

FBSDE with time delayed generators: L^p -solutions, differentiability, representation formulas and path regularity*

Gonalo dos Reis
CMAP
École Polytechnique
Route de Saclay
91128 Palaiseau Cedex
dosreis@cmap.polytechnique.fr

Anthony Réveillac
Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
areveill@math.hu-berlin.de

Jianing Zhang
Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
zhangj@math.hu-berlin.de

August 5, 2010

Abstract

We extend the work of Delong and Imkeller (2010a,b) concerning Backward stochastic differential equations with time delayed generators (delay BSDE). We provide sharper a priori estimates and show that the solution of a delay BSDE is in L^p . We introduce decoupled systems of SDE and delay BSDE (which we term delay FBSDE) and give sufficient conditions for the variational differentiability of their solutions. We connect these derivatives to the Malliavin derivatives of such delay FBSDE via the usual representation formulas which in turn give access to several path regularity results. In particular we prove an extension of the L^2 -path regularity result for delay FBSDE.

2010 AMS subject classifications: Primary: 60H10; Secondary: 60H30, 60H07, 60G17;

Key words and phrases: Backward stochastic differential equation, BSDE, delay, time delayed generators, L^p -solutions, differentiability, calculus of variations, Malliavin Calculus, path regularity.

Introduction

The theory of nonlinear *backward stochastic differential equations* (BSDEs) was introduced by Pardoux and Peng (1990) with its main motivations being mathematical finance (see El Karoui et al. (1997)) and stochastic control theory (see Yong and Zhou (1999)). In the last twenty years much effort has been given to this type of equations and nowadays many classes of BSDEs and results on them are available. Due to tractability, common results are achieved within a Markovian framework. Under certain conditions the BSDE's solution exhibits a Markov structure and hence can be interpreted as an instantaneous transformation of the underlying Markov process that spans the stochastic basis of the underlying probability space. This in turn yields access to the theory of partial differential equations via the non-linear Feynman-Kac formula.

*Gonalo dos Reis kindly acknowledges financial support by the *DFG Research Center MATHEON* to visit the Humboldt-Universität zu Berlin as well of the Chair *Financial risks* of the Risk Foundation sponsored by *Société Générale*. Anthony Réveillac is grateful to *DFG Research Center MATHEON*, project E2 for financial support. Jianing Zhang acknowledges financial support by the *DFG IRTG 1339 SMCP*.

Moving away from the Markovian setting, Delong and Imkeller (2010a,b) introduce a new class of BSDE labeled *backward stochastic differential equations with time delayed generators* (delay BSDEs). The dynamics of these BSDEs are governed by

$$Y_t = \xi + \int_t^T f(s, Y(s), Z(s))ds - \int_t^T Z_s dW_s, \quad t \in [0, T],$$

where the generator f at time $s \in [0, T]$ is allowed to depend on the past values of the solution (Y, Z) over the time interval $[0, s]$ and ξ is a measurable random variable. In these two works the authors answered thoroughly several fundamental questions: existence and uniqueness of a square integrable solution, comparison principles, existence of a measure solution, BMO martingale properties for the control component Z of the solution, Malliavin differentiability for delay BSDEs driven by a Wiener process and a generalized Poisson martingale. To the best of our knowledge the only existence and uniqueness results for this class of BSDEs follow from those two works. As pointed out by Delong (2010), delay BSDEs appear naturally in finance and insurance related problems of pricing and hedging of contracts. In the same work the author analyses a vast scope of contracts to which this class of BSDEs can be applied to.

Paying consideration to and seeking reference from the state of the art of BSDEs with non-time delayed generators, the next step concerning delay BSDEs is to obtain a feasible numerical scheme. Here, the main obstacle is the presence of the control process Z in the generator. This process is usually obtained via the predictable representation property of the underlying stochastic basis, and initially all one knows about Z is that it is a square integrable process. To steer in the direction of a numerical scheme a deeper analysis on the fine properties of the solution of such equations is required. As for numerics for Lipschitz continuous BSDEs (see for example Bouchard and Touzi (2004) or Bender and Denk (2007)) one is usually forced to gather several results concerning the *path regularity* properties of the solution process before being able to give proper convergence results. Such path properties include not only sample path continuity but also estimations on the time increments of the components of the solution by the size of the time increment. For the purpose of establishing such path properties we first need to prove several auxiliary results.

Our agenda consists of refining and extending the existence and uniqueness results obtained in Delong and Imkeller (2010a,b) and then steer into the direction of the smoothness properties of the solution of delay BSDEs. We start by improving the original results of Delong and Imkeller (2010a) concerning their a priori estimates by providing sharper versions of them. In Lemma 2.1 from Delong and Imkeller (2010a), a priori estimates are given expressing the difference (in norm) of solutions of two delay BSDE as the difference of the respective terminal conditions and generators. These a priori estimates fall short of the usual a priori estimates one expects to see due to the presence of the solutions of both delay BSDE on the right hand side of the estimate. We establish an a priori estimate in the classical form where the right hand side of the estimate contains the difference of generators evaluated at their zero spatial state and hence is independent of the BSDEs' solutions. Within the topic of a priori estimates we extend the results of Delong and Imkeller (2010a) in another direction. We show that given extra regularity of the terminal condition and the generator, the solution will inherit this regularity. This allows us to state moment and a priori estimates in general L^p -spaces and not solely in L^2 . The proof of these estimates relies on techniques from Delong and Imkeller (2010a) and on computations carried out for non-time delayed BSDEs in the spirit of Wang et al. (2007). The usual techniques to obtain higher order moment estimates fail in the setting of delay BSDEs, the reason for this can be seen in (10) below - usually the dynamics of Y_t is given by sums of integrals over the interval $[t, T]$ but for delay BSDEs we see from (10) that the dynamics of Y_t depends also on a

integral over the whole interval $[0, T]$. These estimates pave the way to a result of existence and uniqueness of solutions to delay BSDE with Lipschitz continuous generators in general L^p spaces for $p \geq 2$. Inevitably, in analogy to Delong and Imkeller (2010a,b) a compatibility condition on the Lipschitz constant and terminal time is required to obtain existence of solutions (see our Theorem 2.8).

A customary field of application of BSDEs consists in coupling them with SDEs, giving rise (in our case) to systems of delay forward-backward SDEs (delay FBSDEs). We show that when coupling a delay BSDE with a forward diffusion assuming appropriate regularity conditions, we obtain smoothness properties of the solution in terms of the involved parameters, in particular with respect to the initial condition of the forward diffusion. Combining this with the Malliavin differentiability proved in Delong and Imkeller (2010b) enables us to derive the usual representation formulas for FBSDE which display the relationship between the Malliavin derivatives of the solution process and their variational (classical) derivatives. It is somewhat surprising that such a relationship still holds for they are usually consequences of the BSDE's Markov property which due to path dependency clearly fails to materialize in the context of delay FBSDE.

With this collection of results we are finally able to address the path regularity issue of delay BSDE. Using the techniques employed in Imkeller and Dos Reis (2010a,b), we establish path continuity for the components of the solution of delay FBSDE and we give a result that bounds the norm of the increments in time of Y and Z by the size of the time increment. We expect that these results will open the door to the derivation of concrete numerical schemes and their convergence rate and intend to tackle these problems in our future research.

The paper is organized as follows: in Section 1 we fix notations and elaborate on the type of time-delayed BSDEs that we consider. In Section 2 we refine and extend the a priori estimates obtained in Delong and Imkeller (2010a) and then use them to establish existence and uniqueness of solutions in general L^p spaces. In Section 3 we introduce the delay FBSDE framework and use results from the previous sections to obtain the differentiability of the solution process with respect to the initial state of a forward diffusion. The representation formulas and the path regularity results are presented in Section 4.

1 Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a standard d -dimensional Brownian motion W . For a fixed real number $T > 0$ we consider the filtration $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$ generated by W and augmented by all \mathbb{P} -null sets. The filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfies the usual conditions. Depending on whether we work on \mathbb{R}^d or $\mathbb{R}^{d \times m}$, the Euclidean norm respectively the Hilbert-Schmidt operator norm is denoted by $|\cdot|$. Furthermore, ∇ denotes the canonical gradient differential operator and for a function $h(x, y) : \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^n$, we write $\nabla_x h$ or $\nabla_y h$ for the derivatives with respect to x and y . We work with the following topological vector spaces:

- For $p \geq 2$, let $L^p(\mathbb{R}^m)$ be the space of \mathcal{F}_T -measurable random variables $\xi : \Omega \rightarrow \mathbb{R}^m$ normed by $\|\xi\|_{L^p} := \mathbb{E}[|\xi|^p]^{1/p}$.
- For $\beta \geq 0$ and $p \geq 1$, $\mathcal{H}_\beta^p(\mathbb{R}^{d \times m})$ denotes the space of all predictable process φ with values in $\mathbb{R}^{d \times m}$ such that the norm $\|\varphi\|_{\mathcal{H}_\beta^p} := \mathbb{E}\left[\left(\int_0^T e^{\beta s} |\varphi_s|^2 ds\right)^{p/2}\right]^{1/p} < \infty$.
- For $\beta \geq 0$ and $p \geq 2$, $\mathcal{S}_\beta^p(\mathbb{R}^{d \times m})$ denotes the space of all predictable processes η with values in $\mathbb{R}^{d \times m}$ such that the norm $\|\eta\|_{\mathcal{S}_\beta^p} := \mathbb{E}\left[\left(\sup_{0 \leq t \leq T} e^{\beta t} |\eta_t|^2\right)^{p/2}\right]^{1/p} < \infty$.

We omit referencing the range space if no ambiguity arises. It is fairly easy to see that for any $\beta, \bar{\beta} \geq 0$ the norms on \mathcal{H}_β^p , $\mathcal{H}_{\bar{\beta}}^p$ and \mathcal{S}_β^p , $\mathcal{S}_{\bar{\beta}}^p$ are equivalent.

Some notation

We introduce a notational convention which will be used throughout the text: for an arbitrarily given integrable function $f : [0, T] \rightarrow \mathbb{R}^m$, trivially extended to $[-T, 0)$ via $f(t)\mathbb{1}_{[-T, 0)}(t) = 0$, and a given deterministic measure α supported on $[-T, 0]$ which is not necessarily atomless, we denote

$$\begin{aligned}(f \cdot \alpha)(t) &:= \int_{-T}^0 f(t+v)\alpha(dv), \quad t \in [0, T], \\ (f^p \cdot \alpha)(t) &:= \int_{-T}^0 |f(t+v)|^p \alpha(dv), \quad t \in [0, T], \quad p \geq 2.\end{aligned}$$

Similarly, for a given process $(\varphi_t)_{t \in [0, T]}$, extended to $[-T, 0)$ by imposing $\varphi_t = 0$ on $[-T, 0)$, we denote

$$(\varphi \cdot \alpha)(t) := \int_{-T}^0 \varphi_{t+v} \alpha(dv), \quad t \in [0, T], \quad (1)$$

and

$$(\varphi^p \cdot \alpha)(t) := \int_{-T}^0 |\varphi_{t+v}|^p \alpha(dv), \quad t \in [0, T], \quad p \geq 2. \quad (2)$$

We now give a lemma concerning the change of integration order for (1) and (2), which will become useful in the sequel.

Lemma 1.1. *Let φ be a process and α a non-random finite measure supported on $[-T, 0)$. Then we have the following change of integration order: for every $k \geq 1$*

$$\int_t^T (\varphi^k \cdot \alpha)(s) ds = \int_0^T \alpha([r - T, (r - t) \wedge 0)) |\varphi_r|^k dr, \quad \forall t \in [0, T], \quad \mathbb{P} - a.s.$$

Moreover, if we have for $p \geq 1$ that $\varphi \in \mathcal{H}_0^p$, then we also have that

$$\|(\varphi \cdot \alpha)\|_{\mathcal{H}_\beta^p}^p \leq M_p \|\varphi\|_{\mathcal{H}_0^p}^p,$$

where $M_p = (e^{\beta T})^{p/2} (\alpha([-T, 0)))^p$.

Proof. Let t in $[0, T]$ and $k \in [1, +\infty)$. We have that

$$\begin{aligned}\int_t^T (\varphi^k \cdot \alpha)(s) ds &= \int_t^T \int_{-T}^0 |\varphi_{s+v}|^k \alpha(dv) ds = \int_{-T}^0 \int_t^T |\varphi_{s+v}|^k ds \alpha(dv) \\ &= \int_{-T}^0 \int_{(t+v) \vee 0}^{T+v} |\varphi_r|^k dr \alpha(dv) = \int_0^T \int_{(r-T)}^{(r-t) \wedge 0} |\varphi_r|^k \alpha(dv) dr \\ &= \int_0^T \alpha([r - T, (r - t) \wedge 0)) |\varphi_r|^k dr.\end{aligned}$$

The second claim follows by applying Jensen's inequality and changing the integration order as done above, i.e. for any $\beta \geq 0$ and $p \geq 1$ we have

$$\begin{aligned}\mathbb{E} \left[\left(\int_0^T e^{\beta s} |(\varphi \cdot \alpha)(s)|^2 ds \right)^{p/2} \right] &\leq (e^{\beta T} \alpha([-T, 0)))^{p/2} \mathbb{E} \left[\left(\int_0^T (|\varphi|^2 \cdot \alpha)(s) ds \right)^{p/2} \right] \\ &\leq M_p \mathbb{E} \left[\left(\int_0^T |\varphi_s|^2 ds \right)^{p/2} \right] = M_p \|\varphi\|_{\mathcal{H}_0^p}^p,\end{aligned}$$

which concludes the proof. \square

2 General results on BSDE with time delayed generators

In this section we give a brief recapitulation of BSDE with time delayed generators and discuss the setting they are studied under. We then establish convenient a priori estimates on the difference of two solutions to such equations which will play a central role in proving existence and uniqueness of solutions in the more general \mathcal{H}^p -spaces.

2.1 BSDEs with time delayed generators

Let us start with a recap on BSDE with time delayed generators. For two non-random finite measures α_Y, α_Z supported on $[-T, 0)$, we define

$$\alpha := \alpha_Y([-T, 0)) \vee \alpha_Z([-T, 0)). \quad (3)$$

Given $p \geq 2$, we assume that the following holds:

- (H1) ξ is an \mathcal{F}_T -measurable random variable which belongs to $L^p(\mathbb{R}^m)$;
- (H2) the generator $f : \Omega \times [0, T] \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \rightarrow \mathbb{R}^m$ is measurable, \mathbb{F} -adapted and satisfies the following Lipschitz condition: there exists a constant $K > 0$ such that

$$|f(t, y, z) - f(t, y', z')|^2 \leq K(|y - y'|^2 + |z - z'|^2)$$

holds for $d\mathbb{P} \otimes dt$ -almost all $(\omega, t) \in \Omega \times [0, T]$ and for every $(y, z), (y', z') \in \mathbb{R}^m \times \mathbb{R}^{d \times m}$;

- (H3) $\mathbb{E} \left[\left(\int_0^T |f(s, 0, 0)|^2 ds \right)^{p/2} \right] < \infty$;

- (H4) $f(t, \cdot, \cdot) = 0$ if $t < 0$.

Following the notation from equation (1), we write

$$(Y \cdot \alpha_Y)(t) = \int_{-T}^0 Y_{t+v} \alpha_Y(dv) \quad \text{and} \quad (Z \cdot \alpha_Z)(t) = \int_{-T}^0 Z_{t+v} \alpha_Z(dv), \quad 0 \leq t \leq T,$$

for some processes $(Y_t)_{t \in [0, T]}$ and $(Z_t)_{t \in [0, T]}$ satisfying appropriate integrability conditions. Assumption (H2) and Jensen's inequality then imply

$$\begin{aligned} \text{(H2')} \quad & |f(t, (Y \cdot \alpha_Y)(t), (Z \cdot \alpha_Z)(t)) - f(t, (Y' \cdot \alpha_Y)(t), (Z' \cdot \alpha_Z)(t))|^2 \\ & \leq K \{ |((Y - Y') \cdot \alpha_Y)(t)|^2 + |((Z - Z') \cdot \alpha_Z)(t)|^2 \} \\ & \leq L \{ ((Y - Y')^2 \cdot \alpha_Y)(t) + ((Z - Z')^2 \cdot \alpha_Z)(t) \}, \end{aligned}$$

where $L = K\alpha$ with the real number α given by (3). The focus of our study are BSDE with time delayed generators which are of the type

$$Y_t = \xi + \int_t^T f(s, \Gamma(s)) ds - \int_t^T Z_s dW_s, \quad 0 \leq t \leq T, \quad (4)$$

where Γ abbreviates

$$\Gamma(t) := \left(\int_{-T}^0 Y_{t+v} \alpha_Y(dv), \int_{-T}^0 Z_{t+v} \alpha_Z(dv) \right) = \left((Y \cdot \alpha_Y)(t), (Z \cdot \alpha_Z)(t) \right), \quad 0 \leq t \leq T. \quad (5)$$

Using a fixed point argument, Delong and Imkeller (2010a) have shown that a BSDE of the type (4)-(5) admits a unique solution if the parameters of the equation (4) are sufficiently small, i.e. if the Lipschitz constant $K > 0$ or the terminal time $T > 0$ satisfy a smallness condition. The following L^2 existence and uniqueness result is a straightforward modification of Theorem 2.1 from Delong and Imkeller (2010a).

