L2-time regularity of BSDEs with irregular terminal functions

Emmanuel Gobet, Azmi Makhlouf

$\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions

Emmanuel Gobet [a,*], Azmi Makhlouf [a]

[a] *Laboratoire Jean Kuntzmann, Université de Grenoble and CNRS, BP 53, 38041 Grenoble Cedex 9, FRANCE*

Abstract

We study the $\mathbf{L}_2$-time regularity of the Z-component of a Markovian BSDE, whose terminal condition is a function g of a forward SDE $(X_t)_{0\leq t\leq T}$. When g is Lipschitz continuous, Zhang [1] proved that the related squared $\mathbf{L}_2$-time regularity is of order one with respect to the size of the time mesh. We extend this type of result to any function g, including irregular functions such as indicator functions for instance. We show that the order of convergence is explicitly connected to the rate of decreasing of the expected conditional variance of $g(X_T)$ given X_t as t goes to T. This holds true for any Lipschitz continuous generator. The results are optimal.

Key words: backward stochastic differential equations, time regularity, Malliavin calculus, rate of convergence
1991 MSC: 60H10, 65C30

Framework. In the past decade, a lot of attention has been paid to the numerical resolution of Backward Stochastic Differential Equations (BSDEs in short). In this work, we focus on Markovian BSDEs, that is the case where the BSDE is coupled to a forward SDE. For fixed initial condition x_0 and terminal time $T > 0$, it writes

$$\begin{cases} X_0 &= x_0, \\ dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= g(X_T), \end{cases} \tag{1}$$

* Corresponding author
Email addresses: `emmanuel.gobet@imag.fr` (Emmanuel Gobet), `azmi.makhlouf@imag.fr` (Azmi Makhlouf).

where $g(X_T) \in \mathbf{L}_2$, W is a standard Brownian motion. A solution to (1) is a triplet (X, Y, Z) adapted to the filtration of the Brownian motion, and in some appropriate $\mathbf{L}_2$ spaces (defined later). When the generator f equals 0, Y is given by the conditional expectation $Y_t = \mathbb{E}^{\mathcal{F}_t}(g(X_T))$ and Z is the predictable process arising from the predictable representation theorem. This type of closed representation can be extended to f that are linear w.r.t. the variables y and z (called linear BSDEs). In the other cases (truly non linear), usually no closed representation is available and one needs to compute numerical solutions. As explained later, the corner stone to derive a rate of convergence for numerical schemes solving (1) is the $\mathbf{L}_2$-time regularity of Z. It is defined for a given time mesh $\pi = \{0 = t_0 < \cdots < t_i < \cdots < t_N = T\}$ by

$$\mathcal{E}(Z, \pi) = \sum_{i=0}^{N-1} \mathbb{E} \int_{t_i}^{t_{i+1}} |Z_t - \bar{Z}_{t_i}|^2 dt \tag{2}$$

where $\bar{Z}_{t_i} = \frac{1}{t_{i+1}-t_i} \mathbb{E}^{\mathcal{F}_{t_i}} \int_{t_i}^{t_{i+1}} Z_s ds$. Note that $\bar{Z}_{t_i}$ is the projection of $(Z_s)_{t_i \le s \le t_{i+1}}$ on the space of $\mathcal{F}_{t_i}$-measurable random variables, according to the scalar product $< u, v >= \mathbb{E} \int_{t_i}^{t_{i+1}} u_s v_s ds$. The objective of this work is to provide tight estimates of $\mathcal{E}(Z, \pi)$, according to the regularity of the function g and the mesh size $|\pi| = \sup_{0 \le i < N} (t_{i+1} - t_i)$. In all the sequel, we only consider time mesh with N deterministic points.

A brief account on numerical methods for BSDEs. There are three main approaches for the numerical solution of (Y, Z) (the simulation of forward component X is standard). Firstly, under appropriate conditions, $Y_t = v(t, X_t)$ where v solves a semi-linear PDE (and Z is analogously related to the gradient of v) (see [2] for instance): hence one may solve this PDE by deterministic methods and then, we get Y by simulating X. Secondly, one may approach the BSDE by a sequence of linear BSDEs (Picard iteration scheme): this is efficiently achieved by Gobet and Labart [3], by coupling the resolution with iterative control variates that drastically improves the accuracy. The third approach is strongly related to the motivation of this work: it uses a backward dynamic programming equation of the form $(i < N)$

$$\begin{cases} Y^\pi_{t_i} = \mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}} + (t_{i+1} - t_i) f(t_i, X_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})), \\ Z^\pi_{t_i} = \frac{1}{(t_{i+1}-t_i)} \mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}}(W_{t_{i+1}} - W_{t_i})^*), \end{cases} \tag{3}$$

where * denotes the transposition and $Y^\pi_{t_N} = g(X_T)$. In addition, possibly X can be replaced by a process easier to simulate (Euler scheme for instance) and close to X. The equations (3) defines an explicit scheme but it could be implicit as well, replacing in f the quantity $Y^\pi_{t_{i+1}}$ by $Y^\pi_{t_i}$: this does not modify the convergence results. The next big issue would be how to compute the conditional expectations: we do not discuss these aspects here and we refer to [4] for quantization techniques, to [5] for Malliavin calculus tools, to [6] for empirical regression methods. Let us focus on the error estimate between

(Y^π, Z^π) and (Y, Z). Actually under standard Lipschitz assumptions on f, it is now well known (even in the more general case where jumps are included in the equations, see [7]) that the error can be estimated as follows:

$$\begin{aligned} e(Y^\pi - Y, Z^\pi - Z, \pi) &:= \sup_{0\le i\le N} \mathbb{E}(Y^\pi_{t_i} - Y_{t_i})^2 + \sum_{i=0}^{N-1} \mathbb{E}\int_{t_i}^{t_{i+1}} |Z^\pi_{t_i} - Z_t|^2 dt \\ &\le C(|\pi| + \mathcal{E}(Z,\pi)). \end{aligned}$$

Thus, it is clear that the $\mathbf{L}_2$-time regularity of Z plays a crucial role in the rate of convergence of the dynamic programming equation (3).

Known results on the $\mathbf{L}_2$-regularity of Z. In the BSDE framework, the best result to our knowledge has been obtained by Zhang [1]: $\mathcal{E}(Z,\pi)$ is of order $|\pi|$ when g is a Lipschitz continuous function. Consequently, $e(Y^\pi - Y, Z^\pi - Z, \pi)$ is also of order $|\pi|$ and uniform time grids ($t_i = iT/N$) are sufficient for the approximation scheme. However, in practice g may be an indicator function: in that case, one expects that it worsens the rate of convergence $\mathcal{E}(Z,\pi)$ to 0. Actually, this downgrade phenomenon is well known when the generator f is null. This problem is related to the approximation of stochastic integrals and of hedging strategy in finance. In [8], it is proved for instance that for indicator functions in dimension 1, one gets $\mathcal{E}(Z,\pi) = O(N^{-1/2})$ for an uniform time grid. In fact, any rate $N^{-\alpha}$ with $\alpha \in (0,1)$ can be obtained, by picking an appropriate non smooth function g_α. The larger α is, the smoother g_α is. The above results are extended by Geiss and his coauthors (see [9] and references therein) by considering functions g in a Besov space $\mathbf{B}^\alpha_{2,2}$ ($\alpha \in (0,1]$). For a uniform time grid, they prove that the regularity index α exactly gives the rate of convergence: $\mathcal{E}(Z,\pi) = O(N^{-\alpha})$. In addition, to get the rate N^{-1} using a grid with N points, one has to consider points appropriately concentrated near T. We emphasize that their method of proofs is essentially restricted to the Brownian motion case for X because it relies on Hermite polynomials expansion of g, which allows for explicit computations.

The purpose of this work is twofold: firstly, to extend this type of results to general SDEs; secondly, to deal with general BSDEs (i.e. with non null generator).

Summary of our results. For general SDE model, the characterization of the rate of convergence of $\mathcal{E}(Z,\pi)$ in terms of Besov space is no more relevant. It appears more natural to consider the following space ($\alpha \in (0,1]$)

$$\mathbf{L}_{2,\alpha} = \{g \text{ s.t. } \mathbb{E}(g(X_T)^2) + \sup_{0\le t<T} \frac{\mathbb{E}(g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T)))^2}{(T-t)^\alpha} < +\infty\}. \tag{4}$$

It describes the rate of decreasing of the expected conditional variance of $g(X_T)$ given $\mathcal{F}_t$ as t goes to T. If X is a Brownian motion and $T = 1$, $g \in \mathbf{L}_{2,\alpha}$ is equivalent to $g \in \mathbf{B}^\alpha_{2,\infty}$ provided $\alpha \neq 1$ (see Corollary 2.3 in [9]). However, our characterization is more flexible because it is adapted to the process and the time horizon T. In addition, we show that this quantity is intrinsic to the the time regularity of Z (even in the BSDE case). For uniform grids, the rate of convergence is of order $N^{-\alpha}$ (Theorem 21, (a)). Also, one can take non uniform grids to get the rate N^{-1} (Theorem 21, (b)). To achieve these results, we first estimate the error in the null generator case (thus extending the results by Geiss et al. in a non trivial way) (Theorem 8, Theorem 9 and Corollary 10). Then we prove that the non null generator case (involving $(Z_t)_t$) is a perturbation of the null case (with $(z_t)_t$), so that the former results still apply (Theorem 20):

$$\mathcal{E}(Z,\pi) \leq C(\mathcal{E}(z,\pi) + |\pi|).$$

More precisely, we establish that Z is the superposition of z plus a time smoother term (Theorem 12). This result seems to be original in our framework. It allows us to reduce the study of the $\mathbf{L}_2$ time regularity of Z to that of z (the former case) and that of the smoother term (which is easier). The decomposition may be also interesting to get tight estimates on the behavior of Z as t goes to T (Corollary 14). Our proof relies on stochastic analysis techniques combining PDEs, martingales, Itô calculus and BSDEs in $\mathbf{L}_p$ ($p \in (1,2]$). We mention that usually with these tools, g is supposed to have a polynomial growth, ensuring that $g(X_T)$ is in any $\mathbf{L}_p$ for $p > 0$. Here we stress the fact that we only assume $g(X_T) \in \mathbf{L}_2$ which is the minimal condition to discuss the existence and uniqueness of the solution of (1) in $\mathbf{L}_2$ spaces.

We finally discuss the choice of time grids (uniform or alpha dependent) and the optimality of the results.

Preliminaries. Hereafter, W is a q-dimensional Brownian motion, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $(\mathcal{F}_t)_{0\leq t\leq T}$ (T is a fixed terminal time) is the natural filtration of W, augmented with $\mathbb{P}$-null sets.
We denote the conditional expectation $\mathbb{E}(X|\mathcal{F}_t)$ of a random variable X by $\mathbb{E}^{\mathcal{F}_t}(X)$.
$A \leq_c B$ means $A \leq cB$ with a constant c depending on T, b, σ, f and universal constants.
For a $r \times c$ matrix A ($r, c \geq 1$), that will be considered as an element of $\mathbb{R}^{r\times c}$, A^* stands for its transpose, A_j for its jth column, and $|A|$ for its Euclidean norm ($|A| := \sqrt{\mathrm{Tr}(AA^*)}$).
If $\varphi : \mathbb{R}^{p_1} \to \mathbb{R}^{p_2}$ is a differentiable function, its gradient $\nabla_x\varphi(x) := (\partial_{x_1}\varphi(x), ..., \partial_{x_{p_1}}\varphi(x))$ takes values in $\mathbb{R}^{p_2\times p_1}$.
If $p_2 = 1$, $D^2\varphi(x) := (\partial^2_{x_i,x_j}\varphi(x))_{i,j=1...d}$ stands for the Hessian matrix of φ and takes values in $\mathbb{R}^{p_1\times p_1}$.

a) The forward component:

$$\begin{cases} X_0 &= x_0, \\ dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \end{cases} \tag{5}$$

where $X, x_0 \in \mathbb{R}^d$, $\ b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\ \sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$. We will assume that the coefficients of this SDE satisfy the following assumption:

> **($\mathbf{A}_{b,\sigma}$)** The functions b and σ are bounded and twice continuously differentiable with respect to the space variable, with uniformly bounded and γ-Hölder continuous derivatives, for some $\gamma \in (0,1]$. In addition, b and σ are $\frac{1}{2}$-Hölder continuous in time. σ is also assumed to be uniformly elliptic: there exists $\delta > 0$ such that, $\forall (t,x) \in [0,T] \times \mathbb{R}^d$, $[\sigma\sigma^*](t,x) \geq \delta I_d$.

We denote by ∇X_s the gradient of X_s with respect to x_0, and by $(D_t X_s)_{0\leq t\leq s}$ its Malliavin derivative (see [10]). It is known that $(\nabla X_s)_{s\geq 0}$ and $(D_t X_s)_{s\geq t}$ satisfy the following linear SDEs

$$\nabla X_s = I_d + \int_0^s \nabla_x b(r, X_r)\nabla X_r dr + \sum_{j=1}^q \int_0^s \nabla_x \sigma_j(r, X_r)\nabla X_r dW_r^j; \tag{6}$$

$$D_t X_s = \sigma(t, X_t) + \int_t^s \nabla_x b(r, X_r) D_t X_r dr + \sum_{j=1}^q \int_t^s \nabla_x \sigma_j(r, X_r) D_t X_r dW_r^j. \tag{7}$$

The following estimates are standard results in SDE literature.

