NONLINEAR KOLMOGOROV EQUATIONS IN INFINITE DIMENSIONAL SPACES: THE BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS APPROACH AND APPLICATIONS TO OPTIMAL CONTROL

BY MARCO FUHRMAN AND GIANMARIO TESSITORE

Politecnico di Milano and Universitá Di Parma

Solutions of semilinear parabolic differential equations in infinite dimensional spaces are obtained by means of forward and backward infinite dimensional stochastic evolution equations. Parabolic equations are intended in a mild sense that reveals to be suitable also towards applications to optimal control.

1. Introduction. Let us consider a stochastic evolution equation of the form:

$$
(1.1) \quad \begin{cases} dX_\tau = AX_\tau \, d\tau + F(\tau, X_\tau) \, d\tau + G(\tau, X_\tau) \, dW_\tau, & \tau \in [t, T] \subset [0, T], \\ X_t = x \in H, \end{cases}
$$

for a process X in a Hilbert space H, where W is a cylindrical Wiener process in another Hilbert space Ξ, A is the generator of a strongly continuous semigroup of bounded linear operators $\{e^{tA}\}$ in H, F and G are functions with values in H and $L(\Xi, H)$ respectively, satisfying appropriate Lipschitz conditions. Under suitable assumptions, a unique solution $\{X(\tau, t, x), \tau \in [t, T]\}$ exists and defines a Markov process with transition function $\{P_{t,\tau}, 0 \le t \le \tau \le T\}$. $P_{t,\tau}$ acts on measurable functions $\phi : H \to \mathbb{R}$, satisfying suitable growth conditions, according to the formula:

$$
P_{t,\tau}[\phi](x) = \mathbb{E} \phi\big(X(\tau, t, x)\big), \qquad x \in H.
$$

If ϕ is sufficiently regular then the function $v(t, x) = P_{t,T}[\phi](x)$ is a classical solution of the backward Kolmogorov equation

$$
\begin{cases} \dfrac{\partial v(t, x)}{\partial t} + \mathcal{L}_t[v(t, \cdot)](x) = 0, & t \in [0, T], \, x \in H, \\ v(T, x) = \phi(x), \end{cases}
$$

where the linear operator $\mathcal{L}_t$ is defined by

$$
\begin{aligned} \mathcal{L}_t[\phi](x) := {} & \tfrac{1}{2} \operatorname{Trace}\big(G(t, x) G(t, x)^* \nabla^2 \phi(x)\big) \\ & + \langle x, A^* \nabla \phi(x) \rangle_H + \langle F(t, x), \nabla \phi(x) \rangle_H. \end{aligned}
$$

In this formula $\nabla\phi$ and $\nabla^2\phi$ are first and second Gâteaux derivatives of ϕ (identified with elements of H and $L(H, H)$ respectively). When ϕ is not regular, the function $v(t, x) = P_{t,T}[\phi](x)$ can be considered as a generalized solution of the backward Kolmogorov equation. We refer to [9, 10, 41] for a detailed exposition of these facts and related matters.

Here we are interested in a generalization of this equation, written formally as

$$(1.2) \quad \begin{cases} \dfrac{\partial v(t, x)}{\partial t} + \mathcal{L}_t[v(t, \cdot)](x) \\ \quad = \psi\big(t, x, v(t, x), G(t, x)^*\nabla_x v(t, x)\big), \qquad t \in [0, T], \ x \in H, \\ u(T, x) = \phi(x). \end{cases}$$

We will refer to this equation as the nonlinear Kolmogorov equation. $\psi\colon [0, T] \times H \times \mathbb{R} \times \Xi \to \mathbb{R}$ is a given function and $\nabla_x v$, the Gâteaux derivative of v with respect to x, is identified with an element of H, so that $G(t, x)^*\nabla_x v(t, x) \in \Xi$. Notice the occurrence of G in the nonlinear term: this does not imply any loss of generality in the nondegenerate case, that is, when G is boundedly invertible, whereas it involves a genuine restriction in the general case.

One of the main results of this paper, Theorem 6.2, specifies conditions for unique solvability of equation (1.2). Various concepts of solution are known for (linear and) nonlinear parabolic equations in infinite dimensions. In this paper we restrict attention to continuous solutions, but other effective approaches have been proposed, for instance by means of Dirichlet forms [30, 37]. Many investigations have been carried out in connection with the Hamilton–Jacobi–Bellman equation arising in optimal control for nonlinear infinite dimensional stochastic systems, so we postpone references to the discussion on control theoretic applications that will be done below. One possibility to deal with equation (1.2) is to look for classical solutions, that is, functions which are twice differentiable with respect to x and once with respect to t, such that $\mathcal{L}[v(t, \cdot)]$ makes sense for every $t \in [0, T]$ and (1.2) holds. This forces to impose heavy assumptions on the functions ψ and ϕ, involving existence of derivatives up to order two as well as trace conditions on second derivatives. Another possibility, in some sense opposite, is to consider viscosity solutions. Existence and uniqueness of viscosity solutions can be proved under much weaker assumptions on the coefficients. However, in view of applications to optimal control theory, it is important to show the existence of $\nabla_x v$, since this allows to characterize optimal control by feedback laws. Since, in general, viscosity solutions are not differentiable, this characterization is not immediately available. In this paper we will consider solutions in the so called mild sense (already considered in the literature, but not in connection with the backward stochastic equations approach). Namely a mild solution v of equation (1.2) will satisfy the equality

$$(1.3) \quad \begin{aligned} v(t, x) = -\int_t^T & P_{t,\tau}\big[\psi\big(\tau, \cdot, v(\tau, \cdot), G(\tau, \cdot)^*\nabla_x v(\tau, \cdot)\big)\big](x)\, d\tau \\ & + P_{t,T}[\phi](x), \qquad t \in [0, T], x \in H, \end{aligned}$$

which arises formally from (1.2) as the variation of constants formula. We notice that formula (1.3) is meaningful provided v is only once differentiable with respect to x and, of course, provided ψ, v and $\nabla_x v$ satisfy appropriate measurability and growth conditions. Thus, mild solutions are in a sense intermediate between classical and viscosity solutions. We can prove existence and uniqueness of a mild solutions v by requiring existence and boundedness (or growth conditions) of first derivatives of ψ and ϕ: compare Theorem 6.2. We wish to stress that in our assumptions, in contrast to most of the references cited below, the derivatives are understood in the sense of Gâteaux: this is important in view of applications where H is a space of summable functions and nonlinear terms are Nemytskii (evaluation) operators: it is well known that they are not Fréchet differentiable, except in trivial cases. Thus our results can be directly applied to stochastic partial differential equations: compare Example 7.3.1 as an instance of this.

In order to prove existence and uniqueness for the mild solution of (1.2) we generalize a probabilistic technique, based on backward stochastic differential equations, which we believe to have an intrinsic interest. We consider the following backward stochastic evolution equation:

$$(1.4) \qquad \begin{cases} dY_\tau = \langle Z_\tau, dW_\tau \rangle_\Xi + \psi(\tau, X_\tau, Y_\tau, Z_\tau)\, d\tau, & \tau \in [t, T], \\ Y_T = \phi(X_T), \end{cases}$$

where X is the solution of (1.1). Under suitable assumptions on ψ, there exists a unique adapted process (Y, Z) in $\mathbb{R} \times \Xi$, solution of (1.4). The processes X, Y, Z depend on the values of x and t, occurring as initial conditions in (1.1): we may denote them by $X(\tau, t, x)$, $Y(\tau, t, x)$, $Z(\tau, t, x)$, $\tau \in [t, T]$. If we define $v(t, x) = Y(t, t, x)$ then it turns out that the function v is deterministic and it is a solution of the nonlinear Kolmogorov equation.

This connection between backward stochastic equations and nonlinear partial differential equations was proved in the paper [36] for the finite dimensional case, that is, $H = \mathbb{R}^n$, $\Xi = \mathbb{R}^d$ and for classical solutions of the parabolic equation. The extension of these results to the Hilbert space case is far from being straightforward (for instance as we mentioned above the definition of solution of the parabolic equation has to be changed) and it was not investigated before. We are led to a detailed study of equations (1.1) and (1.4), which requires some effort and different arguments than in the finite dimensional case. These extended results have an interest in themselves and constitute another main aim of this article.

Backward stochastic equations in finite dimensions have been intensively studied in recent years, starting from the paper by Pardoux and Peng [35]: we refer the reader to [29] and [11] for an exposition of this subject and of the more general theory of forward–backward systems, as well as a detailed account of the existing literature. We note in passing that equations (1.1) and (1.4) are not a forward–backward system in the most general form, since equation (1.1) does not involve the processes (Y, Z) occurring in (1.4): this level of generality is not needed for our

purposes. In the infinite-dimensional case, however, there are few known results on stochastic nonlinear backward equations: we are only aware of [23], but the results of these paper are not sufficient for us although they are more general in the direction of allowing the presence, in (1.4), of unbounded operators (for results on infinite dimensional linear backward stochastic equations see [40] and [28]).

Next we briefly sketch the main points one needs to face in order to perform the proof of Theorem 6.2 outlined above. We first mention that existence of *some* solution of (1.1) and (1.4) is not sufficient: in order to perform our program we need more precise information. For instance, as in [36], we need to show finiteness of moments of all order, in various senses, for the processes X, Y, Z, and path continuity for X and Y: to this end we combine, adapt and extend arguments from [23] and [36].

Since we aim at proving that $v(t, x) = Y(t, t, x)$ is a solution of (1.2), or better (1.3), we are faced with the issue of dependence of the processes $X(\cdot, t, x)$ $Y(\cdot, t, x)$, $Z(\cdot, t, x)$ on parameters t, x. In contrast to [36], we cannot use the classical Kolmogorov theorem for continuity of random fields, due to the fact that x takes values in a Hilbert space H. Instead, we first investigate existence and continuity of the Gâteaux derivative process $\nabla_x X(\cdot, t, x)$ by means of a parameter depending contraction principle (following [9, 10, 41]). The same argument is then used to investigate dependence of the solution (Y, Z) of (1.4) on the *process X* and the existence of derivatives $\nabla_X Y, \nabla_X Z$ (X is treated as an element of a suitable vector space of processes). Finally, a chain rule is applied to prove existence of the derivatives $\nabla_x Y(\cdot, t, x)$, $\nabla_x Z(\cdot, t, x)$ with respect to $x \in H$ and the required regularity properties of $v(t, x) = Y(t, t, x)$ are obtained.

Another key step in the proof of Theorem 6.2 is the formula

$$Z(s, t, x) = G\big(s, X(s, t, x)\big)^* \nabla_x Y\big(s, s, X(s, t, x)\big) \qquad \text{for almost all } s \in [t, T],$$

which relates the processes X, Y and Z. Following [36], we prove it in two steps, as an application of the Malliavin calculus: denoting by D the Malliavin derivative operator one first proves that

$$Z(s, t, x) = \lim_{\tau \downarrow s} D_s Y(\tau, t, x) \qquad \text{for almost all } s \in [t, T],$$

in mean square, and then that, for almost all s, τ with $t \leq s \leq \tau \leq T$,

$$D_s Y(\tau, t, x) = G\big(s, X(s, t, x)\big)^* \nabla_x Y\big(\tau, s, X(s, t, x)\big).$$

As a preliminary step, one has to prove that X, Y, Z belong to a class of processes for which DX, DY, DZ exist and are suitably regular. To our knowledge, differentiability in the sense of Malliavin for solutions of backward evolution equations in infinite dimensions has not been considered before. Malliavin differentiability has been studied in different contexts for solutions of stochastic partial differential equations: see, for example, [32]. For the case of abstract nonlinear evolution equations we refer to [3, 24] and the bibliography therein.

In these papers, however, the diffusion coefficient is assumed to take values in the space $L_2(\Xi, H)$ of Hilbert–Schmidt operators from Ξ to H. This restriction does not allow for instance to cover the case of white noise, that is, when G equals the identity. We remove this restriction by requiring essentially that for every $s > 0$, $t \in [0, T]$, $x \in H$, the operator $e^{sA}G(t, x)$ belongs to $L_2(\Xi, H)$ and a bound of the form

$$\left| \nabla_x \left(e^{sA} G(t, x) \right) h \right|_{L_2(\Xi, H)} \leq L \, s^{-\gamma} |h|,$$

holds for its directional derivative in every direction $h \in H$, where $L > 0$ and γ are suitable constants and we require $\gamma < 1/2$. A blow-up for the Hilbert–Schmidt norm when $s \to 0$ has to be expected if $G(t, x) \notin L_2(\Xi, H)$ and has been observed in the case of stochastic partial differential equations. We also prove that the following bound holds for DX:

$$\sup_{s \in [0,T]} \mathbb{E} \left(\sup_{\tau \in (s,T]} (\tau - s)^{p\gamma} |D_s X_\tau|^p_{L_2(\Xi, H)} \right) < \infty \qquad \text{for every } p \in [2, \infty).$$

Finally, we note that Malliavin calculus for evolution equations in the special and simpler case of constant G was already considered in [4, 13].

As mentioned above, results of equation (1.2) are suitable for application to problems of nonlinear stochastic optimal control. Let us consider a controlled process X^u solution of

$$(1.5) \quad \begin{cases} dX^u_\tau = AX^u_\tau \, d\tau + F(\tau, X^u_\tau) \, d\tau \\ \qquad + C(\tau, X^u_\tau) u(\tau) \, d\tau + G(\tau, X_\tau) \, dW_\tau, \qquad \tau \in [t, T], \\ X^u_t = x \in H, \end{cases}$$

where the control process u takes values in a given subset $\mathcal{U}$ of another Hilbert space U (possibly, $\mathcal{U}$ coincides with U) and C is a function with values in $L(U, H)$. The aim is to choose a control process u, within a set of admissible controls, in such a way to minimize a cost functional of the form:

$$J(t, x, u) = \mathbb{E} \int_t^T g\big(\sigma, X^u_\sigma, u(\sigma)\big) \, d\sigma + \mathbb{E} \phi(X^u_T),$$

where g and ϕ are given real functions. There is a vast literature on such control problems in infinite dimensions: here we report some of the references that are most closely connected with our approach and we refer the reader to the bibliography therein.

In the book [2], Chapter 4, the authors treat the case where $F = 0$, C is the identity, the diffusion coefficient G is constant and satisfies $\text{Trace} \, GG^* < \infty$, and convexity conditions are imposed on the functions g and ϕ. Under some regularity conditions, the authors are able to find a unique classical solution of the corresponding Hamilton–Jacobi–Bellman equation, using techniques of convex analysis and nonlinear semigroup theory, and to solve the optimal control problem.

In [5, 6] the control problem for (1.5) is solved assuming that A is self-adjoint and C and G equal the identity, under various regularity conditions for F, g, ϕ. Existence and uniqueness of a mild solution v of the Hamilton–Jacobi–Bellman equation is proved in a space of functions possessing a derivative $\nabla_x v$ that blows up at $t = 0$. Notice that the assumption that G equals the identity is a nondegeneracy assumption.

In [16, 17], the results of [5, 6] are extended to general (but still constant) G under weaker nondegeneracy conditions. The proofs are based on corresponding regularity properties of the transition semigroup of the associated Ornstein–Uhlenbeck process, that is, the Markov process Y solution of

$$dY_\tau = AY_\tau \, d\tau + G \, dW_\tau, \qquad \tau \in [t, T] \subset [0, T], \ Y_t = x.$$

In the papers mentioned, the authors provide a direct (classical or mild) solution of the Hamilton–Jacobi–Bellman equation for the value function $v(t, x)$ of the control problem, which is then used to prove that the optimal control u is related to the corresponding optimal trajectory X by a feedback law involving $\nabla_x v$.

In this paper we consider a controlled process X^u solution of

$$\begin{cases} dX_\tau^u = AX_\tau^u \, d\tau + F(\tau, X_\tau^u) \, d\tau + G(\tau, X_\tau^u)R(\tau, X_\tau^u)u(\tau) \, d\tau \\ \qquad + G(\tau, X_\tau) \, dW_\tau, \qquad \tau \in [t, T], \\ X_t^u = x \in H, \end{cases}$$

where R is a function with values in $L(U, \Xi)$. Due to the special structure of the control term, the Hamilton–Jacobi–Bellman equation for the value function is of the form (1.2), provided we set

(1.6)
$$\psi_0(t, x, p) = \inf\{g(t, x, u) + \langle p, u \rangle_U : u \in \mathcal{U}\},$$
$$t \in [0, T], x \in H, p \in U,$$

and $\psi(t, x, z) = -\psi_0(t, x, R(t, x)^*z)$ for $z \in \Xi$. The control problem is understood in the usual weak sense (see [12] and Section 7 below). Under suitable conditions, if we let v denote the unique mild solution of (1.2) then we have $J(t, x, u) \geq v(t, x)$ and the equality holds if and only if the following feedback law is verified by u and X^u:

$$u(\sigma) = \Gamma\big(\sigma, X_\sigma^u, R(\sigma, X_\sigma^u)^* G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u)\big),$$

where $u = \Gamma(t, x, p)$ is the minimizer in (1.6). Thus, we can characterize optimal controls by a feedback law. We refer to Theorem 7.2 for precise statements and additional results.

We underline that we are able to remove the restriction on constancy of the coefficient G. Moreover, due to the special structure of the control problem, no nondegeneracy assumptions of any kind are imposed on G. In the Example 7.3.1 we show that our results can be applied to a model of great interest in mathematical

finance, where absence of nondegeneracy assumptions reveals to be essential. A similar problem is studied in [15] by analytic techniques; in that paper, however, weak nondegeneracy assumptions have still to be required.

We wish to mention that applications to stochastic control are presented here mainly to illustrate the effectiveness of our results on the nonlinear Kolmogorov equation. We believe that much better results can be obtained in combination with direct methods: this will be the subject of further study.

As mentioned before, the notion of viscosity solution, developed by many authors, in particular M. Crandall and P. L. Lions and their collaborators, and suitably applied to the study of control problems, is not discussed here. There exists a huge literature on viscosity solutions (see [7] as a general reference), but here we are only concerned with second order Hamilton–Jacobi–Bellman equations on an infinite-dimensional Hilbert space. Generally speaking, the class of equations that can be treated by this method is much more general than those considered in this paper: it includes fully nonlinear operators and Lipschitz conditions are replaced by weaker regularity assumptions. However, none of the results we know are directly applicable to our situation. In the fundamental papers [8, 25, 27], in particular, the boundedness conditions required on the differential operator is not fulfilled if A is a genuinely unbounded operator. In [39, 38], an unbounded linear part of the drift is taken into consideration, but in order to apply these results one has to require that the diffusion term G takes values in the space of Hilbert–Schmidt operators. Only in [18] such a limitation on G has been weakened exploiting specific properties of special unbounded operators A. For specific control systems much better results are available and viscosity solutions methods have been successfully applied to optimal control problems: we mention [21], where a special form of the Hamilton–Jacobi–Bellman equation is considered, and the papers [26] and [19] which deal with control of the Zakai equation, arising in optimal control of partially observable systems.

We also notice that in all the cited references on viscosity solutions the value function v is not differentiable, hence a characterization of the optimal control through feedback law is not available in general. By our methods, we can prove this kind of regularity results only under structural constraints for the equation and regularity conditions on the coefficients. Although it is apparently impossible to cover, using backward stochastic equations, the case of fully nonlinear operators, differentiability conditions on some of the coefficients can be relaxed, under appropriate additional assumptions; see [14].