Theorem 2.1. *Let $p = 2$ and assume that (H1)-(H4) are satisfied. Assume that the non-negative constants $T, L = K\alpha, \beta$ are such that*

$$(8T + \frac{1}{\beta})L \int_{-T}^0 e^{-\beta u} \rho(du) \max\{1, T\} < 1.$$

where $\rho \in \{\alpha_y, \alpha_z\}$. Then the delay BSDE (4)-(5) has a unique solution $(Y, Z) \in \mathcal{S}_\beta^2(\mathbb{R}^m) \times \mathcal{H}_\beta^2(\mathbb{R}^{d \times m})$.

Remark 2.2. *In Delong and Imkeller (2010a), this result is proved for the one-dimensional case $d = m = 1$. It is clear that by the nature of the fixed point argument, the proof is insensitive to dimension of the equation.*

Remark 2.3. *Given that a compatibility condition is necessary in order to establish existence and uniqueness of solutions and moreover that we will be giving an extended version of it, all the proofs in this section are given with extra detail in order to better control the constants involved in each result.*

2.2 A “sharper” a priori estimate

In Delong and Imkeller (2010a), the authors provide a priori estimates for the time delayed BSDE (4) (see their Lemma 2.1), which estimates the norms of the difference between two BSDE solutions in terms of the terminal condition and the generator. It is worth mentioning that Lemma 2.1 from Delong and Imkeller (2010a) establishes a priori estimates whose right hand side again depends on the solution of the delay BSDE. In the context of Delong and Imkeller (2010a), such a result suffices to establish existence and uniqueness of solutions in $\mathcal{S}_\beta^2 \times \mathcal{H}_\beta^2$, but the situation becomes more intricate when the same issues are considered on $\mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ for $p > 2$. In this section we refine the estimates from Delong and Imkeller (2010a) by providing a right hand side which only depends on the problem’s data (i.e. the differences between the terminal conditions and the corresponding states of the generators at $y = 0$ and $z = 0$).

As a starting observation, we have that if (4) admits a solution (Y, Z) in $\mathcal{H}_\beta^p(\mathbb{R}^m) \times \mathcal{H}_\beta^p(\mathbb{R}^{m \times d})$, then we also have that $Y \in \mathcal{S}_\beta^p(\mathbb{R}^m)$.

Lemma 2.4. *Let $\beta \geq 0, p \geq 2$ and assume that (H1)-(H4) hold. If the delay BSDE (4) admits a solution $(Y, Z) \in \mathcal{H}_\beta^p(\mathbb{R}^m) \times \mathcal{H}_\beta^p(\mathbb{R}^{m \times d})$, then we also have that $Y \in \mathcal{S}_\beta^p(\mathbb{R}^m)$.*

Proof. Throughout let $t \in [0, T]$ and $p \geq 2$. Since all β -norms are equivalent, it suffices to show the result for $\beta = 0$. We drop the β -subscripts in the following. The pair (Y, Z) satisfies

$$Y_t = \xi + \int_t^T f(s, (Y \cdot \alpha_y)(s), (Z \cdot \alpha_z)(s)) ds - \int_t^T Z_s dW_s,$$

hence we have

$$\sup_{0 \leq t \leq T} |Y_t| \leq |\xi| + \int_0^T |f(s, (Y \cdot \alpha_y)(s), (Z \cdot \alpha_z)(s))| ds + \sup_{0 \leq t \leq T} \left| \int_t^T Z_s dW_s \right|.$$

Combining the fact of $Z \in \mathcal{H}^p$ with the inequalities by Young, Doob and Burkholder-Davis-Gundy (BDG), we obtain

$$\begin{aligned}
& \mathbb{E} \left[\left(\sup_{0 \leq t \leq T} \left| \int_t^T Z_s dW_s \right|^2 \right)^{p/2} \right] \\
& \leq 2^{p/2} \mathbb{E} \left[\left(\left| \int_0^T Z_s dW_s \right|^2 + \sup_{0 \leq t \leq T} \left| \int_0^t Z_s dW_s \right|^2 \right)^{p/2} \right] \\
& \leq 2^{p/2} \mathbb{E} \left[\left(2 \sup_{0 \leq t \leq T} \left| \int_0^t Z_s dW_s \right|^2 \right)^{p/2} \right] = 2^p \mathbb{E} \left[\sup_{0 \leq t \leq T} \left| \int_0^t Z_s dW_s \right|^p \right] \\
& \leq 2^p C_p \mathbb{E} \left[\left(\int_0^T |Z_s|^2 ds \right)^{p/2} \right] < \infty.
\end{aligned}$$

Next observe that by the Lipschitz property of the generator f (notice that (H2) implies (H2')), it follows that

$$\begin{aligned}
& \left(\int_0^T |f(s, (Y \cdot \alpha_Y)(s), (Z \cdot \alpha_Z)(s))|^2 ds \right)^{p/2} \\
& \leq 2^{p/2} \left(\int_0^T |f(s, 0, 0)|^2 ds + \int_0^T |f(s, (Y \cdot \alpha_Y)(s), (Z \cdot \alpha_Z)(s)) - f(s, 0, 0)|^2 ds \right)^{p/2} \\
& \leq 2^{p/2} 2^{p/2-1} \left\{ \left(\int_0^T |f(s, 0, 0)|^2 ds \right)^{p/2} + \left(L \int_0^T (|Y|^2 \cdot \alpha_Y)(s) + (|Z|^2 \cdot \alpha_Z)(s) ds \right)^{p/2} \right\}.
\end{aligned}$$

The second term in the bracket can be further estimated by

$$\begin{aligned}
& \left(L \int_0^T (|Y|^2 \cdot \alpha_Y)(s) + (|Z|^2 \cdot \alpha_Z)(s) ds \right)^{p/2} \\
& \leq 2^{p/2-1} L^{p/2} \left\{ \left(\int_0^T (|Y|^2 \cdot \alpha_Y)(s) ds \right)^{p/2} + \left(\int_0^T (|Z|^2 \cdot \alpha_Z)(s) ds \right)^{p/2} \right\} \\
& \leq 2^{p/2-1} L^{p/2} \alpha^{p/2} \left\{ \left(\int_0^T |Y_s|^2 ds \right)^{p/2} + \left(\int_0^T |Z_s|^2 ds \right)^{p/2} \right\},
\end{aligned}$$

where the last line follows from Lemma 1.1. This estimate together with (H3) yields

$$\mathbb{E} \left[\left(\int_0^T |f(s, (Y \cdot \alpha_Y)(s), (Z \cdot \alpha_Z)(s))|^2 ds \right)^{p/2} \right] < \infty.$$

Using hypothesis (H1), i.e. that ξ is in L^p , we can conclude that $Y \in \mathcal{S}^p$ must hold. \square

Let us define the maximum of the weighted measure of $[-T, 0)$ via

$$\tilde{\alpha} := \int_{-T}^0 e^{-\beta s} \alpha_Y(ds) \vee \int_{-T}^0 e^{-\beta s} \alpha_Z(ds), \quad \beta \geq 0. \tag{6}$$

The next results establish a priori estimates for the solutions of two time-delayed BSDEs as given by (4). We distinguish between the cases $p = 2$ and $p > 2$, and we shall start with the case $p = 2$ in the following

Proposition 2.5 (A priori estimates for $p = 2$). *Let $p = 2$ and $\beta, \gamma > 0$. Consider $i \in \{1, 2\}$ and let $(Y^i, Z^i) \in \mathcal{S}_\beta^2 \times \mathcal{H}_\beta^2$ be the solution of the delay BSDE (4) with terminal condition ξ^i*

and generator f^i satisfying (H1)-(H4). Denote by $L > 0$ the Lipschitz constant of f^1 as given in (H2') and set $\delta Y = Y^1 - Y^2$, $\delta Z = Z^1 - Z^2$ and $\delta_2 f_t = f^1(t, (Y^2 \cdot \alpha_Y)(t), (Z^2 \cdot \alpha_Z)(t)) - f^2(t, (Y^2 \cdot \alpha_Y)(t), (Z^2 \cdot \alpha_Z)(t))$ for $t \in [0, T]$. Assume that $\beta, \gamma > 0$ satisfy

$$\gamma > \tilde{\alpha}L \quad \text{and} \quad \beta - \gamma - \frac{\tilde{\alpha}L}{\gamma} > 0. \quad (7)$$

Then there exists a constant $C_2 = C_2(\beta, \gamma, \tilde{\alpha}, L) > 0$ which depends on $\beta, \gamma, \tilde{\alpha}, L$ such that the following a priori estimate holds:

$$\|\delta Y\|_{\mathcal{S}_\beta^2}^2 + \|\delta Y\|_{\mathcal{H}_\beta^2}^2 + \|\delta Z\|_{\mathcal{H}_\beta^2}^2 \leq C_2 \left\{ \mathbb{E} \left[\left(e^{\beta T} |\delta Y_T|^2 \right) \right] + \mathbb{E} \left[\int_0^T e^{\beta s} |\delta_2 f_s|^2 ds \right] \right\}. \quad (8)$$

Proof. Throughout let $t \in [0, T]$, $i \in \{1, 2\}$ and define Γ^i as in (5) for the pair (Y^i, Z^i) . An application of Itô's formula to the semimartingale $e^{\beta t} |\delta Y_t|^2$ for $\beta > 0$ yields

$$\begin{aligned} & e^{\beta t} |\delta Y_t|^2 + \int_t^T \beta e^{\beta s} |\delta Y_s|^2 ds + \int_t^T e^{\beta s} |\delta Z_s|^2 ds \\ &= e^{\beta T} |\delta Y_T|^2 + \int_t^T 2e^{\beta s} \langle \delta Y_s, f^1(s, \Gamma^1(s)) - f^2(s, \Gamma^2(s)) \rangle ds - \int_t^T 2e^{\beta s} \langle \delta Y_s, \delta Z_s dW_s \rangle \\ &\leq e^{\beta T} |\delta Y_T|^2 + \int_t^T \gamma e^{\beta s} |\delta Y_s|^2 ds + \int_t^T \frac{e^{\beta s}}{\gamma} \left(|f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))|^2 \right) ds \\ &\quad + 2 \int_t^T e^{\beta s} \langle \delta Y_s, \delta_2 f_s \rangle ds - \int_t^T 2e^{\beta s} \langle \delta Y_s, \delta Z_s dW_s \rangle. \end{aligned}$$

where the last inequality results from Young's inequality for some $\gamma > 0$. Reorganizing and taking condition (H2') for the generator f^1 into account, we get

$$\begin{aligned} & e^{\beta t} |\delta Y_t|^2 + \int_t^T (\beta - \gamma) e^{\beta s} |\delta Y_s|^2 ds + \int_t^T e^{\beta s} |\delta Z_s|^2 ds \\ &\leq e^{\beta T} |\delta Y_T|^2 + \int_t^T \frac{e^{\beta s}}{\gamma} L \left[(|\delta Y|^2 \cdot \alpha_Y)(s) + (|\delta Z|^2 \cdot \alpha_Z)(s) \right] ds \\ &\quad + 2 \int_t^T e^{\beta s} \langle \delta Y_s, \delta_2 f_s \rangle ds - \int_t^T 2e^{\beta s} \langle \delta Y_s, \delta Z_s dW_s \rangle. \end{aligned}$$

By a change of integration order argument similar to that in the proof of Lemma 1.1 we obtain for $j \in \{Y, Z\}$ and $\phi^Y = \delta Y$, $\phi^Z = \delta Z$

$$\begin{aligned} & \int_t^T e^{\beta s} (|\phi^j|^2 \cdot \alpha_j)(s) ds \\ &= \int_t^T \int_{-T}^0 e^{\beta(s+v)} e^{-\beta v} \mathbb{1}_{\{s+v \geq 0\}} |\phi_{s+v}^j|^2 \alpha_j(dv) ds \\ &= \int_{-T}^0 \int_{(t+v) \vee 0}^{T+v} e^{\beta r} e^{-\beta v} \mathbb{1}_{\{r \geq 0\}} |\phi_r^j|^2 dr \alpha_j(dv) = \int_0^T \int_{r-T}^{(r-t) \wedge 0} e^{\beta r} e^{-\beta v} |\phi_r^j|^2 \alpha_j(dv) dr \\ &= \int_0^T e^{\beta r} |\phi_r^j|^2 \left(\int_{-T}^0 e^{-\beta v} \alpha_j(dv) \right) dr \leq \int_0^T \tilde{\alpha} e^{\beta r} |\phi_r^j|^2 dr, \end{aligned} \quad (9)$$

with $\tilde{\alpha}$ given by (6). Continuing the inequality from above we get

$$\begin{aligned} e^{\beta t}|\delta Y_t|^2 + \int_t^T (\beta - \gamma)e^{\beta s}|\delta Y_s|^2 ds + \int_t^T e^{\beta s}|\delta Z_s|^2 ds &\leq e^{\beta T}|\delta Y_T|^2 + 2 \int_t^T e^{\beta s}\langle \delta Y_s, \delta_2 f_s \rangle ds \\ &+ \int_0^T \frac{\tilde{\alpha}L}{\gamma} e^{\beta s} (|\delta Y_s|^2 + |\delta Z_s|^2) ds - \int_t^T 2e^{\beta s}\langle \delta Y_s, \delta Z_s dW_s \rangle. \end{aligned} \quad (10)$$

Putting $t = 0$ and taking expectations yields

$$\begin{aligned} &(\beta - \gamma - \frac{\tilde{\alpha}L}{\gamma})\mathbb{E}\left[\int_0^T e^{\beta s}|\delta Y_s|^2 ds\right] + (1 - \frac{\tilde{\alpha}L}{\gamma})\mathbb{E}\left[\int_0^T e^{\beta s}|\delta Z_s|^2 ds\right] \\ &\leq \mathbb{E}\left[e^{\beta T}|\delta Y_T|^2\right] + 2\mathbb{E}\left[\int_0^T e^{\beta s}\langle \delta Y_s, \delta_2 f_s \rangle ds\right] \\ &\leq \mathbb{E}\left[e^{\beta T}|\delta Y_T|^2\right] + 2\mathbb{E}\left[\sup_{0 \leq t \leq T} e^{\frac{\beta}{2}t}|\delta Y_t| \int_0^T e^{\frac{\beta}{2}s}|\delta_2 f_s| ds\right] \\ &\leq \mathbb{E}\left[e^{\beta T}|\delta Y_T|^2\right] + \gamma'\mathbb{E}\left[\sup_{0 \leq t \leq T} e^{\beta t}|\delta Y_t|^2\right] + \frac{1}{\gamma'}\mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s}|\delta_2 f_s| ds\right)^2\right] \end{aligned}$$

where we have used Young's inequality with some $\gamma' > 0$ to be specified later. From the last expression we deduce

$$\|\delta Y\|_{\mathcal{H}_\beta^2}^2 + \|\delta Z\|_{\mathcal{H}_\beta^2}^2 \leq C\left\{\mathbb{E}\left[e^{\beta T}|\delta Y_T|^2\right] + \gamma'\|\delta Y\|_{\mathcal{S}_\beta^2}^2 + \frac{1}{\gamma'}\mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s}|\delta_2 f_s| ds\right)^2\right]\right\}, \quad (11)$$

where $C > 0$ is a constant depending $\beta, \gamma, \tilde{\alpha}, L$. In order to obtain the \mathcal{S}_β^2 -estimate for δY , we observe that we have

$$\delta Y_t \leq \delta Y_T + \int_t^T |f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))| ds + \int_t^T |\delta_2 f_s| ds - \int_t^T \delta Z_s dW_s.$$

