# ECOLE POLYTECHNIQUE

## CENTRE DE MATHÉMATIQUES APPLIQUÉES

*UMR CNRS 7641*

91128 PALAISEAU CEDEX (FRANCE). Tél: 01 69 33 41 50. Fax: 01 69 33 30 11

http://www.cmap.polytechnique.fr/


# A Forward-Backward Stochastic Algorithm for Quasi-Linear PDEs.


François Delarue and Stéphane Menozzi



**R.I. N$^0$ 551** *September 2004*

# A Forward-Backward Stochastic Algorithm for Quasi-Linear PDEs

François Delarue[(a),*] and Stéphane Menozzi[(b),(c),†]

(a) Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris VII, UFR de Mathématiques, Case 7012,
2, Place Jussieu, 75251 Paris Cedex 05 - FRANCE.
(b) Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris VI, Case 188,
4, Place Jussieu, 75252 Paris Cedex 05 - FRANCE.
(c) Ecole Polytechnique, Centre de Mathématiques Appliquées
91128 Palaiseau CEDEX - FRANCE.

14th September 2004

**Abstract**

We propose a time-space discretization scheme for quasi-linear parabolic PDEs. The algorithm relies on the theory of fully coupled Forward-Backward SDEs, which provides an efficient probabilistic representation of this type of equations. The derivated algorithm holds for strong solutions defined on any interval of arbitrary length. As a bypass product, we obtain a discretization procedure for the underlying FBSDE. In particular, our work provides an alternative to the method described in Douglas, Ma and Protter [DMP96] and weakens the regularity assumptions required in this reference.

*AMS 2000 Subject Classification.* Primary: 65C30. Secondary: 35K55, 60H10, 60H35.

*Keywords.* Discretization scheme, FBSDEs, Quantization, Quasi-linear PDEs.

# 1 Introduction

Introduced first by Antonelli [Ant93], and then by Ma, Protter and Yong [MPY94], Forward-Backward Stochastic Differential Equations (FBSDEs in short) provide an extension of the Feynman-Kac representation to a certain class of quasi-linear parabolic PDEs. These equations also appear in a large number of application fields such as the Hamiltonian formulation of control problems or the option hedging problem with large investors in financial mathematics (i.e. when the wealth or strategy of an agent has an impact on the volatility). We refer to the monograph of Ma and Yong, [MY99] for details and further applications.

## 1.1 FBSDE Theory and Discretization Algorithm

***Connection between FBSDEs and Quasi-linear parabolic PDEs.*** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a $d$-dimensional Brownian motion $(B_t)_{t\in[0,T]}$, where $T$ denotes an arbitrarily prescribed nonnegative real. For a given initial condition $x_0 \in \mathbb{R}^d$, a Forward-Backward SDE strongly couples a diffusion process

*delarue@math.jussieu.fr
†menozzi@ccr.jussieu.fr, menozzi@cmapx.polytechnique.fr

$U$ to the solution $(V, W)$ of a Backward SDE (as defined in the earlier work of Pardoux and Peng [PP90]):

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

In the whole paper, the coefficients $b$, $f$, $\sigma$ and $H$ are deterministic (and for simplicity also time independent). In this case, Ma, Protter and Yong [MPY94], Pardoux and Tang [PT99] and Delarue [Del02] have investigated in detail the link with the following quasi-linear PDE on $[0,T[\times\mathbb{R}^d$:

$$(\mathcal{E})\begin{cases}\partial_t u(t,x) + \langle b(x, u(t,x), \nabla_x u(t,x)\sigma(x, u(t,x))), \nabla_x u(t,x)\rangle + \frac{1}{2}\mathrm{tr}(a(x,u(t,x))\nabla^2_{x,x}u(t,x))\\ +f(x, u(t,x), \nabla_x u(t,x)\sigma(x,u(t,x))) = 0,\\ u(T,x) = H(x),\end{cases}$$

with $a(x,y) = (\sigma\sigma^*)(x,y),\ (x,y) \in \mathbb{R}^d \times \mathbb{R}$.

***A Probabilistic Numerical Method for FBSDEs and Quasi-Linear PDEs.*** This paper aims to derive from the probabilistic theory of FBSDEs a completely tractable algorithm to approximate the solution of the equation $(\mathcal{E})$. As a bypass product, the procedure also provides a discretization of the triple $(U, V, W)$.

Most of the available numerical methods proposed so far are purely analytic and involve finite-difference or finite-element techniques to approximate the solution $u$ of $(\mathcal{E})$. For example, the discretization procedure for FBSDEs of type $(E)$ given in Douglas, Ma and Protter [DMP96] consists in discretizing first the PDE $(\mathcal{E})$ and then in deriving an approximation of the underlying FBSDE.

At the opposite, we propose in this paper to derive from the FBSDE representation a numerical scheme for quasi-linear equations of type $(\mathcal{E})$. This strategy finds its origin in the earlier work of Chevance [Che97] who introduced a time-space discretization scheme in the decoupled or so-called "pure backward" case. In this latter frame, the coefficients $b$ and $\sigma$ do not depend on $V$ and $W$ and the forward equation reduces to a classical SDE. The process $U$ then appears as an "objective diffusion". Note in this particular case that the time-space discretization scheme and the specific form of the system $(E)$ permit to use a standard "dynamic programming principle".

From a numerical point of view, two other kinds of approaches have been developed in the backward case. The first one is based on Monte-Carlo simulations and Malliavin integration by parts, see Bouchard and Touzi [BT04]. The other one relies on quantization techniques for a discretization scheme of the underlying forward equation. Quantization consists in approximating a random variable by a suitable discrete law. It provides a cheap and numerically efficient alternative to usual Monte-Carlo methods to estimate expectations. In the works of Bally and Pagès [BP03] or Bally, Pagès and Printems [BPP02] on American options, the key idea is to perform an optimal quantization procedure of a discretized version of the underlying diffusion process in order to compute **once for all** by a Monte-Carlo method the corresponding semi-group. Then, the second step consists in doing a dynamic programming descent. For other applications of quantization, we refer to the works of Pagès, Pham and Printems, [PPP04] or Pagès and Printems, [PP04].

***Discretization Strategy.*** In the coupled case, or quasi-linear framework, the diffusion $U$ is not "objective" anymore. Indeed, due to the strong non-linearity of the equation $(\mathcal{E})$, the coefficients of the underlying forward diffusion depend on the solution and on its gradient.

In particular, we can not quantify a discretization scheme of the diffusion process as explained above. This is well understood: without approximating $u$, we do not have any *a priori* knowledge of the optimal shape of the associated grid. Hence, we just focus on the quantization of the Brownian increments appearing in the forward SDE and then choose to define the approximate diffusion on a sequence of truncated $d$-dimensional Cartesian grids. Note that the discretization procedure of $U$ is now coupled to the approximation procedure of $(u, \nabla_x u)$ (denoted in a generic way by $(\overline{u}, \overline{v})$) which is computed along the same sequence of grids. The time-space discretization scheme allows to define $(\overline{u}, \overline{v})$ and the approximations of the transitions of $U$ in order to recover a kind of "dynamic programming principle". Consider indeed a given regular time mesh

$(t_i = ih)_{i\in\{0,...,N\}}$ of $[0,T]$, $h$ being the step size. To every discretization time $t_i$, associate a spatial Cartesian grid $\mathcal{C}_i \equiv \{(x_k^i)_{k\in\mathcal{I}_i}\}$, $\mathcal{I}_i \subset \mathbb{N}^*$, such that $\forall i \in \{0,...,N-1\}$, $\mathcal{C}_i \subset \mathcal{C}_{i+1}$. Starting from $t_N = T$ for which the solution of $(\mathcal{E})$ and its gradient are known, the transition of $U$ from $t_i$ to $t_{i+1}$, $i \in \{0,...,N-1\}$, is then updated iteratively through the Brownian quantized increments and through the values of $\overline{u}(t_{i+1},.)$ and $\overline{v}(t_{i+1},.)$ on the grid $\mathcal{C}_{i+1}$. This permits to express the approximation $\overline{u}(t_i,.)$ through a discretized version of the Feynman-Kac formula.

At this stage, it remains to precise the way we update the approximation of the gradient of the solution $u$. We mention actually that the strategy aims to approximate the product $\nabla_x u(t_k,.)\sigma(.,u(t_k,.))$ instead of $\nabla_x u(t_k,.)$ itself. This explains the specific writing of the PDE $(\mathcal{E})$. We then proceed in two different steps. A first approximation is performed through a martingale increment procedure as done in the discretization scheme of BSDEs explained in Bouchard and Touzi [BT04], or as used in Bally et *al.* [BPP02]. A second step consists in quantizing the Gaussian increments appearing in the former representation. This is an alternative solution to the usual techniques based on Monte-Carlo simulations or on Malliavin integration by parts as employed in [BT04]. Of course, if the matrix $\sigma\sigma^*$ is non-degenerate, the strategy still applies, up to an inversion procedure, to coefficients of the form $(b,f)(x,u(t,x),\nabla_x u(t,x))$.

***Extra References.*** Some of the preliminaries of our approach can be found in Milstein and Tretyakov [MT99] in the specific case where $(b,f)(x,u(t,x),\nabla_x u(t,x)\sigma(t,x,u(t,x)))$ reduces to $(b,f)(x,u(t,x))$. Note however that the proof of the convergence of the underlying numerical scheme proposed in this reference just holds for so-called "equations with small parameter" (i.e. with a small diffusion matrix). Generally speaking, the authors have then to control the regularity properties of the solution of the transport problem associated to the equation $(\mathcal{E})$ (i.e. the same equation as $(\mathcal{E})$, but without any second order terms). Without discussing in detail the basic assumptions made in our paper, note that no condition of this type appears in the sequel: in particular, the matrix $a$ is assumed to be uniformly elliptic. Hence, we feel that the work of Milstein and Tretyakov [MT99] applies to a different framework than ours. For this reason, we avoid any further comparisons between both situations. Add finally for the sake of completeness that Makarov [Mak03] has successfully applied the strategy of Milstein and Tretyakov [MT99] to the case $(b,f) \equiv (b,f)(x,u(t,x),\nabla_x u(t,x)\sigma(t,x,u(t,x)))$ under suitable smoothness properties on the coefficients. Of course, the small parameter condition is then still necessary.

## 1.2 Novelties Brought by the Paper

***A purely probabilistic point of view.*** The proof of the convergence of our algorithm is somehow the first to be essentially of probabilistic nature, since we are able to adapt the usual stability techniques of BSDE theory to the discretized framework. Note in particular that we follow the proof of uniqueness in the *four step scheme* given in Ma, Protter and Yong [MPY94] to handle the strong coupling between the forward and backward components.

In the discretized framework, the gradient terms appearing in $b$ and $f$ bring additional difficulties. Indeed, our gradient approximation does not appear as a representation process given by the martingale representation theorem as the process $W$ in $(E)$. In particular, the strategy introduced by Pardoux and Peng [PP90] to estimate the $L^2$ norm of $W$ over $[0,T]$ fails in the discretized setting. We then propose a specific probabilistic strategy to overcome this deep trouble and thus to handle the nonlinearities of order one, see Subsections 3.3 and 9.3 for details

***Convergence under weak assumptions***. In Douglas, Ma and Protter [DMP96], the authors handle the gradient terms by working under smoothness assumptions that allow them to study the gradient of $u$ as the solution of the differentiated PDE.

Our strategy permits to avoid to differentiate the PDE and thus to really weaken the assumptions required both on the coefficients of $(E)$ and on the smoothness of the solution $u$ of $(\mathcal{E})$ in the above reference. In the previous paper, the coefficients are assumed to be smoothly differentiable and bounded. We just suppose that they are Lipschitz continuous and bounded in $x$. In Douglas and *al.*, the solution $u$ of $(\mathcal{E})$ is at least bounded in $C^{2+\alpha/2,4+\alpha}([0,T]\times\mathbb{R}^d)$, $\alpha \in ]0,1[$. In our paper, we only impose $u$ to belong to $C^{1,2}([0,T]\times\mathbb{R}^d)$ with bounded derivatives of order one in $t$ and one and two in $x$.

***A Completely Tractable Algorithm.*** Furthermore in [DMP96], the authors always take into consideration the case of infinite spatial grids. This turns out to be simpler for the convergence analysis, anyhow it does not provide in all generality a fully implementable algorithm. We discuss the impact of the truncation of the grids and analyze its contribution in the error.

Finally, a linear interpolation procedure is also used in Douglas et *al.* to define the algorithm. This can be heavy in large dimension. The algorithm we propose allows to define the approximate solution only at the nodes of the spatial grid. In this way, we feel that our method is simpler to implement and numerically cheaper. Note moreover that we avoid the inversion of large linear systems associated to "usual" numerical analysis techniques.

### 1.3 Organization of the Paper

In Section 2, we detail general assumption and notation as well as several smoothness properties of the solution $u$ of $(\mathcal{E})$. We also precise the connection between the FBSDE $(E)$ and the quasi-linear PDE $(\mathcal{E})$. Section 3 explains the main algorithmic choices. We present in particular the various steps that led us to the current discretization scheme. The main results are stated and discussed in Section 4. In particular, we give an estimate of the speed of convergence of the algorithm. As a probabilistic counterpart, we estimate the difference between the approximating processes and the initial solution $(U, V, W)$ of $(E)$. Numerical examples are presented in Section 5.

The end of the paper is then mainly devoted to the proof of the convergence results. The proof is divided into three parts. Various *a priori* controls of the discrete objects are stated and proved in Section 6. In Section 7, we adapt the FBSDE machinery to our setting to prove a suitable stability property. Section 8 is then devoted to the last step of the proof and more precisely to a specific refinement of Gronwall's Lemma.

As a conclusion, we compare in Section 9 our strategy to other methods and explain some technical points that motivated the choice of our current algorithm. We also indicate further conceivable extensions.

## 2 Non-Linear Feynman-Kac Formula

In this section, we first give the assumptions on the coefficients of the FBSDE and then briefly recall the connection with quasi-linear PDEs. As detailed later, under these assumptions, the underlying PDE admits a unique strong solution, whose partial derivatives of order one in $t$ and one and two in $x$ are controlled on the whole domain by known parameters. For the sake of simplicity, we also assume that the coefficients do not depend on time.

### 2.1 Coefficients of the Equation

For a given $d \in \mathbb{N}^*$, we consider the following coefficients:

$$b : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d,\ f : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R},\ \sigma : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d \times d},\ H : \mathbb{R}^d \to \mathbb{R}.$$

In the following, we denote by $|.|$ the Euclidean norm of $\mathbb{R}^n$, $n \geq 1$.

**Assumption (A)** We say that the functions $b$, $f$, $H$ and $\sigma$ satisfy Assumption **(A)** if there exist four constants $\alpha > 0$, $\kappa$, $\Lambda$ and $\lambda > 0$ such that:

**(A.1)** $\forall (x, y, z) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d,\ \big|(b, f, \sigma, H)(x, y, z)\big| \leq \Lambda\big(1 + |y| + |z|\big).$

**(A.2)** $\forall (x, y) \in \mathbb{R}^d \times \mathbb{R},\ \forall \zeta \in \mathbb{R}^d,\ \langle \zeta, a(x, y)\zeta\rangle \geq \lambda|\zeta|^2$, where $\forall (x, y) \in \mathbb{R}^d \times \mathbb{R},\ a(x, y) = \sigma\sigma^*(x, y).$

**(A.3)** $\forall (x, y, z),\ (x', y', z') \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$ :

$$\big|(b, f, \sigma, H)(x, y, z) - (b, f, \sigma, H)(x', y', z')\big| \leq \kappa\big(|x - x'| + |y - y'| + |z - z'|\big).$$

**(A.4)** The function $H$ belongs to $\mathcal{C}^{2+\alpha}(\mathbb{R}^d)$ and its $\mathcal{C}^{2+\alpha}$ norm is bounded by $\kappa$.

*From now on, Assumption* **(A)** *is in force.*

## 2.2 Forward-Backward SDE

Consider now a given $T > 0$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a Brownian motion $(B_t)_{0\le t\le T}$ whose natural filtration, augmented with $\mathbb{P}$ null sets, is denoted by $\{\mathcal{F}_t\}_{0\le t\le T}$.

Fix an initial condition $x_0 \in \mathbb{R}^d$ and recall (see Ma, Protter and Yong [MPY94] and Delarue [Del02]) that there exists a unique progressively measurable triple $(U, V, W)$, with values in $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$, such that:

$$\mathbb{E} \sup_{t\in[0,T]} \left(|U_t|^2 + |V_t|^2\right) < +\infty,\ \mathbb{E}\int_0^T |W_t|^2 dt < +\infty,$$

and which satisfies $\mathbb{P}$ almost surely the couple of equations $(E)$.

## 2.3 Quasi-Linear PDE

According to Ladyzhenskaya et *al.* [LSU67], Chapter VI, Theorem 4.1, and to [MPY94] (up to a regularization procedure of the coefficients), we claim that $(\mathcal{E})$ admits a solution $u \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ satisfying:

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

$$\begin{aligned}|u(t,x)| + |\nabla_x u(t,x)| + |\nabla^2_{x,x} u(t,x)| + |\partial_t u(t,x)| \\ + \sup_{t'\in[0,T], t\neq t'} \left[|t-t'|^{-1/2} |\nabla u(t,x) - \nabla u(t',x)|\right] \le C_{2.1}.\end{aligned}$$

*Moreover, $u$ is unique in the class of functions $\widetilde{u} \in \mathcal{C}([0,T] \times \mathbb{R}^d, \mathbb{R}) \cap \mathcal{C}^{1,2}([0,T[\times\mathbb{R}^d, \mathbb{R})$ which satisfy:*

$$\sup_{(t,x)\in[0,T[\times\mathbb{R}^d} \left(|\widetilde{u}(t,x)| + |\nabla_x \widetilde{u}(t,x)|\right) < +\infty.$$

From Ma, Protter and Yong [MPY94], Pardoux and Tang [PT99] and Delarue [Del02], the FBSDE $(E)$ is connected with the PDE $(\mathcal{E})$.

Set $\forall (t,x) \in [0,T[\times\mathbb{R}^d,\ v(t,x) = \nabla_x u(t,x)\sigma(x, u(t,x))$. The relationship between $(E)$ and $(\mathcal{E})$ can be summed up as follows:

$$\forall t \in [0,T],\ V_t = u(t,U_t),\ W_t = v(t,U_t),\ V_t = \mathbb{E}[V_T|\mathcal{F}_t] + \mathbb{E}\left[\int_t^T f(U_s,V_s,W_s)ds|\mathcal{F}_t\right]. \tag{2.1}$$

# 3 Approximation Procedure

In this section, we detail the construction of the approximation algorithm of the solution $u$ of $(\mathcal{E})$. We explain how the final form of the discretization procedure can be derived step by step from the forward-backward representation $(E)$. We also present the quantization techniques used in order to compute expectations related to Brownian increments and we discuss the choice of the underlying spatial grids which appear in the approximating scheme.

## 3.1 Rough Algorithms

***Localization Procedure.*** Recall from Introduction that the forward-backward equation $(E)$ appears as the starting point of our discretization procedure. Indeed, this couple of stochastic equations provides a probabilistic representation of the quasi-linear PDE $(\mathcal{E})$ and summarizes in an integral form the local evolution of the solution $u$. Define now, for a given integer $N \ge 1$, a regular mesh of $[0,T]$ with step $h \equiv T/N$:

$$t_0 \equiv 0,\ t_1 \equiv h = T/N,\ t_2 \equiv 2h, ..., t_N \equiv T.$$

Using $(E)$, for $k \in \{0, ..., N\}$, the evolution of the solution $u$ along the interval $[t_k, t_{k+1}]$ writes from a probabilistic point of view:

$$\begin{cases} U_{t_{k+1}} = U_{t_k} + \int_{t_k}^{t_{k+1}} b(U_s, V_s, W_s) ds + \int_{t_k}^{t_{k+1}} \sigma(U_s, V_s) dB_s, \\ V_{t_k} = V_{t_{k+1}} + \int_{t_k}^{t_{k+1}} f(U_s, V_s, W_s) ds - \int_{t_k}^{t_{k+1}} W_s dB_s, \end{cases} \tag{3.1}$$

where the initial condition $U_{t_k}$ is a square integrable and $\mathcal{F}_{t_k}$-measurable random vector with values in $\mathbb{R}^d$. In particular, conditioning by $U_{t_k} = x \in \mathbb{R}^d$, we deduce:

$$U_{t_{k+1}}^{t_k,x} = x + \int_{t_k}^{t_{k+1}} b\big(U_s^{t_k,x}, V_s^{t_k,x}, W_s^{t_k,x}\big) ds + \int_{t_k}^{t_{k+1}} \sigma\big(U_s^{t_k,x}, V_s^{t_k,x}\big) dB_s, \tag{3.2}$$

and

$$\begin{aligned} V_{t_k}^{t_k,x} &= \mathbb{E}\Big[V_{t_{k+1}}^{t_k,x} + \int_{t_k}^{t_{k+1}} f\big(U_s^{t_k,x}, V_s^{t_k,x}, W_s^{t_k,x}\big) ds\Big], \\ \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} W_s^{t_k,x} ds\Big] &= \mathbb{E}\big[V_{t_{k+1}}^{t_k,x}\big(B_{t_{k+1}} - B_{t_k}\big)\big] + O(h^{3/2}), \end{aligned}$$

where the superscript $(t_k, x)$ denotes the starting point of the diffusion process $U$. The remaining term $O(h^{3/2})$ is a consequence of Assumption **(A.1)**, (2.1) (relationships between $V, W$ and $u$) and Theorem 2.1 (boundedness of $u$ and $\nabla_x u$). Relation (2.1) also yields

$$\begin{aligned} u(t_k, x) &= \mathbb{E}\Big[u\big(t_{k+1}, U_{t_{k+1}}^{t_k,x}\big) + \int_{t_k}^{t_{k+1}} f\big(U_s^{t_k,x}, V_s^{t_k,x}, W_s^{t_k,x}\big) ds\Big], \\ \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} W_s^{t_k,x} ds\Big] &= \mathbb{E}\big[u\big(t_{k+1}, U_{t_{k+1}}^{t_k,x}\big)(B_{t_{k+1}} - B_{t_k})\big] + O(h^{3/2}). \end{aligned} \tag{3.3}$$

In the following, the Brownian increment $B_{t_{k+1}} - B_{t_k}$ is denoted by $\Delta B^k$. In particular, we derive from the above relation that, neglecting the rest, the best constant approximation of $(W_s)_{s\in[t_k,t_{k+1}]}$ in the $L^2([t_k, t_{k+1}] \times \Omega, ds \otimes d\mathbb{P})$ sense is given by:

$$\hat{W}_{t_k}^{t_k,x} \equiv h^{-1}\mathbb{E}[u(t_{k+1}, U_{t_{k+1}}^{t_k,x})\Delta B^k]. \tag{3.4}$$

Relationships (3.2), (3.3) and (3.4) provide a rough background to discretize the local evolution given in (3.1). However, this first form is not satisfactory from an algorithmic point of view. Indeed, because of the strong coupling between the forward and the backward equations, the transition of the diffusion depends on the solution itself, both in the drift term and in the martingale part. At the opposite, in the so-called "pure backward" case, or correspondingly for semi-linear equations, the underlying operator does not depend on the solution. In such a case, the classical Euler machinery applies to discretize the decoupled diffusion $U$.

***Induction Principle.*** Recall that similar difficulties occur to establish the unique solvability of the FBSDE $(E)$. In Delarue [Del02], the first author overcomes the strong coupling between the forward and backward equations by solving by induction the sequence of local FBSDEs (3.1), $k$ running downwards from $N$ to 0. By analogy with this approach, the discretization procedure of the forward component on a step $[t_k, t_{k+1}[$, $0 \le k \le N-1$, must take into account the issues of the former local discretizations of the backward equation, and more specifically the approximations of $u(t_{k+1}, .)$ and $v(t_{k+1}, .)$.

***Predictors.*** Assume to this end that, at time $t_{k+1}$, some approximations $\overline{u}(t_{k+1}, .), \overline{v}(t_{k+1}, .)$ of $u(t_{k+1}, .)$, $v(t_{k+1}, .)$ are available on the whole space. These approximations appear as the "natural" predictors of the true solution and of its gradient on $[t_k, t_{k+1}[$. Introducing the forward approximating transition

$$\mathcal{T}(t_k, x) \equiv b(x, \overline{u}(t_{k+1}, x), \overline{v}(t_{k+1}, x))h + \sigma(x, \overline{u}(t_{k+1}, x))\Delta B^k, \tag{3.5}$$

we derive an associated updating procedure by setting:

$$\begin{aligned} \overline{u}(t_k, x) &\equiv \mathbb{E}\big[\overline{u}\big(t_{k+1}, x + \mathcal{T}(t_k, x)\big)\big] + hf\big(x, \overline{u}(t_{k+1}, x), \overline{v}(t_{k+1}, x)\big), \\ \overline{v}(t_k, x) &\equiv h^{-1}\mathbb{E}\big[\overline{u}\big(t_{k+1}, x + \mathcal{T}(t_k, x)\big)\Delta B^k\big]. \end{aligned} \tag{3.6}$$

Once the predictors are updated, the procedure can be iterated. Of course, at time $T = t_N$ we set: $\overline{u}(t_N, .) \equiv H(.)$ and $\overline{v}(t_N, .) \equiv \nabla_x H(.)\sigma(., H(.))$. Note in particular that the expectations appearing in (3.6) are correctly defined. Indeed, a simple induction procedure shows from Assumptions **(A.1)** and **(A.4)** that $\overline{u}$ and $\overline{v}$ are bounded on $\{t_0, ..., t_N\} \times \mathbb{R}^d$ (but the bound depends on the discretization parameters).