Extending to the infinite dimensional case the arguments in [34] or [29], it seems possible to show that, at least in some cases, the solutions we find here are viscosity solutions (still there are some delicate points related to the choice of test functions, and this will be the subject of future work). On the contrary, as far as we know, probabilistic methods are of no help for the difficult task of proving uniqueness of the viscosity solution.

The plan of the paper is as follows: Section 2 is devoted to notation and preliminary results. In Section 3, equation (1.1) is studied; in particular, regular dependence on parameters t, x and Malliavin differentiability are proved. In Section 4 we study equation (1.4), where X is a given process, and we investigate regular dependence of the solution (Y, Z) on X. In Section 5, equations (1.1) and (1.4) are studied as a system, Malliavin differentiability is proved for (Y, Z), and analytical properties for the function $v(t, x) = Y(t, t, x)$ are treated. In Section 6 we prove our main result on existence and uniqueness of a mild solution of (1.2) and Section 7 is devoted to applications to optimal control.

2. Notation.

2.1. *Vector spaces and stochastic processes.* The norm of an element x of a Banach space E will be denoted $|x|_E$ or simply $|x|$, if no confusion is possible. If F is another Banach space, $L(E, F)$ denotes the space of bounded linear operators from E to F, endowed with the usual operator norm.

The letters Ξ, H, K will always denote Hilbert spaces. Scalar product is denoted $\langle \cdot, \cdot \rangle$, with a subscript to specify the space, if necessary. All Hilbert spaces are assumed to be real and separable. $L_2(\Xi, K)$ is the space of Hilbert–Schmidt operators from Ξ to K, endowed with the Hilbert–Schmidt norm.

By a cylindrical Wiener process with values in a Hilbert space Ξ, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we mean a family $W(t)$, $t \geq 0$, of linear mappings $\Xi \to L^2(\Omega)$ such that:

  (i) for every $u \in \Xi$, $\{W(t)u, t \geq 0\}$ is a real (continuous) Wiener process;
  (ii) for every $u, v \in \Xi$ and $t \geq 0$, $\mathbb{E}(W(t)u \cdot W(t)v) = \langle u, v \rangle_\Xi$.

In the following, all stochastic processes will be defined on subsets of a fixed time interval $[0, T]$. Short-hand writings "a.a. (a.e.)" mean "almost all (almost everywhere) with respect to the Lebesgue measure." $\mathcal{F}_t$, $t \in [0, T]$, will denote, except that in Section 7, the natural filtration of W, augmented with the family $\mathcal{N}$ of $\mathbb{P}$-null sets of $\mathcal{F}_T$:

$$\mathcal{F}_t = \sigma \big(W(s) : s \in [0, t] \big) \vee \mathcal{N}.$$

The filtration $(\mathcal{F}_t)$ satisfies the usual conditions. All the concepts of measurability for stochastic processes (e.g., predictability, etc.) refer to this filtration.

For $[a, b] \subset [0, T]$ we also use the notation

$$\mathcal{F}_{[a,b]} = \sigma \big(W(s) - W(a) : s \in [a, b] \big) \vee \mathcal{N}.$$

By $\mathcal{P}$ we denote the predictable σ-algebra on $\Omega \times [0, T]$ and by $\mathcal{B}(\Lambda)$ the Borel σ-algebra of any topological space Λ.

Next we define several classes of stochastic processes with values in a Hilbert space K.

- $L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$ denotes the space of equivalence classes of processes $Y \in L^2(\Omega \times [0, T]; K)$, admitting a predictable version. $L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$ is endowed with the norm

$$|Y|^2 = \mathbb{E} \int_0^T |Y_\tau|^2 \, d\tau.$$

- $L^p_{\mathcal{P}}(\Omega; L^2([0, T]; K))$ denotes the space of equivalence classes of processes Y such that the norm

$$|Y|^p = \mathbb{E}\left(\int_0^T |Y_\tau|^2 \, d\tau\right)^{p/2}$$

is finite, and Y admits a predictable version.

- $C_{\mathcal{P}}([0, T]; L^2(\Omega; K))$ denotes the space of K-valued processes Y such that $Y \colon [0, T] \to L^2(\Omega; K)$ is continuous and Y has a predictable modification, endowed with the norm

$$|Y|^2 = \sup_{\tau \in [0,T]} \mathbb{E} |Y_\tau|^2.$$

Elements of $C_{\mathcal{P}}([0, T]; L^2(\Omega; K))$ are identified up to modification.

- $L^p_{\mathcal{P}}(\Omega; C([0, T]; K))$ denotes the space of predictable processes Y with continuous paths in K, such that the norm

$$|Y|^p = \mathbb{E} \sup_{\tau \in [0,T]} |Y_\tau|^p$$

is finite. Elements of $L^p_{\mathcal{P}}(\Omega; C([0, T]; K))$ are identified up to indistiguishability.

Given an element Ψ of $L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$, the Itô stochastic integral $\int_0^t \Psi(\sigma) \, d\sigma$, $t \in [0, T]$, can be defined; it is a K-valued martingale belonging to $L^2_{\mathcal{P}}(\Omega; C([0, T]; K))$.

The previous definitions have obvious extensions to processes defined on subintervals of $[0, T]$.

2.2. *The class $\mathcal{G}$.* In this section we introduce a class of maps acting among Banach spaces, possessing suitable continuity and differentiability properties. Many assumptions in the following sections will be stated in terms of membership in this class.

The class we are going to introduce has several useful properties. First, membership in this class is often easy to verify; see Lemmas 2.2 and 2.3 below. Next, it is a well-behaved class as far as chain rules are concerned. Finally, it is sufficiently large to include operators commonly arising in applications to stochastic partial differential equations, such as Nemytskii (evaluation) operators;

it is well known that that the Nemytskii operators are not Fréchet differentiable except in trivial cases.

In this subsection, X, Y, Z, V denote Banach spaces. We recall that for a mapping $F\colon X \to V$ the directional derivative at point $x \in X$ in the direction $h \in X$ is defined as

$$\nabla F(x; h) = \lim_{s \to 0} \frac{F(x + sh) - F(x)}{s},$$

whenever the limit exists in the topology of V. F is called Gâteaux differentiable at point x if it has directional derivative in every direction at point x and there exists an element of $L(X, V)$, denoted $\nabla F(x)$ and called Gâteaux derivative, such that $\nabla F(x; h) = \nabla F(x)h$ for every $h \in X$.

DEFINITION 2.1. We say that a mapping $F\colon X \to V$ belongs to the class $\mathcal{G}^1(X; V)$ if it is continuous, Gâteaux differentiable on X, and $\nabla F\colon X \to L(X, V)$ is strongly continuous.

The last requirement of the definition means that for every $h \in X$ the map $\nabla F(\cdot)h\colon X \to V$ is continuous. Note that $\nabla F\colon X \to L(X, V)$ is not continuous in general if $L(X, V)$ is endowed with the norm operator topology; clearly, if this happens then F is Fréchet differentiable on X. Some features of the class $\mathcal{G}^1(X, V)$ are collected below.

LEMMA 2.1. *Suppose $F \in \mathcal{G}^1(X, V)$. Then*:

(i) $(x, h) \mapsto \nabla F(x)h$ *is continuous from $X \times X$ to V.*
(ii) *If $G \in \mathcal{G}^1(V, Z)$ then $G(F) \in \mathcal{G}^1(X, Z)$ and $\nabla(G(F))(x) = \nabla G(F(x)) \times \nabla F(x)$.*

In addition to the ordinary chain rule in point (ii) above, a chain rule for the Malliavin derivative operator holds: see Section 3.3. Membership of a map in $\mathcal{G}^1(X, V)$ may be conveniently checked as shown in the following lemma.

LEMMA 2.2. *A map $F\colon X \to V$ belongs to $\mathcal{G}^1(X, V)$ provided the following conditions hold*:

(i) *the directional derivatives $\nabla F(x; h)$ exist at every point $x \in X$ and in every direction $h \in X$*;
(ii) *for every h, the mapping $\nabla F(\cdot; h)\colon X \to V$ is continuous*;
(iii) *for every x, the mapping $h \mapsto \nabla F(x; h)$ is continuous from X to V.*

The proofs of these lemmas are left to the reader.

We need to generalize these definitions to functions depending on several variables. For a function $F\colon X \times Y \to V$ the partial directional and Gâteaux

derivatives with respect to the first argument, at point (x, y) and in the direction $h \in X$, are denoted $\nabla_x F(x, y; h)$ and $\nabla_x F(x, y)$ respectively, their definitions being obvious.

DEFINITION 2.2. We say that a mapping $F\colon X \times Y \to V$ belongs to the class $\mathcal{G}^{1,0}(X \times Y; V)$ if it is continuous, Gâteaux differentiable with respect to x on $X \times Y$, and $\nabla_x F\colon X \times Y \to L(X, V)$ is strongly continuous.

As in Lemma 2.1 one can prove that for $F \in \mathcal{G}^{1,0}(X \times Y, V)$ the mapping $(x, y, h) \mapsto \nabla_x F(x, y)h$ is continuous from $X \times Y \times X$ to V, and analogues of the previously stated chain rules hold. The following result is proved as Lemma 2.2 (but note that continuity is explicitly required).

LEMMA 2.3. *A continuous map* $F\colon X \times Y \to V$ *belongs to* $\mathcal{G}^{1,0}(X \times Y, V)$ *provided the following conditions hold*:

(i) *the directional derivatives* $\nabla_x F(x, y; h)$ *exist at every point* $(x, y) \in X \times Y$ *and in every direction* $h \in X$;
(ii) *for every* h, *the mapping* $\nabla F(\cdot, \cdot; h)\colon X \times Y \to V$ *is continuous*;
(iii) *for every* (x, y), *the mapping* $h \mapsto \nabla_x F(x, y; h)$ *is continuous from* X *to* V.

When F depends on additional arguments, the previous definitions and properties have obvious generalizations. For instance, we say that $F\colon X \times Y \times Z \to V$ belongs to $\mathcal{G}^{1,1,0}(X \times Y \times Z; V)$ if it is continuous, Gâteaux differentiable with respect to x and y on $X \times Y \times Z$, and $\nabla_x F\colon X \times Y \times Z \to L(X, V)$ and $\nabla_y F\colon X \times Y \times Z \to L(Y, V)$ are strongly continuous.

We will make systematic use of a parameter depending contraction principle, stated below as Proposition 2.4. It will be used to study regular dependence of solution of stochastic equations on their initial data, which is crucial to investigate regularity properties of the non linear Kolmogorov equation which is the object of our paper. The first part of the following proposition is proved in [41], Theorems 10.1 and 10.2. The second part is an immediate corollary.

PROPOSITION 2.4 (Parameter depending contraction principle). *Let* $F\colon X \times Y \to X$ *be a continuous mapping satisfying*

$$|F(x_1, y) - F(x_2, y)| \le \alpha |x_1 - x_2|,$$

for some $\alpha \in [0, 1)$ *and every* $x_1, x_2 \in X$, $y \in Y$. *Let* $\phi(y)$ *denote the unique fixed point of the mapping* $F(\cdot, y)\colon X \to X$. *Then* $\phi\colon Y \to X$ *is continuous. If, in addition,* $F \in \mathcal{G}^{1,1}(X \times Y, X)$, *then* $\phi \in \mathcal{G}^1(Y, X)$ *and*

$$\nabla \phi(y) = \nabla_x F(\phi(y), y) \nabla \phi(y) + \nabla_y F(\phi(y), y), \qquad y \in Y.$$

More generally, let $F\colon X \times Y \times Z \to X$ *be a continuous mapping satisfying*

$$|F(x_1, y, z) - F(x_2, y, z)| \leq \alpha |x_1 - x_2|,$$

for some $\alpha \in [0, 1)$ *and every* $x_1, x_2 \in X$, $y \in Y$, $z \in Z$. *Let* $\phi(y, z)$ *denote the unique fixed point of the mapping* $F(\cdot, y, z)\colon X \to X$. *Then* $\phi\colon Y \times Z \to X$ *is continuous. If, in addition,* $F \in \mathcal{G}^{1,1,0}(X \times Y \times Z, X)$, *then* $\phi \in \mathcal{G}^{1,0}(Y \times Z, X)$ *and*

$$\nabla_y \phi(y, z) = \nabla_x F(\phi(y, z), y, z) \nabla_y \phi(y, z)$$
$$+ \nabla_y F(\phi(y, z), y, z), \qquad y \in Y, \, z \in Z.$$

3. The forward equation.

3.1. *Existence, uniqueness and regularity.* Let $W(t)$, $t \in [0, T]$, be a cylindrical Wiener process with values in a Hilbert space Ξ, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We fix an interval $[t, T] \subset [0, T]$ and we consider the Itô stochastic differential equation for an unknown process X_τ, $\tau \in [t, T]$, with values in a Hilbert space H:

$$(3.1) \quad \begin{cases} dX_\tau = AX_\tau \, d\tau + F(\tau, X_\tau) \, d\tau + G(\tau, X_\tau) \, dW_\tau, & \tau \in [t, T], \\ X_t = x \in H. \end{cases}$$

The precise notion of solution will be given next. We assume the following:

HYPOTHESIS 3.1. (i) *The operator* A *is the generator of a strongly continuous semigroup* e^{tA}, $t \geq 0$, *in the Hilbert space* H.

(ii) *The mapping* $F\colon [0, T] \times H \to H$ *is measurable and satisfies, for some constant* $L > 0$,

$$|F(t, x) - F(t, y)| \leq L \, |x - y|, \qquad t \in [0, T], \, x, y \in H.$$

(iii) G *is a mapping* $[0, T] \times H \to L(\Xi, H)$ *such that for every* $v \in \Xi$ *the map* $Gv\colon [0, T] \times H \to H$ *is measurable,* $e^{sA}G(t, x) \in L_2(\Xi, H)$ *for every* $s > 0$, $t \in [0, T]$ *and* $x \in H$, *and*

$$|e^{sA}G(t, x)|_{L_2(\Xi, H)} \leq Ls^{-\gamma}(1 + |x|),$$

$$(3.2) \quad |e^{sA}G(t, x) - e^{sA}G(t, y)|_{L_2(\Xi, H)} \leq Ls^{-\gamma}|x - y|,$$

$$s > 0, \, t \in [0, T], \, x, y \in H,$$

$$(3.3) \qquad\qquad |G(t, x)|_{L(\Xi, H)} \leq L(1 + |x|), \qquad t \in [0, T], \, x, y \in H,$$

for some constants $L > 0$ *and* $\gamma \in [0, 1/2)$.

(iv) *For every* $s > 0$, $t \in [0, T]$, $v \in \Xi$,

$$F(t, \cdot) \in \mathcal{G}^1(H, H), \qquad e^{sA}G(t, \cdot) \in \mathcal{G}^1(H, L_2(\Xi, H)).$$

By a solution of equation (3.1) we mean an $(\mathcal{F}_t)$-predictable process X_τ, $\tau \in [t, T]$, with continuous paths in H, such that, $\mathbb{P}$-a.s.,

(3.4)
$$X_\tau = e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-\sigma)A} F(\sigma, X_\sigma)\, d\sigma$$
$$+ \int_t^\tau e^{(\tau-\sigma)A} G(\sigma, X_\sigma)\, dW_\sigma, \qquad \tau \in [t, T].$$

To stress dependence on initial data, we denote the solution by $X(\tau, t, x)$. Note that $X(\tau, t, x)$ is $\mathcal{F}_{[t,T]}$-measurable, hence independent of $\mathcal{F}_t$.

The inequalities (3.3) and Hypothesis 3.1(iv) are needed to have additional regularity for the process X, but they are not used in Proposition 3.2 below. It is a consequence of our assumptions that for every $s > 0$, $t \in [0, T]$, $v \in \Xi$, $x, h \in H$,

(3.5) $|\nabla_x F(t, x)h| \leq L|h|$, $|\nabla_x(e^{sA}G(t, x))h|_{L_2(\Xi, H)} \leq L\, s^{-\gamma}|h|$.

PROPOSITION 3.2.   *Under the assumptions of Hypothesis* 3.1(i)–(iii), *for every $p \in [2, \infty)$ there exists a unique process $X \in L_{\mathcal{P}}^p(\Omega; C([t, T]; H))$ solution of* (3.4). *Moreover,*

(3.6)
$$\mathbb{E} \sup_{\tau \in [t,T]} |X_\tau|^p \leq C(1 + |x|)^p,$$

for some constant C depending only on p, γ, T, L and $M := \sup_{\tau \in [t,T]} |e^{\tau A}|$.

PROOF.   The result is well known; see, for example, [10], Theorem 5.3.1. We include the proof for completeness and because it will be useful in the following. The argument is as follows: we define a mapping Φ from $L_{\mathcal{P}}^p(\Omega; C([t, T]; H))$ to itself by the formula

$$\Phi(X)_\tau = e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-\sigma)A} F(\sigma, X_\sigma)\, d\sigma$$
$$+ \int_t^\tau e^{(\tau-\sigma)A} G(\sigma, X_\sigma)\, dW_\sigma, \qquad \tau \in [t, T],$$

and show that it is a contraction, under an equivalent norm. The unique fixed point is the required solution.

