



## Semilinear Kolmogorov equations on the space of continuous functions via BSDEs

This is a pre print version of the following article:

*Original:*

Masiero, F., Orrieri, C., Tessitore, G., Zanco, G. (2021). Semilinear Kolmogorov equations on the space of continuous functions via BSDEs. STOCHASTIC PROCESSES AND THEIR APPLICATIONS, 136, 1-56 [10.1016/j.spa.2021.01.009].

*Availability:*

This version is available <http://hdl.handle.net/11365/1222421> since 2022-12-16T13:22:41Z

*Published:*

DOI:10.1016/j.spa.2021.01.009

*Terms of use:*

Open Access

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. Works made available under a Creative Commons license can be used according to the terms and conditions of said license.

For all terms of use and more information see the publisher's website.

(Article begins on next page)

# Semilinear Kolmogorov equations on the space of continuous functions via BSDEs

Federica Masiero<sup>a</sup>, Carlo Orrieri<sup>b,1</sup>, Gianmario Tessitore<sup>a</sup>, Giovanni Zanco<sup>c,1,2,\*</sup>

<sup>a</sup>*Dipartimento di Matematica e Applicazioni, Università di Milano-Bicocca. via Cozzi 55, 20125 Milano, Italia*

<sup>b</sup>*Dipartimento di Matematica, Università di Trento. Via Sommarive 14, 38123 Trento, Italia*

<sup>c</sup>*Dipartimento di Economia e Finanza, LUISS Guido Carli. Viale Romania 32, 00197 Roma, Italia*

---

## Abstract

We deal with a class of semilinear parabolic PDEs on the space of continuous functions that arise, for example, as Kolmogorov equations associated to the infinite-dimensional lifting of path-dependent SDEs. We investigate existence of smooth solutions through their representation via forward-backward stochastic systems, for which we provide the necessary regularity theory. Because of the lack of smoothing properties of the parabolic operators at hand, solutions in general will only share the same regularity as the coefficients of the equation. To conclude we exhibit an application to Hamilton-Jacobi-Bellman equations associated to suitable optimal control problems.

---

AMS 2010 Subject classification: 35K58, 60H10, 60H30, 93E20.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Notation and preliminaries</b>	<b>4</b>
2.1	BSDEs toolbox . . . . .	5
<b>3</b>	<b>Setting of the problem and Assumptions</b>	<b>7</b>
3.1	The forward-backward system and the PDE . . . . .	7
3.2	Assumptions . . . . .	10
<b>4</b>	<b>The forward-Backward system</b>	<b>11</b>
4.1	First-order differentiability of the BSDE . . . . .	13
4.2	Second-order differentiability of the BSDE . . . . .	17
<b>5</b>	<b>Solution to the Kolmogorov equation in <math>\mathcal{L}^2</math></b>	<b>20</b>
<b>6</b>	<b>Solution of the Kolmogorov equation in <math>\mathcal{D}</math></b>	<b>21</b>
<b>7</b>	<b>Application to stochastic optimal control</b>	<b>31</b>
<b>8</b>	<b>Appendix</b>	<b>34</b>

## 1. Introduction

The aim of this paper is to address the infinite dimensional semilinear backward Kolmogorov PDE

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) + Du(t, x) [Ax + B(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 u(t, x)] = G(t, x, u(t, x), Du(t, x) \Sigma) , \\ u(T, \cdot) = \Phi \end{cases} \quad (1.1)$$

in the space of continuous functions on a real interval. Under suitable assumptions on the coefficients and on the terminal condition we provide existence of smooth (classical) solutions to (1.1) through the associated

---

\*Corresponding author

<sup>1</sup>C. Orrieri and G. Zanco are supported by the GNAMPA Project 2019 “Trasporto ottimo per dinamiche con interazione”. This research has been performed in the framework of the project Fondazione Cariplo-Regione Lombardia MEGAsTAR “Matematica d’Eccellenza in biologia ed ingegneria come acceleratore di una nuova strateGia per l’ATtRattività dell’ateneo pavese”.

<sup>2</sup>G. Zanco was partly funded by the Austrian Science Fund (FWF) project F 65.

forward-backward stochastic system, extending the methods introduced in Flandoli and Zanco (2016) for the linear equation (i.e.  $G \equiv 0$ ).

PDEs of the above given form naturally arise in connection with path-dependent stochastic differential equations in finite dimension through their infinite-dimensional reformulation in the so-called product space framework, proposed first in Delfour and Mitter (1972) for deterministic systems with delay and in Chojnowska-Michalik (1978) for stochastic ones. In particular semilinear equations as (1.1) describe the value function of optimal control problems for stochastic path dependent state equations.

While the solution theory for path-dependent stochastic systems is classical (see e.g. the monograph Mohammed (1984)) at least when the coefficients are regular enough, the study of associated PDEs is a relatively recent subject for which different approaches have been proposed in the last years.

The recent research activity on path-dependent functionals of stochastic processes and related PDEs originated from insight by Dupire (2009) and investigation by Cont and Fournié (2013) in their development of the so-called functional Itô calculus (a detailed discussion on the relation between this calculus and the product-space approach is carried out in Zanco (2015)). Due to the general lack of regularity of path-dependent functionals (most notably  $\gamma \mapsto \sup_{s \in [0, T]} |\gamma_s|$ ) various authors have introduced different weak notions of solutions for nonlinear path-dependent PDEs, see for example Ekren et al. (2014, 2016); Cosso et al. (2018); Cosso and Russo (2019b); Cordoni et al. (2017, 2019); Bayraktar and Keller (2018); Zhou, Jianjun (2018). In many cases such PDEs arise in connection with stochastic control problems, as in Fuhrman et al. (2010); Tang and Zhang (2015) when dealing with dynamic programming; also the maximum principle approach has been investigated for path-dependent problems, see e.g. Guatteri et al. (2017). Nonetheless, at the current stage there is no complete theory even for regular solutions of nonlinear PDEs (existence of solutions was proved in Cosso et al. (2014) and Cosso and Russo (2019a) only for coefficients with a very specific cylindrical form), and only the linear case has been extensively investigated in this sense (see Flandoli and Zanco (2016) and Di Girolami and Russo (2018)). The appearance of so many different approaches is essentially motivated by the intrinsic infinite-dimensional nature of the problem which then reflects in different notions of differential for functions of paths.

It is by now well understood that the parabolic-type operators associated to path-dependent SDEs do not possess smoothing properties in general (although there is a kind of partial smoothing in some problems with delay and for particular choices of the coefficients, see Gozzi and Masiero (2017) and Rosestolato and Świech (2017)): this affects the regularity of any type of solution, and makes the study of regular solutions nontrivial. In the case we discuss here, a precise investigation of differentiability of the forward-backward system has never been rigorously carried out before. The approach we propose provides, under suitable assumptions on the regularity of the coefficients and on finite-dimensionality of the noise, a solution theory for general PDEs on the space of continuous functions. The just mentioned lack of smoothing properties prevents us from weakening our assumptions on the regularity of the coefficients, that seems to be really essential. Indeed our method has its main goal in providing regular strong solutions to the Kolmogorov PDE, thus, as a consequence, feedback controls in the associated control problems. The drawback is that as far as we understand it cannot be generalized to fully nonlinear equations with irregular coefficients, in contrast to the methods used to build viscosity solutions of Ekren et al. (2014).

Notice that the noise in the infinite-dimensional reformulation of path-dependent stochastic differential equations is naturally finite dimensional. We expect that the method proposed herein can be extended to semilinear Kolmogorov equations associated to stochastic path-dependent PDEs with infinite-dimensional trace class noise, that is, to delay SDEs with values in an infinite-dimensional Hilbert space (some results in this direction are presented in Rosestolato (2016)); certain technical aspect of our proofs should be however adjusted to cover this extension, see e.g. Remark 16.

Let us now briefly introduce our framework and main results, sketching the general lines of the proofs. Fix a finite time horizon  $T > 0$  and consider the path-dependent SDE in  $\mathbb{R}^d$

$$\begin{cases} d\xi_s = b_s(\xi_{[0,s]}) ds + \sigma dW_s, & s \in [t, T], \\ \xi_t = \gamma_t, \end{cases} \quad (1.2)$$

where  $W$  is a  $d_1$ -dimensional Brownian Motion,  $\sigma$  is a  $d \times d_1$  matrix,  $t \in [0, T]$  and  $\gamma_t$  is a given function that belongs to  $D([0, t]; \mathbb{R}^d)$  (the space of càdlàg functions on  $[0, t]$ , endowed with the supremum norm). The value of the solution process  $\xi$  at time  $s$  is denoted by  $\xi_s$ , while its path up to time  $s$  is denoted by  $\xi_{[0,s]}$ . The drift  $b$  at time  $s$  depends on the whole past trajectory of the solution  $\xi_{[0,s]}$  and it is therefore given as a family  $\{b_s\}_{s \in [0, T]}$

$$b_s : D([0, s]; \mathbb{R}^d) \rightarrow \mathbb{R}^d.$$

Note that for different times the drift  $b$  is defined on different spaces of paths; while this is not an issue in the study of the SDE (1.2), it becomes a delicate question for the investigation of the associated Kolmogorov PDE. Furthermore, even if the solution to (1.2) has continuous paths from time  $t$  on, it is convenient (actu-

ally unavoidable) to formulate everything in spaces of càdlàg functions.

The product-space reformulation of (1.2) consists in separating the present state  $\xi_t$  from the past trajectory  $\xi_{[0,t]}$ , rewriting the second one via a time change as a function on  $[-t, 0)$  and then lengthening it towards the past up to  $[-T, 0)$ . In this way it is possible to distinguish between the time  $t$  of the forward equation and the time variable of the past trajectory: for any time  $T$  a process

$$X_t = \left( \begin{array}{c} \xi_t \\ \{\xi_{t+r}\}_{r \in [-T, 0)} \end{array} \right) \in \mathbb{R}^d \times D([-T, 0); \mathbb{R}^d) ,$$

is defined, whose second component is now defined on a fixed functional space. This reformulation allows to recover Markovianity and turns out to be particularly convenient to investigate differentiability properties of the solution of the nonlinear Kolmogorov PDE. As a drawback an additional linear term comes into play, which is given by a first order differential operator  $A$  usually referred to as the *generator of the delay*. Indeed, the process  $X$  turns out to be a solution to the following SDE (the *forward equation* in what follows)

$$\begin{cases} dX_s = AX_s ds + B(s, X_s) ds + \Sigma dW_s , & s \in (t, T) \\ X_t = x , \end{cases} \quad (1.3)$$

where  $B$ ,  $\Sigma$  and  $x$  are suitable infinite-dimensional liftings of  $b$ ,  $\sigma$  and  $\gamma_t$ , respectively.

Given (1.3) it is natural to associate, at least formally, the linear backward Kolmogorov equation

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) + Du(t, x) [Ax + B(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 u(t, x)] = 0, \\ u(T, \cdot) = \Phi \end{cases} \quad (1.4)$$

on  $[0, T] \times (\mathbb{R}^d \times D([-T, 0); \mathbb{R}^d))$ . The terms  $Du$  and  $D^2u$  denote the Fréchet differentials of the solution  $u$  with respect to the variable  $x \in \mathbb{R}^d \times D([-T, 0); \mathbb{R}^d)$  and the terms  $DuB$  and  $\text{Tr} [\Sigma \Sigma^* D^2 u]$  only depends on the  $\mathbb{R}^d$ -component of  $x$  (recall that  $X$  generates from a finite-dimensional SDE).

Then, to account for a nonlinear term  $G$  as in equation (1.1), the introduction of the following backward SDE (BSDE) is essential

$$\begin{cases} dY_s = G(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x}) ds + Z_s dW_s, & s \in [t, T], \\ Y_T = \Phi(X^{t,x}(T)) , \end{cases} \quad (1.5)$$

where the variables  $t$  and  $x$  refer to the initial data of the forward equation satisfied by  $X$ . A solution to (1.5) is a pair of processes  $(Y^{t,x}, Z^{t,x})$  with values in  $\mathbb{R} \times \mathbb{R}^{d_1}$  and the (partially-coupled) system generated by (1.3) and (1.5) goes under the name of *forward-backward system*. Notice that, even if the solution  $(Y, Z)$  is finite-dimensional, it depends in a nontrivial way on the forward process  $X$  that takes values in an infinite-dimensional space.

Our main result is a version of the nonlinear Feynman-Kac formula in terms of backward SDEs.

**Theorem.** *The function*

$$u(t, x) = Y_t^{t,x} ,$$

where  $(Y^{t,x}, Z^{t,x})$  is the unique solution of (1.5), is a classical solution of the semilinear Kolmogorov backward equation with terminal condition  $\Phi$ .

Here by classical solution we mean a function that is two times differentiable with respect to  $x$  and satisfies (1.1) for every  $t \in [0, T]$  and for every  $x$  that belongs to the domain of  $A$ . Since the solution  $u(t, x)$  is represented by  $Y_t^{t,x}$ , where  $(Y^{t,x}, Z^{t,x})$  is the solution to the (1.5), it is crucial to study Fréchet differentiability of the map  $(t, x) \mapsto Y^{t,x}$ , up to the second order with respect to the variable  $x$ . At our best knowledge a precise investigation of differentiability, up to the second order, of the forward-backward system has never been rigorously carried out before in the generality needed here. As a matter of fact higher order differentiability for BSDEs has only been taken into account in Izumi (2018) in a non-Markovian setting and in Malliavin sense. Although the two arguments have several technical similarities it seems that here we cannot use the result in Izumi (2018), as we rely only on Gâteaux and Fréchet differential calculus.

We firstly prove the above theorem in the space  $L^2(-T, 0; \mathbb{R}^d)$  and then extend our results to  $D([-t, 0); \mathbb{R}^d)$ . Note that requiring regularity in  $L^2$ -sense drastically restricts the class of models one can consider so that the  $L^2$ -theory has no much relevance by itself. Nonetheless it is a fundamental intermediate step for studying the PDE on  $D$ . As a marginal remark we notice that the  $L^2$ -theory can be easily adapted to get existence of classical solutions with coefficients in  $L^p$ ,  $p > 2$ , allowing to recover already at this level some interesting examples.

Once the  $L^2$ -theory is established, we proceed as follows: the coefficients  $B$ ,  $G$  and  $\Phi$ , defined on  $D([-T, 0); \mathbb{R}^d)$ ,

are approximated by suitable sequences  $B^n$ ,  $G^n$ ,  $\Phi^n$  defined on  $L^2(-T, 0; \mathbb{R}^d)$ , providing a family of solutions  $u^n$  of the approximated PDEs

$$\begin{cases} \frac{\partial u^n}{\partial t}(t, x) + Du^n(t, x) [Ax + B^n(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 u^n(t, x)] = G^n(t, x, u^n(t, x), Du^n(t, x) \Sigma), \\ u^n(T, \cdot) = \Phi^n. \end{cases}$$

To conclude the proof we need pass to the limit as  $n$  tends to infinity in each term of the PDE. While the derivation of the semilinear PDE in  $L^2$  is similar to the linear case, the passage to the limit shows substantial differences with the corresponding proof for the linear case, and requires a nontrivial analysis of the convergence of the BSDE (1.5) together with its first and second derivative.

The choice to work in  $D([-t, 0]; \mathbb{R}^d)$  is motivated by what we hinted at above: there are very few functions satisfying the needed regularity assumptions in  $L^2$  but many significant examples, most notably those involving pointwise evaluations of the path, can be recovered switching to Banach spaces with a finer topology (see Flandoli and Zanco (2016) for a discussion of several examples meeting our assumptions). To this end, the choice of the space of continuous functions would seem the most natural and appropriate one. However the infinite dimensional reformulation mentioned above has the drawback of creating discontinuities: as an example, the lifted drift term  $B$  has the form

$$B(t, x) = \begin{pmatrix} b_t \\ 0 \end{pmatrix}$$

and has to be interpreted as a càdlàg function that is non-zero only at the current time  $t$ . Consequently, the operator  $A$  introduces a transport effect, explicitly visible in the mild formulation of (1.3), shifting the discontinuity over time. It is therefore convenient to formulate everything in the larger space of càdlàg paths and restrict to the subspace of continuous paths when needed.

The role of the intermediate  $L^2$  step can be informally explained as follows: the natural scheme to investigate existence of regular solutions to (1.1) consists in combining some form of Itô formula with a smoothness result for the solution of the forward-backward system (with respect to the initial data of the forward equation). However, because of the spaces we are working with and of the particular form of the noise, no Itô formula applies to our system and we have to rely on a particular Taylor expansion that exploits the Markovianity recovered through our infinite-dimensional reformulation. Furthermore, to obtain the PDE we need a control over the second order term which is achievable only in  $L^p$  spaces; in particular this allows to show that the second order term is concentrated on the finite-dimensional component, thus providing the trace term as it appears in the equation. The same result cannot be directly proved through estimates with respect to the supremum norm.

Let us finally stress that, as all the technical difficulties related to path-dependency and to the use of càdlàg spaces are already present in the additive noise case, we choose to work in such a setting that considerably simplifies the technical aspects of several points. We however expect our result to hold, under suitable non-degeneracy conditions, when the diffusion coefficient is non-constant as well. For some results on the associated PDE in the linear case see Flandoli et al. (2018).

We eventually apply the result to a stochastic control problem, for the state equation

$$\begin{cases} dX_s^u = AX_s^u ds + B(s, X_s^u) ds + \Sigma u_s ds + \Sigma dW_s, & s \in [t, T] \\ X_t^u = x. \end{cases}$$

We aim at minimizing the cost functional  $\mathcal{J} : [0, T] \times \mathcal{D} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}$

$$\mathcal{J}(t, x, u) := \mathbb{E} \int_t^T [L(s, X_s^{u;t,x}) + Q(u_s)] ds + \mathbb{E} \Upsilon(X_T^{u;t,x}),$$

over all *admissible* controls. The Hamilton-Jacobi-Bellman equation is related to a semilinear Kolmogorov PDE that can be solved in classical sense thanks to the results proved herein; as a consequence we are able to prove the existence of optimal controls in strong formulation.

We briefly outline the structure of the paper. Section 2 contains notation and classical results on BSDE that will be used in the sequel. Section 3 introduces rigorously the product space framework and the assumptions that will stand throughout the paper. In Section 4 we prove some results about regularity of the solution of the stochastic forward-backward system with respect to the initial data of the forward process. Up to this point results are proved in a generic Banach space  $E$ , and possibly specialized to particular spaces when needed. Section 5 is devoted to the proof of existence of a solution to the semilinear backward PDE in  $L^2$ . In Section 6 we carry out the limit procedure and prove the main result. Finally Section 7 contains some applications to optimal control problems.

## 2. Notation and preliminaries

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and fix a time interval  $[0, T]$ . We denote by  $W_t$ ,  $t \geq 0$ , a  $d_1$ -dimensional standard Brownian motion and by  $\mathcal{F}_t$  the associated natural filtration, completed with the null sets in  $\mathcal{F}_T$ . All the measurability properties we refer to have to be intended with respect to this filtration. In the following, given a  $\mathcal{F}_t$ -measurable random variable with finite expectation, we denote by  $\mathbb{E}^{\mathcal{F}_s}(X_t) := \mathbb{E}(X_t | \mathcal{F}_s)$  the conditional expectation of  $X_t$  given  $\mathcal{F}_s$ .

We denote by  $E$  a general Banach space, whose norm is given by  $|\cdot|_E$ , or simply by  $|\cdot|$ , when no confusion can arise. For any pair of Banach spaces  $E, F$ , we write  $L(E, F)$  for the space of linear and bounded operators  $T : E \rightarrow F$ , endowed with the operator norm. In the special case  $F = \mathbb{R}$ , we shorthand  $E' := L(E; \mathbb{R})$ . The operator norm is indicated by  $\|T\|_{L(E, F)}$ , or  $\|T\|$  if no confusion is possible.

Moreover, given two possibly different Banach spaces  $E_1, E_2$  we indicate with  $L(E_1, E_2; F)$  the space of bilinear maps (linear in each argument) from  $E_1 \times E_2 \rightarrow F$ . In the following we will identify  $L(E_1, E_2; F)$  with  $L(E_1; L(E_2; F))$ .

For every  $p, q \geq 1$ , we use the following notation for classes of random variables and stochastic processes with values in a Banach space  $E$ :

- $L_{\mathcal{F}_T}^p(\Omega; E)$ , the set of  $\mathcal{F}_T$ -measurable  $E$ -valued random variable endowed with the norm

$$\|X\|_{L_{\mathcal{F}_T}^p(\Omega; E)} := (\mathbb{E}|X|_E^p)^{1/p};$$

- $L^p(\Omega \times [0, T]; E)$ , the set of progressively measurable  $E$ -valued processes endowed with the norm

$$\|X\|_{L^p(\Omega \times [0, T]; E)} := \left( \mathbb{E} \int_0^T |X_t|_E^p dt \right)^{1/p};$$

- $L^p(\Omega; L^q(0, T; E))$ , the space of progressively measurable  $E$ -valued processes with the norm given by

$$\|X\|_{L^p(\Omega; L^q(0, T; E))} := \left( \mathbb{E} \left( \int_0^T |X_t|_E^q dt \right)^{p/q} \right)^{1/p};$$

- $L^p(\Omega; C([0, T]; E))$ , the space of progressively measurable  $E$ -valued processes such that the map  $t \mapsto X_t$  is a.s. continuous and the norm

$$\|X\|_{L^p(\Omega; C([0, T]; E))} := \left( \mathbb{E} \sup_{t \in [0, T]} |X_t|_E^p \right)^{1/p},$$

is finite.

If  $E = \mathbb{R}$ , to shorten the notation we denote by  $\mathcal{K}_p$  the product space

$$\mathcal{K}_p := L^p(\Omega; C([0, T]; \mathbb{R})) \times L^p(\Omega; L^2(0, T; \mathbb{R}^{d_1})). \quad (2.1)$$

We say that a function  $R : E \rightarrow F$  belongs to  $C^{n, \alpha}(E; F)$  if it is  $n$ -times Fréchet differentiable in  $E$  with measurable differentials  $D^j R$ ,  $j = 1, \dots, n$ , and the map  $x \mapsto D^n R(x)$  is  $\alpha$ -Hölder continuous with measurable norm.

We say that a function  $S : [0, T] \times E \rightarrow F$  belongs to  $C^{1; n, \alpha}$  if for every  $x \in E$  the map  $t \mapsto S(t, x)$  is differentiable with measurable differential and for every  $t \in [0, T]$  the map  $x \mapsto S(t, x)$  belongs to  $C^{n, \alpha}$ . For space-time functions  $R = R(t, x)$  we will denote by  $\frac{\partial R}{\partial t}$  the derivative w.r.t.  $t$  and by  $D^j R$  the Fréchet differentials w.r.t.  $x$ .

In what follows we generally use capital letters  $X, Y, Z, \dots$  to denote random variables, on the contrary we use small letters  $x, y, z, \dots$  to denote deterministic objects. Whenever we write  $a \lesssim b$ , with  $a, b \in \mathbb{R}$ , we mean that there exists a constant  $c > 0$  for which  $a \leq cb$ .

### 2.1. BSDEs toolbox

Here we collect some basic results from the theory of Backward SDEs that will be useful in the sequel. We refer to Pardoux and Răşcanu (2014) for a general introduction to the subject.

Given a  $\mathcal{F}_T$ -measurable real-valued random variable  $\eta$  and a driver  $g : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}$  which is

$\mathcal{F}_t \otimes \mathcal{B}([0, t]) \otimes \mathcal{B}(\mathbb{R} \times \mathbb{R}^{d_1})$ -measurable, we say that a pair of progressively measurable processes  $(Y, Z) \in \mathcal{K}_p$  is a solution to the BSDE associated with  $(g, \eta)$  if  $\mathbb{P}$ -a.s.

$$Y_t = \eta + \int_t^T g(s, Y_s, Z_s) ds - \int_t^T Z_s dW_s, \quad 0 \leq t \leq T. \quad (2.2)$$

In a differential formulation, we also write that  $\mathbb{P}$ -a.s.

$$-dY_t = g(t, Y_t, Z_t) dt - Z_t dW_t, \quad Y_T = \eta, \quad 0 \leq t \leq T. \quad (2.3)$$

Wellposedness results and a priori estimates for solutions to (2.2) hold under specific assumptions on the pair  $(g, \eta)$ . Let us recall here a classical result with uniform Lipschitz hypothesis.

**Proposition 1.** *Let  $p > 1$  and  $\eta \in L^p_{\mathcal{F}_T}(\Omega; \mathbb{R})$ . Moreover, suppose that*

(i) *There exists  $L > 0$  for which*

$$|g(s, y_1, z_1) - g(s, y_2, z_2)| \leq L(|y_1 - y_2| + |z_1 - z_2|), \quad \forall s \in [0, T], \mathbb{P}\text{-a.s.},$$

*for any  $y_1, y_2 \in \mathbb{R}$  and  $z_1, z_2 \in \mathbb{R}^{d_1}$ ;*

(ii)  $\mathbb{E} \left( \int_0^T |g(s, 0, 0)|^2 ds \right)^{p/2} < +\infty$ .

*Then the BSDE (2.2) admits a unique solution  $(Y, Z) \in \mathcal{K}_p$  and for every  $t \in [0, T]$  it holds*

$$\mathbb{E} \sup_{s \in [t, T]} |Y_s|^p + \mathbb{E} \left( \int_t^T |Z_r|^2 dr \right)^{p/2} \leq C \mathbb{E} \left( \int_t^T |g(r, 0, 0)|^2 dr \right)^{p/2} + C \mathbb{E} |\eta|^p, \quad (2.4)$$

where  $C = C(p, L, T)$  is a positive constant.

For a proof of Proposition 1 we refer to (Pardoux and Răşcanu, 2014, Thm. 5.21) where (integrable) time-dependent Lipschitz constants are also taken into account.

In the sequel we will be interested in BSDEs associated with data  $(g, \eta)$  depending on a given stochastic process. Precisely, consider a stochastic process  $X$  with values in a general Banach space  $E$  and assume that  $g : \Omega \times [0, T] \times E \times \mathbb{R} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}$  and  $\eta = \varphi(\cdot)$ ,  $\varphi : E \rightarrow \mathbb{R}$ , are given measurable functions. If we write the equation

$$Y_t = \varphi(X_T) + \int_t^T g(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s dW_s, \quad 0 \leq t \leq T, \quad (2.5)$$

existence and uniqueness of a solution in  $\mathcal{K}_p$  is a consequence of Proposition 1.

Let us now give an explicit formula for one-dimensional BSDEs under general integrability conditions on the driver.

**Lemma 2.** *Let  $p > 1$  and  $\eta \in L^p_{\mathcal{F}_T}(\Omega; \mathbb{R})$ . Suppose that  $(a_t)_{t \geq 0}$ ,  $(b_t)_{t \geq 0}$  are bounded  $\mathbb{R}$ -valued and  $\mathbb{R}^{d_1}$ -valued processes, respectively, and  $c \in L^p(\Omega; L^1(0, T; \mathbb{R}))$ , i.e.*

$$\mathbb{E} \left( \int_0^T |c_s| ds \right)^p < +\infty.$$

*Then the BSDE*

$$-dY_t = (a_t Y_t + b_t Z_t + c_t) dt - Z_t dW_t, \quad Y_T = \eta, \quad (2.6)$$

*admits a unique solution  $(Y, Z) \in \mathcal{K}_p$ . The process  $Y$  can be written as*

$$Y_t = \Gamma_t^{-1} \mathbb{E}^{\mathcal{F}_t} \left[ \Gamma_T \eta + \int_t^T \Gamma_s c_s ds, \right] \quad (2.7)$$

where  $\Gamma$  is given by the formula

$$\Gamma_t = \exp \left[ \int_0^t \left( a_s - \frac{1}{2} |b_s|^2 \right) ds + \int_0^t b_s dW_s \right]. \quad (2.8)$$

Moreover, by setting

$$V_t := \int_0^t |a_s| ds + \frac{1}{1 \wedge (p-1)} \int_0^t |b_s|^2 ds,$$

there exists  $C = C(p) > 0$  such that

$$\mathbb{E} \sup_{t \in [0, T]} |e^{V_t} Y_t|^p + \mathbb{E} \left( \int_0^T e^{2V_r} |Z_r|^2 dr \right)^{p/2} \leq C \mathbb{E} |e^{V_T} \eta|^p + C \mathbb{E} \left( \int_0^T e^{V_r} |c_r| dr \right)^p. \quad (2.9)$$

*Proof.* A proof of this result can be easily derived from (Pardoux and Răşcanu, 2014, Prop. 5.31), in which more general growth conditions on the coefficients are taken into account.  $\square$

### 3. Setting of the problem and Assumptions

In this section we firstly show how PDEs of the form of (1.1) naturally arise in connection with path-dependent stochastic dynamics. Even if path-dependent calculus remains our main motivation, the method we develop here applies to a wide class of equations that do not necessarily originate from path-dependent problems (see the discussion at the end of subsection 3.1). We subsequently introduce the assumptions under which the main results will be valid.

