2^d-1} \chi_{k+1}^{(i)} \int_{t_k}^{t_{k+1}} \frac{1}{2} \text{tr}(H_u(s, X_s^{(i)}) \sigma_k^{(i)} (\gamma^* \gamma - I_d) (\sigma_k^{(i)})^*) ds. \end{aligned}$$

The discrete Itô's formula yields:

$$\begin{aligned} (6.9) \quad \mathbb{E} |\bar{V}_T - Y_T|^2 &= |\bar{V}_0 - Y_0|^2 + 2 \mathbb{E} \sum_{k=1}^N [\bar{V}_{t_{k-1}} - Y_{t_{k-1}}] \left( \sum_{\ell=1}^4 \Delta E_k(\ell) \right) \\ &+ \mathbb{E} \sum_{k=1}^N \left[ \left( \sum_{\ell=1}^4 \Delta E_k(\ell) \right)^2 \right]. \end{aligned}$$

From the above expression, we get:

$$\begin{aligned}
& |\bar{V}_0 - Y_0|^2 + \sum_{k=1}^N \mathbb{E}[(\Delta E_k(1) + \Delta E_k(2) + \Delta E_k(4))^2] \\
&= \mathbb{E}|\bar{V}_T - Y_T|^2 - 2\mathbb{E} \sum_{k=1}^N \left[ [\bar{V}_{t_{k-1}} - Y_{t_{k-1}}] \left( \sum_{\ell=1}^4 \Delta E_k(\ell) \right) \right] \\
&\quad - \sum_{k=1}^N \mathbb{E}[(\Delta E_k(3))^2] - 2 \sum_{k=1}^N \mathbb{E}[\Delta E_k(3)(\Delta E_k(1) + \Delta E_k(2) + \Delta E_k(4))].
\end{aligned}$$

The inequality  $2ab \leq 2a^2 + (1/2)b^2$  yields

$$\begin{aligned}
& |\bar{V}_0 - Y_0|^2 + \frac{1}{2} \sum_{k=1}^N \mathbb{E}[(\Delta E_k(1) + \Delta E_k(2) + \Delta E_k(4))^2] \\
(6.10) \quad & \leq \mathbb{E}|\bar{V}_T - Y_T|^2 \\
& \quad - 2\mathbb{E} \sum_{k=1}^N \left[ [\bar{V}_{t_{k-1}} - Y_{t_{k-1}}] \left( \sum_{\ell=1}^4 \Delta E_k(\ell) \right) \right] + \sum_{k=1}^N \mathbb{E}[(\Delta E_k(3))^2].
\end{aligned}$$

Put:

$$\begin{aligned}
D(1) &\equiv -2\mathbb{E} \sum_{k=1}^N \left[ [\bar{V}_{t_{k-1}} - Y_{t_{k-1}}] \left( \sum_{\ell=1}^4 \Delta E_k(\ell) \right) \right], \\
(6.11) \quad D(2) &\equiv \sum_{k=1}^N \mathbb{E}[(\Delta E_k(3))^2], \\
D(3) &\equiv \sum_{k=1}^N \mathbb{E}[(\Delta E_k(1) + \Delta E_k(2) + \Delta E_k(4))^2].
\end{aligned}$$

*Step Three: Standard BSDE Techniques.* Admit for the moment

**Lemma 6.3.** *There exists a constant  $C_{6.3}$  such that for  $\eta \in ]0, 1[$ :*

$$\begin{aligned}
|D(1)| + D(2) &\leq C\mathcal{E}^2(\text{global}) + Ch\eta^{-1} \sum_{k=1}^N |(\bar{u}_\delta - u)(t_k, \cdot)|_\infty^2 \\
&\quad + h(\eta + h) \sum_{k=1}^N \mathbb{E}[ (|\bar{v} - v|^2)_\delta(t_{k-1}, X_{t_{k-1}}) ].
\end{aligned}$$

**Lemma 6.4.** *There exists a constant  $C_{6.4} > 0$  such that:*

$$\begin{aligned}
D(3) &\geq C_{6.4}^{-1} h \sum_{k=1}^N \mathbb{E}[ (|\bar{v} - v|^2)_\delta(t_{k-1}, X_{t_{k-1}}) ] - C_{6.4} \mathcal{E}^2(\text{global}) \\
&\quad - C_{6.4} h \sum_{k=1}^N |(\bar{u}_\delta - u)(t_k, \cdot)|_\infty^2.
\end{aligned}$$

Note to conclude the proof of Theorem 6.1 that  $Y_T = \bar{V}_T$ . Hence, for  $\eta$  small enough one obtains inequality (6.1) from (6.10), (6.11), and Lemmas 6.3 and 6.4. This completes, up to the proofs of Lemmas 6.3 and 6.4, the proof of Theorem 6.1.  $\square$

*Proof of Lemma 6.3, D(1).* From Theorem 2.1, Lemma 5.1 and Taylor's formula applied to  $\bar{V}_{t_k} - \bar{V}_{t_{k-}} = \sum_{i=0}^{2^d-1} [\chi_k^{(i)}(u(t_k, X_{t_k}^{(i)}) - u(t_k, X_{t_{k-}}^{(i)}))]$ , we derive

$$\begin{aligned}
(6.12) \quad & \mathbb{E}_{k-1}[\Delta E_k(1)] \\
&= \sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} \left\{ \mathbb{E}_{k-1}[\nabla_x u(t_k, X_{t_k}^{(i)})(X_{t_k}^{(i)} - X_{t_{k-}}^{(i)})] \right. \\
&\quad + \mathbb{E}_{k-1} \left[ \int_{t_{k-1}}^{t_k} ds \int_0^1 d\lambda (1-\lambda) \left( \text{tr}(H_u(s, X_s^{(i)}) \sigma_{k-1}^{(i)} (\gamma^* \gamma - I_d) (\sigma_{k-1}^{(i)})^*) \right. \right. \\
&\quad \left. \left. - h^{-1} \text{tr}(H_u(t_k, X_{t_k}^{(i),\lambda}) \sigma_{k-1}^{(i)} \gamma^* [(g(\Delta B^{k-1}) - \Delta B^{k-1})] \right. \right. \\
&\quad \left. \left. \times [(g(\Delta B^{k-1}) - \Delta B^{k-1})]^* \gamma (\sigma_{k-1}^{(i)})^*) \right) \right] \Big\},
\end{aligned}$$