For simplicity, we set $t = 0$ and we treat only the case $F = 0$, the general case being handled in a similar way. Let us introduce the norm $\|X\|^p = \mathbb{E} \sup_{\tau \in [0,T]} e^{-\beta \tau p}|X_\tau|^p$, where $\beta > 0$ will be chosen later. In the space $L^p(\Omega; C([0, T]; H))$ this norm is equivalent to the original one. We will use the so called factorization method; see [10], Theorem 5.2.5. Let us take $p > 2$ and $\alpha \in (0, 1)$ such that

$$\frac{1}{p} < \alpha < \frac{1}{2} - \gamma \quad \text{and let } c_\alpha^{-1} = \int_\sigma^\tau (\tau - s)^{\alpha-1}(s - \sigma)^{-\alpha}\, ds.$$

Then, by the stochastic Fubini theorem,

$$
\begin{aligned}
\Phi(X)_\tau &= e^{\tau A}x \\
&\quad + c_\alpha \int_0^\tau \int_\sigma^\tau (\tau - s)^{\alpha-1}(s-\sigma)^{-\alpha}e^{(\tau-s)A}e^{(s-\sigma)A}\,ds\,G(\sigma, X_\sigma)\,dW_\sigma \\
&= e^{\tau A}x + c_\alpha \int_0^\tau (\tau - s)^{\alpha-1}e^{(\tau-s)A}Y_s\,ds,
\end{aligned}
$$

where

$$
Y_s = \int_0^s (s-\sigma)^{-\alpha}e^{(s-\sigma)A}G(\sigma, X_\sigma)\,dW_\sigma.
$$

By the Hölder inequality, setting $M = \sup_{\tau \in [0,T]}|e^{\tau A}|$, $p' = p/(p-1)$,

$$
\begin{aligned}
e^{-\beta\tau}&\left|\int_0^\tau (\tau-s)^{\alpha-1}e^{(\tau-s)A}Y_s\,ds\right| \\
&\leq \left(\int_0^\tau e^{-p'\beta(\tau-s)}(\tau-s)^{(\alpha-1)p'}\,ds\right)^{1/p'}\left(\int_0^\tau e^{-p\beta s}|e^{(\tau-s)A}Y_s|^p\,ds\right)^{1/p} \\
&\leq M\left(\int_0^T e^{-p'\beta s}s^{(\alpha-1)p'}\,ds\right)^{1/p'}\left(\int_0^T e^{-p\beta s}|Y_s|^p\,ds\right)^{1/p},
\end{aligned}
$$

and we obtain

$$
\|\Phi(X)\| \leq M|x| + Mc_\alpha\left(\int_0^T e^{-p'\beta s}s^{(\alpha-1)p'}\,ds\right)^{1/p'}\left(\mathbb{E}\int_0^T e^{-p\beta s}|Y_s|^p\,ds\right)^{1/p}.
$$

By the Burkholder–Davis–Gundy inequalities, taking into account the assumption (3.2), we have, for some constant c_p depending only on p,

$$
\begin{aligned}
\mathbb{E}|Y_s|^p &\leq c_p\mathbb{E}\left(\int_0^s (s-\sigma)^{-2\alpha}|e^{(s-\sigma)A}G(\sigma, X_\sigma)|^2_{L_2(\Xi,H)}\,d\sigma\right)^{p/2} \\
&\leq L^p c_p\mathbb{E}\left(\int_0^s (s-\sigma)^{-2\alpha-2\gamma}(1+|X_\sigma|)^2\,d\sigma\right)^{p/2} \\
&\leq L^p c_p\mathbb{E}\sup_{\sigma\in[0,s]}[(1+|X_\sigma|)^p e^{-p\beta\sigma}]\left(\int_0^s (s-\sigma)^{-2\alpha-2\gamma}e^{2\beta\sigma}\,d\sigma\right)^{p/2},
\end{aligned}
$$

which implies

$$
\begin{aligned}
e^{-p\beta s}\mathbb{E}|Y_s|^p &\leq L^p c_p(1+\|X\|^p)\left(\int_0^s (s-\sigma)^{-2\alpha-2\gamma}e^{-2\beta(s-\sigma)}\,d\sigma\right)^{p/2} \\
&\leq L^p c_p(1+\|X\|^p)\left(\int_0^T \sigma^{-2\alpha-2\gamma}e^{-2\beta\sigma}\,d\sigma\right)^{p/2}.
\end{aligned}
$$

We conclude that

$$\|\Phi(X)\| \leq M|x| + MLc_\alpha \big(Tc_p(1+\|X\|^p)\big)^{1/p}$$
$$\times \left(\int_0^T e^{-p'\beta s} s^{(\alpha-1)p'}\,ds\right)^{1/p'} \left(\int_0^T \sigma^{-2\alpha-2\gamma} e^{-2\beta\sigma}\,d\sigma\right)^{1/2}.$$

This shows that Φ is a well defined mapping on $L^p(\Omega; C([0,T]; H))$. If X, X^1 are processes belonging to this space, similar passages show that

$$\|\Phi(X) - \Phi(X^1)\| \leq MLc_\alpha (Tc_p)^{1/p} \|X - X^1\|$$
$$\times \left(\int_0^T e^{-p'\beta s} s^{(\alpha-1)p'}\,ds\right)^{1/p'} \left(\int_0^T \sigma^{-2\alpha-2\gamma} e^{-2\beta\sigma}\,d\sigma\right)^{1/2},$$

so that, for β sufficiently large, the mapping Φ is a contraction.

In particular we obtain $\|X\| \leq C(1+|x|)$, which proves the estimate (3.6). $\square$

3.2. *Regular dependence on parameters.* For further developments we need to investigate the dependence of the solution $X(\tau, t, x)$ on the initial data x and t. We first reformulate equation (3.4) as an equation on $[0, T]$. We set

$$(3.7) \qquad S(\tau) = e^{\tau A} \quad \text{for } \tau \geq 0, \qquad S(\tau) = I \quad \text{for } \tau < 0,$$

and we consider the equation

$$(3.8) \quad \begin{aligned} X_\tau &= S(\tau - t)x + \int_0^\tau \mathbb{1}_{[t,T]}(\sigma) S(\tau - \sigma) F(\sigma, X_\sigma)\,d\sigma \\ &\quad + \int_0^\tau \mathbb{1}_{[t,T]}(\sigma) S(\tau - \sigma) G(\sigma, X_\sigma)\,dW_\sigma, \end{aligned}$$

for the unknown process X_τ, $\tau \in [0, T]$. Under the assumptions of Hypothesis 3.1, equation (3.8) has a unique solution $X \in L^p_{\mathcal{P}}(\Omega; C([0,T]; H))$ for every $p \in [2, \infty)$. It clearly satisfies $X_\tau = x$ for $\tau \in [0, t)$, and its restriction to the time interval $[t, T]$ is the unique solution of (3.4). From now on we denote by $X(\tau, t, x)$, $\tau \in [0, T]$, the solution of (3.8).

PROPOSITION 3.3. *Assume Hypothesis* 3.1. *Then, for every* $p \in [2, \infty)$, *the following hold*:

(i) *The map* $(t, x) \mapsto X(\cdot, t, x)$ *belongs to* $\mathcal{G}^{0,1}([0,T] \times H, L^p_{\mathcal{P}}(\Omega; C([0,T]; H)))$.

(ii) *Denoting by* $\nabla_x X$ *the partial Gâteaux derivative, for every direction* $h \in H$ *the directional derivative process* $\nabla_x X(\tau, t, x)h$, $\tau \in [0, T]$ *solves*, $\mathbb{P}$-*a.s., the equation*

$$(3.9) \quad \begin{cases} \nabla_x X(\tau, t, x)h \\ \quad = e^{(\tau-t)A}h + \int_t^\tau e^{(\tau-\sigma)A}\nabla_x F(\sigma, X(\sigma, t, x))\nabla_x X(\sigma, t, x)h \, d\sigma \\ \quad + \int_t^\tau \nabla_x(e^{(\tau-\sigma)A}G(\sigma, X(\sigma, t, x)))\nabla_x X(\sigma, t, x)h \, dW_\sigma, \\ \hspace{9cm} \tau \in [t, T], \\ \nabla_x X(\tau, t, x)h = h, \qquad \tau \in [0, t). \end{cases}$$

(iii) *Finally* $|\nabla_x X(\tau, t, x)h|_{L^p_{\mathcal{P}}(\Omega; C([0,T]; H))} \leq c\,|h|$ *for some constant* c.

PROOF.   Let us consider again the map Φ defined in the proof of Proposition 3.2. In our present notation, Φ can be seen as a mapping from $L^p_{\mathcal{P}}(\Omega; C([0, T];$ $H)) \times [0, T] \times H$ to $L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$:

$$\Phi(X, t, x)_\tau = S(\tau - t)x + \int_0^\tau \mathbb{1}_{[t,T]}(\sigma)S(\tau - \sigma)F(\sigma, X_\sigma)\,d\sigma$$

$$+ \int_0^\tau \mathbb{1}_{[t,T]}(\sigma)S(\tau - \sigma)G(\sigma, X_\sigma)\,dW_\sigma,$$

for $\tau \in [0, T]$. By the arguments of the proof of Proposition 3.2, $\Phi(\cdot, t, x)$ is a contraction in $L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$, under an equivalent norm, uniformly with respect to t, x. The process $X(\cdot, t, x)$ is the unique fixed point of $\Phi(\cdot, t, x)$. So, by the parameter depending contraction principle (Proposition 2.4), it suffices to show that

$$\Phi \in \mathcal{G}^{1,0,1}\big(L^p_{\mathcal{P}}(\Omega; C([0, T]; H)) \times [0, T] \times H, L^p_{\mathcal{P}}(\Omega; C([0, T]; H))\big).$$

By an obvious extension of Lemma 2.3, the proof is concluded by the following steps.

Step 1. Φ is continuous. We have already noticed that $\Phi(\cdot, t, x)$ is a contraction, uniformly with respect to $x \in H$ and $t \in [0, T]$, and so $\Phi(\cdot, t, x)$ is continuous, uniformly in t, x. Moreover for fixed X it is easy to verify that $\Phi(X, \cdot, \cdot)$ is continuous from $[0, T] \times H$ to $L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$.

Step 2.   The directional derivative $\nabla_X \Phi(X, t, x; N)$ in the direction $N \in L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$ is the process given by

$$\nabla_X \Phi(X, t, x; N)_\tau = \int_t^\tau e^{(\tau-\sigma)A}\nabla_x F(\sigma, X_\sigma)N_\sigma\,d\sigma$$

$$+ \int_t^\tau \nabla_x\big(e^{(\tau-\sigma)A}G(\sigma, X_\sigma)\big)N_\sigma\,dW_\sigma, \qquad \tau \in [t, T],$$

$$\nabla_X \Phi(X, t, x; N)_\tau = 0, \qquad \tau \in [0, t);$$

moreover, the mappings $(X, t, x) \mapsto \nabla_X \Phi(X, t, x; N)$ and $N \mapsto \nabla_X \Phi(X, t, x; N)$ are continuous.

We limit ourselves to prove this claim in the special case $F = 0$, the general case being a straightforward extension. For fixed $t \geq 0$ and $x \in H$, for all $\tau \geq t$:

$$
\begin{aligned}
I_\tau^\varepsilon &:= \frac{1}{\varepsilon} \Phi(X + \varepsilon N, t, x)_\tau - \frac{1}{\varepsilon} \Phi(X, t, x)_\tau - \int_t^\tau \nabla_x(e^{(\tau-\sigma)A} G(\sigma, X_\sigma)) N_\sigma dW_\sigma \\
&= \int_t^\tau \left(\int_0^1 (\nabla_x(e^{(\tau-\sigma)A} G(\sigma, X_\sigma + \zeta \varepsilon N_\sigma)) N_\sigma \right. \\
&\qquad\qquad \left. - \nabla_x(e^{(\tau-\sigma)A} G(\sigma, X_\sigma)) N_\sigma) d\zeta \right) dW_\sigma.
\end{aligned}
$$

Proceeding as in the proof of Proposition 3.2 (with $\beta = 0$) we get for $1/p < \alpha < 1/2 - \gamma$ and for a suitable constant c_p

$$
|I^\varepsilon|_{L^p_{\mathscr{P}}(\Omega; C([0,T];H))}^p \leq c_p \mathbb{E} \int_t^T |Y_s^\varepsilon|^p \, ds,
$$

where

$$
\begin{aligned}
Y_s^\varepsilon = \int_t^s (s-\sigma)^{-\alpha} \left(\int_0^1 (\nabla_x(e^{(s-\sigma)A} G(\sigma, X_\sigma + \zeta \varepsilon N_\sigma)) N_\sigma \right. \\
\left. - \nabla_x(e^{(s-\sigma)A} G(\sigma, X_\sigma)) N_\sigma) d\zeta \right) dW_\sigma.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}|Y_s^\varepsilon|^p \leq c\mathbb{E} \left(\int_t^s (s-\sigma)^{-2\alpha} \left| \int_0^1 (\nabla_x(e^{(s-\sigma)A} G(\sigma, X_\sigma + \zeta \varepsilon N_\sigma)) N_\sigma \right. \right. \\
\left. \left. - \nabla_x(e^{(s-\sigma)A} G(\sigma, X_\sigma)) N_\sigma) d\zeta \right|_{L_2(\Xi,H)}^2 d\sigma \right)^{p/2}
\end{aligned}
$$

for a suitable constant c. Since for all ε

$$
\left| \int_0^1 \nabla_x(e^{(s-\sigma)A} G(\sigma, X_\sigma + \zeta \varepsilon N_\sigma)) N_\sigma \, d\zeta \right|_{L_2(\Xi,H)} \leq L(s-\sigma)^{-\gamma} |N|_{C([0,T],H)}
$$

and $\nabla_x(e^{sA} G(t,x)v)$ is continuous in x then, by dominated convergence, we get $\mathbb{E} \int_t^T |Y_s^\varepsilon|^p ds \to 0$ and the claim follows.

Continuity of the mappings $(X, t, x) \mapsto \nabla_X \Phi(X, t, x; N)$ and $N \mapsto \nabla_X \Phi(X, t, x; N)$ can be proved in a similar way.

Step 3. Finally, it is clear that the directional derivative $\nabla_x \Phi(X, t, x; h)$ in the direction $h \in H$ is the process given by

$$
\nabla_x \Phi(X, t, x; h)_\tau = \begin{cases} e^{(\tau-t)A} h, & \tau \in [t, T], \\ h, & \tau \in [0, t), \end{cases}
$$

and that the mappings $(X, t, x) \mapsto \nabla_x \Phi(X, t, x; h)$ and $h \mapsto \nabla_x \Phi(X, t, x; h)$ are continuous.

To complete the proof we notice that the estimate in (iii) is a trivial consequence of equation (3.9) and the fact that $|\nabla_X \Phi|$ is uniformly bounded by a constant < 1, by the contraction property of Φ. $\square$

3.3. *Regularity in the Malliavin spaces.* In order to state the main results we need to recall some basic definitions from the Malliavin calculus. We refer the reader to the book [32] for a detailed exposition; the paper [20] treats the extensions to Hilbert space valued random variables and processes.

For every $h \in L^2([0, T]; \Xi)$ we denote by $W(h)$ the integral $\int_0^T \langle h(t), dW(t) \rangle_\Xi$. Given a Hilbert space K, let S_K be the set of K-valued random variables F of the form

$$F = \sum_{j=1}^m f_j\big(W(h_1), \ldots, W(h_n)\big)e_j,$$

where $h_1, \ldots, h_n \in L^2([0, T]; \Xi)$, $\{e_j\}$ is a basis of K and $f_1, \ldots, f_m$ are infinitely differentiable functions $\mathbb{R}^n \to \mathbb{R}$ bounded together with all their derivatives. The Malliavin derivative DF of $F \in S_K$ is defined as the process $D_s F$, $s \in [0, T]$,

$$D_s F = \sum_{j=1}^m \sum_{k=1}^n \partial_k f_j\big(W(h_1), \ldots, W(h_n)\big)e_j \otimes h_k(s),$$

with values in $L_2(\Xi, K)$; by ∂_k we denote the partial derivatives with respect to the kth variable and by $e_j \otimes h_k(s)$ the operator $u \mapsto e_j \langle h_k(s), u \rangle_\Xi$. It is known that the operator $D \colon S_K \subset L^2(\Omega; K) \to L^2(\Omega \times [0, T]; L_2(\Xi; K))$ is closable. We denote by $\mathbb{D}^{1,2}(K)$ the domain of its closure, and use the same letter to denote D and its closure:

$$D \colon \mathbb{D}^{1,2}(K) \subset L^2(\Omega; K) \to L^2\big(\Omega \times [0, T]; L_2(\Xi; K)\big).$$

The adjoint operator of D,

$$\delta \colon \operatorname{dom}(\delta) \subset L^2\big(\Omega \times [0, T]; L_2(\Xi; K)\big) \to L^2(\Omega; K),$$

is called Skorohod integral. It is known that $\operatorname{dom}(\delta)$ contains $L_{\mathscr{P}}^2(\Omega \times [0, T]; L_2(\Xi; K))$ and the Skorohod integral of a process in this space coincides with the Itô integral; $\operatorname{dom}(\delta)$ also contains the class $\mathbb{L}^{1,2}(L_2(\Xi; K))$, the latter being defined as the space of processes $u \in L^2(\Omega \times [0, T]; L_2(\Xi; K))$ such that $u_r \in \mathbb{D}^{1,2}(L_2(\Xi, K))$ for a.e. $r \in [0, T]$ and there exists a measurable version of $D_s u_r$ satisfying

$$\|u\|_{\mathbb{L}^{1,2}(L_2(\Xi;K))}^2 = \|u\|_{L^2(\Omega \times [0,T]; L_2(\Xi;K))}^2$$
$$+ \mathbb{E} \int_0^T \int_0^T |D_s u_r|_{L_2(\Xi, L_2(\Xi, K))}^2 \, dr \, ds < \infty.$$

Moreover, $\|\delta(u)\|^2_{L^2(\Omega;K)} \leq \|u\|^2_{\mathbb{L}^{1,2}(L_2(\Xi;K))}$. The definition of $\mathbb{L}^{1,2}(K)$ for an arbitrary Hilbert space K is entirely analogous.

We recall that if $F \in \mathbb{D}^{1,2}(K)$ is $\mathcal{F}_t$-adapted then $DF = 0$ a.s. on $\Omega \times (t, T]$.

Finally, we need to define the space $\mathbb{D}^{1,2}_{\mathrm{loc}}(K)$. If $F \in \mathbb{D}^{1,2}(K)$ and $F = 0$ on a measurable subset $A \subset \Omega$ then $\mathbb{1}_A DF = 0$; this follows immediately from the corresponding result for $K = \mathbb{R}^d$ ([33], Lemma 2.6). Therefore the following definition is meaningful: we say that a random variable $F: \Omega \to K$ belongs to the space $\mathbb{D}^{1,2}_{\mathrm{loc}}(K)$ if there exists an increasing sequence of measurable subsets $\Omega_k \subset \Omega$ and elements $F_k \in \mathbb{D}^{1,2}(K)$ such that $\cup_k \Omega_k = \Omega$, $\mathbb{P}$-a.s. and $\mathbb{1}_{\Omega_k} F = \mathbb{1}_{\Omega_k} F_k$. $DF: \Omega \times [0, T] \to L_2(\Xi, K)$ is then defined by requiring $\mathbb{1}_{\Omega_k} DF = \mathbb{1}_{\Omega_k} DF_k$. The following chain rule holds; the proof consists in standard approximation arguments and is left to the reader.

LEMMA 3.4. *Suppose K, H are Hilbert spaces, $\psi \in \mathcal{G}^1(K, H)$ and*

(3.10)
$$\sup_{|x| \leq n} |\nabla \psi(x)|_{L(K,H)} < \infty, \qquad n = 1, 2, \ldots.$$

(i) *If $F \in \mathbb{D}^{1,2}_{\mathrm{loc}}(K)$ then $\psi(F) \in \mathbb{D}^{1,2}_{\mathrm{loc}}(H)$.*

(ii) *If $F \in \mathbb{D}^{1,2}(K)$ and $\sup_{x \in K} |\nabla \psi(x)|_{L(K,H)} < \infty$ then $\psi(F) \in \mathbb{D}^{1,2}(H)$.*

(iii) *More generally, if $F \in \mathbb{D}^{1,2}(K)$, (3.10) holds and*

$$\mathbb{E}|\psi(F)|^2_H < \infty, \qquad \mathbb{E} \int_0^T |\nabla \psi(F) D_s F|^2_{L_2(K,H)} \, ds < \infty,$$

then $\psi(F) \in \mathbb{D}^{1,2}(H)$.

In any of the cases (i)–(iii) *we have $D\psi(F) = \nabla \psi(F) DF$.*

Now let us consider again the process $X = \{X(\tau, t, x), \tau \in [t, T]\}$, denoted simply (X_τ), solution of (3.4), with (t, x) fixed. We set as before $X_\tau = x$, $\tau \in [0, t)$. We will soon prove that X belongs to $\mathbb{L}^{1,2}(H)$. Then it is clear that the equality $D_s X_\tau = 0$, $\mathbb{P}$-a.s., holds for a.a. s, t, τ if $\tau < t$ or $s > \tau$.

PROPOSITION 3.5. *Assume Hypothesis* 3.1. *Then the following properties hold*:

(i) $X \in \mathbb{L}^{1,2}(H)$.