#### 3.1. The forward-backward system and the PDE

In what follows, for a given path  $\xi$  we will denote by  $\xi_t$  the value of  $\xi$  at time  $t$ , while we will use the notation  $\xi_{[0,t]}$  for the path of  $\xi$  up to time  $t$ , that is  $\xi_{[0,t]} = \{\xi(s)\}_{s \in [0,t]}$ . We will denote by  $C([0,t]; \mathbb{R}^d)$  and  $D([0,t]; \mathbb{R}^d)$  the space of  $\mathbb{R}^d$ -valued continuous and càdlàg functions, respectively, defined on the interval  $[0,t]$ .

Let us introduce the path-dependent SDE

$$\begin{cases} d\xi_s = b_s(\xi_{[0,s]}) ds + \sigma dW_s, & s \in [t, T], \\ \xi_{[0,t]} = \gamma \end{cases} \quad (1.2)$$

where  $\gamma \in D([0,t]; \mathbb{R}^d)$  is a given deterministic curve and the drift  $b$  is a family  $\{b_s\}_{s \in [0,T]}$ ,

$$b_s : D([0,s]; \mathbb{R}^d) \rightarrow \mathbb{R}^d. \quad (3.1)$$

A solution to (1.2) will be denoted by  $\xi^{\gamma,t}$ . Some authors define the drift equivalently as a map

$$b : [0, T] \times D([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}^d$$

that is *non-anticipative*:  $b(s, \chi) = b(s, \chi_{[0,s]})$  for every  $s \in [0, T]$ . In this setting, non-anticipativeness is assured requiring  $b$  to be measurable with respect to the  $\sigma$ -algebra induced on  $D([0, T]; \mathbb{R}^d)$  by the metric

$$d_\infty((s, \mu), (t, \chi)) = |s - t| + \sup_{r \in [0, T]} |\mu(r \wedge s) - \chi(r \wedge t)|.$$

We will prefer the first formulation (3.1) in what follow, but everything can be easily adapted to the second one, where the particular topology induced by the metric  $d_\infty$  has to be taken into account.

We now introduce the product space framework (see (Bensoussan et al., 2007, Chap. 4) for a general discussion), where the present state  $\xi_s$  and the past trajectory  $\xi_{[0,s]}$  are seen as separate variables. Setting

$$\begin{aligned} E_0 &:= \left\{ \varphi \in C([-T, 0]; \mathbb{R}^d) : \exists \lim_{r \uparrow 0} \varphi(r) \in \mathbb{R}^d \right\}, \\ D_0 &:= \left\{ \varphi \in D([-T, 0]; \mathbb{R}^d) : \exists \lim_{r \uparrow 0} \varphi(r) \in \mathbb{R}^d \right\}, \end{aligned}$$

we define the spaces

$$\begin{aligned} \mathcal{C} &:= \mathbb{R}^d \times E_0, \\ \widehat{\mathcal{C}} &:= \left\{ x = \begin{pmatrix} y \\ \varphi \end{pmatrix} \in \mathcal{C} \text{ s.t. } y = \lim_{r \uparrow 0} \varphi(r) \right\}, \\ \mathcal{D} &:= \mathbb{R}^d \times D_0, \\ \mathcal{L}^2 &:= \mathbb{R}^d \times L^2(-T, 0; \mathbb{R}^d). \end{aligned} \quad (3.2)$$

The spaces  $\mathcal{C}$ ,  $\widehat{\mathcal{C}}$  and  $\mathcal{D}$  are Banach spaces with respect to the norm  $\| \begin{pmatrix} y \\ \varphi \end{pmatrix} \|^2 = |y|^2 + \|\varphi\|_\infty^2$ , while  $\mathcal{L}^2$  is a Banach space with respect to the norm  $\| \begin{pmatrix} y \\ \varphi \end{pmatrix} \|^2 = |y|^2 + \|\varphi\|_2^2$ . The space  $\mathcal{D}$ , endowed with the topology given by the norm above, is not separable, but this will not undermine the methods used herein.

With these norms we have the natural inclusions

$$\widehat{\mathcal{C}} \subset \mathcal{C} \subset \mathcal{D} \subset \mathcal{L}^2$$

with continuous embeddings. We remark that  $\widehat{\mathcal{C}}$ ,  $\mathcal{C}$  and  $\mathcal{D}$  are dense in  $\mathcal{L}^2$  while neither  $\widehat{\mathcal{C}}$  nor  $\mathcal{C}$  are dense in  $\mathcal{D}$ . The choice of the interval  $[-T, 0]$  is made in accordance with most of the classical literature on delay equations. Note also that the space  $\widehat{\mathcal{C}}$  does not have the structure of a product space, and it is isomorphic



to the space  $C([-T, 0]; \mathbb{R}^d)$ .

The reformulation of equation (1.2) in infinite dimensions is obtained through the family of *restriction operators*

$$\begin{aligned} M_t : \mathcal{D} &\longrightarrow D([0, t]; \mathbb{R}^d) \\ M_t \left( \begin{pmatrix} y \\ \varphi \end{pmatrix} \right) &= \varphi(s-t) \mathbf{1}_{[0, t)}(s) + y \mathbf{1}_{\{t\}}(s) . \end{aligned} \quad (3.3)$$

Using  $M_t$  we can define the operator

$$\begin{aligned} B : [0, T] \times \mathcal{D} &\longrightarrow \mathcal{D} \\ B \left( t, \begin{pmatrix} y \\ \varphi \end{pmatrix} \right) &= \begin{pmatrix} b_t(M_t \left( \begin{pmatrix} y \\ \varphi \end{pmatrix} \right)) \\ 0 \end{pmatrix} . \end{aligned} \quad (3.4)$$

Note that the variable  $t$  appears explicitly in  $B$  even if  $b$  does not depend explicitly on time; such a variable acts here as a selector for  $b_t$  and  $M_t$ . The right inverse of  $M_t$  is the *backward extension operators* defined as

$$\begin{aligned} L^t : D([0, t]; \mathbb{R}^d) &\longrightarrow \mathcal{D} \\ L^t(\chi) &= \begin{pmatrix} \chi(t) \\ \chi(0) \mathbf{1}_{[-T, -t)} + \chi(t + \cdot) \mathbf{1}_{[-t, 0)} \end{pmatrix} ; \end{aligned} \quad (3.5)$$

with these definitions we have that  $M_t L^t \gamma = \gamma$  for every  $\gamma \in D([0, t]; \mathbb{R}^d)$ .

Finally let us introduce the operator

$$\begin{aligned} \text{Dom}(A) &= \left\{ \begin{pmatrix} y \\ \varphi \end{pmatrix} \in \mathcal{L}^2 : \varphi \in W^{1,2}(-T, 0; \mathbb{R}^d), y = \lim_{r \rightarrow 0^-} \varphi(r) \right\} , \\ A &= \begin{pmatrix} 0 & 0 \\ 0 & \frac{d}{dr} \end{pmatrix} , \end{aligned}$$

(we identify an element of  $W^{1,2}(-T, 0; \mathbb{R}^d)$  with its continuous version restricted to  $[-T, 0)$ ) and the space

$$\widehat{\mathcal{C}}^1 := A^{-1}(\{0\} \times E_0) = \left\{ \begin{pmatrix} y \\ \varphi \end{pmatrix} \in \widehat{\mathcal{C}} : \varphi \in C^1([-T, 0); \mathbb{R}^d), \exists \lim_{r \uparrow 0} \varphi'(r) \right\} .$$

The operator  $A$  generates a strongly continuous semigroup  $e^{tA}$  in  $\mathcal{L}^2$  which is explicitly given by the formula

$$e^{tA} \begin{pmatrix} y \\ \varphi \end{pmatrix} = \begin{pmatrix} y \\ \varphi(\cdot + t) \mathbf{1}_{[-T, -t)} + y \mathbf{1}_{[-t, 0)} \end{pmatrix} \quad (3.6)$$

(see Bensoussan et al. (2007) or Yosida (1995) for details). It is evident that such a semigroup is well defined on  $\mathcal{C}$  and  $\mathcal{D}$ , maps  $\mathcal{D}$  into itself, but it is not strongly continuous neither in  $\mathcal{D}$  nor in  $\mathcal{C}$ . Nevertheless it is equibounded in  $\mathcal{D}$ , it maps  $\widehat{\mathcal{C}}$  in itself and it is strongly continuous in  $\widehat{\mathcal{C}}$ .

Consider now a strong solution  $\xi = \xi^{\gamma, t}$  to equation (1.2) and set

$$X_s := L^s \xi_{[0, s]} .$$

$X$  is a  $\mathcal{D}$ -valued process that solves the SDE

$$\begin{cases} dX_s = AX_s ds + B(s, X_s) ds + \Sigma dW_s , & s \in (t, T) \\ X_t = L^t \gamma =: x , \end{cases} \quad (1.3)$$

in mild sense, that is, it satisfies

$$X_s = e^{sA} x + \int_t^s e^{(s-r)A} B(r, X_r) dr + \int_t^s e^{(s-r)A} \Sigma dW_s , \quad s \in [t, T] , \quad (3.7)$$

where  $\Sigma : \mathbb{R}^{d_1} \rightarrow \mathcal{D}$  is the operator given by

$$\Sigma w = \begin{pmatrix} \sigma w \\ 0 \end{pmatrix} \quad (3.8)$$

and

$$\int_t^s e^{(s-r)A} \Sigma dW_s = \int_t^s e^{(s-r)A} \begin{pmatrix} \sigma dW_r \\ 0 \end{pmatrix} = \begin{pmatrix} \int_t^s \sigma dW_r \\ \int_t^s \mathbf{1}_{[-(s-r), 0]}(\cdot) \sigma dW_r \end{pmatrix} = \begin{pmatrix} \sigma (W_s - W_t) \\ \sigma (W_{(s+\cdot) \vee t} - W_t) \end{pmatrix} . \quad (3.9)$$

Conversely, if  $X$  solves (3.7), its first component  $X^1$  solves equation (1.2) and  $X_s^1 = M_s X_s$  for every  $s \in [t, T]$ . In the following we will study the forward equation both in  $\mathcal{D}$  and in  $\mathcal{L}^2$ ; this means that we will consider drift operators  $B$  defined on  $[0, T] \times \mathcal{L}^2$  or on  $[0, T] \times \mathcal{D}$ , depending on the occasion. When needed, the solution to (1.3) will be denoted by  $X^{t, x}$  to stress the dependence on initial data.

**Remark 3.** If we denote by  $Z^t$  the stochastic convolution

$$Z^t(s) = \int_t^s e^{(s-r)A} \Sigma dW_r, \quad s \geq t,$$

then  $s \mapsto Z^t(s)$  is a continuous process with values in  $\widehat{\mathcal{C}}$  for every  $t \in [0, T]$  and  $\mathbb{E} \left[ \|Z^t(s)\|_{\widehat{\mathcal{C}}}^p \right] \lesssim (s-t)^{\frac{p}{2}}$  for every  $p \geq 2$ . Thanks to the continuity of the embedding  $\widehat{\mathcal{C}} \subset \mathcal{L}^2$ , the same properties hold in  $\mathcal{L}^2$  as well. From the explicit form of the semigroup it can be easily seen that  $X_s^{t,x}$  belongs to  $\widehat{\mathcal{C}}$  whenever  $x \in \widehat{\mathcal{C}}$ , whereas it only belongs to  $\mathcal{D}$  if the path  $x \in \mathcal{D}$  is discontinuous at some point. We refer to Flandoli and Zanco (2016) for a detailed discussion.

In Flandoli and Zanco (2016) it was shown that, under some regularity assumptions, the function

$$u(t, x) = \mathbb{E} \left[ \Phi \left( X_T^{t,x} \right) \right]$$

is a regular solution of the linear Kolmogorov backward equation

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) + Du(t, x) [Ax + B(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 u(t, x)] = 0, \\ u(T, \cdot) = \Phi \end{cases} \quad (3.10)$$

where  $\Phi : \mathcal{D} \rightarrow \mathbb{R}$  is a given terminal condition,  $Du(t, x) [Ax + B(t, x)]$  is the duality pairing between  $\mathcal{D}'$  and  $\mathcal{D}$  and the trace term is defined as

$$\text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 v(t, x)] = \sum_{j=1}^d \Sigma \Sigma^* D^2 v(t, x) (e_j, e_j)$$

for an arbitrary orthonormal system  $\{e_j\}_{j=1}^d$  of  $\mathbb{R}^d$ .

Here we are interested in the nonlinear version of (3.10) given by

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) + Du(t, x) [Ax + B(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 u(t, x)] = G(t, x, u(t, x), Du(t, x) \Sigma) \\ u(T, \cdot) = \Phi, \end{cases} \quad (1.1)$$

where  $G : [0, T] \times \mathcal{D} \times \mathbb{R} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}$ .

**Definition 4.** Given  $\Phi \in C(\mathcal{D}, \mathbb{R})$ , we say that  $u : [0, T] \times \mathcal{D} \rightarrow \mathbb{R}$  is a classical solution of the Kolmogorov semilinear backward equation with terminal condition  $\Phi$  if

$$u \in C^{1;2}([0, T] \times \mathcal{D}, \mathbb{R}),$$

and satisfies identity (1.1) for every  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}^1$ .

To find a classical solution to the semilinear Kolmogorov backward equation we introduce the following real-valued BSDE:

$$Y_s^{t,x} + \int_s^T Z_r^{t,x} dW_r = - \int_s^T G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr + \Phi(X_T^{t,x}), \quad t \leq s \leq T, \quad (3.11)$$

where the notation  $(\cdot)^{t,x}$  refers to the initial data  $(t, x)$  of the forward equation. In a differential formulation, we are concerned with the forward-backward system of the form

$$\begin{cases} dX_s^{t,x} = [AX_s^{t,x} + B(s, X_s^{t,x})] ds + \Sigma dW_s \\ dY_s^{t,x} = G(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x}) ds + Z_s^{t,x} dW_s \\ X_t^{t,x} = x \\ Y_T^{t,x} = \Phi(X_T^{t,x}) \end{cases} \quad (3.12)$$

where  $s \in [t, T] \subset [0, T]$ . Our goal is to show that the function

$$u(t, x) = Y_t^{t,x},$$

is a classical solution of the semilinear Kolmogorov backward equation with terminal condition  $\Phi$ . Since the scheme we follow consists in first solving the PDE in  $\mathcal{L}^2$  and then passing to  $\mathcal{D}$ , we will need to study the forward-backward system and the PDE in both these spaces. For this reason we will state some results and assumptions in a general separable Banach space  $E$ , specializing to the cases  $E = \mathcal{L}^2, \mathcal{L}^p, \mathcal{D}, \mathcal{C}, \widehat{\mathcal{C}}$  when needed.

**Remark 5** (Path-dependent case). *When  $G$  and  $\Phi$  are infinite-dimensional lifting of path-dependent functions, i.e.*

$$G(s, x, y, z) = g_s(M_s x, y, z) \quad \text{and} \quad \Phi(x) = \varphi(M_T x)$$

for a family  $\{g_s\}_{s \in [0, T]}$ ,  $g_s : D([0, s]; \mathbb{R}^d) \times \mathbb{R} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}$  and a map  $\varphi : D([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}$  (cf. (3.3)), the PDE (1.1) can be interpreted as the Kolmogorov PDE associated to the path-dependent forward-backward system

$$\begin{cases} d\xi_s = b_s(\xi_{[0, s]}) ds + \sigma dW_s \\ d\psi_s = g_s(\xi_{[0, s]}, \psi_s, \zeta_s) ds + \zeta_s dW_s \\ \xi_{[0, t]} = \gamma_{[0, t]} \\ \psi_T = \varphi(\xi_{[0, T]}) \end{cases} .$$

In this specific situation the PDE actually has the form

$$\frac{\partial u}{\partial t}(t, x) + Du(t, x)Ax = G(t, x, u(t, x), Du(t, x)\Sigma) - Du(t, x)B(t, x) - \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 u^*], \quad (3.13)$$

where the r.h.s depends on  $Du(t, x) \in \mathcal{D}'$  only through its action on the first components of elements in  $\mathcal{D}$ . Moreover, exploiting the so-called functional differential calculus introduced in Cont and Fournié (2013) one can formulate a PDE very similar to (3.13) for which a wellposedness result can also be provided by our approach. We refer again to Flandoli and Zanco (2016) and Zanco (2015) for a detailed discussion about the relations between the two settings and the role played by the operator  $\frac{\partial}{\partial t} + D[\cdot]A$ .

In the following, we essentially provide a solution theory for semilinear PDEs on the space of continuous functions under the assumption that the second order term concentrates on the final dimensional component of  $\mathcal{D}$  (thus ensuring the trace term be well defined) and without requiring that coefficients arise as liftings of path-dependent functions. For these reasons, in the whole presentation we will consider general coefficients, without sticking to the path-dependent formalism.

### 3.2. Assumptions

Let  $E$  be any of the spaces listed in (3.2) and let  $m \geq 0$ . The following sets of assumptions are in force throughout the paper.

**Assumption 1.** *The drift term  $B$  belongs to  $C^{1;2, \alpha}([0, T] \times E; E)$  for some  $\alpha \in (0, 1)$ , with the Hölder norm of  $D^2 B$  bounded uniformly in  $s \in [0, T]$ . Furthermore, there exists a constant  $C \geq 0$  such that*

$$(B.I) \quad |B(s, x)| \leq C(1 + |x|),$$

$$(B.II) \quad |B(s, x_1) - B(s, x_2)| \leq C|x_1 - x_2|,$$

$$(B.III) \quad |D^2 B(s, x)| \leq C(1 + |x|^m),$$

for every  $x, x_1, x_2 \in E$ , uniformly in  $s \in [0, T]$ .

For what concerns the coefficients of the BSDE (1.5), hereinafter we use the notation  $D_i G$  to denote the derivative with respect to the  $i$ -th (spatial) entry of the map  $(x, y, z) \mapsto G(s, x, y, z)$ , and  $D_{i,j}^2 G$  for the second derivatives

**Assumption 2.**  *$G : [0, T] \times E \times \mathbb{R} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}$  is such that for every  $s \in [0, T]$  the map  $G(s, \cdot) \in C^{2, \alpha}(E \times \mathbb{R} \times \mathbb{R}^{d_1}; \mathbb{R})$ . Moreover there exists  $C \geq 0$  such that :*

$$(G.I) \quad |G(s, x, y, z)| \leq C(1 + |x|^m + |y| + |z|);$$

$$(G.II) \quad |G(s, x, y_1, z_1) - G(s, x, y_2, z_2)| \leq C(|y_1 - y_2| + |z_1 - z_2|);$$

$$(G.III) \quad |D_1 G(s, x, y, z)| + |D_{1,1}^2 G(s, x, y, z)| \leq C(1 + |x|^m)(1 + |y| + |z|);$$

$$(G.IV) \quad |D_{1,2}^2 G(s, x, y, z)| + |D_{1,3}^2 G(s, x, y, z)| \leq C(1 + |x|^m)(1 + |y|);$$

$$(G.V) \quad |D_{2,2}^2 G(s, x, y, z)| + |D_{2,3}^2 G(s, x, y, z)| \leq C(1 + |x|^m);$$

$$(G.VI) \quad |D_{3,3}^2 G(s, x, y, z)| \leq C,$$

for every  $x, y, z, y_1, y_2, z_1, z_2 \in E$ , uniformly in  $s \in [0, T]$ .

**Assumption 3.** *The function  $\Phi$  belongs to  $C^{2, \alpha}(E, \mathbb{R})$  for some  $\alpha \in (0, 1)$  and*

$$|\Phi(x)| + |D\Phi(x)| + |D^2\Phi(x)| \leq C(1 + |x|^m).$$

In Section 6, when passing from the PDE in  $\mathcal{L}^2$  to the PDE in  $\mathcal{D}$  we will need to carefully approximate the coefficients. In doing so, it is crucial that if  $B, G, \Phi$  satisfy the above assumptions in  $\mathcal{D}$  then the same could hold for the approximations  $B^n, G^n, \Phi^n$  in  $\mathcal{L}^2$ , possibly with some uniformity with respect to  $n$ . It turns out that a convenient way to build such approximations is to consider a sequence of bounded linear operators  $J^n$  from  $\mathcal{L}^2$  to  $\mathcal{C}$  with the following properties:

- $J^n x \rightarrow x$  in  $\mathcal{C}$  for every  $x \in \mathcal{C}$ ;
- $\sup_n \|J^n x\|_\infty \leq C_J \|x\|_\infty$  for every  $x \in \mathcal{D}$  such that  $M_T(x)$  has at most one jump and is continuous elsewhere, where  $M_T$  is defined in (3.3).

Note that any such sequence converges to the identity uniformly on compact sets of  $\mathcal{C}$ .

An example of  $\{J_n\}$  can be constructed as follows: given any  $\varepsilon \in (0, \frac{T}{2})$  define a function  $\tau_\varepsilon : [-T, 0] \rightarrow [-T, 0]$  as

$$\tau_\varepsilon(x) = \begin{cases} -T + \varepsilon & \text{if } x \in [-T, -T + \varepsilon] \\ x & \text{if } x \in [-T + \varepsilon, -\varepsilon] \\ -\varepsilon & \text{if } x \in [-\varepsilon, 0]. \end{cases}$$

Then choose any function  $\rho \in C^\infty(\mathbb{R}; \mathbb{R})$  such that  $\|\rho\|_1 = 1$ ,  $0 \leq \rho \leq 1$ ,  $\text{supp}(\rho) \subseteq [-1, 1]$  and define a sequence  $\{\rho_n\}$  of mollifiers by  $\rho_n(x) := n\rho(nx)$ . Set, for any  $\varphi \in L^1(-T, 0; \mathbb{R}^d)$

$$\mathcal{J}^n \varphi(x) := \int_{-T}^0 \rho_n(\tau_{\frac{1}{n}}(x) - y) \varphi(y) dy ; \quad (3.14)$$

finally set

$$J^n \begin{pmatrix} a \\ \varphi \end{pmatrix} = \begin{pmatrix} a \\ \mathcal{J}^n \varphi \end{pmatrix} .$$

To ensure the applicability of the limiting procedure, we need one more assumption, that is satisfied by many examples as discussed in Flandoli and Zanco (2016) and Zanco (2015).

**Definition 6.** Let  $F$  be a Banach space,  $R: \mathcal{D} \rightarrow F$  twice Fréchet differentiable and  $\Gamma \subseteq \mathcal{D}$ . We say that  $R$  has one-jump-continuous Fréchet differentials of first and second order on  $\Gamma$  if there exists a sequence of linear continuous operators  $J^n$  as above such that for every  $y \in \Gamma$  and for almost every  $a \in [-T, 0]$  the following hold:

$$\begin{aligned} DR(y) J^n \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} &\longrightarrow DR(y) \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} , \\ D^2 R(y) \left( J^n \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix}, \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} \right) &\longrightarrow 0 , \quad D^2 R(y) \left( \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix}, J^n \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} \right) \longrightarrow 0 , \\ D^2 R(y) \left( J^n \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix}, J^n \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} \right) &\longrightarrow 0 , \end{aligned}$$

where we adopt the convention that  $\begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  when  $a = 0$ .

We will call *smoothing sequence* for  $R$  any sequence  $\{J^n\}$  satisfying the above requirements. By linearity, the above convergences hold true also if  $\begin{pmatrix} 1 \\ \mathbf{1}_{[a,0]} \end{pmatrix}$  is substituted with any  $x \in \mathcal{D}$  with the property that  $M_T(x)$  has at most one jump and it is continuous elsewhere.

**Assumption 4.** For every  $s \in [0, T]$ ,  $B(s, \cdot)$  and  $\Phi$  have one-jump-continuous Fréchet differentials of first and second order on  $\widehat{\mathcal{C}} \subset \mathcal{D}$  and the smoothing sequence of  $B$  does not depend on  $s$ .

**Assumption 5.** For every  $s \in [0, T]$ ,  $y \in \mathbb{R}$ ,  $z \in \mathbb{R}^{d_1}$ ,  $G(s, \cdot, y, z)$  has one-jump continuous Fréchet differential of first order and its smoothing sequence does not depend on  $s$  nor on  $y, z$ .

#### 4. The forward-Backward system

This section is devoted to the forward-backward system (3.12) (FBSDE in the following), that we write below in mild formulation for the reader's convenience:

$$\begin{cases} X_s^{t,x} = e^{(s-t)A} x_{[0,t]} + \int_t^s e^{(s-r)A} B(r, X_r^{t,x}) dr + \int_t^s e^{(s-r)A} \Sigma dW_r \\ Y_s^{t,x} + \int_s^T Z_r^{t,x} dW_r = - \int_s^T G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr + \Phi(X_T^{t,x}) , \end{cases} \quad (4.1)$$

where  $t \leq s \leq T$ . Observe that the system is not fully coupled: the forward equation does not depend on the values of the pair  $(Y, Z)$ . We firstly state some result for the process  $X$ , whose proof can be found in (Flandoli and Zanco, 2016, Thms. 2.2, 2.3, 2.4)

**Proposition 7.** *Under Assumption 1, there exists a set  $\Omega_0 \subseteq \Omega$  of full probability such that:*

- (i) (existence) *for every initial data  $(t, x) \in [0, T] \times E$  and every  $\omega \in \Omega_0$ , equation (3.7) admits a unique solution  $(s, \omega) \mapsto X_s^{t,x}(\omega) \in E$  which is continuous in time if  $E = \mathcal{L}^2$ , while it is only bounded in time if  $E = \mathcal{D}$ ;*
- (ii) (regularity in space) *for every  $\omega \in \Omega_0$ ,  $t \in [0, T]$  and  $s \in [t, T]$  the map  $x \mapsto X_s^{t,x}(\omega)$  is in  $C^{2,\alpha}$ ;*
- (iii) (regularity in time) *if  $E = \mathcal{L}^2$ , for every  $s \in [0, T]$ ,  $x \in E$  and  $\omega \in \Omega_0$  the map  $t \mapsto X_s^{t,x}(\omega)$  ( $t \leq s$ ) is continuous; if  $E = \mathcal{D}$  the same property holds whenever  $x \in \widehat{\mathcal{C}}$ ;*
- (iv) (Markovianity) *if  $E = \mathcal{L}^2$  the solution  $X^{t,x}$  has the markov property.*

From now on we will denote by  $\Omega_0 \subseteq \Omega$  the fixed set given by Proposition 7.

**Theorem 8.** *Assume that  $B : [0, T] \times E \rightarrow E$  satisfies Assumption 1. Fix a time  $t \in [0, T]$  and a  $\mathcal{F}_t$ -measurable  $E$ -valued random variable  $\xi$ , and let  $X^{t,\xi}$  be the unique  $E$ -valued solution to*

$$X_s = e^{(s-t)A}\xi + \int_t^s e^{(s-r)A}B(r, X_r) dr + \int_t^s e^{(s-r)A}\Sigma dW_r. \quad (4.2)$$

For any  $p \geq 1$ , if  $\xi$  has finite  $p$ -th moment then  $X^{t,\xi} \in L^p(\Omega; C([t, T]; E))$  and

$$\mathbb{E} \sup_{s \in [t, T]} |X_s|_E^p \leq c_1 (1 + \mathbb{E} |\xi|^p) \quad (4.3)$$

When  $\xi = x \in E$  is deterministic, for every  $t \in [0, T]$  the map  $x \mapsto X^{t,x}$  is twice Fréchet differentiable as a map from  $E$  to  $L^p(\Omega; C([t, T]; E))$  with continuous differentials; the  $L(E; E)$ -valued process  $D_x X^{t,x}$  is the unique solution to

$$\Xi_s = e^{(s-t)A} + \int_t^s e^{(s-r)A}DB(r, X_r^{t,x})\Xi_r dr, \quad (4.4)$$

while the  $L(E, E; E)$ -valued process  $D_x^2 X^{t,x}$  is the unique solution to

$$\Theta_s = \int_t^s e^{(s-r)A}D^2B(r, X_r^{t,x})(D_x X_r^{t,x}, D_x X_r^{t,x}) dr + \int_t^s e^{(s-r)A}DB(r, X_r^{t,x})\Theta_r dr. \quad (4.5)$$

All the three SDEs above can be solved path-by-path, meaning that for any fixed  $\omega \in \Omega_0$  there exist unique functions  $s \mapsto X_s^{t,\xi}(\omega)$ ,  $s \mapsto D_x X_s^{t,x}(\omega)$  and  $s \mapsto D_x^2 X_s^{t,x}(\omega)$  that satisfy (4.2), (4.4) and (4.5), respectively. Moreover

$$\sup_{s \in [t, T]} \|D_x X_s^{t,x}\|_{L(E; E)} \leq c_2 \quad \text{for a.e. } \omega \in \Omega \quad (4.6)$$

and in particular for any  $E$ -valued random variable  $\eta \in L^p(\Omega; E)$

$$\mathbb{E} \sup_{s \in [t, T]} |D_x X_s^{t,x} \eta|^p \leq c_2 \mathbb{E} |\eta|^p. \quad (4.7)$$

Furthermore

$$\sup_{s \in [t, T]} \|D_x^2 X_s^{t,x}\|_{L(E, E; E)} \leq c_3 (1 + |x|^m) \quad \text{for a.e. } \omega \in \Omega. \quad (4.8)$$

The constants  $c_1, c_2, c_3$  in the inequalities above depend only on  $m, T, D^i B$  with  $i = 0, 1, 2$ , and on the constant  $C$  in Assumption 1.