Lemma 1 *Assume ($\mathbf{A}_{b,\sigma}$). For any $p \geq 2$, there exits a constant C_p such that*

$$\mathbb{E} \sup_{0\leq s\leq T} |X_s|^p \leq C_p(1 + |x_0|^p),$$
$$\mathbb{E}|X_s - X_t|^p \leq C_p |s-t|^{\frac{p}{2}}.$$

From equation (6), one gets the following estimate, that will be used in this work (it is a standard estimate if $t = 0$; one can deduce the estimate for $t \neq 0$ since $\nabla X_s [\nabla X_t]^{-1}$ is the derivative of X_s with respect to X_t):

$$\mathbb{E}^{\mathcal{F}_t} \sup_{t\leq s\leq T} |\nabla X_s [\nabla X_t]^{-1}|^p \leq C_p. \tag{8}$$

Since

$$D_t X_s = \nabla X_s (\nabla X_t)^{-1} \sigma(t, X_t) \mathbb{1}_{t\leq s} \tag{9}$$

and σ is bounded, the same estimate applies to $(D_t X_s)_{t\leq s\leq T}$.

b) The backward component:

$$\begin{cases} -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ \ Y_T &= g(X_T). \end{cases} \tag{10}$$

We define the space $\mathcal{S}^p$ to be the set of continuous adapted processes Y such that

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |Y_t|^p\Big] < +\infty,$$

and $\mathcal{M}^p$ the set of predictable processes Z such that

$$\mathbb{E}\Big[\Big(\int_0^T |Z_s|^2 ds\Big)^{p/2}\Big] < +\infty.$$

In the following, Y is always considered as a one dimensional process, but all our study would remain valid if it were multidimensional.

A solution to (10) is a triplet (X, Y, Z), where X is a continuous adapted $\mathbb{R}^d$-valued process with $\mathbb{E}(\sup_{t\le T} |X_t|^2) < +\infty$, solution to the SDE (5), and $(Y, Z) \in \mathcal{S}^2 \times \mathcal{M}^2$. We make use of the following assumption on the generator:

($\mathbf{A}_f$) The function f is continuous with respect to its four arguments, and continuously differentiable with respect to (x, y, z) with uniformly bounded derivatives. Moreover, $\int_0^T |f(s, 0, 0, 0)|ds < +\infty$.

In Theorem 22, f is assumed to be only Lipschitz continuous in (x, y, z), but not necessarily continuously differentiable.

Under the assumptions **($\mathbf{A}_{b,\sigma}$)** and **($\mathbf{A}_f$)**, and when $\mathbb{E}|g(X_T)|^2 < +\infty$, the FBSDE (1) has a unique solution $(X, Y, Z) \in \mathcal{S}^2 \times \mathcal{S}^2 \times \mathcal{M}^2$.

c) Linear PDE and linear BSDE. Some of our intermediate results require the following boundedness assumption on the terminal function g:

($\mathbf{A}_g$) g is a bounded measurable function.

Under the assumptions **($\mathbf{A}_{b,\sigma}$)** and **($\mathbf{A}_g$)**, and setting $u(t, x) := \mathbb{E}\left[g(X_T^{t,x})\right]$, one has $u(t, x) = \int_{\mathbb{R}^d} p(t, x; T, y)g(y)dy$ where p is the probability transition density function of X. It is well known that p is a smooth function for $t < T$ (see Friedman [11]) and this regularity transfers to u since g is bounded. Indeed, Gaussian type estimates on p and its derivatives enable us to apply the Lebesgue derivation theorem. Of course, boundedness assumptions are too strong for this statement, and sub-exponential growth would be enough. However, assuming only at this stage that g is such that $\mathbb{E}(g^2(X_T)) < +\infty$ leads to technicalities that we have not been able to overcome.
To sum up, under **($\mathbf{A}_g$)**, $\nabla_x u$, $D^2 u$, $\partial_x^3 u$, $\partial_t u$, $\partial_t \nabla_x u$ exist and are continuous

for $t < T$, and u is the smooth solution (on $[0,T)\times\mathbb{R}^d$) of the partial differential equation (PDE)

$$\partial_t u(t,x) + \sum_{i=1}^{d} b_i(t,x)\partial_{x_i} u(t,x) + \frac{1}{2}\sum_{i,j=1}^{d} [\sigma\sigma^*]_{i,j}(t,x)\partial^2_{x_i,x_j} u(t,x) = 0 \text{ for } t < T,$$
$$u(T,x) = g(x). \tag{11}$$

Let $(y_t, z_t)_{0\le t\le T}$ be the solution of the linear BSDE:

$$y_t = g(X_T) - \int_t^T z_s dW_s. \tag{12}$$

Then

$$y_t = u(t,X_t), \quad z_t = \nabla_x u(t,X_t)\sigma(t,X_t). \tag{13}$$

d) The space $\mathbf{L}_{2,\alpha}$. For a measurable function g satisfying $\mathbb{E}|g(X_T)|^2 < +\infty$, we set

$$V_{t,T}(g) := \mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T))\right|^2, \tag{14}$$

and, when g belongs to $\mathbf{L}_{2,\alpha}$, we define $K^\alpha(g)$ as

$$K^\alpha(g) := \mathbb{E}|g(X_T)|^2 + \sup_{t\in[0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha}.$$

Notice that $\bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$ obviously contains uniformly Hölder continuous functions, but also some non-smooth functions, such as the indicator function of a domain (under some conditions on the functions b and σ and on the domain: see Gobet and Munos [12]).

Examples:

- If g is β-Hölder continuous, then $g \in \mathbf{L}_{2,\beta}$.
- If $d = q = 1$, $X = W$ and $g(x) = \mathbb{1}_{[0,+\infty)}(x)$, then $g \in \mathbf{L}_{2,\frac{1}{2}}$ (see paragraph 1.2).
- More generally, for an indicator function of a smooth domain, $g \in \mathbf{L}_{2,\frac{1}{2}}$.

e) The time net. In all what follows, $\pi := (t_k)_{k=0\ldots N}$ is a deterministic time net, such that $0 = t_0 < t_1 < \ldots < t_N = T$, and $|\pi| := \sup_{k=0\ldots N}(t_{k+1} - t_k)$. We shall use the following net ($\beta \in (0,1]$)

$$\pi^{(\beta)} := \Big\{t_k^{(N,\beta)} := T - T\Big(1 - \frac{k}{N}\Big)^{\frac{1}{\beta}}, 0 \le k \le N\Big\}. \tag{15}$$

Note that $\pi^{(1)} = (t_k^{(N,1)})$ coincides with the equidistant net. For $\beta < 1$, the points in $\pi^{(\beta)}$ are more concentrated near T.

f) The constants. We emphasize the fact that, whenever a constant depends on the function g, the dependence will be expressed explicitly, so that all the constants such as C or c do not depend on g, but may depend on b, σ, f, α and other universal constants. They may also depend on T, but remain bounded when $T \to 0$.

1 The BSDE with null generator ($f = 0$) and bounded terminal condition g

1.1 Main results

In this section, we study the solution $(y_t, z_t)_{0\leq t\leq T}$ of the BSDE with null generator, and with terminal condition $g(X_T)$. We derive estimates that will be usefull for the next sections, and in particular we study the $\mathbf{L}_2$-regularity of the integrand $(z_t)_{0\leq t\leq T}$.
It is known for years that the $\mathbf{L}_2$-time regularity of z is strongly related to the rate of explosion of the derivatives of $u(t,x)$ as t goes to T (see [8] and [9]). We give below standard and also new related estimates, that will be useful in the proofs.
The following estimate is standard:

$$\mathbb{E} \sup_{0\leq t\leq T} |y_t|^2 + \mathbb{E}\int_0^T |z_s|^2 ds \leq_c \mathbb{E}|g(X_T)|^2, \tag{16}$$

and, it follows from (13) and (16), under the ellipticity assumption,

$$\mathbb{E} \sup_{0\leq t\leq T} |u(t,X_t)|^2 + \mathbb{E}\int_0^T |\nabla_x u(s,X_s)|^2 ds \leq_c \mathbb{E}|g(X_T)|^2, \tag{17}$$

We now bring together different estimates on $\nabla_x u$ and $D^2 u$ in terms of the suitable integrability of $V_{t,T}(g)$ as t goes to T.

Lemma 2 ($\mathbf{L}_2$-estimates for u and its derivatives) *Assume **($\boldsymbol{A}_{b,\sigma}$)** and **($\boldsymbol{A}_g$)**. Then, there exists a positive constant C, such that, for all $t \in [0,T)$,*

$$\begin{aligned}
\mathbb{E}|u(t,X_t)|^2 &\leq \mathbb{E}|g(X_T)|^2,\\
\mathbb{E}|\nabla_x u(t,X_t)|^2 &\leq C\frac{V_{t,T}(g)}{T-t},\\
\mathbb{E}|D^2 u(t,X_t)|^2 &\leq C\frac{V_{t,T}(g)}{(T-t)^2}.
\end{aligned}$$

For the proof, see section 1.3. The powers of $(T-t)$ appearing in Lemma 2 are standard, but note that the $\mathbf{L}_2$-norms depend on $V_{t,T}(g)$ but not on the supremum norm of g.

The following estimate, which is a consequence of Lemma 2, will be usefull in our work:

Corollary 3 *Assume* $(\boldsymbol{A}_{b,\sigma})$*and* $(\boldsymbol{A}_g)$*. Assume moreover that* $g \in \mathbf{L}_{2,\alpha}$*, for some* $\alpha \in (0,1]$*. Then, there exists a positive constant* C*, such that*

$$\mathbb{E}\left(\int_0^T |\nabla_x u(t,X_t)| + |D^2 u(t,X_t)|dt\right)^2 \leq CT^\alpha K^\alpha(g).$$

For the proof, see section 1.3. We will show (see the proof of Theorem 9), that

$$\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2\, ds \leq_c \frac{1}{N} + \sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} (t_{k+1} - r)\mathbb{E}|D^2 u(r,X_r)|^2 dr$$

(and we have equivalence if $(X_t)_t$ is the Brownian motion).
Now, Geiss and Hujo [9] (Lemma 3.8) showed that, if $\varphi : [0,T) \longrightarrow [0,\infty)$ is a non-decreasing continuous function, then

$$\sum_{k=0}^{N-1}\int_{t_k^{(N,\beta)}}^{t_{k+1}^{(N,\beta)}} (t_{k+1}^{(N,\beta)} - r)\varphi(r)dr \leq \frac{c}{N} \Longleftrightarrow \int_0^T (T-r)^{1-\beta}\varphi(r)dr < \infty. \qquad (18)$$

If $(X_t)_t$ is a Brownian Motion, $(D^2u(r,X_r))_{r<T}$ is an $\mathbf{L}_2$-martingale, which easily implies that $\varphi(r) = \mathbb{E}|D^2u(r,X_r)|^2$ defines a non-decreasing continuous function.
In the following proposition, we give two (more explicit) new characterizations of the integrability of $(T-r)^{1-\beta}\varphi(r)$.

Proposition 4 *Let* $\beta \in (0,1)$*, and assume* $(\boldsymbol{A}_{b,\sigma})$*,* $(\boldsymbol{A}_g)$*. Then the following assertions are equivalent:*
(i) $\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2 dr < +\infty.$
(ii) $\int_0^T (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2 dr < +\infty.$
(iii) $\int_0^T (T-r)^{-1-\beta}V_{r,T}(g)dr < +\infty.$

For the proof, see section 1.3.

Remark 5 *Actually, in the above result, the boundedness assumptions of* g *can be relaxed into a sub-exponential growth condition.*
The characterizations given by Proposition 4 are no longer true when $\beta = 1$*. A counterexample is given by* $g(x) = x$ *with* $(X_t) \equiv (W_t)$*, which gives* $u(t,x) = x$ *: assertion(i) is satisfied, but neither (ii) nor (iii) are. In fact, it can easily be seen that if we take any infinitely smooth but non constant function* g *(always with* $(X_t) \equiv (W_t)$*), the assertion (iii) is never satisfied with* $\beta = 1$*.*
That's why we did not define the space $\mathbf{L}_{2,\beta}$ *as the space of functions satisfying*

the assertion (iii) : otherwise, a Lipschitz continuous function such that $g(x) = x$ would belong to $\mathbf{L}_{2,\beta}$ with $\beta < 1$ but not to $\mathbf{L}_{2,1}$. Thus, it would imply (see below) to work with the non equidistant time grid $\pi^{(\beta)}$ instead of the equidistant one as it has usually been done when one has a Lipschitz continuous terminal condition.
And it is clear that, if $g \in \mathbf{L}_{2,\alpha}$ ($\alpha \leq 1$), then all the three assertions of Proposition 4 are satisfied with $\beta < \alpha$.

Now, we state tight estimates on $\nabla_x u$ and $D^2 u$ according to $g \in \mathbf{L}_{2,\alpha}$ for $\alpha \in (0,1]$ (note that $\alpha = 1$ is allowed).