***Spatial Discretization.*** In practice, it is anyhow impossible to define and to update $\overline{u}, \overline{v}$ on the whole space as done above. The most natural strategy consists in defining the approximations $\overline{u}(t_k, .)$ and $\overline{v}(t_k, .)$ of the true solution and its gradient on a discrete subset of $\mathbb{R}^d$. Those approximations could then be extended to the whole space with a linear interpolation procedure. However, in high dimension, this last operation can be computationally demanding. We thus prefer for simplicity to restrict the approximations to a given spatial grid $\mathcal{C}_k \equiv \{(x_j^k)_{j\in\mathcal{I}_k}, \mathcal{I}_k \subset \mathbb{N}^*\} \subset \mathbb{R}^d$, for $k \in \{0, ..., N\}$. This choice imposes to modify (3.6). Indeed, the "terminal" value $x + \mathcal{T}(t_k, x)$ must belong to the former grid $\mathcal{C}_{k+1}$.

Hence, denoting by $\Pi_{k+1}$ a projection mapping on the grid $\mathcal{C}_{k+1}$, we replace (3.6) by:

$$\begin{aligned} \forall x \in \mathcal{C}_k,\ \overline{u}(t_k, x) &\equiv \mathbb{E}\big[\overline{u}\big(t_{k+1}, \Pi_{k+1}(x + \mathcal{T}(t_k, x))\big)\big] + hf\big(x, \overline{u}(t_{k+1}, x), \overline{v}(t_{k+1}, x)\big), \\ \overline{v}(t_k, x) &\equiv h^{-1}\mathbb{E}\big[\overline{u}\big(t_{k+1}, \Pi_{k+1}(x + \mathcal{T}(t_k, x))\big)\Delta B^k\big]. \end{aligned} \tag{3.7}$$

In the following, we suppose that $\forall (i, j) \in \{0, ..., N\}^2,\ j < i \Rightarrow \mathcal{C}_j \subset \mathcal{C}_i$, so that $\overline{u}(t_{k+1}, x), \overline{v}(t_{k+1}, x)$ are well defined for $x \in \mathcal{C}_k$. Note that if the cardinal of $\mathcal{C}_k$ is finite for every $k$, the above scheme is already implementable up to the computations of the underlying expectations.

We then need to detail the way the Gaussian integrals appearing in (3.7) are computed and to precise the choice of the grids. This is done in Subsections 3.2 and 3.4.

***Global Updating.*** Our updating method using the predictors $\overline{u}(t_{k+1}, .), \overline{v}(t_{k+1}, .)$ is an alternative to the standard fixed point procedure. This latter consists in giving first some global predictors $\overline{u}^0(t_k, .), \overline{v}^0(t_k, .),\ k \in \{0, ..., N\}$. These are used to compute the transitions of the approximating forward process. In this way, we obtain a decoupled forward-backward system, whose solution may be computed by a standard dynamic programming algorithm. A complete descent of this algorithm from $k = N$ to $k = 0$ produces $\overline{u}^1(t_k, .), \overline{v}^1(t_k, .),\ k \in \{0, ..., N\}$, from which we can iterate the previous procedure. In this frame, the underlying distance used to describe the convergence of the fixed point procedure involves all the discretization times and all the spatial points. This strategy appears as a "global updating" one.

From a numerical point of view this seems unrealistic. Indeed, one would need to solve a large number of linear problems. This would either require to use massive Monte-Carlo simulations at each step of the algorithm or to apply, again at each step of the algorithm, a quantization procedure of the approximate diffusion process associated to the current linear problem. Furthermore, it seems intuitively clear that a local updating is far more efficient than a global one.

## 3.2 Quantization

***Expectations Approximation.*** Two methods are conceivable to compute expectations appearing in (3.7).

The first one consists in applying the classical Monte-Carlo procedure for every $k \in \{0, ..., N-1\}$ and for every $x \in \mathcal{C}_k$, and therefore to repeat this argument $\sum_{k=0}^{N-1} |\mathcal{I}_k|$ times. From the central limit Theorem, such a strategy would lead to perform $\sum_{k=0}^{N-1} |\mathcal{I}_k| \times \varepsilon_{\mathrm{MC}}^{-2}$ elementary operations to compute underlying expectations up to the error term $\varepsilon_{\mathrm{MC}}$. This approach seems rather hopeless.

A more efficient method consists in replacing the Gaussian variables appearing in (3.7) by discrete ones with known weights. This procedure is known as "quantization". Consider to this end a probability measure on $\mathbb{R}^d$ with finite support $(y_i)_{i\in\{1,\ldots,M\}}$ and denote by $(p_i)_{i\in\{1,\ldots,M\}}$ the associated weights. Replace then the Gaussian increments in (3.7) by this law. For a given $x \in \mathcal{C}_k$, $0 \leq k \leq N$, the expectations appearing in the induction scheme (3.7) then write:

$$\overline{u}(t_k,x) \equiv \sum_{i=1}^{M}\Big[p_i\overline{u}\big(t_{k+1},\Pi_{k+1}(x+b(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x))h+\sigma(x,\overline{u}(t_{k+1},x))y_i)\big)\Big]$$
$$+ hf(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x)),$$
$$\overline{v}(t,x) \equiv h^{-1}\sum_{i=1}^{M}\Big[p_i\overline{u}\big(t_{k+1},\Pi_{k+1}(x+b(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x))h+\sigma(x,\overline{u}(t_{k+1},x))y_i)\big)y_i\Big],$$

and are explicitly computable.

***Quantization Principle.*** We briefly recall the basic principle of quantization and refer to the monograph of Graf and Lushgy, cf. [GL00], for details. Generally speaking, for a given random variable $\Delta$, the quantization procedure consists in replacing $\Delta$ by its projection on a finite grid $\Lambda(M) \equiv \{(y_i)_{i\in\{1,\ldots,M\}}\} \subset \mathbb{R}^d,\ M \in \mathbb{N}^*$. The projection mapping $G_{\Lambda(M)}$ simply writes:

$$G_{\Lambda(M)}(y) = \sum_{i=1}^{M} y_i \mathbf{1}_{V_i}(y),$$

where, for every $i \in \{1,...,M\}$, $V_i \equiv \{y \in \mathbb{R}^d,\ |y-y_i| = \min_{j\in\{1,\ldots,M\}}|y-y_j|\}$. In quantization theory, $V_i$ is known as the Voronoi tessel of $y_i$.

Define now the quantization of $\Delta$ with respect to the grid $\Lambda(M)$ by: $\hat{\Delta} \equiv G_{\Lambda(M)}(\Delta)$. The weights of $\hat{\Delta}$ then write: $p_i \equiv \mathbb{P}\{\Delta \in V_i\}$, $i \in \{1,...,M\}$. For a given grid, one can compute these values once for all, using e.g. a Monte-Carlo method, so that we may assume to have the "exact" values of these weights.

The crucial step in the quantization procedure therefore lies in the choice of the grid $\Lambda(M)$. To this end, we introduce the so-called "distortion" in order to measure the error associated to the grid $\Lambda(M)$:

$$D_{\Delta,p}\big(\Lambda(M)\big) \equiv \|\Delta - \hat{\Delta}\|_{L^p(\mathbb{P})},\ p \geq 1. \tag{3.8}$$

***Optimal Grids.*** The Bucklew-Wise Theorem, see Theorem 6.2 Chapter II in [GL00] for details, then gives for $\Lambda^*(M)$ achieving the minimum in (3.8):

$$M^{p/d}D^p_{\Delta,p}\big(\Lambda^*(M)\big) \longrightarrow C(p,d) \text{ as } M \to +\infty, \tag{3.9}$$

where $C(p,d)$ is a constant depending on $p,d$ and the variable at hand.

Various algorithms are available to compute an optimal grid $\Lambda^*(M)$. In dimension 1, one may use Lloyd's algorithm, which is deterministic. For $d > 1$, one usually uses the Kohenen algorithm which is a stochastic one, see Bally and Pagès [BP03]. We also recall that, for $d > 1$, the optimal grid is not unique.

Up to a rescaling, the basic object associated to Brownian increments is a $d$-dimensional standard normal random variable. Hence, we assume in the following that an optimal grid $\Lambda^*(M)$ for $\Delta \sim \mathcal{N}(0,\mathbf{I}_d)$ as well as the associated weights $(p_i)_{i\in\{1,\ldots,M\}}$ are given and "perfectly" computed. Let us remark that the Feynman-Kac formula gives an analytical interpretation of $p_i$. Indeed, one has $\forall i \in \{1,...,M\},\ p_i = u_i(0,0)$, where $u_i$ is the solution of the backward heat equation

$$\partial_t u_i(t,x) + \frac{1}{2}\Delta u_i(t,x) = 0, (t,x) \in [0,1[\times\mathbb{R}^d, u_i(1,.) = \mathbf{1}_{V_i}(.).$$

***Quantized Algorithm.*** We are now in position to introduce a more tractable induction principle. Set to this end, for every $k \in \{0,...,N-1\},\ g(\Delta B^k) \equiv h^{1/2}G_{\Lambda^*(M)}(h^{-1/2}\Delta B^k)$. Note in particular from (3.9) that, for every $p \geq 1$, there exists a constant $C_{\text{Quantiz}}(p,d)$ such that:

$$\mathbb{E}\big[|g(\Delta B^k) - \Delta B^k|^p\big]^{1/p} \leq C_{\text{Quantiz}}(p,d)h^{1/2}M^{-1/d}. \tag{3.10}$$

Turn now (3.5) and (3.7) into:

$$\forall x \in \mathcal{C}_k,\ \mathcal{T}(t_k,x) \equiv b(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x))h + \sigma(x,\overline{u}(t_{k+1},x))g(\Delta B^k), \tag{3.11}$$

and,

$$\begin{aligned}\forall x \in \mathcal{C}_k,\ \overline{u}(t_k,x) &\equiv \mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x))\big)\big] + hf\big(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x)\big),\\ \overline{v}(t_k,x) &\equiv h^{-1}\mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x))\big)g(\Delta B^k)\big].\end{aligned} \tag{3.12}$$

To sum up our strategy, the use of predictors allows to recover a kind of standard dynamic programming principle. The quantization gives an easy, cheap and computable algorithm.

### 3.3 Algorithms

The algorithm associated to (3.11) and to (3.12) takes the following form:

**Algorithm 3.1** *For a given sequence of spatial grids $(\mathcal{C}_k)_{0\le k\le N}$, we put:*

$$\begin{aligned}&\forall x \in \mathcal{C}_N,\ \overline{u}(T,x) \equiv H(x),\ \overline{v}(T,x) \equiv \nabla_x H(x)\sigma(x,H(x)),\\ &\forall k \in \{0,...,N-1\},\ \forall x \in \mathcal{C}_k,\\ &\quad \mathcal{T}(t_k,x) \equiv b(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x))h + \sigma(x,\overline{u}(t_{k+1},x))g(\Delta B^k),\\ &\quad \overline{u}(t_k,x) \equiv \mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x))\big)\big] + hf(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x)),\\ &\quad \overline{v}(t_k,x) \equiv h^{-1}\mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x))\big)g(\Delta B^k)\big].\end{aligned}$$

Up to the choice of the underlying grids, this form follows heuristics given in Subsection 3.1: the coefficients of the local transition $\mathcal{T}(t_k,x)$ are expressed in function of the former approximations $\overline{u}(t_{k+1},.)$ and $\overline{v}(t_{k+1},.)$.

For technical reasons detailed in Section 9 we consider for the convergence analysis a slightly different version of the above algorithm. Namely, we need to change, at a given time $t_k$, the discretization of $b$ and $f$ and in particular to replace $\overline{v}(t_{k+1},.)$ by a new predictor. Concerning the driver of the BSDE, we replace $f(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x))$ by $f(x,\overline{u}(t_{k+1},x),\overline{v}(t_k,x))$: the definition of $\overline{v}(t_k,x)$ does not involve $\overline{u}(t_k,x)$.

The story is rather different for $b$. Indeed, the definition of $\overline{v}(t_k,x)$ relies on the choice of the underlying transition. In particular, putting $\overline{v}(t_k,x)$ in $b$ as done in $f$ would lead to an implicit scheme.

Nevertheless, for a given intermediate predictor $\hat{v}(t_k,.)$ of $v(t_k,.)$, we can put:

$$\mathcal{T}(t_k,x) \equiv b\big(x,\overline{u}(t_{k+1},x),\hat{v}(t_k,x)\big)h + \sigma\big(x,\overline{u}(t_{k+1},x)\big)g(\Delta B^k).$$

The whole difficulty is then hidden in the choice of $\hat{v}(t_k,x)$. Our strategy consists in choosing $\hat{v}(t_k,x)$ as the expectation of $\overline{v}(t_{k+1},.)$ with respect to the transition $\mathcal{T}^0(t_k,x) \equiv \sigma(x,\overline{u}(t_{k+1},x))g(\Delta B^k)$. This transition differs from $\mathcal{T}(t_k,x)$ in the drift $b$ and leads to an explicit scheme. Namely, we set:

$$\hat{v}(t_k,x) \equiv \mathbb{E}\big[\overline{v}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}^0(t_k,x))\big)\big]. \tag{3.13}$$

The predictor $\hat{v}(t_k,.)$ in (3.13) appears as a "regularized" version of $\overline{v}(t_{k+1},.)$. Thanks to a Gaussian change of variable, the laws of the underlying transitions $\mathcal{T}^0(t_k,x)$ and $\mathcal{T}(t_k,x)$ are compared in Subsection 7.3.

***Final Algorithm.***

**Algorithm 3.2** *The final algorithm writes:*

$$\begin{aligned}&\forall x \in \mathcal{C}_N,\ \overline{u}(T,x) \equiv H(x),\ \overline{v}(T,x) \equiv \nabla_x H(x)\sigma(x,H(x)),\\ &\forall k \in \{0,...,N-1\},\ \forall x \in \mathcal{C}_k,\\ &\quad \mathcal{T}^0(t_k,x) \equiv \sigma\big(x,\overline{u}(t_{k+1},x)\big)g(\Delta B^k)\\ &\quad \hat{v}(t_k,x) \equiv \mathbb{E}\big[\overline{v}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}^0(t_k,x))\big)\big],\\ &\quad \mathcal{T}(t_k,x) \equiv b\big(x,\overline{u}(t_{k+1},x),\hat{v}(t_k,x)\big)h + \sigma\big(x,\overline{u}(t_{k+1},x)\big)g(\Delta B^k),\\ &\quad \overline{v}(t_k,x) \equiv h^{-1}\mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x))\big)g(\Delta B^k)\big],\\ &\quad \overline{u}(t_k,x) \equiv \mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x))\big)\big] + f\big(x,\overline{u}(t_{k+1},x),\overline{v}(t_k,x)\big)h.\end{aligned}$$

***A Discrete Probabilistic Representation.*** Note that Algorithm 3.2 provides a discretization procedure of the FBSDE $(E)$ just like $(\mathcal{E})$ provides an analytical counterpart to the stochastic system of equations $(E)$. Consider to this end an initial condition $x_0 \in \mathcal{C}_0$ and define a Markov process on the grids $(\mathcal{C}_k)_{0 \le k \le N}$ according to the transitions $(\mathcal{T}(t_k, x))_{k \in \{0,...,N-1\}, x \in \mathcal{C}_k}$:

$$X_0 \equiv x_0, \ \forall k \in \{0, ..., N-1\}, \ X_{t_{k+1}} \equiv \Pi_{k+1}\big(X_{t_k} + \mathcal{T}(t_k, X_{t_k})\big). \tag{3.14}$$

Referring to the connection between $U$ and $(V, W)$, see e.g. (2.1), put now:

$$\forall k \in \{0, ..., N\}, \ Y_{t_k} \equiv \overline{u}(t_k, X_{t_k}), \ Z_{t_k} \equiv \overline{v}(t_k, X_{t_k}). \tag{3.15}$$

Note that $Y$ and $Z$ are correctly defined since $X_{t_k}$ belongs to the grid $\mathcal{C}_k$. The couple $(Y, Z)$ appears as a discrete version of the couple $(V, W)$ in $(E)$. More precisely, one can prove the following discrete Feynman-Kac formula (see Proposition 6.1 for a precise statement):

$$\forall 0 \le k \le N-1, \ Y_{t_k} = \mathbb{E}\Big[H(X_{t_N}) + h \sum_{i=k+1}^{N} f\big(X_{t_{i-1}}, \overline{u}(t_i, X_{t_{i-1}}), Z_{t_{i-1}}\big) | \mathcal{F}_{t_k}\Big]. \tag{3.16}$$

Note anyhow that the process $Z$ does not appear as the martingale part of the process $Y$. However, thanks to the martingale representation theorem, there exists a progressively measurable process $\overline{Z}$, with finite moment of order two, such that:

$$Y_{t_N} + h \sum_{i=1}^{N} f\big(X_{t_{i-1}}, \overline{u}(t_i, X_{t_{i-1}}), Z_{t_{i-1}}\big) = Y_0 + \int_0^{t_N} \overline{Z}_s dB_s. \tag{3.17}$$

Of course, the process $\overline{Z}$ does not match exactly the process $Z$. However, for a given $k \in \{0, ..., N-1\}$, it is readily seen from the above expression that the best $\mathcal{F}_{t_k}$-measurable approximation of $(\overline{Z}_s)_{s \in [t_k, t_{k+1}]}$ in $L^2([t_k, t_{k+1}] \times \Omega, ds \otimes d\mathbb{P})$ is given by:

$$h^{-1}\mathbb{E}\big[Y_{t_{k+1}} \Delta B^k\big].$$

Up to the quantization procedure, this term coincides with $\overline{v}(t_k, X_{t_k})$. In other words, the processes $Z$ and $\overline{Z}$ may be considered as close.

## 3.4 Choice of the Grids

As indicated at the end of Subsection 3.1, it remains to precise the choice of the grids. Because of the strong coupling, little is *a priori* known on the behaviour of the paths of the forward process. Hence, we can not compute a kind of optimal grid for $X$. The most natural choice turns out to be the one of Cartesian grids.

***Unbounded Cartesian Grids.*** Two different choices of grids are conceivable. First, we can treat the case of infinite Cartesian grids:

$$\forall k \in \{0, ..., N\}, \ \mathcal{C}_k \equiv \mathcal{C}_\infty, \ \mathcal{C}_\infty \equiv \delta \mathbb{Z}^d, \tag{3.18}$$

where $\delta > 0$ denotes a spatial discretization parameter. In this case, the projection mapping takes the following form:

$$\forall x \in \mathbb{R}^d, \ \Pi_\infty(x) \equiv \sum_{y \in \mathcal{C}_\infty} \Big[y \prod_{j=1}^{d} \mathbf{1}_{[-\delta/2, \delta/2[}(x_j - y_j)\Big]. \tag{3.19}$$

In other words, for every $j \in \{1, ..., d\}$, the coordinate $j$ of $\Pi_\infty(x)$ is given by $(\Pi_\infty(x))_j = \delta \lfloor \delta^{-1} x_j + 1/2 \rfloor$.

This choice actually simplifies the convergence analysis and allows a direct comparison with the results from the existing literature, see Douglas et *al.* [DMP96]. Note however that it does not provide a fully implementable scheme since the set $\mathcal{C}_\infty$ is infinite.

***Truncated Grids.*** We now discuss the case of truncated grids. Actually, several truncation procedures may be considered, but all need to take into account the specific geometry of a non-degenerate diffusion, or more simply, of the Brownian motion. Set for example, for a given $R > 0$, $\mathcal{C}_0 \equiv \mathcal{C}_\infty \cap \Delta_0$, where:

$$\Delta_0 \equiv \big\{x \in \mathbb{R}^d, \ \forall 1 \le j \le d, \ -\delta \lfloor R\delta^{-1} \rfloor - \delta/2 \le x_j < \delta \lfloor R\delta^{-1} \rfloor + \delta/2\big\}. \tag{3.20}$$

The particular choice of the bounds in the definition of $\Delta_0$ ensures that for all $x \in \mathbb{R}^d$, $\Pi_\infty(x) \in \mathcal{C}_0 \Leftrightarrow x \in \Delta_0$.

Due to the drift part and to the diffusive part of the forward process, it is clear that we need to enlarge the spatial grids as time increases. To this end, fix $\rho > 0$ and define, for every $i \in \{1, ..., N\}$, the truncated grid $\mathcal{C}_i \equiv \mathcal{C}_\infty \cap \Delta_i$, where:

$$\Delta_i \equiv \{x \in \mathbb{R}^d,\ \forall 1 \le j \le d,\ -\delta\lfloor (R+\rho)\delta^{-1}\rfloor - \delta/2 \le x_j < \delta\lfloor (R+\rho)\delta^{-1}\rfloor + \delta/2\}. \tag{3.21}$$

Note in this way that the size of the grid $\mathcal{C}_i$, i.e. of the grid at time $t_i$, does not depend on $t_i$ itself. In other words, the Hölder regularity of the paths of the Brownian motion does not interact with the definition of the grids. To take into account these pathwise properties, the following type of grids could also be used:

$$\Delta_i \equiv \{x \in \mathbb{R}^d,\ \forall 1 \le j \le d,\ -\delta\lfloor (R+\rho t_i^\alpha)\delta^{-1}\rfloor - \delta/2 \le x_j < \delta\lfloor (R+\rho t_i^\alpha)\delta^{-1}\rfloor + \delta/2\},$$

for a given $0 < \alpha < 1/2$. With this definition, the number of points involved in the discretization procedure is smaller. However, since the proof of the convergence of the algorithm is far from being trivial, we prefer, for the sake of simplicity, to keep the first definition of the truncated grids. Hence, for every $i \in \{0, ..., N\}$, $\Pi_i$ writes:

$$\begin{aligned} \forall 0 \le i \le N, \forall x \in \Delta_i,\ \Pi_i(x) &\equiv \mathcal{Q}\big(R+\rho, \Pi_\infty(x)\big) \equiv \Pi_\infty(x), \\ \forall 1 \le i \le N, \forall x \notin \Delta_i,\ \Pi_i(x) &\equiv \mathcal{Q}\big(R+\rho, \Pi_\infty(x)\big), \\ \forall x \notin \Delta_0,\ \Pi_0(x) &\equiv \mathcal{Q}\big(R, \Pi_\infty(x)\big), \end{aligned} \tag{3.22}$$

where for a given $(r, y) \in \mathbb{R}^{+*} \times \mathbb{R}^d$, $\mathcal{Q}(r, y)$ denotes the orthogonal projection of $y$ on the hypercube $[-\delta\lfloor r\delta^{-1}\rfloor, \delta\lfloor r\delta^{-1}\rfloor]^d$:

$$\mathcal{Q}(r, y) \equiv \big((y_i \vee (-\delta\lfloor r\delta^{-1}\rfloor)) \wedge (\delta\lfloor r\delta^{-1}\rfloor)\big)_{1 \le i \le d}.$$

Note finally that $R$ is fixed by the reader once for all in function of the set on which $u$ has to be approximated at the initial time. At the opposite, $\rho$ appears as a discretization parameter chosen by the reader in function of the required precision and of the affordable complexity for Algorithm 3.2.

# 4 Convergence Results

This section is devoted to the convergence analysis of $\overline{u}$ to $u$. As stated in the following theorem, which is the main result of the paper, five different types of errors can be distinguished:

**Theorem 4.1** *Let $p \ge 2$. There exist two constants $c_{4.1}$ and $C_{4.1}$, only depending on $p$, $T$ and on known parameters appearing in* **(A)***, such that for $h < c_{4.1}$, $\delta^2 < h$, $M^{-2/d} < h$ and $\rho \ge 1$:*

$$\sup_{x \in \mathcal{C}_0} |u(0,x) - \overline{u}(0,x)|^2 \le C_{4.1}\big[\mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{trunc}) + \mathcal{E}^2(\text{quantiz}) + \mathcal{E}^2(\text{gradient}, p)\big],$$

*with:*

$$\begin{aligned} &\mathcal{E}(\text{time}) \equiv h^{1/2},\ \mathcal{E}(\text{space}) \equiv h^{-1}\delta,\ \mathcal{E}(\text{trunc}) \equiv R/(R+\rho),\ \mathcal{E}(\text{quantiz}) \equiv h^{-1/2}M^{-1/d}, \\ &\mathcal{E}(\text{gradient}, p) \equiv h^{p/2+d/4-1/2}M^{-p/d}\delta^{-p-d/2} \end{aligned}$$

**Remark 4.1** *The FBSDE counterpart of Theorem 4.1 is given in Subsection 4.3: see Theorems 4.2 and 4.3. Since additional notations are needed to detail these results in a relevant form, we prefer, for the sake of brightness, to postpone their statement (and thus the definition of the associated probabilistic tools) to a different part.*

## 4.1 Classification of Errors

We now detail the meaning of the different errors appearing in Theorem 4.1:

***Temporal Discretization Error.*** $\mathcal{E}(\text{time})$ denotes the temporal discretization error due to the discretization of $[0, T]$ along a regular mesh. In a nutshell, the $1/2$ exponent appearing in the definition of $\mathcal{E}(\text{time})$ corresponds to the Hölder regularity of $u$ and $\nabla_x u$ in time and to the $L^2(\mathbb{P})$ 1/2-Hölder property of the Brownian increments.