*Proof.* Using that  $\mathcal{Z}_s^t$  is a  $E$ -valued martingale, by (Da Prato and Zabczyk, 2014, Thm. 3.9) and Remark 3, we have that

$$\mathbb{E} \sup_{s \in [t, T]} |\mathcal{Z}_s^t|_E^p \leq C \sup_{s \in [t, T]} \mathbb{E} |\mathcal{Z}_s^t|_E^p \leq CT^{\frac{p}{2}}.$$

Therefore, from the uniform estimates on  $e^{tA}$  and Assumption 1 we get that for every  $t \leq R \leq T$

$$\mathbb{E} \sup_{s \in [t, R]} |X_s|^p \lesssim 1 + \mathbb{E} |\xi|^p + \int_t^R \mathbb{E} \left( \sup_{s \in [t, r]} |X_s|^p \right) dr,$$

from which (4.3) follows thanks to Gronwall's lemma. Furthermore, the proof of the Fréchet differentiability of the map  $x \mapsto X_s^{t,x}(\omega)$  given in Flandoli and Zanco (2016) can be easily extended to the required differentiability of  $x \mapsto X^{t,x}$  in the space of  $E$ -valued processes. Well-posedness of (4.4) and (4.5) (and the fact that  $D_x X^{t,x}$  and  $D_x^2 X^{t,x}$  are the required solutions) has been already established in Flandoli and Zanco (2016). Estimates (4.6), (4.7) and (4.8) are then easy consequences of Assumption 1.  $\square$

For what concerns the Backward SDE in (4.1), the following wellposedness result has been given in Fuhrman and Tessitore (2002).

**Proposition 9.** *Under Assumptions 1, 2 and 3, for every  $(t, x) \in [0, T] \times E$ , the BSDE in (4.1) admits a unique solution  $(Y, Z) \in \mathcal{K}_p$ , for every  $p \in [2, +\infty)$ . Moreover, the map  $(t, x) \mapsto (Y^{t,x}, Z^{t,x})$  belongs to  $C([0, T] \times E; \mathcal{K}_p)$  and there exists  $c \geq 0$  such that*

$$\mathbb{E} \sup_{s \in [t, T]} |Y_s^{t,x}|^p + \mathbb{E} \left( \int_t^T |Z_r^{t,x}|^2 dr \right)^{p/2} \leq c(1 + |x|^{pm}). \quad (4.9)$$

**Remark 10.** *The constant  $c \geq 0$  appearing in (4.9) can be chosen independently of  $(t, x)$ . The same applies to Propositions 11, 12 and 15 below. Alternatively, one could set  $(Y_r, Z_r) = 0$  for every  $r \in [0, t]$ .*

#### 4.1. First-order differentiability of the BSDE

Here we investigate the differentiability of the map  $x \mapsto (D_x Y^{t,x}, D_x Z^{t,x})$ . Gâteaux differentiability has been established in a Hilbert setting in Fuhrman and Tessitore (2002) and then extended to a general Banach setting in Masiero (2008) and Masiero and Richou (2014). Our aim is to show that under Assumptions 2 and 3, also Fréchet differentiability takes place.

Let us firstly lighten the notation introducing the shorthand

$$D_i G_r(t, x) := D_i G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}), \quad i = 1, 2, 3,$$

and consider the backward equation satisfied by the pair  $(U^{t,x}, V^{t,x})$ :

$$\begin{aligned} U_s^{t,x} h + \int_s^T V_r^{t,x} h dW_r &= U_T^{t,x} h - \int_s^T D_1 G_r(t, x) D_x X_r^{t,x} h dr \\ &\quad - \int_s^T (D_2 G_r(t, x) U_r^{t,x} h + D_3 G_r(t, x) V_r^{t,x} h) dr, \end{aligned} \quad (4.10)$$

where the terminal condition is given by  $U_T^{t,x} h = D\Phi(X_T^{t,x}) D_x X_T^{t,x} h$ . It turns out that (4.10) admits a unique solution  $(U^{t,x}, V^{t,x})$  which is given by the directional derivatives  $(D_x Y^{t,x} h, D_x Z^{t,x} h)$ , for every  $h \in E$ . This is the content of the next Proposition, whose proof can be found in (Fuhrman and Tessitore, 2002, Prop. 4.8).

**Proposition 11.** *Let Assumptions 1, 2 and 3 hold true. For every  $h \in E$  equation (4.10) admits a unique solution  $(U^{t,x} h, V^{t,x} h) = (D_x Y^{t,x} h, D_x Z^{t,x} h)$ . Moreover, for every  $p > 1$  the map  $(t, x) \mapsto (Y^{t,x}, Z^{t,x})$  is Gâteaux differentiable as a map from  $[0, T] \times E$  to  $\mathcal{K}_p$  and for every  $h \in E$  the directional derivatives  $(D_x Y^{t,x} h, D_x Z^{t,x} h)$  satisfy the BSDE (4.10):*

$$\begin{aligned} D_x Y_s^{t,x} h + \int_s^T D_x Z_r^{t,x} h dW_r &= D\Phi(X_T^{t,x}) D_x X_T^{t,x} h - \int_s^T D_1 G_r(t, x) D_x X_r^{t,x} h dr \\ &\quad - \int_s^T (D_2 G_r(t, x) D_x Y_r^{t,x} h + D_3 G_r(t, x) D_x Z_r^{t,x} h) dr. \end{aligned} \quad (4.11)$$

Finally, for every  $(t, x) \in [0, T] \times E$ , the following estimate holds true

$$\left[ \mathbb{E} \sup_{s \in [t, T]} |D_x Y_s^{t,x} h|^p \right]^{1/p} + \left[ \mathbb{E} \left( \int_t^T |D_x Z_r^{t,x} h|^2 dr \right)^{p/2} \right]^{1/p} \leq C|h| \left( 1 + |x|^{m^2} \right). \quad (4.12)$$

We are now in position to study the Fréchet differentiability of the maps  $t, x \mapsto (D_x Y^{t,x}, D_x Z^{t,x})$ .

**Proposition 12.** *Under Assumptions 1, 2 and 3, the map  $x \mapsto (Y^{t,x}, Z^{t,x})$  (resp.  $t \mapsto (Y^{t,x}, Z^{t,x})$ ) is Fréchet differentiable as a map from  $E$  (resp.  $[0, T]$ ) to  $\mathcal{K}_p$ . Moreover the following estimate holds true*

$$\left[ \mathbb{E} \sup_{s \in [t, T]} \|D_x Y_s^{t,x}\|^p \right]^{1/p} + \left[ \mathbb{E} \left( \int_t^T \|D_x Z_r^{t,x}\|^2 dr \right)^{p/2} \right]^{1/p} \leq C \left( 1 + |x|^{m^2} \right). \quad (4.13)$$

Before entering the details of the proof let us briefly comment on the crucial role played by estimate (2.9). In taking the differences  $(D_x Y^{t,x} - D_x Y^{t,y}, D_x Z^{t,x} - D_x Z^{t,y})$  (hence comparing solutions whose forward process starts at different points), we inevitably end up with the term

$$\int_s^T [D_3 G_r(t, x) - D_3 G_r(t, y)] D_x Z_r^{t,y} h \, dr,$$

leading to the product  $(Z_r^{t,x} - Z_r^{t,y}) D_x Z_r^{t,y} h$  which does not belong to  $L^p(\Omega; L^2(0, T; \mathbb{R}))$ . In this situation standard methods are not effective. Nonetheless, the minimal integrability requirement in Lemma 2 allows to treat with simple tools (see estimate (4.16)) the following term

$$\mathbb{E} \left( \int_s^T |Z_r^{t,x} - Z_r^{t,y}| |D_x Z_r^{t,y} h| \, dr \right)^p.$$

*Proof of Proposition 12.* To shorten the proof we concentrate only on the Fréchet differentiability of the map  $x \mapsto (Y^{t,x}, Z^{t,x})$ , for every  $t \in [0, T]$ . Differentiability in time follows by the very same technique (see e.g. (Masiero and Richou, 2014) for what concerns differentiability in the Gâteaux sense).

The strategy of the proof is as follows: by Theorem 11 we deduce that the pair  $(Y^{t,x}, Z^{t,x})$  is Gâteaux differentiable with respect to  $x$ . Then we show the continuity of  $x \mapsto (D_x Y^{t,x}, D_x Z^{t,x})$  as a map from  $E$  to  $L(E; \mathcal{K}_p)$ , which easily yields the required Fréchet differentiability. To do it, we write the equation for the differences  $D_x Y^{t,x} h - D_x Y^{t,y} h$ ,  $D_x Z^{t,x} h - D_x Z^{t,y} h$  emphasizing its linear character. We employ estimates (2.9) and we show that the r.h.s. vanishes as  $|x - y|_E \rightarrow 0$ , uniformly in  $h \in E$ ,  $|h|_E \leq 1$ .

Given  $x, y, h \in E$ , let us write the equation for the differences  $(D_x Y_s^{t,x} h - D_x Y_s^{t,y} h)$ ,  $(D_x Z_s^{t,x} h - D_x Z_s^{t,y} h)$ :

$$\begin{aligned} & [D_x Y_s^{t,x} - D_x Y_s^{t,y}] h + \int_s^T [D_x Z_r^{t,x} - D_x Z_r^{t,y}] h \, dW_r = [D\Phi(X_T^{t,x}) D_x X_T^{t,x} - D\Phi(X_T^{t,y}) D_x X_T^{t,y}] h \\ & - \int_s^T [D_1 G_r(t, x) D_x X_r^{t,x} - D_1 G_r(t, y) D_x X_r^{t,y}] h \, dr - \int_s^T [D_2 G_r(t, x) D_x Y_r^{t,x} - D_2 G_r(t, y) D_x Y_r^{t,y}] h \, dr \\ & - \int_s^T [D_3 G_r(t, x) D_x Z_r^{t,x} - D_3 G_r(t, y) D_x Z_r^{t,y}] h \, dr \\ & = [D\Phi(X_T^{t,x}) D_x X_T^{t,x} - D\Phi(X_T^{t,y}) D_x X_T^{t,y}] h - \int_s^T [D_1 G_r(t, x) D_x X_r^{t,x} - D_1 G_r(t, y) D_x X_r^{t,y}] h \, dr \\ & - \int_s^T D_2 G_r(t, x) [D_x Y_r^{t,x} - D_x Y_r^{t,y}] h \, dr - \int_s^T D_3 G_r(t, x) [D_x Z_r^{t,x} - D_x Z_r^{t,y}] h \, dr \\ & - \int_s^T [D_2 G_r(t, x) - D_2 G_r(t, y)] D_x Y_r^{t,y} h \, dr - \int_s^T [D_3 G_r(t, x) - D_3 G_r(t, y)] D_x Z_r^{t,y} h \, dr. \end{aligned}$$

If we define

$$\begin{aligned} \Delta Y_r &= (D_x Y_s^{t,x} - D_x Y_s^{t,y}) h, \quad \Delta Z_r = (D_x Z_r^{t,x} - D_x Z_r^{t,y}) h, \\ \xi &= [D\Phi(X_T^{t,x}) D_x X_T^{t,x} - D\Phi(X_T^{t,y}) D_x X_T^{t,y}] h, \\ \lambda(r) &= -\lambda_1(r) - \lambda_2(r) - \lambda_3(r) = -[D_1 G_r(t, x) D_x X_r^{t,x} - D_1 G_r(t, y) D_x X_r^{t,y}] h \\ &\quad - [D_2 G_r(t, x) - D_2 G_r(t, y)] D_x Y_r^{t,y} h - [D_3 G_r(t, x) - D_3 G_r(t, y)] D_x Z_r^{t,y} h, \\ V_s &= \int_t^s |D_2 G_r(t, x)| \, dr + \frac{1}{1 \wedge (p-1)} \int_t^s |D_3 G_r(t, x)|^2 \, dr, \end{aligned} \tag{4.14}$$

the above equation reads

$$\Delta Y_s + \int_s^T \Delta Z_r \, dW_r = \xi - \int_s^T (D_2 G_r(t, x) \Delta Y_r + D_3 G_r(t, x) \Delta Z_r - \lambda(r) \, dr) \, dr.$$

where  $D_2 G_r(t, x)$  and  $D_3 G_r(t, x)$  are bounded processes with values in  $\mathbb{R}$  and  $\mathbb{R}^{d_1}$ , respectively. Since  $V$  is a bounded process as well, estimate (2.9) in Lemma 2 guarantees that

$$\mathbb{E} \sup_{s \in [t, T]} |e^{V_s} \Delta Y_s|^p + \mathbb{E} \left( \int_t^T e^{2V_r} |\Delta Z_r|^2 \, dr \right)^{p/2} \lesssim \mathbb{E} |\xi|^p + \mathbb{E} \left( \int_t^T |\lambda(r)| \, dr \right)^p,$$

and the desired continuity follows as soon as

$$\sup_{\substack{h \in E, \\ |h|_E \leq 1}} \left[ \mathbb{E} |\xi|^p + \mathbb{E} \left( \int_t^T |\lambda(r)| \, dr \right)^p \right] \longrightarrow 0, \quad \text{if } |x - y|_E \rightarrow 0. \quad (4.15)$$

Let us start by showing the convergence for the first term in (4.15). For  $p > 1$

$$\begin{aligned} \mathbb{E} |\xi|^p &\lesssim \mathbb{E} |D\Phi(X_T^{t,x}) D_x X_T^{t,x} h - D\Phi(X_T^{t,y}) D_x X_T^{t,y} h|^p \\ &\lesssim \mathbb{E} |D\Phi(X_T^{t,x}) (D_x X_T^{t,x} h - D_x X_T^{t,y} h)|^p \\ &\quad + \mathbb{E} |(D\Phi(X_T^{t,x}) - D\Phi(X_T^{t,y})) D_x X_T^{t,y} h|^p \\ &= \mathbb{E} (|\xi_1|^p + |\xi_2|^p). \end{aligned}$$

Using Assumption 3 we have

$$|\xi_1| \leq \|D\Phi(X_T^{t,x})\| |D_x X_T^{t,x} h - D_x X_T^{t,y} h| \lesssim \left( 1 + \sup_{s \in [t, T]} |X_s^{t,x}|^m \right) |D_x X_T^{t,x} h - D_x X_T^{t,y} h|,$$

and thanks to estimates (4.3), (4.7) and the Fréchet differentiability of  $x \mapsto X_s^{t,x}(\omega)$  for every  $\omega \in \Omega_0$ ,  $s \in [t, T]$  (see Proposition 7) it holds

$$\begin{aligned} \sup_{\substack{h \in E, \\ |h|_E \leq 1}} \mathbb{E} |\xi_1|^p &\lesssim \sup_{\substack{h \in E, \\ |h|_E \leq 1}} \left[ \mathbb{E} \left( 1 + \sup_{s \in [t, T]} |X_s^{t,x}|^{2mp} \right) \right]^{1/2} \left[ \mathbb{E} \|D_x X_T^{t,x} - D_x X_T^{t,y}\|_{L(E;E)}^{2p} \right]^{1/2} |h|^p \\ &\lesssim \left[ \mathbb{E} \|D_x X_T^{t,x} - D_x X_T^{t,y}\|_{L(E;E)}^{2p} \right]^{1/2} \longrightarrow 0, \quad \text{if } |x - y|_E \rightarrow 0. \end{aligned}$$

A similar argument yields

$$\begin{aligned} \sup_{\substack{h \in E, \\ |h|_E \leq 1}} \mathbb{E} |\xi_2|^p &\leq \sup_{\substack{h \in E, \\ |h|_E \leq 1}} \left[ \mathbb{E} \|D\Phi(X_T^{t,x}) - D\Phi(X_T^{t,y})\|^{2p} \right]^{1/2} \left[ \|D_x X_T^{t,y}\|_{L(E;E)}^{2p} \right]^{1/2} |h|^p \\ &\lesssim \left[ \mathbb{E} \|D\Phi(X_T^{t,x}) - D\Phi(X_T^{t,y})\|^{2p} \right]^{1/2} \longrightarrow 0, \quad \text{if } |x - y|_E \rightarrow 0, \end{aligned}$$

where we employed Vitali convergence theorem. More precisely, Fréchet differentiability of the map  $x \mapsto X_T^{t,x}(\omega)$  for every  $\omega \in \Omega_0$ , along with the continuity of  $D\Phi$ , guarantees the convergence of  $\|D\Phi(X_T^{t,x}) - D\Phi(X_T^{t,y})\| \rightarrow 0$ ; whereas Assumption 3 combined with estimate (4.3) and the choice of a deterministic initial condition  $x \in \mathbb{E}$ , ensure the uniform integrability of  $\|D\Phi(X_T^{t,x}) - D\Phi(X_T^{t,y})\|^{2p}$ .

Concerning the second term in (4.15) we treat separately the three processes  $\lambda_i$  in (4.14). First we write

$$\begin{aligned} |\lambda_1(r)| &= |D_1 G_r(t, x) D_x X_r^{t,x} h - D_1 G_r(t, y) D_x X_r^{t,y} h| \\ &\lesssim |D_1 G_r(t, x) (D_x X_r^{t,x} h - D_x X_r^{t,y} h)| + |[D_1 G_r(t, x) - D_1 G_r(t, y)] D_x X_r^{t,y} h| \\ &= |\lambda_{11}(r)| + |\lambda_{12}(r)|. \end{aligned}$$

Assumption 2 ensures that

$$\begin{aligned} |\lambda_{11}(r)| &\leq \|D_1 G_r(t, x)\| |D_x X_r^{t,x} h - D_x X_r^{t,y} h| \\ &\lesssim \left( 1 + \sup_{r \in [t, T]} |X_r^{t,x}|^m \right) \left( 1 + \sup_{r \in [t, T]} |Y_r^{t,x}| + |Z_r^{t,x}| \right) |D_x X_r^{t,x} h - D_x X_r^{t,y} h|, \end{aligned}$$

and from estimates (4.3) and Proposition 9 we get

$$\mathbb{E} \left( \int_t^T |\lambda_{11}(r)| \, dr \right)^p \lesssim (1 + |x|^{4pm}) |h|^p \left[ \mathbb{E} \left( \int_t^T \|D_x X_r^{t,x} - D_x X_r^{t,y}\|_{L(E;E)}^2 \, dr \right)^p \right]^{1/2}$$

which converges to zero as  $|x - y|_E \rightarrow 0$  uniformly with respect to  $h$ ,  $\|h\| \leq 1$ , thanks to the Fréchet character of the map  $x \mapsto X^{t,x}$  and the Lebesgue dominated convergence theorem (recall estimate (4.6)).



Then we have

$$\begin{aligned}
|\lambda_{12}(r)| &\leq \left| [D_1G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) - D_1G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,y})] D_x X_r^{t,y} h \right| \\
&\quad + \left| [D_1G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,y}) - D_1G(r, X_r^{t,x}, Y_r^{t,y}, Z_r^{t,y})] D_x X_r^{t,y} h \right| \\
&\quad + \left| [D_1G(r, X_r^{t,x}, Y_r^{t,y}, Z_r^{t,y}) - D_1G(r, X_r^{t,y}, Y_r^{t,y}, Z_r^{t,y})] D_x X_r^{t,y} h \right| \\
&= |\lambda_{121}(r)| + |\lambda_{122}(r)| + |\lambda_{123}(r)|.
\end{aligned}$$

Assumption 2 yields

$$\begin{aligned}
|\lambda_{121}(r)| &= \left| \int_0^1 D_{1,3}^2 G(r, X_r^{t,x}, Y_r^{t,x}, \alpha Z_r^{t,x} + (1-\alpha)Z_r^{t,y}) (Z_r^{t,x} - Z_r^{t,y}, D_x X_r^{t,y} h) d\alpha \right| \\
&\lesssim \left( 1 + \sup_{r \in [t, T]} |X_r^{t,x}|^m \right) \left( 1 + \sup_{r \in [t, T]} |Y_r^{t,x}| \right) \left( \sup_{r \in [t, T]} \|D_x X_r^{t,y}\| \right) |h| |Z_r^{t,x} - Z_r^{t,y}|.
\end{aligned}$$

If we apply Holder inequality, Theorem 8 and Proposition 9 we get that, as  $|x - y|_E \rightarrow 0$ ,

$$\sup_{\substack{h \in E, \\ |h|_E \leq 1}} \mathbb{E} \left( \int_t^T |\lambda_{121}(r)| dr \right)^p \lesssim (1 + |x|^{mp}) (1 + |x|^{pm^2}) \left[ \mathbb{E} \left( \int_t^T |Z_r^{t,x} - Z_r^{t,y}|^2 dr \right)^{2p} \right]^{\frac{1}{4}} \rightarrow 0.$$

The same strategy also applies to the terms  $\lambda_{122}(r)$  and  $\lambda_{123}(r)$ . Therefore, uniformly in  $h \in E$ ,  $|h|_E \leq 1$ :

$$\mathbb{E} \left( \int_t^T \beta_1(r) dr \right)^p \leq \mathbb{E} \left( \int_t^T |\lambda_{11}(r)| dr \right)^p + \sum_{i=1}^3 \mathbb{E} \left( \int_t^T |\lambda_{12i}(r)| dr \right)^p \rightarrow 0, \quad \text{if } |x - y|_E \rightarrow 0.$$

We proceed in a similar way for the term  $\lambda_2$ :

$$\begin{aligned}
|\lambda_2(r)| &\leq |D_2G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) - D_2G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,y})| |D_x Y_r^{t,y} h| \\
&\quad + |D_2G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,y}) - D_2G(r, X_r^{t,x}, Y_r^{t,y}, Z_r^{t,y})| |D_x Y_r^{t,y} h| \\
&\quad + |D_2G(r, X_r^{t,x}, Y_r^{t,y}, Z_r^{t,y}) - D_2G(r, X_r^{t,y}, Y_r^{t,y}, Z_r^{t,y})| |D_x Y_r^{t,y} h| \\
&= |\lambda_{21}(r)| + |\lambda_{22}(r)| + |\lambda_{23}(r)|.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E} \left( \int_t^T |\lambda_{21}(r)| dr \right)^p &= \mathbb{E} \left( \int_t^T |D_x Y_r^{t,y} h| \int_0^1 |D_{2,3}^2 G(r, X_r^{t,x}, Y_r^{t,x}, aZ_r^{t,x} + (1-a)Z_r^{t,y})| |Z_r^{t,x} - Z_r^{t,y}| da dr \right)^p \\
&\lesssim \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x Y_r^{t,y} h|^4 \right]^{p/4} \left[ 1 + \mathbb{E} \sup_{r \in [t, T]} |Y_r^{t,x}|^{4m} \right]^{p/4} \left[ \mathbb{E} \int_0^T |Z_r^{t,x} - Z_r^{t,y}|^2 dr \right]^{p/2} \\
&\lesssim |h|^p \mathbb{E} \left( \int_t^T |Z_r^{t,x} - Z_r^{t,y}|^2 dr \right)^{p/2} \rightarrow 0, \quad \text{if } |x - y|_E \rightarrow 0,
\end{aligned}$$

uniformly with respect to  $h \in E$ ,  $|h|_E \leq 1$ . The terms  $\lambda_{22}(r)$  and  $\lambda_{23}(r)$  can be treated in the same manner, so that the required convergence holds for  $\lambda_2(r)$  as well.

It remains to check the term  $\lambda_3(r)$ :

$$\begin{aligned}
|\lambda_3(r)| &\leq |D_3G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) - D_3G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,y})| |D_x Z_r^{t,y} h| \\
&\quad + |D_3G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,y}) - D_3G(r, X_r^{t,x}, Y_r^{t,y}, Z_r^{t,y})| |D_x Z_r^{t,y} h| \\
&\quad + |D_3G(r, X_r^{t,x}, Y_r^{t,y}, Z_r^{t,y}) - D_3G(r, X_r^{t,y}, Y_r^{t,y}, Z_r^{t,y})| |D_x Z_r^{t,y} h| \\
&= |\lambda_{31}(r)| + |\lambda_{32}(r)| + |\lambda_{33}(r)|.
\end{aligned}$$

Exploiting again Assumption 2 we easily derive the required result for each term  $\lambda_{3i}(r)$ ,  $i = 1, 2, 3$ . We present here the estimate involving the increments  $Z_r^{t,x} - Z_r^{t,y}$ :

$$\begin{aligned}
\mathbb{E} \left( \int_t^T |\lambda_{31}(r)| dr \right)^p &\leq \mathbb{E} \left( \int_t^T |D_x Z_r^{t,y} h| \int_0^1 |D_3^2 G(r, X_r^{t,y}, Y_r^{t,x}, aZ_r^{t,x} + (1-a)Z_r^{t,y})| |Z_r^{t,x} - Z_r^{t,y}| da dr \right)^p \\
&\lesssim \left[ \mathbb{E} \left( \int_t^T |D_x Z_r^{t,y} h|^2 dr \right)^p \right]^{1/2} \left[ \mathbb{E} \left( \int_t^T |Z_r^{t,x} - Z_r^{t,y}|^2 dr \right)^p \right]^{1/2} \\
&\lesssim |h|^p (1 + |x|^{m^2 p}) \left[ \mathbb{E} \left( \int_t^T |Z_r^{t,x} - Z_r^{t,y}|^2 dr \right)^p \right]^{1/2}
\end{aligned} \tag{4.16}$$

and the above term converges to zero thanks to Proposition 9.

Summing up, we have that

$$\sup_{\substack{h \in E, \\ |h|_E \leq 1}} \mathbb{E} \left( \int_t^T |\lambda(r)| e^{V_r} dr \right)^p \lesssim \sum_{i=1}^3 \sup_{\substack{h \in E, \\ |h|_E \leq 1}} \mathbb{E} \left( \int_t^T |\lambda_i(r)| dr \right)^p \longrightarrow 0, \quad \text{if } |x - y|_E \rightarrow 0,$$

from which we get the required continuity. For what concerns the estimate (4.13) it simply follows from (4.12) by taking the supremum in  $h \in E$ ,  $|h|_E \leq 1$ .  $\square$

Let us now give a representation result for the solution  $Z^{t,x}$  of (3.11), in terms of the Fréchet differential of the map  $x \mapsto Y^{t,x}$ . This will be crucial in the following, e.g. for the second-order Fréchet differentiability of the map  $x \mapsto (Y^{t,x}, Z^{t,x})$ .