(ii) *There exists a version of DX such that for every $s \in [0, T)$, $\{D_s X_\tau, \tau \in (s, T]\}$ is a predictable process in $L_2(\Xi, H)$ with continuous paths satisfying, for every $p \in [2, \infty)$,*

(3.11)
$$\sup_{s \in [0,T]} \mathbb{E}\left(\sup_{\tau \in (s,T]} (\tau - s)^{p\gamma} |D_s X_\tau|^p_{L_2(\Xi,H)} \right) \leq c,$$

where $c > 0$ depends only on p, L, T, γ and $M = \sup_{\tau \in [0,T]} |e^{\tau A}|$; moreover,
$\mathbb{P}$-*a.s.*,

$$
\begin{aligned}
(3.12) \quad D_s X_\tau = {} & e^{(\tau - s)A} G(s, X_s) + \int_s^\tau e^{(\tau - \sigma)A} \nabla_x F(\sigma, X_\sigma) D_s X_\sigma \, d\sigma \\
& + \int_s^\tau \nabla_x (e^{(\tau - \sigma)A} G(\sigma, X_\sigma)) D_s X_\sigma \, dW_\sigma, \qquad \tau \in (s, T].
\end{aligned}
$$

Moreover, $X_\tau \in \mathbb{D}^{1,2}(H)$ for every $\tau \in [0, T]$.

(iii) *Given any element v of Ξ, the process $Q_{s\tau} = D_s X_\tau v$ is a solution of the equation*

$$
\begin{aligned}
(3.13) \quad Q_{s\tau} = {} & e^{(\tau - s)A} G(s, X_s) v + \int_s^\tau e^{(\tau - \sigma)A} \nabla_x F(\sigma, X_\sigma) Q_{s\sigma} \, d\sigma \\
& + \int_s^\tau \nabla_x (e^{(\tau - \sigma)A} G(\sigma, X_\sigma)) Q_{s\sigma} \, dW_\sigma, \qquad \mathbb{P}\text{-}a.s.
\end{aligned}
$$

for a.a. s, τ with $t \le s \le \tau \le T$. It is unique in the sense that if $\{Q_{s\tau}, t \le s \le \tau \le T\}$ is another process with values in H such that $\{Q_{s\tau}, \tau \in [s, T]\}$ is predictable for every $s \in [t, T]$ and $\mathbb{E} \int_t^T \int_s^T |Q_{s\tau}|^2 \, d\tau \, ds < \infty$ then, for a.a. s, τ, we have $Q_{s\tau} = D_s X_\tau v$, $\mathbb{P}$-a.s.

(iv) *Given $v \in \Xi$ and $p \in [2, \infty)$, there exists a version of DXv such that the map $(s, \tau) \mapsto D_s X_\tau v$, defined for $t \le s \le \tau \le T$, is continuous (hence uniformly continuous and bounded) with values in $L^p(\Omega; H)$. Moreover,*

$$
(3.14) \qquad G(s, X_s) v = \lim_{\tau \downarrow s} D_s X_\tau v
$$

in the norm of $L^p(\Omega; H)$.

In order to prove this proposition we need some preparation. We start with the following lemma.

LEMMA 3.6. *If $X \in \mathbb{L}^{1,2}(H)$ then the random processes*

$$
\int_0^\tau e^{(\tau - r)A} F(r, X_r) \, dr, \qquad \int_0^\tau e^{(\tau - r)A} G(r, X_r) \, dW_r, \qquad \tau \in [0, T],
$$

belong to $\mathbb{L}^{1,2}(H)$ and for a.a. s and τ with $s < \tau$,

$$
D_s \int_0^\tau e^{(\tau - r)A} F(r, X_r) \, dr = \int_s^\tau e^{(\tau - r)A} \nabla_x F(r, X_r) D_s X_r \, dr,
$$

$$
\begin{aligned}
(3.15) \quad & D_s \int_0^\tau e^{(\tau - r)A} G(r, X_r) \, dW_r \\
& \quad = e^{(\tau - s)A} G(s, X_s) + \int_s^\tau \nabla_x (e^{(\tau - r)A} G(r, X_r)) D_s X_r \, dW_r.
\end{aligned}
$$

PROOF. We will prove only (3.15). We need the following fact, proved in [20], Proposition 3.4: if $u \in \mathbb{L}^{1,2}(L_2(\Xi, H))$, and for a.a. s the process $\{D_s u_r, r \in [0, T]\}$ belongs to dom(δ), and the map $s \mapsto \delta(D_s u)$ belongs to $L^2(\Omega \times [0, T]; L_2(\Xi, H))$, then $\delta(u) \in \mathbb{D}^{1,2}(H)$ and $D_s \delta(u) = u_s + \delta(D_s u)$.

We fix τ and we apply this result to the process $u_r = e^{(\tau-r)A} G(r, X_r)$ (we set $u_r = 0$ for $r > \tau$). First notice that

$$\mathbb{E} \int_0^T |u_r|^2 \, dr = \mathbb{E} \int_0^\tau |e^{(\tau-r)A} G(r, X_r)|^2_{L_2(\Xi, H)} \, dr$$

$$\leq L^2 \mathbb{E} \int_0^\tau (\tau - r)^{-2\gamma} (1 + |X_r|)^2 \, dr.$$

The right-hand side is finite for a.a. τ; indeed, by exchanging the integrals we verify that

$$\int_0^T \left(\mathbb{E} \int_0^\tau (\tau - r)^{-2\gamma} (1 + |X_r|)^2 \, dr \right) d\tau$$

$$\leq \int_0^T r^{-2\gamma} \, dr \int_0^T \mathbb{E} (1 + |X_r|)^2 \, dr < \infty,$$

since $X \in \mathbb{L}^{1,2}(H) \subset L^2(\Omega \times [0, T]; H)$. Next, for every r, by the chain rule for Malliavin derivative [Lemma 3.4(ii)], $D_s u_r = \nabla_x (e^{(\tau-r)A} G(r, X_r)) D_s X_r$ for a.a. $s < r$, whereas $D_s u_r = 0$ for a.a. $s > r$, by adaptedness. Next, recalling (3.5),

$$\mathbb{E} \int_0^T |D_s u_r|^2 \, dr = \mathbb{E} \int_s^\tau |\nabla_x (e^{(\tau-r)A} G(r, X_r)) D_s X_r|^2_{L_2(\Xi, L_2(\Xi, H))} \, dr$$

$$\leq L^2 \mathbb{E} \int_s^\tau (\tau - r)^{-2\gamma} |D_s X_r|^2_{L_2(\Xi, H)} \, dr,$$

so that

$$\mathbb{E} \int_0^T \int_0^T |D_s u_r|^2 \, dr \, ds \leq L^2 \mathbb{E} \int_0^\tau \int_s^\tau (\tau - r)^{-2\gamma} |D_s X_r|^2_{L_2(\Xi, H)} \, dr \, ds$$

$$= L^2 \int_0^\tau (\tau - r)^{-2\gamma} \int_0^r \mathbb{E} |D_s X_r|^2_{L_2(\Xi, H)} \, ds \, dr.$$

The right-hand side is finite for a.a. τ; indeed, by exchanging the integrals we verify that

$$\int_0^T \left(\int_0^\tau (\tau - r)^{-2\gamma} \int_0^r \mathbb{E} |D_s X_r|^2_{L_2(\Xi, H)} \, ds \, dr \right) d\tau$$

$$\leq \int_0^T r^{-2\gamma} \, dr \int_0^T \int_0^r \mathbb{E} |D_s X_r|^2_{L_2(\Xi, H)} ds \, dr$$

$$= \int_0^T r^{-2\gamma} \, dr |DX|^2_{L^2(\Omega \times [0,T] \times [0,T]; L_2(\Xi, H))} < \infty,$$

since $X \in \mathbb{L}^{1,2}(H)$. Now we recall that the Skorohod and the Itô integral coincide for adapted integrands, so that

$$\int_0^T \mathbb{E}|\delta(D_s u)|^2 \, ds = \int_0^T \mathbb{E}\left|\int_0^T D_s u_r \, dW_r\right|^2 \, ds$$

$$= \mathbb{E}\int_0^T \int_0^T |D_s u_r|^2 \, dr \, ds < \infty.$$

So for a.a. τ we can apply the result mentioned above and since

$$\delta(u) = \int_0^\tau e^{(\tau-r)A} G(r, X_r) \, dW_r,$$

$$\delta(D_s u) = \int_s^\tau \nabla_x\big(e^{(\tau-r)A} G(r, X_r)\big) D_s X_r \, dW_r,$$

formula (3.15) is proved. The estimate

$$\int_0^T \int_0^\tau \mathbb{E}\left|D_s \int_0^\tau e^{(\tau-r)A} G(r, X_r) \, dW_r\right|^2 \, ds \, d\tau$$

$$\leq 2 \int_0^T \int_0^\tau \mathbb{E}|e^{(\tau-s)A} G(s, X_s)|^2_{L_2(\Xi,H)} \, ds \, d\tau$$

$$+ 2 \int_0^T \int_0^\tau \mathbb{E}\int_s^\tau |\nabla_x\big(e^{(\tau-r)A} G(r, X_r)\big) D_s X_r|^2_{L_2(\Xi,L_2(\Xi,H))} \, dr \, ds \, d\tau$$

$$\leq 2L^2 \int_0^T r^{-2\gamma} \, dr \int_0^T \mathbb{E}(1 + |X_r|)^2 \, dr$$

$$+ 2L^2 \int_0^T r^{-2\gamma} \, dr \, |DX|^2_{L^2(\Omega \times [0,T] \times [0,T]; L_2(\Xi,H))} < \infty,$$

is a consequence of the previous passages, and shows that the process

$$\int_0^\tau e^{(\tau-r)A} G(r, X_r) \, dW_r, \qquad \tau \in [0, T],$$

belongs to $\mathbb{L}^{1,2}(H)$. $\square$

For $s \in [0, T)$ and for arbitrary predictable processes X_τ, Q_τ, $\tau \in [s, T]$, with values in H and $L_2(\Xi, H)$ respectively, we define, for $\tau \in [s, T]$,

$$\Gamma_1(X, Q)_{s\tau} = \int_s^\tau e^{(\tau-r)A} \nabla_x F(r, X_r) Q_r \, dr,$$

$$\Gamma_2(X, Q)_{s\tau} = \int_s^\tau \nabla_x\big(e^{(\tau-r)A} G(r, X_r)\big) Q_r \, dW_r.$$

The same notation will be used when Q_τ, $\tau \in [s, T]$, is a process with values in H.

PROOF OF PROPOSITION 3.5. We fix $t \in [0, T)$. Let us consider the sequence X^n defined as follows: $X^0 = 0$,

$$X_\tau^{n+1} = e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-r)A} F(r, X_r^n)\, dr$$
$$+ \int_t^\tau e^{(\tau-r)A} G(r, X_r^n)\, dW_r, \qquad \tau \in [t, T],$$

and $X_\tau^n = x$ for $\tau < t$. It follows from the proof of Proposition 3.2 that X^n converges to the solution X of equation (3.4) in the space $L_\mathcal{P}^p(\Omega; C([0, T]; H))$ hence, in particular, in the space $L^2(\Omega \times [0, T]; H)$. By Lemma 3.6, $X^n \in \mathbb{L}^{1,2}(H)$ and, for a.a. s and τ with $s < \tau$,

(3.16)
$$D_s X_\tau^{n+1} = e^{(\tau-s)A} G(s, X_s^n) + \int_s^\tau e^{(\tau-r)A} \nabla_x F(r, X_r^n) D_s X_r^n\, dr$$
$$+ \int_s^\tau \nabla_x \big(e^{(\tau-r)A} G(r, X_r^n) \big) D_s X_r^n\, dW_r.$$

Setting $I(X^n)_{s\tau} = e^{(\tau-s)A} G(s, X_s^n)$ for $\tau > s$ and $I(X^n)_{s\tau} = 0$ for $\tau < s$, and recalling the operators introduced above, we may write equality (3.16) as

$$DX^{n+1} = I(X^n) + \Gamma_1(X^n, DX^n) + \Gamma_2(X^n, DX^n).$$

We note that $I(X^n)$ is a bounded sequence in $L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$, since

$$\mathbb{E} \int_0^T \int_0^\tau |e^{(\tau-s)A} G(s, X_s^n)|_{L_2(\Xi,H)}^2\, ds\, d\tau$$
$$\leq L^2 \mathbb{E} \int_0^T \int_0^\tau (\tau - s)^{-2\gamma}(1 + |X_s^n|)^2\, ds\, d\tau$$
$$\leq L^2 \int_0^T \tau^{-2\gamma}\, d\tau \int_0^T \mathbb{E}(1 + |X_s^n|)^2\, ds,$$

and X^n is a bounded sequence in $L^2(\Omega \times [0, T]; H)$. Next we show that there exists an equivalent norm $\|\cdot\|$ in $L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$ such that

(3.17)
$$\|\Gamma_1(X^n, DX^n)\| + \|\Gamma_2(X^n, DX^n)\| \leq \alpha \|DX^n\|,$$

for some $\alpha \in [0, 1)$ independent of n. For simplicity we only consider the operator Γ_2. For a process $(Z_{s\tau}) \in L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$ we introduce the norm

$$\|Z\|^2 = \int_0^T \int_0^T \mathbb{E}|Z_{s\tau}|_{L_2(\Xi,H)}^2 e^{-\beta(\tau-s)}\, d\tau\, ds,$$

where $\beta > 0$ will be chosen later. We have

$$\int_s^T \mathbb{E}|\Gamma_2(X^n, DX^n)_{s\tau}|^2_{L_2(\Xi,H)} e^{-\beta(\tau-s)} \, d\tau$$

$$= \int_s^T \int_s^\tau \mathbb{E}\,|\nabla_x(e^{(\tau-r)A}G(r, X_r^n))D_s X_r^n|^2_{L_2(\Xi,L_2(\Xi,H))} \, dr \, e^{-\beta(\tau-s)} \, d\tau$$

$$\leq L^2 \int_s^T \int_s^\tau (\tau-r)^{-2\gamma} \mathbb{E}\,|D_s X_r^n|^2_{L_2(\Xi,H)} \, dr \, e^{-\beta(\tau-s)} \, d\tau$$

$$= L^2 \int_s^T e^{-\beta(r-s)} \mathbb{E}\,|D_s X_r^n|^2_{L_2(\Xi,H)} \int_r^T (\tau-r)^{-2\gamma} \, e^{-\beta(\tau-r)} \, d\tau \, dr$$

$$\leq L^2 \int_s^T e^{-\beta(r-s)} \mathbb{E}\,|D_s X_r^n|^2_{L_2(\Xi,H)} \, dr \left(\sup_{r\in[s,T]} \int_r^T (\tau-r)^{-2\gamma} \, e^{-\beta(\tau-r)} \, d\tau \right).$$

The supremum on the right-hand side can be estimated by $\int_0^T r^{-2\gamma} e^{-\beta r} \, dr$; so we obtain

$$\|\Gamma_2(X^n, DX^n)\|^2 \leq L^2 \int_0^T r^{-2\gamma} \, e^{-\beta r} \, dr \, \|DX^n\|^2.$$

Now to prove (3.17) it suffices to take β sufficiently large.

From (3.17) and from the fact that $I(X^n)$ is bounded in $L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$, it follows easily that the sequence DX^n is also bounded in this space. Since, as mentioned before, X^n converges to X in $L^2(\Omega \times [0, T]; H)$, it follows from the closedness of the operator D that X belongs to $\mathbb{L}^{1,2}(H)$. Point (i) of Proposition 3.5 is now proved.

By Lemma 3.6, we can compute the Malliavin derivative of both sides of (3.4) and we obtain, for a.a. s and τ with $s < \tau$,

$$(3.18) \qquad D_s X_\tau = I(X)_{s\tau} + \Gamma_1(X, DX)_{s\tau} + \Gamma_2(X, DX)_{s\tau}, \qquad \mathbb{P}\text{-a.s.,}$$

where

$$(3.19) \qquad\qquad\qquad I(X)_{s\tau} = e^{(\tau-s)A}G(s, X_s).$$

Let us introduce the space $\mathcal{K}$ of processes $Q_{s\tau}$, $0 \leq s < \tau \leq T$, such that for every $s \in [t, T)$, $\{Q_{s\tau}, \tau \in (s, T]\}$ is a predictable process in $L_2(\Xi, H)$ with continuous paths, and such that

$$(3.20) \qquad \sup_{s\in[0,T]} \mathbb{E}\left(\sup_{\tau\in(s,T]} e^{-\beta p(\tau-s)}(\tau-s)^{p\gamma}|Q_{s\tau}|^p_{L_2(\Xi,H)} \right) < \infty.$$

Here $p \in [2, \infty)$ is fixed and $\beta > 0$ is a parameter, to be chosen later. Let us consider the equation: for every $s \in [0, T)$, $\mathbb{P}$-a.s.,

$$(3.21) \qquad Q_{s\tau} = I(X)_{s\tau} + \Gamma_1(X, Q)_{s\tau} + \Gamma_2(X, Q)_{s\tau}, \qquad \tau \in (s, T].$$

We are going to prove that there exists a unique $Q \in \mathcal{K}$ solution of this equation. Assume this for a moment. Then, subtracting (3.21) from (3.18), we obtain for a.a. s and τ with $s < \tau$

$$D_s X_\tau - Q_{s\tau} = \Gamma_1(X, DX - Q)_{s\tau} + \Gamma_2(X, DX - Q)_{s\tau}, \qquad \mathbb{P}\text{-a.s.}$$

Repeating the passages that led to (3.17) we obtain

$$\|\Gamma_1(X, DX - Q)\| + \|\Gamma_2(X, DX - Q)\| \leq \alpha \|DX - Q\|,$$

for some $\alpha \in [0, 1)$. This proves that Q is a version of DX. Then equality (3.21) coincides with (3.12), and this proves point (ii) of the proposition, except for the last assertion.

Now we prove unique solvability of (3.21) in the space $\mathcal{K}$. It suffices to show that $I(X) \in \mathcal{K}$ and that $\Gamma_1(X, \cdot) + \Gamma_2(X, \cdot)$ is a contraction in $\mathcal{K}$. Since, for $\tau > s$,

$$|e^{(\tau - s)A} G(s, X_s)|_{L_2(\Xi, H)} \leq L(\tau - s)^{-\gamma}(1 + |X_s|),$$

we have

$$\sup_{s \in [0,T]} \mathbb{E} \sup_{\tau \in (s,T]} (\tau - s)^{p\gamma} |e^{(\tau - s)A} G(s, X_s)|^p_{L_2(\Xi, H)} \leq L^p \sup_{s \in [0,T]} \mathbb{E}(1 + |X_s|)^p,$$

which is finite, since $X \in L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$. This shows that $I(X) \in \mathcal{K}$; the contraction property for $\Gamma_1(X, \cdot) + \Gamma_2(X, \cdot)$ requires a longer argument, and it is postponed to Lemma 3.7 below.

The last assertion of point (ii) is clear for $\tau \in [0, t]$, since $X_\tau = x$. For $\tau \in (t, T]$ we take a sequence $\tau_n \uparrow \tau$ such that $X_{\tau_n} \in \mathbb{D}^{1,2}(H)$ and we note that by (3.11) the sequence $\mathbb{E} \int_0^T |D_s X_{\tau_n}|^2 \, ds$ is bounded by a constant independent of n; since $X_{\tau_n} \to X_\tau$ in $L^2(\Omega; H)$, it follows from the closedness of the operator D that $X_\tau \in \mathbb{D}^{1,2}(H)$.