Lemma 6 *Let $\alpha \in (0,1]$, and assume **($\boldsymbol{A}_{b,\sigma}$)** and **($\boldsymbol{A}_g$)**. Then the three following assertions are equivalent:*
(i) $g \in \mathbf{L}_{2,\alpha}$.
(ii) $\exists C^\alpha(g) > 0$, such that, $\forall t \in [0,T)$,

$$\int_0^t \mathbb{E} \left| D^2 u(s, X_s) \right|^2 ds \leq \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

(iii) $\exists C^\alpha(g) > 0$, such that, $\forall t \in [0,T)$,

$$\mathbb{E} \left| \nabla_x u(t, X_t) \right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

And, if $g \in \mathbf{L}_{2,\alpha}$, one can take $C^\alpha(g) = C K^\alpha(g)$ in (ii) and (iii).
If $\alpha < 1$ (resp. $\alpha = 1$), the previous three assertions are also equivalent to (resp. lead to) the following one:
(iv) $\exists C^\alpha(g) > 0$, such that, $\forall t \in [0,T)$,

$$\mathbb{E} \left| D^2 u(t, X_t) \right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{2-\alpha}}$$

(and one can take $C^\alpha(g) = C K^\alpha(g)$).

Remark 7 *The assumption "$g \in \mathbf{L}_{2,\alpha}$" is natural in our framework, if we want a rate of convergence for $\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 ds$ which is polynomial with respect to the time-step $|\pi|$. In fact, Geiss and Hujo [9] (Theorem 1.3 and Theorem 2.1) showed that, if B is either the Brownian motion or the geometric Brownian motion, then the following assertions are equivalent ($\alpha \in (0,1)$):*
(GH-i) $\exists C > 0, \forall t \in [0,T)$,

$$\mathbb{E} \left| \nabla_x u(t, B_t) \right|^2 \leq \frac{C}{(T-t)^{1-\alpha}}.$$

(GH-ii) $\exists C > 0$ such that, for all time-nets $\pi = (t_k)_{k=0\ldots N}$,

$$\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 ds \leq \frac{C}{N^\alpha}.$$

Lemma 6 shows that the assertion (GH-i), written with a general diffusion X instead of B, is equivalent to the assertion "$g \in \mathbf{L}_{2,\alpha}$" (even for $\alpha = 1$).

The previous estimates are sufficient to assert that, if $g \in \mathbf{L}_{2,\alpha}$, then the equidistant time net provides an $\mathcal{E}(z, \pi^{(1)})$ of order $\frac{1}{N^\alpha}$.

Theorem 8 *Assume $(\mathbf{A}_{b,\sigma})$ and $(\mathbf{A}_g)$. Assume moreover that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0,1]$. Then, with the choice of the equidistant time net,*

$$\mathcal{E}(z, \pi^{(1)}) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k^{(N,1)}}^{t_{k+1}^{(N,1)}} \left| z_s - \bar{z}_{t_k^{(N,1)}} \right|^2 ds \leq C K^\alpha(g) \left(\frac{T}{N} \right)^\alpha$$

(where C does not depend on N).

PROOF. One knows by (13) that $z_s = \nabla_x u(s, X_s) \sigma(s, X_s)$. Thus, by a projection argument, one has

$$\begin{aligned}
\mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 ds &\leq \mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - z_{t_k}|^2 ds \\
&= \mathbb{E} \int_{t_k}^{t_{k+1}} |\nabla_x u(s, X_s) \sigma(s, X_s) - \nabla_x u(t_k, X_{t_k}) \sigma(t_k, X_{t_k})|^2 ds.
\end{aligned}$$

Now, write $\nabla_x u(s, X_s)\sigma(s, X_s) - \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k}) = \nabla_x u(t_k, X_{t_k}) \{\sigma(s, X_s) - \sigma(t_k, X_{t_k})\} + \{\nabla_x u(s, X_s) - \nabla_x u(t_k, X_{t_k})\} \sigma(s, X_s)$. Then, using the assumptions $(\mathbf{A}_{b,\sigma})$, and for $s \in [t_k, t_{k+1}]$,

$$\begin{aligned}
&\mathbb{E}|\nabla_x u(s, X_s)\sigma(s, X_s) - \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})|^2 \\
&\leq_c \mathbb{E}\left\{ \left(|s - t_k|^{\frac{1}{2}} + |X_s - X_{t_k}| \right)^2 |\nabla_x u(t_k, X_{t_k})|^2 \right\} + \mathbb{E}|\nabla_x u(s, X_s) - \nabla_x u(t_k, X_{t_k})|^2 \\
&= E_1 + E_2.
\end{aligned}$$

Clearly, and by means of Lemma 2,

$$E_1 \leq_c |\pi| \mathbb{E}|\nabla_x u(t_k, X_{t_k})|^2 \leq |\pi| \frac{K^\alpha(g)}{(T - t_k)^{1-\alpha}} \leq |\pi| \frac{K^\alpha(g)}{(T - s)^{1-\alpha}}.$$

As in the proof of Proposition 4 (see (22)), one gets the general estimate (under $(\mathbf{A}_g)$)

$$\begin{aligned}
E_2 &\leq_c \int_{t_k}^{s} \mathbb{E}|\nabla_x u(r, X_r)|^2 dr + \int_{t_k}^{s} \mathbb{E}|D^2 u(r, X_r)|^2 dr \\
&\leq_c |\pi| \frac{K^\alpha(g)}{(T - s)^{1-\alpha}} + \int_{t_k}^{s} \mathbb{E}|D^2 u(r, X_r)|^2 dr
\end{aligned}$$

using (iii) Lemma 6. Therefore

$$\begin{aligned}
\mathcal{E}(z,\pi) &\leq_c |\pi| K^\alpha(g) \int_0^T \frac{1}{(T-s)^{1-\alpha}} ds + \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} \int_{t_k}^{s} \mathbb{E}|D^2 u(r,X_r)|^2 dr ds \\
&= |\pi| K^\alpha(g) T^\alpha + \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} (t_{k+1}-r)\mathbb{E}|D^2 u(r,X_r)|^2 dr, \qquad (19)
\end{aligned}$$

where we have used an integration by parts. Note that the above upper bound is available for any time net π.
Now, if π is the equidistant time net, $\mathcal{E}(z,\pi^{(1)})$ is bounded (up to a constant c) by

$$K^\alpha(g)\frac{T^{\alpha+1}}{N} + \frac{T}{N}\int_0^{T-\frac{T}{N}} \mathbb{E}|D^2 u(r,X_r)|^2 dr + \int_{T-\frac{T}{N}}^{T} (T-r)\mathbb{E}|D^2 u(r,X_r)|^2 dr.$$

Using Lemma 2, one gets

$$\begin{aligned}
\mathcal{E}(z,\pi^{(1)}) &\leq_c K^\alpha(g)\frac{T^{\alpha+1}}{N} + \frac{T}{N}\int_0^{T-\frac{T}{N}} \frac{K^\alpha(g)}{(T-r)^{2-\alpha}} dr + \int_{T-\frac{T}{N}}^{T} \frac{K^\alpha(g)}{(T-r)^{1-\alpha}} dr \\
&\leq_c K^\alpha(g)\frac{T^{\alpha+1}}{N} + K^\alpha(g)\frac{T}{N}\left(\left(\frac{T}{N}\right)^{-1+\alpha} - T^{-1+\alpha}\right) + K^\alpha(g)\left(\frac{T}{N}\right)^{\alpha} \\
&\leq_c K^\alpha(g)\left(\frac{T}{N}\right)^{\alpha}.
\end{aligned}$$

□

To get the rate $\frac{1}{N}$ in the case $\alpha \in (0,1)$, one should consider time nets with a higher concentration of points near T to compensate the faster rate of explosion of $(D^2 u)$. For non equidistant time nets, we state the following universal bounds.

Theorem 9 *Assume* $(\boldsymbol{A}_{b,\sigma})$ *and* $(\boldsymbol{A}_g)$*. Assume moreover that* $g \in \mathbf{L}_{2,\alpha}$*, for some* $\alpha \in (0,1]$.
Now, take $\beta = 1$*, if* $\alpha = 1$*, and* $\beta < \alpha$ *otherwise. Then,* $\exists C > 0$ *such that, for any time net* $\pi = \{t_k, k = 0...N\}$*,*

$$\mathcal{E}(z,\pi) \leq C K^\alpha(g) T^\alpha |\pi| + C K^\alpha(g) T^{\alpha-\beta} \sup_{k=0\ldots N-1} \left(\frac{t_{k+1}-t_k}{(T-t_k)^{1-\beta}}\right).$$

PROOF. Owing to inequality (19), $\mathcal{E}(z,\pi)$ is bounded by

$$C\left(|\pi| K^\alpha(g) T^\alpha + \left\{\sup_{k=0\ldots N-1} \sup_{r\in[t_k,t_{k+1})} \left(\frac{t_{k+1}-r}{(T-r)^{1-\beta}}\right)\right\} \int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2 u(r,X_r)|^2 dr\right).$$

Now, for $r \in [t_k, t_{k+1})$,

$$\frac{t_{k+1} - r}{(T-r)^{1-\beta}} = \left(1 - \frac{T - t_{k+1}}{T - r}\right)(T-r)^{\beta}$$
$$\leq \left(1 - \frac{T - t_{k+1}}{T - t_k}\right)(T-t_k)^{\beta} = \frac{t_{k+1} - t_k}{(T-t_k)^{1-\beta}},$$

which leads to $\sup_{r\in[t_k,t_{k+1})}\left(\frac{t_{k+1}-r}{(T-r)^{1-\beta}}\right) = \frac{t_{k+1}-t_k}{(T-t_k)^{1-\beta}}$. Then,

$$\mathcal{E}(z,\pi) \leq_c |\pi| K^{\alpha}(g) T^{\alpha} + \sup_{k=0\ldots N-1}\left(\frac{t_{k+1}-t_k}{(T-t_k)^{1-\beta}}\right)\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2 u(r,X_r)|^2 dr.$$

If $\beta = \alpha = 1$, then, from Lemma 6, one has $\int_0^T \mathbb{E}|D^2u(r,X_r)|^2 dr \leq_c K^{\alpha}(g)$. And, if $\beta < \alpha < 1$, then, from Lemma 2,

$$\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2 dr \leq_c \int_0^T (T-r)^{1-\beta}\frac{K^{\alpha}(g)}{(T-t)^{2-\alpha}} dr \leq K^{\alpha}(g) T^{\alpha-\beta}.$$

We conclude that, in both cases,

$$\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2 dr \leq_c K^{\alpha}(g) T^{\alpha-\beta}.$$

The proof is complete. □

Corollary 10 *Assume* $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_g)$, *and that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$. *Let* β *be as in Theorem 9. Then, with the choice* $\pi^{(\beta)}$ *(defined in (15))*

$$\mathcal{E}(z,\pi^{(\beta)}) = \sum_{k=0}^{N-1}\mathbb{E}\int_{t_k^{(N,\beta)}}^{t_{k+1}^{(N,\beta)}} \left|z_s - \bar{z}_{t_k^{(N,\beta)}}\right|^2 ds \leq C K^{\alpha}(g)\frac{T^{\alpha}}{N}$$

(where C *does not depend on* N*).*

PROOF. Recall that $t_k^{(N,\beta)} = T - T\left(1 - \frac{k}{N}\right)^{\frac{1}{\beta}}$. Since the function $r \mapsto T - T(1-r)^{\frac{1}{\beta}}$ is concave on $[0,1]$, one has

$$t_{k+1}^{(N,\beta)} - t_k^{(N,\beta)} \leq \frac{T}{\beta N}\left(1 - \frac{k}{N}\right)^{\frac{1}{\beta}-1}.$$

Therefore,

$$\frac{t_{k+1}^{(N,\beta)} - t_k^{(N,\beta)}}{(T - t_k^{(N,\beta)})^{1-\beta}} \leq \frac{T}{\beta N}\frac{\left(1-\frac{k}{N}\right)^{\frac{1}{\beta}-1}}{T^{1-\beta}\left(1-\frac{k}{N}\right)^{\frac{1}{\beta}(1-\beta)}} = \frac{T^{\beta}}{\beta N}.$$

This, combined with Theorem 9, proves Corollary 10. □

1.2 Optimality of the time net

One may raise the following question: if $\alpha < 1$ and $g \in \mathbf{L}_{2,\alpha}$, is the time net $\pi^{(\beta)} = \{t_k^{(N,\beta)} : 0 \le k \le N\}$, with $\beta < \alpha$, optimal? In other words, can we take $\beta = \alpha$ so as to have a rate of convergence of $\frac{1}{N}$? It follows from the previous results that the answer is no. Let us give a counterexample.
Let $g(x) = \mathbf{1}_{[0,\infty)}(x)$ and $(X_t) \equiv (W_t)$. Then

$$\begin{aligned} u(t,x) &= \mathbb{P}(x + W_T - W_t \ge 0), \\ u_x^{'}(t,x) &= \frac{1}{\sqrt{2\pi(T-t)}} \exp -\frac{x^2}{2(T-t)}, \\ \mathbb{E}|\nabla_x u(t,X_t)|^2 &= \int_{\mathbb{R}} \frac{1}{2\pi(T-t)} \exp\{-\frac{x^2}{T-t}\} \frac{1}{\sqrt{2\pi t}} \exp\{-\frac{x^2}{2t}\} dx \\ &= \frac{1}{2\pi(T-t)} \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} \exp\{-\frac{T+t}{2t(T-t)} x^2\} dx \\ &= \frac{\sqrt{t}}{2\pi\sqrt{T+t}\sqrt{T-t}}, \end{aligned}$$

which is equivalent to $(T-t)^{-1/2}$, up to a constant, when $t \to T$. Then, it follows from Lemma 6 that $g \in \mathbf{L}_{2,\alpha}$ with $\alpha = \frac{1}{2}$ (but not with $\alpha > \frac{1}{2}$).
However, assertion (ii) of Proposition 4 cannot be satisfied, for this example, with $\beta = \alpha = \frac{1}{2}$, so neither assertion (i) (which is necessary to have the rate $\frac{1}{N}$, in view of the equivalence (18)) is.