**Proposition 13.** *Let Assumptions 1, 2 and 3 be in force. Given the solution  $(Y^{t,x}, Z^{t,x})$  of the BSDE in (4.1), we define the map  $u : [0, T] \times E \rightarrow \mathbb{R}$  as  $u(t, x) := Y_t^{t,x}$ . Then for every  $(t, x) \in [0, T] \times E$  it holds that*

$$Y_s^{t,x} = u(s, X_s^{t,x}), \quad s \in [t, T] \quad (4.17)$$

and, by (4.9), there exists  $c \geq 0$  such that

$$\sup_{t \in [0, T]} |u(t, x)| \leq c(1 + |x|^m). \quad (4.18)$$

Moreover, by Proposition 12 the map  $x \mapsto u(\cdot, x)$  is differentiable and, denoting by  $Du(t, x)$  its gradient with respect to the second variable, we have

$$\sup_{t \in [0, T]} |Du(t, x)| \leq c(1 + |x|^{m^2}). \quad (4.19)$$

Finally, for every  $(t, x) \in [0, T] \times E$  we have the identification

$$Z^{t,x} = DY^{\cdot, X^{\cdot, x}} \Sigma = Du(\cdot, X^{\cdot, x}) \Sigma \quad \text{in } L^p(\Omega; L^2(0, T)), \quad (4.20)$$

where  $DY^{\cdot, X^{\cdot, x}} = D_y Y^{t,y}|_{y=X^{\cdot, x}}$ . In particular for every  $s \in [t, T]$

$$Z_s^{t,x} = \lim_{r \downarrow s} D_x Y_r^{s, X_s^{t,x}}.$$

Notice that (4.20) identifies a specific version  $\tilde{Z}^{t,x} \in L^p(C([0, T]; \mathbb{R}^{d_1}))$  of  $Z^{t,x}$ . This identification will hold throughout the paper and clearly yields

$$\mathbb{E} \left( \sup_{s \in [t, T]} \left| \tilde{Z}_s^{t,x} \right|^p \right) < +\infty. \quad (4.21)$$

*Proof.* For a proof of this result we refer to (Fabbri et al., 2017, Cor. 6.29) for the Hilbert setting and to Masiero (2008), Zhou and Zhang (2011) for the extension to Banach spaces.  $\square$

**Remark 14.** *Let us comment on the particular case in which  $D\Phi$  and  $D_i G$ ,  $i = 1, 2, 3$ , in (4.11) are uniformly bounded. By a standard application of Girsanov theorem, see e.g. (Fabbri et al., 2017, Section 6.7.1), estimate (4.6) yields the boundedness of  $D_x Y^{t,x} h$  in the sense that there exists  $K \geq 0$  such that for every  $h \in E$*

$$|D_x Y_s^{t,x} h| \leq K|h|, \quad \text{for a.e. } \omega \in \Omega, \quad \forall t \leq s \leq T, \quad \forall x \in E. \quad (4.22)$$

Hence, in view of Proposition 13,

$$|Z_s^{t,x}| \leq K|\Sigma|, \quad \text{for a.e. } \omega \in \Omega, \quad \forall t \leq s \leq T, \quad \forall x \in E, \quad (4.23)$$

where the constant  $K$  only depends on  $\sup_x \|D\Phi(x)\|$ ,  $\sup_{s,x,y,z} (|D_1 G(s, x, y, z)| + |D_2 G(s, x, y, z)|)$  but not on  $D_3 G(s, x, y, z)$ , This is crucial in the application to stochastic optimal control in Section 7.

#### 4.2. Second-order differentiability of the BSDE

From the previous section we know that Fréchet derivatives  $(D_x Y^{t,x}, D_x Z^{t,x})$  are well-defined and the pair  $(D_x Y^{t,x} h, D_x Z^{t,x} h)$  solves equation (4.11) for every  $h \in E$ . By exploiting the form of the equation, here we firstly study the Fréchet differentiability of the directional derivatives  $x \mapsto (D_x Y^{t,x} h, D_x Z^{t,x} h)$ , for every  $h \in E$  fixed. Then, using the uniform character of all the estimates, we identify the pair  $(D_x^2 Y^{t,x}, D_x^2 Z^{t,x})$  as the second-order Fréchet differential of the map  $x \mapsto (Y^{t,x}, Z^{t,x})$ . Similarly to the previous subsection we will use the shorthand

$$D_{i,j}^2 G_r(t, x) := D_{i,j}^2 G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) .$$

For every  $h, k \in E$ , let us introduce the backward equation

$$\begin{aligned} F_s^{t,x}(k, h) + \int_s^T H_r^{t,x}(k, h) dW(r) &= F_T^{t,x}(k, h) + \int_s^T L_r^{t,x}(k, h) dr \\ &- \int_s^T [D_2 G_r(t, x) F_r^{t,x}(k, h) + D_3 G_r(t, x) H_r^{t,x}(k, h)] dr , \end{aligned} \quad (4.24)$$

where we used the notation

$$\begin{aligned} F_T^{t,x}(k, h) &:= D^2 \Phi(X_T^{t,x}) (D_x X_T^{t,x} k, D_x X_T^{t,x} h) + D\Phi(X_T^{t,x}) D_x X_T^{t,x}(k, h) ; \\ L_r^{t,x}(k, h) &:= -D_{1,1}^2 G_r(t, x) (D_x X_r^{t,x} k, D_x X_r^{t,x} h) - D_{1,2}^2 G_r(t, x) (D_x Y_r^{t,x} k, D_x X_r^{t,x} h) \\ &- D_{1,3}^2 G_r(t, x) (D_x Z_r^{t,x} k, D_x X_r^{t,x} h) - D_1 G_r(t, x) D_x^2 X_r^{t,x}(k, h) \\ &- D_{2,1}^2 G_r(t, x) (D_x X_r^{t,x} k, D_x Y_r^{t,x} h) - D_{2,2}^2 G_r(t, x) (D_x Y_r^{t,x} k, D_x Y_r^{t,x} h) - D_{2,3}^2 G_r(t, x) (D_x Z_r^{t,x} k, D_x Y_r^{t,x} h) \\ &- D_{3,1}^2 G_r(t, x) (D_x X_r^{t,x} k, D_x Z_r^{t,x} h) - D_{3,2}^2 G_r(t, x) (D_x Y_r^{t,x} k, D_x Z_r^{t,x} h) - D_{3,3}^2 G_r(t, x) (D_x Z_r^{t,x} k, D_x Z_r^{t,x} h) \\ &=: L_{1;r}^{t,x}(k, h) + L_{2;r}^{t,x}(k, h) + L_{3;r}^{t,x}(k, h) \end{aligned}$$

and  $L_{i;r}^{t,x}$  includes all the terms containing  $D_i G$  or  $D_{i,j}^2 G$ , for any  $j$ . The main result of the section is the following

**Proposition 15.** *Let Assumptions 1, 2 and 3 hold true. For every  $t \in [0, T]$ ,  $h, k \in E$ , equation (4.24) admits a unique solution  $(F_s^{t,x}(k, h), H_s^{t,x}(k, h))$ . For every  $p > 1$  the map  $x \mapsto (D_x Y^{t,x} h, D_x Z^{t,x} h)$  (resp.  $t \mapsto (D_x Y^{t,x} h, D_x Z^{t,x} h)$ ) is Gâteaux differentiable as a map from  $E$  (resp.  $[0, T]$ ) to  $\mathcal{K}_p$ . For every  $(k, h) \in E$  the pair  $(D_x^2 Y_s^{t,x}(k, h), D_x^2 Z_s^{t,x}(k, h))$  satisfies the BSDE (4.24). Moreover, for every  $t \in [0, T]$  and  $p > 1$ , the map  $x \mapsto (Y^{t,x}, Z^{t,x})$  is twice Fréchet differentiable as a map from  $E$  to  $\mathcal{K}_p$  with second order Fréchet differential given by  $(D_x^2 Y^{t,x}, D_x^2 Z^{t,x})$  and*

$$\left[ \mathbb{E} \sup_{s \in [t, T]} \|D_x^2 Y_s^{t,x}\|^p \right]^{1/p} + \left[ \mathbb{E} \left( \int_t^T \|D_x^2 Z_s^{t,x}\|^2 dr \right)^{p/2} \right]^{1/p} \leq c(1 + |x|^l) , \quad (4.25)$$

for some  $c, l \geq 0$ .

*Proof.* For what concerns wellposedness of (4.24), let us check that  $F_T^{t,x}(k, h)$  and  $L_r^{t,x}(k, h)$  satisfy the integrability conditions given in Lemma 2.6. The application of Hölder inequality along with Assumption 3 and Theorem 8 immediately give that  $F_T^{t,x}(k, h) \in L_{\mathcal{F}_T}^p(\Omega; \mathbb{R})$ . To prove that  $L^{t,x}(k, h)$  belongs to  $L^p(\Omega; L^1(0, T; \mathbb{R}))$ , for every  $p > 1$ , we profit from the growth of  $G$  (see Assumption 2) and the estimates on  $D_x X_r, D_x Y_r, D_x Z_r$ , given in Theorem 8 and Proposition 11, respectively. Let us give some details for the term  $L_{1;r}^{t,x}(h, k)$ :

$$\begin{aligned} &\int_t^T |D_{1,1}^2 G_r(t, x) (D_x X_r^{t,x} k, D_x X_r^{t,x} h)| dr \\ &\lesssim (1 + \sup_{s \in [t, T]} |X_s^{t,x}|^m) (1 + \sup_{s \in [t, T]} |Y_r^{t,x}|) \sup_{s \in [t, T]} \|D_x X_s^{t,x}\|^2 |h| |k| \int_t^T |Z_r^{t,x}| dr . \end{aligned}$$

Using Hölder inequality and the estimates recalled above we get boundedness in  $L^p(\Omega; L^1(0, T; \mathbb{R}))$ , as required. The other terms in  $L^{t,x}(h, k)$  can be treated in a similar way.

To prove Fréchet differentiability, fixing  $h, k \in E$  and using the equations solved by  $D_x Y_s^{t,x+k} h, D_x Y_s^{t,x} h$  and  $F_s^{t,x}(k, h)$  it can be easily shown that

$$\Upsilon_s^k + \int_s^T \Psi_r^k dW_r = \Upsilon_T^k - \int_s^T (D_2 G_r(t, x) \Upsilon_r^k + D_3 G_r(t, x) \Psi_r^k - M^k(r) dr) dr , \quad (4.26)$$

where

$$\begin{aligned}
\Upsilon_r^k &:= \frac{1}{|k|} [D_x Y_s^{t,x+k} h - D_x Y_s^{t,x} h - F_s^{t,x}(h, k)], & \Psi_r^k &:= \frac{1}{|k|} [D_x Z_s^{t,x+k} h - D_x Z_s^{t,x} h - H_s^{t,x}(h, k)] \\
\Upsilon_T^k &:= \frac{1}{|k|} [U_T^{t,x+k} h - U_T^{t,x} h - F_T^{t,x}(k, h)] \\
M^k &:= M_1^k + M_2^k + M_3^k := -\frac{1}{|k|} [D_1 G_r(t, x+k) D_x X_r^{t,x+k} h - D_1 G_r(t, x) D_x X_r^{t,x} h - L_{1;r}^{t,x}(k, h)] \\
&\quad - \frac{1}{|k|} [(D_2 G_r(t, x+k) - D_2 G_r(t, x)) D_x Y_r^{t,x+k} h - L_{2;r}^{t,x}(k, h)] \\
&\quad - \frac{1}{|k|} [(D_3 G_r(t, x+k) - D_3 G_r(t, x)) D_x Z_r^{t,x+k} h - L_{3;r}^{t,x}(k, h)].
\end{aligned} \tag{4.27}$$

Taking advantage from the linear character of (4.26), thanks to estimate (2.9) and recalling that in this case  $V$  is a bounded process, we have that

$$\mathbb{E} \sup_{s \in [t, T]} |\Upsilon_s^k|^p + \mathbb{E} \left( \int_t^T |\Psi_r^k|^2 dr \right)^{p/2} \lesssim \mathbb{E} |\Upsilon_T^k|^p + \mathbb{E} \left( \int_t^T |M^k(r)| dr \right)^p.$$

The desired Fréchet differentiability follows as soon as

$$\lim_{k \rightarrow 0} \sup_{|h|=1} \left[ \mathbb{E} |\Upsilon_T^k|^p + \mathbb{E} \left( \int_t^T |M^k(r)| dr \right)^p \right] = 0.$$

A detailed computation of all the terms is postponed in the Appendix. Here we only show how to deal with the most delicate one, which we denote by  $M_{311}^k$  to be consistent with the notation of the appendix,

$$M_{311}^k := -\frac{1}{|k|} \int_0^1 [D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, \lambda Z_r^{t,x+k} + (1-\lambda) Z_r^{t,x}) - D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, Z_r^{t,x})],$$

and where the application of Proposition 13 seems to be crucial. Using the notation  $u(t, x) := Y_t^{t,x}$ , from  $\alpha$ -Hölder continuity of  $D_{3,3}^2 G_r(t, x)$  we get, on a set of full probability,

$$\begin{aligned}
\int_t^T |M_{311}^k(r)| dr &\lesssim \frac{1}{|k|} \int_t^T |Z_r^{t,x+k} - Z_r^{t,x}|^\alpha |Z_r^{t,x+k} - Z_r^{t,x}| \|D_x Z_r^{t,x+k}\| |h| dr \\
&\lesssim \sup_{r \in [t, T]} \|Du(r, X_r^{t,x+k}) - Du(r, X_r^{t,x})\|^\alpha |\Sigma| |h| \left( \int_t^T \frac{|Z_r^{t,x+k} - Z_r^{t,x}|^2}{|k|^2} dr \right)^{1/2} \left( \int_t^T \|D_x Z_r^{t,x+k}\|^2 dr \right)^{1/2}.
\end{aligned}$$

To show that  $\mathbb{E} \left( \int_t^T |M_{311}^k(r)| dr \right)^p \rightarrow 0$  we employ Vitali convergence theorem. Taking advantage of the fact that the initial datum  $x$  is deterministic, from (4.13) there exists  $l \geq 0$  such that for any  $p' > p$

$$\mathbb{E} \left( \int_t^T |M_{311}^k(r)| dr \right)^{p'} \lesssim (1 + |x+k|^l + |x|^l) |\Sigma| |h| \left[ \mathbb{E} \left( \int_t^T \frac{|Z_r^{t,x+k} - Z_r^{t,x}|^2}{|k|^2} dr \right)^{2p'} \right]^{1/4} < +\infty,$$

which is bounded thanks to Proposition 12 and estimate (4.13). This guarantees the uniform integrability of the family  $\mathbb{E} \left( \int_t^T |M_{311}^k(r)| dr \right)^p$ , when  $k$  is varying. Hence, it remains to show that for a.e.  $\omega \in \Omega$  and a.e.  $r \in [0, T]$

$$|Du(r, X_r^{t,x+k}) \Sigma - Du(r, X_r^{t,x}) \Sigma| \longrightarrow 0, \quad \text{if } |k| \rightarrow 0. \tag{4.28}$$

To do it, recall the general continuity result for  $y \mapsto D_x Y^{t,y} h$  as a map from  $E$  to  $L^2(\Omega; C([0, T]; \mathbb{R}))$  given in the proof of Proposition 12. When dealing with  $u(t, x) = Y_t^{t,x}$ , which is deterministic, this implies that for every  $y_1, y_2, h \in E$

$$|Du(t, y_1) h - Du(t, y_2) h| \longrightarrow 0, \quad \text{as } |y_1 - y_2| \rightarrow 0. \tag{4.29}$$

Now, given a basis  $e_1, \dots, e_{d_1}$  in  $\mathbb{R}^{d_1}$ , (4.28) is equivalent to the convergence

$$\sup_{j \in \{1, \dots, d_1\}} |Du(r, X_r^{t,x+k}) \Sigma e_j - Du(r, X_r^{t,x}) \Sigma e_j| \longrightarrow 0, \quad \text{if } |k| \rightarrow 0.$$

where  $\Sigma e_j \in E$  for every  $j = 1, \dots, d_1$ . Hence combining the continuity of the map  $x \mapsto X_s^{t,x}(\omega)$  for every  $\omega \in \Omega_0$ ,  $s \in [t, T]$  (see Theorem 8) with the convergence result in (4.29) we easily get (4.28). For a detailed proof of the convergence of all the remaining terms we refer to the Appendix.

To conclude the proof observe that estimate (4.25) is a direct consequence of Lemma 2 applied to (4.24). Indeed, for every  $k, h \in E$ , uniqueness of solutions to (4.24) gives that  $F_s^{t,x}(k, h) = D_x^2 Y_s^{t,x}(k, h)$  and  $H_s^{t,x}(k, h) = D_x^2 Z_s^{t,x}(k, h)$ .  $\square$

**Remark 16.** *The proof of (4.28) as it is performed here exploits the fact that noise in the forward equation is additive, and this simplifies estimates on  $D_x X^{t,x}$ . Our techniques also rely on the fact that the noise is finite dimensional, e.g. when proving the continuity of the map  $y \mapsto Du(t, y)$  in the operator norm as shown above.*

**Corollary 17.** *Setting  $u(t, x) := Y_t^{t,x}$  as in Proposition 13, the map  $x \mapsto u(t, x)$  belongs to  $C^2(E; \mathbb{R})$ , for every  $t \in [0, T]$ . Moreover, for some  $c \geq 0$ ,  $l \in \mathbb{N}$  it holds that*

$$\sup_{t \in [0, T]} |D^2 u(t, x)| \leq c \left(1 + |x|^l\right) .$$

## 5. Solution to the Kolmogorov equation in $\mathcal{L}^2$

In this section we deal with the infinite-dimensional semilinear PDE (1.1) in the space  $\mathcal{L}^2$ . Therefore we assume all the coefficients to be defined on  $\mathcal{L}^2$  and that assumptions 1 and 2 are satisfied with  $E = \mathcal{L}^2$ . This is a quite strong requirement that is seldom satisfied by examples; however it represents only the first step towards establishing the theory for  $E = \widehat{\mathcal{C}}$ , where the same assumptions are much more reasonable and indeed verified by a large class of examples.

**Theorem 18.** *Let Assumptions 1, 2 and 3 hold with  $E = \mathcal{L}^2$ . Then the function  $u(t, x) := Y_t^{t,x}$  is a classical solution, in the sense of Definition 4, to the semilinear Kolmogorov equation (1.1).*

*Proof.* Thanks to the regularity results given in Section 4 we know that the map  $(t, x) \mapsto u(t, x)$  belongs to  $C^{1;2}([0, T] \times \mathcal{L}^2)$ . Hence, it is enough to prove that  $u(t, x) := Y_t^{t,x}$  is a solution of the semilinear Kolmogorov equation in integral form:

$$\begin{aligned} u(t, x) - \Phi(x) + \int_t^T G(s, x, u(s, x), Du(s, x)\Sigma) ds \\ = \int_t^T \left[ Du(s, x) [Ax + B(s, x)] + \frac{1}{2} \sum_{j=1}^d \Sigma \Sigma^* D^2 u(s, x) (e_j, e_j) \right] ds . \end{aligned} \quad (5.1)$$

The standard way of proving such a result goes through an application of Itô formula to the increments of  $u(t, X)$  along a partition; eventually taking expectations, summing along the partition and letting the size of the mesh going to 0 yields the result. The difficulty here lies in the fact that at every time  $t$ ,  $X(t)$  lies almost surely not in the domain of the operator  $A$ . There are different ways to circumvent this difficulty; we detail here one of the possibilities.

Consider two time instants  $0 \leq t_0 \leq t_1 \leq T$  and a point  $x \in \mathcal{L}^2$ . We want to analyse the increment

$$\begin{aligned} u(t_0, x) - u\left(t_1, e^{(t_1-t_0)A}x\right) &= \mathbb{E}u(t_0, x) - u\left(t_1, e^{(t_1-t_0)A}x\right) \\ &= \mathbb{E}Y_{t_0}^{t_0,x} - \mathbb{E}Y_{t_1}^{t_0,x} + \mathbb{E}Y_{t_1}^{t_0,x} - u\left(t_1, e^{(t_1-t_0)A}x\right) . \end{aligned} \quad (5.2)$$

Thanks to the Markov property of  $X^{t_0,x}$  it is not difficult to show (see (Fuhrman and Tessitore, 2002)) that almost surely

$$Y_t^{t_0,x} = Y_t^{t_1, X_{t_1}^{t_0,x}} , \quad Z_t^{t_0,x} = Z_t^{t_1, X_{t_1}^{t_0,x}}$$

for every  $t \in [t_1, T]$ , hence

$$\mathbb{E}Y_{t_1}^{t_0,x} = \mathbb{E}Y_{t_1}^{t_1, X_{t_1}^{t_0,x}} = \mathbb{E}u\left(t_1, X_{t_1}^{t_0,x}\right)$$

and (5.2) yields

$$u(t_0, x) - u\left(t_1, e^{(t_1-t_0)A}x\right) = \mathbb{E}\left[Y_{t_0}^{t_0,x} - Y_{t_1}^{t_0,x}\right] + \mathbb{E}\left[u\left(t_1, X_{t_1}^{t_0,x}\right) - u\left(t_1, e^{(t_1-t_0)A}x\right)\right] . \quad (5.3)$$

Since  $Y$  satisfies the BSDE (3.11), the first expectation on the r.h.s. can be written as

$$\mathbb{E} \left[ - \int_{t_0}^{t_1} G(r, X_r^{t_0, x}, Y_r^{t_0, x}, Z_r^{t_0, x}) dr \right].$$

Now fix  $t \in [0, T]$  and consider a sequence  $\{\pi^n\}$  of partitions of  $[t, T]$  such that each of the  $\pi^n$ 's is given by  $k_n + 1$  points  $t = t_1^n \leq t_2^n \leq \dots \leq t_{k_n+1}^n = T$  and such that  $|\pi^n| \rightarrow 0$  as  $n \rightarrow \infty$ . For each fixed  $n$  and every  $i = 1, \dots, k_n + 1$  we consider (5.3) with  $t_0 = t_i^n$  and  $t_1 = t_{i+1}^n$  and sum over the index  $i$  obtaining

$$u(t, x) - \Phi(x) = - \sum_{i=1}^{k_n+1} \mathbb{E} \left[ \int_{t_i^n}^{t_{i+1}^n} G(r, X_r^{t_i^n, x}, Y_r^{t_i^n, x}, Z_r^{t_i^n, x}) dr \right] + I_n.$$

The term

$$I_n = \sum_{i=1}^{k_n+1} \mathbb{E} \left[ u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) - u(t_{i+1}, e^{(t_{i+1}-t_i)A}x) \right]$$

can be treated as in the proof of Theorem 4.1 of (Flandoli and Zanco, 2016), yielding as  $n \rightarrow \infty$  the linear part of the PDE (i.e. the r.h.s. of (5.1)). The only difference is that in Flandoli and Zanco (2016)  $\Phi$  is assumed to be bounded, but the generalization to the polynomial growth (cf. Assumption 3) is immediate. Concerning the remaining term, we need to prove that

$$\mathbb{E} \int_t^T \sum_{i=1}^{k_n+1} G(r, X_r^{t_i^n, x}, Y_r^{t_i^n, x}, Z_r^{t_i^n, x}) \mathbb{1}_{[t_i^n, t_{i+1}^n)}(r) dr \xrightarrow{n \rightarrow \infty} \int_t^T G(r, x, u(r, x), Du(r, x)\Sigma) dr.$$

Let us write

$$\begin{aligned} & \mathbb{E} \int_t^T \sum_{i=1}^{k_n+1} \left[ G(r, X_r^{t_i^n, x}, Y_r^{t_i^n, x}, Z_r^{t_i^n, x}) - G(r, x, u(r, x), Du(r, x)\Sigma) \right] \mathbb{1}_{[t_i^n, t_{i+1}^n)}(r) dr \\ &= \mathbb{E} \int_t^T \sum_{i=1}^{k_n+1} \left[ G(r, X_r^{t_i^n, x}, Y_r^{t_i^n, x}, Z_r^{t_i^n, x}) - G(r, X_r^{t_i^n, x}, Y_r^{t_i^n, x}, Du(r, x)\Sigma) \right] \mathbb{1}_{[t_i^n, t_{i+1}^n)}(r) dr \\ &+ \mathbb{E} \int_t^T \sum_{i=1}^{k_n+1} \left[ G(r, X_r^{t_i^n, x}, Y_r^{t_i^n, x}, Du(r, x)\Sigma) - G(r, X_r^{t_i^n, x}, u(r, x), Du(r, x)\Sigma) \right] \mathbb{1}_{[t_i^n, t_{i+1}^n)}(r) dr \\ &+ \mathbb{E} \int_t^T \sum_{i=1}^{k_n+1} \left[ G(r, X_r^{t_i^n, x}, u(r, x), Du(r, x)\Sigma) - G(r, x, u(r, x), Du(r, x)\Sigma) \right] \mathbb{1}_{[t_i^n, t_{i+1}^n)}(r) dr \end{aligned}$$

so that, by the Lipschitz character of  $G$  and Proposition 13

$$\begin{aligned} & \lesssim \mathbb{E} \int_t^T \sum_{i=1}^{k_n+1} \left[ \left| Du(r, X_r^{t_i^n, x})\Sigma - Du(r, x)\Sigma \right| + \left| u(r, X_r^{t_i^n, x}) - u(r, x) \right| \right] \mathbb{1}_{[t_i^n, t_{i+1}^n)}(r) dr \\ &+ \mathbb{E} \int_t^T \sum_{i=1}^{k_n+1} \left[ G(r, X_r^{t_i^n, x}, u(r, x), Du(r, x)\Sigma) - G(r, x, u(r, x), Du(r, x)\Sigma) \right] \mathbb{1}_{[t_i^n, t_{i+1}^n)}(r) dr. \end{aligned} \tag{5.4}$$

The last term in (5.4) can be treated as follows. For every  $r \in [t, T]$  fixed, there exists a unique sequence of intervals  $\{[t_{i(r,n)}^n, t_{i(r,n)+1}^n)\}_{n \in \mathbb{N}}$  such that  $r \in [t_{i(r,n)}^n, t_{i(r,n)+1}^n)$  for every  $n \in \mathbb{N}$ . Moreover, for every  $x \in \mathcal{L}^2$ ,  $r \in [t, T]$  and  $\omega \in \Omega_0$ , Proposition 7 guarantees the continuity of the map  $\tau \mapsto X_r^{\tau, x}(\omega)$ , so that

$$\left| X_r^{t_{i(r,n)}^n, x}(\omega) - x \right|_{\mathcal{L}^2} \xrightarrow{n \rightarrow \infty} 0. \tag{5.5}$$

From the regularity of  $G$  (see Assumption 2) it easily follows that for every  $x \in \mathcal{L}^2$ ,  $r \in [t, T]$  and  $\omega \in \Omega_0$

$$\left| G(r, X_r^{t_{i(r,n)}^n, x}(\omega), u(r, x), Du(r, x)\Sigma) - G(r, x, u(r, x), Du(r, x)\Sigma) \right| \xrightarrow{n \rightarrow \infty} 0$$

Thanks to Assumption 2 and estimate (4.3), the application of the Vitali theorem gives the required convergence.

Concerning the first term in (5.4), we employ for every  $r \in [t, T]$  the continuity of the (deterministic) maps  $y \mapsto u(r, y)$  and  $y \mapsto Du(r, y)\Sigma$ , with  $y \in \mathcal{L}^2$ , given by Proposition 13 along with Propositions 9 and 12, respectively. These, in combination with (5.5), give the convergence of the integrand, for a.e.  $\omega \in \Omega$ , for every  $r \in [t, T]$ . Recalling estimates (4.9) and (4.13) and applying again the Vitali convergence theorem we finally get the result.  $\square$

**Remark 19.** *With the very same proof we can easily show existence of solutions also in the space*

$$\mathcal{L}^p := \mathbb{R}^d \times L^p(-T, 0; \mathbb{R}^d) \quad , \quad p > 2 \quad .$$

*This allows to treat coefficients depending on the path in an integral way, which are not smooth in  $\mathcal{L}^2$  but satisfy our assumptions in  $\mathcal{L}^{2+\varepsilon}$ , for any  $\varepsilon > 0$ . For this particular choice, it is then possible to establish wellposedness neither requiring Assumption 5 nor introducing the approximation procedure explained in the next section.*

## 6. Solution of the Kolmogorov equation in $\mathcal{D}$

Here we prove our main result, following the strategy described in the introduction.

For every initial condition  $x \in \widehat{\mathcal{C}}^1$  and every initial time  $t \in [0, T]$  we can find a solution  $(X^{t,x}, Y^{t,x}, Z^{t,x}) \in L^p(\Omega; C([t, T]; \mathcal{D})) \times L^p(\Omega; C([t, T]; \mathbb{R})) \times L^p(\Omega; L^2(0, T; \mathbb{R}^{d_1}))$  of the forward-backward system

$$\begin{cases} dX_s = [AX_s + B(s, X_s)] ds + \Sigma dW_s & \text{in } [t, T] \\ dY_s = G(s, X_s, Y_s, Z_s) ds + Z_s dW_s & \text{in } [t, T] \\ X_t = x \\ Y_T = \Phi(X_T). \end{cases} \quad (6.1)$$

Then we can define the function  $u: [0, T] \times \widehat{\mathcal{C}}^1 \rightarrow \mathbb{R}$  as

$$u(t, x) := Y_t^{t,x}, \quad (6.2)$$

and show that it is a classical solution of the Kolmogorov equation. This is the content of the next theorem.

**Theorem 20.** *Let  $B, G$  and  $\Phi$  satisfy respectively Assumptions 1, 2 and 3 with  $E = \mathcal{D}$ , as well as Assumptions 4 and 5. Assume moreover that  $B$  maps  $\mathcal{C}$  into itself. The function  $u$  defined by (6.2) is a classical solution of the Kolmogorov semilinear equation with terminal condition  $\Phi$ , i.e.  $u \in C^{1;2}([0, T] \times \mathcal{D}, \mathbb{R})$  and*

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) + Du(t, x) [Ax + B(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 v(t, x)] = G(t, x, u(t, x), Du(t, x) \Sigma), \\ u(T, \cdot) = \Phi, \end{cases} \quad (1.1)$$

for every  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}^1$ .