Now we proceed to proving point (iii) of the proposition. Let us fix $v \in \Xi$ and define the space $\mathcal{L}$ of processes $\{Q_{s\tau}, t \leq s \leq \tau \leq T\}$, with values in H, such that $\{Q_{s\tau}, \tau \in [s, T]\}$ is predictable for every $s \in [t, T]$ and the norm

$$\|Q\|^2 = \int_t^T \int_s^T \mathbb{E}|Q_{s\tau}|^2_H e^{-\beta(\tau - s)} \, d\tau \, ds$$

is finite, where $\beta > 0$ is a parameter to be chosen later. Since $I(X)$ [defined in (3.19)] belongs to the space $\mathcal{K}$ introduced above, therefore $I(X)v$ belongs to $\mathcal{L}$ and the equality (3.13) is equivalent to the equality in the space $\mathcal{L}$:

$$(3.22) \qquad Q = I(X)v + \Gamma_1(X, Q) + \Gamma_2(X, Q).$$

It turns out that this equation has a unique solution in $\mathcal{L}$: indeed, $\Gamma_1(X, \cdot) + \Gamma_2(X, \cdot)$ is a contraction in the space $\mathcal{L}$ if β is chosen sufficiently large, as it can be proved by passages almost identical to those leading to (3.17). Finally, DXv belongs to $\mathcal{L}$ since $DX \in L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$, and applying both

sides of (3.12) to v we check that $DXv = I(X)v + \Gamma_1(X, DXv) + \Gamma_2(X, DXv)$. Point (iii) of the proposition is now proved.

To prove point (iv) let us fix $v \in \Xi$ and $p \in [2, \infty)$, let us define $\Delta = \{(s, \tau) : t \leq s \leq \tau \leq T\}$ and introduce the space $\mathcal{H}$ of processes $Q_{s\tau}$, $(s, \tau) \in \Delta$ that are continuous maps $Q \colon \Delta \to L^p(\Omega; H)$ and such that for every s the process $\{Q_{s\tau}, \tau \in [s, T]\}$ is adapted. We are going to show that there exists a unique $Q \in \mathcal{H}$ satisfying the equation: for every $(s, \tau) \in \Delta$,

$$
\begin{aligned}
(3.23) \quad Q_{s\tau} = {} & e^{(\tau-s)A} G(s, X_s)v + \int_s^\tau e^{(\tau-\sigma)A} \nabla_x F(\sigma, X_\sigma) Q_{s\sigma} \, d\sigma \\
& + \int_s^\tau \nabla_x (e^{(\tau-\sigma)A} G(\sigma, X_\sigma)) Q_{s\sigma} \, dW_\sigma, \qquad \mathbb{P}\text{-a.s.}
\end{aligned}
$$

Using the fact that $X \in L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$, and assumption (3.3), it is easy to show that the process $e^{(\tau-s)A} G(s, X_s)v$, $(s, \tau) \in \Delta$, belongs to $\mathcal{H}$. Moreover, we claim that the map that associates to $Q \in \mathcal{H}$ the right-hand side of (3.23) is a linear operator on $\mathcal{H}$ whose operator norm can be shown to be less than 1 provided $\mathcal{H}$ is endowed with the norm

$$
\|Q\|^p = \sup_{(s, \tau) \in \Delta} e^{-\beta p(\tau-s)} \mathbb{E} |Q_{s, \tau}|_H^p,
$$

with $\beta > 0$ sufficiently large. The passages showing this claim are similar to those leading to the estimate (3.17), so we limit ourselves to a sketch.

Consider for instance the process $(s, \tau) \mapsto \int_s^\tau \nabla_x (e^{(\tau-\sigma)A} G(\sigma, X_\sigma)) Q_{s\sigma} \, dW_\sigma$, that we denote $\Gamma_2(X, Q)_{s\tau}$ as before. To prove the contraction property of $\Gamma_2(X, Q)$ we note that by the Burkholder–Davis–Gundy inequalities we have, for some constant c_p,

$$
\begin{aligned}
\mathbb{E} |\Gamma_2(X, Q)_{s\tau}|_H^p &\leq c_p \mathbb{E} \left(\int_s^\tau |\nabla_x (e^{(\tau-r)A} G(\sigma, X_\sigma)) Q_{s\sigma}|_{L_2(\Xi, H)}^2 \, d\sigma \right)^{p/2} \\
&\leq c_p L^p \mathbb{E} \left(\int_s^\tau (\tau-\sigma)^{-2\gamma} |Q_{s\sigma}|_H^2 \, d\sigma \right)^{p/2} \\
&\leq c_p L^p \left(\int_s^\tau (\tau-\sigma)^{-2\gamma} (\mathbb{E} |Q_{s\sigma}|_H^p)^{2/p} \, d\sigma \right)^{p/2} \\
&\leq c_p L^p \|Q\|^p \left(\int_s^\tau (\tau-\sigma)^{-2\gamma} e^{2\beta(\sigma-s)} \, d\sigma \right)^{p/2}.
\end{aligned}
$$

It follows that

$$
e^{-p\beta(\tau-s)} \mathbb{E} |\Gamma_2(X, Q)_{s\tau}|_H^p \leq c_p L^p \|Q\|^p \left(\int_s^\tau (\tau-\sigma)^{-2\gamma} e^{-2\beta(\tau-\sigma)} \, d\sigma \right)^{p/2},
$$

which implies $\|\Gamma_2(X, Q)_{s\tau}\| \leq c_p^{1/2} L \|Q\| (\int_0^T (\tau - \sigma)^{-2\gamma} e^{-2\beta(\tau-\sigma)} d\sigma)^{1/2}$, which shows the required contraction property.

In order to show continuity of the map $\Delta \to L^p(\Omega; H)$, $(s, \tau) \mapsto \Gamma_2(X, Q)_{s\tau}$, it is convenient first to verify continuity of the map

$$(s, \tau) \mapsto \Gamma_2(X, Q)_{s\tau}^\varepsilon := \int_s^{(\tau-\varepsilon)\vee s} \nabla_x (e^{(\tau-\sigma)A} G(\sigma, X_\sigma)) Q_{s\sigma} \, dW_\sigma$$

and then show that $\Gamma_2(X, Q)^\varepsilon \to \Gamma_2(X, Q)$ in $\mathcal{H}$ as $\varepsilon \to 0$. This way the claim is proved, so that in particular the equation (3.23) has a unique solution in $\mathcal{H}$.

Since the solution $Q \in \mathcal{H}$ of equation (3.23) evidently satisfies equation (3.13), it follows from point (iii) that Q is the required version of DXv. Equation (3.14) now follows immediately from (3.23). □

To complete the previous proof, it remains to state and prove the following lemma.

LEMMA 3.7. *For $s \in [0, T)$, let X_τ, $\tau \in [s, T]$, be a predictable process in H and let Q_τ, $\tau \in (s, T]$, be an $L_2(\Xi, H)$-valued continuous adapted process.*

For $p \in [2, \infty)$ sufficiently large and for every $\beta > 0$, the following estimate holds:

$$\mathbb{E}\left(\sup_{\tau \in [s,T]} (\tau - s)^{\gamma p} e^{-\beta p(\tau-s)} \big(|\Gamma_1(X, Q)_{s\tau}|_{L_2(\Xi,H)}^p + |\Gamma_2(X, Q)_{s\tau}|_{L_2(\Xi,H)}^p\big)\right)$$

$$\leq C(\beta)\mathbb{E}\left(\sup_{\tau \in [s,T]} (\tau - s)^{\gamma p} e^{-\beta p(\tau-s)} |Q_\tau|_{L_2(\Xi,H)}^p\right),$$

where $C(\beta)$ depends on β, p, L, γ, T and $M = \sup_{\tau \in [0,T]} |e^{\tau A}|$, and is such that $C(\beta) \to 0$ as $\beta \to +\infty$.

PROOF. For simplicity, we only consider the operator Γ_2. Fixed $s \in [0, T)$ we introduce the space of $L_2(\Xi, H)$-valued continuous adapted processes Q_τ, $\tau \in (s, T]$ such that the norm

$$\|Q\|_s^p = \mathbb{E} \sup_{\tau \in [s,T]} (\tau - s)^{\gamma p} e^{-\beta p(\tau-s)} |Q_\tau|_{L_2(\Xi,H)}^p$$

is finite. We use the factorization method; see [10], Theorem 5.2.5. Let us take $p > 2$ and $\alpha \in (0, 1)$ such that

$$\frac{1}{p} < \alpha < \frac{1}{2} - \gamma \quad \text{and let } c_\alpha^{-1} = \int_r^\tau (\tau - \sigma)^{\alpha-1} (\sigma - r)^{-\alpha} \, d\sigma.$$

Then, by the stochastic Fubini theorem,

$\Gamma_2(X, Q)_{s\tau}$

$$= c_\alpha \int_s^\tau \int_r^\tau (\tau - \sigma)^{\alpha-1} (\sigma - r)^{-\alpha} e^{(\tau-\sigma)A} \nabla_x \big(e^{(\sigma-r)A} G(r, X_r)\big) Q_r \, d\sigma \, dW_r$$

$$= c_\alpha \int_s^\tau (\tau - \sigma)^{\alpha-1} e^{(\tau-\sigma)A} V_\sigma \, d\sigma,$$

where

$$V_\sigma = \int_s^\sigma (\sigma - r)^{-\alpha} \nabla_x \big(e^{(\sigma-r)A} G(r, X_r)\big) Q_r \, dW_r.$$

By the Hölder inequality, setting $M = \sup_{\tau \in [0,T]} |e^{\tau A}|$, $p' = p/(p-1)$,

$$|\Gamma_2(X, Q)_{s\tau}| \leq c_\alpha M \int_s^\tau (\tau - \sigma)^{\alpha-1} |V_\sigma| \, d\sigma$$

$$\leq c_\alpha M \left(\int_s^\tau e^{-p\beta(\sigma-s)} (\sigma - s)^{\gamma p} |V_\sigma|^p \, d\sigma \right)^{1/p}$$

$$\times \left(\int_s^\tau e^{p'\beta(\sigma-s)} (\sigma - s)^{-\gamma p'} (\tau - \sigma)^{(\alpha-1)p'} \, d\sigma \right)^{1/p'}.$$

It follows that

$\|\Gamma_2(X, Q)\|_s^p$

$$\leq c_\alpha^p M^p \int_s^T e^{-p\beta(\sigma-s)} (\sigma - s)^{\gamma p} \mathbb{E} |V_\sigma|^p \, d\sigma \sup_{\tau \in (s,T]} (\tau - s)^{\gamma p} e^{-\beta p(\tau-s)}$$

$$\times \left(\int_s^\tau e^{p'\beta(\sigma-s)} (\sigma - s)^{-\gamma p'} (\tau - \sigma)^{(\alpha-1)p'} \, d\sigma \right)^{p/p'}.$$

Changing σ into $(\sigma - s)/(\tau - s)$, it is easily seen that the supremum on the right-hand side equals

$$\sup_{\tau \in (s,T]} (\tau - s)^{p\alpha-1} e^{-\beta p(\tau-s)} \left(\int_0^1 e^{p'\beta\sigma(\tau-s)} \sigma^{-\gamma p'} (1-\sigma)^{(\alpha-1)p'} \, d\sigma \right)^{p/p'} \leq a(\beta)^p,$$

where we set

$$a(\beta) := \sup_{\lambda \in (0,T]} \lambda^{\alpha-1/p} e^{-\beta\lambda} \left(\int_0^1 e^{p'\beta\sigma\lambda} \sigma^{-\gamma p'} (1-\sigma)^{(\alpha-1)p'} \, d\sigma \right)^{1/p'}.$$

So we arrive at

$$\|\Gamma_2(X, Q)\|_s \leq c_\alpha M a(\beta) \left(\int_s^T e^{-p\beta(\sigma-s)} (\sigma - s)^{\gamma p} \mathbb{E} |V_\sigma|^p \, d\sigma \right)^{1/p}.$$

By the Burkholder–Davis–Gundy inequalities, for some constant c_p depending only on p, we have

$$\mathbb{E}\,|V_\sigma|^p \le c_p \mathbb{E}\bigg(\int_s^\sigma (\sigma - r)^{-2\alpha} \big|\nabla_x\big(e^{(\sigma-r)A} G(r, X_r)\big) Q_r\big|^2_{L_2(\Xi, L_2(\Xi, H))}\, dr\bigg)^{p/2}$$

$$\le L^p c_p \mathbb{E}\bigg(\int_s^\sigma (\sigma - r)^{-2\alpha-2\gamma} |Q_r|^2_{L_2(\Xi, H)}\, dr\bigg)^{p/2}$$

$$\le L^p c_p \|Q\|_s^p \bigg(\int_s^\sigma (\sigma - r)^{-2\alpha-2\gamma} (r - s)^{-2\gamma} e^{2\beta(r-s)}\, dr\bigg)^{p/2}.$$

Changing r into $(r - s)/(\sigma - s)$ and taking into account that $\beta > 0$ and $\alpha + \gamma < 1/2$ we obtain

$$(\sigma - s)^{\gamma p} e^{-p\beta(\sigma-s)} \mathbb{E}\,|V_\sigma|^p \le L^p c_p \|Q\|_s^p (\sigma - s)^{p(-\alpha-\gamma+1/2)}$$

$$\times \bigg(\int_0^1 (1 - r)^{-2\alpha-2\gamma} r^{-2\gamma} e^{-2\beta(1-r)(\sigma-s)}\, dr\bigg)^{p/2}$$

$$\le L^p c_p \|Q\|_s^p T^{p(1/2-\alpha-\gamma)}$$

$$\times \bigg(\int_0^1 (1 - r)^{-2\alpha-2\gamma} r^{-2\gamma}\, dr\bigg)^{p/2}.$$

We conclude that

$$\|\Gamma_2(X, Q)\|_s$$

$$\le c_\alpha M L c_p^{1/p} a(\beta) T^{1/2-\alpha-\gamma+1/p} \bigg(\int_0^1 (1 - r)^{-2\alpha-2\gamma} r^{-2\gamma}\, dr\bigg)^{1/2} \|Q\|_s.$$

This inequality proves the lemma, since the property that $a(\beta) \to 0$ as $\beta \to +\infty$ follows easily from the definition of $a(\beta)$. $\square$

The following result relates the Malliavin derivative of the process X with $\nabla_x X(\tau, t, x)$, the partial Gâteaux derivative with respect to x (compare Proposition 3.3).

PROPOSITION 3.8. *Assume Hypothesis* 3.1. *Then for a.a. s, τ such that $t \le s \le \tau \le T$ we have*

$$(3.24) \quad D_s X(\tau, t, x) = \nabla_x X\big(\tau, s, X(s, t, x)\big) G\big(s, X(s, t, x)\big), \qquad \mathbb{P}\text{-a.s.}$$

Moreover, $D_s X(T, t, x) = \nabla_x X(T, s, X(s, t, x)) G(s, X(s, t, x))$, $\mathbb{P}$-a.s. for a.a. s.

PROOF. Proposition 3.3 states that for every $s \in [0, T]$ and every direction $h \in H$ the directional derivative process $\nabla_x X(\tau, s, x)h$, $\tau \in [s, T]$, solves the

equation: $\mathbb{P}$-a.s.,

$$\nabla_x X(\tau, s, x)h = e^{(\tau-s)A}h + \int_s^\tau e^{(\tau-\sigma)A}\nabla_x F(\sigma, X(\sigma, s, x))\nabla_x X(\sigma, s, x)h\, d\sigma$$

$$+ \int_s^\tau \nabla_x(e^{(\tau-\sigma)A}G(\sigma, X(\sigma, s, x)))$$

$$\times \nabla_x X(\sigma, s, x)h\, dW_\sigma, \qquad \tau \in [s, T].$$

Given $v \in \Xi$ and $t \in [0, s]$, we can replace x by $X(s, t, x)$ and h by $G(s, X(s, t, x))v$ in this equation, since $X(s, t, x)$ is $\mathcal{F}_s$-measurable. Next we note the equality, $\mathbb{P}$-a.s.,

$$X(\sigma, s, X(s, t, x)) = X(\sigma, t, x), \qquad \sigma \in [s, T],$$

which is a consequence of the uniqueness of the solution of (3.4), and we obtain, $\mathbb{P}$-a.s.,

$$\nabla_x X(\tau, s, X(s, t, x))G(s, X(s, t, x))v$$

$$= e^{(\tau-s)A}G(s, X(s, t, x))v$$

$$+ \int_s^\tau e^{(\tau-\sigma)A}\nabla_x F(\sigma, X(\sigma, t, x))\nabla_x X(\sigma, s, X(s, t, x))$$

$$\times G(s, X(s, t, x))v\, d\sigma$$

$$+ \int_s^\tau \nabla_x(e^{(\tau-\sigma)A}G(\sigma, X(\sigma, t, x)))\nabla_x X(\sigma, s, X(s, t, x))$$

$$\times G(s, X(s, t, x))v\, dW_\sigma, \qquad \tau \in [s, T].$$

This shows that the process $\{\nabla_x X(\tau, t, X(s, t, x))G(s, X(s, t, x))v : t \le s \le \tau \le T\}$ is a solution of equation (3.13). Then (3.24) follows from the uniqueness property.

To prove the last assertion, it suffices to take a sequence $\tau_n \uparrow T$ such that (3.24) holds for τ_n and let $n \to \infty$. The conclusion follows from the regularity properties of DX and $\nabla_x X$ stated above, as well as the closedness of the operator D. $\quad\square$

4. The backward equation.

4.1. *Existence, uniqueness and regularity for backward equations in general.* Some of the basic results on backward equations rely on the following well-known representation theorem (see, e.g., [23]). Recall that $(\mathcal{F}_t)$ is the filtration generated by the Wiener process W, augmented in the usual way. We denote by $\mathbb{E}^{\mathcal{F}_\tau}$ the conditional expectation with respect to $\mathcal{F}_\tau$.

PROPOSITION 4.1. *Let K be a Hilbert space and $T > 0$. For arbitrary $\mathcal{F}_T$-measurable $\xi \in L^2(\Omega; K)$ there exists $V \in L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$ such that $\xi = \mathbb{E}\xi + \int_0^T V_\sigma \, dW_\sigma$, $\mathbb{P}$-a.s. Equivalently, for every $\tau \in [0, T]$,*

$$\mathbb{E}^{\mathcal{F}_\tau}\xi = \xi - \int_\tau^T V_\sigma \, dW_\sigma, \qquad \mathbb{P}\text{-a.s.}$$

LEMMA 4.2. *Assume $\eta \in L^2(\Omega; K)$ is $\mathcal{F}_T$-measurable, $f \in L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$. Then there exists a unique pair of processes Y_τ, Z_τ, $\tau \in [0, T]$, such that*:

(i) *$Y \in L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$, $Z \in L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$;*
(ii) *for a.a. $\tau \in [0, T]$,*

(4.1)
$$Y_\tau + \int_\tau^T Z_\sigma \, dW_\sigma = -\int_\tau^T f_\sigma \, d\sigma + \eta.$$

Moreover, Y has a continuous version and for every $\beta \neq 0$,

(4.2)
$$\mathbb{E}\int_0^T e^{2\beta\sigma}|Z_\sigma|^2 d\sigma \leq \frac{4}{\beta}\,\mathbb{E}\int_0^T e^{2\beta\sigma}|f_\sigma|^2 d\sigma + 8\,e^{2\beta T}\mathbb{E}\,|\eta|^2,$$

$$\mathbb{E}\sup_{\tau\in[0,T]} e^{2\beta\tau}|Y_\tau|^2 \leq \frac{4}{\beta}\,\mathbb{E}\int_0^T e^{2\beta\sigma}|f_\sigma|^2 d\sigma + 8\,e^{2\beta T}\mathbb{E}\,|\eta|^2.$$

In particular, $Y \in C_{\mathcal{P}}([0, T]; L^2(\Omega; K))$.