Remark 11 *If the assertion (iii) of Proposition 4 is satisfied for some $\beta < 1$, using analogous arguments we obtain that the rate of convergence is $\frac{1}{N^\beta}$ with the equidistant time net $\pi^{(1)}$, and $\frac{1}{N}$ with the non equidistant net $\pi^{(\beta)}$.*

$\square$

1.3 Proofs

Proof of Lemma 2.
- *Estimate on u.* One has $u(t,X_t) = \mathbb{E}^{\mathcal{F}_t} g(X_T)$, so $\mathbb{E}|u(t,X_t)|^2 \le \mathbb{E}(\mathbb{E}^{\mathcal{F}_t}|g(X_T)|)^2 \le \mathbb{E}|g(X_T)|^2$.
- *First derivative of u.*
Suppose first that $d = q$. Then, under the ellipticity assumption, σ is invertible. It is known that, $\forall t \in [0,T)$, $\nabla u(t,.)$ can be represented as a conditional expectation (see Gobet and Munos [12] among others):

$$\nabla_x u(t,X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T) H_{t,T}^{(1)}\right],$$

where $H^{(1)}_{t,T}$ is the random variable (called Malliavin weight) given by

$$H^{(1)}_{t,T} = \frac{1}{T-t}\int_t^T \sigma^{-1}(s,X_s)\nabla X_s[\nabla X_t]^{-1}dW_s.$$

One uses the estimate (8) to get

$$\mathbb{E}^{\mathcal{F}_t}|H^{(1)}_{t,T}|^2 \le \frac{1}{(T-t)^2}\int_t^T \mathbb{E}^{\mathcal{F}_t}|\nabla X_s[\nabla X_t]^{-1}|^2 ds \le_c \frac{1}{T-t}.$$

Now, since $\mathbb{E}^{\mathcal{F}_t}(H^{(1)}_{t,T}) = 0$, one can write $\nabla_x u(t,X_t) = \mathbb{E}^{\mathcal{F}_t}[(g(X_T) - \mathbb{E}^{\mathcal{F}_t} g(X_T))H^{(1)}_{t,T}]$. The Cauchy-Schwartz inequality yields

$$|\nabla_x u(t,X_t)|^2 \le_c \frac{\mathbb{E}^{\mathcal{F}_t}|g(X_T) - \mathbb{E}^{\mathcal{F}_t} g(X_T)|^2}{T-t}. \tag{20}$$

So, we obtains $\mathbb{E}|\nabla_x u(t,X_t)|^2 \le_c \frac{V_{t,T}(g)}{T-t}$.
If $d \neq q$ (and always under the ellipticity assumption on σ), there exists a $d \times d$ symmetric invertible matrix Σ such that $\sigma\sigma^* = \Sigma^2$ (see Stroock and Varadhan [13], Lemma 5.2.1, to define a square root of $\sigma\sigma^*$). In addition, Σ satisfies the same regularity estimates as σ. Then, one can carry on the proof above, replacing σ by Σ, since the PDE (11) satisfied by u depends on σ only through $\sigma\sigma^*$.
Second derivative of u.
Suppose first that $d = q$, hence σ invertible. It is also known (see again Gobet and Munos [12]) that $\forall t \in [0,T)$, there exists a random variable $H^{(2)}_{t,T}$ such that

$$D^2 u(t,X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T)H^{(2)}_{t,T}\right],$$

and we can prove (as for the previous estimate) that $H^{(2)}_{t,T}$ satisfies,

$$\mathbb{E}^{\mathcal{F}_t}[H^{(2)}_{t,T}] = 0, \qquad \mathbb{E}^{\mathcal{F}_t}|H^{(2)}_{t,T}|^2 \le_c \frac{1}{(T-t)^2}.$$

Then the proof of the estimate of $\mathbb{E}|D^2u(t,X_t)|^2$ is the same as for $\mathbb{E}|\nabla_x u(t,X_t)|^2$. Note that the existence of $H^{(2)}_{t,T}$ relies on the existence of $\nabla(\nabla X)$, which holds under $(\mathbf{A}_{b,\sigma})$ because b and σ are both of class $\mathcal{C}^{2+\gamma}$ $(\gamma > 0)$.
If $d \neq q$ and $\sigma\sigma^*$ is elliptic, we proceed as for $\nabla_x u$ using the matrix $\Sigma = (\sigma\sigma^*)^{-1/2}$. $\square$

Proof of Corollary 3.
First, note that

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^T |\nabla_x u(t,X_t)| + |D^2u(t,X_t)|dt\Big)^2 \\
&\le 2\Big\{\int_0^T \Big(\mathbb{E}|\nabla_x u(t,X_t)|^2\Big)^{\frac12}dt\Big\}^2 + 2\Big\{\int_0^T \Big(\mathbb{E}|D^2u(t,X_t)|^2\Big)^{\frac12}dt\Big\}^2,
\end{aligned}$$

using the generalized Minkowski inequality. Besides, from Lemma 2, and using that $g \in \mathbf{L}_{2,\alpha}$, one obtains

$$\mathbb{E}\left|\nabla_x u(t,X_t)\right|^2 \leq_c \frac{K^\alpha(g)}{(T-t)^{1-\alpha}},\ \mathbb{E}\left|D^2 u(t,X_t)\right|^2 \leq_c \frac{K^\alpha(g)}{(T-t)^{2-\alpha}}.$$

Now, the required result easily follows. □

Proof of Proposition 4.
We prove that (i)⇒ (ii)⇒ (iii)⇒ (i).
(i)⇒ (ii). By Itô's rule,

$$\begin{aligned}\partial_{x_k} u(t,X_t) &= \partial_{x_k} u(0,X_0) + \int_0^t \left\{\nabla(\partial_{x_k} u)\sigma\right\}(s,X_s)dW_s \\ &\quad + \int_0^t \left\{\partial_t\partial_{x_k} u + \nabla(\partial_{x_k} u)b + \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^* D^2(\partial_{x_k} u)\right]\right\}(s,X_s)ds.\end{aligned}$$

In order to get rid of the terms $\partial_t\partial_{x_k} u$ and $D^2(\partial_{x_k} u)$, differentiate the PDE (11) solved by u:

$$\begin{aligned}0 &= \partial_{x_k}\Big(\partial_t u + \nabla u b + \frac{1}{2}\mathrm{Tr}\Big[\sigma\sigma^* D^2 u\Big]\Big) \\ &= \Big(\partial_t\partial_{x_k} u + \nabla(\partial_{x_k} u)b + \frac{1}{2}\mathrm{Tr}\Big[\sigma\sigma^* D^2(\partial_{x_k} u)\Big]\Big) + \Big(\nabla u\partial_{x_k} b + \frac{1}{2}\mathrm{Tr}\Big[\partial_{x_k}(\sigma\sigma^*)D^2 u\Big]\Big).\end{aligned}$$

Consequently

$$\begin{aligned}\partial_{x_k} u(t,X_t) &= \partial_{x_k} u(0,X_0) - \int_0^t \left\{\nabla u\partial_{x_k} b + \frac{1}{2}\mathrm{Tr}\left[\partial_{x_k}(\sigma\sigma^*)D^2 u\right]\right\}(s,X_s)ds \\ &\quad + \int_0^t \left\{\nabla(\partial_{x_k} u)\sigma\right\}(s,X_s)dW_s. \qquad (21)\end{aligned}$$

Then,

$$\begin{aligned}&\mathbb{E}\left|\nabla_x u(t,X_t)\right|^2 = \sum_{k=1}^d \mathbb{E}|\partial_{x_k} u(t,X_t)|^2 \\ &\leq_c \mathbb{E}\left|\nabla_x u(0,X_0)\right|^2 + \mathbb{E}\left(\int_0^t \left\{|\nabla_x u| + |D^2 u|\right\}(s,X_s)ds\right)^2 + \int_0^t \mathbb{E}|D^2 u(s,X_s)|^2 ds \\ &\leq_c \mathbb{E}\left|\nabla_x u(0,X_0)\right|^2 + \int_0^t \mathbb{E}|\nabla_x u(s,X_s)|^2 ds + \int_0^t \mathbb{E}|D^2 u(s,X_s)|^2 ds \qquad (22) \\ &\leq_c \frac{K^\alpha(g)}{T^{1-\alpha}} + \mathbb{E}|g(X_T)|^2 + \int_0^t \mathbb{E}|D^2 u(s,X_s)|^2 ds := \phi(t). \qquad (23)\end{aligned}$$

Then, by integrating by parts, one has

$$\begin{aligned}&\int_0^T (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2 dr \leq_c \overline{\lim_{s\uparrow T}} \int_0^s (T-r)^{-\beta}\phi(r)dr \\ &= \overline{\lim_{s\uparrow T}}\left(-\left[\frac{(T-r)^{1-\beta}}{1-\beta}\phi(r)\right]_0^s + \frac{1}{1-\beta}\int_0^s (T-r)^{1-\beta}\phi'(r)dr\right).\end{aligned}$$

Note that the first term in the limit is bounded by $\frac{T^{1-\beta}}{1-\beta}\left(\frac{K^\alpha(g)}{T^{1-\alpha}}+\mathbb{E}|g(X_T)|^2\right)$. The second term is bounded by $\frac{1}{1-\beta}\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2dr$, which is finite because assertion (i) is in force.

(ii)$\Rightarrow$ (iii): the proof is similar to the previous one. In view of (12-13), one has

$$\mathbb{E}\left|g(X_T)-\mathbb{E}^{\mathcal{F}_r}(g(X_T))\right|^2 \leq_c \int_r^T \mathbb{E}|\nabla_x u(s,X_s)|^2 ds := \Psi(r).$$

Then, using an integration by parts, one gets

$$\begin{aligned}
&\int_0^T (T-r)^{-1-\beta}\mathbb{E}\left|g(X_T)-\mathbb{E}^{\mathcal{F}_r}(g(X_T))\right|^2 dr\\
&\leq_c \varlimsup_{s\uparrow T}\int_0^s (T-r)^{-1-\beta}\Psi(r)dr\\
&\leq_c \varlimsup_{s\uparrow T}\left(\left[\frac{(T-r)^{-\beta}}{\beta}\Psi(r)\right]_0^s+\frac{1}{\beta}\int_0^s (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2dr\right).
\end{aligned}$$

The second term is finite according to assertion (ii). The first one is bounded by

$$\varlimsup_{s\uparrow T}\frac{(T-s)^{-\beta}}{\beta}\int_s^T \mathbb{E}|\nabla_x u(r,X_r)|^2dr \leq \varlimsup_{s\uparrow T}\frac{1}{\beta}\int_s^T (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2dr,$$

because $(T-r)^{-\beta}$ is increasing with respect to r. The limit above equals then to 0 since the related integral is convergent.

(iii)$\Rightarrow$ (i). From Lemma 2, one has $\mathbb{E}|D^2u(r,X_r)|^2 \leq_c \frac{V_{r,T}(g)}{(T-r)^2}$, from which we deduce (using assertion (iii)) that

$$\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2dr \leq_c \int_0^T (T-r)^{-1-\beta}V_{r,T}(g)dr < +\infty.$$

□

Proof of Lemma 6.
We prove that (i)$\Rightarrow$ (ii)$\Rightarrow$ (iii)$\Rightarrow$(i).

(i)$\Rightarrow$ (ii):

$$\begin{aligned}
\int_0^t \mathbb{E}|D^2u(s,X_s)|^2ds &= \sum_{k=1}^d \int_0^t \mathbb{E}|\nabla(\partial_{x_k}u)|^2(s,X_s)ds\\
&\leq_c \sum_{k=1}^d \int_0^t \mathbb{E}|\nabla(\partial_{x_k}u)\sigma|^2(s,X_s)ds,
\end{aligned}$$

by the ellipticity assumption. Then, by the isometry property and equality (21), one gets that $\int_0^t \mathbb{E}|D^2u(s,X_s)|^2 ds$ is bounded (up to a constant) by

$$\begin{aligned}
&\sum_{k=1}^d \mathbb{E}\Big(\int_0^t \nabla(\partial_{x_k}u)\sigma(s,X_s)dW_s\Big)^2 \\
&= \sum_{k=1}^d \mathbb{E}\Big(\partial_{x_k}u(t,X_t) - \partial_{x_k}u(0,X_0) + \int_0^t \Big\{\nabla u \partial_{x_k} b + \frac{1}{2}\mathrm{Tr}\Big[\partial_{x_k}(\sigma\sigma^*)D^2u\Big]\Big\}(s,X_s)ds\Big)^2 \\
&\leq_c \mathbb{E}\Big|\nabla_x u(t,X_t)\Big|^2 + \mathbb{E}\Big|\nabla_x u(0,X_0)\Big|^2 + \mathbb{E}\Big(\int_0^t \Big\{|\nabla_x u| + |D^2 u|\Big\}(s,X_s)ds\Big)^2 \\
&\leq_c \frac{K^\alpha(g)}{(T-t)^{1-\alpha}} + \frac{K^\alpha(g)}{T^{1-\alpha}} + K^\alpha(g)T^\alpha,
\end{aligned}$$

where we have used Lemma 2 and Corollary 3.

<u>(ii)⇒ (iii)</u>: this is an immediate consequence of the inequality (23).