Multiplying by the monotone increasing function $e^{\frac{\beta}{2}t}$ and taking the conditional expectation with respect to \mathcal{F}_t , we get

$$\begin{aligned} e^{\frac{\beta}{2}t}\delta Y_t &\leq \mathbb{E}\left[e^{\frac{\beta}{2}t}|\delta Y_T| + e^{\frac{\beta}{2}t} \int_t^T |f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))| ds + e^{\frac{\beta}{2}t} \int_t^T |\delta_2 f_s| ds \middle| \mathcal{F}_t\right] \\ &\leq \mathbb{E}\left[e^{\frac{\beta}{2}T}|\delta Y_T| + \int_t^T e^{\frac{\beta}{2}s}|f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))| ds \right. \\ &\quad \left. + \int_0^t e^{\frac{\beta}{2}s}|f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))| ds + \int_t^T e^{\frac{\beta}{2}s}|\delta_2 f_s| ds + \int_0^t e^{\frac{\beta}{2}s}|\delta_2 f_s| ds \middle| \mathcal{F}_t\right] \\ &= \mathbb{E}\left[e^{\frac{\beta}{2}T}|\delta Y_T| + \int_0^T e^{\frac{\beta}{2}s}|f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))| ds + \int_0^T e^{\frac{\beta}{2}s}|\delta_2 f_s| ds \middle| \mathcal{F}_t\right]. \end{aligned}$$

Using Doob's inequality, we obtain

$$\begin{aligned} &\|\delta Y\|_{\mathcal{S}_\beta^2}^2 \\ &\leq 4 \mathbb{E}\left[\left(\mathbb{E}\left[e^{\frac{\beta}{2}T}|\delta Y_T| + \int_0^T e^{\frac{\beta}{2}s}|f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))| ds + \int_0^T e^{\frac{\beta}{2}s}|\delta_2 f_s| ds \middle| \mathcal{F}_T\right]\right)^2\right] \\ &\leq 12 \mathbb{E}\left[e^{\beta T}|\delta Y_T|^2 + T \int_0^T e^{\beta s}|f^1(s, \Gamma^1(s)) - f^1(s, \Gamma^2(s))|^2 ds + \left(\int_0^T e^{\frac{\beta}{2}s}|\delta_2 f_s| ds\right)^2\right], \end{aligned}$$

where the last line follows by Jensen's inequality. Since f^1 satisfies (H2'), an application of Lemma 1.1 yields

$$\|\delta Y\|_{\mathcal{S}_\beta^2}^2 \leq 12 \left\{ \mathbb{E} \left[e^{\beta T} |\delta Y_T|^2 \right] + \tilde{\alpha} T L \left(\|\delta Y\|_{\mathcal{H}_\beta^2}^2 + \|\delta Z\|_{\mathcal{H}_\beta^2}^2 \right) + \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds \right)^2 \right] \right\}.$$

Hence, plugging into (11) we find

$$\begin{aligned} & (1 - 12C\gamma'\tilde{\alpha}TL) \mathbb{E} \left[\sup_{0 \leq t \leq T} e^{\beta t} |\delta Y_t|^2 \right] \\ & \leq 12 \left\{ (1 + C\tilde{\alpha}TL) \mathbb{E} \left[e^{\beta T} |\delta Y_T|^2 \right] + (1 + C\gamma'^{-1}\tilde{\alpha}TL) \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds \right)^2 \right] \right\}. \end{aligned}$$

Choosing γ' small enough so that $(1 - 12C\gamma'\tilde{\alpha}TL) > 0$ is satisfied, we conclude that estimate (8) holds for a constant $C_2 = C_2(\beta, \gamma, \tilde{\alpha}, L)$. \square

The proof for the case $p > 2$ is more involved and utilizes techniques from the proof of Proposition 2.5. The main reason for the proof to be more involved can be seen in (10). Usually the dynamics of Y_t is described by integrals over the interval $[t, T]$ but for delay BSDEs we see from (10) that the dynamics of Y_t depends also on an integral over the whole interval $[0, T]$.

Proposition 2.6. *Let $p > 2$ and $\beta, \gamma > 0$. Consider $i \in \{1, 2\}$ and denote by $(Y^i, Z^i) \in \mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ the solution of the delay BSDE (4) with terminal condition ξ^i and generator f^i satisfying (H1)-(H4). Denote by $L > 0$ the Lipschitz constant of f^1 in (H2') and set $\delta Y = Y^1 - Y^2$, $\delta Z = Z^1 - Z^2$ and $\delta_2 f_t = f^1(t, Y^2(t), Z^2(t)) - f^2(t, Y^2(t), Z^2(t))$. Assume that $\beta, \gamma > 0$ satisfy (7). Then there exist constants $\gamma_2, \gamma_3 > 0$ such that for L and T small enough (i.e. chosen such that the constants D_i , $i \in \{1, \dots, 5\}$, specified in the proof are positive) there exists a constant $C_p = C_p(\beta, \gamma, \gamma_2, \gamma_3, \tilde{\alpha}, L, T) > 0$ satisfying the a priori estimate*

$$\|\delta Y\|_{\mathcal{S}_\beta^p}^p + \|\delta Y\|_{\mathcal{H}_\beta^p}^p + \|\delta Z\|_{\mathcal{H}_\beta^p}^p \leq C_p \left\{ \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] + \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds \right)^p \right] \right\}. \quad (12)$$

Proof. Throughout let $t \in [0, T]$, $i \in \{1, 2\}$ and set $D_1 := \beta - \gamma - \frac{\tilde{\alpha}L}{\gamma}$ and $D_2 := 1 - \frac{\tilde{\alpha}L}{\gamma}$. Recall (10) from Proposition 2.5:

$$\begin{aligned} & e^{\beta t} |\delta Y_t|^2 + \int_t^T (\beta - \gamma) e^{\beta s} |\delta Y_s|^2 ds + \int_t^T e^{\beta s} |\delta Z_s|^2 ds \leq e^{\beta T} |\delta Y_T|^2 + 2 \int_t^T e^{\beta s} \langle \delta Y_s, \delta_2 f_s \rangle ds \\ & + \int_0^T \frac{\tilde{\alpha}L}{\gamma} e^{\beta s} (|\delta Y_s|^2 + |\delta Z_s|^2) ds - \int_t^T 2e^{\beta s} \langle \delta Y_s, \delta Z_s \rangle dW_s. \end{aligned} \quad (13)$$

We say that a constant $C > 0$ depends on the data if $C = C(\beta, \gamma, \gamma_2, \tilde{\alpha}, L, T)$. We carry out the proof in several steps.

Step 1: We claim that

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\delta Z_s|^2 ds \right)^{p/2} \right] & \leq D_3^{-1} \left\{ 2^{p/2-1} \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] + 2^{3p/2-2} d_{p/2} \gamma_2 \|\delta Y\|_{\mathcal{S}_\beta^p}^p \right. \\ & \left. + 2^{3p/2-2} \mathbb{E} \left[\left| \int_0^T e^{\beta s} \langle \delta Y_s, \delta_2 f_s \rangle ds \right|^{p/2} \right] \right\}, \end{aligned} \quad (14)$$

with $D_3 := ((1 - \frac{\tilde{\alpha}L}{\gamma})^{p/2} - \frac{2^{3p/2-2}}{\gamma_2} d_{p/2})$, where $d_{p/2} > 0$ is a given constant appearing in the Burkholder-Davis-Gundy inequality which only depends on $p > 2$. Estimate (14) can be deduced

as follows: putting $t = 0$ in (13) and noticing that by (7) the constants D_1 and D_2 are positive we get

$$\begin{aligned} \left(1 - \frac{\tilde{\alpha}L}{\gamma}\right) \int_0^T e^{\beta s} |\delta Z_s|^2 ds &\leq (\beta - \gamma - \frac{\tilde{\alpha}L}{\gamma}) \int_0^T e^{\beta s} |\delta Y_s|^2 ds + \left(1 - \frac{\tilde{\alpha}L}{\gamma}\right) \int_0^T e^{\beta s} |\delta Z_s|^2 ds \\ &\leq e^{\beta T} |\delta Y_T|^2 + 2 \int_0^T e^{\beta s} \langle \delta Y_s, \delta_2 f_s \rangle ds - 2 \int_0^T e^{\beta s} \langle \delta Y_s, \delta Z_s \rangle dW_s. \end{aligned}$$

Now raising both sides to the power $p/2 > 1$, making use of the fact that for $a, b, c \in \mathbb{R}$

$$\begin{aligned} |a + 2b - 2c|^{p/2} &\leq 2^{p/2-1} \left(|a|^{p/2} + |2b - 2c|^{p/2} \right) \\ &\leq 2^{p/2-1} \left(|a|^{p/2} + 2^{p/2-1} (|2b|^{p/2} + |2c|^{p/2}) \right) \\ &= 2^{p/2-1} |a|^{p/2} + 2^{3p/2-2} |b|^{p/2} + 2^{3p/2-2} |c|^{p/2} \end{aligned}$$

and taking expectations, we get

$$\begin{aligned} \left(1 - \frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\delta Z_s|^2 ds \right)^{p/2} \right] &\leq 2^{p/2-1} \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] \\ &\quad + 2^{3p/2-2} \mathbb{E} \left[\left| \int_0^T e^{\beta s} \langle \delta Y_s, \delta_2 f_s \rangle ds \right|^{p/2} \right] + 2^{3p/2-2} \mathbb{E} \left[\left| \int_0^T e^{\beta s} \langle \delta Y_s, \delta Z_s \rangle dW_s \right|^{p/2} \right]. \end{aligned} \quad (15)$$

An application of the BDG inequality yields that for a given constant $d_{p/2} > 0$, we have

$$\begin{aligned} \mathbb{E} \left[\left| \int_0^T e^{\beta s} \langle \delta Y_s, \delta Z_s \rangle dW_s \right|^{p/2} \right] &\leq d_{p/2} \mathbb{E} \left[\left(\int_0^T e^{2\beta s} |\delta Y_s|^2 |\delta Z_s|^2 ds \right)^{p/4} \right] \\ &\leq d_{p/2} \mathbb{E} \left[\left(\sup_{0 \leq t \leq T} e^{\beta t} |\delta Y_t|^2 \right)^{p/4} \left(\int_0^T e^{\beta s} |\delta Z_s|^2 ds \right)^{p/4} \right] \\ &\leq d_{p/2} \left\{ \gamma_2 \|\delta Y\|_{\mathcal{S}_\beta^p}^p + \frac{1}{\gamma_2} \|\delta Z\|_{\mathcal{H}_\beta^p}^p \right\}, \end{aligned} \quad (16)$$

where the last line follows from Young's inequality with $\gamma_2 > 0$. Now we choose $\gamma_2 > 0$ such that $D_3 > 0$. Note that with this γ_2 , if one replaces L by L' with $L' < L$ then the quantity $(1 - \frac{2\tilde{\alpha}L'}{\gamma})^{p/2} - \frac{2^{3p/2-2}}{\gamma_2} d_{p/2} > D_3$ is still positive. Plugging (16) into (15), we get

$$\begin{aligned} D_3 \|\delta Z\|_{\mathcal{H}_\beta^p}^p &\leq 2^{p/2-1} \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] + 2^{3p/2-2} \mathbb{E} \left[\left| \int_0^T e^{\beta s} \langle \delta Y_s, \delta_2 f_s \rangle ds \right|^{p/2} \right] \\ &\quad + 2^{3p/2-2} d_{p/2} \gamma_2 \|\delta Y\|_{\mathcal{S}_\beta^p}^p, \end{aligned}$$

which proves the claim.

Step 2: We claim that

$$\begin{aligned} D_4 \|\delta Y\|_{\mathcal{S}_\beta^p}^p &\leq \left(\frac{p}{p-2} \right)^{p/2} \left\{ (2^{p-2} + 2^{3p/2-3} \left(\frac{\tilde{\alpha}L}{\gamma} \right)^{p/2} D_3^{-1}) \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] \right. \\ &\quad \left. + (2^{3p/2-2} + 2^{5p/2-4} \left(\frac{\tilde{\alpha}L}{\gamma} \right)^{p/2} D_3^{-1}) \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds \right)^{p/2} \right] \right\}, \end{aligned} \quad (17)$$

holds for

$$D_4 := 1 - 2^{5/4p-4} \gamma_2 d_{p/2} \left(\frac{p}{p-2}\right)^{p/2} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1} - \left(\frac{\tilde{\alpha}L}{\gamma} T\right)^{p/2} \left(\frac{p}{p-2}\right)^{p/2} 2^{p-2}.$$

Before showing this estimate we stress that we can choose L and T such that $D_4 > 0$. More precisely, the constants β, γ, γ_2 are already fixed and depend on L . But if one replaces L by L' with $L' < L$ then it is clear that $D'_1, D'_2, D'_3 > 0$ (where D'_i denotes D_i with L replaced by L') for the same β, γ, γ_2 since $D'_i > D_i$, $i = 1, 2, 3$. Thus we can choose L and T small enough making $D_1, \dots, D_4 > 0$.

Now we prove (17). For this we go back to (13), take conditional expectation with respect to \mathcal{F}_t and the sup in $t \in [0, T]$, raise to the power $p/2$, apply Doob's inequality and find

$$\begin{aligned} & \mathbb{E} \left[\sup_{0 \leq t \leq T} (e^{\beta t} |\delta Y_t|^2)^{p/2} \right] \\ & \leq \mathbb{E} \left[\sup_{0 \leq t \leq T} \left(\mathbb{E} \left[e^{\beta T} |\delta Y_T|^2 + 2 \int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds \right. \right. \right. \\ & \quad \left. \left. \left. + \int_0^T \frac{\tilde{\alpha}L}{\gamma} e^{\beta s} (|\delta Y_s|^2 + |\delta Z_s|^2) ds \middle| \mathcal{F}_t \right] \right)^{p/2} \right] \\ & \leq \left(\frac{p}{p-2}\right)^{p/2} \mathbb{E} \left[\left(e^{\beta T} |\delta Y_T|^2 + 2 \int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds + \int_0^T \frac{\tilde{\alpha}L}{\gamma} e^{\beta s} (|\delta Y_s|^2 + |\delta Z_s|^2) ds \right)^{p/2} \right] \\ & \leq \left(\frac{p}{p-2}\right)^{p/2} \left\{ 2^{p-2} \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] + 2^{3p/2-2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds \right)^{p/2} \right] \right. \\ & \quad \left. + 2^{p-2} \mathbb{E} \left[\left(\int_0^T \frac{\tilde{\alpha}L}{\gamma} e^{\beta s} |\delta Y_s|^2 ds \right)^{p/2} \right] + 2^{p-2} \mathbb{E} \left[\left(\int_0^T \frac{\tilde{\alpha}L}{\gamma} e^{\beta s} |\delta Z_s|^2 ds \right)^{p/2} \right] \right\}. \quad (18) \end{aligned}$$

Note that we made use of the fact that for $a, b, c, d \in \mathbb{R}$ and $p > 2$, we have

$$\begin{aligned} |a + 2b + c + d|^{p/2} & \leq 2^{p/2-1} (|a + 2b|^{p/2} + |c + d|^{p/2}) \\ & \leq 2^{p-2} |a|^{p/2} + 2^{3p/2-2} |b|^{p/2} + 2^{p-2} |c|^{p/2} + 2^{p-2} |d|^{p/2}. \end{aligned}$$

Plugging (14) into (18), we get

$$\begin{aligned} \|\delta Y\|_{\mathcal{S}_\beta^p}^p & \leq \left(\frac{p}{p-2}\right)^{p/2} \left\{ 2^{p-2} \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] + 2^{3p/2-2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds \right)^{p/2} \right] \right. \\ & \quad + 2^{p-2} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} \|\delta Y\|_{\mathcal{H}_\beta^p}^p + \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1} \times 2^{p-2} \left\{ 2^{p/2-1} \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] \right. \\ & \quad \left. \left. + 2^{3p/2-2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds \right)^{p/2} \right] + 2^{3p/2-2} d_{p/2} \gamma_2 \|\delta Y\|_{\mathcal{S}_\beta^p}^p \right\} \right\} \\ & \leq \left(\frac{p}{p-2}\right)^{p/2} \left\{ (2^{p-2} + 2^{3p/2-3} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1}) \mathbb{E} \left[(e^{\beta T} |\delta Y_T|^2)^{p/2} \right] \right. \\ & \quad + (2^{3/2p-2} + 2^{5p/2-4} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1}) \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds \right)^{p/2} \right] \\ & \quad \left. + \left(2^{p-2} \left(\frac{\tilde{\alpha}L}{\gamma} T\right)^{p/2} + 2^{5p/2-4} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1} d_{p/2} \gamma_2 \right) \|\delta Y\|_{\mathcal{S}_\beta^p}^p \right\}, \end{aligned}$$

from which (17) readily follows.