If, in addition, there exists $p \in [2, \infty)$ such that

$$\mathbb{E}\left(\int_0^T |f_\sigma|^2 \, d\sigma\right)^{p/2} < \infty, \qquad \mathbb{E}\,|\eta|^p < \infty,$$

then for every δ such that $0 \leq T - \delta < T$ we have

(4.3)
$$\mathbb{E}\sup_{\tau\in[T-\delta,T]}|Y_\tau|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_\sigma|^2 \, d\sigma\right)^{p/2}$$
$$\leq c_p\delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |f_\sigma|^2 \, d\sigma\right)^{p/2} + c_p\mathbb{E}\,|\eta|^p,$$

where c_p is a positive constant, depending only on p.

PROOF. We modify the argument in [23].

Uniqueness. Assume that (4.1) holds. Then, taking conditional expectation with respect to $\mathcal{F}_\tau$ we obtain, for a.e. τ,

(4.4)
$$Y_\tau = \mathbb{E}^{\mathcal{F}_\tau}\eta - \int_\tau^T \mathbb{E}^{\mathcal{F}_\tau} f_\sigma \, d\sigma.$$

If $\eta = 0$, $f = 0$ this equality implies that $Y = 0$; from (4.1) it follows that $\int_\tau^T Z_\sigma \, dW_\sigma = 0$, which implies $Z = 0$ as well.

Existence.   Define $\xi = \eta - \int_0^T f_\sigma \, d\sigma$. Since $\xi \in L^2(\Omega; K)$ is $\mathcal{F}_T$-measurable, by Proposition 4.1 there exists $Z \in L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$ such that

$$\mathbb{E}^{\mathcal{F}_\tau} \xi = \xi - \int_\tau^T Z_\sigma \, dW_\sigma,$$

for every $\tau \in [0, T]$. Now it suffices to define $Y_\tau = \mathbb{E}^{\mathcal{F}_\tau} \xi$ and equation (4.1) is satisfied. The existence of a continuous version is immediate, since (4.1) implies

$$Y_\tau - Y_0 = \int_0^\tau Z_\sigma \, dW_\sigma + \int_0^\tau f_\sigma \, d\sigma.$$

Estimates (4.2).   Since $\eta \in L^2(\Omega; K)$ is $\mathcal{F}_T$-measurable, by Proposition 4.1 there exists $L \in L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$ such that

$$(4.5) \qquad\qquad \mathbb{E}^{\mathcal{F}_\tau} \eta = \eta - \int_\tau^T L_\theta \, dW_\theta,$$

for every $\tau \in [0, T]$. Similarly, for a.a. σ there exists a predictable process $\{K(\theta, \sigma), \theta \in [0, \sigma]\}$ in $L^2_{\mathcal{P}}(\Omega \times [0, \sigma]; L_2(\Xi, K))$ such that

$$(4.6) \qquad\qquad \mathbb{E}^{\mathcal{F}_\tau} f_\sigma = f_\sigma - \int_\tau^\sigma K(\theta, \sigma) \, dW_\theta,$$

for $\tau \in [0, \sigma]$. We set $K(\theta, \sigma) = 0$ for $\theta \in (\sigma, T]$ and we can verify that the map $K \colon \Omega \times [0, T] \times [0, T]$ can be taken to $\mathcal{P} \times \mathcal{B}([0, T])$-measurable, where $\mathcal{P}$ is the predictable σ-algebra on $\Omega \times [0, T]$ and $\mathcal{B}([0, T])$ denotes the Borel subsets of $[0, T]$; the existence of such a version of K can be proved by approximating f by simple processes and by a monotone class argument (or one can argue as in [23], proof of Lemma 2.1). Substituting into (4.4) and applying the stochastic Fubini theorem gives

$$Y_\tau = \eta - \int_\tau^T L_\theta \, dW_\theta - \int_\tau^T \left(f_\sigma - \int_\tau^\sigma K(\theta, \sigma) \, dW_\theta \right) d\sigma$$

$$= \eta - \int_\tau^T f_\sigma \, d\sigma - \int_\tau^T L_\theta \, dW_\theta + \int_\tau^T \left(\int_\theta^T K(\theta, \sigma) \, d\sigma \right) dW_\theta.$$

Comparing with the backward equation, we conclude by uniqueness that for a.a. θ,

$$Z_\theta = L_\theta - \int_\theta^T K(\theta, \sigma) \, d\sigma.$$

Now let $\beta \neq 0$. From (4.5) we deduce

$$\mathbb{E} \int_0^T e^{2\beta\theta} |L_\theta|^2 \, d\theta \leq e^{2\beta T} \mathbb{E} \left| \int_0^T L_\theta \, d\theta \right|^2 = e^{2\beta T} \mathbb{E} |\eta - \mathbb{E}^{\mathcal{F}_0} \eta|^2$$

$$\leq 2e^{2\beta T} \mathbb{E} |\eta|^2 + 2e^{2\beta T} \mathbb{E} |\mathbb{E}^{\mathcal{F}_0} \eta|^2 \leq 4e^{2\beta T} \mathbb{E} |\eta|^2.$$

Next note that

$$\left| \int_\theta^T K(\theta, \sigma) \, d\sigma \right|^2 \le \int_\theta^T e^{-2\beta\sigma} \, d\sigma \int_\theta^T e^{2\beta\sigma} |K(\theta, \sigma)|^2 \, d\sigma$$

$$\le \frac{e^{-2\beta\theta}}{2\beta} \int_\theta^T e^{2\beta\sigma} |K(\theta, \sigma)|^2 \, d\sigma,$$

so that

$$\mathbb{E} \int_0^T e^{-2\beta\theta} \left| \int_\theta^T K(\theta, \sigma) \, d\sigma \right|^2 d\theta \le \frac{1}{2\beta} \mathbb{E} \int_0^T \int_\theta^T e^{2\beta\sigma} |K(\theta, \sigma)|^2 \, d\sigma \, d\theta$$

$$= \frac{1}{2\beta} \int_0^T e^{2\beta\sigma} \mathbb{E} \int_0^\sigma |K(\theta, \sigma)|^2 \, d\theta \, d\sigma.$$

Since (4.6) yields

$$\mathbb{E} \int_0^\sigma |K(\theta, \sigma)|^2 \, d\theta = \mathbb{E} \left| \int_0^\sigma K(\theta, \sigma) \, dW_\theta \right|^2 = \mathbb{E} |f_\sigma - \mathbb{E}^{\mathcal{F}_0} f_\sigma|^2$$

$$\le 2\mathbb{E} |f_\sigma|^2 + 2\mathbb{E} |\mathbb{E}^{\mathcal{F}_\tau} f_\sigma|^2 \le 4\mathbb{E} |f_\sigma|^2,$$

the proof of the first inequality in (4.2) is complete. Now we prove the second one, estimating separately the two terms on the right-hand side of (4.4). By the Doob inequality for martingales,

$$\mathbb{E} \sup_{\tau \in [0,T]} e^{2\beta\tau} |\mathbb{E}^{\mathcal{F}_\tau} \eta|^2 \le e^{2\beta T} 4 \, \mathbb{E} |\eta|^2.$$

Next, since

$$\left(\int_\tau^T |f_\sigma| \, d\sigma \right)^2 \le \int_\tau^T e^{-2\beta\sigma} \, d\sigma \int_\tau^T e^{2\beta\sigma} |f_\sigma|^2 \, d\sigma \le \frac{e^{-2\beta\tau}}{2\beta} \int_\tau^T e^{2\beta\sigma} |f_\sigma|^2 \, d\sigma,$$

we obtain

$$e^{\beta\tau} \left| \int_\tau^T \mathbb{E}^{\mathcal{F}_\tau} f_\sigma \, d\sigma \right| \le \mathbb{E}^{\mathcal{F}_\tau} \left(e^{\beta\tau} \int_\tau^T |f_\sigma| \, d\sigma \right)$$

$$\le \frac{1}{\sqrt{2\beta}} \mathbb{E}^{\mathcal{F}_\tau} \left(\int_\tau^T e^{2\beta\sigma} |f_\sigma|^2 \, d\sigma \right)^{1/2}$$

and by the Doob inequality,

$$\mathbb{E} \sup_{\tau \in [0,T]} e^{2\beta\tau} \left| \int_\tau^T \mathbb{E}^{\mathcal{F}_\tau} f_\sigma \, d\sigma \right|^2 \le \frac{4}{2\beta} \mathbb{E} \int_0^T e^{2\beta\sigma} |f_\sigma|^2 \, d\sigma.$$

Estimates (4.3).   Since, for $\tau \in [T - \delta, T]$

$$\int_\tau^T |f_\sigma| \, d\sigma \leq \left(\int_\tau^T |f_\sigma|^2 \, d\sigma\right)^{1/2} (T - \tau)^{1/2} \leq \left(\int_\tau^T |f_\sigma|^2 \, d\sigma\right)^{1/2} \delta^{1/2},$$

it follows from (4.4) that

$$\mathbb{E} \sup_{\tau \in [T-\delta,T]} |Y_\tau|^p \leq c_p \mathbb{E} \sup_{\tau \in [T-\delta,T]} |\mathbb{E}^{\mathcal{F}_\tau} \eta|^p$$

$$+ c_p \delta^{p/2} \mathbb{E} \sup_{\tau \in [T-\delta,T]} \left| \mathbb{E}^{\mathcal{F}_\tau} \left(\int_\tau^T |f_\sigma|^2 \, d\sigma\right)^{1/2} \right|^p$$

$$\leq c_p \mathbb{E} |\eta|^p + c_p \delta^{p/2} \mathbb{E} \left(\int_{T-\delta}^T |f_\sigma|^2 \, d\sigma\right)^{p/2},$$

which proves the desired inequality on the process Y. To obtain a similar estimate on Z we first set $Z_\theta^1 = -\int_\theta^T K(\theta, \sigma) \, d\sigma$, so that $Z_\theta = L_\theta + Z_\theta^1$. From (4.5) it follows that $\mathbb{E}^{\mathcal{F}_\tau} \eta - \mathbb{E}^{\mathcal{F}_{T-\delta}} \eta = \int_{T-\delta}^\tau L_\theta \, dW_\theta$, so by the Burkholder–Davis–Gundy and the Doob inequalities,

$$\mathbb{E} \left(\int_{T-\delta}^T |L_\theta|^2 d\theta\right)^{p/2} \leq c_p \mathbb{E} \sup_{\tau \in [T-\delta,T]} \left|\int_{T-\delta}^\tau L_\theta \, dW_\theta\right|^p$$

$$= c_p \mathbb{E} \sup_{\tau \in [T-\delta,T]} |\mathbb{E}^{\mathcal{F}_\tau} \eta - \mathbb{E}^{\mathcal{F}_{T-\delta}} \eta|^p \leq c_p \mathbb{E} |\eta|^p.$$

In order to prove a similar estimate for Z^1 we first note that, setting $Y_\tau^1 = -\int_\tau^T \mathbb{E}^{\mathcal{F}_\tau} f_\sigma \, d\sigma$, the pair (Y^1, Z^1) is the solution corresponding to $\eta = 0$. Therefore

$$Y_\tau^1 - Y_{T-\delta}^1 = \int_{T-\delta}^\tau Z_\sigma^1 \, dW_\sigma + \int_{T-\delta}^\tau f_\sigma \, d\sigma.$$

So we obtain

$$\mathbb{E} \left(\int_{T-\delta}^T |Z_\sigma^1|^2 d\sigma\right)^{p/2} \leq c_p \mathbb{E} \sup_{\tau \in [T-\delta,T]} \left|\int_{T-\delta}^\tau Z_\sigma^1 \, dW_\sigma\right|^p$$

$$\leq c_p \mathbb{E} \sup_{\tau \in [T-\delta,T]} |Y_\tau^1|^p + c_p \mathbb{E} \left(\int_{T-\delta}^T |f_\sigma| \, d\sigma\right)^p.$$

For Y^1 we can use the estimate proved above with $\eta = 0$:

$$\mathbb{E} \sup_{\tau \in [T-\delta,T]} |\mathbb{E}^{\mathcal{F}_\tau} Y_\tau^1|^p \leq c_p \delta^{p/2} \mathbb{E} \left(\int_{T-\delta}^T |f_\sigma|^2 \, d\sigma\right)^{p/2}.$$

Finally, the required estimate follows from

$$\int_{T-\delta}^T |f_\sigma| \, d\sigma \leq \left(\int_{T-\delta}^T |f_\sigma|^2 \, d\sigma\right)^{1/2} \delta^{1/2}. \qquad \square$$

Now we are concerned with the equation

$$(4.7) \qquad Y_\tau + \int_\tau^T Z_\sigma \, dW_\sigma = -\int_\tau^T f(\sigma, Y_\sigma, Z_\sigma) \, d\sigma + \eta.$$

In the following proposition, K is a Hilbert space, the mapping $f \colon \Omega \times [0, T] \times K \times L_2(\Xi, K) \to K$ is assumed to be measurable with respect to $\mathcal{P} \times \mathcal{B}([0, T] \times K \times L_2(\Xi, K))$ and $\mathcal{B}(K)$ respectively (we recall that by $\mathcal{P}$ we denote the predictable σ-algebra on $\Omega \times [0, T]$ and by $\mathcal{B}(\Lambda)$ the Borel σ-algebra of any topological space Λ). $\eta \colon \Omega \to K$ is assumed to be $\mathcal{F}_T$-measurable,

PROPOSITION 4.3. *Assume that*:

(i) *there exists $L > 0$ such that*

$$|f(\sigma, y_1, z_1) - f(\sigma, y_2, z_2)| \le L(|y_1 - y_2| + |z_1 - z_2|),$$

$\mathbb{P}$-*a.s. for every* $\sigma \in [0, T]$, $y_1, y_2 \in K$, $z_1, z_2 \in L_2(\Xi, K)$;
(ii) $\mathbb{E} \int_0^T |f(\sigma, 0, 0)|^2 \, d\sigma < \infty$, $\mathbb{E} |\eta|^2 < \infty$.

Then there exists a unique pair of processes $Y_\tau, Z_\tau, \tau \in [0, T]$, *such that* $Y \in C_{\mathcal{P}}([0, T]; L^2(\Omega; K))$, $Z \in L_{\mathcal{P}}^2(\Omega \times [0, T]; L_2(\Xi, K))$ *and* (4.7) *holds for* $\tau \in [0, T]$. *Moreover, Y has a continuous version and* $\mathbb{E} \sup_{\tau \in [0, T]} |Y_\tau|^2 < \infty$.

If, in addition, there exists $p \in [2, \infty)$ such that

$$(4.8) \qquad \mathbb{E} \left(\int_0^T |f(\sigma, 0, 0)|^2 \, d\sigma \right)^{p/2} < \infty, \qquad \mathbb{E} |\eta|^p < \infty,$$

then we have $Y \in L_{\mathcal{P}}^p(\Omega; C([0, T]; K))$, $Z \in L_{\mathcal{P}}^p(\Omega; L^2([0, T]; L_2(\Xi, K)))$ *and*

$$(4.9) \qquad \begin{aligned} &\mathbb{E} \sup_{\tau \in [0, T]} |Y_\tau|^p + \mathbb{E} \left(\int_0^T |Z_\sigma|^2 \, d\sigma \right)^{p/2} \\ &\qquad \le c \, \mathbb{E} \left(\int_0^T |f(\sigma, 0, 0)|^2 \, d\sigma \right)^{p/2} + c \, \mathbb{E} |\eta|^p, \end{aligned}$$

for some constant $c > 0$ depending only on p, L, T.

Finally assume that, for all λ in a metric space Λ, a function f_λ is given satisfying (4.8) *and verifying assumption* (i) *with L independent on λ. Also assume that, as $\lambda \to \lambda_0$,*

$$(4.10) \qquad \mathbb{E} \left(\int_0^T |f_\lambda(\sigma, Y_\sigma, Z_\sigma) - f_{\lambda_0}(\sigma, Y_\sigma, Z_\sigma)|^2 \, d\sigma \right)^{p/2} \to 0$$

for all $Y \in L_{\mathcal{P}}^p(\Omega; C([0, T]; K))$, $Z \in L_{\mathcal{P}}^p(\Omega; L^2([0, T]; L_2(\Xi, K)))$.