<u>(iii)⇒ (i)</u>: from equation (12), one gets

$$V_{t,T}(g) \leq_c \int_t^T \mathbb{E}\left|\nabla_x u(s,X_s)\right|^2 ds \leq_c \int_t^T \frac{C^\alpha(g)}{(T-s)^{1-\alpha}} ds \leq_c C^\alpha(g)(T-t)^\alpha,$$

which means that $g \in \mathbf{L}_{2,\alpha}$.

The fact that (i) leads to (iv) follows from Lemma 2, and it is clear that, when $\alpha < 1$, (iv) leads to (ii). □

In the following section, we state some results on $Z_t - z_t$ that, put together with those of section 1, will be crucial to study the $\mathbf{L}_2$-regularity of the Z-component of BSDEs with non null generator.

2 A representation and an estimate of $Z_t - z_t$ when the terminal condition g is bounded and belongs to $\mathbf{L}_{2,\alpha}$

2.1 The main result

We define

$$Y_t^0 := Y_t - y_t, \qquad Z_t^0 := Z_t - z_t.$$

Then, the process (Y^0, Z^0) is the solution, in $\mathcal{S}^2 \times \mathcal{M}^2$ (because (Y,Z) and (y,z) are in such spaces), of the BSDE with null terminal condition and generator

$$f^0(t,x,y,z) := f\left(t,x,y+u(t,x), z + \nabla_x u(t,x)\sigma(t,x)\right),$$

i.e.

$$Y_t^0 = \int_t^T f^0(s, X_s, Y_s^0, Z_s^0)ds - \int_t^T Z_s^0 dW_s.$$

We set

$$\begin{aligned} a_r^0 &:= \nabla_x f^0(r, X_r, Y_r^0, Z_r^0); \\ b_r^0 &:= \nabla_y f^0(r, X_r, Y_r^0, Z_r^0); \\ c_r^0 &:= \nabla_z f^0(r, X_r, Y_r^0, Z_r^0). \end{aligned}$$

These quantities play a key role in the further estimates.
Note that

$$\begin{aligned} &|a_r^0| \le C\left(1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)|\right); \qquad (24)\\ &|b_r^0| \le C; \\ &|c_r^0| \le C. \end{aligned}$$

Hence, f^0 is Lipschitz continuous with respect to y and z, but not with respect to x because $\nabla_x u$ and $D^2 u$ may explode as t goes to T.
Our purpose is to estimate $Z - z = Z^0$, and it is known that usually the Z^0-component is related to the Malliavin derivative of the Y^0-component (see Proposition 5.3. in [14]). But this is stated under strong integrability conditions: namely in Prop.5.3.[14], it is required that $\mathbb{E}\int_0^T \int_0^T |D_\theta f^0(s, X_s, y, z)|^2 ds d\theta < \infty$, for any y and z. This is not satisfied in our case (since it essentially means that $\mathbb{E}\int_0^T |D^2 u(s, X_s)|^2 ds < +\infty$). However, we are going to prove that the expected result (relating Z^0 to Malliavin derivatives) holds in our setting ($g \in \mathbf{L}_{2,\alpha}$). We proceed by a localization of the generator (see paragraph 2.2).
Before giving our main result, we introduce (U, V) the solution of the linear BSDE

$$\begin{aligned} U_t = &\int_t^T \left\{ a_r^0 + U_r \left(b_r^0 I_d + b_r^{'} + \sum_{j=1}^q c_{j,r}^0 \sigma_{j,r}^{'} \right) + \sum_{j=1}^q V_r^j \left(c_{j,r}^0 I_d + \sigma_{j,r}^{'} \right) \right\} dr \\ &- \sum_{j=1}^q \int_t^T V_r^j dW_r^j, \end{aligned} \qquad (25)$$

where $b_r^{'}$ and $\sigma_{j,r}^{'}$ denote respectively $\nabla_x b(r, X_r)$ and $\nabla_x \sigma_j(r, X_r)$, and $c_{j,r}^0$ is the j-th component of c_r^0. It is well defined in $\mathcal{S}^2 \times \mathcal{M}^2$ (see Lemma 25 in Appendix A) because it follows from Lemma 3 and inequality (24) that

$$\mathbb{E}\left(\int_0^T |a_r^0| dr\right)^2 \le_c T^\alpha K^\alpha(g) < +\infty. \qquad (26)$$

Our main result is stated as follows:

Theorem 12 *Assume **($A_{b,\sigma}$)**, **(A_g)** and **(A_f)**. Assume moreover that $g \in$*

$\mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$. *Then,* $\mathrm{d}\mathbb{P} \otimes dt - a.s.$, *one has*

$$Z_t - z_t = U_t \sigma(t, X_t).$$

In particular, since $z_t + U_t\sigma(t, X_t)$ is continuous, Z has a continuous version : this extends the results by Ma and Zhang [15], in the case when g is continuously differentiable with bounded derivative. We work with this version in the sequel.

Remark 13 *$Z_t - z_t$ has also a closed representation as a conditional expectation: see equation (38) in the proof.*

It is now easy to derive pointwise and $\mathbf{L}_2$-estimates of $Z_t - z_t$ as t goes to T. We will not use the following estimates in the sequel, but we guess that they are interesting for themselves.

Corollary 14 *Assume $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_g)$ and $(\boldsymbol{A}_f)$. Assume moreover that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0,1]$. Then, for all $t \in [0,T)$, the following pointwise estimate and $\mathbf{L}_2$-estimate hold:*

$$|Z_t - z_t| \leq C \int_t^T \frac{\sqrt{\mathbb{E}^{\mathcal{F}_t}\left[\left(g(X_T) - \mathbb{E}^{\mathcal{F}_s} g(X_T) \right)^2 \right]}}{T-s} ds + C(T-t);$$

$$\mathbb{E}\, |Z_t - z_t|^2 \leq C K^\alpha(g)(T-t)^\alpha + C(T-t)^2.$$

Remark 15 *When g is bounded and α-Hölder continuous (i.e. $|g(x)-g(y)| \leq C(g)|x-y|^\alpha$), the pointwise estimate in Corollary 14 leads to $|Z_t - z_t| \leq_c C(g)(T-t)^{\frac{\alpha}{2}} + (T-t)$.*
Since $(z_t)_{0 \leq t \leq T}$ may explode at time T, this is a way to assert that Z and z are equivalent for times close to T.

Proof of Corollary 14. Theorem 12 and the conditional version of estimate (A.2) yield

$$|Z_t - z_t| \leq_c |U_t| \leq_c \left\{ \mathbb{E}^{\mathcal{F}_t} \Big(\int_t^T |a_r^0| dr \Big)^2 \right\}^{\frac{1}{2}} \leq_c \int_t^T \left(\mathbb{E}^{\mathcal{F}_t} |a_r^0|^2 \right)^{\frac{1}{2}} dr$$

using the generalized Minkowski inequality for the last inequality. From (24), one has

$$\mathbb{E}^{\mathcal{F}_t} |a_r^0|^2 \leq_c 1 + \mathbb{E}^{\mathcal{F}_t} |\nabla_x u(r, X_r)|^2 + \mathbb{E}^{\mathcal{F}_t} |D^2 u(r, X_r)|^2.$$

Therefore, by means of the pointwise estimates obtained in the proof of Lemma 2 (see inequality (20) for $\nabla_x u(r, X_r)$), one gets (for $t \leq r < T$)

$$\mathbb{E}^{\mathcal{F}_t} |a_r^0|^2 \leq_c 1 + \frac{\mathbb{E}^{\mathcal{F}_t}[(g(X_T) - \mathbb{E}^{\mathcal{F}_r} g(X_T))^2]}{(T-r)^2},$$

and

$$|Z_t - z_t| \leq_c \int_t^T 1 + \frac{\sqrt{\mathbb{E}^{\mathcal{F}_t}[(g(X_T) - \mathbb{E}^{\mathcal{F}_r} g(X_T))^2]}}{T-r} dr,$$

proving the pointwise estimate. Consequently, using the generalized Minkowski inequality and $g \in \mathbf{L}_{2,\alpha}$, one has

$$\begin{aligned}
\mathbb{E}\,|Z_t - z_t|^2 &\leq_c \Big(\int_t^T \frac{\sqrt{\mathbb{E}[(g(X_T) - \mathbb{E}^{\mathcal{F}_r} g(X_T))^2]}}{T-r} dr \Big)^2 + (T-t)^2 \\
&\leq_c \Big(\sqrt{K^\alpha(g)} \int_t^T \frac{1}{(T-r)^{1-\frac{\alpha}{2}}} dr \Big)^2 + (T-t)^2 \\
&\leq_c K^\alpha(g)(T-t)^\alpha + (T-t)^2.
\end{aligned}$$

□

2.2 Proof of Theorem 12

Since $a_r^0 = \nabla_x f^0(r, X_r, Y_r^0, Z_r^0)$ may explode as t goes to T, we proceed by a time localization of f^0 as follows: for $\varepsilon > 0$, we define:

$$f^\varepsilon(t,x,y,z) = f^0(t,x,y,z) \mathbb{1}_{t \leq T-\varepsilon}$$

and $(Y_t^\varepsilon, Z_t^\varepsilon)$ the solution, in $\mathcal{S}^2 \times \mathcal{M}^2$, of the localized BSDE:

$$Y_t^\varepsilon = \int_t^T f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon) ds - \int_t^T Z_s^\varepsilon dW_s.$$

As for a_s^0, b_s^0 and c_s^0, we define

$$\begin{aligned}
a_s^\varepsilon &:= \nabla_x f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon), \\
b_s^\varepsilon &:= \nabla_y f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon), \\
c_s^\varepsilon &:= \nabla_z f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon).
\end{aligned}$$

We assume $(\mathbf{A}_{b,\sigma})$, $(\mathbf{A}_g)$, $(\mathbf{A}_f)$, and that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0,1]$.
The idea of our proof of Theorem 12 is the following: we prove that Z^ε converges to Z^0 as ε goes to 0 (Lemma 16) and that $D_t Y_t^\varepsilon$ converges to some $\mathcal{D}_t Y_t^0$ satisfying a linear BSDE (Lemma 19). Then, since $Z_t^\varepsilon = D_t Y_t^\varepsilon$ (Lemma 17), we conclude that $Z_t^0 = \mathcal{D}_t Y_t^0$. Finally, we derive the BSDE (25) satisfied by $(U_t)_{0 \leq t \leq T}$ from that satisfied by $(\mathcal{D}_t Y_t^0)_{0 \leq t \leq T}$.

Step 1: Stability

Lemma 16

$$\lim_{\varepsilon \to 0} \mathbb{E}\left[\sup_{s \in [0,T]} |Y_s^\varepsilon - Y_s^0|^2 + \int_0^T |Z_s^\varepsilon - Z_s^0|^2 ds \right] = 0.$$

PROOF. We denote $Y_s^\varepsilon - Y_s^0$ by δY_s and $Z_s^\varepsilon - Z_s^0$ by δZ_s. Then $(\delta Y_s, \delta Z_s)_{0\le s\le T}$ is the solution in $\mathcal{S}^2 \times \mathcal{M}^2$ to the BSDE with a null terminal condition and the (random) generator

$$\begin{aligned}\delta f(t,y,z) &:= f^\varepsilon(t,X_t,y+Y_t^0,z+Z_t^0) - f^0(t,X_t,Y_t^0,Z_t^0)\\ &= \left[f^0(t,X_t,y+Y_t^0,z+Z_t^0) - f^0(t,X_t,Y_t^0,Z_t^0)\right]\mathbb{1}_{t\le T-\varepsilon}\\ &- f^0(t,X_t,Y_t^0,Z_t^0)\mathbb{1}_{t>T-\varepsilon}.\end{aligned}$$

Since the function $f^0(t,x,.,.)$ is Lipschitz continuous (and its Lipschitz constant is the same as that of f),

$$|\delta f(t,y,z)| \le C|y| + C|z| + \mathbb{1}_{t>T-\varepsilon}|f^0(t,X_t,Y_t^0,Z_t^0)|.$$

Then, thanks to a standard stability result (see Proposition 3.2 in [16]), one obtains the following estimate:

$$\mathbb{E}\left[\sup_{s\in[0,T]}|\delta Y_s|^2 + \int_0^T|\delta Z_s|^2 ds\right] \le C\mathbb{E}\left(\int_0^T \mathbb{1}_{s>T-\varepsilon}|f^0(s,X_s,Y_s^0,Z_s^0)|ds\right)^2.$$

Now, one has $f^0(s,X_s,Y_s^0,Z_s^0) = f(s,X_s,Y_s,Z_s)$, which is square integrable, since (Y,Z) is the solution in $\mathcal{S}^2\times\mathcal{M}^2$ of BSDE (1). Then, by the monotone convergence theorem, the above upper bound converges to 0 as ε goes to 0. □

Step 2: Malliavin derivatives

Lemma 17 *For any fixed $\varepsilon > 0$, $(Y^\varepsilon, Z^\varepsilon)$ belongs to $\mathbf{L}^2\left((0,T),\mathbb{D}^{1,2}\times(\mathbb{D}^{1,2})^q\right)$. Denoting their Malliavin derivatives by $(D_tY_s^\varepsilon, D_tZ_s^\varepsilon)$, one has, a.s., for all $t\in[0,T)$ (see [14]),*

$$Z_t^\varepsilon = D_tY_t^\varepsilon.$$

PROOF. This is a direct application of Proposition 5.9 of [14]. It remains to show that f^ε has bounded derivatives w.r.t (x,y,z). One has, for $w=x,y,z$,

$$\nabla_w f^\varepsilon(t,x,y,z) = \nabla_w f^0(t,x,y,z)\mathbb{1}_{t\le T-\varepsilon},$$

and, if one defines $\theta := (t,x,y+u(t,x),z+\nabla_x u(t,x)\sigma(t,x))$, then

$$\begin{aligned}&\nabla_x f^0(t,x,y,z) = \nabla_x f(\theta) + \nabla_y f(\theta)\nabla_x u(t,x) + \nabla_z f(\theta)\nabla_x[(\nabla_x u\sigma)^*](t,x),\\ &\nabla_y f^0(t,x,y,z) = \nabla_y f(\theta),\\ &\nabla_z f^0(t,x,y,z) = \nabla_z f(\theta).\end{aligned}$$