**Remark 21.** *The requirement that  $B$  maps  $\mathcal{C}$  into itself is automatically satisfied if  $B$  is the lifting of a path-dependent function as described in Subection 3.1. The regularity  $x \in \widehat{\mathcal{C}}^1$  is necessary to give sense to the term  $Ax$  and it is a standard requirement in the framework of classical solutions.*

*Proof.* We give here a complete scheme of the proof and postpone most of the technicalities to Lemmas 22-26 below. We will assume for simplicity that  $B, \Phi$  and  $G$  have the same smoothing sequences, but the proof applies with almost no modifications also when the smoothing sequences are different.

To lighten the notation, we will write  $X_s^n = X_s^{n;t,x}(\omega)$ ,  $X_s = X_s^{t,x}(\omega)$  (and similarly for  $Y_s^n, Z_s^n$ ). We will also take  $m \geq 1$  in Assumptions 1 and 2; this guarantees that the exponents  $mp$  in all the estimates below are larger than 1. The general case  $m \geq 0$  follows from a further application of Hölder's inequality.

Firstly observe that, by Proposition 13 and Corollary 17,  $u$  has the required regularity and

$$Du(t, x) = D_x Y_t^{t,x} \quad \text{and} \quad D^2 u(t, x) = D_x^2 Y_t^{t,x} \quad .$$

Then, given  $B, G$  and  $\Phi$  we define for every  $n \in \mathbb{N}$

$$\begin{aligned} B^n(t, x) &:= B(t, J^n x) \\ G^n(t, x, y, z) &:= G(t, J^n x, y, z) \\ \Phi^n(x) &:= \Phi(J^n x) \quad ; \end{aligned}$$

it is immediate to check that also  $B^n, G^n, \Phi^n$  satisfy Assumptions 1-5 on  $\mathcal{L}^2$  with constants that *do not depend on  $n$* . Moreover

$$\begin{aligned} D\Phi^n(x)\bar{x} &= D\Phi(J^n x)J^n \bar{x} \quad , \\ D^2\Phi^n(x)(\bar{x}_1, \bar{x}_2) &= D^2\Phi(J^n x)(J^n \bar{x}_1, J^n \bar{x}_2) \quad , \\ D_1 G^n(r, x, y, z)\bar{x} &= D_1 G(r, J^n x, y, z)J^n \bar{x} \quad , \\ D_{1,1}^2 G^n(r, x, y, z)(\bar{x}_1, \bar{x}_2) &= D_{1,1}^2 G(r, J^n x, y, z)(J^n \bar{x}_1, J^n \bar{x}_2) \quad , \\ D_{1,2} G^n(r, x, y, z)(\bar{x}, \bar{y}) &= D_{1,2} G(r, J^n x, y, z)(J^n \bar{x}, \bar{y}) \quad , \\ D_2 G^n(r, x, y, z)\bar{y} &= D_2 G(r, J^n x, y, z)\bar{y} \quad , \\ D_{2,2}^2 G^n(r, x, y, z)(\bar{y}_1, \bar{y}_2) &= D_{2,2}^2 G(r, J^n x, y, z)(\bar{y}_1, \bar{y}_2) \quad , \end{aligned} \quad (6.3)$$

with similar identities for the derivatives of  $G^n$  with respect to the variable  $z$  and for the derivatives of  $B^n$ . We actually have that  $B^n$  maps  $\mathcal{L}^2$  into  $\mathcal{C} \subset \mathcal{D}$  and, thanks to the properties of  $\{J^n\}_{n \in \mathbb{N}}$  (see Section 3.2) and the Lipschitz character of  $B$ , for every  $t \in [0, T]$  and  $x \in \mathcal{C}$  it holds that

$$B^n(t, x) \xrightarrow{n \rightarrow \infty} B(t, x) \text{ in } \mathcal{C} \quad (6.4)$$

and for every  $(t, x, y, z) \in [0, T] \times \mathcal{C} \times \mathbb{R} \times \mathbb{R}^{d_1}$

$$\Phi^n(x) \xrightarrow{n \rightarrow \infty} \Phi(x), \quad G^n(t, x, y, z) \xrightarrow{n \rightarrow \infty} G(t, x, y, z) \quad \text{in } \mathbb{R}. \quad (6.5)$$

For every  $x \in \mathcal{L}^2$ ,  $t \in [0, T]$  and for each  $n \in \mathbb{N}$  we can solve the forward-backward system in  $\mathcal{L}^2$

$$\begin{cases} dX_s^n = [AX_s^n + B^n(s, X_s^n)] ds + \Sigma dW_s & \text{in } [t, T] \\ dY_s^n = G^n(s, X_s^n, Y_s^n, Z_s^n) ds + Z_s^n dW_s & \text{in } [t, T] \\ X_t^n = x \\ Y_T^n = \Phi^n(X_T^n) \end{cases} \quad (6.6)$$

thus obtaining a sequence of solutions  $(X^{n;t,x}, Y^{n;t,x}, Z^{n;t,x})_n$ . Note that all the estimates in Theorem 8, Proposition 9 and Propositions 11-15 hold *uniformly* in  $n$  due to the equiboundedness of the  $J^n$ 's; this is a crucial feature for the proof.

Thanks to Theorem 18, the deterministic function

$$u^n(t, x) := Y_t^{n;t,x}$$

is twice Fréchet differentiable with

$$Du^n(t, x) = D_x Y_t^{n;t,x}, \quad D^2 u^n(t, x) = D_x^2 Y_t^{n;t,x}, \quad (6.7)$$

and it solves the backward PDE<sub>n</sub> in  $\mathcal{L}^2$

$$\begin{cases} \frac{\partial u^n}{\partial t}(t, x) + Du^n(t, x) [Ax + B^n(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 u^n(t, x)] = G^n(t, x, u^n(t, x), Du^n(t, x) \Sigma), \\ u^n(T, x) = \Phi^n(x). \end{cases} \quad (6.8)$$

By choosing  $x \in \widehat{\mathcal{C}}^1 \subset \widehat{\mathcal{C}}$  also in the system (6.6), we know that for every  $n \in \mathbb{N}$  and  $s \in [t, T]$  the random variable  $X_s^{n;t,x}$  belongs to  $\widehat{\mathcal{C}}$  (cf. Proposition 7) and in particular it is differentiable as a random variable with values in  $\mathcal{L}^2$ . To conclude the proof, it remains to show that  $u^n(t, x)$  converges to  $u(t, x)$  for every  $t \in [0, T]$  and that each term in the PDE<sub>n</sub> converges to the corresponding term in the PDE (1.1) as  $n \rightarrow \infty$ . Convergence of  $u^n(t, x)$  to  $u(t, x)$ , for every  $(t, x) \in [0, T] \times \widehat{\mathcal{C}}$ , is a consequence of the (more general) convergence  $Y^n \rightarrow Y$  in  $L^p(\Omega; C([t, T]; \mathbb{R}))$  given in Lemma 24 below.

Regarding the first derivative of  $u$ , Lemma 25 guarantees that for any  $h \in \mathcal{C}$

$$DY^n h \rightarrow DY h \quad \text{in } L^p(\Omega; C([t, T]; \mathbb{R}));$$

therefore  $Du^n(t, x)h \rightarrow Du(t, x)h$  for every  $h \in \mathcal{C}$ . Writing

$$Du^n(t, x)B^n(t, x) - Du(t, x)B(t, x) = Du^n(t, x) [B^n(t, x) - B(t, x)] + [Du^n(t, x) - Du(t, x)] B(t, x),$$

the convergence of the second term on the r.h.s. is straightforward. Using estimate (4.19) for  $u^n$  (which is indeed uniform in  $n$ ), the first term goes to zero by (6.4). Since  $Ax \in \mathcal{C}$ , this implies that the linear first order term  $Du^n(t, x) [Ax + B^n(t, x)]$  in PDE<sub>n</sub> converges to the corresponding term in the limit PDE.

For what concerns the second order term in PDE<sub>n</sub>, in Lemma 26 we exploit the identification result obtained in Proposition 13 to show that for any  $h, k \in \mathcal{C}$  it holds

$$\begin{aligned} D_x^2 Y^n(k, h) &\rightarrow D_x^2 Y(k, h) && \text{in } L^p(\Omega; C([t, T]; \mathbb{R})), \\ D_x^2 Z^n(k, h) &\rightarrow D_x^2 Z(k, h) && \text{in } L^p(\Omega; L^2([t, T]; \mathbb{R}^{d_1})), \end{aligned}$$

which is sufficient thanks to (6.7). Finally, since  $Y_t, Y_t^n, D_x Y_t$  and  $D_x Y_t^n$  are all deterministic, from continuity of  $G$  and Lemmas 24, 25 it follows that

$$G^n(t, x, u^n(t, x), D_x Y_t^n \Sigma) \rightarrow G(t, x, u(t, x), Du(t, x) \Sigma),$$

and this concludes the proof.  $\square$



In Lemmas 22–26 below we will always let the assumptions of Theorem 20 to hold, without explicitly write it in every statement. The only difference concerns the less stringent requirement  $x \in \widehat{\mathcal{C}}$  (instead of  $x \in \widehat{\mathcal{C}}^1$ ) which turns out to be sufficient for all the convergences.

We will use the notation  $a \lesssim b$  meaning  $a \leq Cb$  for some positive constant  $C$  only when the hidden constant  $C$  does not depend on  $n$  nor on the time variables. All the convergences has to be intended as  $n$  goes to  $+\infty$ .

**Lemma 22.** *Let  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}$ . Then  $X^{n;t,x} \rightarrow X^{t,x}$  and  $J^n X^{n;t,x} \rightarrow X^{t,x}$  P-a.s. in  $C([0, T]; \mathcal{C})$  and also in  $L^p(\Omega; C([0, T]; \mathcal{C}))$ .*

*Proof.* Recall that  $\Omega_0 \subset \Omega$  is the subset of full probability where  $X^{(t,x)}$  has continuous trajectories. Given  $x \in \widehat{\mathcal{C}}$  and  $\omega \in \Omega_0$  the continuity of the map  $[t, T] \ni s \mapsto X_s \in \widehat{\mathcal{C}}$  guarantees the compactness in  $\widehat{\mathcal{C}}$  of the set  $\{X_s(\omega)\}_s$ . Therefore  $J^n X_s \rightarrow X_s$  uniformly in  $s$  almost surely, i.e.

$$\sup_{s \in [t, T]} |J^n X_s - X_s| \xrightarrow{n \rightarrow \infty} 0, \quad \text{P-a.s. .}$$

Thanks to Assumption 1 and to the properties of the semigroup  $e^{tA}$  we have almost surely

$$\begin{aligned} |X_\tau^n - X_\tau| &= \left| \int_t^\tau e^{(\tau-r)A} [B^n(r, X_r^n) - B(r, X_r)] dr \right| \\ &\lesssim \int_t^\tau [|B(r, J^n X_r^n) - B(r, J^n X_r)| + |B(r, J^n X_r) - B(r, X_r)|] dr \\ &\lesssim \int_t^\tau [|X_r^n - X_r| + |J^n X_r - X_r|] dr ; \end{aligned}$$

therefore

$$\sup_{\tau \in [t, s]} |X_\tau^n - X_\tau| \lesssim \int_t^s \sup_{\tau \in [t, r]} |X_\tau^n - X_\tau| dr + \int_t^s \sup_{\tau \in [t, r]} |J^n X_\tau - X_\tau| dr$$

and by Gronwall's lemma

$$\sup_{\tau \in [t, s]} |X_\tau^n - X_\tau| \xrightarrow{n \rightarrow \infty} 0, \quad \text{P-a.s. .} \quad (6.9)$$

Since for every  $s$  a.s.

$$\begin{aligned} |J^n X_s^n - X_s|_{\widehat{\mathcal{C}}} &\leq |J^n X_s^n - J^n X_s|_{\widehat{\mathcal{C}}} + |J^n X_s - X_s|_{\widehat{\mathcal{C}}} \\ &\leq |J^n|_{L(\mathcal{L}^2; \mathcal{C})} |X_s^n - X_s|_{\widehat{\mathcal{C}}} + |J^n X_s - X_s|_{\widehat{\mathcal{C}}} \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

the equiboundedness of the  $J^n$ 's implies that

$$\sup_{\tau \in [t, s]} |J^n X_s^n - X_s|_{\widehat{\mathcal{C}}} \xrightarrow{n \rightarrow \infty} 0 .$$

The second claim follows by estimate (4.3) (here the initial datum  $x$  is deterministic). Indeed

$$\sup_{s \in [t, T]} |X_s^n| + \sup_{s \in [t, T]} |X_s| \leq \gamma_T, \quad \text{P-a.s.}$$

where  $\gamma_T$  is a random variable with  $\mathbb{E}\gamma_T^p < \infty$ , for every  $p \geq 1$ . □

**Lemma 23.** *Let  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}$ . For every  $h \in \mathcal{C}$ ,  $D_x X^{n;t,x} h \rightarrow D_x X^{t,x} h$  P-a.s. in  $C([0, T]; \mathcal{C})$  and in  $L^p(\Omega; C([0, T]; \mathcal{C}))$ . Moreover, for every  $h, k \in \mathcal{C}$ ,  $D_x^2 X^{n;t,x}(k, h) \rightarrow D_x^2 X^{t,x}(k, h)$  in  $L^p(\Omega, C([0, T]; \mathcal{C}))$ .*

*Proof.* First note that in general we cannot expect  $J^n D_x X_s h$  to converge to  $D_x X_s h$  when  $h \notin \widehat{\mathcal{C}}$ ; this is due to the action of the semigroup  $e^{tA}$  on  $h$  (see equation (4.4)). We prove here only the first part of the statement, for what concerns second order derivatives we refer to the appendix.

Thanks to the equiboundedness of the  $J^n$ 's and (4.6), we can find a constant  $c = c(B, \Sigma, T)$  such that

$$|D_x X_s(t, x) h| \vee \sup_{n \in \mathbb{N}} |D_x X_s^n(t, x) h| \leq c |h| \quad \forall h \in \mathcal{C}, \quad \forall s \in [t, T]. \quad (6.10)$$

By properties of  $B$ ,  $e^{tA}$  and  $J^n$  we also have, for  $h \in \mathcal{C}$ ,

$$|D_x X_\tau^n h - D_x X_\tau h|^p \lesssim \int_t^\tau |DB^n(r, X_r^n) D_x X_r^n h - DB(r, X_r) D_x X_r h|^p dr$$

$$\begin{aligned}
&\lesssim \int_t^\tau |DB(r, J^n X_r^n) (J^n D_x X_r^n h - J^n D_x X_r h)|^p dr + \int_t^\tau |(DB(r, J^n X_r^n) - DB(r, X_r)) J^n D_x X_r h|^p dr \\
&\quad + \int_t^\tau |DB(r, X_r) (J^n D_x X_r h - D_x X_r h)|^p dr \\
&\lesssim \int_t^\tau |D_x X_r^n h - D_x X_r h|^p dr + \int_t^\tau |h|^p |J^n X_r^n - X_r|^p \int_0^1 |D^2 B(r, aJ^n X_r^n + (1-a)X_r)|^p da dr \\
&\quad + \int_t^\tau |DB(r, X_r) (J^n D_x X_r h - D_x X_r h)|^p dr \\
&\lesssim \int_t^\tau |D_x X_r^n h - D_x X_r h|^p dr + \int_t^\tau |h|^p |J^n X_r^n - X_r|^p \int_0^1 (1 + |aJ^n X_r^n + (1-a)X_r|^{mp}) da dr \\
&\quad + \int_t^\tau |DB(r, X_r) (J^n D_x X_r h - D_x X_r h)|^p dr .
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{E} \sup_{\tau \in [t, T]} |D_x X_\tau^n h - D_x X_\tau h|^p &\lesssim \int_t^T \mathbb{E} \sup_{\tau \in [t, r]} |D_x X_\tau^n h - D_x X_\tau h|^p dr \\
&\quad + \mathbb{E} \int_t^T |h|^p |J^n X_r^n - X_r|^p \int_0^1 (1 + |aJ^n X_r^n + (1-a)X_r|^{mp}) da dr \\
&\quad + \mathbb{E} \int_t^T |DB(r, X_r) (J^n D_x X_r h - D_x X_r h)|^p dr .
\end{aligned}$$

Now the second term is bounded by

$$\mathbb{E} \left[ |h|^p \left( 1 + \sup_{r \in [t, T]} |X_r^n|^{mp} + \sup_{r \in [t, T]} |X_r|^{mp} \right) \sup_{r \in [t, T]} |J^n X_r^n - X_r|^p \right],$$

which goes to zero thanks to Hölder inequality, estimates (4.3) and Lemma 22. Exploiting estimate (4.7) and Assumption 4 the same holds for the third term. From the Gronwall's lemma we get the convergence in  $L^p(\Omega; C([0, T]; \mathcal{C}))$ . Finally, by the very same technique, the a.s. convergence in  $C([0, T]; \mathcal{C})$  follows directly exploiting the a.s. convergence of  $J^n X^{n;t,x}$  given in Lemma 22.  $\square$

**Lemma 24.** *Let  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}$ . Then  $Y^{n;t,x} \rightarrow Y^{t,x}$  in  $L^p(\Omega; C([t, T]; \mathbb{R}))$  and  $Z^{n;t,x} \rightarrow Z^{t,x}$  in  $L^p(\Omega; L^2(t, T; \mathbb{R}^{d_1}))$ .*

*Proof.* We first show that, for every  $s \in [t, T]$ ,  $Y_s^{n;t,x} \rightarrow Y_s^{t,x}$  in  $L^p(\Omega; \mathbb{R})$ . Given  $p \geq 2$ , for every  $s \in [t, T]$ ,  $Y_s^n - Y_s$  and  $Z_s^n - Z_s$  satisfy the identity

$$Y_s^n - Y_s + \int_s^T [Z_r^n - Z_r] dW_r = \Phi^n(X_T^n) - \Phi(X_T) + \int_s^T \hat{G}_r^n dr,$$

where  $\hat{G}_r^n$  is the process

$$\hat{G}_r^n = G^n(r, X_r^n, Y_r^n, Z_r^n) - G(r, X_r, Y_r, Z_r).$$

By Itô formula and taking expectation we get

$$\begin{aligned}
\mathbb{E} |Y_s^n - Y_s|^p &+ \frac{p(p-1)}{2} \mathbb{E} \int_s^T |Y_r^n - Y_r|^{p-2} |Z_r^n - Z_r|^2 dr \\
&\leq \mathbb{E} |\Phi^n(X_T^n) - \Phi(X_T)|^p + p \mathbb{E} \int_s^T |Y_r^n - Y_r|^{p-1} |\hat{G}_r^n| dr .
\end{aligned} \tag{6.11}$$

Since  $G^n, G$  satisfy Assumption 2, for the last integral in (6.11) we have

$$\begin{aligned}
\mathbb{E} \int_s^T |Y_r^n - Y_r|^{p-1} |\hat{G}_r^n| dr &\leq \left( \frac{C^2}{2} + C + \frac{p-1}{p} \right) \int_s^T |Y_r^n - Y_r|^p dr \\
&\quad + \frac{1}{2} \int_s^T |Y_r^n - Y_r|^{p-2} |Z_r^n - Z_r|^2 dr + \frac{1}{p} \int_s^T |G(r, J^n X_r^n, Y_r, Z_r) - G(r, X_r, Y_r, Z_r)|^p dr
\end{aligned}$$

where  $C$  is the constant provided by assumption 2. Therefore

$$\begin{aligned}
\mathbb{E} |Y_s^n - Y_s|^p &\lesssim \int_s^T \mathbb{E} |Y_r^n - Y_r|^p dr + \mathbb{E} |\Phi^n(X_T^n) - \Phi(X_T)|^p \\
&\quad + \mathbb{E} \int_s^T |G(r, J^n X_r^n, Y_r, Z_r) - G(r, X_r, Y_r, Z_r)|^p dr
\end{aligned} \tag{6.12}$$

and since  $\int_s^T |G(r, J^n X_r^n, Y_r, Z_r) - G(r, X_r, Y_r, Z_r)|^p dr$  is decreasing in  $s$ , by Gronwall's lemma

$$\mathbb{E} |Y_s^n - Y_s|^p \lesssim \left[ \mathbb{E} |\Phi^n(X_T^n) - \Phi(X_T)|^p + \mathbb{E} \int_t^T |G(r, J^n X_r^n, Y_r, Z_r) - G(r, X_r, Y_r, Z_r)|^p dr \right] \quad (6.13)$$

The first term on the r.h.s. of (6.13) can be easily shown to converge to 0 thanks to the properties of  $\Phi$ , the uniform bound on  $J^n$  and the convergence proved in Lemma 22. For the second term on the r.h.s. of (6.13), recall that  $G$  is continuous and by Lemma 22  $J^n X_s^n(\omega)$  converges to  $X_s(\omega)$  for every  $s \in [t, T]$  and a.e.  $\omega \in \Omega$ ; then Assumption 2, estimates (4.3) and Propositions 9 and 13 yield

$$\mathbb{E} \int_t^T |G(r, J^n X_r^n, Y_r, Z_r) - G(r, X_r, Y_r, Z_r)|^p dr \xrightarrow{n \rightarrow \infty} 0 ,$$

thanks to Vitali convergence theorem. Note that, by Tonelli's theorem and the dominated convergence theorem,  $\mathbb{E} \int_s^T |Y_r^n - Y_r|^p dr \rightarrow 0$  for every  $s \in [t, T]$ ; so that

$$\mathbb{E} \left( \int_t^T |Y_r^n - Y_r|^2 dr \right)^{p/2} \xrightarrow{n \rightarrow \infty} 0 .$$

For what concerns  $Z^{n;t,x}$ , starting by (6.11) with  $p = 2$  it is easily seen that  $\mathbb{E} \int_t^T |Z_r^n - Z_r|^2 dr \rightarrow 0$ ; moreover,

$$\mathbb{E} \left( \int_s^T |Z_r^n - Z_r|^2 dr \right)^{\frac{p}{2}} \leq \left[ \mathbb{E} \int_s^T |Z_r^n - Z_r|^2 dr \right]^{\frac{1}{2}} \left[ \mathbb{E} \left( \int_s^T |Z_r^n - Z_r|^2 dr \right)^{p-1} \right]^{\frac{1}{2}} ,$$

so that the result holds for arbitrary  $p \geq 2$  since the last term can be estimated uniformly in  $n \in \mathbb{N}$  thanks to Proposition 9.

Let finally show the refined convergence  $Y^{n;t,x} \rightarrow Y^{t,x}$  in  $L^p(\Omega; C([t, T]; \mathbb{R}))$ . It is easy to show that

$$\mathbb{E} \left( \int_t^T |\hat{G}_r^n|^2 dr \right)^{p/2} \lesssim 1 + |x|^{mp} ;$$

hence we can apply estimate (2.4) in Proposition 1 to obtain

$$\mathbb{E} \sup_{s \in [t, T]} |Y_r^n - Y_r|^p + \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^2 dr \right)^{p/2} \lesssim \mathbb{E} \left( \int_t^T |\hat{G}_r^n|^2 dr \right)^{p/2} + \mathbb{E} |\Phi^n(X_T^n) - \Phi(X_T)|^p . \quad (6.14)$$

From Assumption 2 it holds

$$\begin{aligned} \mathbb{E} \left( \int_t^T |\hat{G}_r^n|^2 dr \right)^{p/2} &\lesssim \mathbb{E} \left( \int_t^T |Y_r^n - Y_r|^2 dr \right)^{p/2} + \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^2 dr \right)^{p/2} \\ &\quad + \mathbb{E} \left( \int_t^T |G(r, J^n X_r^n, Y_r, Z_r) - G(r, X_r, Y_r, Z_r)|^2 dr \right)^{p/2} , \end{aligned}$$

therefore the r.h.s. of (6.14) converges to 0 as  $n \rightarrow +\infty$  thanks to Lemma 24 and the uniform convergence of the  $X^n$ .  $\square$

**Lemma 25.** *Let  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}$ . For any  $h \in \mathcal{C}$ ,  $DY^{n;t,x} h \rightarrow DY^{t,x} h$  in  $L^p(\Omega; C([t, T]; \mathbb{R}))$  and  $DZ^{n;t,x} h \rightarrow DZ^{t,x} h$  in  $L^p(\Omega; L^2([t, T]; \mathbb{R}^{d_1}))$ .*

*Proof.* Similarly to the proof of Proposition 12, for any  $h \in \mathcal{C} \subset \mathcal{L}^2$ , we consider the equations satisfied by  $\Delta Y_r^n = (D_x Y_r^n - D_x Y_r) h$ ,  $\Delta Z_r^n = (D_x Z_r^n - D_x Z_r) h$ :

$$\Delta Y_s^n + \int_s^T \Delta Z_r^n dW_r = \eta^n + \int_s^T \alpha_r^n \Delta Y_r^n dr + \int_s^T \beta^n(r) dr + \int_s^T \gamma_r^n \Delta Z_r dr$$

where

$$\eta^n := [D_x \Phi^n(X_T^n) D_x X_T^n - D_x \Phi(X_t) D_x X_T] h ,$$

$$\begin{aligned}
\alpha_r^n &:= -D_2 G^n(r, X_r^n, Y_r^n, Z_r^n) , \\
\beta^n(r) &:= -[D_1 G^n(r, X_r^n, Y_r^n, Z_r^n) D_x X_r^n - D_1 G(r, X_r, Y_r, Z_r) D_x X_r] h \\
&\quad - [D_2 G^n(r, X_r^n, Y_r^n, Z_r^n) - D_2 G(r, X_r, Y_r, Z_r)] D_x Y_r h \\
&\quad - [D_3 G^n(r, X_r^n, Y_r^n, Z_r^n) - D_3 G(r, X_r, Y_r, Z_r)] D_x Z_r h \\
&=: -\beta_1^n(r) - \beta_2^n(r) - \beta_3^n(r) , \quad \gamma_r^n := -D_3 G^n(r, X_r^n, Y_r^n, Z_r^n) .
\end{aligned}$$

and

$$V_s^n = \int_t^s |\alpha_r^n| dr + \frac{1}{1 \wedge (p-1)} \int_t^s |\gamma_r^n|^2 dr , \quad (6.15)$$

By Lemma 2 we have, for every  $n \in \mathbb{N}$ , the estimate

$$\mathbb{E} \sup_{t \in [t, T]} \left| e^{V_t^n} \Delta Y_t^n \right|^p + \mathbb{E} \left( \int_t^T e^{2V_r^n} |\Delta Z_r^n|^2 dr \right)^{\frac{p}{2}} \lesssim \mathbb{E} \left| e^{V_t^n} \eta^n \right|^p + \mathbb{E} \left( \int_t^T e^{V_r^n} |\beta^n(r)| dr \right)^p ;$$

to get the desired convergence we need to show that the r.h.s. of the above inequality goes to 0 as  $n \rightarrow +\infty$ . By the uniform boundedness of  $V_s^n$ ,  $n \in \mathbb{N}$ ,  $s \in [t, T]$ , it holds, for  $p \geq 2$

$$\begin{aligned}
\mathbb{E} |\eta^n|^p &\lesssim \mathbb{E} |D\Phi^n(X_T^n) D_x X_T^n h - D\Phi(X_T) D_x X_T h|^p \\
&\lesssim \mathbb{E} |D\Phi^n(X_T^n) (D_x X_T^n h - D_x X_T h)|^p + \mathbb{E} |(D\Phi^n(X_T^n) - D\Phi(X_T)) D_x X_T h|^p \\
&= \mathbb{E} ([B_1^n]^p + [B_2^n]^p) .
\end{aligned}$$

Recalling (6.3) we have, by the equiboundedness of the  $J^n$ 's,

$$B_1^n \leq \|D\Phi(J^n X_T^n)\| |J^n D_x X_T^n h - J^n D_x X_T h| \lesssim \left( 1 + \sup_{s \in [t, T]} |X_s^n|^m \right) |D_x X_T^n h - D_x X_T h|$$

so that, by Lemma 23

$$\mathbb{E} ([B_1^n]^p) \lesssim \left[ \mathbb{E} \left( 1 + \sup_{s \in [t, T]} |X_s^n|^{2mp} \right) \right]^{\frac{1}{2}} \left[ \mathbb{E} |D_x X_T^n h - D_x X_T h|^{2p} \right]^{\frac{1}{2}} \xrightarrow{n \rightarrow \infty} 0 .$$

Concerning  $B_2^n$  we have

$$\begin{aligned}
\mathbb{E} (B_2^n)^p &\leq \mathbb{E} \|D\Phi(J^n X_T^n) - D\Phi(X_T)\|^p |J^n D_x X_T h|^p + \mathbb{E} |D\Phi(X_T) (J^n D_x X_T h - D_x X_T h)|^p \\
&\lesssim (1 + |x|^{mp}) |h|^p
\end{aligned}$$

Hence by Lemma 23, continuity of  $\Phi$ , equiboundedness of  $J^n$  and Assumption 4,  $\mathbb{E} (B_2^n)^p$  goes to zero, implying that  $\mathbb{E} |\eta^n|^p \rightarrow 0$ , as  $n \rightarrow +\infty$ .