If we denote by $(Y(\lambda, \eta), Z(\lambda, \eta))$ the solution of (4.7) *corresponding to $f = f_\lambda$ and to the final data $\eta \in L^p(\Omega, \mathbb{R})$ then the map $(\lambda, \eta) \to (Y(\lambda, \eta), Z(\lambda, \eta))$ is continuous from $\Lambda \times L^p(\Omega; \mathbb{R})$ to $L_{\mathcal{P}}^p(\Omega; C([0, T]; K)) \times L_{\mathcal{P}}^p(\Omega; L^2([0, T]; L_2(\Xi, K)))$.*

PROOF.   We denote $\mathcal{K} = C_{\mathcal{P}}([0, T]; L^2(\Omega; K)) \times L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$ and we define a mapping $\Gamma: \mathcal{K} \to \mathcal{K}$ by setting $(Y, Z) = \Gamma(U, V)$ if (Y, Z) is the pair satisfying

$$(4.11) \qquad Y_\tau + \int_\tau^T Z_\sigma \, dW_\sigma = -\int_\tau^T f(\sigma, U_\sigma, V_\sigma) \, d\sigma + \eta,$$

compare Lemma 4.2. The estimates (4.2) show that Γ is well defined, and it is a contraction if $\mathcal{K}$ is endowed with the norm

$$|(Y, Z)|^2_{\mathcal{K}} = \mathbb{E} \int_0^T e^{2\beta\sigma}(|Y_\sigma|^2 + |Z_\sigma|^2) d\sigma,$$

provided β is sufficiently large. For simplicity we only verify the contraction property: if $(U^1, V^1) \in \mathcal{K}$, $(Y^1, Z^1) = \Gamma(U^1, V^1)$ and let $\overline{Y} = Y - Y^1$, $\overline{Z} = Z - Z^1$, $\overline{U} = U - U^1$, $\overline{V} = V - V^1$, $\overline{f}_\sigma = \psi(\sigma, U_\sigma, V_\sigma) - \psi(\sigma, U^1_\sigma, V^1_\sigma)$, have

$$(4.12) \qquad \overline{Y}_\tau + \int_\tau^T \overline{Z}_\sigma \, dW_\sigma = -\int_\tau^T \overline{f}_\sigma \, dW_\sigma,$$

so that by (4.2),

$$\begin{aligned}
|(\overline{Y}, \overline{Z})|^2_{\mathcal{K}} &\leq T \, \mathbb{E} \sup_{\tau \in [0,T]} e^{2\beta\tau} |\overline{Y}_\tau|^2 + \mathbb{E} \int_0^T e^{2\beta\sigma} |\overline{Z}_\sigma|^2 \, d\sigma \\
&\leq \frac{8(1+T)}{\beta} \mathbb{E} \int_0^T e^{2\beta\sigma} |\overline{f}_\sigma|^2 \, d\sigma \\
&\leq \frac{8(1+T)L^2}{\beta} \mathbb{E} \int_0^T e^{2\beta\sigma} (|\overline{U}_\sigma| + |\overline{V}_\sigma|)^2 \, d\sigma \\
&\leq \frac{16(1+T)L^2}{\beta} |(\overline{U}, \overline{V})|^2_{\mathcal{K}}.
\end{aligned}$$

Now we prove the estimate (4.9). We denote $\mathcal{K}_{p,\delta} = L^p(\Omega; C([T-\delta, T]; \mathbb{R})) \times L^p(\Omega; L^2([T-\delta, T]; L_2(\Xi, \mathbb{R})))$ and define $\Gamma: \mathcal{K}_{p,\delta} \to \mathcal{K}_{p,\delta}$ setting $(Y, Z) = \Gamma(U, V)$ if (Y, Z) is the pair satisfying equation (4.11) for $\tau \in [T - \delta, T]$. It is easily verified that Γ is well defined and it is a contraction in $\mathcal{K}_{p,\delta}$, provided $\delta > 0$ is chosen sufficiently small; indeed, arguing as before, we deduce from (4.12) and from (4.3) the inequalities

$$\begin{aligned}
|(\overline{Y}, \overline{Z})|^p_{\mathcal{K}} &= \mathbb{E} \sup_{\tau \in [T-\delta, T]} |\overline{Y}_\tau|^p + \mathbb{E} \left(\int_{T-\delta}^T |\overline{Z}_\sigma|^2 \, d\sigma \right)^{p/2} \\
&\leq c_p \delta^{p/2} L^p \, \mathbb{E} \left(\int_{T-\delta}^T (|\overline{U}_\sigma| + |\overline{V}_\sigma|)^2 \, d\sigma \right)^{p/2} \\
&\leq c_p 2^{p/2} \delta^p L^p \delta \, \mathbb{E} \sup_{\tau \in [T-\delta, T]} |\overline{U}_\tau|^p + c_p (2\delta)^{p/2} L^p \mathbb{E} \left(\int_{T-\delta}^T |\overline{V}_\sigma|^2 d\sigma \right)^{p/2} \\
&\leq c_p (2\delta)^{p/2} L^p (1 + \delta^{p/2}) |(\overline{U}, \overline{V})|^p_{\mathcal{K}},
\end{aligned}$$

and the contraction property holds provided $c_p(2\delta)^{p/2}L^p(1 + \delta^{p/2}) < 1$. Repeating this argument on intervals $[T - \delta, T - 2\delta]$, $[T - 2\delta, T - 3\delta]$, etc., shows that $Y \in L^p(\Omega; C([0, T]; \mathbb{R}))$ and $Z \in L^p(\Omega; L^2([0, T]; L_2(\Xi, \mathbb{R})))$.

Next note that it follows from our assumptions that

$$|f(\sigma, x, y)| \leq |f(\sigma, 0, 0)| + L(|x| + |y|).$$

Applying estimate (4.3) to equation (4.7) we obtain

$$\mathbb{E} \sup_{\tau \in [T-\delta, T]} |Y_\tau|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_\sigma|^2 \, d\sigma\right)^{p/2}$$

$$\leq c_p \delta^{p/2} \mathbb{E}\left(\int_{T-\delta}^T |f(\sigma, Y_\sigma, Z_\sigma)|^2 \, d\sigma\right)^{p/2} + c_p \mathbb{E} |\eta|^p$$

$$\leq c_p \mathbb{E} |\eta|^p + c_p 3^{p-1} \delta^{p/2} \mathbb{E}\left(\int_{T-\delta}^T |f(\sigma, 0, 0)|^2 \, d\sigma\right)^{p/2}$$

$$+ c_p 3^{p-1} L^p \delta^{p/2} \mathbb{E}\left(\int_{T-\delta}^T |Y_\sigma|^2 \, d\sigma\right)^{p/2}$$

$$+ c_p 3^{p-1} L^p \delta^{p/2} \mathbb{E}\left(\int_{T-\delta}^T |Z_\sigma|^2 \, d\sigma\right)^{p/2}$$

$$\leq c_p \mathbb{E} |\eta|^p + c_p 3^{p-1} \delta^{p/2} \mathbb{E}\left(\int_{T-\delta}^T |f(\sigma, 0, 0)|^2 \, d\sigma\right)^{p/2}$$

$$+ c_p 3^{p-1} L^p \delta^p \mathbb{E} \sup_{\tau \in [T-\delta, T]} |Y_\tau|^p$$

$$+ c_p 3^{p-1} L^p \delta^{p/2} \mathbb{E}\left(\int_{T-\delta}^T |Z_\sigma|^2 \, d\sigma\right)^{p/2}.$$

Choosing $\delta > 0$ so small that $\alpha := c_p 3^{p-1} L^p(\delta^p + \delta^{p/2}) < 1$ we obtain

$$\mathbb{E} \sup_{\tau \in [T-\delta, T]} |Y_\tau|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_\sigma|^2 \, d\sigma\right)^{p/2}$$

$$(4.13) \qquad \leq c_p \mathbb{E} |\eta|^p + c_p 3^{p-1} \delta^{p/2} \mathbb{E}\left(\int_{T-\delta}^T |f(\sigma, 0, 0)|^2 \, d\sigma\right)^{p/2}$$

$$+ \alpha\left[\mathbb{E} \sup_{\tau \in [T-\delta, T]} |Y_\tau|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_\sigma|^2 \, d\sigma\right)^{p/2}\right],$$

and it follows that

$$\mathbb{E} \sup_{\tau \in [T-\delta, T]} |Y_\tau|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_\sigma|^2 \, d\sigma\right)^{p/2}$$

$$\leq c \mathbb{E} |\eta|^p + c \mathbb{E}\left(\int_{T-\delta}^T |f(\sigma, 0, 0)|^2 \, d\sigma\right)^{p/2},$$

with c depending only on p and L. Next we note that for $\tau \leq T - \delta$,

$$Y_\tau + \int_\tau^{T-\delta} Z_\sigma \, dW_\sigma = - \int_\tau^{T-\delta} f(\sigma, Y_\sigma, Z_\sigma) \, d\sigma + Y_{T-\delta},$$

and proceeding as before we obtain

$$\mathbb{E} \sup_{\tau \in [T-2\delta, T-\delta]} |Y_\tau|^p + \mathbb{E} \left(\int_{T-2\delta}^{T-\delta} |Z_\sigma|^2 \, d\sigma \right)^{p/2}$$

$$\leq c\mathbb{E} \, |Y_{T-\delta}|^p + c\mathbb{E} \left(\int_{T-2\delta}^{T-\delta} |f(\sigma, 0, 0)|^2 \, d\sigma \right)^{p/2},$$

with the same choice of δ and the same value of c. After a finite number of steps we arrive at (4.9).

Finally the proof of the last assertion can be done in a straightforward way repeating the above argument. □

REMARK 4.4.   The mapping Γ defined in the previous proof was shown to be a contraction in the space $\mathcal{K} = C_{\mathcal{P}}([0, T]; L^2(\Omega; K)) \times L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$. In a similar way, the estimates (4.2) allow to show that Γ is well defined and it is a contraction in the space $L^p_{\mathcal{P}}(\Omega; C([0, T]; K)) \times L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$ as well as in the space $L^2_{\mathcal{P}}(\Omega \times [0, T]; K) \times L^2_{\mathcal{P}}(\Omega \times [0, T]; L_2(\Xi, K))$. In particular, uniqueness holds for equation (4.7) in the latter space, too.

The following lemma is needed in the proof of Proposition 5.4.

LEMMA 4.5.   *Under the assumptions of Proposition* 4.3 [*in particular*, (4.8)], *the sequence defined setting* $Y^0 = 0$, $Z^0 = 0$ *and* (Y^{n+1}, Z^{n+1}) *to be the pair such that*

$$Y_\tau^{n+1} + \int_\tau^T Z_\sigma^{n+1} \, dW_\sigma = - \int_\tau^T f(\sigma, Y_\sigma^n, Z_\sigma^n) \, d\sigma + \eta,$$

satisfies

$$\sup_n \mathbb{E} \sup_{\tau \in [0,T]} |Y_\tau^n|^p + \sup_n \mathbb{E} \left(\int_0^T |Z_\sigma^n|^2 \, d\sigma \right)^{p/2} < \infty.$$

PROOF.   Let us set $b_n(s, t) = \mathbb{E} \sup_{\tau \in [s,t]} |Y_\tau|^p + \mathbb{E}(\int_s^t |Z_\sigma|^2 \, d\sigma)^{p/2}$. Arguing as in the previous proof, instead of (4.13) we arrive at the inequality

$$b_{n+1}(T - \delta, T) \leq c_p \mathbb{E} \, |\eta|^p + c_p 3^{p-1} \delta^{p/2} \mathbb{E} \left(\int_{T-\delta}^T |f(\sigma, 0, 0)|^2 \, d\sigma \right)^{p/2}$$

$$+ \alpha b_n(T - \delta, T)$$

where $\alpha < 1$ and $\delta > 0$ depend only on p and L. It follows that $\sup_n b_n(T - \delta, T) < \infty$; in particular, $\sup_n \mathbb{E}\,|Y_{T-\delta}^n|^p < \infty$. In a similar way,

$$
\begin{aligned}
b_{n+1}(T - 2\delta, T - \delta) \leq{}& c_p \mathbb{E}\,|Y_{T-\delta}^n|^p \\
&+ c_p 3^{p-1} \delta^{p/2} \mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |f(\sigma, 0, 0)|^2 \, d\sigma\right)^{p/2} \\
&+ \alpha b_n(T - 2\delta, T - \delta)
\end{aligned}
$$

and so $\sup_n b_n(T - 2\delta, T - \delta) < \infty$. The required result follows by iteration. $\square$

4.2. *Backward equations depending on parameters: regular dependence.* Now we are dealing with the backward equation

$$(4.14) \qquad Y_\tau + \int_\tau^T Z_\sigma \, dW_\sigma = -\int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) \, d\sigma + \eta,$$

on the time interval $[0, T]$, where η is a given $\mathcal{F}_T$-measurable K-valued random variable and X_τ, $\tau \in [0, T]$, is a given predictable process. The mapping $\psi \colon [0, T] \times H \times K \times L_2(\Xi, K) \to K$ is assumed to be Borel measurable. The solution we are looking for is a pair of predictable processes Y_τ, Z_τ, $\tau \in [0, T]$, with values in K and $L_2(\Xi, K)$ respectively.

We fix the following assumptions on ψ.

HYPOTHESIS 4.6. (i) *There exists $L > 0$ such that*

$$|\psi(\sigma, x, y_1, z_1) - \psi(\sigma, x, y_2, z_2)| \leq L(|y_1 - y_2| + |z_1 - z_2|),$$

for every $\sigma \in [0, T]$, $x \in H$, $y_1, y_2 \in K$, $z_1, z_2 \in L_2(\Xi, K)$.
  (ii) *For every $\sigma \in [0, T]$, $\psi(\sigma, \cdot, \cdot, \cdot) \in \mathcal{G}^{1,1,1}(H \times K \times L_2(\Xi, K), K)$.*
  (iii) *There exist $L > 0$ and $m \geq 0$ such that*

$$|\nabla_x \psi(\sigma, x, y, z)h| \leq L|h|(1 + |z|)(1 + |x| + |y|)^m,$$

for every $\sigma \in [0, T]$, $x, h \in H$, $y \in K$, $z \in L_2(\Xi, K)$.
  (iv) *There exists $L > 0$ such that $|\psi(\sigma, 0, 0, 0)| \leq L$ for every $\sigma \in [0, T]$.*

Conditions (i) and (ii) imply that the Gâteaux derivatives of ψ with respect to y and z are uniformly bounded: for every point (x, y, z) and every directions $k \in K$, $v \in L_2(\Xi, K)$,

$$|\nabla_y \psi(\sigma, x, y, z)k| \leq L\,|k|, \qquad |\nabla_z \psi(\sigma, x, y, z)v| \leq L\,|v|.$$

Moreover, conditions (i)–(iv) imply that

$$(4.15) \qquad\qquad |\psi(\sigma, x, y, z)| \leq L(1 + |x|^{m+1} + |z| + |y|).$$

Finally, conditions (i), (ii) and (iii) imply

$$(4.16) \quad \begin{aligned} &|\psi(\sigma, x_1, y, z) - \psi(\sigma, x_2, y, z)| \\ &\leq L(1 + |z|)(1 + |x_1|^m + |x_2|^m + |y|^m)|x_2 - x_1|. \end{aligned}$$

To start we need the following general lemma that generalizes (with identical proof) the classical result on continuity of evaluation operators; see, for example, [1].

LEMMA 4.7. *Let K_1, K_2 and K_3 be Banach spaces and $\ell: [0, T] \times K_1 \times K_2 \to K_3$ be a measurable map such that, for all $t \in [0, T]$, $\ell(t, \cdot): K_1 \times K_2 \to K_3$ is continuous.*

(i) *Suppose that for some $c > 0$ and $\mu \geq 1$,*

$$|\ell(t, v_1, v_2)|_{K_3} \leq c(1 + |v_1|_{K_1}^\mu)(1 + |v_2|_{K_2}), \qquad t \in [0, T], \ v_1 \in K_1, \ v_2 \in K_2.$$

For all $U \in L_{\mathcal{P}}^{r_1}(\Omega; C([0, T]; K_1))$, $V \in L_{\mathcal{P}}^{r_2}(\Omega; L^2([0, T]; K_2))$ with $r_1, r_2 \geq 1$, let us define in the natural way the evaluation operator $\ell(U, V)(t, \omega) = \ell(U(t, \omega), V(t, \omega))$.

If $\mu/r_1 + 1/r_2 = 1/r_3$ and $r_1 \geq \mu$ then the evaluation operator is continuous from $L_{\mathcal{P}}^{r_1}(\Omega; C([0, T]; K_1)) \times L_{\mathcal{P}}^{r_2}(\Omega; L^2([0, T]; K_2))$ to $L_{\mathcal{P}}^{r_3}(\Omega; L^2([0, T]; K_3))$.

(ii) *Similarly, if*

$$|\ell(t, v_1, v_2)|_{K_3} \leq c(1 + |v_1|_{K_1}^\mu + |v_2|_{K_2}), \qquad t \in [0, T], \ v_1 \in K_1, \ v_2 \in K_2,$$

and $r_2 = \mu r_1$ then the evaluation operator is continuous from $L_{\mathcal{P}}^{r_1}(\Omega; L^2([0, T]; K_2)) \times L_{\mathcal{P}}^{r_2}(\Omega; C([0, T]; K_1))$ to $L_{\mathcal{P}}^{r_1}(\Omega; L^2([0, T]; K_3))$.

We are now in a position of showing existence and uniqueness and regular dependence on data of the solution of equation (4.14). For $p \geq 2$ we denote:

$$\mathcal{K}_p = L_{\mathcal{P}}^p(\Omega; C([0, T]; K)) \times L_{\mathcal{P}}^p(\Omega; L^2([0, T]; L_2(\Xi, K)))$$

endowed with the natural norm.

PROPOSITION 4.8. *Assume Hypotheses* 3.1 *and* 4.6.

(i) *If $X \in L_{\mathcal{P}}^\rho(\Omega; C([0, T]; H))$, $\eta \in L^r(\Omega; K)$ with $\rho = r(m + 1)$, $r \geq 2$, then there exists a unique solution in $\mathcal{K}_r$ of equation (4.14) that will be denoted by $(Y(\cdot, X, \eta), Z(\cdot, X, \eta))$.*

(ii) *The following estimate holds:*

$$(4.17) \quad \begin{aligned} &\mathbb{E} \sup_{\tau \in [0, T]} |Y(\tau, X, \eta)|^r + \left(\mathbb{E} \int_0^T |Z(\sigma, X, \eta)|^2 \, d\sigma \right)^{r/2} \\ &\leq c\big(1 + |X|_{L_{\mathcal{P}}^\rho(\Omega; C([0, T]; H))}^\rho\big) + c\mathbb{E}|\eta|^r \end{aligned}$$

for a suitable constant c depending only on ρ, r and ψ.

(iii) *The map $(X, \eta) \to (Y(\cdot, X, \eta), Z(\cdot, X, \eta))$ is continuous from $L_{\mathcal{P}}^{\rho}(\Omega; C([0, T]; H)) \times L^r(\Omega; K)$ to $\mathcal{K}_r$.*

(iv) *The map $(X, \eta) \to (Y(\cdot, X, \eta), Z(\cdot, X, \eta))$ is in $\mathcal{G}^{1,1}(L_{\mathcal{P}}^{\rho}(\Omega; C([0, T]; H)) \times L^r(\Omega; \mathbb{R}), \mathcal{K}_p)$ with $r = (m+2)p$, $p \geq 2$ [consequently, $\rho = p(m+1)(m+2)$].*

Moreover, for all $X \in L_{\mathcal{P}}^{\rho}(\Omega; C([0, T]; H))$, $\eta \in L^r(\Omega; K)$ the directional derivative in the direction (N, ζ) with $N \in L_{\mathcal{P}}^{\rho}(\Omega; C([0, T]; H))$ and $\zeta \in L^r(\Omega; K)$ that we will denote by $(\nabla_{X,\eta} Y(\cdot, X, \eta)(N, \zeta), \nabla_{X,\eta} Z(\cdot, X, \eta)(N, \zeta))$ is the unique solution in $\mathcal{K}_p$ of

$$\nabla_{X,\eta} Y(\tau, X, \eta)(N, \zeta) + \int_{\tau}^{T} \nabla_{X,\eta} Z(\sigma, X, \eta)(N, \zeta) \, dW_{\sigma}$$

$$= -\int_{\tau}^{T} \nabla_x \psi(\sigma, X_{\sigma}, Y(\sigma, X, \eta), Z(\sigma, X, \eta)) N_{\sigma} \, d\sigma$$

$$- \int_{\tau}^{T} \nabla_y \psi(\sigma, X_{\sigma}, Y(\sigma, X, \eta), Z(\sigma, X, \eta)) \nabla_{X,\eta} Y(\sigma, X, \eta)(N, \zeta) \, d\sigma$$

$$- \int_{\tau}^{T} \nabla_z \psi(\sigma, X_{\sigma}, Y(\sigma, X, \eta), Z(\sigma, X, \eta)) \nabla_{X,\eta} Z(\sigma, X, \eta)(N, \zeta) \, d\sigma + \zeta.$$

(v) *Finally the following estimate holds*:

$$\mathbb{E} \sup_{\tau \in [0, T]} |\nabla_{X,\eta} Y(\tau, X, \eta)(N, \zeta)|^p$$

(4.18)
$$+ \mathbb{E}\left(\int_0^T |\nabla_{X,\eta} Z(\sigma, X, \eta)(N, \zeta)|^2 \, d\sigma \right)^{p/2}$$
$$\leq c |N|_{L_{\mathcal{P}}^p(\Omega; C([0, T]; H))}^p \left(1 + |X|_{L_{\mathcal{P}}^\rho(\Omega; C([0, T]; H))}^{(m+1)^2} + |\eta|_{L^r(\Omega; K)}^{m+1} \right)^p$$
$$+ c |\zeta|_{L^p(\Omega; K)}^p.$$

PROOF. Let $\Lambda = L_{\mathcal{P}}^{\rho}(\Omega; C([0, T]; H))$ and, for every $X \in \Lambda$,

$$f_X(\sigma, y, z) = \psi(\sigma, X_{\sigma}, y, z).$$

By (4.15) and Lemma 4.7(ii) applied with $K_1 = H$, $K_2 = K \times L_2(\Xi, K)$, $U = X$, $V = (Y, Z)$ we obtain that for all $(Y, Z) \in \mathcal{K}_r$ the map $X \to f_X(Y, Z)$ is continuous from Λ to $L_{\mathcal{P}}^r(\Omega; L^2([0, T]; K))$ and

$$\mathbb{E}\left(\int_0^T |f_X(\sigma, 0, 0)|^2 \, d\sigma \right)^{r/2} \leq c\left(1 + \mathbb{E}\left(\sup_{\sigma \in [0, T]} |X_{\sigma}|^{r(m+1)} \right) \right).$$

Therefore points (i)–(iii) of the claim follow immediately from Proposition 4.3.