***Spatial Discretization Error.*** $\mathcal{E}(\text{space})$ denotes the spatial discretization error. This quantity highly depends on the ratio between the spatial and the temporal steps. This connection between $\delta$ and $h$ can be explained as follows: the drift part of the transitions $(\mathcal{T}(t_k,.))_{0\le k\le N}$ is of order $h$ and the diffusive one is of order $h^{1/2}$. Thus, to take into account the influence of the drift at the local level, the spatial discretization parameter must be smaller than $h$. In other words, $\delta h^{-1}$ must be small.

***Quantization Error.*** $\mathcal{E}(\text{quantiz})$ denotes the error due to the quantization procedure of the Brownian increments. This error depends on the ratio between the distortion and the temporal step. The quantity $\mathcal{E}(\text{quantiz})$ represents the typical bound between $\overline{v}(t_k,x)$ and the best constant approximation of the process $(\overline{Z}_s)_{s\in[t_k,t_{k+1}]}$, i.e. between $\overline{v}(t_k,x)$ and:

$$h^{-1}\mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x))\big)\Delta B^k\big].$$

Note indeed that the distance between $\Delta B^k$ and $g(\Delta B^k)$ is of order $h^{1/2}M^{-1/d}$, see (3.10). Since the underlying expectation is divided by $h^{-1}$, this leads to a term in $h^{-1/2}M^{-1/d}$.

***Truncation Error.*** $\mathcal{E}(\text{trunc})$ denotes the error associated to the truncation procedure. As written in Theorem 4.1, it depends on $R$ and $\rho$, where $R$ denotes the radius of the initial grid $\mathcal{C}_0$ and $R+\rho$ the radius of the grids $(\mathcal{C}_k)_{1\le k\le N}$. If $\rho$ tends to $+\infty$, i.e. if the grids are not truncated, this error term reduces to zero.

Generally speaking, $\mathcal{E}(\text{trunc})$ appears as the Bienaymé-Chebychev estimate of the probability that the approximating process $X$ stays inside the grids $(\mathcal{C}_k)_{0\le k\le N}$. The lack of relevant estimates of the discretized version of the drift $b$ (recall that the function $b$ is not bounded), and more specially of the discretized gradient $\overline{v}$ explains the reason why the Bienaymé-Chebychev estimate applies in this framework and not better ones (as Bernstein inequality). We also recall that the unboundedness of the coefficients is the most common case in the applications, see e.g. Burgers equation in Section 5.

***Gradient Error.*** $\mathcal{E}(\text{gradient},p)$ denotes an extra error generated by the lack of estimates of the discretized gradient $\overline{v}$. This term follows from the specific choice of the predictor $\hat{v}$ made in Subsection 3.3 and appears in the second step of the proof of Theorem 4.1, see more precisely Subsections 7.1 and 7.3.

The convergence of $\mathcal{E}(\text{gradient},p)$ towards 0 relies on the term $h^{p/2}M^{-p/d}\delta^{-p}$, $M$ being chosen large enough and $p$ as large as necessary. In short, this reduced form represents the probability that the distance between the Gaussian increment and its quantization exceeds the spatial step $\delta$. Note indeed from (3.10) that for every $p\ge 2$:

$$\mathbb{P}\big\{|\Delta B^k - g(\Delta B^k)| > \delta\big\} \le C_{\text{Quantiz}}(p,d)h^{p/2}M^{-p/d}\delta^{-p}. \tag{4.1}$$

Thus, the error term $\mathcal{E}(\text{gradient},p)$ depends on the ratio between the spatial discretization step and the quantization distortion of the underlying Gaussian increments.

The probability (4.1) appears in the control of the distance between the predictor $\hat{v}$ and the true gradient $v$. In this frame, the strategy consists in writing the predictor $\hat{v}$ as an expectation with respect to the Gaussian kernel and not to its quantized version. Generally speaking, this strategy holds when the quantized transition $\mathcal{T}(t_k,x)$ and its Gaussian counterpart belong to the same cell of the spatial grid, i.e. when the distance between the Brownian increment and the quantized one is of the same order as the length of a given cell. Since the spatial grid step is given by $\delta$, we then need to control the probability that the difference between the increments exceeds $\delta$.

Of course, when $b$ does not depend on $z$, there is no reason to define $\hat{v}$. In such a case, $\mathcal{E}(\text{gradient},p)$ reduces to 0.

## 4.2 Comments on the Rate of Convergence

***Error in function of*** $h$***.*** To detail in a more explicit way the rate of convergence given by Theorem 4.1, we give an example in which $\rho$ ($\rho<+\infty$), $\delta$ and $M$ are expressed as powers of $h$. Assume indeed that $\rho$, $\delta$ and $M$ are chosen in the following way:

$$\rho = Rh^{-1/2},\ \ \delta \equiv h^{1+\alpha},\ \ M^{-2/d} \equiv h^{1+\beta}, \alpha,\beta \ge 0.$$

In such a case:

$$\mathcal{E}(\text{gradient}, p) = h^{p/2+d/4-1/2} M^{-p/d} \delta^{-p-d/2} = \exp\big[\ln(h)\big[p(\beta/2-\alpha) - (d/2+1+\alpha d)/2\big]\big].$$

To ensure the convergence of the algorithm we then need to choose:

$$p(\beta/2-\alpha) - (d/2+1+\alpha d)/2 > 0 \Longleftrightarrow \beta > 2\alpha + (1/p)(d/2+1+\alpha d).$$

Put finally $\beta = 2\alpha + (d/2+1+\alpha d)/p + \eta$, $\eta > 0$. The rate of convergence of the fully implementable algorithm is given by:

$$\sup_{x \in \mathcal{C}_0} |u(0,x) - \overline{u}(0,x)|^2 \leq C_{4.1}\big[h + h^{2\alpha} + h^{\beta} + h^{p\eta}\big].$$

Taking $\alpha = 1/2$ and $\eta = 1/p$ then yields:

$$\sup_{x \in \mathcal{C}_0} |u(0,x) - \overline{u}(0,x)|^2 \leq C_{4.1} h.$$

In particular, for $p$ large enough, the exponent $\beta$ is close to 1 and the number $M$ of points needed to quantify the Brownian increments is close to $h^{-d}$. Here is the limit of the method: for a large $d$ and a small $h$, we need a rather large number of points for the Gaussian quantization. Recall anyhow that the Gaussian grids are computed once for all. Thus, the numerical effort to get sharp quantization grids can be made apart from our Algorithm.

***Estimates of*** $\nabla_x u$. The reader might wonder about the estimate of the gradient of $u$. Note in this framework that two strategies are conceivable.

First, the probabilistic counterpart of Theorem 4.1 given in Subsection 4.3 provides an $L^2$ estimate of the distance between $\overline{v}$ and the gradient of the true solution. Note however that the underlying $L^2$ norm is taken with respect to the distribution of the discrete process $X$ (cf. (3.14)).

To get a joint estimate of the solution and of its gradient with respect to the *supremum* norm, the reader can apply the following strategy: differentiate if possible the PDE ($\mathcal{E}$) and apply, once again if possible, Algorithm 3.2 to $(u, \nabla_x u)$ seen as the solution of a system of parabolic quasi-linear PDEs. Such a strategy is applied in Section 5 to the solution of the porous media equation and to its gradient. Note that this approach coincides with the one followed by Douglas et *al.* [DMP96].

### 4.3 Estimates of the Discrete Processes

We now translate Theorem 4.1 in a more probabilistic way. Recall indeed that, in several situations (e.g. in financial mathematics), the knowledge of the triple $(U, V, W)$ is as crucial as the knowledge of the couple $(u, \nabla_x u)$.

Recall from (3.14), (3.15) and (3.16) that the discrete process $(X, Y, Z)$ given by:

$$\begin{aligned}
&X_0 \equiv x_0,\ \forall k \in \{0, ..., N-1\},\ X_{t_{k+1}} \equiv \Pi_{k+1}\big(X_{t_k} + \mathcal{T}(t_k, X_{t_k})\big),\\
&\forall k \in \{0, ..., N\},\ Y_{t_k} \equiv \overline{u}(t_k, X_{t_k}),\ Z_{t_k} \equiv \overline{v}(t_k, X_{t_k}),
\end{aligned}$$

provides a discretization of $(U, V, W)$.

We then prove that $(X, Y, Z)$ and $(U, V, W)$ get closer in a suitable sense as $h$, $\delta$, $M^{-1}$ and $\rho^{-1}$ vanish. Note however that we are not able to prove that the distance between $(X, Y, Z)$ and $(U, V, W)$ over the whole interval $[0, T]$ tends to zero. Indeed, since the projections $(\Pi_i)_{0 \leq i \leq N}$ map every point outside the sets $(\Delta_i)_{0 \leq i \leq N}$ onto the boundaries of $(\mathcal{C}_i)_{0 \leq i \leq N}$ (see e.g. (3.22)), we do not control efficiently the transition of the process $X$ after the first hitting time of the boundaries of the grids by $X$. It is then well understood that we have to stop the triple $(X, Y, Z)$ at this first hitting time. Put to this end:

$$\tau_\infty \equiv \inf\big\{t_k,\ 1 \leq k \leq N,\ X_{t_{k-1}} + \mathcal{T}(t_{k-1}, X_{t_{k-1}}) \notin \Delta_k\big\},\ \inf(\emptyset) = +\infty. \tag{4.2}$$

First, as a bypass product of the proof of Theorem 4.1, the function $\overline{v}$ provides an approximation of $v$ in the following $L^2$ sense:

**Theorem 4.2** *Let $p \geq 2$. Then, there exist two constants $c_{4.2}$ and $C_{4.2}$, only depending on $p$ and on known parameters appearing in* **(A)**, *such that for $h < c_{4.2}$, $\delta^2 < h$, $M^{-2/d} < h$ and $\rho \geq 1$:*

$$\begin{aligned} &h \sum_{i=0}^{N-1} \mathbb{E}\big[|\overline{v}(t_i, X_{t_i})\mathbf{1}_{\{t_i<\tau_\infty\}} - v(t_i, X_{t_i})|^2\big] \\ &\qquad \leq C_{4.2}\big[\mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz}) + \mathcal{E}^2(\text{trunc}) + \mathcal{E}^2(\text{gradient}, p)\big]. \end{aligned}$$

Moreover, the triple $(X, Y, Z)$ stopped at time $\tau_\infty$ satisfies:

**Theorem 4.3** *Let $p \geq 2$. Then, there exist two constants $c_{4.3}$ and $C_{4.3}$, only depending on $p$ and on known parameters appearing in* **(A)**, *such that for $h < c_{4.3}$, $\delta^2 < h$, $M^{-2/d} < h$ and $\rho \geq 1$:*

$$\begin{aligned} &\mathbb{E}\big[\sup_{i\in\{0,\dots,N\}} |X_{t_i\wedge\tau_\infty} - U_{t_i}|^2\big] + \mathbb{E}\big[\sup_{i\in\{0,\dots,N\}} |Y_{t_i\wedge\tau_\infty} - V_{t_i}|^2\big] + h\sum_{i=0}^{N-1} \mathbb{E}\big[|Z_{t_i}\mathbf{1}_{\{t_i<\tau_\infty\}} - W_{t_i}|^2\big] \\ &\qquad \leq C_{4.3}\big[\mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz}) + \mathcal{E}^2(\text{trunc}) + \mathcal{E}^2(\text{gradient}, p)\big]. \end{aligned}$$

# 5 Numerical Examples

In this section, we illustrate the behaviour of the algorithm with numerical examples. To this end, we choose equations that admit an explicit solution. This permits to compare the results obtained with our algorithm to a reference value. In this frame, we focus on three examples: the one-dimensional Burgers equation, the deterministic KPZ equation in dimension two and the one-dimensional porous media equation.

## 5.1 One Dimensional Burgers Equation

Consider first the backward Burgers equation:

$$\begin{aligned} &\partial_t u(t,x) - (u\partial_x u)(t,x) + \frac{\varepsilon^2}{2}\partial^2_{x,x}u(t,x) = 0, (t,x) \in [0,T[\times\mathbb{R},\ \varepsilon > 0, \\ &u(T,x) = H(x),\ x \in \mathbb{R},\ H \in C_b^{2+\alpha}(\mathbb{R}),\ \alpha \in ]0,1[. \end{aligned} \tag{5.1}$$

Using a non-linear transformation, one can derive an explicit expression of the solution of (5.1). This is known as the Cole-Hopf factorization, see Whitham [Whi68], Chapter IV, or Woyczyńsky [Woy98], Chapter III, for details. The solution of (5.1) then writes:

$$\forall (t,x) \in [0,T] \times \mathbb{R},\ u(t,x) = \frac{\mathbb{E}[H(x+\varepsilon B_{T-t})\phi(x+\varepsilon B_{T-t})]}{\mathbb{E}[\phi(x+\varepsilon B_{T-t})]}, \tag{5.2}$$

where $B$ is a standard Brownian motion and $\forall y \in \mathbb{R},\ \phi(y) \equiv \exp\left(-\varepsilon^{-2}\int_0^y H(u)du\right)$.

From the explicit representation (5.2), we can derive numerically, using e.g. a Riemann sum, a Monte-Carlo method, or a quantized version of the expectation (5.2), a reference solution to test the algorithm.

The reader may object that Burgers equation is actually semi-linear and not quasi-linear. Actually, it depends on whether we consider the non-linear term as a drift or as a second member. We describe below the algorithms associated to these two points of view, even if the coupled case is the only one to fulfill Assumption **(A)**.

Moreover, in the forward-backward representation of Burgers equation, the estimation procedure of the gradient is not necessary to compute the approximate solution $\overline{u}$. Numerically, this case turns out to be the most robust. Finally, in both cases, the intermediate predictor $\hat{v}$ is useless: in the coupled case, the drift of the diffusion $U$ reduces to $V$ (and thus does not depend on $W$), and in the decoupled one, the drift vanishes.

### 5.1.1 Explicit Expression of the Algorithms

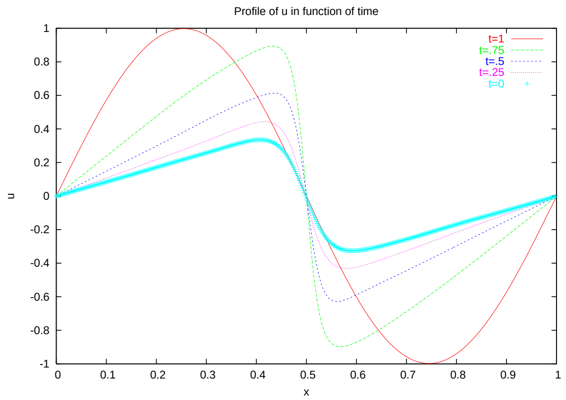
For a given final condition $H \in C_b^{2+\alpha}(\mathbb{R}),\ \alpha \in ]0,1[$, we write:

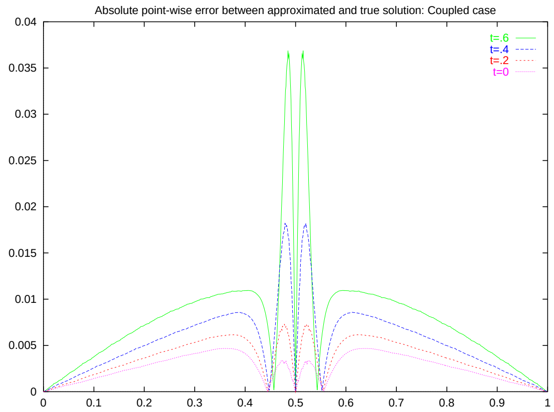
**Algorithm 5.1 (Coupled case)**
$\forall x \in \mathcal{C}_N,\ \overline{u}(T,x) \equiv H(x),$
$\forall k \in \{0,...,N-1\},\ \forall x \in \mathcal{C}_k,\ \overline{u}(t_k,x) \equiv \mathbb{E}[\overline{u}(t_{k+1}, \Pi_{k+1}(x - \overline{u}(t_{k+1},x)h + \varepsilon g(\Delta B^k)))],$
$\forall k \in \{0,...,N-1\},\ \forall x \in \mathcal{C}_k,\ \overline{v}(t_k,x) \equiv h^{-1}\mathbb{E}[\overline{u}(t_{k+1}, \Pi_{k+1}(x - \overline{u}(t_{k+1},x)h + \varepsilon g(\Delta B^k)))g(\Delta B^k)].$

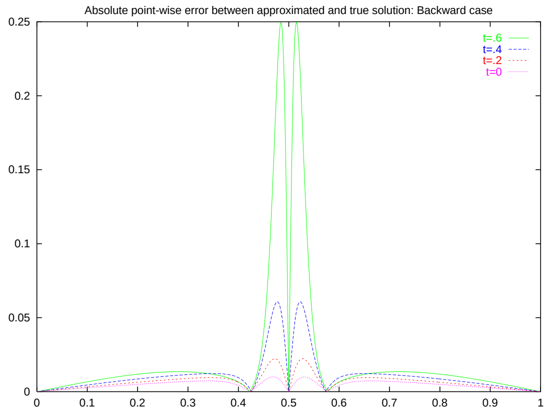
**Algorithm 5.2 (Pure Backward case)**
$\forall x \in \mathcal{C}_N,\ \overline{u}(T,x) \equiv H(x),$
$\forall k \in \{0,...,N-1\},\ \forall x \in \mathcal{C}_k,\ \overline{u}(t_k,x) \equiv \mathbb{E}[\overline{u}(t_{k+1}, \Pi_{k+1}(x + \varepsilon g(\Delta B^k)))] - h\varepsilon^{-1}\overline{u}(t_{k+1},x)\overline{v}(t_k,x),$
$\forall k \in \{0,...,N-1\},\ \forall x \in \mathcal{C}_k,\ \overline{v}(t_k,x) \equiv h^{-1}\mathbb{E}[\overline{u}(t_{k+1}, \Pi_{k+1}(x + \varepsilon g(\Delta B^k)))g(\Delta B^k)].$

### 5.1.2 Numerical Results

In order to avoid first to truncate the grids, we choose a periodic initial solution. Put to this end $H(x) = \sin(2\pi x)$ and derive from (5.2) that $u$ is 1-periodic. This allows to define $\overline{u}(t_k, .)$ on $\mathcal{C}_\infty$ by setting $\forall x \in \mathcal{C}_\infty,\ \overline{u}(t_k,x) \equiv \overline{u}(t_k, x - \lfloor x \rfloor)$. Hence, we can set $\mathcal{C}_k \equiv \mathcal{C}_\infty$ for $k \in \{0,...,N-1\}$. For $T = 1, \delta = 10^{-3}, h = .01, M = 160, \varepsilon = .15$, we present below the results of the previous algorithms. The explicit solution given by (5.2) is approximated by quantization techniques with a 500 points grid. We plot below some profiles of the reference value for various discretization times as well as the pointwise absolute error between this reference solution and the approximations obtained with our algorithms.



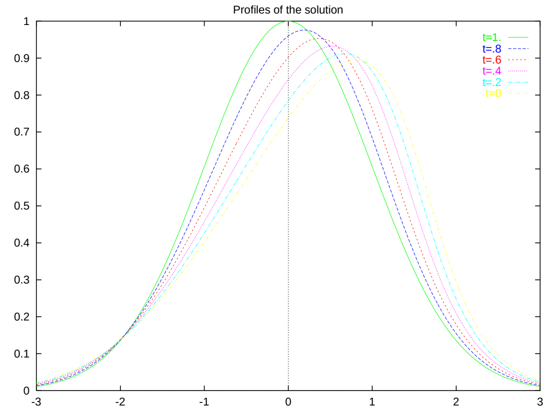

On the profiles of the explicit solution, the abscisses of the peaks of the initial sinusoidal wave are going closer to each other up to a given time $t_0$. This is a typical shocking wave behaviour. Because of the viscosity, i.e. $\varepsilon$ is non zero, there is no shock and the amplitude of the wave decays when $t$ goes to zero.

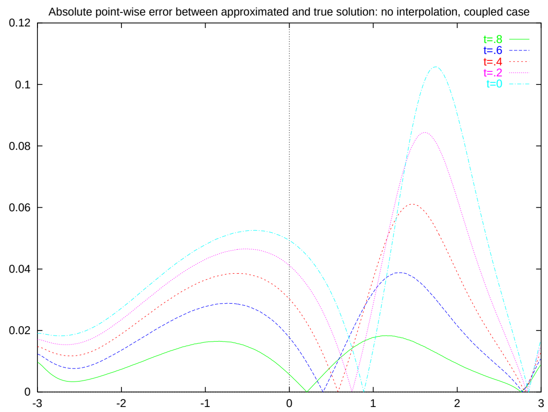
From a numerical point of view, the coupled case provides several advantages. First, the convergence of Algorithm 5.1 does not rely on the discretization procedure of the gradient. In short, there is no reason to update the gradient in order to obtain the approximate solution with the first algorithm. The computation of $\overline{v}$ just provides in this case an $L^2$ estimate of the gradient. At the opposite, this computation is necessary in Algorithm 5.2.

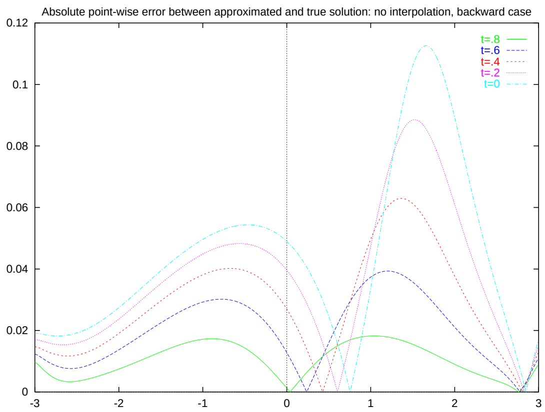
Moreover, since the coefficient $f(y,z) = \varepsilon^{-1}yz$ is not globally Lipschitz in the pure backward case, it is then another story to establish the convergence of Algorithm 5.2.

These theoritical remarks are confirmed by the above pictures. Even though Algorithm 5.2 does not behave too poorly, it is still less precise than Algorithm 5.1. The factor between the absolute pointwise errors of the two algorithms is approximately 5.

***Truncation error.*** We now illustrate the effects of truncation and deal with a non periodic final data. Namely, we take $H(x) = \exp(-x^2/2)$, $T = 1$, $h = .02$, $\rho = 3$, $\delta = \rho/500$, $M = 250$. The reference value, see profiles below, is computed from the Cole-Hopf explicit solution by quantization techniques with a 500 points grid. We run Algorithm 3.2 with the previous parameters to obtain:



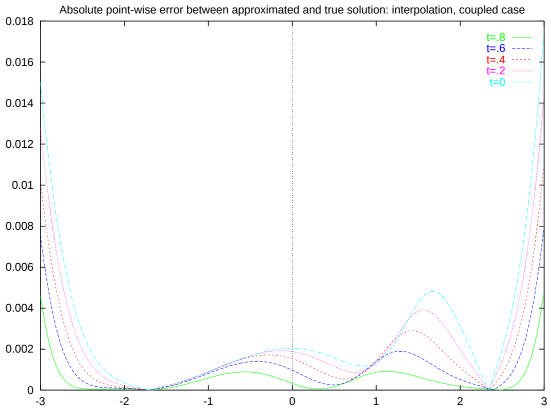



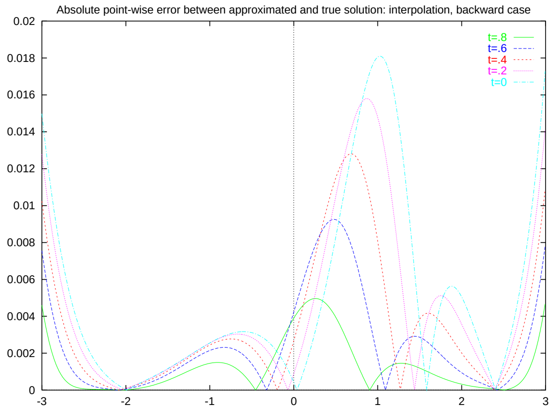




Choose now $R = 1$: the expected truncated error $\mathcal{E}(\text{trunc})$ is given by .25 whereas the absolute point-wise error between both solutions is bounded by .05 on $[-1, 1]$. This emphasizes the difficulty to control the truncation procedure in our algorithm. There are two possible arguments to explain this difference between .25 and .05. First, as explained in Subsection 9.1, our way to estimate $\mathcal{E}(\text{trunc})$ is suitable for unbounded drifts $b$, and more particulary for drifts depending on the gradient. In our case, the drift is bounded (since the solution is bounded by 1), and most relevant estimates could apply. Second, the fast decay of the final condition $H$ may explain the low influence of distant points on the values of the solution on $[-1, 1]$.

Note also that the relative error is close to .1 on $[-1, 1]$. A possible strategy to decrease it would consist in refining the spatial mesh.

We also feel that the choice of the rough projection mappings $(\Pi_k)_{0 \le k \le N}$ deeply affects the global error. To investigate more precisely their influence, we replace them by standard linear interpolation procedures (which are defined in an obvious way since the underlying space is one dimensional). In short, this permits to extend continuously the approximated solution $\overline{u}$ to the whole space. With the same parameters as above, we then get:




Numerically, the interpolation can thus be really relevant to improve the convergence (see Subsection 9.2 for further details and explanations on this point). To obtain the same precision without interpolation we need to refine significantly the parameters (taking e.g. $\delta = 2 \times 10^{-4}$). Let us finally mention that the results obtained with the coupled representation and the linear interpolation are still more accurate than with the backward one.