To prove that  $\mathbb{E} \left( \int_0^T |\beta^n(r)| dr \right)^p \rightarrow 0$  first note that

$$\mathbb{E} \left( \int_0^T |\beta^n(r)| e^{\lambda V_r^n} dr \right)^p \lesssim \sum_{i=1}^3 \mathbb{E} \left( \int_0^T |\beta_i^n(r)| dr \right)^p .$$

We detail the computations for the term  $\beta_1^n(r)$ , the remaining terms being very similar.

$$\begin{aligned}
|\beta_1^n(r)| &= |D_1 G(r, J^n X_r^n, Y_r^n, Z_r^n) J^n D_x X_r^n h - D_1 G(r, X_r, Y_r, Z_r) D_x X_r h| \\
&\lesssim |D_1 G(r, J^n X_r^n, Y_r^n, Z_r^n) J^n (D_x X_r^n h - D_x X_r h)| \\
&\quad + |[D_1 G(r, J^n X_r^n, Y_r^n, Z_r^n) J^n - D_1 G(r, X_r, Y_r, Z_r)] D_x X_r h| \\
&= C_1^n(r) + C_2^n(r) .
\end{aligned}$$

Using Assumption 2 we get

$$\begin{aligned}
C_1^n(r) &\leq |D_1 G(r, J^n X_r^n, Y_r^n, Z_r^n)| |J^n| |D_x X_r^n h - D_x X_r h| \\
&\lesssim \left( 1 + \sup_{r \in [s, T]} |X_r^n|^m \right) \left( 1 + \sup_{r \in [s, T]} |Y_r^n| \right) |D_x X_r^n h - D_x X_r h| ,
\end{aligned}$$

therefore, reasoning as before,

$$\mathbb{E} \int_s^T [C_1^n(r)]^p dr \lesssim (1 + |x|^{2pm}) \left[ \mathbb{E} \left( \int_s^T |D_x X_r^n h - D_x X_r h| dr \right)^{3p} \right]^{\frac{1}{3}}$$

and the r.h.s. goes to 0 as  $n \rightarrow +\infty$  by Lemma 23. Then we have

$$\begin{aligned} C_2^n(r) &\leq |D_1 G(r, J^n X_r^n, Y_r^n, Z_r^n) J^n D_x X_r h - D_1 G(r, J^n X_r^n, Y_r^n, Z_r) J^n D_x X_r h| \\ &\quad + |D_1 G(r, J^n X_r^n, Y_r^n, Z_r) J^n D_x X_r h - D_1 G(r, J^n X_r^n, Y_r, Z_r) J^n D_x X_r h| \\ &\quad + |D_1 G(r, J^n X_r^n, Y_r, Z_r) J^n D_x X_r h - D_1 G(r, X_r, Y_r, Z_r) D_x X_r h| \\ &= C_{21}^n(r) + C_{22}^n(r) + C_{23}^n(r), \end{aligned}$$

thanks to Assumption 2 we can further bound

$$\begin{aligned} C_{21}^n(r) &= \left| \int_0^1 D_{1,3}^2 G(r, J^n X_r^n, Y_r^n, aZ_r^n + (1-a)Z_r) (Z_r^n - Z_r, J^n D_x X_r h) da \right| \\ &\lesssim \left( 1 + \sup_{r \in [s, T]} |X_r^n|^m \right) \left( 1 + \sup_{r \in [s, T]} |Y_r^n| \right) \left( \sup_{r \in [s, T]} |D_x X_r h| \right) |Z_r^n - Z_r| \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \left( \int_s^T C_{21}^n(r) dr \right)^p &\lesssim \left[ 1 + \mathbb{E} \sup_{r \in [0, T]} |X_r^n|^{4mp} \right]^{\frac{1}{4}} \left[ 1 + \mathbb{E} \sup_{r \in [0, T]} |Y_r^n|^{4p} \right]^{\frac{1}{4}} \\ &\quad \times \left[ \mathbb{E} \sup_{r \in [0, T]} |D_x X_r h|^{4p} \right]^{\frac{1}{4}} \left[ \mathbb{E} \left( \int_s^T |Z_r^n - Z_r|^2 dr \right)^{2p} \right]^{\frac{1}{4}} \\ &\lesssim (1 + |x|^{mp}) |h|^p (1 + |x|^{pm^2}) \left[ \mathbb{E} \left( \int_s^T |Z_r^n - Z_r|^2 dr \right)^{2p} \right]^{\frac{1}{4}} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

thanks to Lemma 24. The same holds for  $C_{22}^n(r)$ , while

$$\begin{aligned} C_{23}^n(r) &\leq |[D_1 G(r, J^n X_r^n, Y_r, Z_r) - D_1 G(r, X_r, Y_r, Z_r)] J^n D_x X_r h| \\ &\quad + |D_1 G(r, X_r, Y_r, Z_r) (J^n D_x X_r h - D_x X_r h)| \end{aligned}$$

which goes to 0 as  $n \rightarrow +\infty$  thanks to continuity of  $D_1 G$ , equiboundedness of  $J^n$ , Lemma 23 and Assumption 5. By the uniform bound (6.10) and Vitali convergence theorem also  $\mathbb{E} \left( \int_s^T C_{23}^n(r) dr \right)^p$  goes to 0 as  $n \rightarrow +\infty$ . This immediately yields that

$$\mathbb{E} \left( \int_s^T C_2^n(r) dr \right)^p \leq \sum_{i=1}^3 \mathbb{E} \left( \int_s^T C_{2i}^n(r) dr \right)^p \xrightarrow{n \rightarrow \infty} 0,$$

which concludes the proof.  $\square$

The next lemma concerns the convergence of second order derivatives and it is the most delicate one.

**Lemma 26.** *Let  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}$ . For any  $h, k \in \mathcal{C}$ ,  $D_x^2 Y^{n;t,x}(k, h) \rightarrow D_x^2 Y^{t,x}(k, h)$  in  $L^p(\Omega; C([t, T]; \mathbb{R}))$  and  $D_x^2 Z^{n;t,x}(k, h) \rightarrow D_x^2 Z^{t,x}(k, h)$  in  $L^p(\Omega; L^2([t, T]; \mathbb{R}^{d_1}))$ .*

*Proof.* Here we just focus on the most difficult (and interesting) term to deal with, the other terms are discussed in the Appendix. Recalling the equation satisfied by  $D_x^2 Y_s(k, h)$  (analogously by  $D_x^2 Y_s^n(k, h)$ ) and denoting

$$\Delta^2 Y_s^n(k, h) = D_x^2 Y_s^n(k, h) - D_x^2 Y_s(k, h), \quad \Delta^2 Z_s^n(k, h) = D_x^2 Z_s^n(k, h) - D_x^2 Z_s(k, h),$$

it holds that

$$\Delta^2 Y_s^n(k, h) + \int_s^T \Delta^2 Z_r^n(k, h) dW_r = \bar{\eta}^n + \int_s^T \bar{\alpha}_r^n \Delta^2 Y_r^n(k, h) dr + \int_s^T \bar{\beta}_r^n dr + \int_s^T \bar{\gamma}_r^n \Delta^2 Z_r^n(k, h) dr,$$

where  $\bar{\eta}^n$ ,  $\bar{\alpha}^n$ ,  $\bar{\gamma}^n$  and  $\bar{\beta}^n$  are suitable coefficients, whose definition is given in the Appendix. Exploiting the linear character of the equation (see also the proof of Proposition 15) we need to check that for some  $p \geq 2$

$$\mathbb{E} |\bar{\eta}^n|^p + \mathbb{E} \left( \int_t^T |\bar{\beta}_r^n| \, dr \right)^p \xrightarrow{n \rightarrow \infty} 0 .$$

Let us show how to deal with one of the term involved in the definition of  $\bar{\beta}^n$ , namely, for  $s \in [t, T]$ ,

$$\textcircled{12} := \int_s^T [D_{3,3}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x Z_r^n k, D_x Z_r^n h) - D_{3,3}^2 G(r, X_r, Y_r, Z_r) (D_x Z_r k, D_x Z_r h)] \, dr ,$$

where, for sake of consistency, we used the same notation as in the Appendix. Then

$$\begin{aligned} \mathbb{E} |\textcircled{12}|^p &\leq \mathbb{E} \left( \int_t^T |[D_{3,3}^2 G_r(n) - D_{3,3}^2 G_r(\cdot)] (D_x Z_r^n k, D_x Z_r^n h)| \, dr \right)^p \\ &\quad + \mathbb{E} \left( \int_t^T |D_{3,3}^2 G_r(\cdot) (D_x Z_r^n k, D_x Z_r^n h - D_x Z_r h)| \, dr \right)^p \\ &\quad + \mathbb{E} \left( \int_t^T |D_{3,3}^2 G_r(\cdot) (D_x Z_r^n k - D_x Z_r k, D_x Z_r h)| \, dr \right)^p \\ &= \mathbb{E} [\bar{F}_{331}^n]^p + \mathbb{E} [\bar{F}_{332}^n]^p + \mathbb{E} [\bar{F}_{333}^n]^p . \end{aligned}$$

By Assumptpion 2

$$\begin{aligned} \mathbb{E} [\bar{F}_{331}^n]^p &\lesssim \mathbb{E} \left( \int_t^T (|J^n X_r^n - X_r|^\alpha + |Y_r^n - Y_r|^\alpha + |Z_r^n - Z_r|^\alpha) |D_x Z_r^n k| |D_x Z_r^n h| \, dr \right)^p \\ &\lesssim \mathbb{E} \left[ \left( \sup_{r \in [t, T]} |J^n X_r^n - X_r|^{\alpha p} + \sup_{r \in [t, T]} |Y_r^n - Y_r|^{\alpha p} \right) \right. \\ &\quad \left. \times \left( \int_t^T |D_x Z_r^n k|^2 \, dr \right)^{p/2} \left( \int_t^T |D_x Z_r^n h|^2 \, dr \right)^{p/2} \right] \\ &\quad + \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r^n k| |D_x Z_r^n h| \, dr \right)^p \\ &\lesssim \left( \left[ \mathbb{E} \sup_{r \in [t, T]} |J^n X_r^n - X_r|^{\bar{\nu} \alpha p} \right]^{1/\bar{\nu}} + \left[ \mathbb{E} \sup_{r \in [t, T]} |Y_r^n - Y_r|^{\bar{\nu} \alpha p} \right]^{1/\bar{\nu}} \right) \\ &\quad \times \left[ \mathbb{E} \left( \int_t^T |D_x Z_r^n k|^2 \, dr \right)^{\nu p/2} \right]^{1/\nu} \left[ \mathbb{E} \left( \int_t^T |D_x Z_r^n h|^2 \, dr \right)^{\nu p/2} \right]^{1/\nu} \\ &\quad + \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r^n k| |D_x Z_r^n h| \, dr \right)^p . \end{aligned}$$

The first two terms go to 0 as  $n \rightarrow \infty$  by Lemmas 22 and 24, respectively. It remains to estimate

$$\begin{aligned} \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r^n k| |D_x Z_r^n h| \, dr \right)^p &\lesssim \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r k| |D_x Z_r h| \, dr \right)^p \\ &\quad + \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r^n k - D_x Z_r k| |D_x Z_r^n h - D_x Z_r h| \, dr \right)^p \\ &\quad + \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r k| |D_x Z_r^n h - D_x Z_r h| \, dr \right)^p \\ &\quad + \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r^n k - D_x Z_r k| |D_x Z_r h| \, dr \right)^p \end{aligned}$$

Thanks to (4.21), we get that

$$\begin{aligned}
& \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r^n k - D_x Z_r k| |D_x Z_r^n h - D_x Z_r h| \, dr \right)^p \\
& \leq \mathbb{E} \sup_{r \in [t, T]} |Z_r^n - Z_r|^\alpha \left( \int_t^T |D_x Z_r^n k - D_x Z_r k| |D_x Z_r^n h - D_x Z_r h| \, dr \right)^p \\
& \lesssim \left( \mathbb{E} \int_t^T |D_x Z_r^n k - D_x Z_r k|^2 \, dr \right)^{\frac{1}{2p}} \left( \mathbb{E} \int_t^T |D_x Z_r^n h - D_x Z_r h|^2 \, dr \right)^{\frac{1}{2p}} \xrightarrow{n \rightarrow \infty} 0,
\end{aligned}$$

where the last convergence follows by Lemma 25. Similarly,

$$\begin{aligned}
& \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r k| |D_x Z_r^n h - D_x Z_r h| \, dr \right)^p \\
& \leq \mathbb{E} \sup_{r \in [t, T]} |Z_r^n - Z_r|^\alpha \left( \int_t^T |D_x Z_r k| |D_x Z_r^n h - D_x Z_r h| \, dr \right)^p \\
& \lesssim \left( \mathbb{E} \int_t^T |D_x Z_r k|^2 \, dr \right)^{\frac{1}{2p}} \left( \mathbb{E} \int_t^T |D_x Z_r^n h - D_x Z_r h|^2 \, dr \right)^{\frac{1}{2p}} \xrightarrow{n \rightarrow \infty} 0,
\end{aligned}$$

and we can proceed in the same way for the term

$$\mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r^n k - D_x Z_r k| |D_x Z_r h| \, dr \right)^p.$$

It remains to study the convergence of

$$\mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^\alpha |D_x Z_r k| |D_x Z_r h| \, dr \right)^p \tag{6.16}$$

We first notice that for every  $\bar{q} \geq 1$

$$\mathbb{E} \sup_{r \in [t, T]} |D_x Z_r h|^{\bar{q}} < +\infty. \tag{6.17}$$

Indeed, by (4.20)

$$\begin{aligned}
D_x Z_r &= D_x [Du(t, X_r^{t,x}) \Sigma] \\
&= D^2 u(t, X_r^{t,x}) D_x X_r^{t,x} \Sigma,
\end{aligned}$$

and thanks to Corollary 17 and (4.6) we immediately get (6.17). Therefore, to show that (6.16) converges to 0 as  $n \rightarrow +\infty$ , by a standard application of Hölder inequality with respect to  $\omega$ , it is enough to prove that

$$\mathbb{E} \int_t^T |Z_r^n - Z_r|^q \, dr \longrightarrow 0, \tag{6.18}$$

for some  $q \geq 1$  (it actually holds for every  $q \geq 1$ ). If we write

$$\begin{aligned}
\mathbb{E} \int_t^T |Z_r^n - Z_r|^q \, dr &= \mathbb{E} \int_t^T |Du^n(r, X_r^n) \Sigma - Du(r, X_r) \Sigma|^q \, dr \\
&\leq \mathbb{E} \int_t^T |Du^n(r, X_r^n) \Sigma - Du^n(r, X_r) \Sigma|^q \, dr + \mathbb{E} \int_0^T |Du^n(r, X_r) \Sigma - Du(r, X_r) \Sigma|^q \, dr;
\end{aligned}$$

for what concerns the first term, by Corollary 17 we have

$$\begin{aligned}
& \mathbb{E} \sup_{r \in [t, T]} |Du^n(r, X_r^n) \Sigma - Du^n(r, X_r) \Sigma|^q \\
& \lesssim \left[ 1 + \mathbb{E} \sup_{r \in [t, T]} |X_r^n|^{2lq} + \mathbb{E} \sup_{r \in [t, T]} |X_r|^{2lq} \right]^{\frac{1}{2}} \left[ \mathbb{E} \sup_{r \in [t, T]} |X_r^n - X_r|^{2q} \right]^{\frac{1}{2}}
\end{aligned}$$

which converges to 0 as  $n$  goes to  $+\infty$  by (4.3) and Lemma 22.

As for the second term, the convergence provided by Lemma 25 yields

$$Du^n(s, X_s^{t,x}(\omega))\Sigma \xrightarrow{n \rightarrow +\infty} Du(s, X_s^{t,x}(\omega))\Sigma \quad \text{for a.e. } (s, \omega) \in [t, T] \times \Omega, \forall x \in \widehat{\mathcal{C}}.$$

Indeed, for every  $y \in \widehat{\mathcal{C}}$ ,  $D_x Y_s^{s,y}$  is deterministic and (4.20) along with Lemma 25 imply that for a.e.  $s \in [t, T]$

$$Du^n(s, y)\Sigma \rightarrow Du(s, y)\Sigma;$$

the required convergence follows by the evaluation  $y = X_s^{t,x}(\omega) \in \widehat{\mathcal{C}}$ , for a.e.  $\omega \in \Omega$ . A final application of Lebesgue dominated convergence theorem, together with (4.19) applied to  $u^n$  and  $u$  (we remark here that the constant in that estimate does not depend on  $n$ ) and the bound (4.6) yields

$$\mathbb{E} \int_t^T |Du^n(r, X_r)\Sigma - Du(r, X_r)\Sigma|^q dr \rightarrow 0,$$

hence the required convergence in (6.18).  $\square$

## 7. Application to stochastic optimal control

Here we apply the results obtained in the previous sections to semilinear Kolmogorov equations arising as Hamilton-Jacobi-Bellman (HJB) equations associated to some control problems.

We describe the evolution of the state with the forward controlled dynamics in  $\mathcal{D}$

$$\begin{cases} dX_s^u = AX_s^u ds + B(s, X_s^u) ds + \Sigma u_s ds + \Sigma dW_s, & s \in [t, T] \\ X_t^u = x, \end{cases} \quad (7.1)$$

where  $\Sigma : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^d \times \{0\} \subset \mathcal{D}$ ,  $B : \mathcal{D} \rightarrow \mathcal{D}$  is such that  $B(\mathcal{C}) \subseteq \mathcal{C}$ ,  $u : \Omega \times [0, T] \rightarrow \mathbb{R}^{d_1}$  is the control action and  $A, W$  are as in Subsection 3.1. As before, the solution to (7.1) has to be intended in mild sense and will be denoted also by  $X^{u;t,x}$  to emphasize the dependence on the initial data.

Besides equation (7.1) we define the cost functional  $\mathcal{J} : [0, T] \times \mathcal{D} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}$

$$\mathcal{J}(t, x, u) := \mathbb{E} \int_t^T [L(s, X_s^{u;t,x}) + Q(u_s)] ds + \mathbb{E}\Upsilon(X_T^{u;t,x}) \quad (7.2)$$

for real-valued functions  $L, Q, \Upsilon$ , defined respectively on  $[0, T] \times \mathcal{D}$ ,  $\mathbb{R}^{d_1}$  and  $\mathcal{D}$ , and the class of admissible controls

$$\mathcal{A} := \left\{ u : \Omega \times [0, T] \rightarrow \mathbb{R}^{d_1}, (\mathcal{F}_s)_s\text{-predictable} : \|u\|_{L^\infty((0,T) \times \Omega)} < +\infty \right\}.$$

The control problem consists in minimizing the functional  $\mathcal{J}$  over the admissible controls  $u \in \mathcal{A}$ .

**Remark 27.** For example, our control problem arises as infinite-dimensional lifting of a finite-dimensional path-dependent control problem. Indeed, let  $\xi^u$  be a solution to the path-dependent state equation

$$\begin{cases} d\xi^u(s) = b_s(\xi_{[0,s]}^u) ds + \sigma u_s ds + \sigma dW_s, & s \in [t, T], \\ \xi_{[0,t]}^u = \gamma, \end{cases} \quad (7.3)$$

with  $b_s : D([0, s]; \mathbb{R}^d) \rightarrow \mathbb{R}$ , for every  $s \in [t, T]$ ,  $\sigma : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^d$  and  $\gamma \in D([0, t]; \mathbb{R}^d)$  fixed. To obtain the forward SDE (7.1) we just set  $x := L^t \gamma$  (with  $L^t$  as in (3.5)) and define  $B, \Sigma$  as in (3.4), (3.8), respectively (see Subsection 3.1 for more details). Furthermore, given a path-dependent functional

$$l = \{l_s\}_{s \in [t, T]}, \quad l_s : D([0, s]; \mathbb{R}^d) \rightarrow \mathbb{R}$$

and functions

$$q : \mathbb{R}^{d_1} \rightarrow \mathbb{R}, \quad \varphi : D([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R},$$

we can define the path-dependent cost functional

$$j(t, \gamma, u) := \mathbb{E} \int_t^T [l_s(\xi_{[0,s]}^u) + q(u_s)] ds + \mathbb{E}\varphi(x_{[0,T]}^u). \quad (7.4)$$

Introducing the liftings

$$\begin{aligned} L : [0, T] \times \mathcal{D} &\rightarrow \mathbb{R}, & L(s, x) &= l_s(M_s x), \\ Q = q, \quad \Upsilon : \mathcal{D} &\rightarrow \mathbb{R}, & \Upsilon(x) &= \varphi(M_T x), \end{aligned}$$

we exactly recover (7.2) with the property  $\mathcal{J}(t, x, u) = j(t, M_t x, u)$ , for every  $(t, x, u) \in [0, T] \times \mathcal{D} \times \mathcal{A}$ . For the sake of generality, in the following we deal with abstract problems in  $\mathcal{D}$ , without exploiting the path-dependent structure behind it.



The following assumptions on the optimal control problem will be in force throughout.

**Assumption 6.** *There exists  $\alpha \in (0, 1)$  and constants  $a > 0, b, c, R, C \geq 0$  such that*

(J.I)  $Q : \mathbb{R}^{d_1} \rightarrow \mathbb{R}$  is continuous and

$$|Q(u)| \geq a|u|^2 - b, \quad \forall u \in \mathbb{R}^{d_1}, \quad |Q(u)| \leq c|u|^2, \text{ for } |u| \geq R; \quad (7.5)$$

(J.II)  $L : [0, T] \times \mathcal{D} \rightarrow \mathbb{R}$  is continuous,  $L(s, \cdot)$  belongs to  $C^{2,\alpha}(\mathcal{D}, \mathbb{R})$  for every  $s \in [0, T]$  and

$$|L(s, x)| \leq C(1 + |x|), \quad |DL(s, x)| + |D^2L(s, x)| \leq C, \quad \forall s \in [0, T];$$

(J.III)  $\Upsilon : \mathcal{D} \rightarrow \mathbb{R}$  belongs to  $C^{2,\alpha}(\mathcal{D}, \mathbb{R})$  and

$$|\Upsilon(x)| + |D\Upsilon(x)| + |D^2\Upsilon(x)| \leq C;$$

The Hamiltonian of the problem is defined as

$$\mathcal{H}(z) := \inf_{u \in \mathbb{R}^{d_1}} \{Q(u) + zu\} \quad \forall z \in \mathbb{R}^{d_1},$$

and we denote with  $\Gamma(z)$  the set of minimizers

$$\Gamma(z) = \{u : \mathcal{H}(z) = Q(u) + zu\}.$$

**Assumption 7.** *The Hamiltonian  $\mathcal{H} : \mathbb{R}^{d_1} \rightarrow \mathbb{R}$  belongs to  $C^{2,\alpha}(\mathbb{R}^{d_1})$ .*

Note that  $\mathcal{H}(z) = -Q^*(-z)$ ,  $Q^*$  being the Fenchel conjugate of  $Q$ , and from Assumption 6 it is easily seen that  $\mathcal{H}$  has quadratic growth. Moreover, a sufficient condition for Assumption 7 to hold is to require  $Q \in C^3(\mathbb{R}^{d_1})$  strictly convex (along with the superlinearity given by Assumption 6), see e.g. (Fathi, 2008, Prop. 2.6.3) for a general result.

Denoting with  $\Psi : [0, T] \times \mathcal{D} \times \mathbb{R} \rightarrow \mathbb{R}$  the map  $\Psi(t, x, z) := L(t, x) + \mathcal{H}(z)$ , let us now introduce the BSDE

$$\begin{cases} dY_s^{t,x} = \Psi(s, X_s^{t,x}, Z_s^{t,x}) ds + Z_s^{t,x} dW_s, & s \in [t, T], \\ Y_T^{t,x} = \Upsilon(X_T^{t,x}); \end{cases} \quad (7.6)$$

where  $X_s^{t,x}$  solves the forward state equation (7.1) for every  $t, x \in [0, T] \times \mathcal{D}$  with  $u = 0$ .

**Proposition 28.** *Let Assumptions 1, 6 and 7 hold; then for every  $(t, x) \in [0, T] \times \mathcal{D}$  the BSDE (7.6) admits a unique solution  $(Y^{t,x}, Z^{t,x}) \in \mathcal{K}_p$ , for every  $p > 1$ . Moreover, the map*

$$x \mapsto Y^{t,x}, \quad \mathcal{D} \rightarrow L^p(\Omega; C([0, T]; \mathbb{R}))$$

*is twice differentiable and there exists  $K \geq 0$  such that*

$$|Z_s^{t,x}| \leq K|\Sigma|, \quad \forall s \in [t, T], \quad \mathbb{P}\text{-a.s.}$$

*In addition, if  $|D^2B(s, x)| \leq C$  for every  $(s, x) \in [0, T] \times \mathcal{D}$ , there exists  $c \geq 0$  such that*

$$\mathbb{E} \sup_{s \in [t, T]} \|D_x^2 Y_s^{t,x}\| \leq c. \quad (7.7)$$

*Proof.* Let us start by setting  $\mathcal{H}_M = \mathcal{H}(\rho_M(\cdot))$ , where  $\rho_M(z)$  is a smooth function such that  $\rho(z) = z$  if  $|z| < M$  and  $\rho(z) = 0$  if  $|z| > M + 1$ . The truncated Hamiltonian  $\mathcal{H}_M \in C^{2,\alpha}(\mathbb{R}^{d_1})$  has bounded derivatives,  $\Psi_M(t, x, z) := L(t, x) + \mathcal{H}_M(z)$  complies with Assumption 2 and thanks to Proposition 9 the BSDE

$$\begin{cases} dY_s^{M;t,x} = \Psi_M(s, X_s^{t,x}, Z_s^{M;t,x}) ds + Z_s^{M;t,x} dW_s, & s \in [t, T], \\ Y_T^{M;t,x} = \Upsilon(X_T^{t,x}), \end{cases} \quad (7.8)$$

admits a unique solution  $(Y^{M;t,x}, Z^{M;t,x})$  in  $\mathcal{K}_p$ . In view of Assumptions 6 and 7, the application of Propositions 12 and 15 guarantees that the map

$$x \mapsto Y^{M;t,x}, \quad \mathcal{D} \rightarrow L^p(\Omega; C([0, T]; \mathbb{R}))$$

is twice differentiable. Furthermore, thanks to Remark 14 it follows that

$$|Z_s^{M;t,x}| \leq K|\Sigma|, \quad \forall s \in [t, T], \quad \mathbb{P}\text{-a.s.},$$

for some constant  $K \geq 0$  which is *independent* on  $M$ . Thus, choosing  $M > K|\Sigma|$ , it follows that  $\mathcal{H}_M(Z_s^{t,x}) = \mathcal{H}(Z_s^{t,x})$  for every  $s \in [t, T]$  and the pair  $(Y^{M;t,x}, Z^{M;t,x})$  solves also equation (7.6). By the uniqueness of solutions of (7.6) with bounded second component it easily follows that  $(Y^{M;t,x}, Z^{M;t,x}) \equiv (Y^{t,x}, Z^{t,x})$  whenever  $M > K|\Sigma|$ , hence the boundedness of  $Z^{t,x}$ .