Hence, only the boundedness of $\nabla_x f^\varepsilon(t,x,y,z)$ needs to be justified. This readily follows from $|\nabla_x u(t,x)| + |D^2u(t,x)| \le C\|g\|_\infty/(T-t) \le C\|g\|_\infty/\varepsilon$, for $t\le T-\varepsilon$, where the first inequality is proved as in Lemma 2 (using the

boundedness of g).
Then, $dt \otimes d\mathbb{P}$-a.s., one has $Z_t^\varepsilon = D_t Y_t^\varepsilon$. In addition, in Ma and Zhang[15], it is proved that the above processes have a continuous version, which enables to pass to an a.s. equality for any t. $\square$

Note that, always from [14], $(D_t Y_s^\varepsilon)_{s\geq t}$ $(\varepsilon > 0)$ satisfies to the following linear BSDE:

$$D_t Y_s^\varepsilon = \int_s^T \left\{ a_r^\varepsilon D_t X_r + b_r^\varepsilon D_t Y_r^\varepsilon + \sum_{j=1}^q c_{j,r}^\varepsilon D_t Z_{j,r}^\varepsilon \right\} dr - \sum_{j=1}^q \int_s^T D_t Z_{j,r}^\varepsilon dW_r^j. \quad (27)$$

and, for $s \in [0,t), (D_t Y_s^\varepsilon, D_t Z_s^\varepsilon) = (0,0)$.
We introduce $(\mathcal{D}_t Y_s^0, \mathcal{D}_t Z_s^0)_{t\leq s\leq T}$, as the unique solution, in $\mathcal{S}^p \times \mathcal{M}^p$ (for $p \in (1,2)$) to the following BSDE:

$$\mathcal{D}_t Y_s^0 = \int_s^T \left\{ a_r^0 D_t X_r + b_r^0 \mathcal{D}_t Y_r^0 + \sum_{j=1}^q c_{j,r}^0 \mathcal{D}_t Z_{j,r}^0 \right\} dr - \sum_{j=1}^q \int_s^T \mathcal{D}_t Z_{j,r}^0 dW_r^j. \quad (28)$$

For $s \in [0,t), (\mathcal{D}_t Y_s^0, \mathcal{D}_t Z_s^0) := (0,0)$.
Note that BSDE (28) is well defined, applying Lemma 25. In fact, b_r^0 et c_r^0 are uniformly bounded, and from (24),

$$\mathbb{E}\left(\int_t^T \left| a_r^0 D_t X_r \right| dr \right)^p \leq_c \mathbb{E}\left[\sup_{0\leq r\leq T} |D_t X_r|^p \left(\int_0^T 1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)| dr \right)^p \right]. \quad (29)$$

This upper bound is finite using Hölder's inequality. Indeed, $\sup_{0\leq r\leq T} |D_t X_r|^p$ is in any $\mathbf{L}_q$ (see remark after inequality (9)) and the integral term is in $\mathbf{L}_2$ (Corollary 3).
Note that $\mathcal{D}_t Y_s^0$ is given by the following closed formula (which is standard for linear BSDEs, see e.g. [14])

$$\mathcal{D}_t Y_s^0 = \mathbb{E}^{\mathcal{F}_s} \int_s^T \Gamma_r^s a_r^0 D_t X_r dr, \quad (30)$$

where $(\Gamma_r^s)_{r\geq s}$ is the adjoint process defined by the forward linear SDE

$$d\Gamma_r^s = \Gamma_r^s \left(b_r^0 dr + c_r^0 dW_r \right), \qquad \Gamma_s^s = 1.$$

Furthermore, define $(\nabla Y_s^0, \nabla Z_s^0)_{0\leq s\leq T}$, to be the unique solution, in $\mathcal{S}^p \times \mathcal{M}^p$ $(1 < p < 2)$, to the following BSDE:

$$\nabla Y_s^0 = \int_s^T \left\{ a_r^0 \nabla X_r + b_r^0 \nabla Y_r^0 + \sum_{j=1}^q c_{j,r}^0 \nabla Z_{j,r}^0 \right\} dr - \sum_{j=1}^q \int_s^T \nabla Z_{j,r}^0 dW_r^j. \quad (31)$$

This is a slight abuse of notation because $(\nabla Y^0, \nabla Z^0)$ solves the BSDE obtained by differentiating with respect to x_0 the BSDE solved by (Y^0, Z^0), but we do not prove that $(\nabla Y^0, \nabla Z^0)$ are the gradients of (Y^0, Z^0) with respect to x_0 (however, this is true, using extra computations as before).
Then, from (9) it follows that

$$\mathcal{D}_t Y_s^0 = \nabla Y_s^0 [\nabla X_t]^{-1} \sigma(t, X_t), \qquad \mathcal{D}_t Z_s^0 = \nabla Z_s^0 [\nabla X_t]^{-1} \sigma(t, X_t). \tag{32}$$

Step 3: Convergence of $D_t Y_s^\varepsilon$ to $\mathcal{D}_t Y_s^0$ as ε goes to 0
To justify this step, we repeatedly use the lemma below.

Lemma 18 *Let φ be a bounded continuous deterministic function. Let (Θ^ε), (Θ) and (γ) be processes such that*

$$(i)\ \mathbb{E}\int_0^T |\Theta_s^0|^2 ds < +\infty;$$
$$(ii)\ \lim_{\varepsilon\to 0} \mathbb{E}\int_0^T |\Theta_s^\varepsilon - \Theta_s^0|^2 ds = 0;$$
$$(iii)\ \exists p \in (1,2) \text{ such that } \mathbb{E}\Big(\int_0^T |\gamma_s| ds\Big)^p < +\infty.$$

Then, $\lim_{\varepsilon\to 0} \mathbb{E}\Big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)||\gamma_s| ds\Big)^p = 0.$

PROOF. Let K_1, K_2 and δ be positive constants. It is clear that

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)||\gamma_s| ds\Big)^p \\
&\leq \big(2\|\varphi\|_\infty\big)^p \mathbb{E}\Big(\int_0^T |\gamma_s| \mathbb{1}_{|\gamma_s|>K_1} ds\Big)^p + K_1^p \big(2\|\varphi\|_\infty\big)^p \mathbb{E}\Big(\int_0^T \mathbb{1}_{|\Theta_s^0|>K_2} ds\Big)^p \\
&+ K_1^p \big(2\|\varphi\|_\infty\big)^p \mathbb{E}\Big(\int_0^T \mathbb{1}_{|\Theta_s^\varepsilon - \Theta_s^0|>\delta} ds\Big)^p \\
&+ K_1^p \mathbb{E}\Big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)| \mathbb{1}_{|\Theta_s^0|\leq K_2, |\Theta_s^\varepsilon - \Theta_s^0|\leq\delta} ds\Big)^p
\end{aligned} \tag{33}$$

Let $\eta > 0$. Firstly, by assumption (iii), K_1 can be chosen large enough to ensure that the first term is bounded by $\frac{\eta}{4}$. Besides, from Chebyshev inequality, one has

$$\mathbb{E}\mathbb{1}_{|\Theta_s^0|>K_2} + \mathbb{E}\mathbb{1}_{|\Theta_s^\varepsilon - \Theta_s^0|>\delta} \leq \frac{\mathbb{E}|\Theta_s^0|^{2/p}}{K_2^{2/p}} + \frac{\mathbb{E}|\Theta_s^\varepsilon - \Theta_s^0|^{2/p}}{\delta^{2/p}},$$

so that, owing to assumption (i), one can choose K_2 large enough to make the second term in (33) bounded by $\frac{\eta}{4}$. Thirdly, since φ is continuous on the compact $[-K_2 - 1, K_2 + 1]$, it is also uniformly continuous on the same compact, and one can choose δ small enough to ensure that the last term in

(33) is bounded by $\frac{\eta}{4}$. Finally, owing to assumption (ii), one can choose ε small enough to make the third term in (33) bounded by $\frac{\eta}{4}$.
Therefore, we have proved that for any $\eta > 0$, $\mathbb{E}\Big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)||\gamma_s|ds\Big)^p \leq \eta$, provided that ε is small enough. $\square$

Lemma 19

$$\lim_{\varepsilon\to 0} \mathbb{E}[\sup_{s\in[t,T]} |D_t Y_s^\varepsilon - \mathcal{D}_t Y_s^0|^p] = 0.$$

PROOF. $(D_t Y_s^\varepsilon - \mathcal{D}_t Y_s^0, D_t Z_s^\varepsilon - \mathcal{D}_t Z_s^0)_{t\leq s\leq T}$ satisfies the linear BSDE:

$$\begin{aligned} D_t Y_s^\varepsilon - \mathcal{D}_t Y_s^0 = &\int_s^T [(a_r^\varepsilon - a_r^0) D_t X_r + (b_r^\varepsilon - b_r^0)\mathcal{D}_t Y_r^0 + \sum_{j=1}^q (c_{j,r}^\varepsilon - c_{j,r}^0)\mathcal{D}_t Z_{j,r}^0] dr \\ &+ \int_s^T [b_r^\varepsilon \Big(D_t Y_r^\varepsilon - \mathcal{D}_t Y_r^0\Big) + \sum_{j=1}^q c_{j,r}^\varepsilon \Big(D_t Z_r^\varepsilon - \mathcal{D}_t Z_{j,r}^0\Big)] dr \\ &- \sum_{j=1}^q \int_s^T \Big(D_t Z_r^\varepsilon - \mathcal{D}_t Z_{j,r}^0\Big) dW_r^j. \end{aligned}$$

Set $\eta_r^{\varepsilon,a} := (a_r^\varepsilon - a_r^0) D_t X_r$, $\eta_r^{\varepsilon,b} := (b_r^\varepsilon - b_r^0)\mathcal{D}_t Y_r^0$ and $\eta_r^{\varepsilon,c} := \sum_{j=1}^q (c_{j,r}^\varepsilon - c_{j,r}^0)\mathcal{D}_t Z_{j,r}^0$. Using the a priori $\mathbf{L}_p$-estimate in Lemma 25, one has

$$\mathbb{E}[\sup_{s\in[t,T]} |D_t Y_s^\varepsilon - \mathcal{D}_t Y_s^0|^p] \leq_c \mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,a}| dr\Big)^p + \mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,b}| dr\Big)^p + \mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,c}| dr\Big)^p. \tag{34}$$

Let us prove that each contribution with $\eta_.^{\varepsilon,a}$, $\eta_.^{\varepsilon,b}$ and $\eta_.^{\varepsilon,c}$ converges to 0.
<u>Contribution with $\eta_.^{\varepsilon,a}$</u>. It is clear that

$$\begin{aligned} |\eta_r^{\varepsilon,a}| \leq &|\nabla_x f^0(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon) - \nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_t X_r| \\ &+ \mathbb{1}_{r>T-\varepsilon} |\nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_t X_r|. \end{aligned}$$

First, note that $|\nabla_x f^0(t,x,y,z)| \leq_c 1 + |\nabla_x u(t,x)| + |D^2 u(t,x)|$. This implies that, uniformly in ε, $\mathbb{1}_{r>T-\varepsilon}|\nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_t X_r| \leq_c \Big(1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)|\Big)|D_t X_r|$, whose integral w.r.t. r belongs to $\mathbf{L}_p$ (this has been established in the proof of the existence of BSDE (28)), and is consequently a.s. finite. It readily follows from the dominated convergence theorem that $\mathbb{E}\Big(\int_0^T \mathbb{1}_{r>T-\varepsilon}|\nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_t X_r|\Big)^p$ converges to 0 as ε goes to 0.
Next, setting $\Theta_r^\varepsilon := \Big(r, X_r, Y_r^\varepsilon + u(r, X_r), Z_r^\varepsilon + \nabla_x u(r, X_r)\sigma(r, X_r)\Big)$, $\varphi_w(\Theta) := \nabla_w f(\Theta)$, for $w = x, y, z$, and $\gamma_r := \Big(1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)|\Big)|D_t X_r|$, one has

$$|\nabla_x f^0(\Theta_r^\varepsilon) - \nabla_x f^0(\Theta_r^0)||D_t X_r| \leq_c \sum_{w=x,y,z} |\varphi_w(\Theta_r^\varepsilon) - \varphi_w(\Theta_r^0)|\gamma_r.$$