## 5.2 Quadratic Backward Equation: Deterministic KPZ Equation

In this subsection, we focus on the so-called "deterministic KPZ" equation (see e.g. Kardar, Parisi and Zhang [KPZ86] and Woyczyński [Woy98], Chapter I, for a physical interpretation):

$$\begin{aligned} &\partial_t u(t,x) + \frac{1}{2}\mathrm{tr}(\sigma\sigma^* \nabla^2_{x,x} u(t,x)) + \frac{\nu}{2}\langle \sigma\sigma^* \nabla_x u(t,x), \nabla_x u(t,x)\rangle = 0,\ (t,x) \in [0,T[\times \mathbb{R}^d,\\ &u(T,x) = H(x),\ x \in \mathbb{R}^d, \end{aligned} \tag{5.3}$$

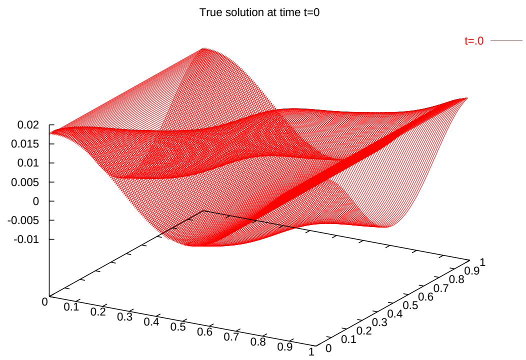
where $\nu \in \mathbb{R}^{+*}$ is a given parameter and $\sigma$ a given constant matrix such that $\sigma\sigma^*$ is positive definite.

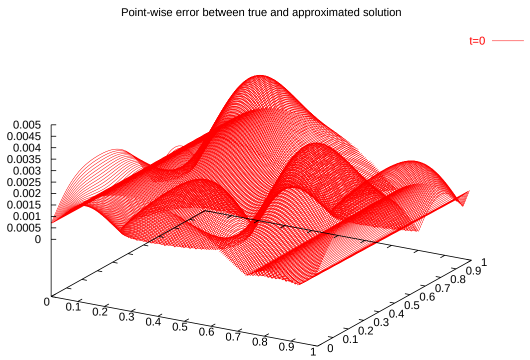
Such an equation admits too a "Cole-Hopf explicit solution", see again [KPZ86], that writes:

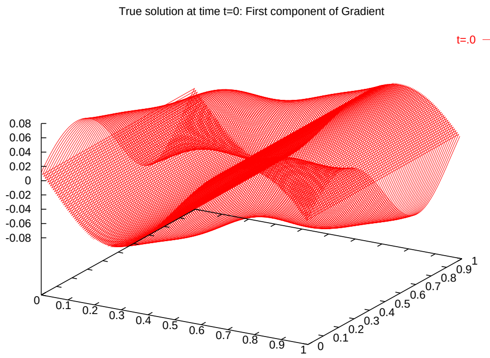
$$u(t,x) = \frac{\log(\mathbb{E}[\exp(\nu H(x + \sigma B_{T-t}))])}{\nu}. \tag{5.4}$$

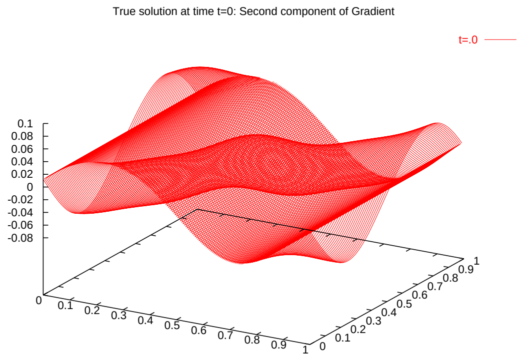
We then apply Algorithm 3.2 to Equation (5.3) seen as a true quasi-linear equation (so-called "coupled case" in the former subsection).

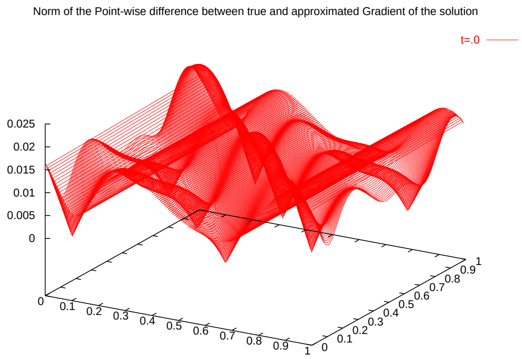
Concerning the initial condition, we choose $H(x) = \prod_{i=1}^d \sin(2\pi x_i)$. By construction, we have $\forall x \in \mathbb{R}^d$, $\forall k \in \mathbb{Z}^d$, $u(t, x+k) = u(t,x)$. Since the solution is periodic, $\overline{u}$ can be defined on the whole grid $\mathcal{C}_\infty$ (see also Paragraph 5.1.2). We now present the results for $d = 2$, $\nu = .3, T = .5$, $h = .02, \delta = 5 \times 10^{-4}, M = 160$ and $\sigma\sigma^* = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\rho = .8$. The reference value and its gradient have been derived from (5.4) using quantization techniques with a 500 points grid. At $t = 0$, one has:






The relative error between the approximate and true solutions is at most .25. The explanation seems rather simple: the explicit solution quickly decays as time decreases. Anyway, we feel that our algorithm manages to catch this specific decreasing phenomenon.

Let us also mention that the last picture represents the pointwise difference of the true and approximated gradients, but the control given by Theorem 4.2 just holds in $L^2$.

## 5.3 Porous Media Equation

To conclude this section, we focus on the equation (this example is taken from Makarov [Mak03]):

$$\begin{aligned}
&\partial_t u(t,x) + (u\partial^2_{x,x}u)(t,x) + (\partial_x u)^2(t,x) + u^2(t,x) = 0,\ (t,x) \in ]0,T] \times \mathbb{R}, \\
&u(T,x) = T^{-1}\frac{4}{3}\cos^2(\frac{\pi x}{L}),\ L = 2\sqrt{2}\pi,
\end{aligned} \tag{5.5}$$

which admits the $L$-periodic explicit solution $u(t,x) = t^{-1}\frac{4}{3}\cos^2(\frac{\pi x}{L})$.

Note that (5.5) does not fulfill Assumption **(A)**. In the sequel, we choose without any rigorous justifications to apply Algorithm 3.2 on $[T/2, T]$ (note however for a rough explanation that the quadratic growth of the coefficients ensures that Theorem 4.1 holds on a suitable interval $[t, T]$, for $t$ close enough to $T$, and, in the same way, Theorem 2.1 applies away from 0).

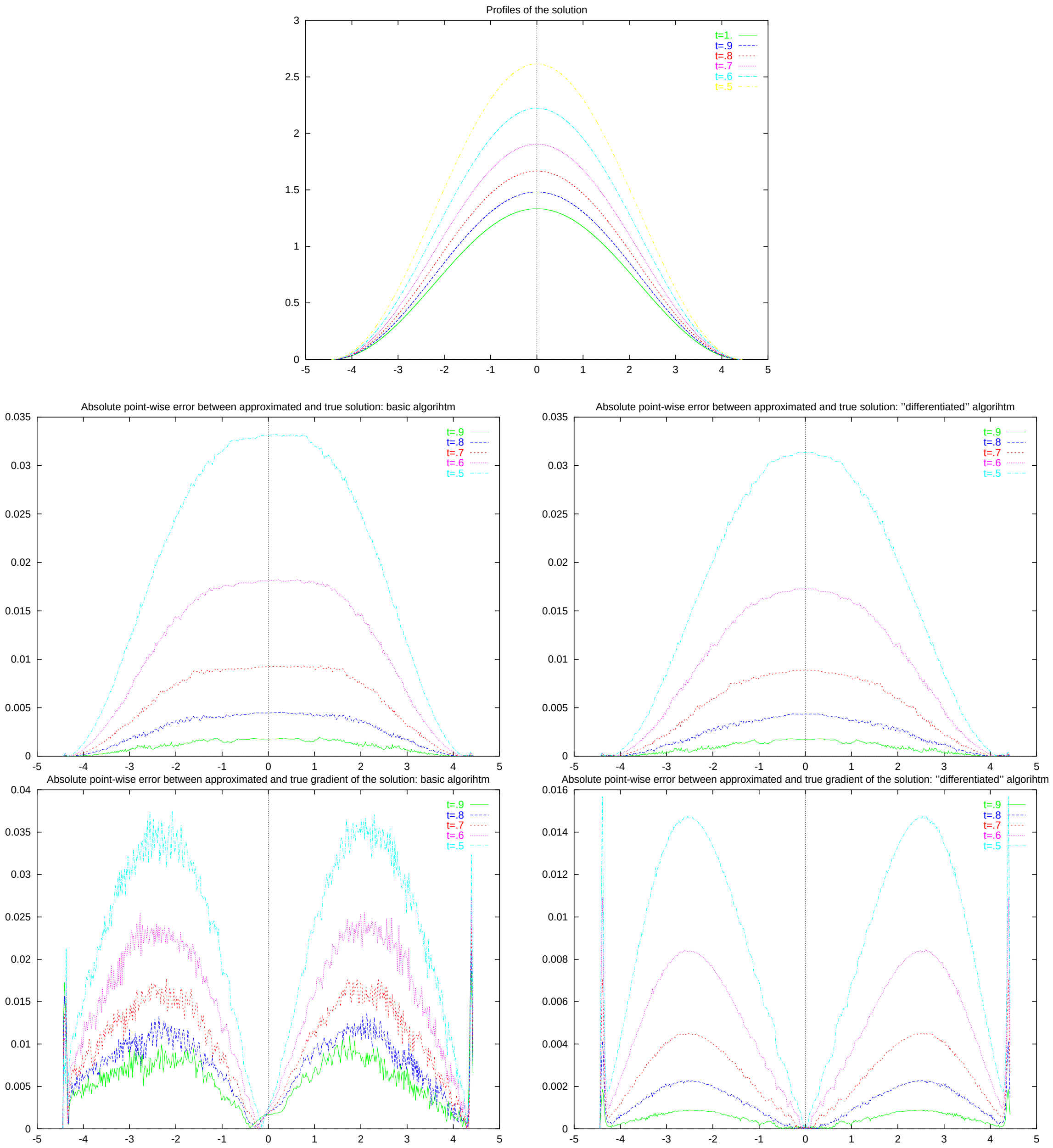
Nevertheless, as explained in Subsection 4.2, this procedure just provides an $L^2$-estimate of $\nabla_x u$. In this framework, we have decided to apply the so-called "differentiated" approach, described in Subsection 4.2, to obtain a pointwise estimate of $\nabla_x u$ (see Algorithm 5.3 below).

Note finally from the periodicity of $u$ that $\overline{u}$ can be defined on the whole grid $\mathcal{C}_\infty$ as in the previous example (see also Paragraph 5.1.2).

**Algorithm 5.3 (Differentiated Algorithm)**
$\forall x \in \mathcal{C}_N,\ \overline{u}(T,x) = T^{-1}\frac{4}{3}\cos^2(\frac{\pi x}{L}),\ \overline{w}(T,x) = T^{-1}\left(-\frac{8\pi}{3L}\cos(\frac{\pi x}{L})\sin(\frac{\pi x}{L})\right),$
$\forall k \in \{0,...,N-1\}, \forall x \in \mathcal{C}_k,$

$$\overline{u}(t_k,x) = \mathbb{E}\big[\overline{u}\big(t_{k+1},\Pi_{k+1}(x+\overline{w}(t_{k+1},x)h+\sqrt{2\overline{u}(t_{k+1},x)}g(\Delta B^k))\big)\big] + h\overline{u}(t_{k+1},x)^2,$$
$$\overline{w}(t_k,x) = \mathbb{E}\big[\overline{w}\big(t_{k+1},\Pi_{k+1}(x+3\overline{w}(t_{k+1},x)h+\sqrt{2\overline{u}(t_{k+1},x)}g(\Delta B^k))\big)\big] + 2h\overline{u}(t_{k+1},x)\overline{w}(t_{k+1},x).$$

For $T=1, h=.02, \delta=L/500, M=160$, we present below the results obtained first with Algorithm 3.2 (the approximation of the gradient with this algorithm is undefined at $x=\pm L/2$ and we thus arbitrarily set it to zero) and then with Algorithm 5.3. On $[-L/2, L/2]$ it comes:




We first observe that the approximated solutions obtained with the two algorithms are not significantly different. The main advantage of the differentiated algorithm is, as expected, for the pointwise approximation of the gradient. Indeed, in that case there is a factor 4 between the absolute pointwise errors associated to the two methods. Let us also indicate that both methods present some "singularity" in the neigbourhood of $x=\pm L/2$ for the estimation of the gradient. This could be expected for Algorithm 3.2 since the estimation of the gradient is obtained by dividing $\overline{v}$ by a term that goes to 0 when $x \longrightarrow \pm L/2$. It is a bit more surprising for Algorithm 5.3.

# 6 Proof, First Step: *A Priori* Controls of the Discrete Objects

In this section, we give various *a priori* estimates of the couple $(Y, Z)$ introduced in (3.15) and of the approximate diffusion $X$ defined in (3.14). These controls are necessary to establish Theorems 4.1, 4.2 and 4.3.

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

***Conditions on Parameters.*** Recall that $p$ denotes in Theorems 4.1, 4.2 and 4.3 a real larger than 2. It is from now on fixed. Furthermore, we assume that the conditions of Theorem 4.1 on $h$, $\delta$, $M$ and $\rho$ are fulfilled. Namely, the statements of the following Propositions and Lemmas hold for $h$, $h^{-1}\delta^2$, $h^{-1}M^{-2/d}$ and $\rho^{-1}$ small enough even if these conditions are not explicitly written.

## 6.1 Discrete Backward Equation and Associated *a priori* Estimates

***Discrete Feynman-Kac Formula.***

**Proposition 6.1** *With the notations of Algorithm 3.2 and Subsection 4.3, the sequence $(Y_{t_k})_{0\le k\le N}$ satisfies the discrete Feynman-Kac representation:*

$$\forall 0 \le k \le N-1,\ Y_{t_k} = \mathbb{E}\Big[H(X_{t_N}) + h\sum_{i=k+1}^{N} f\big(X_{t_{i-1}}, \overline{u}(t_i, X_{t_{i-1}}), Z_{t_{i-1}}\big)|\mathcal{F}_{t_k}\Big]. \tag{6.1}$$

**Proof.** Recall first that $\overline{u}$ and $\overline{v}$ are bounded (but estimates are not uniform with respect to the parameters): see Subsection 3.1, (3.6). In particular, the r.h.s. in the above expression is correctly defined. Note moreover, by definition of $X$, $Y$ and $Z$, that for a given $k \in \{0, ..., N-1\}$:

$$Y_{t_k} = \mathbb{E}\Big[Y_{t_{k+1}} + hf\big(X_{t_k}, \overline{u}(t_{k+1}, X_{t_k}), Z_{t_k}\big)|\mathcal{F}_{t_k}\Big].$$

The proof of Proposition 6.1 follows by iteration of this last identity. □

Recall now from the Martingale Representation Theorem, that there exists a progressively measurable process $\overline{Z}$, with finite moment of order two, such that:

$$Y_{t_N} + h\sum_{i=1}^{N} f\big(X_{t_{i-1}}, \overline{u}(t_i, X_{t_{i-1}}), Z_{t_{i-1}}\big) = Y_0 + \int_0^{t_N} \overline{Z}_s dB_s. \tag{6.2}$$

This representation permits to apply the BSDE machinery to our frame. However, as well-know in the literature devoted to SDEs (or equivalently to PDEs), several *a priori* estimates of the solution are necessary to apply this strategy.

Propositions 6.2 and 6.3 provide *a priori* estimates of the *supremum* norm of $\overline{u}$ and of the $L^2$ norms of $Z$ and $\overline{Z}$, as well as a pointwise upper bound of the predictors $\overline{v}$ and $\hat{v}$ defined in Algorithm 3.2. Lemma 6.4 gives a crucial estimate of the difference between $Z$ and $\overline{Z}$. The proofs are postponed to Subsection 6.3.

**Proposition 6.2** *There exists a constant $C_{6.2}$ such that:*

$$\sup_{i=0\ldots N}\big[\sup_{x\in\mathcal{C}_i}|\overline{u}(t_i, x)|^2\big] \le C_{6.2}.$$

**Proposition 6.3** *There exists a constant $C_{6.3}$ such that:*

$$\mathbb{E}\Big[\int_0^T |\overline{Z}_s|^2 ds\Big] + h\sum_{i=0}^{N-1}\mathbb{E}\big[|Z_{t_i}|^2\big] + h\sup_{i=0\ldots N}\big[\sup_{x\in\mathcal{C}_i}|\overline{v}(t_i, x)|^2\big] + h\sup_{i=0\ldots N-1}\big[\sup_{x\in\mathcal{C}_i}|\hat{v}(t_i, x)|^2\big] \le C_{6.3}.$$

The distance between $Z$ and $\overline{Z}$ can be estimated as follows:

**Lemma 6.4** *There exists a constant $C_{6.4}$ such that for $k \in \{1, ..., N\}$:*

$$\mathbb{E}\bigg|hZ_{t_{k-1}} - \mathbb{E}\Big[\int_{t_{k-1}}^{t_k} \overline{Z}_s ds | \mathcal{F}_{t_{k-1}}\Big]\bigg|^2 \leq C_{6.4} h^2 \mathcal{E}^2(\text{quantiz}).$$

## 6.2 Approximate Diffusion

***Extension of the "Discrete Diffusion".*** Recall from (3.14) that, up to now, the approximate diffusion $X$ is defined at the discretization times $(t_k)_{0\leq k\leq N}$. For the proof, we need to extend the definition of $X$ to the whole set $[0, T]$. Put for all $k \in \{0, ..., N-1\}$ and $t \in [t_k, t_{k+1}[$:

$$X_t \equiv X_{t_k} + b\big(X_{t_k}, \overline{u}(t_{k+1}, X_{t_k}), \hat{v}(t_k, X_{t_k})\big)(t - t_k) + \sigma\big(X_{t_k}, \overline{u}(t_{k+1}, X_{t_k})\big)\big[B_t - B_{t_k}\big]. \tag{6.3}$$

Hence, the extended process $(X_t)_{0\leq t\leq T}$ is discontinuous at times $(t_k)_{1\leq k\leq N-1}$. At a given time $t_k$, $1 \leq k \leq N$, the size of the jump performed by the process depends on the quantization error and on the spatial projection error. The first error is easily controlled by the distortion. Concerning the second one, the projection error is close to the spatial step $\delta$ when the grids are infinite. For truncated grids, the story is slightly different. In fact, as soon as the process stays inside $(\Delta_k)_{0\leq k\leq N}$, the projection error is close to the step $\delta$ of the interior mesh of the grid $(\mathcal{C}_k)_{0\leq k\leq N}$. At the opposite, outside $(\Delta_k)_{0\leq k\leq N}$, the jump of the process may take large values.

***Hitting Time of the Boundaries of the Grids.*** It is then well understood that we need to control the size of these jumps. In particular, we need to control the first hitting time of the boundaries of the sets $(\Delta_k)_{0\leq k\leq N}$ by the discrete process $(X_{t_k})_{0\leq k\leq N}$. This is the reason why the stopping time $\tau_\infty$, defined in (4.2), appears in the statement of Theorems 4.2 and 4.3.

***Increments of the Forward Process.***

**Lemma 6.5** *There exists a constant $C_{6.5}$ such that for every $k \in \{0, ..., N-1\}$:*

$$\forall t \in [t_k, t_{k+1}[, \ \mathbb{E}\big[|X_t - X_{t_k}|^2 | \mathcal{F}_{t_k}\big] \leq C_{6.5} h.$$

**Proof.** The proof just follows from Assumption **(A.1)**, i.e. $b$ and $\sigma$ have linear growth, and from the boundedness of $\overline{u}$ and $h|\hat{v}|^2$ (see Propositions 6.2 and 6.3). □

**Lemma 6.6** *For a given $k \in \{0, ..., N-1\}$, the norm of the increment $X_{t_{k+1}} - X_{t_k}$ is always bounded by $|\mathcal{T}(t_k, X_{t_k})| + \delta$. In particular, there exists a constant $C_{6.6}$ such that:*

$$\mathbb{E}\big[|X_{t_{k+1}} - X_{t_k}|^2 | \mathcal{F}_{t_k}\big] \leq C_{6.6}\big[h + \delta^2\big].$$

**Proof.** Since $X_{t_k} \in \mathcal{C}_\infty$, one has $\Pi_\infty(X_{t_k} + \mathcal{T}(t_k, X_{t_k})) = X_{t_k} + \Pi_\infty(\mathcal{T}(t_k, X_{t_k}))$ (invariance by translation of the grid $\mathcal{C}_\infty$). By definition of $X_{t_{k+1}}$, we get $X_{t_{k+1}} = \mathcal{Q}(R + \rho, \Pi_\infty(X_{t_k} + \mathcal{T}(t_k, X_{t_k}))) = \mathcal{Q}(R + \rho, X_{t_k} + \Pi_\infty(\mathcal{T}(t_k, X_{t_k})))$, where $\mathcal{Q}$ is defined in Section 3.4. Now, for every $y$ in the image of the projection $\mathcal{Q}(R+\rho, .)$ and for every $z \in \mathbb{R}^d$, the distance $|\mathcal{Q}(R + \rho, y + z) - y|$ is bounded by $|z|$. Hence:

$$\begin{aligned} |X_{t_{k+1}} - X_{t_k}| &= \big|\mathcal{Q}\big(R + \rho, X_{t_k} + \Pi_\infty(\mathcal{T}(t_k, X_{t_k}))\big) - X_{t_k}\big| \\ &\leq |\Pi_\infty(\mathcal{T}(t_k, X_{t_k}))| \\ &\leq |\mathcal{T}(t_k, X_{t_k})| + \delta. \end{aligned} \tag{6.4}$$

Thanks to Propositions 6.2 and 6.3, we are able to bound the drift $b$ appearing in the transition. Since $\mathbb{E}[|g(\Delta B^k)|^2] \leq Ch$, from Assumption **(A.1)** and Proposition 6.2, we also control the martingale part of the transition. This completes the proof. □

***Auxiliary Controls of the Forward Process.*** The time continuous extension of $X$ remains close to the discrete version of $X$ up to time $\tau_\infty$:

**Lemma 6.7** *There exists a constant $C_{6.7}$ such that:*

$$\sum_{i=0}^{N-1} \mathbb{E}\big[\mathbf{1}_{\{t_{i+1}<\tau_\infty\}} |X_{t_{i+1}} - X_{t_{i+1}-}|^2\big] \leq C_{6.7} h\big(\mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz})\big).$$

The proof is postponed to Subsection 6.3.

## 6.3 Proofs of the *A Priori* Controls

***Discrete BSDE***

This paragraph is devoted to the proof of Propositions 6.2, 6.3 and Lemma 6.4. We first give a control of the $L^2$ norm between $Z_{t_{k-1}}$ and the conditional expectation of $\int_{t_{k-1}}^{t_k} \overline{Z}_s ds$ appearing in Lemma 6.4. This preliminary estimate permits to prove Proposition 6.2. We then derive the complete proofs of Proposition 6.3 and Lemma 6.4.

*Step One: Preliminary Control in Lemma 6.4.*

The strategy just follows from the local BSDE writing explained in the proof of Proposition 6.1. Indeed, from (6.2), write for a given $k \in \{0, ..., N-1\}$:

$$Y_{t_{k+1}} + hf\big(X_{t_k}, \overline{u}(t_{k+1}, X_{t_k}), Z_{t_k}\big) = Y_{t_k} + \int_{t_k}^{t_{k+1}} \overline{Z}_s dB_s.$$

Multiplying this identity by $\Delta B^k$ and taking the conditional expectation w.r.t. $\mathcal{F}_{t_k}$ we obtain:

$$\mathbb{E}\big[Y_{t_{k+1}} \Delta B^k | \mathcal{F}_{t_k}\big] = \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \overline{Z}_s ds | \mathcal{F}_{t_k}\Big].$$

Recall now that:

$$\begin{aligned} hZ_{t_k} &= h\overline{v}(t_k, X_{t_k}) \\ &= \mathbb{E}\big[\overline{u}(t_{k+1}, X_{t_{k+1}}) g(\Delta B^k) | \mathcal{F}_{t_k}\big] = \mathbb{E}\big[Y_{t_{k+1}} g(\Delta B^k) | \mathcal{F}_{t_k}\big]. \end{aligned}$$

Hence, we deduce that:

$$hZ_{t_k} - \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \overline{Z}_s ds | \mathcal{F}_{t_k}\Big] = \mathbb{E}\big[Y_{t_{k+1}} \big(g(\Delta B^k) - \Delta B^k\big) | \mathcal{F}_{t_k}\big]. \tag{6.5}$$

Recall now from (3.10) that there exists $C > 0$ s.t.:

$$\mathbb{E}[|g(\Delta B^k) - \Delta B^k|^2] \le ChM^{-2/d}. \tag{6.6}$$

From (6.5) and (6.6) we derive:

$$\mathbb{E}\Big[\Big|hZ_{t_k} - \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \overline{Z}_s ds | \mathcal{F}_{t_k}\Big]\Big|^2\Big] \le ChM^{-2/d} \mathbb{E}\big[Y_{t_{k+1}}^2\big]. \tag{6.7}$$

As already explained, this preliminary estimate (6.7) is necessary to prove Proposition 6.2 from which we will derive $\mathbb{E}[Y_{t_{k+1}}^2] \le C$, and thus complete the proof of Lemma 6.4.