Concerning estimate (7.7), fixing  $\bar{M} > K|\Sigma|$  big enough we firstly observe that  $D_x^i Y^{\bar{M};t,x} = D_x^i Y^{t,x}$ , for  $i = 1, 2$ , so that we can directly work with  $\mathcal{H}$  instead of  $\mathcal{H}_{\bar{M}}$ . Then, for every  $h \in \mathcal{D}$ , the application of estimate (2.9) to the linear BSDE solved by the pair  $(DY^{t,x}h, DZ^{t,x}h)$ , immediately gives  $D_x Z^{t,x} \in L^p(\Omega; L^2(0, T; \mathbb{R}^{d_1}))$ . Moreover, for every  $(k, h) \in \mathcal{D}$  the equation for the second derivatives reads as

$$\begin{aligned} D_x^2 Y_s^{t,x}(k, h) + \int_s^T D_x^2 Z_r^{t,x}(k, h) dW(r) &= \Xi_T(k, h) - \int_s^T [D^2 L(r, X^{u;t,x}) (D_x X_r^{t,x} k, D_x X_r^{t,x} h) \\ &+ DL(r, X^{u;t,x}) D_x^2 X_r^{t,x}(k, h) + D^2 \mathcal{H}(Z_r^{t,x}) (D_x Z_r^{t,x} k, D_x Z_r^{t,x} h)] dr \\ &- \int_s^T [D\mathcal{H}(Z_r^{t,x}) D_x^2 Z_r^{t,x}(k, h)] dr, \end{aligned} \quad (7.9)$$

where  $\Xi_T(k, h) := D^2 \Upsilon(X_T^{t,x})(D_x X_T^{t,x} k, D_x X_T^{t,x} h) + D\Upsilon(X_T^{t,x}) D_x^2 X_T^{t,x}(k, h)$  is uniformly bounded thanks to Assumption 6 and estimate (4.6).

The boundedness of  $DL, D^2L$  and the uniform bound on  $D_x^2 X^{t,x}$  given by (4.8) (with  $m = 0$ ) finally guarantee the validity of (7.7) by the application of estimate (2.9) to equation (7.9). This concludes the proof.  $\square$

For every  $(t, x) \in [0, T] \times \mathcal{D}$ , let us now introduce the value function  $v$  associated to the cost functional  $\mathcal{J}$ :

$$v(t, x) = \inf_{u \in \mathcal{A}} \mathcal{J}(t, x, u),$$

and consider the associated HJB equation

$$\begin{cases} \frac{\partial v}{\partial t}(t, x) + Dv(t, x) [Ax + B(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 v(t, x)] = \Psi(t, x, Dv(t, x) \Sigma), \\ v(T, \cdot) = \Upsilon. \end{cases} \quad (7.10)$$

This is a semilinear Kolmogorov equation with the same structure as equation (1.1), for which we already obtained a wellposedness result.

**Proposition 29.** *Let  $B$  satisfies Assumptions 1, 4 with  $E = \mathcal{D}$  and suppose that  $B$  maps  $\mathcal{C}$  into itself. Let also Assumptions 6, 7 hold and suppose that  $\Upsilon$  has one-jump-continuous Fréchet differential of first and second order on  $\widehat{\mathcal{C}} \subset \mathcal{D}$ . Then the function  $v(t, x) := Y_t^{t,x}$  (where  $Y^{t,x}$  solves (7.6)) is a classical solution to the HJB equation (7.10) in the sense of Definition 4, for every  $t \in [0, T]$  and  $x \in \widehat{\mathcal{C}}^1$ .*

*Proof.* Recall that  $\Psi_M := L(t, x) + \mathcal{H}_M(z)$  satisfies Assumption 2. Hence, for  $M > K|\Sigma|$ , Theorem 20 and Proposition 28 guarantee that  $v(t, x) := Y_t^{M;t,x} = Y_t^{t,x}$  is a classical solution of the HJB equation

$$\begin{cases} \frac{\partial v}{\partial t}(t, x) + Dv(t, x) [Ax + B(t, x)] + \frac{1}{2} \text{Tr}_{\mathbb{R}^d} [\Sigma \Sigma^* D^2 v(t, x)] = \Psi_M(t, x, Dv(t, x) \Sigma), \\ v(T, \cdot) = \Upsilon. \end{cases}$$

Since  $|Dv(t, x) \Sigma| \leq K|\Sigma|$ ,  $\Psi_M \equiv \Psi$  and  $v$  is also a classical solution of (7.10) in the sense of Definition 4.  $\square$

In order to derive the so-called *fundamental relation* for the value function, it is useful to introduce a family of auxiliary problems. For every  $\Lambda \in \mathbb{R}$ , we define

$$\mathcal{A}^\Lambda := \left\{ u : \Omega \times [0, T] \rightarrow \mathbb{R}^{d_1}, (\mathcal{F}_s)_s\text{-predictable} : \|u\|_{L^\infty((0, T) \times \Omega)} \leq \Lambda \right\},$$

with the corresponding value function and Hamiltonian

$$v^\Lambda(t, x) = \inf_{u \in \mathcal{A}^\Lambda} \mathcal{J}(t, x, u), \quad \mathcal{H}^\Lambda(z) = \inf_{u \in \mathbb{R}^{d_1}, |u| \leq \Lambda} \{Q(u) + zu\} \quad \forall z \in \mathbb{R}^{d_1}.$$

**Remark 30.** *Note that  $\mathcal{H}^\Lambda$  is Lipschitz but, in general, does not belong to  $C^{2,\alpha}(\mathbb{R}^{d_1})$ . A counterexample is given by  $Q(u) = \frac{1}{2}|u|^2$ , for which  $-\mathcal{H}(z) = \frac{1}{2}|z|^2$  and*

$$-\mathcal{H}^\Lambda(z) = \begin{cases} \frac{1}{2}|z|^2 & |z| \leq \Lambda, \\ \Lambda|z| - \frac{\Lambda^2}{2} & |z| > \Lambda; \end{cases} \quad (7.11)$$

*which is not  $C^2$ -regular on  $|z| = \Lambda$ .*

**Proposition 31.** *Under the same assumptions of Proposition 29, for every  $t \in [0, T]$ ,  $x \in \widehat{\mathcal{C}}^1$  and for every  $u \in \mathcal{A}$  it holds*

$$v(t, x) = \mathcal{J}(t, x, u) + \mathbb{E} \int_t^T [\mathcal{H}(Dv(X_s^u)\Sigma) - Dv(X_s^u)\Sigma u_s - Q(u_s)] ds, \quad (7.12)$$

so that  $v(t, x) \leq \mathcal{J}(t, x, u)$  and equality holds if and only if  $u_s \in \Gamma(Dv(s, X_s^u)\Sigma)$  for a.e.  $s \in [t, T]$ ,  $\omega \in \Omega$ . Furthermore, if  $|D^2B(s, x)| \leq C$  for every  $(s, x) \in [0, T] \times \mathcal{D}$ , and  $\Gamma_0 : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_1}$  is a measurable selection of  $\Gamma$  with  $\Gamma_0$  Lipschitz continuous, then the closed-loop equation

$$\begin{cases} dX_s^u = AX_s^u ds + B(s, X_s^u) ds + \Sigma \Gamma_0(Dv(s, X_s^u)\Sigma) ds + \Sigma dW_s, & s \in [t, T], \\ X_t^u = x, \end{cases} \quad (7.13)$$

admits a unique solution denoted by  $X^*$ , and the pair  $(X^*, \Gamma_0(Dv(\cdot, X^*)\Sigma))$  is optimal.

*Proof.* For  $M > K|\Sigma|$  there exists  $\Lambda \geq 0$  such that  $\mathcal{H}_M(z) = \mathcal{H}^\Lambda(z)$  for every  $|z| \leq M$ . Moreover, there exists an increasing function  $\tilde{\rho} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  with  $\lim_{\Lambda \rightarrow +\infty} \tilde{\rho}(\Lambda) = +\infty$  such that  $\mathcal{H}^\Lambda(z) = \mathcal{H}(z)$  for every  $|z| < \tilde{\rho}(\Lambda)$ . Thus, for  $\Lambda, M \geq 0$  with  $\tilde{\rho}(\Lambda) > M$  it holds

$$\mathcal{H}_M(z) = \mathcal{H}^\Lambda(z) = \mathcal{H}(z), \quad \forall |z| \leq M. \quad (7.14)$$

Thanks to Proposition 28,  $(Y^{t,x}, Z^{t,x})$  solves equation (7.6) either with  $\Psi$ ,  $\Psi_M$  or  $\Psi^\Lambda := L + \mathcal{H}^\Lambda$ . Then,  $v(t, x) := Y_t^{t,x}$  is a classical solution of (7.10), and by (Fuhrman and Tessitore, 2002, Thm. 7.2) (see also (Fuhrman et al., 2010, Section 6)) it can be easily proved that equality (7.12) holds for every  $u \in \mathcal{A}^\Lambda$ . The extension to any  $u \in \mathcal{A}$  follows again by (7.14).

For what concerns the closed loop equation (7.13), existence of a (unique) solution follows by the Lipschitz continuity of the selection  $\Gamma_0$  along with the estimate (7.7) applied to  $D^2v$  (which in turn ensures the Lipschitz continuity of  $Dv$ ).  $\square$

## 8. Appendix

We collect here a detailed version of the proofs of Proposition 15, Lemma 23 and Lemma 26.

*Proof of Proposition 15.* We detail here the Fréchet differentiability of the map  $x \mapsto (D_x Y^{t,x} h, D_x Z^{t,x} h)$ . We fix  $h, k \in E$  and use the equations solved by  $D_x Y_s^{t,x+k} h$ ,  $D_x Y_s^{t,x} h$  and  $F_s^{t,x}(h, k)$  to write:

$$\begin{aligned} & [D_x Y_s^{t,x+k} h - D_x Y_s^{t,x} h - F_s^{t,x}(h, k)] + \int_s^T [D_x Z_s^{t,x+k} h - D_x Z_s^{t,x} h - H_s^{t,x}(h, k)] dW(r) \\ &= \left[ U_T^{t,x+k} h - U_T^{t,x} h - F_T^{t,x}(k, h) \right] \\ & \quad - \int_s^T [D_1 G_r(t, x+k) D_x X_r^{t,x+k} h - D_1 G_r(t, x) D_x X_r^{t,x} h - L_{1;r}^{t,x}(k, h)] dr \\ & \quad - \int_s^T [D_2 G_r(t, x+k) D_x Y_r^{t,x+k} h - D_2 G_r(t, x) D_x Y_r^{t,x} h - L_{2;r}^{t,x}(k, h) - D_2 G_r(t, x) F_r^{t,x}(k, h)] dr \\ & \quad - \int_s^T [D_3 G_r(t, x+k) D_x Z_r^{t,x+k} h - D_3 G_r(t, x) D_x Z_r^{t,x} h - L_{3;r}^{t,x}(k, h) - D_3 G_r(t, x) H_r^{t,x}(k, h)] dr \\ &= \left[ U_T^{t,x+k} h - U_T^{t,x} h - F_T^{t,x}(k, h) \right] \\ & \quad - \int_s^T [D_1 G_r(t, x+k) D_x X_r^{t,x+k} h - D_1 G_r(t, x) D_x X_r^{t,x} h - L_{1;r}^{t,x}(k, h)] dr \\ & \quad - \int_s^T [(D_2 G_r(t, x+k) - D_2 G_r(t, x)) D_x Y_r^{t,x+k} h - L_{2;r}^{t,x}(k, h)] dr \\ & \quad - \int_s^T D_2 G_r(t, x) [D_x Y_r^{t,x+k} h - D_x Y_r^{t,x} h - D_2 G_r(t, x) F_r^{t,x}(k, h)] dr \\ & \quad - \int_s^T [(D_2 G_r(t, x+k) - D_3 G_r(t, x)) D_x Z_r^{t,x+k} h - L_{3;r}^{t,x}(k, h)] dr \\ & \quad - \int_s^T D_3 G_r(t, x) [D_x Z_r^{t,x+k} h - D_x Z_r^{t,x} h - D_3 G_r(t, x) H_r^{t,x}(k, h)] dr. \end{aligned}$$

If we divide both left and right hand side by  $|k|$  and we shorten the notation as in (4.27) we end up with

$$\Upsilon_s^k + \int_s^T \Psi_r^k dW_r = \Upsilon_T^k - \int_s^T (D_2 G_r(t, x) \Upsilon_r^k + D_3 G_r(t, x) \Psi_r^k - M^k(r) dr) dr .$$

Exploiting estimate (2.9) in Lemma 2 we have to show that

$$\lim_{k \rightarrow 0} \sup_{|h|=1} \left[ \mathbb{E} |\Upsilon_T^k|^p + \mathbb{E} \left( \int_t^T |M^k(r)| dr \right)^p \right] = 0 .$$

Firstly, dealing with the final datum, let us set for convenience  $\tilde{\Upsilon}_T^k := |k| \Upsilon_T^k$ ; then almost surely

$$\begin{aligned} \tilde{\Upsilon}_T^k &= U_T^{t,x+k} h - U_T^{t,x} h - F_T^{t,x}(k, h) \\ &= D\Phi(X_T^{t,x+k}) D_x X_T^{t,x+k} h - D\Phi(X_T^{t,x}) D_x X_T^{t,x} h \\ &\quad - D^2\Phi(X_T^{t,x}) (D_x X_T^{t,x} k, D_x X_T^{t,x} h) - D\Phi(X_T^{t,x}) D_x^2 X_T^{t,x}(k, h) \\ &= \left[ D\Phi(X_T^{t,x+k}) - D\Phi(X_T^{t,x}) \right] D_x X_T^{t,x+k} h - D^2\Phi(X_T^{t,x}) (D_x X_T^{t,x} k, D_x X_T^{t,x} h) \\ &\quad + D\Phi(X_T^{t,x}) \left[ D_x X_T^{t,x+k} h - D_x X_T^{t,x} h - D_x^2 X_T^{t,x}(k, h) \right] \\ &=: \tilde{\Upsilon}_{T,1}^k + \tilde{\Upsilon}_{T,2}^k . \end{aligned} \tag{8.1}$$

For what concerns  $\Upsilon_{T,1}^k$  we use Assumption 3 to write

$$\begin{aligned} \tilde{\Upsilon}_{T,1}^k &= \int_0^1 \left[ D^2\Phi(\lambda X_T^{t,x+k} + (1-\lambda) X_T^{t,x}) - D^2\Phi(X_T^{t,x}) \right] \left( X_T^{t,x+k} - X_T^{t,x}, D_x X_T^{t,x+k} h \right) d\lambda \\ &\quad + D^2\Phi(X_T^{t,x}) \left( X_T^{t,x+k} - X_T^{t,x} - D_x X_T^{t,x} k, D_x X_T^{t,x+k} h \right) \\ &\quad + D^2\Phi(X_T^{t,x}) \left( D_x X_T^{t,x} k, D_x X_T^{t,x+k} h - D_x X_T^{t,x} h \right) \\ &=: \tilde{\Upsilon}_{T,11}^k + \tilde{\Upsilon}_{T,12}^k + \tilde{\Upsilon}_{T,13}^k . \end{aligned} \tag{8.2}$$

Using  $\alpha$ -Hölder continuity of  $D^2\Phi(\cdot)$  (see Assumption 3),  $\Upsilon_{T,11}^k$  can be estimated as follows:

$$\begin{aligned} \mathbb{E} |\Upsilon_{T,11}^k|^p &\lesssim \left( \mathbb{E} \left| X_T^{t,x+k} - X_T^{t,x} \right|^{4p\alpha} \right)^{1/4} \mathbb{E} \left( \left| D_x X_T^{t,x+k} \right|^{4p} \right)^{1/4} |h|^p \\ &\quad \times \left[ \mathbb{E} \left( \frac{\left| X_T^{t,x+k} - X_T^{t,x} \right|}{|k|} \right)^{2p} \right]^{1/2} \longrightarrow 0 , \quad \text{if } |k| \rightarrow 0 , \end{aligned} \tag{8.3}$$

thanks to the Fréchet differentiability of  $x \mapsto X_T^{t,x}$  and estimate (4.7) (see Theorem 8). A similar argument can be used for  $\Upsilon_{T,12}^k$  and  $\Upsilon_{T,13}^k$ :

$$\begin{aligned} \mathbb{E} |\Upsilon_{T,12}^k|^p &\lesssim (1 + \mathbb{E} |X_T^{t,x}|^{4p})^{1/4} \left[ \mathbb{E} \left( \frac{\left| X_T^{t,x+k} - X_T^{t,x} - D_x X_T^{t,x} h \right|}{|k|} \right)^{4p} \right]^{1/4} \\ &\quad \times |h|^p \left[ \mathbb{E} \left| D_x X_T^{t,x+k} \right|^{2p} \right]^{1/2} \longrightarrow 0 , \quad \text{if } |k| \rightarrow 0 , \end{aligned} \tag{8.4}$$

$$\begin{aligned} \mathbb{E} |\Upsilon_{T,13}^k|^p &\lesssim (1 + \mathbb{E} |X_T^{t,x}|^{4p})^{1/4} \mathbb{E} \left( \left| D_x X_T^{t,x+k} - D_x X_T^{t,x} \right|^{4p} \right)^{1/4} |h|^p \\ &\quad \times \left[ \mathbb{E} \left\| D_x X_T^{t,x} \right\|^{2p} \right]^{1/2} \longrightarrow 0 , \quad \text{if } |k| \rightarrow 0 , \end{aligned} \tag{8.5}$$

where we used Assumption 3, estimate (4.3), continuity of the map  $x \mapsto D_x X_T^{t,x}$  and the bound (4.7).

Finally,  $\Upsilon_{T,2}^k$  can be shown to go to zero, as  $|k| \rightarrow 0$ , thanks to the second Fréchet differentiability of  $x \mapsto X_T^{t,x}$  along with the bound (4.3). Indeed

$$\sup_{|h|=1} \mathbb{E} |\Upsilon_{T,2}^k|^p \lesssim (1 + \mathbb{E} |X_T^{t,x}|^{2p})^{1/2} \left[ \mathbb{E} \left( \frac{\left| D_x X_T^{t,x+k} - D_x X_T^{t,x} - D_x^2 X_T^{t,x} k \right|}{|k|} \right)^{2p} \right]^{1/2} \longrightarrow 0 , \quad \text{if } |k| \rightarrow 0 .$$

In the rest of the proof we only concentrate on  $M_3^k$ , being the more intricate term (it involves the derivatives  $D_x Z$  for which the topology of the estimates is weaker). The other terms can be treated in a similar, and simpler, way.

To lighten the notation we introduce the shorthand  $\mathcal{M}_3^k := -|k|M_i^k(r)$ .

$$\begin{aligned}
\mathcal{M}_3^k &= [D_3 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, Z_r^{t,x+k}) - D_3 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, Z_r^{t,x})] D_x Z_r^{t,x+k} h \\
&\quad + D_{3,3}^2 G_r(t, x) (D_x Z_r^{t,x} k, D_x Z_r^{t,x} h) \\
&\quad + [D_3 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, Z_r^{t,x}) - D_3 G(r, X_r^{t,x+k}, Y_r^{t,x}, Z_r^{t,x})] D_x Z_r^{t,x+k} h \\
&\quad + D_{3,2}^2 G_r(t, x) (D_x Y_r^{t,x} k, D_x Z_r^{t,x} h) \\
&\quad + [D_3 G(r, X_r^{t,x+k}, Y_r^{t,x}, Z_r^{t,x}) - D_3 G(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})] D_x Z_r^{t,x+k} h \\
&\quad + D_{1,3}^2 G_r(t, x) (D_x X_r^{t,x} k, D_x Z_r^{t,x} h) \\
&=: \mathcal{M}_{31}^k + \mathcal{M}_{32}^k + \mathcal{M}_{33}^k.
\end{aligned} \tag{8.6}$$

Thanks to the Fréchet differentiability of  $D_3 G$ , the first term can be rewritten as

$$\begin{aligned}
\mathcal{M}_{31}^k &= \int_0^1 [D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, \lambda Z_r^{t,x+k} + (1-\lambda)Z_r^{t,x}) - D_{3,3}^2 G_r(t, x)] \\
&\quad (Z_r^{t,x+k} - Z_r^{t,x}, D_x Z_r^{t,x+k} h) d\lambda + D_{3,3}^2 G_r(t, x) (Z_r^{t,x+k} - Z_r^{t,x}, D_x Z_r^{t,x+k} h) \\
&\quad - D_{3,3}^2 G_r(t, x) (D_x Z_r^{t,x} k, D_x Z_r^{t,x+k} h) + D_{3,3}^2 G_r(t, x) (D_x Z_r^{t,x} k, D_x Z_r^{t,x+k} h - D_x Z_r^{t,x} h)
\end{aligned}$$

from which

$$\begin{aligned}
\mathcal{M}_{31}^k &= \int_0^1 [D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, \lambda Z_r^{t,x+k} + (1-\lambda)Z_r^{t,x}) - D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, Z_r^{t,x})] \\
&\quad (Z_r^{t,x+k} - Z_r^{t,x}, D_x Z_r^{t,x+k} h) d\lambda \\
&\quad + [D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x+k}, Z_r^{t,x}) - D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x}, Z_r^{t,x})] (Z_r^{t,x+k} - Z_r^{t,x}, D_x Z_r^{t,x+k} h) \\
&\quad + [D_{3,3}^2 G(r, X_r^{t,x+k}, Y_r^{t,x}, Z_r^{t,x}) - D_{3,3}^2 G_r(t, x)] (Z_r^{t,x+k} - Z_r^{t,x}, D_x Z_r^{t,x+k} h) \\
&\quad + D_{3,3}^2 G_r(t, x) (Z_r^{t,x+k} - Z_r^{t,x} - D_x Z_r^{t,x} k, D_x Z_r^{t,x+k} h) \\
&\quad + D_{3,3}^2 G_r(t, x) (D_x Z_r^{t,x} k, D_x Z_r^{t,x+k} h - D_x Z_r^{t,x} h) \\
&=: \mathcal{M}_{311}^k + \mathcal{M}_{312}^k + \mathcal{M}_{313}^k + \mathcal{M}_{314}^k + \mathcal{M}_{315}^k.
\end{aligned}$$

The term  $\mathcal{M}_{311}^k$  has been already discussed in the main proof. For what concerns  $\mathcal{M}_{314}^k$  and  $\mathcal{M}_{315}^k$  we have

$$\begin{aligned}
\mathbb{E} \left( \int_t^T |M_{314}^k(r)| dr \right)^p &\lesssim \mathbb{E} \left( \int_t^T |Z_r^{t,x+k} - Z_r^{t,x} - D_x Z_r^{t,x} k| |D_x Z_r^{t,x+k} h| dr \right)^p \\
&\lesssim \left[ \mathbb{E} \left( \int_t^T \frac{|Z_r^{t,x+k} - Z_r^{t,x} - D_x Z_r^{t,x} k|^2}{|k|^2} dr \right)^p \right]^{1/2} \left[ \mathbb{E} \left( \int_t^T \|D_x Z_r^{t,x+k}\|^2 dr \right)^p \right]^{1/2} |h| \longrightarrow 0, \quad \text{if } |k| \rightarrow 0,
\end{aligned} \tag{8.7}$$

uniformly in  $h$ ,  $|h| = 1$ , thanks to the Fréchet differentiability of  $x \mapsto Z^{t,x}$  and estimate (4.13). Finally

$$\begin{aligned}
\mathbb{E} \left( \int_t^T |M_{315}^k(r)| dr \right)^p &\lesssim \left[ \mathbb{E} \left( \int_t^T |D_x Z_r^{t,x+k} h - D_x Z_r^{t,x} h|^2 dr \right)^p \right]^{1/2} \\
&\quad \times \left[ \mathbb{E} \left( \int_t^T \|D_x Z_r^{t,x+k}\|^2 dr \right)^p \right]^{1/2} \longrightarrow 0, \quad \text{if } |k| \rightarrow 0,
\end{aligned} \tag{8.8}$$

thanks to the continuity of the map  $x \mapsto D_x Z^{t,x} h$  from  $E \rightarrow \mathcal{K}_p$  given in Proposition 12 and the estimate (4.13).

Coming back to  $\mathcal{M}_3^k$ , the terms  $\mathcal{M}_{32}^k$  and  $\mathcal{M}_{33}^k$  can be treated in the same way as  $\mathcal{M}_{31}^k$  (in this cases there is no need for the identification theorem), taking advantage of the estimates (4.9), (4.13) for  $Y^{t,x}$ ,  $DY^{t,x}$  in  $L^p(C([0, T]; \mathbb{R}))$  and  $L^p(C([0, T]; E'))$ , respectively. For what concerns the terms  $\mathcal{M}_1^k$  and  $\mathcal{M}_2^k$  the argument of the proof is exactly the same, due to the symmetry of the construction.

Summing up the above computations we finally have

$$\sup_{|h|=1} \left[ \mathbb{E} |\Upsilon_T^k|^p + \mathbb{E} \left( \int_t^T |M^k(r)| dr \right)^p \right] \longrightarrow 0, \quad \text{if } |k| \rightarrow 0,$$

which is the required result. □

*Prof of Lemma 23.* Let us prove the second part of the statement, concerning second order derivatives. For  $0 \leq t \leq \tau \leq s \leq T$  we have

$$\begin{aligned} & |D_x^2 X_\tau^n(k, h) - D_x^2 X_\tau(k, h)| \\ & \lesssim \int_t^\tau |D^2 B^n(r, X_r^n)(D_x X_r^n k, D_x X_r^n h) - D^2 B(r, X_r)(D_x X_r k, D_x X_r h)| \, dr \\ & \quad + \int_t^\tau |DB^n(r, X_r^n) D_x^2 X_r^n(k, h) - DB(r, X_r) D_x^2 X_r(k, h)| \, dr \\ & = \int_t^\tau (A_1^n(r) + A_2^n(r)) \, dr \end{aligned}$$

so that

$$\mathbb{E} \sup_{\tau \in [t, s]} |D_x^2 X_\tau^n(k, h) - D_x^2 X_\tau(k, h)|^p \lesssim \int_t^s (\mathbb{E}[A_1^n(r)]^p + \mathbb{E}[A_2^n(r)]^p) \, dr .$$

Choose now  $\varepsilon > \max\left\{1 - \frac{1}{\alpha p}, 0\right\}$  and set

$$\bar{\nu} = \frac{1}{\alpha p} + \varepsilon, \quad \nu = 2 \frac{1 + \varepsilon \alpha p}{1 + \alpha p(\varepsilon - 1)} ;$$

$\bar{\nu}$  and  $\nu$  will be used as exponents and are chosen to ensure that all the inequalities exploited below are correct for any choice of  $p > 1$  and  $\alpha \in (0, 1)$ . From the bound

$$\begin{aligned} A_1^n(r) & \leq |D^2 B(r, J^n X_r^n)(J^n D_x X_r^n k, J^n D_x X_r^n h - J^n D_x X_r h)| \\ & \quad + |D^2 B(r, J^n X_r^n)(J^n D_x X_r^n k - J^n D_x X_r k, J^n D_x X_r h)| \\ & \quad + |[D^2 B(r, J^n X_r^n) - D^2 B(r, X_r)](J^n D_x X_r k, J^n D_x X_r h)| \\ & \quad + |D^2 B(r, X_r)(J^n D_x X_r k - D_x X_r k, J^n D_x X_r h - D_x X_r h)| \\ & \quad + |D^2 B(r, X_r)(D_x X_r k, J^n D_x X_r h - D_x X_r h)| \\ & \quad + |D^2 B(r, X_r)(J^n D_x X_r k - D_x X_r k, D_x X_r h)| \\ & \lesssim \left(1 + \sup_{r \in [t, T]} |X_r^n|^m\right) \sup_{r \in [t, T]} |D_x X_r^n k| \sup_{r \in [t, T]} |D_x X_r^n h - D_x X_r h| \\ & \quad + \left(1 + \sup_{r \in [t, T]} |X_r^n|^m\right) \sup_{r \in [t, T]} |D_x X_r h| \sup_{r \in [t, T]} |D_x X_r^n k - D_x X_r k| \\ & \quad + \sup_{r \in [t, T]} |D_x X_r k| \sup_{r \in [t, T]} |D_x X_r h| \sup_{r \in [t, T]} |J^n X_r^n - X_r|^\alpha \\ & \quad + |D^2 B(r, X_r)(J^n D_x X_r k - D_x X_r k, J^n D_x X_r h - D_x X_r h)| \\ & \quad + |D^2 B(r, X_r)(D_x X_r k, J^n D_x X_r h - D_x X_r h)| \\ & \quad + |D^2 B(r, X_r)(J^n D_x X_r k - D_x X_r k, D_x X_r h)| \end{aligned}$$

we get

$$\begin{aligned} \mathbb{E}[A_1^n(r)]^p & \lesssim \left[ \mathbb{E} \left( 1 + \sup_{r \in [t, T]} |X_r^n|^{3mp} \right) \right]^{1/3} \left[ \mathbb{E} \sup_{r \in [t, T]} \|D_x X_r^n\|^{3p} \right]^{1/3} \\ & \quad \times \left( |k| \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x X_r^n h - D_x X_r h|^{3p} \right]^{1/3} + |h| \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x X_r^n k - D_x X_r k|^{3p} \right]^{1/3} \right) \\ & \quad + |h|^p |k|^p \left[ \mathbb{E} \sup_{r \in [t, T]} \|D_x X_r\|^{\nu p} \right]^{2/\nu} \left[ \mathbb{E} \sup_{r \in [t, T]} |J^n X_r^n - X_r|^{\bar{\nu} \alpha p} \right]^{1/\bar{\nu}} \\ & \quad + \mathbb{E} |D^2 B(r, X_r)(J^n D_x X_r k - D_x X_r k, J^n D_x X_r h - D_x X_r h)|^p \\ & \quad + \mathbb{E} |D^2 B(r, X_r)(D_x X_r k, J^n D_x X_r h - D_x X_r h)|^p \end{aligned}$$

$$+ \mathbb{E} \left| D^2 B(r, X_r) (J^n D_x X_r k - D_x X_r k, D_x X_r h) \right|^p .$$