with  $X_{t_k}^{(i),\lambda} \equiv X_{t_k}^{(i)} + \lambda(X_{t_k}^{(i)} - X_{t_{k-}}^{(i)})$ . Exploiting (3.5), we get

$$\begin{aligned}
& \mathbb{E}_{k-1}[\nabla_x u(t_k, X_{t_k}^{(i)})(X_{t_k}^{(i)} - X_{t_{k-}}^{(i)})] \\
&= \mathbb{E}_{k-1}[\nabla_x u(t_k, X_{t_k}^{(i)}) \sigma_{k-1}^{(i)} \gamma^* \mathbb{E}[g(\Delta B^{k-1}) - \Delta B^{k-1} | g(\Delta B^{k-1})]] = 0.
\end{aligned}$$

Thus with obvious notations

$$(6.13) \quad \mathbb{E}_{k-1}[\Delta E_k(1)] \equiv \sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} \int_{t_{k-1}}^{t_k} ds \int_0^1 d\lambda (1-\lambda) \mathbb{E}_{k-1}[I_{k-1}^{(i)}(s, \lambda)].$$

Noting that  $h^{-1} \mathbb{E}_{k-1}[(g(\Delta B^{k-1}) - \Delta B^{k-1})(g(\Delta B^{k-1}) - \Delta B^{k-1})^*] = I_d - (\gamma \gamma^*)^{-1}$ , we derive

$$\begin{aligned}
& \mathbb{E}_{k-1}[I_{k-1}^{(i)}(s, \lambda)] \\
&= \mathbb{E}_{k-1}[\text{tr}([H_u(s, X_s^{(i)}) - H_u(t_{k-1}, X_{t_{k-1}+}^{(i)})] \sigma_{k-1}^{(i)} (\gamma^* \gamma - I_d) (\sigma_{k-1}^{(i)})^*)] \\
&\quad - \mathbb{E}_{k-1}[h^{-1} \text{tr}([H_u(t_k, X_{t_k}^{(i),\lambda}) - H_u(t_{k-1}, X_{t_{k-1}+}^{(i)})] \\
&\quad \times \sigma_{k-1}^{(i)} \gamma^* [(g(\Delta B^{k-1}) - \Delta B^{k-1})] [(g(\Delta B^{k-1}) - \Delta B^{k-1})]^* \gamma (\sigma_{k-1}^{(i)})^*)].
\end{aligned}$$

From (3.4), note first that  $|\gamma \gamma^* - I_d| \leq CM^{-2/d}$ . The key idea is now to use the  $\alpha$ -Hölder continuity of the Hessian matrix of  $u$ , cf. Theorem 2.1, combined with the fact that for  $\beta \in [2, d+2]$ , the square Gaussian quantization is still rate optimal in  $L^\beta(\mathbb{P})$ , cf. [10], that is  $(\mathbb{E}[|g(\Delta B^{k-1}) - \Delta B^{k-1}|^\beta])^{\beta^{-1}} \leq Ch^{1/2} M^{-1/d}$ . Hence, putting  $\zeta > 1$  s.t.  $\zeta^{-1} + (\beta/2)^{-1} = 1$ , the Hölder inequality yields:

$$\begin{aligned}
& |\mathbb{E}_{k-1}[I_{k-1}^{(i)}(s, \lambda)]| \\
&\leq C \mathbb{E}_{k-1}[(h^{\alpha/2} + \sup_{t_{k-1} < s < t_k} |X_s^{(i)} - X_{t_{k-1}+}^{(i)}|^\alpha + |X_{t_k}^{(i)} - X_{t_k}^{(i)}|^\alpha \zeta) \zeta^{-1}] \\
&\quad \times [M^{-2/d} + h^{-1} \mathbb{E}_{k-1}[|\Delta B^{k-1} - g(\Delta B^{k-1})|^\beta]^{2\beta^{-1}}] \leq Ch^{\alpha/2} M^{-2/d}.
\end{aligned}$$

Plug the above control into (6.13) to obtain  $|\mathbb{E}_{k-1}[\Delta E_k(1)]| \leq Ch^{1+\alpha/2} M^{-2/d}$ .

Turn to  $\Delta E_k(2)$ . Apply the Taylor formula to the function  $u$ .

$$\begin{aligned}
\mathbb{E}_{k-1}[\Delta E_k(2)] &= \mathbb{E}_{k-1} \left[ \sum_{i=0}^{2^d-1} \chi_k^{(i)} (u(t_{k-1}, X_{t_{k-1}+}^{(i)}) - u(t_{k-1}, X_{t_{k-1}})) \right] \\
&= \sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} [u(t_{k-1}, X_{t_{k-1}+}^{(i)}) - u(t_{k-1}, X_{t_{k-1}})] \\
&= \sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} \left[ \nabla_x u(t_{k-1}, X_{t_{k-1}}) (X_{t_{k-1}+}^{(i)} - X_{t_{k-1}}) \right. \\
&\quad \left. + \int_0^1 d\lambda (1-\lambda) \text{tr} [H_u(t_{k-1}, X_{t_{k-1}} + \lambda(X_{t_{k-1}+}^{(i)} - X_{t_{k-1}})) \right. \\
&\quad \left. \times (X_{t_{k-1}+}^{(i)} - X_{t_{k-1}}) (X_{t_{k-1}+}^{(i)} - X_{t_{k-1}})^*] \right] \\
&\equiv \sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} [\Delta E_k^{(i)}(2, 1) + \Delta E_k^{(i)}(2, 2)].
\end{aligned}$$

Since the shape functions  $(\phi_z)_{z \in \mathcal{C}_\infty}$  are chosen to interpolate exactly the polynomials of order less than one, it comes  $\sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} \Delta E_k^{(i)}(2, 1) = 0$ . Moreover, the regularity properties of  $u$  yield  $|\sum_{i=0}^{2^d-1} \phi_{k-1}^{(i)} \Delta E_k^{(i)}(2, 2)| \leq C\delta^2$ , so that  $|\mathbb{E}_{k-1}[\Delta E_k(2)]| \leq Ch\mathcal{E}(\text{space})$ .