Owing to assumption $(\mathbf{A}_f)$, φ_w is continuous and bounded. Assumptions (i) and (ii) of Lemma 18 hold thanks to Lemma 16 and to the fact that $\Theta^0_r = (r, X_r, Y_r, Z_r)$. Assumption (iii) is checked since $\mathbb{E}(\int_0^T \gamma_s ds)^p < +\infty$, for $p \in (1,2)$ (see inequality (29) previously proved). It follows from Lemma 18 that $\mathbb{E}(\int_0^T |\nabla_x f^0(\Theta^\varepsilon_r) - \nabla_x f^0(\Theta^0_r)||D_t X_r| dr)^p$ converges to 0 as ε goes to 0. Finally,

$$\mathbb{E}\Big(\int_0^T |\eta^{\varepsilon,a}_r| dr\Big)^p \to 0. \tag{35}$$

<ins>Contribution with $\eta^{\varepsilon,b}_\cdot$</ins>. One has

$$|\eta^{\varepsilon,b}_r| \leq_c |\nabla_y f(\Theta^\varepsilon_r) - \nabla_y f(\Theta^0_r)||\mathcal{D}_t Y^0_r| + \mathbb{1}_{r>T-\varepsilon}|\nabla_y f^0(r, X_r, Y^0_r, Z^0_r)||\mathcal{D}_t Y^0_r|.$$

Then, we follow exactly the same proof as that of the contribution of $\eta^{\varepsilon,a}_\cdot$. One has only to check that $\gamma_r := |\mathcal{D}_t Y^0_r|$ satisfies assumption (iii) of Lemma 18. This readily follows from $(\mathcal{D}_t Y^0_r)_{t \leq r \leq T} \in \mathcal{S}^p$, which ensures that

$$\mathbb{E}\Big(\int_0^T \gamma_r dr\Big)^p \leq_c \mathbb{E} \sup_{r \in [t,T]} |\mathcal{D}_t Y^0_r|^p < +\infty.$$

Thus,

$$\mathbb{E}\Big(\int_0^T |\eta^{\varepsilon,b}_r| dr\Big)^p \to 0. \tag{36}$$

<ins>Contribution with $\eta^{\varepsilon,c}_\cdot$</ins>. One has

$$|\eta^{\varepsilon,c}_r| \leq_c |\nabla_z f(\Theta^\varepsilon_r) - \nabla_z f(\Theta^0_r)||\mathcal{D}_t Z^0_r| + \mathbb{1}_{r>T-\varepsilon}|\nabla_y f^0(r, X_r, Y^0_r, Z^0_r)||\mathcal{D}_t Z^0_r|.$$

Similarly, we check the integrability of $\gamma_r := |\mathcal{D}_t Z^0_r|$. Since $(\mathcal{D}_t Z^0_r)_{t \leq r \leq T} \in \mathcal{M}^p$,

$$\mathbb{E}\Big(\int_0^T \gamma_r dr\Big)^p \leq_c \mathbb{E}\Big(\int_0^T |\mathcal{D}_t Z^0_r|^2 dr\Big)^{\frac{p}{2}} < +\infty.$$

This gives

$$\mathbb{E}\Big(\int_0^T |\eta^{\varepsilon,c}_r| dr\Big)^p \to 0. \tag{37}$$

From (34), (35), (36) and (37), the proof is complete. $\square$

<ins>**Step 4: Proof of Theorem 12**</ins>

From Lemma 17 and Lemma 19, we know that $\lim_{\varepsilon \to 0} \mathbb{E}|Z^\varepsilon_t - D_t Y^0_t|^p = 0$. Besides, from Lemma 16, there is a subsequence $(\varepsilon_n)_{n \geq 1}$ decreasing to 0 such that $\lim_{n \to +\infty} Z^{\varepsilon_n}_t = Z^0_t$, $(d\mathbb{P} \otimes dt) - a.s.$. Thus, we conclude that, $(d\mathbb{P} \otimes dt) - a.s.$,

$$Z^0_t = D_t Y^0_t = \mathbb{E}^{\mathcal{F}_t} \int_t^T \Gamma^t_r a^0_r D_t X_r dr, \tag{38}$$

taking advantage of the explicit representation of $D_t Y^0_t$ in (30).

Then, from (32) and setting $U_t := \nabla Y_t^0 [\nabla X_t]^{-1}$, we have proved $Z_t^0 = U_t \sigma(t, X_t)$. It remains to show that U satisfies the BSDE (25). It is clear that

$$-dU_t = (-d\nabla Y_t^0)[\nabla X_t]^{-1} + \nabla Y_t^0(-d[\nabla X_t]^{-1}) - d < \nabla Y_t^0, [\nabla X_t]^{-1} >_t .$$

Besides, it is known (see e.g. [12]) that

$$-d[\nabla X_t]^{-1} = [\nabla X_t]^{-1}\Big\{\Big(b_t^{'} - \sum_{j=1}^{q}(\sigma_{j,t}^{'})^2\Big)dt + \sum_{j=1}^{q}\sigma_{j,t}^{'}dW_t^j\Big\}.$$

Then, from the expression of $d\nabla Y_t^0$ in (31), it follows that

$$\begin{aligned}
-dU_t &= \Big(a_t^0 + b_t^0 \nabla Y_t^0 [\nabla X_t]^{-1} + \sum_{j=1}^{q} c_{j,t}^0 \nabla Z_{j,t}^0 [\nabla X_t]^{-1}\Big)dt - \sum_{j=1}^{q} \nabla Z_{j,t}^0 [\nabla X_t]^{-1} dW_t^j \\
&\quad + \nabla Y_t^0 [\nabla X_t]^{-1}\Big\{\Big(b_t^{'} - \sum_{j=1}^{q}(\sigma_{j,t}^{'})^2\Big)dt + \sum_{j=1}^{q}\sigma_{j,t}^{'}dW_t^j\Big\} \\
&\quad + \sum_{j=1}^{q} \nabla Z_{j,t}^0 [\nabla X_t]^{-1}\sigma_{j,t}^{'}dt \\
&\quad = a_t^0 dt + U_t\Big(b_t^0 I_d + b_t^{'} + \sum_{j=1}^{q} c_{j,t}^0 \sigma_{j,t}^{'}\Big)dt \\
&\quad + \sum_{j=1}^{q}\Big(\nabla Z_{j,t}^0 [\nabla X_t]^{-1} - U_t\sigma_{j,t}^{'}\Big)\Big(c_{j,t}^0 I_d + \sigma_{j,t}^{'}\Big)dt \\
&\quad - \sum_{j=1}^{q}\Big(\nabla Z_{j,t}^0 [\nabla X_t]^{-1} - U_t\sigma_{j,t}^{'}\Big)dW_t^j.
\end{aligned}$$

By setting $V_t^j := \nabla Z_{j,t}^0 [\nabla X_t]^{-1} - U_t \sigma_{j,t}^{'}$, we obtain

$$-dU_t = \Big\{a_t^0 + U_t\Big(b_t^0 I_d + b_t^{'} + \sum_{j=1}^{q} c_{j,t}^0\sigma_{j,t}^{'}\Big) + \sum_{j=1}^{q} V_t^j\Big(c_{j,t}^0 I_d + \sigma_{j,t}^{'}\Big)\Big\}dt - \sum_{j=1}^{q} V_t^j dW_t^j.$$

□

3 $\mathbf{L}_2$-regularity of Z_t when the terminal condition $g \in \mathbf{L}_{2,\alpha}$, but is not necessarily bounded

3.1 The main results

In this section, we aim to establish an $\mathbf{L}_2$-regularity of the process $(Z_t)_t$, more precisely, to have a good rate of convergence of $\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|Z_s - \bar{Z}_{t_k}\right|^2 ds$, where

$$\bar{Z}_{t_k} := \frac{1}{t_{k+1} - t_k} \mathbb{E}^{\mathcal{F}_{t_k}} \int_{t_k}^{t_{k+1}} Z_u du.$$

In Zhang [1], it was shown that, for BSDEs with Lipschitz continuous terminal conditions, this rate is linear with respect to the time step $|\pi|$.

Here, we extend this result to non-Lipschitz terminal functions g: we show that, if we suppose that $g \in \bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$, we can obtain the same rate $|\pi|^\alpha$ for the equidistant time net $\pi^{(1)}$ or the rate $|\pi|$ for an appropriate choice of the time net.
In fact, we show that this $\mathbf{L}_2$-regularity of $(Z_t)_{0\le t\le T}$ can be deduced, under the assumption above on g, from that of the process $(z_t)_{0\le t\le T}$ (Theorem 20).

This is an interesting fact, since the study of the martingale integrand of the initial nonlinear BSDE can be reduced to that of the martingale integrand of the linear simpler BSDE with a null generator. We can then derive the desired rate (Theorem 21). We state below these two main results, that hold even if g is not bounded. Their proofs are postponed to the next paragraph.

Theorem 20 *Assume* $(\boldsymbol{A}_{b,\sigma})$*,* $(\boldsymbol{A}_f)$ *and that* $g \in \mathbf{L}_{2,\alpha}$*, for some* $\alpha \in (0,1]$*. Then, there is a positive constant* C *such that, for any time net* $\pi = \{t_k : k = 0...N\}$

$$\mathcal{E}(Z,\pi) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left| Z_s - \bar{Z}_{t_k} \right|^2 ds \le C\mathcal{E}(z,\pi) + C\Big(K^\alpha(g)T^\alpha + T^2\Big)|\pi|.$$

Theorem 21 *Assume* $(\boldsymbol{A}_{b,\sigma})$*,* $(\boldsymbol{A}_f)$ *and that* $g \in \mathbf{L}_{2,\alpha}$*, for some* $\alpha \in (0,1]$*. Then, there is a positive constant* C *(which does not depend on* N*) such that*

a) for the choice of the equidistant time net $\pi^{(1)}$*,*

$$\mathcal{E}(Z,\pi^{(1)}) \le C\frac{K^\alpha(g)T^\alpha + T^3N^{-1+\alpha}}{N^\alpha};$$

b) for the choice of $\pi^{(\beta)}$*, with* β *as in Corollary 10,*

$$\mathcal{E}(Z,\pi^{(\beta)}) \le C\frac{K^\alpha(g)T^\alpha + T^3}{N}.$$

Since all the bounds depend on the regularity of f only through $\|\nabla_x f\|_\infty$, $\|\nabla_y f\|_\infty$ and $\|\nabla_z f\|_\infty$, we can state the following theorem.

Theorem 22 *Assume* $(\boldsymbol{A}_{b,\sigma})$ *and that* $g \in \mathbf{L}_{2,\alpha}$*, for some* $\alpha \in (0,1]$*. Then, Theorem 21 still holds when the generator* f *is uniformly Lipschitz continuous in* x*,* y *and* z *(but not necessarily continuously differentiable).*

The above theorem is derived from Theorem 21 using standard stability results, we omit details.

3.2 Proof of Theorem 20

In order to be able to use the results of the previous sections, we begin by assuming that g is bounded.

Step 1: Proof when g is bounded

Suppose for the moment that g is bounded (and belongs to $\mathbf{L}_{2,\alpha}$).
Recalling that $Z^0_t = Z_t - z_t = U_t\sigma(t, X_t)$, we are going to exploit the BSDE (25) satisfied by U in order to bound $\mathbb{E}\int_{t_k}^{t_{k+1}} |Z^0_s - \bar{Z}^0_{t_k}|^2 ds$.
From the BSDE (25) satisfied by $(U_t, V_t)_t$ (Theorem 12), and using Lemma 25, one obtains the following estimate:

$$\mathbb{E}\Big[\sup_{r\in[0,T]} |U_r|^2 + \int_0^T |V_r|^2 dr\Big] \leq_c \mathbb{E}\Big(\int_0^T |a^0_r| dr\Big)^2.$$

Since $|a^0_r| \leq_c 1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)|$, we use Corollary 3 to obtain

$$\mathbb{E}\Big[\sup_{r\in[0,T]} |U_r|^2 + \int_0^T |V_r|^2 dr\Big] \leq_c T^\alpha K^\alpha(g) + T^2. \tag{39}$$

Let $0 \leq t \leq s \leq T$. Always from BSDE (25), and using the estimate (39), one has

$$\begin{aligned}\mathbb{E}|U_s - U_t|^2 &\leq_c \mathbb{E}\Big(\int_t^s |a^0_r| dr\Big)^2 + (s-t)^2 \mathbb{E}\sup_r |U_r|^2 + \mathbb{E}\int_t^s |V_r|^2 dr\\ &\leq_c \mathbb{E}\Big(\int_t^s |a^0_r| dr\Big)^2 + \Big[K^\alpha(g)T^\alpha + T^2\Big](s-t)^2 + \mathbb{E}\int_t^s |V_r|^2 dr. \end{aligned}\tag{40}$$

We go back to the regularity of Z. As we did in the proof of Theorem 8, we write, using Theorem 12 to have $Z^0_s = U_s\sigma(s, X_s)$, $(d\mathbb{P}\otimes dt) - a.s.$,

$$\mathbb{E}\int_{t_k}^{t_{k+1}} |Z^0_s - \bar{Z}^0_{t_k}|^2 ds \leq \mathbb{E}\int_{t_k}^{t_{k+1}} |U_s\sigma(s, X_s) - U_{t_k}\sigma(t_k, X_{t_k})|^2 ds.$$

Now, for $s \in [t_k, t_{k+1}]$,

$$\begin{aligned}\mathbb{E}|U_s\sigma(s, X_s) - U_{t_k}\sigma(t_k, X_{t_k})|^2 &= \mathbb{E}\Big|\Big(U_s - U_{t_k}\Big)\sigma(s, X_s) - U_{t_k}\Big(\sigma(t_k, X_{t_k}) - \sigma(s, X_s)\Big)\Big|^2\\ &\leq_c \mathbb{E}|U_s - U_{t_k}|^2 + |\pi|\mathbb{E}|U_{t_k}|^2.\end{aligned}$$