Step 3: At this stage, estimating $\mathbb{E}\left[\left(\int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds\right)^{p/2}\right]$ will yield (12) which follows as a consequence from equation(17). Applying Young's inequality with some $\gamma_3 > 0$ we get

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T e^{\beta s} |\langle \delta Y_s, \delta_2 f_s \rangle| ds\right)^{p/2}\right] &\leq \mathbb{E}\left[\left(\int_0^T e^{\beta s} |\delta Y_s| |\delta_2 f_s| ds\right)^{p/2}\right] \\ &\leq \mathbb{E}\left[\sup_{0 \leq t \leq T} (e^{\frac{\beta}{2}t} |\delta Y_t|)^{p/2} \left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds\right)^{p/2}\right] \\ &\leq \gamma_3 \|\delta Y\|_{\mathcal{S}_\beta^p}^p + \frac{1}{\gamma_3} \mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds\right)^p\right]. \end{aligned} \quad (19)$$

Together with (17), inequality (19) leads to

$$\begin{aligned} D_5 \|\delta Y\|_{\mathcal{S}_\beta^p}^p &\leq \left(\frac{p}{p-2}\right)^{p/2} \left\{ (2^{p-2} + 2^{3p/2-3} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1}) \mathbb{E}[(e^{\beta T} |\delta Y_T|^2)^{p/2}] \right. \\ &\quad \left. + (2^{3/2p-2} + 2^{5p/2-4} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1}) \gamma_3^{-1} \mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds\right)^p\right] \right\}, \end{aligned} \quad (20)$$

where $D_5 := D_4 - \left(\frac{p}{p-2}\right)^{p/2} \gamma_3 \left(2^{3/2p-2} + 2^{5p/2-4} \left(\frac{\tilde{\alpha}L}{\gamma}\right)^{p/2} D_3^{-1}\right)$ and $\gamma_3 > 0$ is chosen such that $D_5 > 0$. Finally putting together the previous steps, the estimate (12) follows. Note that (20) implies

$$\|\delta Y\|_{\mathcal{S}_\beta^p}^p \leq C_p \left\{ \mathbb{E}[(e^{\beta T} |\delta Y_T|^2)^{p/2}] + \mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds\right)^p\right] \right\}, \quad (21)$$

where $C_p > 0$ is a constant depending on the data. This leads to

$$\|\delta Y\|_{\mathcal{H}_\beta^p}^p \leq T^{p/2} \|\delta Y\|_{\mathcal{S}_\beta^p}^p \leq C_p \left\{ \mathbb{E}[(e^{\beta T} |\delta Y_T|^2)^{p/2}] + \mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds\right)^p\right] \right\}.$$

Finally, observing that (14) from step 1 yields

$$\|\delta Z\|_{\mathcal{H}_\beta^p}^p \leq C_p \left\{ \mathbb{E}[(e^{\beta T} |\delta Y_T|^2)^{p/2}] + \mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 f_s| ds\right)^p\right] + \|\delta Y\|_{\mathcal{S}_\beta^p}^p \right\},$$

and applying (21), it follows that (12) is valid. This finishes the proof. \square

As a by-product of the two previous theorems we obtain under their respective assumptions a result on the moment estimates for the solution of BSDE (4).

Corollary 2.7 (Moment estimates). *Under the assumptions of Proposition 8 and Proposition 2.6, we have for $p \geq 2$*

$$\|Y\|_{\mathcal{S}_\beta^p}^p + \|Y\|_{\mathcal{H}_\beta^p}^p + \|Z\|_{\mathcal{H}_\beta^p}^p \leq C_p \left\{ \mathbb{E}[(e^{\beta T} |\delta Y_T|^2)^{p/2}] + \mathbb{E}\left[\left(\int_0^T e^{\beta s} |f(s, 0, 0)|^2 ds\right)^p\right] \right\}.$$

The moment and a priori estimates in Delong and Imkeller (2010a) are tailor-made for a Picard iteration procedure in $\mathcal{H}^2 \times \mathcal{H}^2$. The right-hand side of their estimates depends on the solution process but such an estimate suffices in their context. The a priori estimates from Proposition 2.5 and Proposition 2.6 are somewhat “sharper” in the sense that they are the usual a priori

estimates one expects to obtain when dealing with BSDE, i.e. they exhibit a right-hand side which solely depends on the data δY_T and $\delta_2 f_t$.

With estimate (12) at hand, we now proceed to show the existence and uniqueness of solutions to (4) in $\mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ for $p > 2$. For $p = 2$, Theorem 2.1 from Delong and Imkeller (2010a) (recalled in our Theorem 2.1) yields a sufficient condition which guarantees the standard Picard iteration to converge and proves the existence and uniqueness of solutions to (4). We will show in the following result that for $p > 2$, the convergence of the same Picard iteration is retained. What is needed to achieve this goal is to put up some extra effort for showing that the Picard iterates (Y^n, Z^n) satisfy the corresponding $\mathcal{S}_\beta^p, \mathcal{H}_\beta^p$ -integrability properties.

Theorem 2.8. *Let $p > 2$ and assume that (H1)-(H4) hold. Assume that $\beta, \gamma, L, T, \gamma_2, \gamma_3$ are chosen like in Proposition 2.6 such that condition (7) holds and let C_p denote the constant appearing in the a priori estimate (12). If*

$$2^{p/2-1} C_p (L\tilde{\alpha})^{p/2} \max\{1, T\}^p < 1, \quad (22)$$

where $\tilde{\alpha}$ is given by (6), then the BSDE (4) admits a unique solution in $\mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$.

Proof. Throughout let $t \in [0, T]$ and $p > 2$. The proof is based on the standard Picard iteration: we initialize by $Y^0 = 0$ and $Z^0 = 0$ and define recursively

$$Y_t^{n+1} = \xi + \int_t^T f(s, \Gamma^n(s)) ds - \int_t^T Z_s^{n+1} dW_s, \quad 0 \leq t \leq T, \quad (23)$$

with $\Gamma^n(s) = (\int_{-T}^0 Y_{s+v}^n \alpha_Y(dv), \int_{-T}^0 Z_{s+v}^n \alpha_Z(dv))$ for $s \in [0, T]$ and $n \in \mathbb{N}$. In the following, $C > 0$ will denote some generic constant which may vary from line to line but always independent of $n \in \mathbb{N}$. We proceed by induction. For $n \geq 1$, assume that $(Y^n, Z^n) \in \mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ is already shown, and we prove that (23) has a unique solution $(Y^{n+1}, Z^{n+1}) \in \mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$. Note that because of

$$\begin{aligned} & \mathbb{E} \left[\left(\int_0^T |f(s, \Gamma^n(s))| ds \right)^p \right] \\ & \leq \mathbb{E} \left[\left(\int_0^T |f(s, 0, 0)| ds + \int_0^T |f(s, \Gamma^n(s)) - f(s, 0, 0)| ds \right)^p \right] \\ & \leq 2^{p-1} \mathbb{E} \left[\left(\int_0^T |f(s, 0, 0)| ds \right)^p + \left(T \int_0^T |f(s, \Gamma^n(s)) - f(s, 0, 0)|^2 ds \right)^{p/2} \right] \\ & \leq 2^{p-1} \mathbb{E} \left[\left(\int_0^T |f(s, 0, 0)| ds \right)^p \right. \\ & \quad \left. + L^{p/2} T^{p/2} \left\{ \int_0^T \int_{-T}^0 |Y_{s+v}^n|^2 \alpha_Y(dv) ds + \int_0^T \int_{-T}^0 |Z_{s+v}^n|^2 \alpha_Z(dv) ds \right\}^{p/2} \right] \\ & \leq 2^{p-1} \mathbb{E} \left[\left(\int_0^T |f(s, 0, 0)| ds \right)^p + (\alpha K T)^{p/2} \left\{ \int_0^T |Y_s^n|^2 ds + \int_0^T |Z_s^n|^2 ds \right\}^{p/2} \right] \\ & \leq 2^{p-1} \mathbb{E} \left[\left(\int_0^T |f(s, 0, 0)| ds \right)^p \right] + 2^{p/2-1} (2\alpha K T)^{p/2} \left(T^{p/2} \|Y^n\|_{\mathcal{S}_0^p}^p + \|Z^n\|_{\mathcal{H}_0^p}^p \right) \\ & < \infty, \end{aligned} \quad (24)$$

the martingale representation yields a uniquely determined process $Z^{n+1} \in \mathcal{H}_0^2$ such that

$$\mathbb{E} \left[\xi + \int_0^T f(s, \Gamma^n(s)) ds \middle| \mathcal{F}_t \right] = \mathbb{E} \left[\xi + \int_0^T f(s, \Gamma^n(s)) ds \right] + \int_0^T Z_s^{n+1} dW_s.$$

It is then standard to choose Y^{n+1} to be a continuous version of

$$Y_t^{n+1} = \mathbb{E}\left[\xi + \int_t^T f(s, \Gamma^n(s))ds \middle| \mathcal{F}_t\right].$$

Let us first show that $Y^{n+1} \in \mathcal{S}_\beta^p$:

$$\begin{aligned} \|Y^{n+1}\|_{\mathcal{S}_\beta^p}^p &= \mathbb{E}\left[\sup_{t \in [0, T]} |Y_t^{n+1}|^p\right] \leq \mathbb{E}\left[\sup_{t \in [0, T]} \left(\mathbb{E}[|\xi| + \int_0^T |f(s, \Gamma^n(s))|ds \middle| \mathcal{F}_t]\right)^p\right] \\ &\leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[\left(|\xi| + \int_0^T |f(s, \Gamma^n(s))|ds\right)^p\right] \\ &\leq 2^{p-1} \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|\xi|^p + \left(\int_0^T |f(s, \Gamma^n(s))|ds\right)^p\right] < \infty, \end{aligned}$$

where the last inequality follows from $\xi \in L^p$ and (24) which proves that $Y^{n+1} \in \mathcal{S}_0^p$. Since all $\|\cdot\|_{\mathcal{S}_\beta^p}$ -norms are equivalent it follows that $Y^{n+1} \in \mathcal{S}_\beta^p$. To see that $Z^{n+1} \in \mathcal{H}_\beta^p$, recall that Itô's formula applied to $e^{\beta t}|Y_t^{n+1}|^2$ yields

$$\begin{aligned} e^{\beta t}|Y_t^{n+1}|^2 + \int_t^T \beta e^{\beta s}|Y_s^{n+1}|^2 ds + \int_t^T e^{\beta s}|Z_s^{n+1}|^2 ds \\ = e^{\beta T}|\xi|^2 + \int_t^T 2e^{\beta s}\langle Y_s^{n+1}, f(s, \Gamma^n(s)) \rangle ds - \int_t^T 2e^{\beta s}\langle Y_s^{n+1}, Z_s^{n+1}dW_s \rangle \end{aligned}$$

which implies

$$\begin{aligned} &\left(\int_0^T e^{\beta s}|Z_s^{n+1}|^2 ds\right)^{p/2} \\ &\leq \left(e^{\beta T}|\xi|^2 + \int_0^T 2e^{\beta s}|Y_s^{n+1}| |f(s, \Gamma^n(s))| ds + \sup_{t \in [0, T]} \left|\int_0^t 2e^{\beta s}\langle Y_s^{n+1}, Z_s^{n+1}dW_s \rangle\right|\right)^{p/2} \\ &\leq 2^{p/2-1} (e^{\beta T}|\xi|^2)^{p/2} + 2^{p-2} \left(\int_0^T 2e^{\beta s}|Y_s^{n+1}| |f(s, \Gamma^n(s))| ds\right)^{p/2} \\ &\quad + 2^{3p/2-2} \left(\sup_{t \in [0, T]} \left|\int_0^t e^{\beta s}\langle Y_s^{n+1}, Z_s^{n+1}dW_s \rangle\right|\right)^{p/2}. \end{aligned} \tag{25}$$

On the one hand, we have

$$\begin{aligned} &\mathbb{E}\left[\left(\int_0^T 2e^{\beta s}|Y_s^{n+1}| |f(s, \Gamma^n(s))| ds\right)^{p/2}\right] \\ &\leq \mathbb{E}\left[\left(\int_0^T 2e^{\beta s}|Y_s^{n+1}| |f(s, \Gamma^n(s)) - f(s, 0, 0)| ds + \int_0^T 2e^{\beta s}|Y_s^{n+1}| |f(s, 0, 0)| ds\right)^{p/2}\right] \\ &\leq \|Y^{n+1}\|_{\mathcal{S}_\beta^2}^2 + \mathbb{E}\left[\left(\int_0^T e^{\beta s}|f(s, \Gamma^n(s)) - f(s, 0, 0)|^2 ds + \int_0^T 2e^{\beta s}|Y_s^{n+1}| |f(s, 0, 0)| ds\right)^{p/2}\right] \\ &\leq C \left\{ \|Y^{n+1}\|_{\mathcal{S}_\beta^p}^p + \mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s}|f(s, 0, 0)| ds\right)^p\right] + \|Y^n\|_{\mathcal{S}_\beta^p}^p + \|Z^n\|_{\mathcal{H}_\beta^p}^p \right\} \\ &< \infty, \end{aligned}$$

where we have used that

$$\int_0^T 2e^{\beta s}|Y_s^{n+1}| |f(s, 0, 0)| ds \leq \sup_{0 \leq t \leq T} e^{\beta t}|Y_t^{n+1}|^2 + \left(\int_0^T e^{\frac{\beta}{2}s}|f(s, 0, 0)| ds\right)^2.$$

On the other hand, we use the Burkholder-Davis-Gundy inequality to get

$$\begin{aligned}
\mathbb{E} \left[\sup_{t \in [0, T]} \left| \int_0^t e^{\beta s} \langle Y_s^{n+1}, Z_s^{n+1} dW_s \rangle \right|^{p/2} \right] &\leq d_{p/2} \mathbb{E} \left[\left(\int_0^T e^{2\beta s} |Y_s^{n+1}|^2 |Z_s^{n+1}|^2 ds \right)^{p/4} \right] \\
&\leq d_{p/2} \mathbb{E} \left[\left(\sup_{t \in [0, T]} e^{\beta t} |Y_t^{n+1}|^2 \int_0^T e^{\beta s} |Z_s^{n+1}|^2 ds \right)^{p/4} \right] \\
&\leq d_{p/2} \left\{ \kappa \|Y^{n+1}\|_{\mathcal{S}_\beta^p}^p + \frac{1}{\kappa} \|Z^{n+1}\|_{\mathcal{H}_\beta^p}^p \right\}, \tag{26}
\end{aligned}$$

where the last line follows from Young's inequality with some arbitrary constant $\kappa > 0$. Now choosing $\kappa > 0$ such that $1 - 2^{2p-2} d_{p/2} \kappa^{-1} > 0$, it follows from (25) and (26) that

$$\begin{aligned}
&\left(1 - \frac{2^{2p-2} d_{p/2}}{\kappa}\right) \|Z^{n+1}\|_{\mathcal{H}_\beta^p}^p \\
&\leq C \left\{ \mathbb{E} \left[(e^{\beta T} |\xi|^2)^{p/2} \right] + \|Y^{n+1}\|_{\mathcal{S}_\beta^p}^p + \mathbb{E} \left[\left(\int_0^T |f(s, 0, 0)| ds \right)^p \right] + \|Y^n\|_{\mathcal{S}_\beta^p}^p + \|Z^n\|_{\mathcal{H}_\beta^p}^p \right\} \\
&< \infty,
\end{aligned}$$

proving that $Z^{n+1} \in \mathcal{H}_\beta^p$.