Due to (5.7), (5.9) and (6.8), the contribution of  $\Delta E(4)$  in  $D(1)$  is null.

Note now from Theorem 2.1 (boundedness of the first and second order derivatives) that  $\Delta E(3)$  may be seen as a ‘‘Hölder/Lipschitz’’ difference. Using additionally the above controls for the conditional expectations of  $\Delta E(1)$  and  $\Delta E(2)$ , we derive that there exists a constant  $C$ , such that:

$$\begin{aligned}
|D(1)| &\leq C\mathbb{E} \sum_{k=1}^N \left\{ |\bar{V}_{t_{k-1}} - Y_{t_{k-1}}| \left[ h\mathcal{E}(\text{global}) + \sum_{i=0}^{2^d-1} \left[ \phi_{k-1}^{(i)} \right. \right. \right. \\
&\quad \times \left( \int_{t_{k-1}}^{t_k} [|X_s^{(i)} - X_{t_{k-1}+}^{(i)}|^\alpha + |\bar{V}_s^{(i)} - \bar{u}(t_k, X_{t_{k-1}+}^{(i)})|] ds \right. \\
&\quad \left. \left. \left. + \int_{t_{k-1}}^{t_k} [|\bar{W}_s^{(i)} - \bar{v}(t_{k-1}, X_{t_{k-1}+}^{(i)})|] ds \right) \right] \right] \right\}.
\end{aligned}$$

Recall that  $\bar{V}_s^{(i)} = u(s, X_s^{(i)})$ ,  $\bar{W}_s^{(i)} = v(s, X_s^{(i)})$ . From Theorem 2.1 (Hölder regularity of  $u$  in  $t$ ), (6.6) (regularity of  $\bar{V}$  and  $\bar{W}$ ) and Lemma 6.2 (control of the increments of  $X$ ), we then deduce:

$$\begin{aligned}
(6.14) \quad |D(1)| &\leq C\mathbb{E} \sum_{k=1}^N \left[ |\bar{V}_{t_{k-1}} - Y_{t_{k-1}}| \left( h\mathcal{E}(\text{global}) + h^{1+\alpha/2} \right. \right. \\
&\quad \left. \left. + h \{ |(\bar{u}_\delta - u)(t_k, \cdot)|_\infty + |\bar{v} - v|_\delta(t_{k-1}, X_{t_{k-1}}) \} \right) \right].
\end{aligned}$$

Since  $h^{1+\alpha/2} = h\mathcal{E}(\text{time})$ , we derive, from Young’s inequality, the required control for  $D(1)$ .  $\square$



*Proof for  $D(2)$ .* Recall that  $D(2) = \mathbb{E} \sum_{k=1}^N [(\Delta E_k(3))^2]$ . Again, the terms  $\Delta E_k(3)$  are “Hölder/Lipschitz” differences, so that the strategy used to obtain (6.14) applies. This provides the required bound for  $D(2)$ .  $\square$

6.2.1. *Proof of Lemma 6.4.* Recall that  $D(3) = \sum_{k=1}^N \mathbb{E} [(\Delta E_k(1) + \Delta E_k(2) + \int_{t_{k-1}}^{t_k} [W_s \gamma^* - \bar{Z}_s] dB_s + \Delta R_k(2))^2]$ . Now, sticking to the notations introduced for the proof of Lemma 6.3 (see (6.12)), Taylor’s formula applied to  $\Delta E_k(1)$  yields:

$$\begin{aligned}
& \int_{t_{k-1}}^{t_k} [W_s \gamma^* - \bar{Z}_s] dB_s + \Delta E_k(1) \\
&= \left\{ \sum_{i=0}^{2^d-1} \chi_k^{(i)} \left( v(t_{k-1}, X_{t_{k-1}+}^{(i)}) \gamma^* g(\Delta B^{k-1}) - \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} dB_s \right) \right\} \\
&+ \left\{ \sum_{i=0}^{2^d-1} \chi_k^{(i)} [\nabla_x u(t_k, X_{t_k}^{(i)}) \sigma_{k-1}^{(i)} - v(t_{k-1}, X_{t_{k-1}+}^{(i)})] \gamma^* (g(\Delta B^{k-1}) - \Delta B^{k-1}) \right. \\
&- \sum_{i=0}^{2^d-1} \chi_k^{(i)} \int_0^1 d\lambda (1-\lambda) \text{tr} [H_u(t_k, X_{t_k}^{(i),\lambda}) \sigma_{k-1}^{(i)} \gamma^* (g(\Delta B^{k-1}) - \Delta B^{k-1}) \\
&\quad \times (g(\Delta B^{k-1}) - \Delta B^{k-1})^* \gamma(\sigma_{k-1}^{(i)})^*] \\
&\left. + \sum_{i=0}^{2^d-1} \chi_k^{(i)} \int_{t_{k-1}}^{t_k} [W_s^{(i)} - v(t_{k-1}, X_{t_{k-1}+}^{(i)})] \gamma^* dB_s + \Delta R_k(3) \right\} \\
&\equiv A_k(1) + A_k(2).
\end{aligned}$$