*Step Two: Proof of Proposition 6.2 (Boundedness of the Approximate Solution)*

To estimate the *supremum* norm of $\overline{u}$ over the grids $\mathcal{C}_0, ..., \mathcal{C}_N$, we follow the basic strategy of the BSDE theory and therefore apply a discrete version of Itô's formula to the discrete BSDE formula given in Proposition 6.1. Such a formula can be found in Shiryaev [Shi96], Chapter VII, Subsection 9:

**Lemma 6.8** *Let $(A_k)_{0 \le k \le n}$ be a sequence of vectors with values in $\mathbb{R}^q$, $q \ge 1$. Then :*

$$|A_k|^2 = |A_0|^2 + 2\sum_{i=1}^k \langle \Delta A_i, A_{i-1}\rangle + \sum_{i=1}^k |\Delta A_i|^2,$$

*where* $\forall 1 \le i \le n,\ \Delta A_i \equiv A_i - A_{i-1}$.

Apply now Lemma 6.8 to the sequence $(Y_{t_k})_{0\le k\le N}$. We obtain:

$$|Y_T|^2 = |Y_0|^2 + 2\sum_{i=1}^N \langle Y_{t_i} - Y_{t_{i-1}}, Y_{t_{i-1}}\rangle + \sum_{i=1}^N |Y_{t_i} - Y_{t_{i-1}}|^2.$$

Recall from Subsection 6.1 that:

$$Y_{t_i} - Y_{t_{i-1}} = -hf(X_{t_{i-1}}, \overline{u}(t_i, X_{t_{i-1}}), Z_{t_{i-1}}) + \int_{t_{i-1}}^{t_i} \overline{Z}_s dB_s.$$

We deduce that:

$$\begin{aligned}\mathbb{E}|Y_T|^2 = |Y_0|^2 &+ 2h\sum_{i=1}^N \mathbb{E}\langle -f(X_{t_{i-1}}, \overline{u}(t_i, X_{t_{i-1}}), Z_{t_{i-1}}), Y_{t_{i-1}}\rangle \\ &+ h^2\sum_{i=1}^N \mathbb{E}\big[f^2(X_{t_{i-1}}, \overline{u}(t_i, X_{t_{i-1}}), Z_{t_{i-1}})\big] + \mathbb{E}\int_0^T |\overline{Z}_s|^2 ds.\end{aligned}$$

Thanks to Assumption **(A.1)**, there exists a constant $C$ such that:

$$|Y_0|^2 + \mathbb{E}\int_0^T |\overline{Z}_s|^2 ds \le \mathbb{E}|Y_T|^2 + Ch\sum_{i=1}^N \mathbb{E}\Big[|Y_{t_{i-1}}|\big(1 + |\overline{u}(t_i, X_{t_{i-1}})| + |Z_{t_{i-1}}|\big)\Big].$$

From Young's inequality, we derive for every $\eta > 0$:

$$\begin{aligned}|Y_0|^2 + \mathbb{E}\int_0^T |\overline{Z}_s|^2 ds \le \mathbb{E}|Y_T|^2 &+ C\eta^{-1}h\sum_{i=1}^N \mathbb{E}\big[|Y_{t_{i-1}}|^2\big] \\ &+ C\eta h\sum_{i=1}^N \mathbb{E}\big[1 + |\overline{u}(t_i, X_{t_{i-1}})|^2 + |Z_{t_{i-1}}|^2\big].\end{aligned} \tag{6.8}$$

From (6.7) we get that for every $i \in \{1, ..., N\}$:

$$h\mathbb{E}\big[|Z_{t_{i-1}}|^2\big] \le 2\Big[\mathbb{E}\int_{t_{i-1}}^{t_i} |\overline{Z}_s|^2 ds + CM^{-2/d}\mathbb{E}\big[Y_{t_i}^2\big]\Big].$$

Hence, from (6.8) and the above identity:

$$\begin{aligned}|Y_0|^2 + \frac{h}{2}\sum_{i=0}^{N-1} \mathbb{E}\big[|Z_{t_i}|^2\big] \le \mathbb{E}|Y_T|^2 &+ C(1+\eta^{-1})(h + M^{-2/d})\sum_{i=0}^N \mathbb{E}\big[|Y_{t_i}|^2\big] \\ &+ C\eta h\sum_{i=1}^N \mathbb{E}\big[1 + |\overline{u}(t_i, X_{t_{i-1}})|^2 + |Z_{t_{i-1}}|^2\big].\end{aligned} \tag{6.9}$$

Choose $\eta = (4C)^{-1}$ and deduce (with a new constant $C$) that:

$$|Y_0|^2 \le \mathbb{E}|Y_T|^2 + C(h + M^{-2/d})\sum_{i=0}^N \mathbb{E}\big[|Y_{t_i}|^2\big] + Ch\sum_{i=1}^N \mathbb{E}\big[1 + |\overline{u}(t_i, X_{t_{i-1}})|^2\big].$$

Recall that $|Y_T| \le |H|_\infty$. Thus, we claim (recall that $M^{-2/d} < h$):

$$\sup_{x\in\mathcal{C}_0} |\overline{u}(0,x)|^2 \le C + Ch\sum_{i=0}^N \sup_{x\in\mathcal{C}_i} |\overline{u}(t_i,x)|^2.$$

Therefore, there exists a constant $c > 0$ such that for $h < c$ (recall indeed that $h$ is small):

$$\sup_{x\in\mathcal{C}_0} |\overline{u}(0,x)|^2 \le C + Ch\sum_{i=1}^N \sup_{x\in\mathcal{C}_i} |\overline{u}(t_i,x)|^2.$$

As usual in BSDE theory, we could establish in a similar way that for every initial condition $(t_k, x)$, $1 \leq k \leq N$:

$$\forall k \in \{0, ..., N-1\}, \ \sup_{x \in \mathcal{C}_k} |\overline{u}(t_k, x)|^2 \leq C + Ch \sum_{i=k+1}^{N} \sup_{x \in \mathcal{C}_i} |\overline{u}(t_i, x)|^2.$$

A discrete version of Gronwall's Lemma yields the result. □

*Step Three: Proofs of Proposition 6.3 and Lemma 6.4.*

*Proof of Proposition 6.3.* The $L^2$-estimate of $Z$ follows from Proposition 6.2 and (6.9). Then, the $L^2$-estimate of $\overline{Z}$ follows from (6.8).

Finally, as a consequence of Proposition 6.2 and the definitions of $\overline{v}$ and $\hat{v}$, see Algorithm 3.2, we deduce the estimates of the *supremum* norms of $\overline{v}$ and $\hat{v}$. □

*Proof of Lemma 6.4.* Lemma 6.4 follows from (6.7) and Proposition 6.2. □

***Approximate Forward Diffusion: Proof of Lemma 6.7.***

Recall from (6.3) that the difference $X_{t_{i+1}} - X_{t_{i+1}-}$ writes:

$$\begin{aligned} X_{t_{i+1}} &- X_{t_{i+1}-} \\ &= \big[\Pi_{i+1}\big(X_{t_i} + \mathcal{T}(t_i, X_{t_i})\big) - \big(X_{t_i} + \mathcal{T}(t_i, X_{t_i})\big)\big] + \sigma\big(X_{t_i}, \overline{u}(t_{i+1}, X_{t_i})\big)\big[g(\Delta B^i) - \Delta B^i\big] \\ &\equiv E_1(i+1) + E_2(i+1). \end{aligned} \tag{6.10}$$

$E_1(i+1)$ appears as a projection error and $E_2(i+1)$ as a quantization one. There is no difficulty to estimate the second term: it is readily seen from (3.10) that $\mathbb{E}[|E_2(i+1)|^2|\mathcal{F}_{t_i}] \leq ChM^{-2/d}$. To estimate the first term, note that $X_{t_i} + \mathcal{T}(t_i, X_{t_i})$ belongs to $\Delta_{i+1}$ on the event $\{t_{i+1} < \tau_\infty\}$. Thus, the distance between $X_{t_i} + \mathcal{T}(t_i, X_{t_i})$ and $\Pi_{i+1}(X_{t_i} + \mathcal{T}(t_i, X_{t_i}))$ is then bounded by the step $\delta$. Deduce that the term $E_1(i+1)$ is bounded by $\delta$ on $\{t_{i+1} < \tau_\infty\}$. Finally,

$$\sum_{i=1}^{N} \mathbb{E}\big[\mathbf{1}_{\{t_{i+1} < \tau_\infty\}} |X_{t_{i+1}} - X_{t_{i+1}-}|^2\big] \leq Ch^{-1}\big[\delta^2 + hM^{-2/d}\big]. \tag{6.11}$$

This completes the proof. □

# 7 Proof, Second Step: Stability Properties

This section focuses on the second step of the proof of Theorems 4.1, 4.2 and 4.3, and aims to establish more specifically a suitable intermediate inequality, close to usual stability properties of FBSDEs.

***Strategy.*** Recall first that two main strategies are conceivable in the theoretical framework to establish classical stability theorems for FBSDEs.

Denote to this end by $(U', V', W')$ a solution of another FBSDE of type $(E)$ with different coefficients. The associated PDE solution is just denoted by $u'$. In order to compare $u'$ with $u$, recall that the following approaches have been employed in the literature devoted to FBSDE:

1. First, the recent induction principle given in Delarue [Del02] can be applied. In short, $u$ and $u'$ are compared on a neighborhood of the boundary $T$ with classical arguments of stochastic analysis and the estimate of the difference between these solutions is then extended by induction from the final bound $T$ to the initial bound 0. The local estimates consist in studying the distance between $U$ and $U'$ and between $(V, W)$ and $(V', W')$. This strategy has been successfully applied to establish the existence and uniqueness of solutions to FBSDEs under a non-degeneracy assumption, see again [Del02], or to establish convergence properties arising in homogenization of quasilinear PDEs [Del04].

2. A second approach follows the earlier *Four Step Scheme* of Ma, Protter and Yong [MPY94]. In a nutshell, instead of studying the difference between $U$ and $U'$ and between $(V, W)$ and $(V', W')$, the process $(u(t, U'_t))_{0 \leq t \leq T}$ is written with Itô's formula as the solution of a BSDE. This BSDE is then compared with the one satisfied by $(V', W')$. In particular, these BSDEs are both written with respect

to the same diffusion $U'$. Generally speaking, this strategy holds when $u$ is smooth enough (e.g. if $u$ satisfies Theorem 2.1). It is then more direct than the previous one.

Thanks to Theorem 2.1, we will apply the second strategy: we will compare the process $Y$ with the process $(u(t,X_t))_{0\le t\le T\wedge\tau_\infty}$ (see (6.3) for the definition of the extension of $X$).

### 7.1 Statements of the Stability Results

***First Stability Property.*** Applying the usual FBSDE machinery, we are able to establish in Subsection 7.2 the following first inequality:

**Proposition 7.1** *There exists a constant $C_{7.1}$ such that for $\eta$ small enough:*

$$
\begin{aligned}
&|(\overline{u}-u)(0,x_0)|^2 + C_{7.1}^{-1} h \sum_{j=1}^{N} \mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\
&\quad \le C_{7.1}\Bigg[\mathbb{P}\{\tau_\infty<+\infty\} + \mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz}) \\
&\qquad + \eta^{-1} h \sum_{j=1}^{N} \mathbb{E}\big[|(\overline{u}-u)(t_j,X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\
&\qquad + \eta^{-1} h \sum_{j=1}^{N} \mathbb{E}\big[|(\overline{u}-u)(t_{j-1},X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\
&\qquad + (\eta+h) h \sum_{j=1}^{N} \mathbb{E}\big[|(\hat{v}-v)(t_{j-1},X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Bigg].
\end{aligned}
\tag{7.1}
$$

*When the drift $b$ does not depend on $z$, the last term of the r.h.s. does not appear.*

***Estimates of the Gradient Increment.*** Assume for the moment that Proposition 7.1 holds. Note that the main problem then remains to estimate the last term in the r.h.s. of (7.1). Thanks to the specific choice of $\hat{v}$ in Subsection 3.3, we are able to establish in Section 7.3 the following control:

**Proposition 7.2** *There exists a constant $C_{7.2}$ such that, for $k\in\{0,...,N-1\}$, on $\{t_k<\tau_\infty\}$:*

$$
|(\hat{v}-v)(t_k,X_{t_k})| \le C_{7.2}\Big[\mathcal{E}(\text{gradient},p) + \mathcal{E}(\text{time}) + h\mathcal{E}(\text{space}) + \mathbb{E}\big[|(\overline{v}-v)(t_{k+1},X_{t_{k+1}})|^2|\mathcal{F}_{t_k}\big]^{1/2}\Big].
$$

***Main Stability Theorem.*** From Propositions 7.1 and 7.2, we claim:

**Theorem 7.3** *There exists a constant $C_{7.3}$ such that for $\eta$ small enough:*

$$
\begin{aligned}
&|(\overline{u}-u)(0,x_0)|^2 + C_{7.3}^{-1} h \sum_{j=1}^{N} \mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\
&\quad \le C_{7.3}\Bigg[\mathbb{P}\{\tau_\infty<+\infty\} + \mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz}) + \mathcal{E}^2(\text{gradient},p) \\
&\qquad + \eta^{-1} h \sum_{j=1}^{N} \mathbb{E}\big[|(\overline{u}-u)(t_j,X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\
&\qquad + \eta^{-1} h \sum_{j=1}^{N} \mathbb{E}\big[|(\overline{u}-u)(t_{j-1},X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\
&\qquad + (\eta+h) h \sum_{j=1}^{N} \mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_j})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Bigg].
\end{aligned}
\tag{7.2}
$$

Application of Theorem 7.3 to the proof of Theorems 4.1, 4.2 and 4.3 is given in Section 8.

## 7.2 Proof of Proposition 7.1

***Starting Point: Time Continuous Backward Processes.***

Recall that we aim to apply the second strategy exposed in the beginning of this section and thus to apply Itô's formula to $(u(t,X_t))_{0\le t\le T}$. Referring to the structure of the PDE $(\mathcal{E})$, set to this end for notational convenience:

$$\forall t \in [0,T],\ \overline{V}_t \equiv u(t,X_t),\ \overline{W}_t \equiv \nabla_x u(t,X_t)\sigma(X_t,\overline{V}_t). \tag{7.3}$$

Note moreover that the martingale part of $(\overline{V}_t)_{0\le t\le T}$ is driven by:

$$\forall t \in [0,T[,\ \hat{W}_t \equiv \nabla_x u(t,X_t)\sigma\big(X_{\phi(t)},\overline{u}(\phi(t)+h,X_{\phi(t)})\big). \tag{7.4}$$

where $\phi(t)=t_k$ for $t_k\le t<t_{k+1},\ k\in\{0,...,N-1\}$.

From Theorem 2.1, we derive the following *a priori estimates* of $\overline{V},\overline{W}$:

$$\forall k\in\{0,...,N-1\},\ \forall s\in[t_k,t_{k+1}[,\ \mathbb{E}\big[|\overline{V}_s-\overline{V}_{t_k}|+|\overline{W}_s-\overline{W}_{t_k}|\ |\mathcal{F}_{t_k}\big]\le C\mathbb{E}\big[(s-t_k)^{1/2}+|X_s-X_{t_k}|\ |\mathcal{F}_{t_k}\big].$$

Hence, from Lemma 6.5, we get for $s\in[t_k,t_{k+1}[$ (recall that $h$ is small) :

$$\mathbb{E}\big[|\overline{V}_s-\overline{V}_{t_k}|+|\overline{W}_s-\overline{W}_{t_k}|\ |\mathcal{F}_{t_k}\big]\le Ch^{1/2}. \tag{7.5}$$

***Step One: Itô's formula for $\bar{V}$.***

Using Itô's formula and the equation satisfied by $u$, we obtain for $t\in[t_i,t_{i+1}[,\ i\in\{0,...,N-1\}$:

$$\begin{aligned}
\overline{V}_t = \overline{V}_{t_i} &+ \int_{t_i}^t \langle \nabla_x u(s,X_s), b\big(X_{t_i},\overline{u}(t_{i+1},X_{t_i}),\hat{v}(t_i,X_{t_i})\big)-b(X_s,\overline{V}_s,\overline{W}_s)\rangle ds\\
&+\frac{1}{2}\int_{t_i}^t \mathrm{tr}\Big(\big[a(X_{t_i},\overline{u}(t_{i+1},X_{t_i}))-a(X_s,\overline{V}_s)\big]\nabla^2_{x,x}u(s,X_s)\Big)ds\\
&-\int_{t_i}^t f\big(X_s,\overline{V}_s,\overline{W}_s\big)ds+\int_{t_i}^t \nabla_x u(s,X_s)\sigma\big(X_{t_i},\overline{u}(t_{i+1},X_{t_i})\big)dB_s.
\end{aligned}$$

Let $t$ tend to $t_{i+1}$ and deduce with obvious notation that:

$$\begin{aligned}
&\overline{V}_{t_{i+1}}-\overline{V}_{t_i}\\
&\quad=\overline{V}_{t_{i+1}}-\overline{V}_{t_{i+1}-}+\int_{t_i}^{t_{i+1}}\big[F\big(s,X_s,X_{t_i},\overline{u}(t_{i+1},X_{t_i}),\hat{v}(t_i,X_{t_i})\big)-F\big(s,X_s,X_s,\overline{V}_s,\overline{W}_s\big)\big]ds\\
&\quad-\int_{t_i}^{t_{i+1}} f\big(X_s,\overline{V}_s,\overline{W}_s\big)ds+\int_{t_i}^{t_{i+1}}\hat{W}_s dB_s.
\end{aligned}$$

***Step Two: Difference of the Processes.***

The strategy is well-known: we aim to make the difference between $\overline{V}$ and $Y$ and then to apply the usual BSDE machinery to estimate the distance between these processes. Hence, we claim from (6.2):

$$\begin{aligned}
\overline{V}_{t_{i+1}}-Y_{t_{i+1}}-\big[\overline{V}_{t_i}-Y_{t_i}\big] &= \overline{V}_{t_{i+1}}-\overline{V}_{t_{i+1}-}\\
&+\int_{t_i}^{t_{i+1}}\big[F\big(s,X_s,X_{t_i},\overline{u}(t_{i+1},X_{t_i}),\hat{v}(t_i,X_{t_i})\big)-F\big(s,X_s,X_s,\overline{V}_s,\overline{W}_s\big)\big]ds\\
&-\int_{t_i}^{t_{i+1}}\big[f\big(X_s,\overline{V}_s,\overline{W}_s\big)-f(X_{t_i},\overline{u}(t_{i+1},X_{t_i}),Z_{t_i})\big]ds\\
&+\int_{t_i}^{t_{i+1}}\big[\hat{W}_s-\overline{Z}_s\big]dB_s\\
&\equiv \Delta E_{i+1}(1)+\Delta E_{i+1}(2)+\Delta E_{i+1}(3)+\Delta E_{i+1}(4).
\end{aligned}$$

As explained above, we now aim to estimate the distance between the processes $\overline{V}$ and $Y$ up to time $\tau_\infty \wedge T$. Lemma 6.8 yields:

$$
\begin{aligned}
&\mathbb{E}|\overline{V}_{T\wedge\tau_\infty} - Y_{T\wedge\tau_\infty}|^2 = |\overline{V}_0 - Y_0|^2 \\
&\quad + 2\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[\overline{V}_{t_{j-1}} - Y_{t_{j-1}}\big]\big[\Delta E_j(1) + \Delta E_j(2) + \Delta E_j(3) + \Delta E_j(4)\big]\Big] \\
&\quad + \mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[\Delta E_j(1) + \Delta E_j(2) + \Delta E_j(3) + \Delta E_j(4)\big]^2\Big].
\end{aligned}
\tag{7.6}
$$

Denote $\forall j \in \{1, ..., N\},\ E_j \equiv \Delta E_j(1) + \Delta E_j(2) + \Delta E_j(3)$. From the above expression, we get:

$$
\begin{aligned}
&|\overline{V}_0 - Y_0|^2 + \sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\Delta E_j(4)^2] = \mathbb{E}|\overline{V}_{T\wedge\tau_\infty} - Y_{T\wedge\tau_\infty}|^2 \\
&- 2\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[\overline{V}_{t_{j-1}} - Y_{t_{j-1}}\big]E_j\Big] - \sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}(E_j^2 + 2E_j\Delta E_j(4))].
\end{aligned}
$$

From Young's inequality we derive:

$$
\begin{aligned}
&|\overline{V}_0 - Y_0|^2 + \frac{1}{2}\sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\Delta E_j(4)^2] \le \mathbb{E}|\overline{V}_{T\wedge\tau_\infty} - Y_{T\wedge\tau_\infty}|^2 \\
&- 2\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[\overline{V}_{t_{j-1}} - Y_{t_{j-1}}\big]E_j\Big] + \sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}E_j^2].
\end{aligned}
\tag{7.7}
$$

Put:

$$
\begin{aligned}
D(1) &\equiv -2\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[\overline{V}_{t_{j-1}} - Y_{t_{j-1}}\big]E_j\Big],\ D(2) \equiv \sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}E_j^2], \\
D(3) &\equiv \sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\Delta E_j(4)^2].
\end{aligned}
\tag{7.8}
$$

***Step Three: Standard BSDE Techniques.***

Following the BSDE techniques, we have to upper bound $D(1), D(2)$ (resp. lower bound $D(3)$) by terms appearing in the r.h.s. (resp. l.h.s.) of (7.1). The following lemmas whose proofs are postponed to the end of the subsection give the needed controls.

**Lemma 7.4** *There exists a constant $C_{7.4}$ such that for $\eta \in ]0, 1]$:*

$$
\begin{aligned}
&|D(1)| + D(2) \le C\big[\mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz}) + \mathbb{P}\{\tau_\infty < +\infty\}\big] \\
&\quad + Ch\sum_{j=1}^{N}\Big[\eta^{-1}\mathbb{E}\big[|(\overline{u} - u)(t_{j-1}, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] + \mathbb{E}\big[|(\overline{u} - u)(t_j, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Big] \\
&\quad + h(\eta + h)\sum_{j=1}^{N}\Big[\mathbb{E}\big[|(\hat{v} - v)(t_{j-1}, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] + \mathbb{E}\big[|(\overline{v} - v)(t_{j-1}, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Big].
\end{aligned}
$$

**Lemma 7.5** *There exists a constant $C_{7.5} > 0$ such that:*

$$
\begin{aligned}
D(3) \ge\ & C_{7.5}^{-1}h\sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|(\overline{v} - v)(t_{j-1}, X_{t_{j-1}})|^2] - C_{7.5}\mathcal{E}^2(\text{quantiz}) - C_{7.5}\mathcal{E}^2(\text{time}) \\
&- C_{7.5}h\sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|(\overline{u} - u)(t_j, X_{t_{j-1}})|^2].
\end{aligned}
$$

Note to conclude the proof of Proposition 7.1 that $Y_T = \overline{V}_T$. Hence, from Theorem 2.1 and Proposition 6.2 (boundedness of $u$ and $\overline{u}$), $\mathbb{E}|\overline{V}_{T\wedge\tau_\infty} - Y_{T\wedge\tau_\infty}|^2 \leq C\mathbb{P}\{\tau_\infty < T\} \leq C\mathbb{P}\{\tau_\infty < +\infty\}$. Choose finally $\eta$ small enough to obtain inequality (7.1) from (7.7), (7.8), and Lemmas 7.4 and 7.5. This completes, up to the proofs of Lemmas 7.4 and 7.5, the proof of Proposition 7.1. □

***Proof of Lemma 7.4.***

Recall that:

$$D(1) = -2\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[\overline{V}_{t_{j-1}} - Y_{t_{j-1}}\big]\big[\Delta E_j(1) + \Delta E_j(2) + \Delta E_j(3)\big]\Big]. \tag{7.9}$$

Note from Theorem 2.1 that $\Delta E(2)$ and $\Delta E(3)$ may be seen as "Lipschitz" differences since the partial derivatives of $u$ of order one and two in $x$ are bounded. Hence, it is readily seen that there exists a constant $C$, such that:

$$\begin{aligned}|D(1)| \leq C\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|\overline{V}_{t_{j-1}} - Y_{t_{j-1}}| \times \Big(&|\overline{V}_{t_j} - \overline{V}_{t_j-}| \\ &+ \int_{t_{j-1}}^{t_j}\big[|X_s - X_{t_{j-1}}| + |\overline{V}_s - \overline{u}(t_j, X_{t_{j-1}})| + |\overline{W}_s - \hat{v}(t_{j-1}, X_{t_{j-1}})|\big]ds \\ &+ \int_{t_{j-1}}^{t_j}\big[|X_s - X_{t_{j-1}}| + |\overline{V}_s - \overline{u}(t_j, X_{t_{j-1}})| + |\overline{W}_s - Z_{t_{j-1}}|\big]ds\Big)\Big].\end{aligned}$$

Recall that $\overline{V}_s = u(s, X_s)$. From Theorem 2.1 (Hölder regularity of $u$ in $t$), (7.5) (regularity of $\overline{V}$ and $\overline{W}$) and Lemma 6.5 (control of the increments of $X$), we then deduce:

$$\begin{aligned}|D(1)| \leq C\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|\overline{V}_{t_{j-1}} - Y_{t_{j-1}}| \times \Big(&|\overline{V}_{t_j} - \overline{V}_{t_j-}| \\ &+ \int_{t_{j-1}}^{t_j}\big[h^{1/2} + |(\overline{u} - u)(t_j, X_{t_{j-1}})|\big]ds \\ &+ \int_{t_{j-1}}^{t_j}\big[|\overline{W}_{t_{j-1}} - \hat{v}(t_{j-1}, X_{t_{j-1}})| + |\overline{W}_{t_{j-1}} - Z_{t_{j-1}}|\big]ds\Big)\Big].\end{aligned} \tag{7.10}$$

Recall now that $\overline{W}_{t_{j-1}} = v(t_{j-1}, X_{t_{j-1}})$ and $Z_{t_{j-1}} = \overline{v}(t_{j-1}, X_{t_{j-1}})$. Thus, applying Young's inequality to (7.10), it comes for every $\eta \in ]0,1]$:

$$\begin{aligned}|D(1)| &\leq C\mathcal{E}^2(\text{time}) + C\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[|\overline{V}_{t_{j-1}} - Y_{t_{j-1}}||\overline{V}_{t_j} - \overline{V}_{t_j-}|\big]\Big] \\ &+ Ch\sum_{j=1}^{N}\Big[\eta^{-1}\mathbb{E}\big[|(\overline{u} - u)(t_{j-1}, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] + \mathbb{E}\big[|(\overline{u} - u)(t_j, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Big] \\ &+ \eta h\sum_{j=1}^{N}\Big[\mathbb{E}\big[|(\hat{v} - v)(t_{j-1}, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] + \mathbb{E}\big[|(\overline{v} - v)(t_{j-1}, X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Big].\end{aligned} \tag{7.11}$$

It now remains to estimate the second term in the r.h.s. of (7.11). Note first that for all $j \in \{1, ..., N\}$, $\{t_{j-1} < \tau_\infty\} = \{t_j < \tau_\infty\} \cup \{t_j = \tau_\infty\}$. Hence, thanks to the boundedness of $u$ and $\overline{u}$ (see Theorem 2.1 and Proposition 6.2):

$$\begin{aligned}&\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[|\overline{V}_{t_{j-1}} - Y_{t_{j-1}}||\overline{V}_{t_j} - \overline{V}_{t_j-}|\big]\Big] \\ &\quad\leq \Big[\mathbb{E}\sum_{j=1}^{N}\big[\mathbf{1}_{\{t_j<\tau_\infty\}}|\overline{V}_{t_{j-1}} - Y_{t_{j-1}}|^2\big]\Big]^{1/2}\Big[\mathbb{E}\sum_{j=1}^{N}\big[\mathbf{1}_{\{t_j<\tau_\infty\}}|\overline{V}_{t_j} - \overline{V}_{t_j-}|^2\big]\Big]^{1/2} \\ &\qquad+ C\mathbb{P}\{\tau_\infty < +\infty\}.\end{aligned} \tag{7.12}$$

Deduce from Lemma 6.7 (jumps of the process $X$) and the global Lipschitz property of $u$ (see Theorem 2.1) that:

$$\begin{aligned}
&\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[|\overline{V}_{t_{j-1}}-Y_{t_{j-1}}||\overline{V}_{t_j}-\overline{V}_{t_j-}|\big]\Big]\\
&\quad\le C\Big[\mathbb{E}\sum_{j=1}^{N}\big[\mathbf{1}_{\{t_j<\tau_\infty\}}|(\overline{u}-u)(t_{j-1},X_{t_{j-1}})|^2\big]\Big]^{1/2}h^{1/2}\big(\mathcal{E}^2(\text{space})+\mathcal{E}^2(\text{quantiz})\big)^{1/2}\\
&\qquad+C\mathbb{P}\{\tau_\infty<+\infty\}\\
&\quad\le C\big(\mathcal{E}^2(\text{space})+\mathcal{E}^2(\text{quantiz})\big)+Ch\mathbb{E}\sum_{j=1}^{N}\big[\mathbf{1}_{\{t_j<\tau_\infty\}}|(\overline{u}-u)(t_{j-1},X_{t_{j-1}})|^2\big]\\
&\qquad+C\mathbb{P}\{\tau_\infty<+\infty\}.
\end{aligned}\tag{7.13}$$

Plug (7.13) in (7.11) to derive the required control for $D(1)$.