Then, using (39) and (40),

$$\mathbb{E}|U_s\sigma(s, X_s) - U_{t_k}\sigma(t_k, X_{t_k})|^2 \leq_c \mathbb{E}\Big(\int_{t_k}^{t_{k+1}} |a^0_r| dr\Big)^2 + \mathbb{E}\int_{t_k}^{t_{k+1}} |V_r|^2 dr + \Big[T^\alpha K^\alpha(g) + T^2\Big]|\pi|.$$

Therefore,

$$\begin{aligned}
&\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\Big|Z_s^0-\bar{Z}_{t_k}^0\Big|^2 ds\\
&\le_c |\pi|\sum_{k=0}^{N-1}\Big\{\mathbb{E}\Big(\int_{t_k}^{t_{k+1}}|a_r^0|dr\Big)^2+\mathbb{E}\int_{t_k}^{t_{k+1}}|V_r|^2dr\Big\}+\Big[T^\alpha K^\alpha(g)+T^2\Big]|\pi|\\
&\le_c |\pi|\Big\{\mathbb{E}\Big(\int_0^T|a_r^0|dr\Big)^2+\mathbb{E}\int_0^T|V_r|^2dr+(T^\alpha K^\alpha(g)+T^2)\Big\}\\
&\le_c |\pi|\Big(T^\alpha K^\alpha(g)+T^2\Big).
\end{aligned}$$

Now, since clearly $\Big|Z_s-\bar{Z}_{t_k}\Big|^2=\Big|(z_s+Z_s^0)-(\bar{z}_{t_k}+\bar{Z}_{t_k}^0)\Big|^2\le 2\Big|z_s-\bar{z}_{t_k}\Big|^2+2\Big|Z_s^0-\bar{Z}_{t_k}^0\Big|^2$, we conclude that

$$\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\Big|Z_s-\bar{Z}_{t_k}\Big|^2ds\le_c\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}|z_s-\bar{z}_{t_k}|^2ds+|\pi|\Big(T^\alpha K^\alpha(g)+T^2\Big). \tag{41}$$

□

Now, we suppose that g is not necessarily bounded (and g belongs to $\mathbf{L}_{2,\alpha}$). We use the following bounded approximation of g. For $M>0$ and $y\in\mathbb{R}$, we set

$$\phi_M(y):=-M\vee y\wedge M, \tag{42}$$

and

$$g_M:=\phi_M\circ g. \tag{43}$$

It is clear that, when $M\to+\infty$, $g_M(x)\to g(x)$ for all $x\in\mathbb{R}$ such that $|g(x)|<+\infty$, and $g_M(X_T)\to g(X_T)$ in $\mathbf{L}_2$.
We denote by (Y^M,Z^M) (resp. (y^M,z^M)) the solution to BSDE (1) (resp. BSDE (12)) with $g_M(X_T)$ as terminal condition instead of $g(X_T)$.

Step 2: Some stability results when M goes to $+\infty$

Lemma 23 *Assume ($\boldsymbol{A}_{b,\sigma}$) and that $g\in\mathbf{L}_{2,\alpha}$, for some $\alpha\in(0,1]$. Then, $g_M\in\mathbf{L}_{2,\alpha}$ and*

$$K^\alpha(g_M)\le K^\alpha(g).$$

PROOF. Recall that $y_t=\mathbb{E}^{\mathcal{F}_t}g(X_T)$, and set $y_t^M:=\mathbb{E}^{\mathcal{F}_t}g_M(X_T)=\mathbb{E}^{\mathcal{F}_t}\phi_M(y_T)$, where ϕ_M is the function already defined by (42). Note that ϕ_M is Lipschitz-continuous, with a Lipschitz constant equal to 1. One has

$$\begin{aligned} V_{t,T}(g_M) &= \mathbb{E}\Big|y_T^M - y_t^M\Big|^2 = \mathbb{E}\Big|\phi_M(y_T) - \mathbb{E}^{\mathcal{F}_t}\phi_M(y_T)\Big|^2 \\ &\le \mathbb{E}\Big|\phi_M(y_T) - \phi_M(y_t)\Big|^2 \le \mathbb{E}|y_T - y_t|^2 = V_{t,T}(g), \end{aligned}$$

where we used a projection property on $\mathbf{L}_2(\mathcal{F}_t)$ for the first inequality.
In addition, clearly $|g_M(X_T)| \le |g(X_T)|$, which readily finishes the proof. $\square$

Lemma 24

$$\lim_{M\to+\infty} \sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|z_s^M - \bar{z}_{t_k}^M\right|^2 ds = \sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2\, ds; \tag{44}$$

$$\lim_{M\to+\infty} \sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|Z_s^M - \bar{Z}_{t_k}^M\right|^2 ds = \sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|Z_s - \bar{Z}_{t_k}\right|^2 ds. \tag{45}$$

PROOF. We only prove (44), since for (45), the arguments are the same. Write

$$\begin{aligned} &\Big|\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|z_s^M - \bar{z}_{t_k}^M\right|^2 ds - \sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2\, ds\Big| \\ &= \Big|\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} \Big(\left|z_s - \bar{z}_{t_k}\right| - \left|z_s^M - \bar{z}_{t_k}^M\right|\Big)\Big(\left|z_s - \bar{z}_{t_k}\right| + \left|z_s^M - \bar{z}_{t_k}^M\right|\Big) ds\Big| \\ &\le \sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|(z_s - z_s^M) - (\bar{z}_{t_k} - \bar{z}_{t_k}^M)\right|\Big(\left|z_s - \bar{z}_{t_k}\right| + \left|z_s^M - \bar{z}_{t_k}^M\right|\Big) ds \\ &\le \sum_{k=0}^{N-1} \sqrt{\mathbb{E}\int_{t_k}^{t_{k+1}} \left|(z_s - z_s^M) - (\bar{z}_{t_k} - \bar{z}_{t_k}^M)\right|^2 ds}\sqrt{\mathbb{E}\int_{t_k}^{t_{k+1}} \Big(\left|z_s - \bar{z}_{t_k}\right| + \left|z_s^M - \bar{z}_{t_k}^M\right|\Big)^2 ds}. \end{aligned}$$

Now,

$$\begin{aligned} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|(z_s - z_s^M) - (\bar{z}_{t_k} - \bar{z}_{t_k}^M)\right|^2 ds &\le 2\mathbb{E}\int_0^T \left|z_s - z_s^M\right|^2 ds + 2\mathbb{E}\int_0^T \left|\bar{z}_{t_k} - \bar{z}_{t_k}^M\right|^2 ds \\ &\le 4\mathbb{E}\int_0^T \left|z_s - z_s^M\right|^2 ds, \end{aligned}$$

where we used a projection argument. By classical stability results for BSDEs, the term above tends to 0 when $M \to +\infty$. Besides,

$$\begin{aligned} \mathbb{E}\int_{t_k}^{t_{k+1}} \Big(\left|z_s - \bar{z}_{t_k}\right| + \left|z_s^M - \bar{z}_{t_k}^M\right|\Big)^2 ds &\le 4\mathbb{E}\int_0^T \Big(\left|z_s\right|^2 + \left|\bar{z}_{t_k}\right|^2 + \left|z_s^M\right|^2 + \left|z_{t_k}^M\right|^2\Big) ds \\ &\le C\mathbb{E}|g(X_T)|^2, \end{aligned}$$

where C does not depend on M (still using the classical a priori estimate for BSDEs). Thus, we have proved (44). $\square$

Step 3: Proof when g is not necessarily bounded

Applying (41) in Step 1 with g_M, one has

$$\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left|Z_s^M - \bar{Z}_{t_k}^M\right|^2 ds \leq_c \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left|z_s^M - \bar{z}_{t_k}^M\right|^2 ds + |\pi|\Big(T^\alpha K^\alpha(g_M) + T^2\Big),$$

and, since $K^\alpha(g_M) \leq K^\alpha(g)$ (Lemma 23),

$$\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left|Z_s^M - \bar{Z}_{t_k}^M\right|^2 ds \leq_c \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left|z_s^M - \bar{z}_{t_k}^M\right|^2 ds + |\pi|\Big(T^\alpha K^\alpha(g) + T^2\Big). \tag{46}$$

Passing to the limit when $M \to +\infty$ and using Lemma 24, we prove Theorem 20. □

3.3 Proof of Theorem 21

a) Equidistant time net $\pi^{(1)}$. As a direct consequence of Theorem 20 and Theorem 8, one has

$$\begin{aligned}\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k^{(N,1)}}^{t_{k+1}^{(N,1)}} \left|Z_s^M - \bar{Z}_{t_k^{(N,1)}}^M\right|^2 ds &\leq_c K^\alpha(g_M)\Big(\frac{T^\alpha}{N^\alpha} + \frac{T^{1+\alpha}}{N}\Big) + \frac{T^3}{N} \\ &\leq_c K^\alpha(g)\frac{T^\alpha}{N^\alpha} + \frac{T^3}{N}\end{aligned}$$

using Lemma 23. Passing to the limit when $M \to +\infty$ (Lemma 24), we prove estimate (a) of Theorem 21.
b) Time net $\pi^{(\beta)}$. The proof is the same, using Theorem 20, Corollary 10 and Lemma 23. □

A Appendix

The following lemma gives an a priori estimate for linear BSDEs, and is a direct consequence of Proposition 3.2 in [16] (applied with $f(\omega, t, y, z)$ of the form $a_t + b_t y + c_t z$).

Lemma 25 *Consider the linear BSDE*

$$U_t = \xi + \int_t^T \Big(a_r + U_r b_r + \sum_{j=1}^q V_{j,r} c_{j,r}\Big) dr - \sum_{j=1}^q \int_t^T V_{j,r} dW_r^j \tag{A.1}$$

where $\xi \in \mathbb{R}^{1\times d}, a_r \in \mathbb{R}^{1\times d}, b_r \in \mathbb{R}^{d\times d}, c_{j,r} \in \mathbb{R}^{d\times d}, U_r \in \mathbb{R}^{1\times d}, V_{j,r} \in \mathbb{R}^{1\times d}, W_r \in \mathbb{R}^q$*, for some progressively measurable coefficients* $(a_r)_r$*,* $(b_r)_r$*,* $(c_r)_r$ *and a* $\mathcal{F}_T$*-measurable terminal condition* ξ*.*
If $|b_r|$*,* $|c_{j,r}|$ *are uniformly bounded, and if* $\mathbb{E}|\xi|^p + \mathbb{E}\Big(\int_0^T |a_r|dr\Big)^p < +\infty$*, then there exists an unique solution* (U,V) *in* $\mathcal{S}^p \times \mathcal{M}^p$ *to BSDE (A.1), and the following estimate holds:*

$$\mathbb{E}\sup_{r\in[0,T]} |U_r|^p + \mathbb{E}\Big(\int_0^T |V_r|^2 dr\Big)^{\frac{p}{2}} \leq C\Big\{\mathbb{E}|\xi|^p + \mathbb{E}\Big(\int_0^T |a_r|dr\Big)^p\Big\}. \tag{A.2}$$

References

[1] J. Zhang, A numerical scheme for BSDEs, The Annals of Applied Probability 14 (1) (2004) 459–488.

[2] E. Pardoux, Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order, in: Stochastic analysis and related topics, VI (Geilo, 1996), Vol. 42 of Progr. Probab., Birkhäuser Boston, Boston, MA, 1998, pp. 79–127.

[3] E. Gobet, C. Labart, Solving BSDEs with adaptive control variate, in preparation.

[4] V. Bally, G. Pagès, Error analysis of the optimal quantization algorithm for obstacle problems, Stochastic Processes and their Applications 106 (1) (2003) 1–40.

[5] B. Bouchard, N. Touzi, Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations, Stochastic Processes and their Applications 111 (2004) 175–206.

[6] J. P. Lemor, E. Gobet, X. Warin, Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations, Bernoulli 12 (5) (2006) 889–916.

[7] E. Gobet, J. P. Lemor, Numerical simulation of bsdes using empirical regression methods: theory and practice, in: Proceedings of the Fifth Colloquium on BSDEs (29th May - 1st June 2005, Shangai) - Available on http://hal.archives-ouvertes.fr/hal-00291199/fr/, 2006.

[8] E. Gobet, E. Temam, Discrete time hedging errors for options with irregular pay-offs, Finance and Stochastics 5 (3) (2001) 357–367.

[9] S. Geiss, M. Hujo, Interpolation and approximation in $L_2(\gamma)$, Journal of Approximation Theory 144 (2007) 213–232.

[10] D. Nualart, Malliavin calculus and related topics, 2nd edition, Springer-Verlag, New York, 2005.

[11] A. Friedman, Partial differential equations of parabolic type, Prentice-Hall, 1964.

[12] E. Gobet, R. Munos, Sensitivity analysis using Itô-Malliavin calculus and martingales, and application to stochastic optimal control, SIAM Journal on control and optimization 43 (5) (2005) 1676–1713.

[13] D. W. Stroock, S. R. S. Varadhan, Multidimensional diffusion processes, Springer-Verlag, Berlin, 1979.

[14] N. El Karoui, S. Peng, M. C. Quenez, Backward stochastic differential equations in finance, Mathematical Finance 7 (1) (1997) 1–71.

[15] J. Ma, J. Zhang, Representation theorems for backward stochastic differential equations, The Annals of Applied Probability 12 (4) (2002) 1390–1418.

[16] P. Briand, B. Delyon, Y. Hu, E. Pardoux, L. Stoica, l^p solutions of backward stochastic differential equations., Stochastic Processes and their Applications 108 (1) (2003) 109–129.