From Theorem 2.1, (6.8) and standard estimates for the increments of  $X$ , it comes  $\sum_{k=1}^N \mathbb{E} [(A_k(2))^2] \leq Ch$ . Thus,  $D(3) \geq (1/2) \sum_{j=1}^N \mathbb{E} [(A_k(1) + \Delta R_k(2) + \Delta E_k(2))^2] - Ch$ . From (5.11), (6.8) and Lemma 5.1, we already know that the covariance between the stochastic integral (resp. the quantized Brownian increment) and  $\Delta R_k(2)$  vanishes. Moreover,

$$\begin{aligned}
& \mathbb{E}_{k-1} [A_k(1) \Delta E_k(2)] \\
&= \sum_{i=0}^{2^d-1} \mathbb{E}_{k-1} \left[ \chi_k^{(i)} \left( v(t_{k-1}, X_{t_{k-1}+}^{(i)}) \gamma^* g(\Delta B^{k-1}) \right. \right. \\
&\quad \left. \left. - \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} dB_s \right) (u(t_{k-1}, X_{t_{k-1}+}^{(i)}) - u(t_{k-1}, X_{t_{k-1}})) \right] = 0.
\end{aligned}$$

We finally obtain

$$\begin{aligned}
(6.15) \quad D(3) &\geq \frac{1}{2} \sum_{k=1}^N \sum_{i=0}^{2^d-1} \mathbb{E} \left[ \phi_{k-1}^{(i)} \mathbb{E}_{k-1} \left[ \left( v(t_{k-1}, X_{t_{k-1}+}^{(i)}) \gamma^* g(\Delta B^{k-1}) \right. \right. \right. \\
&\quad \left. \left. \left. - \int_{t_{k-1}}^{t_k} \bar{Z}_s^{(i)} dB_s \right)^2 \right] \right] - Ch.
\end{aligned}$$

A new application of the Taylor formula of order one to  $u(t_k, X_{t_k}^{(i)})$  with respect to the initial value  $(t_{k-1}, X_{t_{k-1}+}^{(i)})$  yields, for all  $i \in \llbracket 0, 2^d \rrbracket$  and  $k \in \llbracket 0, N \rrbracket$ ,  $u(t_k, X_{t_k}^{(i)}) =$

$u(t_{k-1}, X_{t_{k-1}}^{(i)}) + \nabla_x u(t_{k-1}, X_{t_{k-1}}^{(i)})(X_{t_k}^{(i)} - X_{t_{k-1}}^{(i)}) + O(h)$ , so that

$$(6.16) \quad \mathbb{E}_{k-1} \left[ \left( u(t_k, X_{t_k}^{(i)}) - \mathbb{E}_{k-1}[u(t_k, X_{t_k}^{(i)})] - v(t_{k-1}, X_{t_{k-1}}^{(i)}) \gamma^* g(\Delta B^{k-1}) \right)^2 \right] \leq Ch^2.$$

Hence, denoting by  $\mathbb{V}_{k-1}$  the conditional variance with respect to  $\mathcal{F}_{t_{k-1}}$ , it comes from (6.15) and (5.5)

$$(6.17) \quad D(3) \geq \frac{1}{4} \sum_{k=1}^N \sum_{i=0}^{2^d-1} \mathbb{E}[\phi_{k-1}^{(i)} \mathbb{V}_{k-1}[(u - \bar{u}_\delta)(t_k, X_{t_k}^{(i)})]] - Ch.$$

Since  $v(t_{k-1}, X_{t_{k-1}}^{(i)}) = h^{-1} \mathbb{E}_{k-1}[v(t_{k-1}, X_{t_{k-1}}^{(i)}) \gamma^* g(\Delta B^{k-1}) g(\Delta B^{k-1})^* \gamma]$ , the Cauchy-Schwarz inequality and (6.16) yield

$$\begin{aligned} & \left| \mathbb{E}_{k-1} \left[ h^{-1} u(t_k, X_{t_k}^{(i)}) g(\Delta B^{k-1})^* \gamma \right] - v(t_{k-1}, X_{t_{k-1}}^{(i)}) \right| \\ &= h^{-1} \left| \mathbb{E}_{k-1} \left[ \left( u(t_k, X_{t_k}^{(i)}) - \mathbb{E}_{k-1}[u(t_k, X_{t_k}^{(i)})] - v(t_{k-1}, X_{t_{k-1}}^{(i)}) \gamma^* g(\Delta B^{k-1}) \right) g(\Delta B^{k-1})^* \gamma \right] \right| \leq Ch^{1/2}, \end{aligned}$$

so that

$$\begin{aligned} & h |(v - \bar{v})(t_{k-1}, X_{t_{k-1}}^{(i)})|^2 \\ & \leq Ch^{-1} \left| \mathbb{E}_{k-1} \left[ (u - \bar{u}_\delta)(t_k, X_{t_k}^{(i)}) g(\Delta B^{k-1})^* \gamma \right] \right|^2 + Ch^2 \\ & \leq C \mathbb{V}_{k-1}[(u - \bar{u}_\delta)(t_k, X_{t_k}^{(i)})] + Ch^2. \end{aligned}$$

Due to (6.17), this completes the proof.  $\square$

## 7. A BOUND WHEN $b$ DEPENDS ON THE GRADIENT

As announced in Section 3, we now provide a bound for  $\mathcal{E}(\text{global})$  when  $b$  depends on  $v$ . As already explained, the exhibited control is the same as in Theorem 3.2 provided  $\delta^2 \leq h$  and  $\delta^{-3} h^{3/2} M^{-2/d} \leq 1$ . For the typical value  $\delta = h^{1/2+\alpha/4}$ , this amounts to choose  $M$  equal (up to a constant) to  $h^{-3d\alpha/8}$ . Of course, this is not satisfactory.

What we feel is the following: we expect  $\delta = h^{1/2}$  to be the good value for the spatial parameter provided an interpolation of order  $1 + \lceil \alpha \rceil$ ,  $2 + \alpha$  denoting in a more general way than in Assumption **(A)** the spatial regularity of the true solution and being possibly greater than three. In this frame,  $\delta^{-3} h^{3/2} M^{-2/d}$  is less than  $M^{-2/d}$  and is thus well controlled. Again, we are not able to do so because the Lagrange interpolation procedures of order greater than two involve signed weights.