Let us now turn to the estimation of $D(2)$. Recall that $D(2)$ is given by:

$$D(2)=\mathbb{E}\sum_{j=1}^{N}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big[\Delta E_j(1)+\Delta E_j(2)+\Delta E_j(3)\big]^2\Big].$$

We first give a control of $\Delta E_j(1)$, for a given $j\in\{1,...,N\}$. To this end, note again from Lemma 6.7 (jumps of the process $X$) that:

$$\begin{aligned}
\sum_{j=1}^{N}\mathbb{E}\big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}(\Delta E_j(1))^2\big]&=\sum_{j=1}^{N}\mathbb{E}\big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|u(t_j,X_{t_j})-u(t_j,X_{t_j-})|^2\big]\\
&\le C\sum_{j=1}^{N}\mathbb{E}\big[\mathbf{1}_{\{t_j<\tau_\infty\}}|X_{t_j}-X_{t_j-}|^2\big]+C\mathbb{P}\{\tau_\infty<+\infty\}.\\
&\le Ch\big(\mathcal{E}^2(\text{space})+\mathcal{E}^2(\text{quantiz})\big)+C\mathbb{P}\{\tau_\infty<+\infty\}.
\end{aligned}\tag{7.14}$$

We recall that $\Delta E_j(2)$ and $\Delta E_j(3)$ are "Lipschitz" differences. Hence, as done in (7.10) to estimate $D(1)$, we can derive that:

$$\begin{aligned}
&\sum_{j=1}^{N}\mathbb{E}\big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big(\Delta E_j(2)+\Delta E_j(3)\big)^2\big]\\
&\quad\le Ch^2\Bigg(1+\sum_{j=1}^{N}\mathbb{E}\big[|(\overline{u}-u)(t_j,X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\\
&\quad+\sum_{j=1}^{N}\Big[\mathbb{E}\big[|(\hat{v}-v)(t_{j-1},X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]+\mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Big]\Bigg).
\end{aligned}\tag{7.15}$$

Equations (7.14) and (7.15) give the required control for $D(2)$. This completes the proof of Lemma 7.4. □

***Proof of Lemma 7.5.***

Write first:

$$
\begin{aligned}
&h\sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|\overline{v}(t_{j-1},X_{t_{j-1}})-v(t_{j-1},X_{t_{j-1}})|^2]\\
&\quad\le Ch\sum_{j=1}^{N}\Bigg\{\mathbb{E}\Bigg[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\Bigg|\overline{v}(t_{j-1},X_{t_{j-1}})-\frac{1}{h}\mathbb{E}\Big[\int_{t_{j-1}}^{t_j}\overline{Z}_s ds\big|\mathcal{F}_{t_{j-1}}\Big]\Bigg|^2\Bigg]\\
&\qquad+\mathbb{E}\Bigg[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\Bigg|\frac{1}{h}\mathbb{E}\Big[\int_{t_{j-1}}^{t_j}[\overline{Z}_s-\hat{W}_s]ds\big|\mathcal{F}_{t_{j-1}}\Big]\Bigg|^2\Bigg]\\
&\qquad+\mathbb{E}\Bigg[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\Bigg|\frac{1}{h}\mathbb{E}\Big[\int_{t_{j-1}}^{t_j}[\hat{W}_s-v(t_{j-1},X_{t_{j-1}})]ds\big|\mathcal{F}_{t_{j-1}}\Big]\Bigg|^2\Bigg]\Bigg\}\\
&\quad\equiv A(1)+A(2)+A(3).
\end{aligned}
$$

From Lemma 6.4 (distance between $Z$ and $\overline{Z}$) we then derive $A(1)\le C\mathcal{E}^2(\text{quantiz})$. For the term $A(2)$, the Cauchy-Schwarz inequality yields:

$$
A(2)\le C\sum_{j=1}^{N}\mathbb{E}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\int_{t_{j-1}}^{t_j}|\overline{Z}_s-\hat{W}_s|^2 ds\Big]=C\sum_{j=1}^{N}\mathbb{E}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big(\Delta E_j(4)\big)^2\Big]=CD(3).
$$

Concerning $A(3)$ it comes:

$$
\begin{aligned}
A(3)&\le C\sum_{j=1}^{N}\mathbb{E}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\int_{t_{j-1}}^{t_j}|\hat{W}_s-v(t_{j-1},X_{t_{j-1}})|^2 ds\Big]\\
&=C\sum_{j=1}^{N}\mathbb{E}\Big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\int_{t_{j-1}}^{t_j}|\nabla_x u(s,X_s)\sigma\big(X_{t_{j-1}},\overline{u}(t_j,X_{t_{j-1}})\big)\\
&\qquad-\nabla_x u(t_{j-1},X_{t_{j-1}})\sigma\big(X_{t_{j-1}},u(t_{j-1},X_{t_{j-1}})\big)|^2 ds\Big].
\end{aligned}
$$

Following the techniques employed in the previous proof, relying on the smoothness of the true solution, cf. Theorem 2.1, on the boundedness of the approximate solution, cf. Proposition 6.2, and on intermediate controls of the process $X$, cf. Lemma 6.5, we get:

$$
A(3)\le Ch\Big[1+\sum_{j=1}^{N}\mathbb{E}\big[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|\overline{u}(t_j,X_{t_{j-1}})-u(t_j,X_{t_{j-1}})|^2\big]\Big].
$$

The above estimates of $A(1),A(2),A(3)$ give:

$$
\begin{aligned}
D(3)\ge\ &C^{-1}h\sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|\overline{v}(t_{j-1},X_{t_{j-1}})-v(t_{j-1},X_{t_{j-1}})|^2]\\
&-Ch\sum_{j=1}^{N}\mathbb{E}[\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}|\overline{u}(t_j,X_{t_{j-1}})-u(t_j,X_{t_{j-1}})|^2]-C\mathcal{E}^2(\text{quantiz})-C\mathcal{E}^2(\text{time}).
\end{aligned}
$$

This completes the proof. □

## 7.3 Proof of Proposition 7.2 (Difference of the Gradients)

***Strategy.***

In Proposition 7.2, we aim to control the quantity $|(\hat{v}-v)(t_k,X_{t_k})|$ for $t_k<\tau_\infty$. Recall to this end (see Algorithm 3.2):

$$
\hat{v}(t_k,X_{t_k})=\mathbb{E}\big[\overline{v}\big(t_{k+1},\Pi_{k+1}(X_{t_k}+\mathcal{T}^0(t_k,X_{t_k}))\big)|\mathcal{F}_{t_k}\big].
$$

We first write $v(t_k, X_{t_k})$ in a similar way to study the difference $(\hat{v} - v)(t_k, X_{t_k})$. From Theorem 2.1 (regularity of $u$) and from the proof of Lemma 6.6 (regularity of $X$), we claim:

$$
\begin{aligned}
&\big|\mathbb{E}\big[v(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k})))|\mathcal{F}_{t_k}\big] - v(t_k, X_{t_k})\big| \\
&\quad \le C\Big[h^{1/2} + \big|\mathbb{E}\big[v(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k})))|\mathcal{F}_{t_k}\big] - v(t_{k+1}, X_{t_k})\big|\Big] \\
&\quad \le C\Big[h^{1/2} + \mathbb{E}\big[|\Pi_{k+1}(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k})) - X_{t_k}|\ |\mathcal{F}_{t_k}\big]\Big] \\
&\quad \le C\big[h^{1/2} + \delta\big].
\end{aligned}
$$

Hence:

$$
|(\hat{v} - v)(t_k, X_{t_k})| \le C\mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k}))\big)\big|\ |\mathcal{F}_{t_k}\big] + C(h^{1/2} + \delta). \tag{7.16}
$$

Proposition 7.2 directly follows from (7.16) and from the next theorem:

**Theorem 7.6** *There exists a constant $C_{7.6}$ such that on $\{t_k < \tau_\infty\}$:*

$$
\begin{aligned}
&\mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k}))\big)\big|\ |\mathcal{F}_{t_k}\big] \\
&\quad \le C_{7.6}\mathcal{E}(\text{gradient}, p) + C_{7.6}\mathbb{E}\big[|(\overline{v} - v)(t_{k+1}, X_{t_{k+1}})|^2|\mathcal{F}_{t_k}\big]^{1/2}.
\end{aligned}
$$

The main difficulty to prove Theorem 7.6 lies in the lack of regularity of $\overline{v}$. To overcome this point, note first that

$$
\mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k}))\big)\big|\ |\mathcal{F}_{t_k}\big] \tag{7.17}
$$

and

$$
\mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}(t_k, X_{t_k}))\big)\big|^2|\mathcal{F}_{t_k}\big]^{1/2} \tag{7.18}
$$

write as expectations of a given function with respect to two different kernels. We then aim to compare these underlying kernels. Recall that for a given $x \in \mathcal{C}_k$, both $\mathcal{T}^0(t_k, x)$ and $\mathcal{T}(t_k, x)$ are, up to a quantization procedure, Gaussian random variables with same covariance matrices but different means. The strategy then consists in applying a Gaussian change of variable to pass from the first kernel to the second one.

***Step One: Proof of Theorem 7.6, Exhibition of Underlying Kernels.***

We first write (7.17) with respect to the underlying kernel $\mathcal{T}^0$. Note in this frame, with the notations of Subsection 3.4, that for every $x \in \mathbb{R}^d$, $\Pi_{k+1}(x) = \Pi_{k+1} \circ \Pi_\infty(x)$ since $\Pi_\infty(x) \in \Delta_k \Leftrightarrow x \in \Delta_k$. Thus, using the invariance by translation of $\mathcal{C}_\infty$ to pass from the second to the third line, (7.17) writes:

$$
\begin{aligned}
&\mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k}))\big)\big|\ |\mathcal{F}_{t_k}\big] \\
&\quad = \mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(\Pi_\infty(X_{t_k} + \mathcal{T}^0(t_k, X_{t_k})))\big)\big|\ |\mathcal{F}_{t_k}\big] \\
&\quad = \mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(X_{t_k} + \Pi_\infty(\mathcal{T}^0(t_k, X_{t_k})))\big)\big|\ |\mathcal{F}_{t_k}\big] \\
&\quad = \sum_{y \in \mathcal{C}_\infty} \Big[\big|(\overline{v} - v)(t_{k+1}, \Pi_{k+1}(X_{t_k} + y))\big|\mathbb{P}\big\{\Pi_\infty(\mathcal{T}^0(t_k, X_{t_k})) = y|\mathcal{F}_{t_k}\big\}\Big].
\end{aligned} \tag{7.19}
$$

In the same way, the square of (7.18) writes:

$$
\begin{aligned}
&\mathbb{E}\big[\big|(\overline{v} - v)\big(t_{k+1}, \Pi_{k+1}(X_{t_k} + \mathcal{T}(t_k, X_{t_k}))\big)\big|^2\ |\mathcal{F}_{t_k}\big] \\
&\quad = \sum_{y \in \mathcal{C}_\infty} \Big[\big|(\overline{v} - v)(t_{k+1}, \Pi_{k+1}(X_{t_k} + y))\big|^2\mathbb{P}\big\{\Pi_\infty(\mathcal{T}(t_k, X_{t_k})) = y|\mathcal{F}_{t_k}\big\}\Big].
\end{aligned} \tag{7.20}
$$

Equations (7.19) and (7.20) provide relevant writings to estimate (7.17) and (7.18). Indeed, it is sufficient to bound for a given $x \in \mathcal{C}_k$ and a given $y \in \mathcal{C}_\infty$ the probability $\mathbb{P}\{\Pi_\infty(\mathcal{T}^0(t_k, x)) = y\}$ by (up to a multiplicative constant) the probability $\mathbb{P}\{\Pi_\infty(\mathcal{T}(t_k, x)) = y\}$. Recall to this end that:

$$
\begin{aligned}
&\mathcal{T}^0(t_k, x) = \sigma(x, \overline{u}(t_{k+1}, x))g(\Delta B^k), \\
&\mathcal{T}(t_k, x) = b(x, \overline{u}(t_{k+1}, x), \hat{v}(t_k, x))h + \sigma(x, \overline{u}(t_{k+1}, x))g(\Delta B^k) \\
&\hat{v}(t_k, x) = \mathbb{E}[\overline{v}(t_{k+1}, \Pi_{k+1}(x + \mathcal{T}^0(t_k, x)))].
\end{aligned}
$$

For notational convenience, we set:

$$\Sigma(t_{k+1},x) = \sigma(x,\overline{u}(t_{k+1},x)),$$
$$\mu(t_{k+1},x) = b(x,\overline{u}(t_{k+1},x),\hat{v}(t_k,x)).$$

We also introduce $\|\Sigma\|_\infty = \sup_{k\in\{0,\dots,N\}}\big[\sup_{x\in\mathcal{C}_k}|\Sigma(t_k,x)|\big]$ and $\|\mu\|_\infty = \sup_{k\in\{0,\dots,N\}}\big[\sup_{x\in\mathcal{C}_k}\ |\mu(t_k,x)|\big]$. From Assumption **(A.1)** and Propositions 6.2 and 6.3 (boundedness of $\overline{u}$ and $h^{1/2}\hat{v}$), $\|\Sigma\|_\infty + h^{1/2}\|\mu\|_\infty \le C$.

***Step Two: Proof of Theorem 7.6, Comparison of Kernels.***

The following proposition, whose proof is postponed to the end of the section, establishes a connection between the previous probabilities $\mathbb{P}\big\{\Pi_\infty(\mathcal{T}^0(t_k,x)) = y\big\}$ and $\mathbb{P}\big\{\Pi_\infty(\mathcal{T}(t_k,x)) = y\big\}$.

**Proposition 7.7** *There exists a constant $C_{7.7} > 0$ such that for every $y \in \mathcal{C}_\infty$:*

$$\mathbb{P}\big\{\Pi_\infty(\mathcal{T}^0(t_k,x)) = y\big\} \le \alpha_k(y) + \beta(y)(\mathbb{P}^{1/2}\big\{\Pi_\infty(\mathcal{T}(t_k,x)) = y\big\} + \eta_k)$$

*where*

$$\begin{aligned}
\alpha_k(y) &\equiv \mathbb{P}\big\{|\Sigma(t_{k+1},x)g(\Delta B^k) - y|_\infty \le \delta/2,\ |g(\Delta B^k) - \Delta B^k|_\infty > \delta/(2\|\Sigma\|_\infty)\big\},\\
\beta(y) &\equiv C_{7.7}\delta^{d/2}h^{-d/4}\exp\big[-C_{7.7}^{-1}h^{-1}|y|^2\big],\ \eta_k \equiv \mathbb{P}^{1/2}\big\{|g(\Delta B^k) - \Delta B^k|_\infty > \delta/(4\|\Sigma\|_\infty)\big\}.
\end{aligned}$$

*In the above expression, for all $z \in \mathbb{R}^d$, $|z|_\infty \equiv \max_{i\in\{1,\dots,d\}}|z_i|$.*

From Proposition 6.3, $h^{1/2}\overline{v}$ is bounded by a known constant. Owing to Proposition 7.7 and (7.19), we then get:

$$\begin{aligned}
&\sum_{y\in\mathcal{C}_\infty}\Big[\big[|(\overline{v}-v)(t_{k+1},\Pi_{k+1}(x+y))|\big]\mathbb{P}\big\{\Pi_\infty(\mathcal{T}^0(t_k,x)) = y\big\}\Big]\\
&\quad\le Ch^{-1/2}\mathbb{P}\big\{|g(\Delta B^k) - \Delta B^k|_\infty > \delta/(2\|\Sigma\|_\infty)\big\}\\
&\qquad + C\delta^{d/2}h^{-d/4-1/2}\mathbb{P}^{1/2}\big\{|g(\Delta B^k) - \Delta B^k|_\infty > \delta/(4\|\Sigma\|_\infty)\big\}\Big[\sum_{y\in\mathcal{C}_\infty}\exp\big[-C^{-1}h^{-1}|y|^2\big]\Big]\\
&\qquad + C\Big[\sum_{y\in\mathcal{C}_\infty}\big[\delta^d h^{-d/2}\exp\big[-C^{-1}h^{-1}|y|^2\big]\big]\Big]^{1/2}\\
&\qquad\quad\times\Big[\sum_{y\in\mathcal{C}_\infty}|(\overline{v}-v)(t_{k+1},\Pi_{k+1}(x+y))|^2\mathbb{P}\big\{\Pi_\infty(\mathcal{T}(t_k,x)) = y\big\}\Big]^{1/2}\\
&\quad\equiv T(1) + T(2) + T(3).
\end{aligned} \tag{7.21}$$

Thanks to (3.10) and to the Bienaymé-Chebychev inequality:

$$T(1) \le Ch^{p/2-1/2}\delta^{-p}M^{-p/d}. \tag{7.22}$$

Thanks again to (3.10) (applied to the exponent $2p$):

$$T(2) \le Ch^{p/2-d/4-1/2}\delta^{-p+d/2}M^{-p/d}\Big[\sum_{y\in\mathcal{C}_\infty}\exp\big[-C^{-1}h^{-1}|y|^2\big]\Big]. \tag{7.23}$$

Note now from (7.20) that:

$$\begin{aligned}
T(3) = C&\Big[\sum_{y\in\mathcal{C}_\infty}\big[\delta^d h^{-d/2}\exp\big[-C^{-1}h^{-1}|y|^2\big]\big]\Big]^{1/2}\\
&\times\mathbb{E}\big[|(\overline{v}-v)(t_{k+1},\Pi_{k+1}(x+\mathcal{T}(t_k,x)))|^2\big]^{1/2}.
\end{aligned} \tag{7.24}$$

Since $h^{-1}\delta^2$ is small, note now that:

$$\sum_{y\in\mathcal{C}_\infty}(\delta h^{-1/2})^d\exp(-C^{-1}h^{-1}|y|^2)$$
$$=\left(\sum_{j\in\mathbb{Z}}\delta h^{-1/2}\exp(-C^{-1}j^2\delta^2h^{-1})\right)^d\le C'\left(\int_{\mathbb{R}}\delta h^{-1/2}\exp(-C^{-1}z^2\delta^2h^{-1})dz\right)^d\le C''.$$

Hence:

$$\sum_{y\in\mathcal{C}_\infty}\exp\big[-C^{-1}h^{-1}|y|^2\big]\le C''\delta^{-d}h^{d/2}. \tag{7.25}$$

From (7.21), (7.22), (7.23), (7.24) and (7.25), we complete the proof of Theorem 7.6 (recall again that $h^{-1}\delta^2$ is small to dominate $T(1)$ by $\mathcal{E}(\text{gradient},p)$). □

***Proof of Proposition 7.7, Gaussian Change of Variable.***

It now remains to prove Proposition 7.7. Note first that:

$$\begin{aligned}
&\mathbb{P}\big\{\Pi_\infty(\mathcal{T}^0(t_k,x))=y\big\}\\
&\quad\le\mathbb{P}\big\{|\mathcal{T}^0(t_k,x)-y|_\infty\le\delta/2\big\}\\
&\quad=\mathbb{P}\big\{|\Sigma(t_{k+1},x)g(\Delta B^k)-y|_\infty\le\delta/2\big\}\\
&\quad=\mathbb{P}\big\{|\Sigma(t_{k+1},x)g(\Delta B^k)-y|_\infty\le\delta/2,\ |g(\Delta B^k)-\Delta B^k|_\infty>\delta/(2\|\Sigma\|_\infty)\big\}\\
&\qquad+\mathbb{P}\big\{|\Sigma(t_{k+1},x)g(\Delta B^k)-y|_\infty\le\delta/2,\ |g(\Delta B^k)-\Delta B^k|_\infty\le\delta/(2\|\Sigma\|_\infty)\big\}\\
&\quad\equiv P(1)+P(2).
\end{aligned}\tag{7.26}$$

Note that $P(1)$ matches $\alpha_k(y)$ in Proposition 7.7. Hence, we just have to estimate $P(2)$:

$$\begin{aligned}
P(2)&=\mathbb{P}\big\{|\Sigma(t_{k+1},x)g(\Delta B^k)-y|_\infty\le\delta/2,\ |g(\Delta B^k)-\Delta B^k|_\infty\le\delta/(2\|\Sigma\|_\infty)\big\}\\
&\le\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k-y|_\infty\le\delta\big\}.
\end{aligned}\tag{7.27}$$

The strategy is now clear. A Gaussian change of variable permits to introduce artificially the drift appearing in the definition of $\mathcal{T}(t_k,x)$. It comes:

$$\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k-y|_\infty\le\delta\big\}=\mathbb{E}^{\mathbb{Q}_k}\left[\frac{d\mathbb{P}}{d\mathbb{Q}_k}\mathbf{1}_{\{|\Sigma(t_{k+1},x)\Delta B^k-y|_\infty\le\delta\}}\right]$$

where $\dfrac{d\mathbb{Q}_k}{d\mathbb{P}}=\exp\left(\langle(\Sigma^{-1}\mu)(t_{k+1},x),\Delta B^k\rangle-\dfrac{|(\Sigma^{-1}\mu)(t_{k+1},x)|^2h}{2}\right)$. Under $\mathbb{Q}_k$, $\Delta B^k$ is a $d$-dimensional Gaussian random variable with mean $(\Sigma^{-1}\mu)(t_{k+1},x)h$ and covariance matrix $h\mathbf{I}_d$. Hence,

$$\begin{aligned}
&\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k-y|_\infty\le\delta\big\}\\
&\quad=\mathbb{E}\left[\exp\left(-\langle(\Sigma^{-1}\mu)(t_{k+1},x),\Delta B^k\rangle-\frac{|(\Sigma^{-1}\mu)(t_{k+1},x)|^2h}{2}\right)\mathbf{1}_{\{|\Sigma(t_{k+1},x)\Delta B^k+\mu(t_{k+1},x)h-y|_\infty\le\delta\}}\right]\\
&\quad\le C\exp\big[C(1+h^{-1/2}|y|)\big]\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k+\mu(t_{k+1},x)h-y|_\infty\le\delta\big\},
\end{aligned}$$

where the last identity follows from Assumption **(A)** and Propositions 6.2 and 6.3.