In the next step, we prove that the sequence (Y^n, Z^n) converges in $\mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$. Using the a priori estimate (12), we get

$$\begin{aligned}
&\|Y^{n+1} - Y^n\|_{\mathcal{S}_\beta^p}^p + \|Z^{n+1} - Z^n\|_{\mathcal{H}_\beta^p}^p \\
&\leq C_p \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |f(s, \Gamma^n(s)) - f(s, \Gamma^{n-1}(s))| ds \right)^p \right] \\
&\leq C_p T^{p/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |f(s, \Gamma^n(s)) - f(s, \Gamma^{n-1}(s))|^2 ds \right)^{p/2} \right].
\end{aligned}$$

In analogy to the calculation carried out in paragraph (2.7) in Delong and Imkeller (2010a)[Proof of Theorem 2.1], it is easy to see that we have

$$\begin{aligned}
&\|Y^{n+1} - Y^n\|_{\mathcal{S}_\beta^p}^p + \|Z^{n+1} - Z^n\|_{\mathcal{H}_\beta^p}^p \\
&\leq C_p T^{p/2} \mathbb{E} \left[\left(L \max \left\{ \int_{-T}^0 e^{-\beta s} \alpha_Y(ds), \int_{-T}^0 e^{-\beta s} \alpha_Z(ds) \right\} \right. \right. \\
&\quad \times \left. \left(T \sup_{t \in [0, T]} e^{\beta t} |Y_t^n - Y_t^{n-1}|^2 + \int_0^T e^{\beta s} |Z_s^n - Z_s^{n-1}|^2 ds \right) \right)^{p/2} \right] \\
&\leq C_p T^{p/2} 2^{p/2-1} \left(L \max \left\{ \int_{-T}^0 e^{-\beta s} \alpha_Y(ds), \int_{-T}^0 e^{-\beta s} \alpha_Z(ds) \right\} \right)^{p/2} \\
&\quad \times \left(T^{p/2} \|Y^n - Y^{n-1}\|_{\mathcal{S}_\beta^p}^p + \|Z^n - Z^{n-1}\|_{\mathcal{H}_\beta^p}^p \right) \\
&\leq C_p 2^{p/2-1} \left(L \max \left\{ \int_{-T}^0 e^{-\beta s} \alpha_Y(ds), \int_{-T}^0 e^{-\beta s} \alpha_Z(ds) \right\} \right)^{p/2} \max \{1, T\}^p \\
&\quad \times \left(\|Y^n - Y^{n-1}\|_{\mathcal{S}_\beta^p}^p + \|Z^n - Z^{n-1}\|_{\mathcal{H}_\beta^p}^p \right).
\end{aligned}$$

Hence, by (22), the standard fixed point argument yields that (Y^n, Z^n) converges in $\mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$, which finishes the proof. \square

3 Decoupled FBSDE with time delayed generators

The objective of this section is to extend the results from Delong and Imkeller (2010a,b) to the case of decoupled forward-backward stochastic differential equations. For measurable functions b, σ, g, f , specified in more detail below, we study the time delayed FBSDE

$$X_t^x = x + \int_0^t b(s, X_s^x) ds + \int_0^t \sigma(s, X_s^x) dW_s, \quad x \in \mathbb{R}^d, \quad (27)$$

$$Y_t^x = g(X_T^x) + \int_t^T f(s, \Theta^x(s)) ds - \int_t^T Z_s^x dW_s, \quad 0 \leq t \leq T, \quad (28)$$

where for $t \in [0, T]$, we write

$$\begin{aligned} \Theta^x(t) &= ((X^x \cdot \alpha_x)(t), (Y^x \cdot \alpha_y)(t), (Z^x \cdot \alpha_z)(t)) \\ &= \left(\int_{-T}^0 X_{t+v}^x \alpha_x(dv), \int_{-T}^0 Y_{t+v}^x \alpha_y(dv), \int_{-T}^0 Z_{t+v}^x \alpha_z(dv) \right), \end{aligned} \quad (29)$$

with given deterministic finite measures α_x, α_y and α_z supported on $[-T, 0)$. The coefficients b, σ, g, f appearing in (27)-(28) are assumed to satisfy smoothness and integrability conditions such that the backward equation (28) falls back into the setting of (H1)-(H4) from section 2.1. More precisely, we assume the following to hold:

- (F1) $g : \mathbb{R}^d \rightarrow \mathbb{R}^m$ has bounded first order partial derivatives;
- (F2) $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{d \times m} \rightarrow \mathbb{R}^m$ is continuously differentiable and its first order partial derivatives $\nabla_x f, \nabla_y f, \nabla_z f$ are bounded;
- (F3) $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ are continuously differentiable functions with bounded derivatives; $b(\cdot, 0)$ and $\sigma(\cdot, 0)$ are bounded; σ is elliptic.
- (F4) $\mathbb{E} \left[\left(\int_0^T |f(s, 0, 0, 0)|^2 ds \right)^{p/2} \right] < \infty$ for $p \geq 2$
- (F5) $f(t, \cdot, \cdot, \cdot) \mathbb{1}_{(-\infty, 0)}(t) = 0$.

Condition (F3) is a standard assumption which guarantees the existence and uniqueness of solutions to the forward diffusion (27). Furthermore, condition (F2) implies that the generator is uniformly Lipschitz continuous in $(x, y, z) \in \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{d \times m}$. In analogy to conditions (H2) and (H2') from section 2.1, let us write down the following implication of the Lipschitz condition (F2): there exist constants $K, L > 0$ such that for any $t \in [0, T]$ and any sufficiently integrable vector or matrix valued processes u, u', y, y' and z, z' it holds that

$$\begin{aligned} (\text{F2}') \quad & \left| f(t, (u \cdot \alpha_x)(t), (y \cdot \alpha_y)(t), (z \cdot \alpha_z)(t)) - f(t, (u' \cdot \alpha_x)(t), (y' \cdot \alpha_y)(t), (z' \cdot \alpha_z)(t)) \right|^2 \\ & \leq K \left(|(u \cdot \alpha_x)(t) - (u' \cdot \alpha_x)(t)|^2 \right. \\ & \quad \left. + |(y \cdot \alpha_y)(t) - (y' \cdot \alpha_y)(t)|^2 + |(z \cdot \alpha_z)(t) - (z' \cdot \alpha_z)(t)|^2 \right) \\ & \leq L \left(((x - x')^2 \cdot \alpha_x)(t) + ((y - y')^2 \cdot \alpha_y)(t) + ((z - z')^2 \cdot \alpha_z)(t) \right). \end{aligned}$$

For a fixed $x \in \mathbb{R}^d$, the existence and uniqueness of the backward equation (28) in $\mathcal{S}_\beta^2 \times \mathcal{H}_\beta^2$ is guaranteed under the assumptions (F1)-(F5) together with the compatibility criterion from

Theorem 2.1 on the terminal time and the Lipschitz constant, i.e.

$$(8T + \frac{1}{\beta})L \int_{-T}^0 e^{-\beta s} \rho(ds) \max\{1, T\} < 1,$$

where $\rho \in \{\alpha_Y, \alpha_Z\}$. To lift the existence and uniqueness into $\mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ for $p > 2$, one only needs to replace the condition above by the compatibility condition from Theorem 2.8,

$$2^{p/2-1}C_p \left(L \int_{-T}^0 e^{-\beta s} \rho(ds) \right)^{p/2} \max\{1, T\}^p < 1,$$

where $\rho \in \{\alpha_Y, \alpha_Z\}$. Throughout this section, given $p \geq 2$, we will assume that for every $x \in \mathbb{R}^d$, the FBSDE (27)-(28) admits a unique solution $(X^x, Y^x, Z^x) \in \mathcal{S}_\beta^q \times \mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ for all $q \geq 2$. To avoid a notation overload for the rest of this work we assume $m = 1$.

3.1 Norm differentiability

In this section we investigate the variational differentiability of the solution (X^x, Y^x, Z^x) of the time delayed FBSDE (27)-(28) with respect to the Euclidean parameter $x \in \mathbb{R}^d$, i.e. with respect to the initial condition of the forward diffusion. By a well known result (see e.g. Protter (2005)), (F3) implies that the forward component X^x is differentiable with respect to the parameter $x \in \mathbb{R}^d$. It is natural to pose the question whether this smoothness is carried over to (Y^x, Z^x) in the setting of FBSDE with time delayed generators. Our goal is to show that the variational equations of (27)-(28) are given by

$$\nabla X_t^x = Id + \int_0^t \nabla b(s, X_s^x) \nabla X_s^x ds + \int_0^t \nabla \sigma(s, X_s^x) \nabla X_s^x dW_s, \quad (30)$$

$$\nabla Y_t^x = \nabla g(X_T^x) \nabla X_T^x - \int_t^T \nabla Z_s^x dW_s + \int_t^T \langle (\nabla f)(s, \Theta^x(s)), (\nabla \Theta^x)(s) \rangle ds, \quad (31)$$

where the notation $(\nabla \Theta^x)(t)$ is to be understood in the same fashion as in (29), i.e.

$$(\nabla \Theta^x)(t) = ((\nabla X^x \cdot \alpha_X)(t), (\nabla Y^x \cdot \alpha_Y)(t), (\nabla Z^x \cdot \alpha_Z)(t)), \quad t \in [0, T]. \quad (32)$$

Note that (F3) implies that (30) admits a unique solution in \mathcal{S}_β^p for every $p \geq 2$. Let (X, Y, Z) and ∇X solve (27)-(28) and (30) respectively and let Θ^x be as defined by (29). Now consider the BSDE with the linear time delayed generator

$$\begin{aligned} P_t &= \nabla g(X_T^x) \nabla X_T^x - \int_t^T Q_s dW_s \\ &\quad + \int_t^T \langle (\nabla f)(s, \Theta^x(s)), ((\nabla X^x \cdot \alpha_X)(s), (P \cdot \alpha_Y)(s), (Q \cdot \alpha_Z)(s)) \rangle ds. \end{aligned} \quad (33)$$

The existence and uniqueness of solutions to this BSDE follows from Theorem 2.1 for $p = 2$ and from Theorem 2.8 for $p > 2$. From now on we prove the main results of this section for $p > 2$ which are also valid when $p = 2$ under slightly different conditions on β , L and T which are made more precise in Remark 3.3.

Corollary 3.1. *Let $p > 2$ and let (F1)-(F5) be satisfied. Let $L > 0$ be as in (F2') and assume that $T, L, \beta > 0$ are chosen like in Proposition 2.6 and satisfy in addition*

$$2^{p/2-1}C_p \left(L \int_{-T}^0 e^{-\beta s} \rho(ds) \right)^{p/2} \max\{1, T\}^p < 1,$$

where $\rho \in \{\alpha_Y, \alpha_Z\}$. Then for every fixed x in \mathbb{R}^d , the BSDE (33) has a unique solution $(P, Q) \in \mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$.

Proof. Let $p \geq 2$. Given the hypothesis, $(X^x, Y^x, Z^x) \in \mathcal{S}_\beta^q \times \mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ is the solution of (27)-(28) for every $q \geq 2$. Since ∇g is bounded, the terminal condition $\nabla g(X_T^x) \nabla X_T^x$ is square integrable. Assumption (F2) implies that the linear generator of (33) has bounded coefficients, and hence is Lipschitz continuous. This Lipschitz constant is the same as the Lipschitz constant of f . Hence Theorem 2.8 can be applied which yields the result. \square

The uniqueness of solutions of equation (33) implies that the solutions to (31) and (33) coincide, i.e. $(\nabla Y^x, \nabla Z^x) = (P, Q)$ holds almost surely. For the rest of the section, we assume that all assumptions ensuring the existence and uniqueness of the variational equations (30)-(31) are fulfilled, i.e. we assume that the assumptions of Corollary 3.1 hold. In our next result we show the mapping $x \mapsto (Y^x, Z^x)$ is differentiable in an adequate sense.

Proposition 3.2. *Let β, γ satisfy the assumptions of Proposition 2.6. Let $p > 2$ and assume that the hypotheses (F1)-(F5) hold. Then, for any $x \in \mathbb{R}^d$, the solution (X^x, Y^x, Z^x) of the FBSDE (27)-(28) is norm-differentiable in the following sense:*

$$\lim_{\varepsilon \rightarrow 0} \left\| \frac{Y^{x+\varepsilon h} - Y^x}{\varepsilon} - \nabla Y^x \right\|_{\mathcal{S}_\beta^p}^p = \lim_{\varepsilon \rightarrow 0} \left\| \frac{Z^{x+\varepsilon h} - Z^x}{\varepsilon} - \nabla Z^x \right\|_{\mathcal{H}_\beta^p}^p = 0, \quad \forall h \in \mathbb{R}^d \setminus \{0\},$$

where $(\nabla Y^x, \nabla Z^x)$ is the unique solution of the BSDE

$$\nabla Y_t^x = \nabla g(X_T^x) \nabla X_T^x - \int_t^T \nabla Z_s^x dW_s + \int_t^T \langle (\nabla f)(s, \Theta^x(s)), (\nabla \Theta^x)(s) \rangle ds,$$

with Θ^x and $\nabla \Theta^x$ defined in (29) and (32).

Proof. Let $x \in \mathbb{R}^d$, $h \in \mathbb{R}^d \setminus \{0\}$, $t, s \in [0, T]$ and $\varepsilon > 0$. We use the following notations

$$\begin{aligned} A_{s,x} &:= \int_0^1 \nabla_x f\left(s, (X^x \cdot \alpha_x)(s) + \theta((X^{x+\varepsilon h} - X^x) \cdot \alpha_x)(s), \right. \\ &\quad \left. (Y^{x+\varepsilon h} \cdot \alpha_y)(s), (Z^{x+\varepsilon h} \cdot \alpha_z)(s)\right) d\theta, \\ A_{s,y} &:= \int_0^1 \nabla_y f\left(s, (X^x \cdot \alpha_x)(s), \right. \\ &\quad \left. (Y^x \cdot \alpha_y)(s) + \theta((Y^{x+\varepsilon h} - Y^x) \cdot \alpha_y)(s), (Z^{x+\varepsilon h} \cdot \alpha_z)(s)\right) d\theta, \\ A_{s,z} &:= \int_0^1 \nabla_z f\left(s, (X^x \cdot \alpha_x)(s), \right. \\ &\quad \left. (Y^x \cdot \alpha_y)(s), (Z^x \cdot \alpha_z)(s) + \theta((Z^{x+\varepsilon h} - Z^x) \cdot \alpha_z)(s)\right) d\theta. \end{aligned} \quad (34)$$

By assumption (F2), note that $|A_{s,*}| \leq L$ for $*$ = x, y, z and for every s in $[0, T]$. We denote by (P, Q) the solution to the BSDE (33) which coincides with $(\nabla Y, \nabla Z)$. We also write $U := \frac{Y^{x+\varepsilon h} - Y^x}{\varepsilon} - P$, $V := \frac{Z^{x+\varepsilon h} - Z^x}{\varepsilon} - Q$, $\tilde{X} := \frac{X^{x+\varepsilon h} - X^x}{\varepsilon} - \nabla X^x$ and $\xi := \frac{g(X_T^{x+\varepsilon h}) - g(X_T^x)}{\varepsilon} - \nabla g(X_T^x) \nabla X_T^x$, and we claim that

$$\lim_{\varepsilon \rightarrow 0} \|U\|_{\mathcal{S}_\beta^p}^p = \lim_{\varepsilon \rightarrow 0} \|V\|_{\mathcal{H}_\beta^p}^p = 0, \quad \text{for arbitrary } x \in \mathbb{R}^d, h \in \mathbb{R}^d \setminus \{0\}$$

which obviously proves the norm differentiability. To start with, we have

$$\begin{aligned} U_t &= \xi + \int_t^T \frac{f(s, \Theta^{x+\varepsilon h}(s)) - f(s, \Theta^x(s))}{\varepsilon} ds \\ &\quad - \int_t^T \langle (\nabla f)(s, \Theta^x(s)), ((\nabla X^x \cdot \alpha_x)(s), (P \cdot \alpha_y)(s), (Q \cdot \alpha_z)(s)) \rangle ds - \int_t^T V_s dW_s. \end{aligned}$$

Using the identity $\phi(x) - \phi(y) = (x - y) \int_0^1 \nabla \phi(y + \theta(x - y)) d\theta$ for a differentiable function $\phi : \mathbb{R} \rightarrow \mathbb{R}$, the previous equation leads to

$$\begin{aligned} U_t &= \xi + \frac{1}{\varepsilon} \int_t^T [A_{s,x}((X^{x+\varepsilon h} - X^x) \cdot \alpha_x)(s) \\ &\quad + A_{s,y}((Y^{x+\varepsilon h} - Y^x) \cdot \alpha_y)(s) + A_{s,z}((Z^{x+\varepsilon h} - Z^x) \cdot \alpha_z)(s)] ds \\ &\quad - \int_t^T \langle (\nabla f)(s, \Theta^x(s)), ((\nabla X^x \cdot \alpha_x)(s), (P \cdot \alpha_y)(s), (Q \cdot \alpha_z)(s)) \rangle ds - \int_t^T V_s dW_s \\ &= \xi + \int_t^T \Phi(s, (\tilde{X} \cdot \alpha_x)(s), (U \cdot \alpha_y)(s), (V \cdot \alpha_z)(s)) ds - \int_t^T V_s dW_s, \end{aligned} \quad (35)$$

with $\Phi(s, x, y, z) := R_s + xA_{s,x} + yA_{s,y} + zA_{s,z}$, and

$$\begin{aligned} R_s &:= -\langle (\nabla f)(s, \Theta^x(s)), ((\nabla X^x \cdot \alpha_x)(s), (P \cdot \alpha_y)(s), (Q \cdot \alpha_z)(s)) \rangle \\ &\quad + A_{s,x}(\nabla X^x \cdot \alpha_x)(s) + A_{s,y}(P \cdot \alpha_y)(s) + A_{s,z}(Q \cdot \alpha_z)(s). \end{aligned}$$

Applying the a priori estimate of Proposition 2.6 or the moment estimate from Corollary 2.7 to the BSDE (35) in (U, V) and taking into account that Φ satisfies (F2), we get

$$\begin{aligned} \|U\|_{\mathcal{S}_\beta^p}^p + \|V\|_{\mathcal{H}_\beta^p}^p &\leq C_p \left\{ \mathbb{E} \left[(e^{\beta T} |\xi|^2)^{p/2} \right] + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\Phi(s, (\tilde{X} \cdot \alpha_x)(s), 0, 0)| ds \right)^p \right] \right\} \\ &\leq C \left\{ \mathbb{E} \left[(e^{\beta T} |\xi|^p)^{p/2} \right] + \|\tilde{X}\|_{\mathcal{H}_\beta^p}^2 + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |R_s| ds \right)^p \right] \right\}, \end{aligned} \quad (36)$$

for some generic constant $C > 0$ (where we have used that $A_{\cdot,x}$ is uniformly bounded). We proceed to compute the limit of each term on the right hand side of (36) as ε goes to zero.