The other point to note is that we are not even able to establish the convergence of Algorithm 3.1 in the general frame  $b = b(x, u, v)$ . The reason was already discussed in our previous paper [8]: one way or another, the analysis of Algorithm 3.1 leads to investigate the regularity of the approximated gradient  $\bar{v}$ . This task is certainly far from being trivial.

The common strategy consists in introducing an intermediate predictor for  $v$ , denoted by  $\hat{v}$ , for which the underlying estimates are easier to establish. In [8],  $\hat{v}(t_k, \cdot)$ ,  $k \in \llbracket 0, N \rrbracket$ , is given as a regularized version of  $\bar{v}(t_{k+1}, \cdot)$ . Our choice in the

current paper is slightly different. For  $k \in \llbracket 0, N \rrbracket$  and  $x \in \mathcal{C}_\infty$ , set:

$$(7.1) \quad \begin{aligned} \hat{v}(t_k, x) &\equiv h^{-1} \mathbb{E}[\bar{u}_\delta(t_{k+1}, x + \mathcal{T}^0(t_k, x))g(\Delta B^k)^*]\gamma, \\ \mathcal{T}^0(t_k, x) &\equiv \Sigma(t_k, x)\gamma^*g(\Delta B^k). \end{aligned}$$

Define the transition  $\mathcal{T}(t_k, x) \equiv \beta(t_k, x)h + \Sigma(t_k, x)\gamma^*g(\Delta B^k)$  with  $\beta(t_k, x) = b(x, \bar{u}(t_{k+1}, x), \hat{v}(t_k, x))$ . The definitions of  $\bar{u}(t_k, x)$  and  $\bar{v}(t_k, x)$  then remain the same (according to the new choice for  $\mathcal{T}(t_k, x)$ ).

The above definition is in fact rather natural: when replacing  $\bar{u}_\delta$  by the true solution  $u$ , the first approximation of  $h^{-1} \mathbb{E}[u(t_{k+1}, x + \mathcal{T}^0(t_k, x))g(\Delta B^k)^*]\gamma$  remains  $v(t_k, x)$  itself.

To investigate the convergence of the above algorithm, we also need to change the way we interpolate the approximated solution at the nodes of the grid. Indeed, for our analysis, we require  $\bar{u}_\delta$  to be sufficiently smooth in  $x$ . A possible strategy consists in choosing a  $B$ -spline basis instead of a family of Lagrange kernels of order one, see e.g. the monograph of de Boor [5]. Namely, choose  $\Phi$  in (3.1) as  $\Phi(t) = 2/3 - t^2 + |t|^3/2$  for  $|t| \leq 1$ ,  $\Phi(t) = (1/6)(2 - |t|)^3$  for  $1 \leq |t| \leq 2$  and  $\Phi(t) = 0$  for  $|t| \geq 2$ . Of course, this changes the number of “useful neighbors” in (5.1): there are now  $4^d$  instead of  $2^d$  neighbors to consider. Anyhow, this  $\Phi$  is nonnegative and the sum  $\sum_{z \in \mathcal{C}_\infty} \phi_z(x)$  still matches one for  $x \in \mathbb{R}^d$ , so that the probabilistic interpretation of Lemma 5.1 still holds.

As expected, the approximated solution  $\bar{u}_\delta$  then belongs to  $\mathcal{C}^2$  with Lipschitz derivatives of order one and two.

We claim that the main steps of the proof of Theorem 3.2 still hold in this new frame. Anyhow, several differences are to be quoted:

- First, the interpolation procedure associated to the spline basis is exact for constant functions, but is not anymore for polynomials of order one. In fact, this is not a big deal: the underlying error is of order  $\delta^2$ , and induces, due to the propagation phenomenon along the time mesh, an additional error of the same order as  $\mathcal{E}(\text{space})$ .
- Second, the drift of the approximate transition is not bounded anymore, since it now depends on  $\hat{v}$ . Following the proof of Proposition 5.5, we could establish that the *supremum* norm of  $\hat{v}$  is bounded by  $Ch^{-1/2}$ . In particular, the drift of the approximate transition is bounded by  $Ch^{1/2}$ , and is thus of the same order as the stochastic part. This is sufficient to apply the previous machinery (think for example to Lemma 6.2).
- Third, the estimates for  $D(1)$  and  $D(2)$  in Lemma 6.3 involve a new term that refers to the difference between  $\hat{v}$  and  $v$ . This leads to a new version for Theorem 6.1. Following [8], it now writes with  $Ch \sum_{k=0}^{N-1} \mathbb{E}[(\hat{v} - \bar{v})_\delta^2(t_k, X_{t_k})]$  in addition to the right hand side.

The best that we can prove for the third point is that for  $x \in \mathbb{R}^d$  and  $k \in \llbracket 0, N \rrbracket$

$$(7.2) \quad \begin{aligned} (|\hat{v} - \bar{v}|^2)_\delta(t_k, x) &\leq C(h^{-1}\delta^4 + h) \\ &\quad + C|(\bar{u} - u)(t_{k+1}, \cdot)|_\infty^2 [1 + (|\hat{v}|^2)_\delta(t_k, x)] \mathcal{B}^2(\text{gradient}), \end{aligned}$$

with  $\mathcal{B}(\text{gradient}) = 1 + \delta^{-3}h^{3/2}M^{-2/d}$  (the proof is postponed to the end of the section). Up to a modification of the constant  $C$ , we can also write:

$$\begin{aligned} (|\hat{v} - \bar{v}|^2)_\delta(t_k, x) &\leq C(h^{-1}\delta^4 + h) \\ &\quad + C|(\bar{u} - u)(t_{k+1}, \cdot)|_\infty^2 (|\hat{v} - \bar{v}|^2)_\delta(t_k, x) \mathcal{B}^2(\text{gradient}) \\ &\quad + C|(\bar{u} - u)(t_{k+1}, \cdot)|_\infty^2 [1 + (|\bar{v}|^2)_\delta(t_k, x)] \mathcal{B}^2(\text{gradient}). \end{aligned}$$