We deduce from (7.27) and the above control that:

$$P(2)\le C\exp\big[C(1+h^{-1/2}|y|)\big]\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k+\mu(t_{k+1},x)h-y|_\infty\le\delta\big\}. \tag{7.28}$$

It now remains to quantify the Gaussian increment in the latter quantity with the converse procedure. The first step is to replace the upper bound $\delta$ in (7.28) by $\delta/4$. Note to this end that, for a given $y\in\mathcal{C}_\infty$, we can find a finite subset $(y_i)_{1\le i\le N(d)}$ included in $\mathbb{R}^d$, $N(d)=4^d$, such that:

$$\begin{aligned}
&\forall i\in\{1,...,N(d)\},\ |y-y_i|_\infty\le\delta,\\
&\big\{|\Sigma(t_{k+1},x)\Delta B^k+\mu(t_{k+1},x)h-y|_\infty\le\delta\big\}=\bigcup_{i=1}^{N(d)}\big\{|\Sigma(t_{k+1},x)\Delta B^k+\mu(t_{k+1},x)h-y_i|_\infty\le\delta/4\big\}.
\end{aligned}$$

Hence:

$$\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty \le \delta\big\} \le \sum_{i=1}^{N(d)} \mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y_i|_\infty \le \delta/4\big\}. \quad (7.29)$$

In order to recover $y$ instead of $y_i$ for each term of the sum in the r.h.s. of the above expression, we use once again a Gaussian change of variable. Namely, $\forall i \in \{1, ..., N(d)\}$, put:

$$\frac{d\mathbb{Q}_k(i)}{d\mathbb{P}} = \exp\left(\langle \Sigma^{-1}(t_{k+1},x)(y_i - y), \Delta B^k\rangle - \frac{|\Sigma^{-1}(t_{k+1},x)(y_i-y)|^2 h}{2}\right).$$

Set $G_k(i) = \Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y_i$. Under $\mathbb{Q}_k(i)$, $G_k(i) \sim \mathcal{N}(\mu(t_{k+1},x)h - y, \Sigma\Sigma^*(t_{k+1},x)h)$. Hence, from (7.29):

$$\begin{aligned}
&\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty \le \delta\big\}\\
&\quad\le \sum_{i=1}^{N(d)} \mathbb{E}\bigg[\exp\left(-\langle \Sigma^{-1}(t_{k+1},x)(y_i - y), \Delta B^k\rangle - \frac{|\Sigma^{-1}(t_{k+1},x)(y_i-y)|^2 h}{2}\right)\\
&\qquad\qquad \times \mathbf{1}_{\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty \le \delta/4\}}\bigg].
\end{aligned}$$

Arguments similar to the ones used at the preceding Gaussian change of variable combined with the inequality $|y - y_i|_\infty \le \delta$ then give:

$$\begin{aligned}
&\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty \le \delta\big\}\\
&\quad\le CN(d)\exp(C\delta|y|)\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty \le \delta/4\big\}.
\end{aligned} \quad (7.30)$$

Write now:

$$\begin{aligned}
&\mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty \le \delta/4\big\}\\
&\quad= \mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty < \delta/4\big\}\\
&\quad\le \mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty < \delta/4,\ |\Delta B^k - g(\Delta B^k)|_\infty \le \delta/(4\|\Sigma\|_\infty)\big\}\\
&\qquad\quad + \mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty < \delta/4,\ |\Delta B^k - g(\Delta B^k)|_\infty > \delta/(4\|\Sigma\|_\infty)\big\}\\
&\quad\le \mathbb{P}\big\{|\Sigma(t_{k+1},x)g(\Delta B^k) + \mu(t_{k+1},x)h - y|_\infty < \delta/2,\ |\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty < \delta/4\big\}\\
&\qquad\quad + \mathbb{P}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty < \delta/4,\ |\Delta B^k - g(\Delta B^k)|_\infty > \delta/(4\|\Sigma\|_\infty)\big\}\\
&\quad\equiv P(3) + P(4).
\end{aligned} \quad (7.31)$$

Note first from the Cauchy-Schwarz inequality, Assumption **(A)** and the explicit expression of the Gaussian kernel:

$$\begin{aligned}
P(3) &\le \mathbb{P}^{1/2}\big\{|\Sigma(t_{k+1},x)g(\Delta B^k) + \mu(t_{k+1},x)h - y|_\infty < \delta/2\big\}\\
&\qquad \times \mathbb{P}^{1/2}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty < \delta/4\big\}\\
&\le C\mathbb{P}^{1/2}\big\{\Pi_\infty(\mathcal{T}(t_k,x)) = y\big\}\\
&\qquad \times \left(\int_{B_\infty(0,\delta/4)} \exp\left(-h^{-1}|\Sigma^{-1}(t_{k+1},x)(z - (\mu(t_{k+1},x)h - y))|^2\right)\frac{dz}{(2\pi h)^{d/2}}\right)^{1/2}\\
&\le C\big[\delta^d h^{-d/2}\exp\big[-C^{-1}h^{-1}|y|^2\big]\big]^{1/2}\mathbb{P}^{1/2}\big\{\Pi_\infty(\mathcal{T}(t_k,x)) = y\big\}.
\end{aligned} \quad (7.32)$$

From the same arguments, we also derive:

$$\begin{aligned}
P(4) &\le \mathbb{P}^{1/2}\big\{|\Sigma(t_{k+1},x)\Delta B^k + \mu(t_{k+1},x)h - y|_\infty < \delta/4\big\}\mathbb{P}^{1/2}\big\{|\Delta B^k - g(\Delta B^k)|_\infty > \delta/(4\|\Sigma\|_\infty)\big\}\\
&\le C\big[\delta^d h^{-d/2}\exp\big[-C^{-1}h^{-1}|y|^2\big]\big]^{1/2}\mathbb{P}^{1/2}\big\{|\Delta B^k - g(\Delta B^k)|_\infty > \delta/(4\|\Sigma\|_\infty)\big\}.
\end{aligned} \quad (7.33)$$

Thus, from (7.32) and (7.33):

$$P(3) + P(4) \le \beta(y)\big(\mathbb{P}^{1/2}\big\{\Pi_\infty(\mathcal{T}(t_k,x)) = y\big\} + \eta_k\big). \quad (7.34)$$

Apply Young's inequality to the product $h^{-1/2}|y|$ and $\delta|y|$ in the exponential terms in (7.28) and (7.30) and deduce finally from (7.28), (7.30), (7.31) and (7.34) that:

$$P(2) \le C\delta^{d/2}h^{-d/4}\exp\big[-C^{-1}h^{-1}|y|^2\big]\big[\mathbb{P}^{1/2}\{\Pi_\infty(\mathcal{T}(t_k,x))=y\}$$
$$+\,\mathbb{P}^{1/2}\big\{|\Delta B^k - g(\Delta B^k)|_\infty > \delta/(4\|\Sigma\|_\infty)\big\}\big].$$

Thanks to (7.26), this completes the proof of Proposition 7.7. □

# 8 Proof, Third Step: Gronwall's Lemma

This section is devoted to the final step of the proof of Theorems 4.1, 4.2 and 4.3. For notational convenience, we set:

$$\mathcal{E}^2(\mathrm{global}) \equiv \mathcal{E}^2(\mathrm{time}) + \mathcal{E}^2(\mathrm{space}) + \mathcal{E}^2(\mathrm{trunc}) + \mathcal{E}^2(\mathrm{quantiz}) + \mathcal{E}^2(\mathrm{gradient}, p).$$

## 8.1 Proof of Theorem 4.1, Infinite Grids

We first explain how to derive Theorem 4.1 from Theorem 7.3 when $\rho = +\infty$, i.e. when $\tau_\infty = +\infty$ a.s.. In this framework, the term $\mathcal{E}^2(\mathrm{trunc})$ in $\mathcal{E}^2(\mathrm{global})$ reduces to 0. The general case is detailed in the next subsection. For infinite grids, we obtain from Theorem 7.3:

$$\begin{aligned}
&|(\overline{u}-u)(0,x_0)|^2 + C^{-1}h\sum_{j=1}^N \mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2\big]\\
&\le C\Big[\mathcal{E}^2(\mathrm{global}) + \eta^{-1}h\sum_{j=1}^N \mathbb{E}\big[|(\overline{u}-u)(t_j,X_{t_{j-1}})|^2\big] + \eta^{-1}h\sum_{j=1}^N \mathbb{E}\big[|(\overline{u}-u)(t_{j-1},X_{t_{j-1}})|^2\big]\\
&\quad + (\eta+h)h\sum_{j=1}^N \mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_j})|^2\big]\Big].
\end{aligned}$$

Note that, for every $x \in \mathcal{C}_N$, $\overline{v}(T,x) = v(T,x)$. Hence, choose $\eta$ and $h$ small enough to deduce that:

$$|(\overline{u}-u)(0,x_0)|^2 + C^{-1}h\sum_{j=1}^N \mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2\big] \le C\Big[\mathcal{E}^2(\mathrm{global}) + h\sum_{j=0}^N \sup_{x\in\mathcal{C}_j}|(\overline{u}-u)(t_j,x)|^2\Big]. \quad (8.1)$$

As usual in BSDE theory, note that the estimate (8.1) holds actually for any starting point $(t_k,x)$, $0 \le k \le N$, $x \in \mathcal{C}_k$. Hence:

$$\sup_{x\in\mathcal{C}_k}|(\overline{u}-u)(t_k,x)|^2 \le C\Big[\mathcal{E}^2(\mathrm{global}) + h\sum_{j=k}^N \sup_{x\in\mathcal{C}_j}|(\overline{u}-u)(t_j,x)|^2\Big].$$

Since $h$ is small, for every $k \in \{0,...,N-1\}$:

$$\sup_{x\in\mathcal{C}_k}|(\overline{u}-u)(t_k,x)|^2 \le C\Big[\mathcal{E}^2(\mathrm{global}) + h\sum_{j=k+1}^N \sup_{x\in\mathcal{C}_j}|(\overline{u}-u)(t_j,x)|^2\Big].$$

Apply now Gronwall's Lemma and deduce that:

$$\sup_{x\in\mathcal{C}_0}|\overline{u}(0,x)-u(0,x)|^2 \le C\mathcal{E}^2(\mathrm{global}).$$

This completes the proof of Theorem 4.1 when $\rho = +\infty$. □

## 8.2 Proof of Theorem 4.1, General Case

We now turn to the case of truncated grids. Generally speaking, most of the approach given in the former subsection still applies in the general framework. It is however impossible to mimic word for word the arguments given above and we need to refine the previous Gronwall argument.

***First Step.*** We first aim to get rid of the difference $\overline{v}-v$ appearing in the r.h.s. of (7.2). Due to the functions $(\mathbf{1}_{\{t_{j-1}<\tau_\infty\}})_{j=1,\dots,N}$, the machinery used in the previous subsection does not apply. To overcome this difficulty, we write $\{t_{j-1}<\tau_\infty\}=\{t_j<\tau_\infty\}\cup\{t_j=\tau_\infty\}$. Indeed, since $\overline{v}(T,x)=v(T,x)$ for $x\in\mathcal{C}_N$ and $h|\overline{v}-v|^2$ is bounded (see Theorem 2.1 and Proposition 6.3), note that:

$$\begin{aligned}
&h\sum_{j=1}^{N}\mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_j})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\\
&\quad=h\sum_{j=1}^{N-1}\mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_j})|^2\mathbf{1}_{\{t_j<\tau_\infty\}}\big]+h\sum_{j=1}^{N-1}\mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_j})|^2\mathbf{1}_{\{t_j=\tau_\infty\}}\big]\\
&\quad\le h\sum_{j=2}^{N}\mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]+C\mathbb{P}\{\tau_\infty<+\infty\}.
\end{aligned}\tag{8.2}$$

Hence, (7.2) and (8.2) give for $\eta$ and $h$ small enough:

$$\begin{aligned}
&|(\overline{u}-u)(0,x_0)|^2+C^{-1}h\sum_{j=1}^{N}\mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\\
&\quad\le C\Big[\mathbb{P}\{\tau_\infty<+\infty\}+\mathcal{E}^2(\text{global})+h\sum_{j=1}^{N}\mathbb{E}\big[|(\overline{u}-u)(t_j,X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\\
&\qquad+h\sum_{j=2}^{N}\mathbb{E}\big[|(\overline{u}-u)(t_{j-1},X_{t_{j-1}})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Big].
\end{aligned}\tag{8.3}$$

The term $\mathcal{E}^2(\text{time})+\mathcal{E}^2(\text{space})+\mathcal{E}^2(\text{quantiz})+\mathcal{E}^2(\text{gradient},p)$ appearing in (7.2) has been replaced, for notational convenience, by $\mathcal{E}^2(\text{global})$ in the above reference. Nevertheless, mention carefully that the origin of the term $\mathcal{E}^2(\text{trunc})$ has not been explained yet. It is in the following lines.

***Second Step.*** Note that (8.3) still holds if $X$ starts at a given time $t_i$, $i\in\{0,\dots N\}$, from an $\mathcal{F}_{t_i}$-measurable and square integrable random variable $\xi$ with values in $\mathcal{C}_i$. In such a case:

$$\begin{aligned}
&\mathbb{E}\big[|(\overline{u}-u)(t_i,\xi)|^2\big]\\
&\quad\le C\Big[\mathbb{P}\{\tau_\infty^{t_i,\xi}<+\infty\}+\mathcal{E}^2(\text{global})+h\sum_{j=i+1}^{N}\mathbb{E}\big[|(\overline{u}-u)(t_j,X_{t_{j-1}}^{t_i,\xi})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty^{t_i,\xi}\}}\big]\\
&\qquad+h\sum_{j=i+2}^{N}\mathbb{E}\big[|(\overline{u}-u)(t_{j-1},X_{t_{j-1}}^{t_i,\xi})|^2\mathbf{1}_{\{t_{j-1}<\tau_\infty^{t_i,\xi}\}}\big]\Big],
\end{aligned}\tag{8.4}$$

where the superscript $(t_i,\xi)$ denotes the initial condition of the process $X$. Due to the shift between $t_{j-1}$ and $t_j$ in the r.h.s., there is no possible choice of $\xi$ to recover the same form of terms in the left and right hand sides. In particular, Gronwall's Lemma does not apply at this stage of the proof. Note in fact that the same problem occured in Subsection 8.1: this was the reason why the supremum was taken in the r.h.s. of (8.1). To mimic this strategy, choose $\xi=x\in\mathcal{C}_i$ and take the supremum over these $x$:

$$\sup_{x\in\mathcal{C}_i}|(\overline{u}-u)(t_i,x)|^2\le C\Big[\sup_{x\in\mathcal{C}_i}\mathbb{P}\{\tau_\infty^{t_i,x}<+\infty\}+\mathcal{E}^2(\text{global})+h\sum_{j=i+1}^{N}\sup_{x\in\mathcal{C}_j}|(\overline{u}-u)(t_j,x)|^2\Big].\tag{8.5}$$

It is then readily seen that $\mathbb{P}\{\tau_\infty^{t_i,x}<+\infty\}$ is far from being negligible when $x$ tends to the boundary of the grid $\mathcal{C}_i$. In particular, there is no hope to prove Theorem 4.1 in the case $\rho<+\infty$ with the arguments used in Subsection 8.1.

***Strategy.*** The reason why the latter method fails is rather clear: the way the supremum is taken in (8.5) is too rough to be really efficient. Our strategy then consists in applying (8.4) to a suitable choice of $\xi$. We then have to estimate the probability $\mathbb{P}\{\tau_\infty^{t_i,\xi}<+\infty\}$ for a random initial condition $(t_i,\xi)$, $\xi\in L^2(\Omega,\mathcal{F}_{t_i},\mathbb{P})$ with values in $\mathcal{C}_i$. To this end, we need to control efficiently the tails of the variables $(X_{t_j\wedge\tau_\infty^{t_i,\xi}}^{t_i,\xi})_{i\le j\le N}$. Since the drift $b$ is not bounded, the best we can do consists in estimating the $L^2$ norms of these variables.

*An $L^2$ Control of the Process $X$.*

The following lemma is proved at the end of the subsection:

**Lemma 8.1** *For all $k \in \{0,...,N\}$, put $\tau_k = \tau_\infty \wedge t_k$. Then, there exists a constant $C_{8.1}$ such that for all $i \in \{0,...,N\}$ and $\xi \in L^2(\Omega, \mathcal{F}_{t_i}, \mathbb{P})$ with values in $\mathcal{C}_i$:*

$$\forall k \in \{i,...,N\},\ \mathbb{E}\big[|X^{t_i,\xi}_{\tau_k}|^2\big] \le C_{8.1}\big[\mathbb{E}|\xi|^2 + 1 + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{gradient}, p)\big].$$

*Estimate of the Probability of Hitting the Boundary.*

Thanks to the previous lemma, we are now able to estimate the probability $\mathbb{P}\{\tau^{t_i,\xi}_\infty < +\infty\}$, with $(i,\xi)$ as in Lemma 8.1. Indeed, $\{\tau^{t_i,\xi}_\infty < +\infty\} \subset \{|X^{t_i,\xi}_{\tau_N}|_\infty + \delta \ge R + \rho\}$. Thanks to the Bienaymé-Chebychev inequality and to Lemma 8.1 (with $k = N$), it comes:

$$\begin{aligned}\mathbb{P}\{\tau^{t_i,\xi}_\infty < +\infty\} &\le C(R+\rho)^{-2}\big[\mathbb{E}\big[|X^{t_i,\xi}_{\tau_N}|^2\big] + \delta^2\big] \\ &\le C\big[(R+\rho)^{-2}\mathbb{E}[|\xi|^2] + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{trunc}) + \mathcal{E}^2(\text{gradient}, p)\big].\end{aligned} \tag{8.6}$$

Plug now (8.6) into (8.4) to obtain:

$$\begin{aligned}&\mathbb{E}\big[|(\overline{u} - u)(t_i, \xi)|^2\big] \\ &\quad\le C\Bigg[(R+\rho)^{-2}\mathbb{E}[|\xi|^2] + \mathcal{E}^2(\text{global}) + h\sum_{j=i+1}^{N}\mathbb{E}\big[|(\overline{u} - u)(t_j, X^{t_i,\xi}_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau^{t_i,\xi}_\infty\}}\big] \\ &\qquad + h\sum_{j=i+2}^{N}\mathbb{E}\big[|(\overline{u} - u)(t_{j-1}, X^{t_i,\xi}_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau^{t_i,\xi}_\infty\}}\big]\Bigg].\end{aligned} \tag{8.7}$$

*A Refined Gronwall Argument.*

The key idea is to find by induction a sequence of constants $c_i(1)$, $c_i(2)$, $i \in \{0,...,N\}$, such that for any $\xi \in L^2(\Omega, \mathcal{F}_{t_i}, \mathbb{P})$ with values in $\mathcal{C}_i$:

$$\mathbb{E}\big[|(\overline{u} - u)(t_i, \xi)|^2\big] \le c_i(1)\mathcal{E}^2(\text{global}) + c_i(2)(R+\rho)^{-2}\mathbb{E}[|\xi|^2]. \tag{8.8}$$

If we can exhibit two sequences $c_i(1)$ and $c_i(2)$, $i \in \{0,...,N\}$, satisfying (8.8) and uniformly bounded by a constant $C$, then (8.8) yields:

$$\mathbb{E}\big[|(\overline{u} - u)(t_i, \xi)|^2\big] \le C\big(\mathcal{E}^2(\text{global}) + (R+\rho)^{-2}\mathbb{E}[|\xi|^2]\big). \tag{8.9}$$

Choosing $i = 0$ and $\xi = x_0 \in \mathcal{C}_0$, we then complete the proof of Theorem 4.1.

*Construction of the Sequences $c_i(1), c_i(2)$.*

For $i = N$, we put: $c_N(1) = c_N(2) = 0$.

Consider now a given $i \in \{0,...,N-1\}$ and assume that $c_j(1)$ and $c_j(2)$ are known for $j = i+1,...,N$. Note first from (8.8), with $i = j$ and $\xi = X^{t_i,\xi}_{t_{j-1}}$, and from Lemma 8.1, with $k = j-1$, that:

$$\begin{aligned}\mathbb{E}\big[|(\overline{u} - u)(t_j, X^{t_i,\xi}_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau^{t_i,\xi}_\infty\}}\big] &\le c_j(1)\mathcal{E}^2(\text{global}) + c_j(2)(R+\rho)^{-2}\mathbb{E}\big[|X^{t_i,\xi}_{\tau_{j-1}}|^2\big] \\ &\le c_j(1)\mathcal{E}^2(\text{global}) + c_j(2)C_{8.1}(R+\rho)^{-2}\big[\mathbb{E}[|\xi|^2] + 1 + \mathcal{E}^2(\text{global})\big] \\ &\le \big(c_j(1) + 2c_j(2)C_{8.1}\big)\mathcal{E}^2(\text{global}) + c_j(2)C_{8.1}(R+\rho)^{-2}\mathbb{E}[|\xi|^2].\end{aligned} \tag{8.10}$$

Note that the same holds for $\mathbb{E}[|(\overline{u} - u)(t_{j-1}, X^{t_i,\xi}_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau^{t_i,\xi}_\infty\}}]$ with respect to $c_{j-1}(1)$ and $c_{j-1}(2)$, $j \in \{i+2,...,N\}$. Thus, from (8.7) and (8.10):

$$\begin{aligned}\mathbb{E}\big[|(\overline{u} - u)(t_i, \xi)|^2\big] \le C\Bigg[&(R+\rho)^{-2}\mathbb{E}[|\xi|^2] + \mathcal{E}^2(\text{global}) \\ &+ 2h\sum_{j=i+1}^{N}\big[\big(c_j(1) + 2c_j(2)C_{8.1}\big)\mathcal{E}^2(\text{global}) + c_j(2)C_{8.1}(R+\rho)^{-2}\mathbb{E}[|\xi|^2]\big]\Bigg]\end{aligned} \tag{8.11}$$

Set:

$$c_i(1) = C + 2Ch \sum_{j=i+1}^{N} \big[c_j(1) + 2C_{8.1} c_j(2)\big], \ c_i(2) = C + 2CC_{8.1} h \sum_{j=i+1}^{N} \big[c_j(2)\big]. \tag{8.12}$$

Gronwall's Lemma directly proves that the second sequence is bounded. This allows to derive the boundedness of the first sequence. □

It now remains to prove Lemma 8.1.

***Proof of Lemma 8.1.*** We remove the superscript $(t_i, \xi)$ in the writing of $X$. Then:

$$\begin{aligned}
X_{\tau_k} &= \xi + \sum_{j=i}^{k-1} \big[(X_{t_{j+1}} - X_{t_j}) \mathbf{1}_{\{t_j < \tau_\infty\}}\big] \\
&= \xi + \sum_{j=i}^{k-1} \big[\mathcal{T}(t_j, X_{t_j}) \mathbf{1}_{\{t_j < \tau_\infty\}}\big] \\
&\quad + \sum_{j=i}^{k-1} \big[\big(\Pi_{j+1}(X_{t_j} + \mathcal{T}(t_j, X_{t_j})) - X_{t_j} - \mathcal{T}(t_j, X_{t_j})\big) \mathbf{1}_{\{t_{j+1} < \tau_\infty\}}\big] \\
&\quad + \sum_{j=i}^{k-1} \big[\big(\Pi_{j+1}(X_{t_j} + \mathcal{T}(t_j, X_{t_j})) - X_{t_j} - \mathcal{T}(t_j, X_{t_j})\big) \mathbf{1}_{\{t_{j+1} = \tau_\infty\}}\big] \\
&\equiv \xi + S(1) + S(2) + S(3).
\end{aligned} \tag{8.13}$$

Deal first with $S(2)$ and $S(3)$. Note to this end that the distance between $X_{t_j} + \mathcal{T}(t_j, X_{t_j})$ and $\Pi_{j+1}(X_{t_j} + \mathcal{T}(t_j, X_{t_j}))$ is bounded by $\delta$ on $\{t_{j+1} < \tau_\infty\}$. Deduce in particular that $\mathbb{E}[|S(2)|^2] \le \delta^2 (k-i)^2 \le C\mathcal{E}^2(\text{space})$. Note moreover that the distance between $X_{t_j} + \mathcal{T}(t_j, X_{t_j})$ and $\Pi_{j+1}(X_{t_j} + \mathcal{T}(t_j, X_{t_j}))$ is always bounded by $2|\mathcal{T}(t_j, X_{t_j})| + \delta$ (see Lemma 6.6). Thus, applying the Cauchy-Schwarz inequality, it is readily seen that:

$$\begin{aligned}
\mathbb{E}\big[|S(3)|^2\big] &\le C\delta^2 + C \sum_{j=i}^{k-1} \mathbb{E}\big[|\mathcal{T}(t_j, X_{t_j})|^2 \mathbf{1}_{\{t_{j+1} = \tau_\infty\}}\big] \\
&\le Ch^2 \mathcal{E}^2(\text{space}) + C\mathbb{P}\{\tau_\infty < +\infty\} + C \sum_{j=i}^{k-1} \mathbb{E}\big[|\mathcal{T}(t_j, X_{t_j})|^4\big].
\end{aligned} \tag{8.14}$$

From Propositions 6.2 and 6.3, we can prove that $\mathbb{E}[|\mathcal{T}(t_j, X_{t_j})|^4] \le Ch^2$. We finally deduce ($h$ being small):

$$\mathbb{E}\big[|S(2)|^2\big] + \mathbb{E}\big[|S(3)|^2\big] \le C\big[\mathbb{P}\{\tau_\infty < +\infty\} + \mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space})\big]. \tag{8.15}$$

A simpler strategy in (8.14) consists in bounding $\mathbb{E}[|\mathcal{T}(t_j, X_{t_j})|^2]$ by $Ch$ and thus $\mathbb{E}[|S(3)|^2]$ by $C(1+\mathcal{E}^2(\text{space}))$. However, the control obtained in (8.14) turns out to be useful in the next subsection.