We first deal with the second term of the right hand side of (36). Note that \tilde{X} is solution to the linear SDE

$$\tilde{X}_t = J_t + \int_0^t [\nabla \sigma(\bar{X}_s) \tilde{X}_s] dW_s + \int_0^t [\nabla b(\bar{X}_s) \tilde{X}_s] ds,$$

where \bar{X}_s denotes some random point between X_s^x and $X_s^{x+\varepsilon h}$ and J is defined as

$$J_t := \int_0^t [\nabla X_s^x (\nabla \sigma(\bar{X}_s) - \nabla \sigma(X_s^x))] dW_s + \int_0^t [\nabla X_s^x (\nabla b(\bar{X}_s) - \nabla b(X_s^x))] ds.$$

In order to apply Lemma V.3.1 of Protter (2005) we must check that $\|J\|_{\mathcal{S}_0^p} < \infty$. Indeed, Doob's inequality leads to

$$\begin{aligned} &\mathbb{E} \left[\left(\sup_{t \in [0, T]} \left| \int_0^t [\nabla X_s^x (\nabla \sigma(\bar{X}_s) - \nabla \sigma(X_s^x))] dW_s \right|^2 \right)^{p/2} \right] \\ &\leq C \mathbb{E} \left[\left(\int_0^T |\nabla X_s^x (\nabla \sigma(\bar{X}_s) - \nabla \sigma(X_s^x))|^2 ds \right)^{p/4} \right] < \infty. \end{aligned}$$

Moreover, note that by Lebesgue's dominated convergence theorem

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left[\left(\int_0^T |\nabla X_s^x (\nabla \sigma(\bar{X}_s) - \nabla \sigma(X_s^x))|^2 ds \right)^{p/4} \right] = 0.$$

Similarly, using Jensen's inequality the finite variation part of J is an element of $\mathcal{S}_0^p(\mathbb{R})$ and

$$\lim_{\varepsilon \rightarrow 0} \|J\|_{\mathcal{S}_0^p} = 0.$$

From (Protter, 2005, Lemma V.3.1) it holds that $\|\tilde{X}\|_{\mathcal{S}_0^p} \leq 2\|J\|_{\mathcal{S}_0^p}$ and thus $\lim_{\varepsilon \rightarrow 0} \|\tilde{X}\|_{\mathcal{H}_\beta^p} = 0$.

Let us consider the terminal condition term. Denoting once again \bar{X}_T a random point between X_T^x and $X_T^{x+\varepsilon h}$ (that is componentwise) it holds that

$$\begin{aligned} & \mathbb{E}[(e^{\beta T}|\xi|^2)^{p/2}] \\ &= \mathbb{E}\left[\left(e^{\beta T}\left|\nabla g(\bar{X}_T)\left(\frac{X_T^{x+\varepsilon h} - X_T^x}{\varepsilon} - \nabla X_T^x\right) + \nabla X_T^x(\nabla g(\bar{X}_T) - \nabla g(X_T^x))\right|^2\right)^{p/2}\right] \\ &\leq C \mathbb{E}\left[\left(e^{\beta T}\left|\left(\frac{X_T^{x+\varepsilon h} - X_T^x}{\varepsilon} - \nabla X_T^x\right)\right|^2\right)^{p/2}\right] + C \mathbb{E}\left[\left(e^{\beta T}|\nabla X_T^x|^2|\nabla g(\bar{X}_T) - \nabla g(X_T^x)|^2\right)^{p/2}\right] \\ &\leq C e^{\beta T} \|\tilde{X}\|_{\mathcal{S}_0^p} + C \mathbb{E}\left[\left(e^{\beta T}|\nabla X_T^x|^2|\nabla g(\bar{X}_T) - \nabla g(X_T^x)|^2\right)^{p/2}\right] \xrightarrow{\varepsilon \rightarrow 0} 0, \end{aligned}$$

where we have used Lebesgue's dominated convergence theorem for the second summand and the estimate obtained above on the norm of \tilde{X} for the first one.

Now, let us consider the last term on the right hand side of (36). We have that

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T e^{\beta s} |R_s| ds\right)^p\right] &\leq C \mathbb{E}\left[\left(\int_0^T e^{\beta s} |(A_{s,x} - \nabla_x f(s, \Theta^x(s))) (\nabla X^x \cdot \alpha_x)(s)| ds\right)^p\right] \\ &\quad + C \mathbb{E}\left[\left(\int_0^T e^{\beta s} |(A_{s,y} - \nabla_y f(s, \Theta^x(s))) (P \cdot \alpha_y)(s)| ds\right)^p\right] \\ &\quad + C \mathbb{E}\left[\left(\int_0^T e^{\beta s} |(A_{s,z} - \nabla_z f(s, \Theta^x(s))) (Q \cdot \alpha_z)(s)| ds\right)^p\right]. \end{aligned}$$

Standard arguments yield

$$A_{t,x} \longrightarrow \nabla_x f(t, \Theta^x(t)) \quad \text{as } \varepsilon \rightarrow 0 \text{ in probability,}$$

for dt-a.a. $t \in [0, T]$. Note that $\varepsilon > 0$ is implicitly contained in $A_{t,x}$. Moreover, Proposition 2.6 and the previous calculations show that

$$\begin{aligned} & \|Y^{x+\varepsilon h} - Y^x\|_{\mathcal{S}_\beta^p}^p + \|Z^{x+\varepsilon h} - Z^x\|_{\mathcal{H}_\beta^p}^p \\ & \leq C \left\{ \mathbb{E}[(e^{\beta T}|g(X^{x+\varepsilon h}) - g(X^x)|^2)^{p/2}] + \|X^{x+\varepsilon h} - X^x\|_{\mathcal{H}_\beta^p}^p \right\} \xrightarrow{\varepsilon \rightarrow 0} 0, \end{aligned}$$

with $C > 0$ being a generic constant. This implies for dt-a.a. $t \in [0, T]$

$$Y_t^{x+\varepsilon h} \rightarrow Y_t^x, \quad Z_t^{x+\varepsilon h} \rightarrow Z_t^x, \quad \text{as } \varepsilon \rightarrow 0 \text{ in probability.}$$

Since $\nabla_y f, \nabla_z f$ are continuous, it follows that

$$\begin{aligned} A_{t,y} &\longrightarrow \nabla_y f(t, \Theta^x(t)), \quad \text{as } \varepsilon \rightarrow 0 \text{ in probability,} \\ A_{t,z} &\longrightarrow \nabla_z f(t, \Theta^x(t)), \quad \text{as } \varepsilon \rightarrow 0 \text{ in probability,} \end{aligned}$$

for dt-a.a. $t \in [0, T]$. Thus, using Lemma 1.1 and the fact that P and Q are square integrable, Lebesgue's dominated convergence theorem (which also holds, if almost sure convergence is replaced by convergence in probability, cf. Shiryaev (1995), remark on page 258) yields $\lim_{\varepsilon \rightarrow 0} \mathbb{E}[(\int_0^T e^{\beta s} |R_s| ds)^p] = 0$. Now (36) yields that

$$\lim_{\varepsilon \rightarrow 0} \{\|U\|_{\mathcal{S}_\beta^p}^p + \|V\|_{\mathcal{H}_\beta^p}^p\} = 0,$$

which proves the claim. \square

Remark 3.3. *The conclusions of Corollary 3.1 and Proposition 3.2 are still valid if one replaces $p > 2$ by $p = 2$. Then the assumptions on $T, L, \beta > 0$ have to be replaced by those of Proposition 2.5 and of Theorem 2.1.*

3.2 Strong differentiability

All previous assumptions on existence and uniqueness remain in force. In this section, we concentrate on the smoothness properties of the paths associated to the processes (Y^x, Z^x) . A first result is obtained in the following

Proposition 3.4. *Under the assumptions of Proposition 3.2, we have for $x, x' \in \mathbb{R}^d$*

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} |X_t^x - X_t^{x'}|^q\right] \leq C|x - x'|^q, \quad \text{for any } q \geq 2,$$

and for any $p > 2$

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} (e^{\beta t} |Y_t^x - Y_t^{x'}|^2)^{p/2}\right] + \mathbb{E}\left[\left(\int_0^T e^{\beta s} |Z_s^x - Z_s^{x'}|^2 ds\right)^{p/2}\right] \leq C|x - x'|^p.$$

Thus for every $x \in \mathbb{R}^d$,

- the mapping $x \mapsto Y^x$ from \mathbb{R}^d to the space of càdlàg functions equipped with the topology given by the uniform convergence on compact sets is continuous \mathbb{P} -almost surely and
- the mapping $x \mapsto Z^x$ is continuous from \mathbb{R}^d to $L^2([0, T])$ \mathbb{P} -almost surely.

In particular, for every $x \in \mathbb{R}^d$,

- the mapping $x \mapsto Y_t^x$ from \mathbb{R}^d to \mathbb{R} is continuous for all $t \in [0, T]$, \mathbb{P} -almost surely and
- the mapping $x \mapsto Z_t^x(\omega)$ is continuous for every $x \in \mathbb{R}^d$ and $dt \otimes d\mathbb{P}$ -almost all (t, ω) .

Proof. The estimate on the forward process is classical (see *e.g.* (Protter, 2005, Theorem V.37 Equation (***) p. 309)). In this proof, $C > 0$ denotes a generic constant which may differ from line to line. We apply the a priori estimate from Proposition 2.6 and get

$$\begin{aligned} & \mathbb{E}\left[\sup_{0 \leq t \leq T} (e^{\beta t} |Y_t^x - Y_t^{x'}|^2)^{p/2}\right] + \mathbb{E}\left[\left(\int_0^T e^{\beta s} |Z_s^x - Z_s^{x'}|^2 ds\right)^{p/2}\right] \\ & \leq C_p \left\{ \mathbb{E}\left[(e^{\beta T} |g(X_T^x) - g(X_T^{x'})|^2)^{p/2}\right] \right. \\ & \quad \left. + \mathbb{E}\left[\left(\int_0^T e^{\frac{\beta}{2}s} |f(s, (X^x \cdot \alpha_x)(s), \zeta(s)) - f(s, (X^{x'} \cdot \alpha_x)(s), \zeta(s))|^2 ds\right)^p\right] \right\} \\ & \leq C \left\{ \mathbb{E}\left[(e^{\beta T} |g(X_T^x) - g(X_T^{x'})|^2)^{p/2}\right] \right. \\ & \quad \left. + \mathbb{E}\left[\left(\int_0^T e^{\beta s} |f(s, (X^x \cdot \alpha_x)(s), \zeta(s)) - f(s, (X^{x'} \cdot \alpha_x)(s), \zeta(s))|^2 ds\right)^{p/2}\right] \right\}, \end{aligned}$$

with $\zeta(\cdot) := ((Y^{x'} \cdot \alpha_y)(\cdot), (Z^{x'} \cdot \alpha_z)(\cdot))$. Using the mean value theorem and the boundedness

of ∇f and ∇g (i.e. the Lipschitz property of f and g), we deduce

$$\begin{aligned}
& \mathbb{E} \left[\sup_{0 \leq t \leq T} (e^{\beta t} |Y_t^x - Y_t^{x'}|^2)^{p/2} \right] + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |Z_s^x - Z_s^{x'}|^2 ds \right)^{p/2} \right] \\
& \leq C \left\{ \mathbb{E} \left[(e^{\beta T} |X_T^x - X_T^{x'}|^2)^{p/2} \right] + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |((X^x - X^{x'}) \cdot \alpha_\chi)(s)|^2 ds \right)^{p/2} \right] \right\} \\
& \leq C \left\{ \mathbb{E} \left[(e^{\beta T} |X_T^x - X_T^{x'}|^2)^{p/2} \right] + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |X_s^x - X_s^{x'}|^2 ds \right)^{p/2} \right] \right\} \\
& \leq C |x - x'|^p,
\end{aligned}$$

where the last two lines follow by applying the change of integration from (9) and the first claim of the proposition. The continuity properties of the mappings $x \mapsto Y^x$ and $x \mapsto Z^x$ are now obtained by an application of Kolmogorov's continuity criterion (see for example (Protter, 2005, IV.7 Corollary 1)). \square

If the generator exhibits additional regularity, it even turns out that the paths of $x \mapsto Y^x$ are continuously differentiable.

Theorem 3.5. *Assume that the assumptions of Proposition 2.6 are satisfied for some $p > 2$ and assume that all second order partial derivatives of the generator f are bounded. Then, for $(x, \varepsilon), (x', \varepsilon') \in \mathbb{R}^d \times (0, \infty)$ and $h \in \mathbb{R}^d$ it holds that*

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} \left(e^{\beta t} \left| \frac{Y_t^{x+\varepsilon h} - Y_t^x}{\varepsilon} - \frac{Y_t^{x'+\varepsilon' h} - Y_t^{x'}}{\varepsilon'} \right|^2 \right)^{p/2} \right] \leq C (|x - x'|^2 + |\varepsilon - \varepsilon'|^2)^{p/2}.$$

Thus $\nabla_x Y^x$ belongs to \mathcal{H}^p and the mapping $x \mapsto Y_t^x(\omega)$ is continuously differentiable for all $t \in [0, T]$, \mathbb{P} -almost surely.

Proof. As in the previous proof, $C > 0$ denotes a generic constant which can differ from line to line. Let $p > 2$, $t, s \in [0, T]$ and $h \in \mathbb{R}^d \setminus \{0\}$. For $(x, \varepsilon) \in \mathbb{R}^d \times (0, \infty)$ let $U_s^{x, \varepsilon} := \frac{Y_s^{x+\varepsilon h} - Y_s^x}{\varepsilon}$, $V_s^{x, \varepsilon} := \frac{Z_s^{x+\varepsilon h} - Z_s^x}{\varepsilon}$, $\xi^{x, \varepsilon} := \frac{g(X_T^{x+\varepsilon h}) - g(X_T^x)}{\varepsilon}$ and $\tilde{X}_s^{x, \varepsilon} := \frac{X_s^{x+\varepsilon h} - X_s^x}{\varepsilon}$. Using the notation from the proof of Proposition 3.2, the pair $(U^{x, \varepsilon}, V^{x, \varepsilon})$ satisfies the BSDE

$$U_t^{x, \varepsilon} = \xi^{x, \varepsilon} + \int_t^T \Phi(s, \zeta^{x, \varepsilon}(s)) ds - \int_t^T V_s^{x, \varepsilon} dW_s,$$

with $\zeta^{x, \varepsilon}(s) := ((U^{x, \varepsilon} \cdot \alpha_\gamma)(s), (V^{x, \varepsilon} \cdot \alpha_\gamma)(s))$ and $\Phi(s, y, z) := (\tilde{X}^{x, \varepsilon} \cdot \alpha_\chi)(s) A_{s, x}^{x, \varepsilon} + y A_{s, y}^{x, \varepsilon} + z A_{s, z}^{x, \varepsilon}$. Note that the terms $A_{s, *}$ with $*$ $\in \{x, y, z\}$ are given by (34). Let another pair $(x', \varepsilon') \in \mathbb{R}^d \times (0, \infty)$ be given. Using Proposition 2.6, we obtain

$$\|U^{x, \varepsilon} - U^{x', \varepsilon'}\|_{\mathcal{S}_\beta^p}^p \leq C_p \left\{ \mathbb{E} \left[(e^{\beta T} |\xi^{x, \varepsilon} - \xi^{x', \varepsilon'}|^2)^{p/2} \right] + \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |\delta_2 \Phi(s)| ds \right)^p \right] \right\},$$

with

$$\begin{aligned}
\delta_2 \Phi(s) &:= (\tilde{X}^{x, \varepsilon} \cdot \alpha_\chi)(s) A_{s, x}^{x, \varepsilon} - (\tilde{X}^{x', \varepsilon'} \cdot \alpha_\chi)(s) A_{s, x}^{x', \varepsilon'} \\
&\quad + (U^{x', \varepsilon'} \cdot \alpha_\gamma)(s) (A_{s, y}^{x, \varepsilon} - A_{s, y}^{x', \varepsilon'}) + (V^{x', \varepsilon'} \cdot \alpha_\gamma)(s) (A_{s, z}^{x, \varepsilon} - A_{s, z}^{x', \varepsilon'}).
\end{aligned}$$