Combine this bound with the new version of Theorem 6.1. For  $\mathcal{B}(\text{gradient}) \leq 2$ , obtain from Gronwall's lemma (up to the modification of the initial condition) the following system of inequalities:

$$\begin{aligned} (7.3) \quad &|(\bar{u} - u)(t_k, \cdot)|_\infty^2 \leq C\mathcal{E}^2(\text{global}) + Ch \sum_{j=k}^{N-1} \mathbb{E}[(|\bar{v} - \hat{v}|^2)_\delta(t_j, X_{t_j})] \\ &\forall x \in \mathbb{R}^d, (|\hat{v} - \bar{v}|^2)_\delta(t_k, x) \leq C(h^{-1}\delta^4 + h) \\ &\quad + 4C|(\bar{u} - u)(t_{k+1}, \cdot)|_\infty^2 (|\hat{v} - \bar{v}|^2)_\delta(t_k, x) \\ &\quad + 4C|(\bar{u} - u)(t_{k+1}, \cdot)|_\infty^2 [1 + (|\bar{v}|^2)_\delta(t_k, x)]. \end{aligned}$$

Assume now that  $(C + 2C^2)\mathcal{E}^2(\text{global}) \exp(8C^2T + 8C^2C_{5.5}) \leq 1/(8C)$ . We prove by induction that for all  $k \in \llbracket 0, N \rrbracket$ ,

$$(7.4) \quad |(\bar{u} - u)(t_k, \cdot)|_\infty^2 \leq (C + 2C^2)\mathcal{E}^2(\text{global}) \exp(8C^2T + 8C^2C_{5.5}),$$

Inequality (7.4) clearly holds for  $k = N$ . Assume that it is true at given ranks  $k+1, k+2, \dots, N$ ,  $k \in \llbracket 0, N \rrbracket$ , and prove that it holds at the rank  $k$ . Due to (7.3) (second inequality), (7.4) and the assumed bound for  $\mathcal{E}(\text{global})$ , it comes for all  $j \in \llbracket k, N \rrbracket$  and  $x \in \mathbb{R}^d$ :

$$\begin{aligned} (|\hat{v} - \bar{v}|^2)_\delta(t_j, x) &\leq C(h^{-1}\delta^4 + h) + (1/2)(|\hat{v} - \bar{v}|^2)_\delta(t_j, x) \\ &\quad + 4C|(\bar{u} - u)(t_{j+1}, \cdot)|_\infty^2 [1 + (|\bar{v}|^2)_\delta(t_j, x)], \end{aligned}$$

so that,

$$(7.5) \quad (|\hat{v} - \bar{v}|^2)_\delta(t_j, x) \leq 2C(h^{-1}\delta^4 + h) + 8C|(\bar{u} - u)(t_{j+1}, \cdot)|_\infty^2 [1 + (|\bar{v}|^2)_\delta(t_j, x)].$$

Prove now (7.4). Referring to the first line in (7.3) and to (7.5), it comes:

$$\begin{aligned} &|(\bar{u} - u)(t_k, \cdot)|_\infty^2 \\ &\leq C\mathcal{E}^2(\text{global}) \\ &\quad + Ch \sum_{j=k}^{N-1} \{2C(h^{-1}\delta^4 + h) + 8C|(\bar{u} - u)(t_{j+1}, \cdot)|_\infty^2 [1 + \mathbb{E}[(|\bar{v}|^2)_\delta(t_j, X_{t_j})]]\} \\ &\leq (C + 2C^2)\mathcal{E}^2(\text{global}) \\ &\quad + 8C^2h \sum_{j=k}^{N-1} |(\bar{u} - u)(t_{j+1}, \cdot)|_\infty^2 [1 + \mathbb{E}[(|\bar{v}|^2)_\delta(t_j, X_{t_j})]]. \end{aligned}$$

A discrete version of Gronwall's lemma and Proposition 5.5 yield the result.  $\square$

*Proof of (7.2).* For  $x \in \mathcal{C}_\infty$ , it comes from the very definition of  $\hat{v}$  and  $\bar{v}$  (see Algorithm 3.1 and (7.1)):

$$\begin{aligned}
 & (\hat{v} - \bar{v})(t_k, x) \\
 &= h^{-1} \mathbb{E}[(\bar{u}_\delta(t_{k+1}, x + \mathcal{T}^0(t_k, x)) - \bar{u}_\delta(t_{k+1}, x + \mathcal{T}(t_k, x)))g(\Delta B^k)^*] \gamma \\
 &= h^{-1} \mathbb{E}[(\bar{u} - u)_\delta(t_{k+1}, x + \mathcal{T}^0(t_k, x)) \\
 (7.6) \quad & - (\bar{u} - u)_\delta(t_{k+1}, x + \mathcal{T}(t_k, x))]g(\Delta B^k)^*] \gamma \\
 &+ h^{-1} \mathbb{E}[(u_\delta(t_{k+1}, x + \mathcal{T}^0(t_k, x)) - u_\delta(t_{k+1}, x + \mathcal{T}(t_k, x)))g(\Delta B^k)^*] \gamma \\
 &\equiv G(1, x) + G(2, x).
 \end{aligned}$$

Start with  $G(1, x)$ .

$$\begin{aligned}
 G(1, x) &= -\mathbb{E} \left[ \int_0^1 [\nabla_x (\bar{u} - u)_\delta(t_{k+1}, x + \lambda \beta(t_k, x)h + \Sigma(t_k, x)\gamma^* g(\Delta B^k)) \right. \\
 (7.7) \quad & \left. \times \beta(t_k, x)g(\Delta B^k)^*] d\lambda \right] \gamma.
 \end{aligned}$$

Admit for the moment the following Lemma:

**Lemma 7.1.** *There exists a constant  $C_{7.1}$  such that for every  $\ell \in \llbracket 1, d \rrbracket$  and for every bounded function  $\varphi \in C^2(\mathbb{R}^d, \mathbb{R})$  with bounded derivatives of order one and two and with Lipschitz continuous second order derivatives*

$$\begin{aligned}
 & \left| \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_\ell} (g(\Delta B^k))g(\Delta B^k)^* \right] \right| \\
 & \leq C_{7.1} [|\varphi|_\infty + hM^{-2/d}|\nabla^{(2)}\varphi|_\infty + h^{3/2}M^{-2/d}|\nabla^{(3)}\varphi|_\infty].
 \end{aligned}$$

Apply Lemma 7.1 to  $G(1, x)$  or more specifically to the function  $y \in \mathbb{R}^d \mapsto (\bar{u} - u)_\delta(t_{k+1}, x + \lambda \beta(t_k, x)h + \Sigma(t_k, x)\gamma^* y)\beta_i(t_k, x)$ , for  $i \in \llbracket 1, d \rrbracket$  and  $\lambda \in (0, 1)$ . The function  $(\bar{u} - u)_\delta(t_{k+1}, \cdot)$  belongs to  $C^2(\mathbb{R}^d, \mathbb{R})$  with bounded and Lipschitz continuous derivatives of order one and two, and for  $i \in \llbracket 2, 3 \rrbracket$ ,  $|\nabla^{(i)}(\bar{u} - u)_\delta(t_{k+1}, \cdot)|_\infty \leq C|(\bar{u} - u)_\delta(t_{k+1}, \cdot)|_\infty \delta^{-i}$ . It comes

$$(7.8) \quad |G(1, x)| \leq C|(\bar{u} - u)_\delta(t_{k+1}, \cdot)|(1 + |\hat{v}(t_k, x)|)\mathcal{B}(\text{gradient}).$$

Turn now to  $G(2, x)$ . Since  $u(t_{k+1}, \cdot)$  belongs to  $C^2(\mathbb{R}^d, \mathbb{R})$ , with bounded derivatives, the distance between  $u$  and  $u_\delta$  is worth  $C\delta^2$ , so that

$$\begin{aligned}
 |G(2, x)| &\leq C(h^{-1/2}\delta^2 \\
 &+ h^{-1}|\mathbb{E}[(u(t_{k+1}, x + \mathcal{T}^0(t_k, x)) - u(t_{k+1}, x + \mathcal{T}(t_k, x)))g(\Delta B^k)^*]|).
 \end{aligned}$$

Expand the last term as  $G(1, x)$  in (7.7) and substract  $\nabla_x u(t_{k+1}, x + \lambda \beta(t_k, x)h)$  to  $\nabla_x u(t_{k+1}, x + \lambda \beta(t_k, x)h + \Sigma(t_k, x)\gamma^* g(\Delta B^k))$  (recall that  $g(\Delta B^k)$  is centered). We obtain  $|G(2, x)| \leq C(h^{-1/2}\delta^2 + h|\beta(t_k, x)|) \leq C(h^{-1/2}\delta^2 + h^{1/2})$  since  $|\beta(t_k, x)| \leq Ch^{-1/2}$  (see Proposition 5.5). Thanks to (7.6) and (7.8), we derive (7.2).  $\square$

*Proof of Lemma 7.1*

$$\begin{aligned}
 & \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_\ell} (g(\Delta B^k))g(\Delta B^k)^* \right] \stackrel{(3.5)}{=} \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_\ell} (g(\Delta B^k))(\Delta B^k)^* \right] \\
 &= \mathbb{E} \left[ \left( \frac{\partial \varphi}{\partial x_\ell} (g(\Delta B^k)) - \frac{\partial \varphi}{\partial x_\ell} (\Delta B^k) \right) (\Delta B^k)^* \right] + \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_\ell} (\Delta B^k) (\Delta B^k)^* \right] \\
 &\equiv \Phi(1) + \Phi(2).
 \end{aligned}$$

Investigate first  $\Phi(1)$

$$\begin{aligned}
\Phi(1) &= \mathbb{E}\left[\left(\frac{\partial\varphi}{\partial x_\ell}(g(\Delta B^k)) - \frac{\partial\varphi}{\partial x_\ell}(\Delta B^k)\right)((\Delta B^k)^* - g(\Delta B^k)^*)\right] \\
&\quad + \mathbb{E}\left[\left(\frac{\partial\varphi}{\partial x_\ell}(g(\Delta B^k)) - \frac{\partial\varphi}{\partial x_\ell}(\Delta B^k)\right)g(\Delta B^k)^*\right] \\
&= \mathbb{E}\left[\left(\frac{\partial\varphi}{\partial x_\ell}(g(\Delta B^k)) - \frac{\partial\varphi}{\partial x_\ell}(\Delta B^k)\right)((\Delta B^k)^* - g(\Delta B^k)^*)\right] \\
&\quad - \int_0^1 \left[\nabla_x\left(\frac{\partial\varphi}{\partial x_\ell}\right)(g(\Delta B^k) + \lambda(\Delta B^k - g(\Delta B^k)))\right. \\
&\quad \quad \left.\times (\Delta B^k - g(\Delta B^k))g(\Delta B^k)^*\right] d\lambda \\
&= \mathbb{E}\left[\left(\frac{\partial\varphi}{\partial x_\ell}(g(\Delta B^k)) - \frac{\partial\varphi}{\partial x_\ell}(\Delta B^k)\right)((\Delta B^k)^* - g(\Delta B^k)^*)\right] \\
&\quad - \int_0^1 \left[\left(\nabla_x\left(\frac{\partial\varphi}{\partial x_\ell}\right)(g(\Delta B^k) + \lambda(\Delta B^k - g(\Delta B^k))) - \nabla_x\left(\frac{\partial\varphi}{\partial x_\ell}\right)(g(\Delta B^k))\right)\right. \\
&\quad \quad \left.\times (\Delta B^k - g(\Delta B^k))g(\Delta B^k)^*\right] d\lambda
\end{aligned}$$