Thanks to Lemmas 22-23, (4.3) and (4.7), the first three terms goes to zero. Concerning the remaining three terms, we employ Assumption 4. More precisely, we have

$$\left[ \left( 1 + \mathbb{E} \sup_{r \in [t, T]} |X_r|^{3mp} \right) \mathbb{E} \sup_{r \in [t, T]} |D_x X_r k|^{3p} \mathbb{E} \sup_{r \in [t, T]} |D_x X_r h|^{3p} \right]^{1/3},$$

and recalling again (4.3) and (4.7) we get a uniform bound in  $r \in [0, T]$ . We can then pass to the limit exploiting Vitali convergence theorem. Similarly

$$\begin{aligned} A_2^n(r) &\leq \left| [DB(r, J^n X_r^n) - DB(r, X_r)] J^n D_x^2 X_r^n(k, h) \right| \\ &\quad + \left| DB(r, X_r) J^n [D_x^2 X_r^n(k, h) - D_x^2 X_r(k, h)] \right| \\ &\quad + \left| DB(r, X_r) [J^n D_x^2 X_r(k, h) - D_x^2 X_r(k, h)] \right| \\ &\lesssim \sup_{r \in [t, T]} |D_x^2 X_r^n(k, h)| \left( 1 + \sup_{r \in [t, T]} |X_r^n|^m + \sup_{r \in [t, T]} |X_r|^m \right) \sup_{r \in [t, T]} |J^n X_r^n - X_r| \\ &\quad + |D_x^2 X_r^n(k, h) - D_x^2 X_r(k, h)| \\ &\quad + |DB(r, X_r) [J^n D_x^2 X_r(k, h) - D_x^2 X_r(k, h)]| , \end{aligned}$$

where we used the  $C^1$ -regularity of  $x \mapsto DB(\cdot, x)$  along with the growth of  $D^2 B$  given in Assumption 1. Taking the expectation of the  $p$ -th power, the first and last term converge to 0 for a.e.  $r \in [0, T]$  as above. Therefore

$$\mathbb{E} \sup_{\tau \in [t, s]} |D_x^2 X_\tau^n(k, h) - D_x^2 X_\tau(k, h)|^p \lesssim N^n(s) + \int_t^s \mathbb{E} \sup_{\tau \in [t, r]} |D_x^2 X_\tau^n(k, h) - D_x^2 X_\tau(k, h)|^p dr$$

where  $N^n(s)$  contains all the other terms and  $N^n(s) \rightarrow 0$  for every  $s \in [t, T]$ . The application of the Gronwall lemma concludes the proof.  $\square$

*Proof of Lemma 26.* Recalling the equation satisfied by  $D_x^2 Y_s(k, h)$  (analogously by  $D_x^2 Y_s^n(k, h)$ ) we set

$$\Delta^2 Y_s^n(k, h) = D_x^2 Y_s^n(k, h) - D_x^2 Y_s(k, h) , \quad \Delta^2 Z_s^n(k, h) = D_x^2 Z_s^n(k, h) - D_x^2 Z_s(k, h)$$

to obtain

$$\begin{aligned} \Delta^2 Y_s^n(k, h) + \int_s^T \Delta^2 Z_r^n(k, h) dW_r = & \\ - \int_s^T [D_1 G^n(r, X_r^n, Y_r^n, Z_r^n) D_x^2 X_r^n(k, h) - D_1 G(r, X_r, Y_r, Z_r) D_x^2 X_r(k, h)] dr & \quad (1) \\ - \int_s^T [D_{1,1}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x X_r^n k, D_x X_r^n h) - D_{1,1}^2 G(r, X_r, Y_r, Z_r) (D_x X_r k, D_x X_r h)] dr & \quad (2) \\ - \int_s^T [D_{1,2}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x Y_r^n k, D_x X_r^n h) - D_{1,2}^2 G(r, X_r, Y_r, Z_r) (D_x Y_r k, D_x X_r h)] dr & \quad (3) \\ - \int_s^T [D_{1,3}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x Z_r^n k, D_x X_r^n h) - D_{1,3}^2 G(r, X_r, Y_r, Z_r) (D_x Z_r k, D_x X_r h)] dr & \quad (4) \\ - \int_s^T [D_2 G^n(r, X_r^n, Y_r^n, Z_r^n) D_x^2 Y_r^n(k, h) - D_2 G(r, X_r, Y_r, Z_r) D_x^2 Y_r(k, h)] dr & \quad (5) \\ - \int_s^T [D_{2,1}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x X_r^n k, D_x Y_r^n h) - D_{2,1}^2 G(r, X_r, Y_r, Z_r) (D_x X_r k, D_x Y_r h)] dr & \quad (6) \\ - \int_s^T [D_{2,2}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x Y_r^n k, D_x Y_r^n h) - D_{2,2}^2 G(r, X_r, Y_r, Z_r) (D_x Y_r k, D_x Y_r h)] dr & \quad (7) \\ - \int_s^T [D_{2,3}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x Z_r^n k, D_x Y_r^n h) - D_{2,3}^2 G(r, X_r, Y_r, Z_r) (D_x Z_r k, D_x Y_r h)] dr & \quad (8) \\ - \int_s^t [D_3 G^n(r, X_r^n, Y_r^n, Z_r^n) D_x^2 Z_r^n(k, h) - D_3 G(r, X_r, Y_r, Z_r) D_x^2 Z_r(k, h)] dr & \quad (9) \end{aligned}$$

$$- \int_s^T [D_{3,1}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x X_r^n k, D_x Z_r^n h) - D_{3,1}^2 G(r, X_r, Y_r, Z_r) (D_x X_r k, D_x Z_r h)] dr \quad (10)$$

$$- \int_s^T [D_{3,2}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x Y_r^n k, D_x Z_r^n h) - D_{3,2}^2 G(r, X_r, Y_r, Z_r) (D_x Y_r k, D_x Z_r h)] dr \quad (11)$$

$$- \int_s^T [D_{3,3}^2 G^n(r, X_r^n, Y_r^n, Z_r^n) (D_x Z_r^n k, D_x Z_r^n h) - D_{3,3}^2 G(r, X_r, Y_r, Z_r) (D_x Z_r k, D_x Z_r h)] dr \quad (12)$$

$$+ D^2 \Phi^n(X_T^n) (D_x X_T^n k, D_x X_T^n h) - D^2 \Phi(X_T) (D_x X_T k, D_x X_T h) \quad (13)$$

$$+ D\Phi^n(X_T^n) D_x^2 X_T^n(k, h) - D\Phi(X_T) D_x^2 X_T(k, h). \quad (14)$$

We rephrase the above BSDE as

$$\begin{aligned} & \Delta^2 Y_s^n(k, h) + \int_s^T \Delta^2 Z_r^n(k, h) dW_r = \\ & \textcircled{1} + \textcircled{2} + \textcircled{3} + \textcircled{4} \\ & - \int_s^T D_2 G^n(r, X_r^n, Y_r^n, Z_r^n) \Delta^2 Y_r^n(k, h) dr \\ & - \int_s^T [D_2 G^n(r, X_r^n, Y_r^n, Z_r^n) - D_2 G(r, X_r, Y_r, Z_r)] D_x^2 Y_r(k, h) dr \quad (5a) \\ & + \textcircled{6} + \textcircled{7} + \textcircled{8} \\ & - \int_s^T D_3 G^n(r, X_r^n, Y_r^n, Z_r^n) \Delta^2 Z_r^n(k, h) dr \\ & - \int_s^T [D_3 G^n(r, X_r^n, Y_r^n, Z_r^n) - D_3 G(r, X_r, Y_r, Z_r)] D_x^2 Z_r(k, h) dr \quad (9a) \\ & + \textcircled{10} + \textcircled{11} + \textcircled{12} + \textcircled{13} + \textcircled{14}, \end{aligned}$$

which is of the form

$$\Delta^2 Y_s^n(k, h) + \int_s^T \Delta^2 Z_r^n(k, h) dW_r = \bar{\eta}^n + \int_s^T \bar{\alpha}_r^n \Delta^2 Y_r^n(k, h) dr + \int_s^T \bar{\beta}_r^n dr + \int_s^T \bar{\gamma}_r^n \Delta^2 Z_r^n(k, h) dr$$

with

$$\begin{aligned} \bar{\eta}^n & := \textcircled{13} + \textcircled{14}, \\ \bar{\alpha}_r^n & := -D_2 G^n(r, X_r^n, Y_r^n, Z_r^n), \\ \bar{\gamma}_r^n & := -D_3 G^n(r, X_r^n, Y_r^n, Z_r^n), \\ \int_t^T \bar{\beta}_r^n dr & = \textcircled{1} + \textcircled{2} + \textcircled{3} + \textcircled{4} + \textcircled{5a} + \textcircled{6} + \textcircled{7} + \textcircled{8} + \textcircled{9a} + \textcircled{10} + \textcircled{11} + \textcircled{12}. \end{aligned}$$

Defining  $V$  as in (6.15), the equation being linear, we can apply the same strategy as in the proof of Lemma 25. Hence it suffices to check that for some  $p \geq 2$

$$\mathbb{E} |\bar{\eta}^n|^p + \mathbb{E} \left( \int_t^T |\bar{\beta}_r^n| dr \right)^p \xrightarrow{n \rightarrow \infty} 0.$$

The convergence of  $\bar{\eta}^n$  under Assumption 4 was proved in Flandoli and Zanco (2016) for  $p = 1$  and  $\Phi \in C_b^{2,\alpha}$ ; the extension to  $\Phi$  with polynomial growth relies on Theorem 8 and on the uniform integrability of  $\Phi^n(X^n)$  in  $L^p$ . By Hölder inequality we also obtain convergence of  $\bar{\eta}^n$  in  $L^p(\Omega)$  for any  $p \geq 2$ .

We now show how to deal with some of the addends defining  $\bar{\beta}^n$ . The same technique can be used also for the remaining terms.

$$\begin{aligned} \mathbb{E} |\textcircled{1}|^p & \leq \mathbb{E} \left( \int_t^T |D_1 G^n(r, X_r^n, Y_r^n, Z_r^n) [D_x^2 X_r^n(k, h) - D_x^2 X_r(k, h)]| dr \right)^p \\ & + \mathbb{E} \left( \int_t^T |[D_1 G^n(r, X_r^n, Y_r^n, Z_r^n) - D_1 G(r, X_r, Y_r, Z_r)] D_x^2 X_r(k, h)| dr \right)^p \end{aligned}$$



$$= \mathbb{E} \left( \left[ \overline{C}_1^n \right]^p + \left[ \overline{C}_2^n \right]^p \right) .$$

Thanks to the uniform bounds on  $X^n$ ,  $Y^n$ ,  $Z^n$  and Assumption 2 we have

$$\begin{aligned} \left[ \overline{C}_1^n \right]^p &\lesssim \left( 1 + \sup_{r \in [t, T]} |X_r^n|^{mp} \right) \sup_{r \in [t, T]} \left| [D_x^2 X_r^n - D_x^2 X_r] (k, h) \right|^p \\ &\quad \times \left( 1 + \sup_{r \in [t, T]} |Y_r^n|^p + \left( \int_t^T |Z_r^n|^2 dr \right)^{p/2} \right) , \end{aligned}$$

hence  $\mathbb{E} \left[ \overline{C}_1^n \right]^p \rightarrow 0$  by Hölder's inequality and Lemma 23. The convergence of  $\overline{C}_2^n$  is identical to the one of  $\int C_2^n$  in the proof of Lemma 25. By using the shorthand

$$D_{ij}^2 G_r(n) := D_{ij}^2 G(r, J^n X_r^n, Y_r^n, Z_r^n) , D_{ij}^2 G_r(\cdot) := D_{ij}^2 G(r, X_r, Y_r, Z_r) ,$$

for  $i, j = 1, 2, 3$ , for the term  $\textcircled{2}$  we get

$$\begin{aligned} \mathbb{E} \left| \textcircled{2} \right|^p &\lesssim \mathbb{E} \left( \int_t^T |D_{1,1}^2 G_r(n) (J^n D_x X_r^n k, J^n D_x X_r^n h) - D_{1,1}^2 G_r(n) (J^n D_x X_r k, J^n D_x X_r h)| dr \right)^p \\ &\quad + \mathbb{E} \left( \int_t^T |D_{1,1}^2 G_r(n) (J^n D_x X_r k, J^n D_x X_r h) - D_{1,1}^2 G_r(\cdot) (J^n D_x X_r k, J^n D_x X_r h)| dr \right)^p \\ &\quad + \mathbb{E} \left( \int_t^T |D_{1,1}^2 G_r(\cdot) (J^n D_x X_r k, J^n D_x X_r h) - D_{1,1}^2 G_r(\cdot) (D_x X_r k, D_x X_r h)| dr \right)^p \\ &= \mathbb{E} \left( \left[ \overline{F}_{111}^n \right]^p + \left[ \overline{F}_{112}^n \right]^p + \left[ \overline{F}_{113}^n \right]^p \right) . \end{aligned}$$

$\overline{F}_{112}^n$  can be studied again as  $\int C_2^n$  in the proof of Lemma 25, while

$$\begin{aligned} \left[ \overline{F}_{111}^n \right]^p &\leq \left( \int_t^T |D_{1,1}^2 G_r(n) [(J^n D_x X_r^n, J^n D_x X_r^n h - J^n D_x X_r h) \right. \\ &\quad \left. + (J^n D_x X_r^n k - J^n D_x X_r k, J^n D_x X_r h)]| dr \right)^p \\ &\lesssim \left[ \sup_{r \in [t, T]} |D_x X_r^n k|^p \sup_{r \in [t, T]} |D_x X_r^n h - D_x X_r h|^p \right. \\ &\quad \left. + \sup_{r \in [t, T]} |D_x X_r k|^p \sup_{r \in [t, T]} |D_x X_r^n h - D_x X_r h|^p \right] \\ &\quad \times \left( 1 + \sup_{r \in [t, T]} |X_r^n|^{mp} \right) \left( 1 + \sup_{r \in [t, T]} |Y_r^n|^p + \left( \int_t^T |Z_r^n|^2 dr \right)^{p/2} \right) , \end{aligned}$$

and

$$\begin{aligned} \left[ \overline{F}_{113}^n \right]^p &\leq \int_t^T |D_{1,1}^2 G_r(\cdot) (J^n D_x X_r k - D_x X_r k, J^n D_x X_r h - D_x X_r h)|^p dr \\ &\quad + \int_t^T |D_{1,1}^2 G_r(\cdot) (D_x X_r k, J^n D_x X_r h - D_x X_r h)|^p dr \\ &\quad + \int_t^T |D_{1,1}^2 G_r(\cdot) (J^n D_x X_r k - D_x X_r k, D_x X_r h)|^p dr ; \end{aligned}$$

the desired convergence then follows by taking expectation, applying Hölder's inequality and using Assumption 4, the uniform bound (6.10) and the Vitali convergence theorem.

$$\mathbb{E} \left| \textcircled{3} \right|^p \lesssim \mathbb{E} \left( \int_t^T |D_{12}^2 G_r(n) (D_x Y_r^n k, J^n D_x X_r^n h) - D_{12}^2 G_r(\cdot) (D_x Y_r k, J^n D_x X_r h)| dr \right)^p$$

$$\begin{aligned}
& + \mathbb{E} \left( \int_t^T |D_{12}^2 G_r(\cdot)(D_x Y_r k, J^n D_x X_r h) - D_{12}^2 G_r(\cdot)(D_x Y_r k, D_x X_r h)| \, dr \right)^p \\
& = \mathbb{E} \left( [\bar{\mathbb{F}}_{121}^n]^p + [\bar{\mathbb{F}}_{122}^n]^p \right)
\end{aligned}$$

Choose again any  $\varepsilon > \max \left\{ 1 - \frac{1}{\alpha p}, 0 \right\}$  and set

$$\bar{\nu} = \frac{1}{\alpha p} + \varepsilon, \quad \nu = 2 \frac{1 + \varepsilon \alpha p}{1 + \alpha p(\varepsilon - 1)};$$

then by Hölder character of the second derivatives of  $G$

$$\begin{aligned}
\mathbb{E} [\bar{\mathbb{F}}_{121}^n]^p & \lesssim \mathbb{E} \left( \int_t^T |D_{12}^2 G_r(n)(D_x Y_r^n k, J^n D_x X_r^n h) - D_{12}^2 G_r(n)(D_x Y_r^n k, J^n D_x X_r h)| \, dr \right)^p \\
& + \mathbb{E} \left( \int_t^T |D_{12}^2 G_r(n)(D_x Y_r^n k, J^n D_x X_r h) - D_{12}^2 G_r(n)(D_x Y_r k, J^n D_x X_r h)| \, dr \right)^p \\
& + \mathbb{E} \left( \int_t^T |[D_{12}^2 G_r(n) - D_{12}^2 G_r(\cdot)](D_x Y_r k, J^n D_x X_r h)| \, dr \right)^p \\
& \lesssim \left( 1 + \left[ \mathbb{E} \sup_{r \in [t, T]} |X_r^n|^{4mp} \right]^{1/4} \right) \left( 1 + \left[ \mathbb{E} \sup_{r \in [t, T]} |Y_r^n|^{4p} \right]^{1/4} \right) \\
& \quad \times \left( \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x Y_r^n k|^4 \right]^{1/4} \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x X_r^n h - D_x X_r h|^{4p} \right]^{1/4} \right. \\
& \quad \left. + \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x X_r h|^4 \right]^{1/4} \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x Y_r^n k - D_x Y_r k|^{4p} \right]^{1/4} \right) \\
& \quad + \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x Y_r k|^{\nu p} \right]^{1/\nu} \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x X_r h|^{\nu p} \right]^{1/\nu} \\
& \quad \times \left( \left[ \mathbb{E} \sup_{r \in [t, T]} |Y_r^n - Y_r|^{\bar{\nu} \alpha p} \right]^{1/\bar{\nu}} + \left[ \mathbb{E} \int_t^T |J^n X_r^n - X_r|^{\bar{\nu} \alpha p} \, dr \right]^{1/\bar{\nu}} \right) \\
& \quad + \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x Y_r k|^{3p} \right]^{1/3} \left[ \mathbb{E} \sup_{r \in [t, T]} |D_x X_r h|^{3p} \right]^{1/3} \left[ \mathbb{E} \left( \int_t^T |Z_r^n - Z_r|^2 \, dr \right)^{3p/2} \right]^{1/3}
\end{aligned}$$

which goes to 0 by Lemmas 22, 23 and 24, thanks to the choice of  $\bar{\nu}$  and  $\nu$ . The term  $\bar{\mathbb{F}}_{122}^n$  can be controlled as before using Assumption 4.

The estimates for (4) are almost identical. Terms (5a) and (9a) are treated in the same way as  $\beta_2^n$  and  $\beta_3^n$  in Lemma 25, respectively. We have already shown how to deal with (12) above, the remaining terms (6), (7), (8), (10), (11) are then simple adaptations of (1), (2), (3) and (12).  $\square$

## References

- Bayraktar, E., Keller, C., 2018. Path-dependent Hamilton-Jacobi equations in infinite dimensions. *Journal of Functional Analysis* 275, 2096–2161. URL: <http://www.sciencedirect.com/science/article/pii/S002212361830274X>.
- Bensoussan, A., Da Prato, G., Delfour, M.C., Mitter, S.K., 2007. Representation and control of infinite dimensional systems. *Systems & Control: Foundations & Applications*. second ed., Birkhäuser Boston, Inc., Boston, MA. URL: <https://doi.org/10.1007/978-0-8176-4581-6>, doi:10.1007/978-0-8176-4581-6.
- Chojnowska-Michalik, A., 1978. Representation theorem for general stochastic delay equations. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* 26, 635–642.

- Cont, R., Fournié, D.A., 2013. Functional Itô calculus and stochastic integral representation of martingales. *Ann. Probab.* 41, 109–133. URL: <https://doi.org/10.1214/11-AOP721>, doi:10.1214/11-AOP721.
- Cordoni, F., Di Persio, L., Maticiuc, L., Zălinescu, A., 2019. A stochastic approach to path-dependent nonlinear kolmogorov equations via bsdes with time-delayed generators and applications to finance. preprint arXiv URL: <https://arxiv.org/abs/1602.05793>.
- Cordoni, F., Di Persio, L., Oliva, I., 2017. A nonlinear kolmogorov equation for stochastic functional delay differential equations with jumps. *Nonlinear Differential Equations and Applications NoDEA* 24, 16. URL: <https://doi.org/10.1007/s00030-017-0440-3>, doi:10.1007/s00030-017-0440-3.
- Cosso, A., Di Girolami, C., Russo, F., 2014. Calculus via regularizations in banach spaces and kolmogorov-type path-dependent equations. preprint arXiv URL: <https://arxiv.org/abs/1411.8000>.
- Cosso, A., Federico, S., Gozzi, F., Rosestolato, M., Touzi, N., 2018. Path-dependent equations and viscosity solutions in infinite dimension. *Ann. Probab.* 46, 126–174. URL: <https://doi.org/10.1214/17-AOP1181>, doi:10.1214/17-AOP1181.
- Cosso, A., Russo, F., 2019a. Functional Itô versus Banach space stochastic calculus and strict solutions of semilinear path-dependent equations. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 19.
- Cosso, A., Russo, F., 2019b. Strong-viscosity solutions: Classical and path-dependent pdes. *Osaka J. Math.* 56, 323–373. URL: <https://projecteuclid.org:443/euclid.ojm/1554278428>.
- Da Prato, G., Zabczyk, J., 2014. Stochastic equations in infinite dimensions. volume 152 of *Encyclopedia of Mathematics and its Applications*. Second ed., Cambridge University Press, Cambridge. URL: <https://doi.org/10.1017/CB09781107295513>.
- Delfour, M.C., Mitter, S.K., 1972. Hereditary differential systems with constant delays. I. General case. *J. Differential Equations* 12, 213–235; erratum, *ibid.* 14 (1973), 397. URL: [https://doi.org/10.1016/0022-0396\(72\)90030-7](https://doi.org/10.1016/0022-0396(72)90030-7), doi:10.1016/0022-0396(72)90030-7.
- Di Girolami, C., Russo, F., 2018. About classical solutions of the path-dependent heat equation. preprint arXiv URL: <https://arxiv.org/abs/1804.03845>.
- Dupire, B., 2009. Functional itô calculus. Portfolio Research Paper 2009-04, Bloomberg .
- Ekren, I., Keller, C., Touzi, N., Zhang, J., 2014. On viscosity solutions of path dependent pdes. *Ann. Probab.* 42, 204–236. URL: <https://doi.org/10.1214/12-AOP788>, doi:10.1214/12-AOP788.
- Ekren, I., Touzi, N., Zhang, J., 2016. Viscosity solutions of fully nonlinear parabolic path dependent pdes: Part i. *Ann. Probab.* 44, 1212–1253. URL: <https://doi.org/10.1214/14-AOP999>, doi:10.1214/14-AOP999.
- Fabrizi, G., Gozzi, F., Świech, A., 2017. Stochastic optimal control in infinite dimension. volume 82 of *Probability Theory and Stochastic Modelling*. Springer, Cham. URL: <https://doi.org/10.1007/978-3-319-53067-3>, doi:10.1007/978-3-319-53067-3. dynamic programming and HJB equations, With a contribution by Marco Fuhrman and Gianmario Tessitore.
- Fathi, A., 2008. Weak kam theorem in lagrangian dynamics preliminary version number 10. by CUP .
- Flandoli, F., Russo, F., Zanco, G., 2018. Infinite-dimensional calculus under weak spatial regularity of the processes. *J. Theoret. Probab.* 31, 789–826. URL: <https://doi.org/10.1007/s10959-016-0724-2>, doi:10.1007/s10959-016-0724-2.
- Flandoli, F., Zanco, G., 2016. An infinite-dimensional approach to path-dependent Kolmogorov equations. *Ann. Probab.* 44, 2643–2693. URL: <http://dx.doi.org/10.1214/15-AOP1031>, doi:10.1214/15-AOP1031.
- Fuhrman, M., Masiero, F., Tessitore, G., 2010. Stochastic equations with delay: optimal control via BSDEs and regular solutions of Hamilton-Jacobi-Bellman equations. *SIAM J. Control Optim.* 48, 4624–4651. URL: <https://doi.org/10.1137/080730354>, doi:10.1137/080730354.
- Fuhrman, M., Tessitore, G., 2002. Nonlinear Kolmogorov equations in infinite dimensional spaces: the backward stochastic differential equations approach and applications to optimal control. *Ann. Probab.* 30, 1397–1465. URL: <http://dx.doi.org/10.1214/aop/1029867132>, doi:10.1214/aop/1029867132.

- Gozzi, F., Masiero, F., 2017. Stochastic optimal control with delay in the control I: Solving the HJB equation through partial smoothing. *SIAM J. Control Optim.* 55, 2981–3012. URL: <https://doi.org/10.1137/16M1070128>, doi:10.1137/16M1070128.
- Guatteri, G., Masiero, F., Orrieri, C., 2017. Stochastic maximum principle for SPDEs with delay. *Stochastic Process. Appl.* 127, 2396–2427. URL: <https://doi.org/10.1016/j.spa.2016.11.007>, doi:10.1016/j.spa.2016.11.007.
- Izumi, Y., 2018. Higher order differentiability of solutions to backward stochastic differential equations. *Stochastics* 90, 102–150. URL: <https://doi.org/10.1080/17442508.2017.1315119>, doi:10.1080/17442508.2017.1315119.
- Masiero, F., 2008. Stochastic optimal control problems and parabolic equations in Banach spaces. *SIAM J. Control Optim.* 47, 251–300. URL: <https://doi.org/10.1137/050632725>, doi:10.1137/050632725.
- Masiero, F., Richou, A., 2014. HJB equations in infinite dimensions with locally Lipschitz Hamiltonian and unbounded terminal condition. *J. Differential Equations* 257, 1989–2034. URL: <http://dx.doi.org/10.1016/j.jde.2014.05.026>, doi:10.1016/j.jde.2014.05.026.
- Mohammed, S., 1984. *Stochastic Functional Differential Equations*. Chapman & Hall/CRC research notes in mathematics series, Pitman Advanced Pub. Program. URL: <https://books.google.it/books?id=C6fzoTRUti0C>.
- Pardoux, E., Răşcanu, A., 2014. Stochastic differential equations, backward SDEs, partial differential equations. volume 69 of *Stochastic Modelling and Applied Probability*. Springer, Cham. URL: <https://doi.org/10.1007/978-3-319-05714-9>.
- Rosestolato, M., 2016. Functional Itô calculus in Hilbert spaces and application to path-dependent kolmogorov equations. arXiv preprint arXiv:1606.06326 .
- Rosestolato, M., Świech, A., 2017. Partial regularity of viscosity solutions for a class of kolmogorov equations arising from mathematical finance. *Journal of Differential Equations* 262, 1897–1930.
- Tang, S., Zhang, F., 2015. Path-dependent optimal stochastic control and viscosity solution of associated bellman equations. *Discrete & Continuous Dynamical Systems - A* 35, 5521–5553. URL: <http://aimsciences.org//article/id/01e710a7-b825-4b45-a498-4a0da2cad94a>, doi:10.3934/dcds.2015.35.5521.
- Yosida, K., 1995. *Functional Analysis*. Classics in Mathematics, Springer Berlin Heidelberg. URL: <https://books.google.at/books?id=QqNpbTQwKXMC>.
- Zanco, G., 2015. Infinite-dimensional methods for path-dependent stochastic differential equations. Ph.D. thesis. Università di Pisa.
- Zhou, J., Zhang, Z., 2011. Optimal control problems for stochastic delay evolution equations in Banach spaces. *Internat. J. Control* 84, 1295–1309. URL: <https://doi.org/10.1080/00207179.2011.592999>, doi:10.1080/00207179.2011.592999.
- Zhou, Jianjun, 2018. A class of infinite-horizon stochastic delay optimal control problems and a viscosity solution to the associated hjb equation. *ESAIM: COCV* 24, 639–676. URL: <https://doi.org/10.1051/cocv/2017042>, doi:10.1051/cocv/2017042.