Using the hypotheses on f (i.e. all partial derivatives up to order two are bounded), we find

$$\begin{aligned} |\delta_2 \Phi(s)| \leq C & \left\{ |((\tilde{X}^{x,\varepsilon} - \tilde{X}^{x',\varepsilon'}) \cdot \alpha_\chi)(s)| |A_{s,x}^{x,\varepsilon}| + |(\tilde{X}^{x',\varepsilon'} \cdot \alpha_\chi)(s)| |A_{s,x}^{x,\varepsilon} - A_{s,x}^{x',\varepsilon'}| \right. \\ & \left. + |(U^{x',\varepsilon'} \cdot \alpha_\gamma)(s)| |A_{s,y}^{x,\varepsilon} - A_{s,y}^{x',\varepsilon'}| + |(V^{x',\varepsilon'} \cdot \alpha_z)(s)| |A_{s,z}^{x,\varepsilon} - A_{s,z}^{x',\varepsilon'}| \right\}. \end{aligned}$$

As a consequence

$$\begin{aligned} & \|U^{x,\varepsilon} - U^{x',\varepsilon'}\|_{\mathcal{S}_\beta^p}^p \\ & \leq C \left\{ \mathbb{E} \left[(e^{\beta T} |\xi^{x,\varepsilon} - \xi^{x',\varepsilon'}|^2)^{p/2} \right] + \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |((\tilde{X}^{x,\varepsilon} - \tilde{X}^{x',\varepsilon'}) \cdot \alpha_\chi)(s)| |A_{s,x}^{x,\varepsilon}| ds \right)^p \right] \right. \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |(\tilde{X}^{x',\varepsilon'} \cdot \alpha_\chi)(s)| |A_{s,x}^{x,\varepsilon} - A_{s,x}^{x',\varepsilon'}| ds \right)^p \right] \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |(U^{x',\varepsilon'} \cdot \alpha_\gamma)(s)| |A_{s,y}^{x,\varepsilon} - A_{s,y}^{x',\varepsilon'}| ds \right)^p \right] \\ & \quad \left. + \mathbb{E} \left[\left(\int_0^T e^{\frac{\beta}{2}s} |(V^{x',\varepsilon'} \cdot \alpha_z)(s)| |A_{s,z}^{x,\varepsilon} - A_{s,z}^{x',\varepsilon'}| ds \right)^p \right] \right\} \\ & \leq C \left\{ \mathbb{E} \left[(e^{\beta T} |\xi^{x,\varepsilon} - \xi^{x',\varepsilon'}|^2)^{p/2} \right] \right. \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |((\tilde{X}^{x,\varepsilon} - \tilde{X}^{x',\varepsilon'}) \cdot \alpha_\chi)(s)|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,x}^{x,\varepsilon}|^2 ds \right)^p \right]^{1/2} \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |(\tilde{X}^{x',\varepsilon'} \cdot \alpha_\chi)(s)|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,x}^{x,\varepsilon} - A_{s,x}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |(U^{x',\varepsilon'} \cdot \alpha_\gamma)(s)|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,y}^{x,\varepsilon} - A_{s,y}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \\ & \quad \left. + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |(V^{x',\varepsilon'} \cdot \alpha_z)(s)|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,z}^{x,\varepsilon} - A_{s,z}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \right\}, \end{aligned}$$

where for each term we used the Cauchy-Schwarz's inequality twice and that $e^{\frac{\beta}{2}s} \leq e^{\beta s}$. Applying (9) to the previous expression we get

$$\begin{aligned} \|U^{x,\varepsilon} - U^{x',\varepsilon'}\|_{\mathcal{S}_\beta^p}^p & \leq C \left\{ \mathbb{E} \left[(e^{\beta T} |\xi^{x,\varepsilon} - \xi^{x',\varepsilon'}|^2)^{p/2} \right] \right. \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\tilde{X}_s^{x,\varepsilon} - \tilde{X}_s^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,x}^{x,\varepsilon}|^2 ds \right)^p \right]^{1/2} \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\tilde{X}_s^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,x}^{x,\varepsilon} - A_{s,x}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |U_s^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,y}^{x,\varepsilon} - A_{s,y}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \\ & \quad \left. + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |V_s^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,z}^{x,\varepsilon} - A_{s,z}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \right\}. \end{aligned}$$

Since $(U^{x',\varepsilon'}, V^{x',\varepsilon'})$ is a solution in $\mathcal{S}_\beta^p \times \mathcal{H}_\beta^p$ of a BSDE, it follows from Corollary 2.7 that the quantities $\mathbb{E} \left[\left(\int_0^T e^{\beta s} |U_s^{x',\varepsilon'}|^2 ds \right)^p \right]$ and $\mathbb{E} \left[\left(\int_0^T e^{\beta s} |V_s^{x',\varepsilon'}|^2 ds \right)^p \right]$ are finite and uniformly bounded in ε' . By the assumptions on b and σ , we have

$$\mathbb{E} \left[\left(\int_0^T e^{\beta s} |\tilde{X}_s^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} < \infty.$$

In addition, by the boundedness of ∇f we have that $|A_{s,*}^{x,\varepsilon}|$ and $|A_{s,*}^{x',\varepsilon'}|$ are uniformly bounded with $* \in \{x, y, z\}$. Thus the estimate reduces to

$$\begin{aligned} & \|U^{x,\varepsilon} - U^{x',\varepsilon'}\|_{\mathcal{S}_\beta^p}^p \\ & \leq C \left\{ \mathbb{E} \left[(e^{\beta T} |\xi^{x,\varepsilon} - \xi^{x',\varepsilon'}|^2)^{p/2} \right] \right. \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\tilde{X}_s^{x,\varepsilon} - \tilde{X}_s^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,x}^{x,\varepsilon} - A_{s,x}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \\ & \quad \left. + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,y}^{x,\varepsilon} - A_{s,y}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |A_{s,z}^{x,\varepsilon} - A_{s,z}^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} \right\}. \end{aligned} \quad (37)$$

Using the mean value theorem and the fact that the second order partial derivatives are bounded it holds that

$$\begin{aligned} & |A_{s,x}^{x,\varepsilon} - A_{s,x}^{x',\varepsilon'}| + |A_{s,y}^{x,\varepsilon} - A_{s,y}^{x',\varepsilon'}| + |A_{s,z}^{x,\varepsilon} - A_{s,z}^{x',\varepsilon'}| \\ & \leq C \left\{ (|X^{x+\varepsilon h} - X^{x'+\varepsilon' h}| \cdot \alpha_x)(s) + (|Y^{x+\varepsilon h} - Y^{x'+\varepsilon' h}| \cdot \alpha_y)(s) \right. \\ & \quad + (|Z^{x+\varepsilon h} - Z^{x'+\varepsilon' h}| \cdot \alpha_z)(s) + (|X^x - X^{x'}| \cdot \alpha_x)(s) \\ & \quad \left. + (|Y^x - Y^{x'}| \cdot \alpha_y)(s) + (|Z^x - Z^{x'}| \cdot \alpha_z)(s) \right\}. \end{aligned}$$

Plugging the right hand side of this inequality in (37) and using Lemma 1.1 one gets

$$\begin{aligned} & \mathbb{E} \left[\left(\sup_{0 \leq t \leq T} e^{\beta t} |U_t^{x,\varepsilon} - U_t^{x',\varepsilon'}|^2 \right)^{p/2} \right] \\ & \leq C \left\{ \mathbb{E} \left[(e^{\beta T} |\xi^{x,\varepsilon} - \xi^{x',\varepsilon'}|^2)^{p/2} \right] \right. \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |\tilde{X}_s^{x,\varepsilon} - \tilde{X}_s^{x',\varepsilon'}|^2 ds \right)^p \right]^{1/2} + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |X_s^x - X_s^{x'}|^2 ds \right)^p \right]^{1/2} \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |X_s^{x+\varepsilon h} - X_s^{x'+\varepsilon' h}|^2 ds \right)^p \right]^{1/2} + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |Y_s^{x+\varepsilon h} - Y_s^{x'+\varepsilon' h}|^2 ds \right)^p \right]^{1/2} \\ & \quad + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |Z_s^{x+\varepsilon h} - Z_s^{x'+\varepsilon' h}|^2 ds \right)^p \right]^{1/2} + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |Y_s^x - Y_s^{x'}|^2 ds \right)^p \right]^{1/2} \\ & \quad \left. + \mathbb{E} \left[\left(\int_0^T e^{\beta s} |Z_s^x - Z_s^{x'}|^2 ds \right)^p \right]^{1/2} \right\}. \end{aligned}$$

We recall another estimate for the forward process

$$\mathbb{E} \left[|\xi^{x,\varepsilon} - \xi^{x',\varepsilon'}|^p \right] \leq C(|x - x'|^2 + |\varepsilon - \varepsilon'|^2)^{p/2},$$

which is proved for example in (Ankirchner et al., 2007, Lemma 7.4). This result combined with Proposition 3.4 leads to

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} (e^{\beta t} |U_t^{x,\varepsilon} - U_t^{x',\varepsilon'}|^2)^{p/2} \right] \leq C(|x - x'|^2 + |\varepsilon - \varepsilon'|^2)^{p/2}.$$

The last claim of the theorem follows using Kolmogorov's continuity criterion (see for example (Protter, 2005, IV.7 Corollary 1)). \square

4 Representation formulas and path regularity

One of the fundamental results of FBSDE concerns the relationship between the Malliavin and the variational (classical) derivatives of the solution process: the Malliavin derivative of the solution of the BSDE can be explicitly expressed as the product of its variational derivative (with respect to the initial parameter of the SDE) and the variational derivatives of the solution of the forward diffusion. This relationship holds both in the standard Lipschitz generator (see Proposition 5.9 of El Karoui et al. (1997)) and the quadratic generator case (see e.g. Theorem 2.9 of Imkeller and Dos Reis (2010a)) for classical BSDE without time delayed generators.

In this section we show that this relationship still holds for decoupled FBSDE with time delayed generators. Such a result is somewhat surprising since it is normally attached to the Markov setting of non-time delayed BSDE which clearly fails to materialize for this class of BSDE. Imperative for this relationship to hold is the fact that the forward process X is Markovian along with a good behavior of the terminal condition.

As in the previous section, whenever we consider the delay FBSDE (27)-(28), we assume that all conditions to ensure the existence of a unique solution (X, Y, Z) are in force. Moreover, since for $\beta \geq 0$, all β -norms are equivalent, in the following we content ourselves with giving results for $\beta = 0$. In addition, for the sake of simplicity, once again in this section we take $m = 1$, in other words the stochastic process Y solution of our BSDEs takes values in \mathbb{R} .

Malliavin's differentiability of FBSDE with time delayed generators

We recall Theorem 4.1 of Delong and Imkeller (2010b), modified to our the FBSDE setting. Theorem 4.1 from Delong and Imkeller (2010b) shows that the solutions of time delayed BSDE are Malliavin differentiable, and as a consequence, it can be deduced that the solution of the time delayed FBSDE (27)-(28) is also Malliavin differentiable. Under the condition (F3) on the coefficients of the forward equation (27), the Malliavin differentiability of the forward process X is a standard result, see for instance Theorem 2.2.1 in Nualart (1995). We denote the solution to the equations (27)-(28) by $(X, Y, Z) := (X^x, Y^x, Z^x)$. The next result states the Malliavin differentiability of (X, Y, Z) . Using notation introduced in Section 3, we define for $0 \leq u \leq t \leq T$

$$\begin{aligned} (D_u \Theta)(t) &= ((D_u X \cdot \alpha_X)(t), (D_u Y \cdot \alpha_Y)(t), (D_u Z \cdot \alpha_Z)(t)) \\ &= \left(\int_{-T}^0 D_u X_{t+v} \alpha_X(dv), \int_{-T}^0 D_u Y_{t+v} \alpha_Y(dv), \int_{-T}^0 D_u Z_{t+v} \alpha_Z(dv) \right). \end{aligned}$$

We define in the canonical way¹ the space $\mathbb{L}_{1,2}$ as the space of progressively measurable processes, $X \in \mathcal{H}^2$, that are Malliavin differentiable and normed by $\|X\|_{\mathbb{L}_{1,2}} = \mathbb{E}[\int_0^T |X_s|^2 ds + \int_0^T \int_0^T |D_u X_s|^2 ds du]^{1/2}$.

Theorem 4.1. *Under the conditions of Corollary 3.1 the Malliavin derivatives (DX, DY, DZ) of (X, Y, Z) solves uniquely in $\mathbb{L}_{1,2} \times \mathbb{L}_{1,2} \times \mathbb{L}_{1,2}$ the following time delayed FBSDE:*

$$D_u X_t = \sigma(u, X_u) + \int_u^t \nabla_x b(s, X_s) D_u X_s ds + \int_u^t \nabla_x \sigma(s, X_s) D_u X_s dW_s, \quad (38)$$

$$D_u Y_t = \nabla g(X_T) D_u X_T - \int_t^T D_u Z_s dW_s + \int_t^T \langle (\nabla f)(s, \Theta(s)), (D_u \Theta)(s) \rangle ds, \quad (39)$$

¹See Section 2.2 of Imkeller and Dos Reis (2010a), Section 5.2 of El Karoui et al. (1997) or simply Nualart (1995)

for $0 \leq u \leq t \leq T$ (zero otherwise) and where $\alpha_x, \alpha_y, \alpha_z$ are given non-random finite measures supported on $[-T, 0]$ with Θ as given by (29). Furthermore, $\{D_t Y_t : t \in [0, T]\}$ is a version of $\{Z_t : t \in [0, T]\}$.

Proof. The results concerning the forward component are well known, see Nualart (1995) or Imkeller and Dos Reis (2010a). The conditions of Corollary 3.1 ensure that Theorem 4.1 from Delong and Imkeller (2010b) can be applied. Hence Y and Z are Malliavin differentiable. \square

The representation formulas

We now present the representation formulas for (38), (39) which are effectively expressed in terms of the variational $\nabla X, \nabla Y$ and ∇Z .

Theorem 4.2. *Let the conditions of Theorem 4.1 hold. Let (X, Y, Z) , $(\nabla X, \nabla Y, \nabla Z)$ and (DX, DY, DZ) denote the solutions of FBSDE (27)-(28), (30)-(31) and (38)-(39) respectively. Then the following representation formulas hold:*

$$D_u X_t = \nabla X_t (\nabla X_u)^{-1} \sigma(u, X_u) \mathbb{1}_{\{u \leq t\}}, \quad t, u \in [0, T], \text{ d}\mathbb{P} - a.s. \quad (40)$$

$$D_u Y_t = \nabla Y_t (\nabla X_u)^{-1} \sigma(u, X_u) \mathbb{1}_{\{u \leq t\}}, \quad t, u \in [0, T], \text{ d}\mathbb{P} - a.s.$$

$$Z_t = \nabla Y_t (\nabla X_t)^{-1} \sigma(t, X_t), \quad t \in [0, T], \text{ d}\mathbb{P} \otimes dt - a.a. \quad (41)$$

$$D_u Z_t = \nabla Z_t (\nabla X_u)^{-1} \sigma(t, X_u) \mathbb{1}_{\{u \leq t\}}, \quad t, u \in [0, T], \text{ d}\mathbb{P} \otimes dt - a.a.$$

Proof. As in Theorem 4.1, the Malliavin differentiability of the forward component is well known, see Nualart (1995) or Imkeller and Dos Reis (2010a). Theorem 4.1 ensures that (DX, DY, DZ) is the unique solution of the time delayed FBSDE (38)-(39). Throughout let $t \in [0, T]$ and $u \in [0, t]$. We define the processes

$$U_{u,t} = \nabla Y_t (\nabla X_u)^{-1} \sigma(X_u) \mathbb{1}_{\{u \leq t\}} \quad \text{and} \quad V_{u,t} = \nabla Z_t (\nabla X_u)^{-1} \sigma(X_u) \mathbb{1}_{\{u \leq t\}},$$

and for $s \in [0, T]$, we set

$$\begin{aligned} D_u X(s) &= \int_{-T}^0 D_u X_{s+v} \alpha_x(dv), \\ U_u(s) &= \int_{-T}^0 U_{u,s+v} \alpha_y(dv) = \int_{-T}^0 \nabla Y_{s+v} (\nabla X_u)^{-1} \sigma(u, X_u) \mathbb{1}_{\{u \leq s+v\}} \alpha_y(dv), \\ V_u(s) &= \int_{-T}^0 V_{u,s+v} \alpha_z(dv) = \int_{-T}^0 \nabla Z_{s+v} (\nabla X_u)^{-1} \sigma(u, X_u) \mathbb{1}_{\{u \leq s+v\}} \alpha_z(dv), \end{aligned}$$

compare also with the notation in (1). Multiplying the BSDE (31) with $(\nabla X_u)^{-1} \sigma(u, X_u)$ and then using (40) we obtain for any $0 \leq u \leq t \leq T$ d \mathbb{P} -a.s. that

$$\begin{aligned} U_{u,t} &= \nabla g(X_T) D_u X_T - \int_t^T V_{u,s} dW_s \\ &\quad + \int_t^T \langle (\nabla f)(s, \Theta(s)), (D_u X(s), U_u(s), V_u(s)) \rangle ds, \end{aligned}$$

where Θ is given by $\Theta(\cdot) = ((X \cdot \alpha_x)(\cdot), (Y \cdot \alpha_y)(\cdot), (Z \cdot \alpha_z)(\cdot))$ (compare with (29) from section 3). Now, Theorem 4.1 states that the solution of BSDE (39) is unique, hence (U, V) must coincide with (DY, DZ) . Another way to see this would be to use the a priori estimates of Proposition 2.6 with (39) and the above BSDE.