It now remains to deal with $S(1)$. The strategy is slightly different: thanks to Propositions 6.3 and 7.2, we are able to estimate the drift, and thanks to the independence of the Brownian increments, we easily bound the martingale part. From Assumption **(A)**, we deduce that there exists a constant $C$ such that (recall that $h$ is small):

$$\mathbb{E}\big[|S(1)|^2\big] \le Ch(k-i) \left[1 + h \sum_{j=i}^{k-1} \mathbb{E}\big[|\hat{v}(t_j, X_{t_j})|^2 \mathbf{1}_{\{t_j < \tau_\infty\}}\big]\right].$$

Apply now Proposition 7.2 concerning the difference $\hat{v} - v$. Since $v$ is bounded, deduce that (recall that $\mathcal{E}^2(\text{time}) = h$ and $h^2 \mathcal{E}^2(\text{space}) = \delta^2$ are small):

$$\mathbb{E}\big[|S(1)|^2\big] \le Ch(k-i) \left[1 + \mathcal{E}^2(\text{gradient}, p) + h \sum_{j=i}^{k-1} \mathbb{E}\big[|\overline{v}(t_{j+1}, X_{t_{j+1}})|^2 \mathbf{1}_{\{t_j < \tau_\infty\}}\big]\right]. \tag{8.16}$$

Recall now from Proposition 6.3 that $h \sum_{j=i}^{N-1} \mathbb{E}[|\overline{v}(t_{j+1}, X_{t_{j+1}})|^2]$ is bounded.

Thus, from (8.16), we claim:

$$\mathbb{E}\big[|S(1)|^2\big] \le C\big[1 + \mathcal{E}^2(\text{gradient}, p)\big]. \tag{8.17}$$

Thanks to (8.13), (8.15) and (8.17), we complete the proof of Lemma 8.1. □

## 8.3 Proofs of Theorems 4.2 and 4.3

We turn to the proof of Theorems 4.2 and 4.3. The initial condition of the process $X$ is given by $X_0 = x_0$, $x_0 \in \mathcal{C}_0$, as in (3.14).

***Proof of Theorem 4.2.*** From inequalities (8.3) (deriving from the stability theorem), (8.6) (probability of hitting the boundary of the grids) and (8.10) (estimate of $\overline{u} - u$, recall that $c_j(1)$, $c_j(2)$, $j \in \{0, ..., N\}$, are uniformly bounded), Theorem 4.2 holds with $v(t_i, X_{t_i})\mathbf{1}_{\{t_i < \tau_\infty\}}$ instead of $v(t_i, X_{t_i})$. Since $v$ is bounded (see Theorem 2.1) and since the probability of hitting the boundaries of the grids is controlled (see again (8.6)), we easily complete the proof of Theorem 4.2. □

***Proof of Theorem 4.3.*** It just remains to study the convergence of $(X_{t_k}, Y_{t_k}, Z_{t_k})_{0 \le t_k \le \tau_\infty \wedge T}$ towards the solution $(U, V, W)$ of $(E)$. Thanks to the Lipschitz properties of $b$ and $\sigma$, we first deduce by standard computations the analogue of Proposition 7.1. Recall to this end that, for all $k \in \{0, ..., N\}$, $\tau_k = \tau_\infty \wedge t_k$.

**Proposition 8.2** *There exists a constant $C_{8.2}$ such that for $k \in \{1, ..., N\}$:*

$$\begin{aligned}
\mathbb{E}|X_{\tau_k} - U_{\tau_k}|^2 \le C_{8.2} \Bigg[ & \mathbb{P}\{\tau_\infty < +\infty\} + \mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz}) \\
& + h \sum_{j=0}^{k-1} \Big[ \mathbb{E}\big[|X_{t_j} - U_{t_j}|^2 \mathbf{1}_{\{t_j < \tau_\infty\}}\big] + \mathbb{E}\big[|(\overline{u} - u)(t_{j+1}, X_{t_j})|^2 \mathbf{1}_{\{t_j < \tau_\infty\}}\big] \Big] \\
& + h \sum_{j=0}^{k-1} \mathbb{E}\big[|(\hat{v} - v)(t_j, X_{t_j})|^2 \mathbf{1}_{\{t_j < \tau_\infty\}}\big] \Bigg].
\end{aligned} \tag{8.18}$$

**Sketch of the Proof.** Note that $X_{\tau_k} - U_{\tau_k}$ writes:

$$\begin{aligned}
X_{\tau_k} - U_{\tau_k} &= \sum_{j=0}^{k-1} \big[[(X_{t_{j+1}} - U_{t_{j+1}}) - (X_{t_j} - U_{t_j})]\mathbf{1}_{\{t_j < \tau_\infty\}}\big] \\
&= \sum_{j=0}^{k-1} \big[[\mathcal{T}(t_j, X_{t_j}) - (U_{t_{j+1}} - U_{t_j})]\mathbf{1}_{\{t_j < \tau_\infty\}}\big] \\
&\quad + \sum_{j=0}^{k-1} \big[[\Pi_{j+1}(X_{t_j} + \mathcal{T}(t_j, X_{t_j})) - (X_{t_j} + \mathcal{T}(t_j, X_{t_j}))]\mathbf{1}_{\{t_{j+1} < \tau_\infty\}}\big] \\
&\quad + \sum_{j=0}^{k-1} \big[[\Pi_{j+1}(X_{t_j} + \mathcal{T}(t_j, X_{t_j})) - (X_{t_j} + \mathcal{T}(t_j, X_{t_j}))]\mathbf{1}_{\{t_{j+1} = \tau_\infty\}}\big] \\
&\equiv R(1) + R(2) + R(3).
\end{aligned}$$

Note that $R(2)$ and $R(3)$ exactly match $S(2)$ and $S(3)$ in the proof of Lemma 8.1, with $i = 0$. In particular, (8.15) applies. Due to the Lipschitz properties of $b$ and $\sigma$ and to the smoothness of $u$ (see Theorem 2.1), the term $R(1)$ is treated in a classical way. □

Recall now from Proposition 7.2 (estimate of $\hat{v} - v$), Theorem 4.2 ($L^2$ estimate of $\overline{v} - v$) and (8.6) (probability of hitting the boundary of the grids):

$$
\begin{aligned}
& h\sum_{j=0}^{k-1}\mathbb{E}\big[|(\hat{v}-v)(t_j,X_{t_j})|^2\mathbf{1}_{\{t_j<\tau_\infty\}}\big] \\
& \quad \leq C\Big[\mathcal{E}^2(\text{time})+\mathcal{E}^2(\text{space})+\mathcal{E}^2(\text{gradient},p)+h\sum_{j=0}^{k-1}\mathbb{E}\big[|(\overline{v}-v)(t_{j+1},X_{t_{j+1}})|^2\mathbf{1}_{\{t_j<\tau_\infty\}}\big]\Big] \\
& \quad \leq C\Big[\mathcal{E}^2(\text{time})+\mathcal{E}^2(\text{space})+\mathcal{E}^2(\text{gradient},p)+h\sum_{j=1}^{k}\mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_j})|^2\mathbf{1}_{\{t_j<\tau_\infty\}}\big] \\
& \qquad +h\sum_{j=1}^{k}\mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_j})|^2\mathbf{1}_{\{t_j=\tau_\infty\}}\big]\Big] \\
& \quad \leq C\big[\mathcal{E}^2(\text{global})+\mathbb{P}\{\tau_\infty<+\infty\}\big]\leq C\mathcal{E}^2(\text{global}).
\end{aligned}
\tag{8.19}
$$

Apply now inequality (8.10) (estimate of $\overline{u}-u$) and (8.19) to (8.18) and deduce from Gronwall's Lemma:

$$
\sup_{k\in\{0,\dots,N\}}\mathbb{E}|X_{\tau_k}-U_{\tau_k}|^2\leq C\mathcal{E}^2(\text{global}). \tag{8.20}
$$

Finally, according to Theorem 2.1, to Theorem 4.2 ($L^2$ estimate of $\overline{v}-v$) and to (8.10) (estimate of $\overline{u}-u$), we deduce the following intermediate estimate:

$$
\sup_{k\in\{0,\dots,N\}}\mathbb{E}|X_{\tau_k}-U_{\tau_k}|^2+\sup_{k\in\{0,\dots,N\}}\mathbb{E}|Y_{\tau_k}-V_{\tau_k}|^2+h\sum_{j=0}^{N-1}\mathbb{E}\big[|Z_{t_j}-W_{t_j}|^2\mathbf{1}_{\{t_j<\tau_\infty\}}\big]\leq C\mathcal{E}^2(\text{global}). \tag{8.21}
$$

Applying Doob's inequality, we easily obtain the upper bound for:

$$
\mathbb{E}\big[\sup_{k\in\{0,\dots,N\}}|X_{\tau_k}-U_{\tau_k}|^2\big]+\mathbb{E}\big[\sup_{k\in\{0,\dots,N\}}|Y_{\tau_k}-V_{\tau_k}|^2\big]+h\sum_{j=0}^{N-1}\mathbb{E}\big[|Z_{t_j}-W_{t_j}|^2\mathbf{1}_{\{t_j<\tau_\infty\}}\big]. \tag{8.22}
$$

It finally remains to prove that (8.22) holds with $(U_{t_k},V_{t_k},W_{t_k})_{0\leq k\leq N}$ instead of $(U_{\tau_k},V_{\tau_k},W_{t_k}\mathbf{1}_{\{t_k<\tau_\infty\}})_{0\leq k\leq N}$. Since the same arguments apply for $V$ and $W$, we just detail the case of $U$. Note indeed that for every $k\in\{0,\dots,N\}$:

$$
\sup_{k\in\{0,\dots,N\}}|X_{\tau_k}-U_{t_k}|^2\leq C\sup_{k\in\{0,\dots,N\}}|X_{\tau_k}-U_{\tau_k}|^2+C\sup_{k\in\{0,\dots,N\}}|U_{\tau_k}-U_{t_k}|^2.
$$

Thanks to Burkholder, Davis and Gundy inequalities, it is readily seen that:

$$
\mathbb{E}\big[\sup_{k\in\{0,\dots,N\}}|U_{\tau_k}-U_{t_k}|^2\big]\leq C\mathbb{E}\big[t_N-\tau_\infty\big]\leq CT\mathbb{P}\{\tau_\infty<+\infty\}.
$$

Referring to (8.6), we easily complete the proof of Theorem 4.3. □

# 9 Conclusion

As a conclusion, we first give in Subsection 9.1 further comments on Theorem 4.1 and compare in particular the global error with the one obtained by Douglas et *al.* in [DMP96]. We then give some easy extensions in Subsection 9.2. Finally, we detail in Subsection 9.3 the technical difficulties associated to the natural Algorithm 3.1.

## 9.1 Comments and Comparisons with Other Methods

We discuss in this subsection the total complexity and the rate of convergence of Algorithm 3.2.

***Complexity of the Algorithm.*** Note first that the order of the total complexity of the algorithm is $h^{-1}\times M\times(2\delta^{-1}(\rho+R))^d$.

***Rate of Convergence.*** Recall also that the global error of the algorithm is given by Theorem 4.1. Comparing with the results in [DMP96], this global error is worse in our case. There are two reasons to explain this difference. The first one does not depend on the algorithm, but is a consequence of our working assumptions. Indeed, under suitable smoothness properties of the coefficients $b, f, \sigma$ and of the solution $u$, standard Itô developments in (7.9) would lead to $\mathcal{E}^2(\text{time}) = h^2$ as in [DMP96].

At the opposite, the second reason for which the global error is worse in our case, depends on the specific structure of the algorithm. Indeed, our choice to avoid linear interpolation procedures induces a rather large projection error $\mathcal{E}^2(\text{space})$. To reach a term of order one with respect to $h$ for $\mathcal{E}^2(\text{space})$, we then need to take $\delta \equiv h^{3/2}$. This choice is far from being satisfactory and highly increases the complexity when the dimension $d$ increases. Intuitively, there is no specific reason for such a relationship between $\delta$ and $h$: as explained in Subsection 4.1, $\delta$ has just to be small in front of $h$ to take into account the influence of the drift $b$ at the local level. For this reason, we aim to study in further investigations the convergence analysis of the algorithm when using a suitable "smooth" interpolation operator instead of a rough projection mapping. This point is discussed in a detailed way in the next subsection.

***Further Comments on Errors.*** To conclude this subsection, we investigate the three last error terms $\mathcal{E}(\text{trunc})$, $\mathcal{E}(\text{quantiz})$ and $\mathcal{E}(\text{gradient}, p)$.

The truncation error decays linearly when the grid size increases. This control may seem rather poor to the reader. Recall indeed that $\mathcal{E}(\text{trunc})$ appears, up to the discretization procedure, as the probability that a diffusion process leaves a given bounded set. In the case of elliptic diffusions with bounded coefficients, it is well known that this probability decays exponentially fast as the size of the underlying set increases. Recall in this frame from Theorem 2.1 that the coefficients of the elliptic diffusion $U$ are bounded. Note however that this rough argument fails in the discretized setting since there is no *a priori* sharp estimate of the approximate gradient $\overline{v}$ and thus of the associated approximate drift. This explains why our strategy to estimate $\mathcal{E}(\text{trunc})$ lies on the Bienaymé-Chebychev inequality and thus provides the current form given by Theorem 4.1.

Note finally that the errors associated to the quantization procedure, $\mathcal{E}^2(\text{quantiz})$, and to the probabilistic approximation of the gradient, $\mathcal{E}^2(\text{gradient}, p)$, are explicitly controlled in terms of $M$, $h$, and $\delta$. They emphasize the price to pay to weaken the assumptions: we have to assume that the quantization grid is rather small compared to the spatial discretization one. Obviously, this increases the number of elementary operations of the algorithm and thus its total complexity. However, this does not affect so much the discretization procedure of the Gaussian law itself since quantization grids can be computed once for all apart from the implementation procedure of the algorithm.

## 9.2 Extensions and Further Investigations

We now discuss some possible extensions of our work.

***Interpolation Procedure.*** As announced in the latter subsection, we first investigate the assets and liabilities of a smooth interpolation procedure. One of the main advantages of the spatial discretization proposed in Section 3.4, and then used in Algorithm 3.2, lies in its simplicity of implementation. However, from a purely mathematical point of view, this procedure may be rather awkward, since it ignores more or less the deep smoothness of the true solution $u$.

Note in this framework that the function $\Pi_\infty$ may be seen as an operator acting on functions from $\mathbb{R}^d$ into $\mathbb{R}$. For such a function, the operator provides a rough interpolation of order 0 depending on the values of the function on the spatial mesh $\mathcal{C}_\infty$. As mentioned above, this interpolation procedure does not preserve the smoothness properties of the underlying function: in any cases, except if the function is constant, the interpolation procedure induces jumps of size of order $\delta$. As a consequence, the distance between the function and the interpolated one is also of order $\delta$.

A relevant strategy would consist in replacing the projection $\Pi_\infty$ by a smoother interpolation operator. In our framework, an interpolation operator is said to be smooth if the distance between a given function $\ell$ and the interpolated one decreases with the regularity order of $\ell$. For example, in dimension 1, the linear interpolation operator:

$$\ell \mapsto \big(x \mapsto \delta^{-1}(\delta + \delta\lfloor\delta^{-1}x\rfloor - x)\ell\big(\delta\lfloor\delta^{-1}x\rfloor\big) + \delta^{-1}(x - \delta\lfloor\delta^{-1}x\rfloor)\ell\big(\delta\lfloor\delta^{-1}x\rfloor + \delta\big)\big),$$

maps a $\mathcal{C}^2(\mathbb{R},\mathbb{R})$ function into a piecewise smooth function and the distance between them is of order $\delta^2$.

Algorithm 3.2 can be written with respect to this new choice, but we also believe that the proof would be more difficult to detail. Moreover, smooth interpolation procedures in higher dimension slow down the running of the underlying algorithm.

***Weakening Assumption.*** Note to conclude this subsection that some assumptions could be weakened. First, Theorem 2.1 still holds if $b$ and $f$ are just Hölder in $x$: in such a case, usual estimates of the gradient of $u$ hold and Schauder's theory still applies. In particular, the reader can verify that Theorems 4.1 and 4.2 are still valid in this case (but Theorem 4.3 given in Subsection 4.3 is not).

Moreover, Algorithm 3.2 still converges if $b, f$, and $\sigma$ depend on $t$ in a Hölder way.

Finally, the following extension is conceivable. For $H \in C^{1+\alpha}$, $\alpha \in ]0,1[$, the partial derivatives of order two of $u$ have an integrable singularity in the neighborhood of $T$. In this frame, it would be interesting to adapt the Gronwall arguments given in Section 8.

## 9.3 Justification of Algorithm 3.2

We finally explain why we are not able to show the convergence of Algorithm 3.1.

***Convergence of Algorithm 3.1.*** Recall that the main difference between Algorithm 3.1 and Algorithm 3.2 lies in the definition of the forward transitions. Indeed, in Algorithm 3.1:

$$\begin{aligned} \mathcal{T}(t_k,x) &\equiv b(x,\overline{u}(t_{k+1},x),\overline{v}(t_{k+1},x))h + \sigma(x,\overline{u}(t_{k+1},x))g(\Delta B^k), \\ X_0 &\equiv x_0,\ \forall k \in \{0,...,N-1\},\ X_{t_{k+1}} = \Pi_{k+1}\big(X_{t_k} + \mathcal{T}(t_k,X_{t_k})\big). \end{aligned}$$

Unfortunately, in this case, the well known BSDE machinery fails under Assumption **(A)**. At first sight, this could seem rather amazing. Indeed, recall that very strong *a priori* estimates of the solution $u$ and of its partial derivatives hold in our framework. In particular, we could expect the discretization procedure of $u$ and of its gradient to converge under such smoothness properties.

The main difficulty encountered to establish the convergence of Algorithm 3.1 appears in Section 7. More precisely, the lack of *a priori* controls of the regularity of $\overline{u}$ and $\overline{v}$ makes the stability strategy fruitless. Note indeed that inequality (7.1) becomes in the frame of Algorithm 3.1:

$$\begin{aligned} &|(\overline{u}-u)(0,x_0)|^2 + C_{7.1}^{-1} h \sum_{j=1}^N \mathbb{E}\big[|(\overline{v}-v)(t_{j-1},X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\ &\le C_{7.1}\Bigg[\mathbb{P}\{\tau_\infty < +\infty\} + \mathcal{E}^2(\text{time}) + \mathcal{E}^2(\text{space}) + \mathcal{E}^2(\text{quantiz}) \\ &\quad + \eta^{-1} h \sum_{j=1}^N \mathbb{E}\big[|(\overline{u}-u)(t_j,X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\ &\quad + \eta^{-1} h \sum_{j=1}^N \mathbb{E}\big[|(\overline{u}-u)(t_{j-1},X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big] \\ &\quad + (\eta+h) h \sum_{j=1}^N \mathbb{E}\big[|(\overline{v}-v)(t_j,X_{t_{j-1}})|^2 \mathbf{1}_{\{t_{j-1}<\tau_\infty\}}\big]\Bigg]. \end{aligned} \tag{9.1}$$

Inequalities (7.1) and (9.1) just differ in the last term: $\hat{v}(t_{j-1},X_{t_{j-1}})$ becomes $\overline{v}(t_j,X_{t_{j-1}})$. Note to be complete that a similar shift occurs in $v$, but, due to Theorem 2.1, it can be removed without any difficulties. To apply the strategy used in Section 7, and in particular to derive an equivalent of Theorem 7.3 from (9.1), we then need to investigate the regularity in space of $\overline{v}$. According to the definition of $\overline{v}$, a first step then consists in studying the regularity in space of $\overline{u}$.

***Lipschitz Control of $\overline{u}$.*** Note that the natural strategy to control the oscillations of $\overline{u}$ would consist in applying the usual FBSDE machinery to the triples $(X^{t_k,x},Y^{t_k,x},Z^{t_k,x})$ and $(X^{t_k,y},Y^{t_k,y},Z^{t_k,y})$ for $k \in \{0,...,N-1\}$ and $x,y \in \mathcal{C}_k$. Of course, superscripts $(t_k,x)$ and $(t_k,y)$ denote the initial conditions of the

Markov process $X$.

Nevertheless, we are not able to apply the strategies used in [Del02] and [Del03] to derive from the forward-backward writing local and global estimates of the discrete gradient of $\overline{u}$. There are two reasons to explain this failure.

First, the rough projection mapping chosen in Algorithms 3.1 and 3.2 induces an irreducible error greater than $\delta$ when estimating the difference between $\overline{u}(t_k, x)$ and $\overline{u}(t_k, y)$ in function of the parameters appearing in **(A)**. The strategy to overcome this difficulty is well known: the projection mapping has to be replaced by a smoother interpolation operator.

Second, any probabilistic strategy to estimate the Lipschitz constant of $\overline{u}$ in $x$ such as the one exposed in [Del02] leads one way or another to the same difficulty as the one encountered to apply the stability procedure to Algorithm 3.1. More precisely, studying the difference between the triples $(X^{t_k,x}, Y^{t_k,x}, Z^{t_k,x})$ and $(X^{t_k,y}, Y^{t_k,y}, Z^{t_k,y})$, for $k \in \{0, ..., N-1\}$ and $x, y \in \mathcal{C}_k$, leads to investigate the regularity of $\overline{v}$. In short, one needs to estimate first the regularity of $\overline{v}$ to derive the one of $\overline{u}$. Intuitively, it is well understood that this is hopeless.


## Acknowledgements

We would like to thank Professor Vlad Bally for his thoughtful attentive comments.


## References


[Ant93] F. Antonelli. Backward-Forward Stochastic Differential Equations. *Ann. Appl. Probab.*, 3-3:777–793, 1993.

[BP03] V. Bally and G. Pagès. A quantization algorithm for solving discrete time multi-dimensional optimal stopping problems. *Bernoulli*, 9–6:1003–1049, 2003.

[BPP02] V. Bally, G. Pagès, and J. Printems. A quantization tree method for pricing and hedging multi-dimensional American options. *Tech. report of the university Paris VI, Lab. Prob. et Mod. Al., To appear in Math. Fin.*, 753, 2002.

[BT04] B. Bouchard and N. Touzi. Discrete time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stoch. Proc. and App.*, 111:175–206, 2004.

[Che97] D. Chevance. Numerical methods for Backward Stochastic Differential Equations. *Publ. Newton Inst., Cambridge University Press*, pages 232–234, 1997.

[Del02] F. Delarue. On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case. *Stoch. Proc. and App.*, 99:209–286, 2002.

[Del03] F. Delarue. Estimates of the solutions of a system of quasilinear PDEs. A probabilistic scheme. *Séminaire de Probabilités*, XXXVII:290–332, 2003.

[Del04] F. Delarue. Auxiliary SDEs for homogenization of quasilinear PDEs with periodic coefficients. *Annals of Probability*, 32:2305–2361, 2004.

[DMP96] J. Douglas, J. Ma, and P. Protter. Numerical methods for Forward-Backward Stochastic Differential Equations. *Ann. Appl. Prob.*, 6:940–968, 1996.

[GL00] S. Graf and H. Lushgy. *Foundations of quantization for random vectors.* LNM-1730, Springer-Verlag, 2000.

[KPZ86] M. Kardar, G. Parisi, and Y.-C. Zhang. Dynamic scaling of growing interfaces. *Phys. Rev. Lett.*, 56:889–892, 1986.

[LSU67] 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.

[Mak03] R. N. Makarov. Numerical solution of quasilinear parabolic equations and Backward Stochastic Differential Equations. *Russian J. Numer. Anal. Math. Modelling*, 18–5:397–412, 2003.

[MPY94] J. Ma, P. Protter, and J. Yong. Solving Forward-Backward Stochastic Differential Equations explicitly - a four step scheme. *Prob. Th. Rel. Fields*, 98:339–359, 1994.

[MT99] G. N. Milstein and M. V. Tretyakov. Numerical algorithms for semilinear parabolic equations with small parameter based on approximation of stochastic equations. *Math. Comp.*, 69–229:237–267, 1999.

[MY99] J. Ma and J. Yong. *Forward-Backward Stochastic Differential Equations and their applications.* LNM-1702, Springer-Verlag, 1999.

[PP90] E. Pardoux and S.G. Peng. Adapted solution of a Backward Stochastic Differential Equation. *Systems Control Lett*, 14-1:55–61, 1990.

[PP04] G. Pagès and J. Printems. Functional quantization for numerics: pricing Asian options. *Tech. report of the university Paris VI, Lab. Prob. et Mod. Al.*, 900, 2004.

[PPP04] G. Pagès, H. Pham, and J. Printems. Optimal quantization methods and applications to numerical problems in finance. To appear in Handbook on Numerical Methods in Finance (S. Rachev, ed.), Birkhauser, 2003. *Tech. report of the university Paris VI, Lab. Prob. et Mod. Al.*, 813, 2004.

[PT99] E. Pardoux and S. Tang. Forward-Backward Stochastic Differential Equations and quasilinear parabolic PDEs. *Probab. Theory Related Fields*, 114-2:123–150, 1999.

[Shi96] A.N. Shiryaev. *Probability, Second Edition.* Graduate Texts in Mathematics, 95. Springer-Verlag, New York., 1996.

[Whi68] N. Whitham. *Non linear waves.* Prentice Hall, 1968.

[Woy98] W. A. Woyczyński. *Burgers-KPZ turbulence.* LNM-1700, Springer-Verlag, 1998.