using (3.5) for the last equality. It comes  $|\Phi(1)| \leq C(|\nabla^{(2)}\varphi|_\infty + |\nabla^{(3)}\varphi|_\infty h^{1/2})h \times M^{-2/d}$ . Deal finally, for  $j \in \llbracket 1, d \rrbracket$ , with the  $j^{\text{th}}$  coordinate of  $\Phi(2)$ :

$$\begin{aligned}
\Phi_j(2) &= (2\pi)^{-d/2} h^{1/2} \int_{\mathbb{R}^d} \frac{\partial\varphi}{\partial y_\ell}(h^{1/2}y) y_j \exp(-|y|^2/2) dy \\
&= -(2\pi)^{-d/2} \int_{\mathbb{R}^d} \varphi(h^{1/2}y) \frac{\partial}{\partial y_\ell}(y_j \exp(-|y|^2/2)) dy.
\end{aligned}$$

Hence,  $|\Phi(2)| \leq C|\varphi|_\infty$ . This completes the proof.  $\square$

#### REFERENCES

1. F. Antonelli, *Backward-Forward Stochastic Differential Equations*, Ann. Appl. Prob. **3-3** (1993), 777–793.
2. V. Bally, G. Pagès, and J. Printems, *A quantization tree method for pricing and hedging multi-dimensional American options*, Math. Finance **15** (2005), 119–168.
3. B. Bouchard and N. Touzi, *Discrete time approximation and Monte-Carlo simulation of Backward Stochastic Differential Equations.*, Stoch. Proc. Appl. **111** (2004), 175–206.
4. S. Brenner and L. Scott, *The Mathematical Theory of Finite Element Methods*, Second edition. Texts in Applied Mathematics, Springer-Verlag, New York, 2002.
5. C. de Boor, *A pratical guide to splines. revised edition*, Springer-Verla, New York, 2001.
6. F. Delarue, *On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case.*, Stoch. Proc. Appl. **99** (2002), 209–286.
7. F. Delarue and G. Guatteri, *Weak existence and uniqueness for FBSDEs.*, To appear in Stoch. Proc. Appl. (2006).
8. F. Delarue and S. Menozzi, *A Forward Backward Stochastic Algorithm for Quasilinear PDEs.*, Ann. Appl. Prob. **16-1** (2006), 140–184.
9. J. Douglas, J. Ma, and P. Protter, *Numerical methods for Forward-Backward Stochastic Differential Equations*, Ann. Appl. Prob. **6** (1996), 940–968.
10. S. Graf, H. Luschgy, and G. Pagès, *The (r,s)-problem in the quantization of probability measures*, Preprint (2006).
11. S. Graf and H. Lushgy, *Foundations of quantization for random vectors*, LNM-1730, Springer-Verlag, 2000.
12. Y. Hu, P. Imkeller, and M. Müller, *Utility maximization in incomplete markets*, Ann. Appl. Prob. **15** (2005).
13. J. Jacod and A.N. Shiryaev, *Limit theorems for stochastic processes, second edition*, Springer Verlag, 2004.

14. M. Kobylanski, *Backward stochastic differential equations and partial differential equations with quadratic growth*, Ann. Prob. **28** (2000).
15. O.A. Ladyzhenskaya, V.A. Solonnikov, and N.N. Ural'ceva, *Linear and quasilinear equations of parabolic type.*, Translations of Mathematical Monographs, Vol. 23, American Mathematical Society, Providence., 1967.
16. J.P. Lemor, E. Gobet, and X. Warin, *Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations*, Technical Report of CMAP, Ecole Polytechnique, Palaiseau, France **574** (2005).
17. ———, *A regression-based monte-carlo method to solve backward stochastic differential equations*, Ann. Appl. Prob. **15** (2005), 2172–2002.
18. J. Ma, P. Protter, and J. Yong, *Solving Forward-Backward Stochastic Differential Equations explicitly - a four step scheme*, Prob. Th. Rel. Fields **98** (1994), 339–359.
19. J. Ma and J. Yong, *Forward-backward stochastic differential equations and their applications*, LNM-1702, Springer-Verlag, 1999.
20. G. N. Milstein and M. V. Tretyakov, *Numerical algorithms for semilinear parabolic equations with small parameter based on approximation of stochastic equations.*, Math. Comp. **69–229** (1999), 237–267.
21. ———, *Numerical algorithms for forward-backward stochastic differential equations connected with semilinear parabolic equations*, Technical Report, School of Mathematics and Computer Science, University of Leicester, U.K. **21** (2004).
22. ———, *Solving forward-backward stochastic differential equations and related quasilinear parabolic equations*, Technical Report, School of Mathematics and Computer Science, University of Leicester, U.K. **35** (2004).
23. G.N. Milstein and M.V. Tretyakov, *Stochastic numerics for mathematical physics*, Springer-Verlag, Berlin, 2004.
24. E. Pardoux and S.G. Peng, *Adapted solution of a Backward Stochastic Differential Equation.*, Systems Control Lett. **14-1** (1990), 55–61.
25. A.N. Shiryaev, *Probability, second edition.*, Graduate Texts in Mathematics, 95. Springer-Verlag, New York., 1996.
26. W.A. Woyczyński, *Burgers-KPZ turbulence*, LNM-1700, Springer-Verlag, 1998.

UNIVERSITÉ PARIS 7, UFR DE MATHÉMATIQUES, CASE 7012, 2, PLACE JUSSIEU, 75251 PARIS CEDEX 05, FRANCE

*Current address:* Université Paris 7

*E-mail address:* delarue@math.jussieu.fr

UNIVERSITÉ PARIS 7, UFR DE MATHÉMATIQUES, CASE 7012, 2, PLACE JUSSIEU, 75251 PARIS CEDEX 05, FRANCE

*Current address:* Université Paris 7

*E-mail address:* menozzi@math.jussieu.fr