Formula (41) follows easily from a combination of the representation formula for $D_u Y_t$ combined with $D_t Y_t = Z_t$, d $\mathbb{P} \otimes dt$ -a.a. (see Theorem 4.1). \square

Implications of the representation formula

The representation formulas in the previous theorem allow for a deeper analysis of the control process Z concerning its path properties.

Theorem 4.3. *Let $|f(\cdot, 0, 0, 0)|$ be uniformly bounded and the conditions of Theorem 2.8 be in force (ensuring the existence of a solution to the BSDE (33)). Then for $p \geq 2$, the mapping $t \mapsto Z_t$ is continuous $d\mathbb{P}$ -a.s. If moreover we have $p > 2$, then we also have*

$$\|Z\|_{\mathcal{S}_0^q} < \infty \quad \text{for } q \in [2, p).$$

In particular, for $p > 2$ we have for every $s, t \in [0, T]$ that $\mathbb{E}[|Y_t - Y_s|^p] \leq C|t - s|^{p/2}$ and that Y has continuous paths.

Proof. It is fairly easy to show that $(\nabla Y_t(\nabla X_t)^{-1}\sigma(t, X_t))_{t \in [0, T]}$ is continuous. By assumption, σ is a continuous function and it is well known that both processes $(\nabla X)^{-1}$ and X have continuous paths. ∇Y is continuous because its dynamics is given as a sum of a stochastic integral of a predictable process against a Brownian motion (so a continuous martingale) and a Lebesgue integral with well behaved integrand. If two processes are versions of each other and one is continuous then they are in fact modifications of each other and hence Z has continuous paths. Now since Z has continuous paths, then the representation formula (41) does not only hold $d\mathbb{P} \otimes dt$ -almost surely but in fact holds for all $t \in [0, T]$ and \mathbb{P} -almost all $\omega \in \Omega$. Using that $\nabla Y \in \mathcal{S}_0^p$ for some $p > 2$ (see Corollary 3.1 and Proposition 3.2), $(\nabla X)^{-1}, \sigma(\cdot, X) \in \mathcal{S}_0^q$ for any $q \geq 2$ and Hölder's inequality, we conclude that $Z \in \mathcal{S}_0^q$ for every $q \in [2, p)$.

The property concerning the increments of Y is easy to prove since $X, Y, Z \in \mathcal{S}_0^p$ for some $p > 2$. For $0 \leq s \leq t \leq T$, we have (recall that $|f(\cdot, \Theta(\cdot))| \leq |f(\cdot, \Theta(\cdot)) - f(\cdot, 0, 0, 0)| + |f(\cdot, 0, 0, 0)|$ and that $|f(\cdot, 0, 0, 0)|$ is uniformly bounded)

$$Y_t - Y_s = 0 + \int_s^t f(u, \Theta(u))du - \int_s^t Z_u dW_u,$$

so using the assumptions and the Burkholder-Davis-Gundy inequality, we get for a generic constant C which may vary from line to line and some $p > 2$

$$\begin{aligned} \mathbb{E}[|Y_t - Y_s|^p] &\leq C \mathbb{E}\left[\left|\int_s^t f(u, \Theta(u))du\right|^p + \left|\int_s^t Z_u dW_u\right|^p\right] \\ &\leq C|t - s|^{p/2}(1 + \|X\|_{\mathcal{S}_0^p}^p + \|Y\|_{\mathcal{S}_0^p}^p + \|Z\|_{\mathcal{S}_0^p}^p) + \mathbb{E}\left[\left(\int_s^t |Z_u|^2 du\right)^{p/2}\right] \\ &\leq C|t - s|^{p/2}. \end{aligned}$$

This in particular yields the applicability of Kolmogorov's continuity criterion to Y . \square

The L^2 -regularity result

We finish this section with the L^2 -regularity result for the control component Z of the solution of the time delayed FBSDE. Let π be a partition of the time interval $[0, T]$ with N points and mesh size $|\pi|$. We define a set of random variables via

$$\bar{Z}_{t_i}^\pi = \frac{1}{t_{i+1} - t_i} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_s ds \mid \mathcal{F}_{t_i}\right], \quad \text{for all partition points } t_i, \quad 0 \leq i \leq N - 1.$$

It is well known that $\bar{Z}_{t_i}^\pi$ is the \mathcal{F}_{t_i} -measurable least square approximation of $\frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} Z_s ds$, i.e.

$$\mathbb{E}\left[\left|\frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} Z_s ds - \bar{Z}_{t_i}^\pi\right|^2\right] = \inf_{V \in L^2(\mathcal{F}_{t_i})} \mathbb{E}\left[\left|\frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} Z_s ds - V\right|^2\right]. \quad (42)$$

We associate the process $(\bar{Z}_t^\pi)_{t \in [0, T]}$ to $\{\bar{Z}_{t_i}^\pi\}_{i=0, \dots, N-1}$ via $\bar{Z}_t^\pi = \bar{Z}_{t_i}^\pi$ for $t \in [t_i, t_{i+1})$, $0 \leq i \leq N-1$. Similarly, for the set of random variables $\{Z_{t_i}^\pi\}_{i=0, \dots, N-1} = \{Z_{t_i} : t_i \in \pi\}$, we associate the process $(Z_t^\pi)_{t \in [0, T]}$ via $Z_t^\pi = Z_{t_i}^\pi$ for $t \in [t_i, t_{i+1})$, $0 \leq i \leq N-1$. The definition of the conditional expectation implies that for every $i = 0, \dots, N-1$, we have

$$\mathbb{E}[|Z_{t_i}^\pi|^2] - 2\mathbb{E}[Z_{t_i}^\pi \bar{Z}_{t_i}^\pi] \geq -\mathbb{E}[|\bar{Z}_{t_i}^\pi|^2],$$

from which it follows that \bar{Z}^π is the best \mathcal{H}^2 -approximation of Z , leading to

$$\|Z - \bar{Z}^\pi\|_{\mathcal{H}^2} \leq \|Z - Z^\pi\|_{\mathcal{H}^2} \rightarrow 0, \quad \text{as } |\pi| \rightarrow 0.$$

Using Theorem 4.3 we are able to determine explicitly the rate of convergence of the above limit. The following result extends Theorem 5.6 from Imkeller and Dos Reis (2010a) to the setting of FBSDE with time delayed generators.

Theorem 4.4 (L^2 -regularity). *Assume that the conditions of Theorem 4.3 hold for some $p > 2$ and assume further that σ is $\frac{1}{2}$ -Hölder continuous function in its time variable. Then*

$$\max_{0 \leq i \leq N-1} \left\{ \sup_{t_i \leq t \leq t_{i+1}} \mathbb{E}[|Y_t - Y_{t_i}|^2] \right\} + \sum_{i=0}^{N-1} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |Z_s - \bar{Z}_{t_i}^\pi|^2 ds\right] \leq C|\pi|.$$

Proof. The result concerning the Y component follows immediately from Theorem 4.3. As for the result for Z , let us remark that since \bar{Z}^π is the best \mathcal{H}^2 -approximation of Z over π in the sense of (42), it follows that

$$\sum_{i=0}^{N-1} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |Z_s - \bar{Z}_{t_i}^\pi|^2 ds\right] \leq \sum_{i=0}^{N-1} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}|^2 ds\right] = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}[|Z_s - Z_{t_i}|^2] ds,$$

where the last equality follows from the use of Fubini's theorem to switch the integration order (recall that $Z \in \mathcal{S}_0^p$ for some $p > 2$). Theorem 4.3 allows to use (41) to rewrite the difference inside the expectation. We have $Z_s - Z_{t_i} = I_1 + I_2 + I_3$ with $I_1 = [\nabla Y_s - \nabla Y_{t_i}](\nabla X_{t_i})^{-1} \sigma(t_i, X_{t_i})$, $I_2 = \nabla Y_s[(\nabla X_s)^{-1} - (\nabla X_{t_i})^{-1}] \sigma(t_i, X_{t_i})$, $I_3 = \nabla Y_s(\nabla X_s)^{-1} [\sigma(s, X_s) - \sigma(t_i, X_{t_i})]$ and $s \in [t_i, t_{i+1}]$.

From the proof of part (ii) of Theorem 5.8 in Imkeller and Dos Reis (2010b) one obtains that

$$\sum_{i=0}^{N-1} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |I_2|^2 ds + \int_{t_i}^{t_{i+1}} |I_3|^2 ds\right] \leq C|\pi|.$$

The calculations that lead to the above result are quite easy to carry out. They rely on known estimates for SDEs found for instance in Theorem 2.3 and 2.4 of Imkeller and Dos Reis (2010a) combined with the fact that $\nabla Y \in \mathcal{S}^p$ for some $p > 2$.

To handle the term I_1 one needs to proceed with more care. Let us start with a simple trick:

$$\mathbb{E}\left[|(\nabla Y_s - \nabla Y_{t_i})(\nabla X_{t_i})^{-1} \sigma(t_i, X_{t_i})|^2\right] = \mathbb{E}\left[\mathbb{E}[|\nabla Y_s - \nabla Y_{t_i}|^2 | \mathcal{F}_{t_i}] |(\nabla X_{t_i})^{-1} \sigma(t_i, X_{t_i})|^2\right]. \quad (43)$$

Writing the BSDE for the difference $\nabla Y_s - \nabla Y_{t_i}$ for $s \in [t_i, t_{i+1}]$ we get for a generic constant $C > 0$ that

$$\begin{aligned} \mathbb{E} \left[|\nabla Y_s - \nabla Y_{t_i}|^2 \middle| \mathcal{F}_{t_i} \right] &\leq C \mathbb{E} \left[\left| \int_{t_i}^s \langle (\nabla f)(r, \Theta(r)), (\nabla \Theta)(r) \rangle dr \right|^2 + \left| \int_{t_i}^s \nabla Z_r dW_r \right|^2 \middle| \mathcal{F}_{t_i} \right] \\ &\leq C \mathbb{E} \left[|\pi| \int_{t_i}^{t_{i+1}} |(\nabla \Theta)(r)|^2 dr + \int_{t_i}^{t_{i+1}} |\nabla Z_r|^2 dr \middle| \mathcal{F}_{t_i} \right], \end{aligned}$$

where we used the uniform boundedness of the derivatives of f , Jensen's inequality, Itô's isometry and proceeded to maximize over the time interval $[t_i, t_{i+1}]$. Combining the last line with (43) and using the tower property, we obtain

$$\begin{aligned} &\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[\mathbb{E} \left[|\nabla Y_s - \nabla Y_{t_i}|^2 \middle| \mathcal{F}_{t_i} \right] |(\nabla X_{t_i})^{-1} \sigma(t_i, X_{t_i})|^2 \right] ds \\ &\leq C \sum_{i=0}^{N-1} |\pi| \mathbb{E} \left[\left(|\pi| \int_{t_i}^{t_{i+1}} |(\nabla \Theta)(r)|^2 dr + \int_{t_i}^{t_{i+1}} |\nabla Z_r|^2 dr \right) |(\nabla X_{t_i})^{-1} \sigma(t_i, X_{t_i})|^2 \right] \\ &\leq |\pi| \mathbb{E} \left[\sup_{0 \leq t \leq T} |(\nabla X_t)^{-1} \sigma(t, X_t)|^2 \sum_{i=0}^{N-1} \left(|\pi| \int_{t_i}^{t_{i+1}} |(\nabla \Theta)(r)|^2 dr + \int_{t_i}^{t_{i+1}} |\nabla Z_r|^2 dr \right) \right] \\ &= |\pi| \mathbb{E} \left[\sup_{0 \leq t \leq T} |(\nabla X_t)^{-1} \sigma(t, X_t)|^2 \left(|\pi| \int_0^T |(\nabla \Theta)(r)|^2 dr + \int_0^T |\nabla Z_r|^2 dr \right) \right] \\ &\leq C |\pi|, \end{aligned}$$

where in the last line we used the fact that $\nabla X, (\nabla X)^{-1}, X \in \mathcal{S}_0^q$ for every $q \geq 2$ and that $\nabla Y, \nabla Z \in \mathcal{H}_0^p$ for some $p > 2$ (in combination with Hölder's inequality) to conclude the finiteness of the expectation. Combining this estimate with the ones for I_2 and I_3 finishes the proof. \square

Having established a path regularity result for FBSDE with time-delayed generators one can now start discussing a working numerical scheme. Given the nature of this class of BSDE one is naturally inclined to propose a scheme based on a discretization of a Picard iteration. Roughly speaking such a scheme follows the footsteps of Bender and Denk (2007). The scheme proposed in Bender and Denk (2007) is shown to converge to the scheme proposed by Bouchard and Touzi (2004) whose convergence is known. For delay FBSDE the discussion is more involved since the relevant question seems to be not the convergence of the Bender and Denk (2007) scheme to that of Bouchard and Touzi (2004) but the convergence of Bouchard and Touzi (2004) to the original solution. A concrete scheme as well as convergence results is left for future research.

References

- S. Ankirchner, P. Imkeller, and G. Dos Reis. Classical and variational differentiability of BSDEs with quadratic growth. *Electron. J. Probab.*, 12:1418–1453, 2007.
- C. Bender and R. Denk. A forward simulation of backward SDEs. *Stochastic Processes and their Applications*, 117(12):1793–1812, December 2007. doi: doi:10.1016/j.spa.2007.03.005.
- B. Bouchard and N. Touzi. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.*, 111(2):175–206, 2004. ISSN 0304-4149.

- L. Delong. Applications of time-delayed backward stochastic differential equations to pricing, hedging and management of insurance and financial risks. Preprint - arXiv:1005.4417v1, 2010.
- L. Delong and P. Imkeller. Backward stochastic differential equations with time delayed generator - Results and counterexamples. To appear in: The Annals of Applied Probability, 2010a.
- L. Delong and P. Imkeller. On Malliavin's differentiability of BSDE with time delayed generators driven by Brownian motion and Poisson random measures. To appear in: Stochastic Processes and their Applications, 2010b.
- N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.
- P. Imkeller and G. Dos Reis. Path regularity and explicit convergence rate for BSDE with truncated quadratic growth. *Stochastic Processes Appl.*, 120(3):348–379, 2010a.
- P. Imkeller and G. Dos Reis. Corrigendum to “Path regularity and explicit convergence rate for BSDE with truncated quadratic growth” [stochastic process. appl. 120 (2010) 348-379]. To appear in: Stochastic Processes and their Applications, 2010b.
- D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995. ISBN 0-387-94432-X.
- E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Syst. Control Lett.*, 14(1):55–61, 1990.
- P. E. Protter. *Stochastic integration and differential equations*. Applications of Mathematics (New York). Springer-Verlag, 2nd edition, 2005. Version 2.1.
- A. N. Shiryaev. *Probability. Transl. from the Russian by R. P. Boas. 2nd ed.* Graduate Texts in Mathematics. 95. New York, NY: Springer-Verlag., 1995.
- J. Wang, Q. Ran, and Q. Chen. L^p solutions of BSDEs with stochastic Lipschitz condition. *J. Appl. Math. Stoch. Anal.*, pages Art. ID 78196, 14, 2007. ISSN 1048-9533.
- J. Yong and X. Y. Zhou. *Stochastic controls. Hamiltonian systems and HJB equations*. Applications of Mathematics. 43. New York, NY: Springer. xx, 438 p., 1999.