To deal with point (iv) it is convenient now to introduce another backward stochastic equation; we will eventually show that it is satisfied by the derivatives of (Y, Z) with respect to X and η. For all $\zeta \in L^p(\Omega; K)$, $X, N \in L_{\mathcal{P}}^r(\Omega; C([0, T]; H))$, $(Y, Z) \in \mathcal{K}_r$ we look for $(\widehat{Y}(X, N, Y, Z, \zeta), \widehat{Z}(X, N, Y, Z, \zeta)) \in \mathcal{K}_p$ solving

$$\widehat{Y}_\tau + \int_\tau^T \widehat{Z}_\sigma \, dW_\sigma = -\int_\tau^T \nabla_x \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) N_\sigma \, d\sigma$$

(4.19)
$$-\int_\tau^T \nabla_y \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) \widehat{Y}_\sigma \, d\sigma$$

$$-\int_\tau^T \nabla_z \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) \widehat{Z}_\sigma \, d\sigma + \zeta.$$

By Hypothesis 4.6(iii) we have

$$\mathbb{E}\left(\int_0^T |\nabla_x \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) N_\sigma|^2 \, d\sigma \right)^{p/2}$$

$$\leq L |N|_{L_\mathcal{P}^r(\Omega; C([0,T];H))}^p \left(1 + |Z|_{L_\mathcal{P}^r(\Omega; L^2([0,T];L_2(\Xi,K)))} \right)^p$$

$$\times \left(1 + |X|_{L_\mathcal{P}^r(\Omega; C([0,T];H))}^m + |Y|_{L_\mathcal{P}^r(\Omega; C([0,T];H))}^m \right)^p$$

for a suitable constant L. Since $\nabla_y \psi$ and $\nabla_z \psi$ are bounded, by Proposition 4.3 the equation (4.19) admits a unique solution in $\mathcal{K}_p$. Moreover, by Lemma 4.7(i), the map $(X, N, Y, Z) \to \nabla_x \psi(\cdot, X_{(\cdot)}, Y_{(\cdot)}, Z_{(\cdot)}) N_{(\cdot)}$ is continuous from the space

$$K^\# := L_\mathcal{P}^r\big(\Omega; C([0,T];H)\big) \times L_\mathcal{P}^r\big(\Omega; C([0,T];H)\big) \times \mathcal{K}_r$$

to $L_\mathcal{P}^p(\Omega; L^2([0,T];K))$. Therefore, taking into account once more the boundedness of $\nabla_y \psi$ and $\nabla_z \psi$, we can apply the final statement of Proposition 4.3 with $\Lambda = K^\#$ and conclude that the map $(X, N, Y, Z, \zeta) \to (\widehat{Y}(X, N, Y, Z, \zeta), \widehat{Z}(X, N, Y, Z, \zeta))$ is continuous from $K^\# \times L^p(\Omega; K)$ to $\mathcal{K}_p$ and the estimate

$$\mathbb{E}\left(\sup_{\sigma \in [0,T]} |\widehat{Y}_\sigma|^p \right) + \mathbb{E}\left(\int_0^T |\widehat{Z}_\sigma|^2 d\sigma \right)^{p/2}$$

(4.20)
$$\leq c |N|_{L_\mathcal{P}^r(\Omega; C([0,T];H))}^p \left(1 + |Z|_{L_\mathcal{P}^r(\Omega; L^2([0,T];L_2(\Xi,K)))} \right)^p$$

$$\times \left(1 + |X|_{L_\mathcal{P}^r(\Omega; C([0,T];H))}^m + |Y|_{L_\mathcal{P}^r(\Omega; C([0,T];H))}^m \right)^p + c\mathbb{E}|\zeta|^p$$

holds for some constant $c > 0$.

It remains to prove that if $X, N \in L_\mathcal{P}^\rho(\Omega; C([0,T];H))$ and $\eta, \zeta \in L^r(\Omega; K)$ then the directional derivative of $(Y(X, \eta), Z(X, \eta))$ in the direction (N, ζ) is given by

$$\big(\widehat{Y}(X, N, Y(X, \eta), Z(X, \eta), \zeta), \widehat{Z}(X, N, Y(X, \eta), Z(X, \eta), \zeta) \big).$$

Let us define

$$\overline{Y}^\varepsilon := \frac{1}{\varepsilon} [Y(X + \varepsilon N, \eta + \varepsilon \zeta) - Y(X, \eta)] - \widehat{Y}(X, N, Y(X, \eta), Z(X, \eta), \zeta),$$

$$\overline{Z}^\varepsilon := \frac{1}{\varepsilon} [Z(X + \varepsilon N, \eta + \varepsilon \zeta) - Z(X, \eta)] - \widehat{Z}(X, N, Y(X, \eta), Z(X, \eta), \zeta).$$

For $\varepsilon \to 0$ we have $\overline{Y}_\varepsilon \to 0$ in $L^p_{\mathcal{P}}(\Omega; C([0, T]; K))$ and $\overline{Z}_\varepsilon \to 0$ in $L^p_{\mathcal{P}}(\Omega; L^2([0, T]; L_2(\Xi, K)))$. For short we let $Y = Y(X, \eta), Z = Z(X, \eta), Y^\varepsilon = Y(X + \varepsilon N, \eta + \varepsilon \zeta), Z^\varepsilon = Z(X + \varepsilon N, \eta + \varepsilon \zeta), \widehat{Y} = \widehat{Y}(X, N, Y(X, \eta), Z(X, \eta), \zeta), \widehat{Z} = \widehat{Z}(X, N, Y(X, \eta), Z(X, \eta), \zeta)$.

The proof will be done by induction, dividing the interval $[0, T]$ into subintervals $[T - \delta, T], [T - 2\delta, T - \delta]$ and so on, for a suitable δ depending only on ψ and p. All the subintervals are treated in the same way (the proof for $[T - \delta, T]$ being even easier), so we concentrate on the second one, namely $[T - 2\delta, T - \delta]$. On such interval we have

$$\overline{Y}^\varepsilon_\tau + \int_\tau^{T-\delta} \overline{Z}^\varepsilon_\sigma \, d\sigma = -\int_\tau^{T-\delta} v^\varepsilon(\sigma) \, d\sigma + \overline{Y}^\varepsilon_{T-\delta},$$

where $v^\varepsilon = v^\varepsilon_1 + v^\varepsilon_2$ and

$$v^\varepsilon_1(\sigma) = \frac{1}{\varepsilon}[\psi(\sigma, X_\sigma + \varepsilon N_\sigma, Y^\varepsilon_\sigma, Z^\varepsilon_\sigma) - \psi(\sigma, X_\sigma, Y^\varepsilon_\sigma, Z^\varepsilon_\sigma)]$$

$$- \nabla_x \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) N_\sigma,$$

$$v^\varepsilon_2(\sigma) = \frac{1}{\varepsilon}[\psi(\sigma, X_\sigma, Y^\varepsilon_\sigma, Z^\varepsilon_\sigma) - \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma)] - \nabla_y \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) \widehat{Y}_\sigma$$

$$- \nabla_z \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) \widehat{Z}_\sigma.$$

By Proposition 4.3 we have

$$\mathbb{E} \sup_{\tau \in [T-2\delta, T-\delta]} |\overline{Y}^\varepsilon_\tau|^p + \mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |\overline{Z}^\varepsilon_\sigma|^2 \, d\sigma\right)^{p/2}$$

$$\leq c_p \delta^{p/2} \sum_{i=1}^2 \mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |v^\varepsilon_i(\sigma)|^2 \, d\sigma\right)^{p/2} + c_p \mathbb{E}|\overline{Y}^\varepsilon_{T-\delta}|^p$$

and by the inductive assumption $\mathbb{E}|\overline{Y}^\varepsilon_{T-\delta}|^p \to 0$.

We start to evaluate the integral terms on the right. We can write

$$v^\varepsilon_1(\sigma) = \int_0^1 \nabla_x \psi(\sigma, X_\sigma + \varepsilon \tau N_\sigma, Y^\varepsilon_\sigma, Z^\varepsilon_\sigma) N_\sigma \, d\tau$$

$$- \int_0^1 \nabla_x \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) N_\sigma \, d\tau.$$

For all $x, g, n \in H$, $y \in K$, $z \in L_2(\Xi, K)$ let $\chi(x, g, n, y, z) = \int_0^1 \nabla_x \psi(x + \tau g, y, z) n \, d\tau$, so that $v^\varepsilon_1(\sigma) = \chi(X_\sigma, \varepsilon N_\sigma, N_\sigma, Y^\varepsilon_\sigma, Z^\varepsilon_\sigma) - \chi(X_\sigma, 0, N_\sigma, Y_\sigma, Z_\sigma)$. Moreover $|\chi(x, g, n, y, z)| \leq L|n|(1 + |z|)(1 + |x|^m + |g|^m + |y|^m)$ and χ is a continuous map. Applying Lemma 4.7(i) with $K_1 = H^{\times 3} \times K$, $K_2 = L_2(\Xi, K)$, $r_1 = r_2 = r$, $\mu = m + 1$ and taking into account that $(X, \varepsilon N, N, Y^\varepsilon) \to$

$(X, 0, N, Y)$ in $L^r_{\mathscr{P}}(\Omega, C([T-2\delta, T-\delta], K_1))$ and $Z^\varepsilon \to Z$ in $L^r_{\mathscr{P}}(\Omega, L^2([T-2\delta, T-\delta], K_2))$ we immediately obtain $\mathbb{E}(\int_{T-2\delta}^{T-\delta} |v_1^\varepsilon(\sigma)|^2 d\sigma)^{p/2} \to 0$.

Dealing now with v_2^ε we can rewrite $v_2^\varepsilon = v_{2.1}^\varepsilon + v_{2.2}^\varepsilon$ where

$$
v_{2.1}^\varepsilon(\sigma) = \int_0^1 \big(\nabla_y \psi\big(\sigma, X_\sigma, Y_\sigma + \tau(Y_\sigma^\varepsilon - Y_\sigma), Z_\sigma + \tau(Z_\sigma^\varepsilon - Z_\sigma)\big)\widehat{Y}_\sigma
$$
$$
- \nabla_y \psi\big(\sigma, X_\sigma, Y_\sigma, Z_\sigma\big)\widehat{Y}_\sigma\big)\, d\tau
$$
$$
+ \int_0^1 \big(\nabla_z \psi\big(\sigma, X_\sigma, Y_\sigma + \tau(Y_\sigma^\varepsilon - Y_\sigma), Z_\sigma + \tau(Z_\sigma^\varepsilon - Z_\sigma)\big)\widehat{Z}_\sigma
$$
$$
- \nabla_z \psi\big(\sigma, X_\sigma, Y_\sigma, Z_\sigma\big)\widehat{Z}_\sigma\big)\, d\tau,
$$
$$
v_{2.2}^\varepsilon(\sigma) = \int_0^1 \nabla_y \psi\big(\sigma, X_\sigma, Y_\sigma + \tau(Y_\sigma^\varepsilon - Y_\sigma), Z_\sigma + \tau(Z_\sigma^\varepsilon - Z_\sigma)\big)\overline{Y}_\sigma^\varepsilon\, d\tau
$$
$$
+ \int_0^1 \nabla_z \psi\big(\sigma, X_\sigma, Y_\sigma + \tau(Y_\sigma^\varepsilon - Y_\sigma), Z_\sigma + \tau(Z_\sigma^\varepsilon - Z_\sigma)\big)\overline{Z}_\sigma^\varepsilon\, d\tau.
$$

Since $\nabla_y \psi$ and $\nabla_z \psi$ are bounded, by the dominated convergence theorem we immediately obtain $\mathbb{E}(\int_{T-2\delta}^{T-\delta} |v_{2.1}^\varepsilon(\sigma)|^2 d\sigma)^{p/2} \to 0$. Moreover,

$$
\mathbb{E}\bigg(\int_{T-2\delta}^{T-\delta} |v_{2.2}^\varepsilon(\sigma)|^2\, d\sigma\bigg)^{p/2}
$$
$$
\leq \overline{c}\bigg(\mathbb{E}\sup_{\tau \in [T-2\delta, T-\delta]} |\overline{Y}_\tau^\varepsilon|^p + \mathbb{E}\bigg(\int_{T-2\delta}^{T-\delta} |\overline{Z}_\sigma^\varepsilon|^2\, d\sigma\bigg)^{p/2}\bigg)
$$

for a suitable constant $\overline{c}$ depending only on ψ, p, T. Choosing δ such that $c_p \overline{c} \delta^{p/2} < 1$ the claim follows immediately.

Finally (4.18) follows plugging (4.17) into (4.20). $\quad\square$

REMARK 4.9. If Hypothesis 4.6(iii) is replaced by the stronger requirement:
(iii-bis) There exists $L > 0$ such that $|\nabla_x \psi(\sigma, x, y, z)h| \leq L|h|$, for every $\sigma \in [0, T]$, $x, h \in H$, $y \in K$, $z \in L_2(\Xi, K)$.

Then instead of (4.18) we can obtain, with identical proof, the following stronger estimate:

$$
\mathbb{E}\sup_{\tau \in [0,T]} |\nabla_{X,\eta} Y(\tau, X, \eta)(N, \zeta)|^p
$$
$$
(4.21) \qquad + \mathbb{E}\bigg(\int_0^T |\nabla_{X,\eta} Z(\sigma, X, \eta)(N, \zeta)|^2\, d\sigma\bigg)^{p/2}
$$
$$
\leq c|N|^p_{L^p_{\mathscr{P}}(\Omega; C([0,T];H))} + c|\zeta|^p_{L^p(\Omega;K)}.
$$

5. The backward–forward system. In this section we consider the system of stochastic differential equations

$$
(5.1) \quad
\begin{cases}
X_\tau = e^{(\tau-t)A}x + \displaystyle\int_t^\tau e^{(\tau-\sigma)A} F(\sigma, X_\sigma)\, d\sigma \\
\qquad + \displaystyle\int_t^\tau e^{(\tau-\sigma)A} G(\sigma, X_\sigma)\, dW_\sigma, \\
Y_\tau + \displaystyle\int_\tau^T Z_\sigma\, dW_\sigma = -\displaystyle\int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma)\, d\sigma + \phi(X_T),
\end{cases}
$$

for τ varying on the time interval $[t, T] \subset [0, T]$. As in Section 2 we extend the domain of the solution setting $X(\tau, t, x) = x$ for $\tau \in [0, t)$. We assume that $\psi \colon [0, T] \times H \times \mathbb{R} \times L_2(\Xi, \mathbb{R}) \to \mathbb{R}$ verifies Hypothesis 4.6 with $K = \mathbb{R}$. On the function $\phi \colon H \to \mathbb{R}$ we make the following assumptions.

HYPOTHESIS 5.1. (i) *There exists $L > 0$ such that, for every $x_1, x_2 \in H$,*

$$|\phi(x_1) - \phi(x_2)| \le L\,|x_1 - x_2|;$$

(ii) $\phi \in \mathcal{G}^1(H, \mathbb{R})$.

Notice that the system is decoupled, that is, the first equation does not contain the solution (Y, Z) of the second one. Therefore, under the assumptions of Hypotheses 3.1, 4.6, 5.1 by Propositions 3.2 and 4.8 there exists a unique solution of (5.1). We remark that the process X is $\mathcal{F}_{[t,T]}$-measurable, so that Y_t is measurable both with respect to $\mathcal{F}_{[t,T]}$ and $\mathcal{F}_t$; it follows that Y_t is indeed deterministic (see also [11]).

We denote by $(X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x))$, $\tau \in [t, T]$ the solution, in order to stress dependence on the parameters $t \in [0, T]$ and $x \in H$.

For later use we notice two useful identities: for $t \le s \le T$ the equality, $\mathbb{P}$-a.s.,

$$
(5.2) \quad X\big(\tau, s, X(s, t, x)\big) = X(\tau, t, x), \qquad \tau \in [s, T],
$$

is a consequence of the uniqueness of the solution of (3.4). Since the solution of the backward equation is uniquely determined on an interval $[s, T]$ by the values of the process X on the same interval, for $t \le s \le T$ we have, $\mathbb{P}$-a.s.,

$$
(5.3) \quad
\begin{aligned}
Y\big(\tau, s, X(s, t, x)\big) &= Y(\tau, t, x) &&\text{for } \tau \in [s, T], \\
Z\big(\tau, s, X(s, t, x)\big) &= Z(\tau, t, x) &&\text{for a.a. } \tau \in [s, T].
\end{aligned}
$$

5.1. *Regularity with respect to parameters.* To investigate regularity properties of the dependence on t and x, we notice that with the notation of Propositions 3.3 and 4.8:

$$
\begin{aligned}
Y(\sigma, t, x) &= Y\big(\sigma, X(\cdot, t, x), \phi(X(T, t, x))\big), \\
Z(\sigma, t, x) &= Z\big(\sigma, X(\cdot, t, x), \phi(X(T, t, x))\big).
\end{aligned}
$$

Moreover, as a consequence of Hypothesis 5.1, it can be easily proved that the map $\eta \mapsto \phi(\eta)$ belongs to the space $\mathcal{G}^1(L^p(\Omega; H), L^p(\Omega; \mathbb{R}))$, for every $p \in [2, \infty)$. The following proposition is then an immediate consequence of Propositions 3.2, 3.3 and 4.8, and the chain rule for the class $\mathcal{G}$, stated in Lemma 2.1.

PROPOSITION 5.2.   *Assume Hypotheses* 3.1, 4.6 *and* 5.1. *Recall the notation*

$$\mathcal{K}_p = L^p_{\mathcal{P}}(\Omega; C([0, T]; \mathbb{R})) \times L^p_{\mathcal{P}}(\Omega; L^2([0, T]; L_2(\Xi, \mathbb{R}))).$$

Then the map $(t, x) \mapsto (Y(\cdot, t, x), Z(\cdot, t, x))$ *belongs to* $\mathcal{G}^{0,1}([0, T] \times H, \mathcal{K}_p)$ *for all* $p \in [2, \infty)$.

Denoting by $\nabla_x Y$, $\nabla_x Z$ *the partial Gâteaux derivatives with respect to* x, *the directional derivative process in the direction* $h \in H$, $\{\nabla_x Y(\tau, t, x)h, \nabla_x Z(\tau, t, x)h, \tau \in [0, T]\}$ *solves the equation*, $\mathbb{P}$-*a.s.*,

$$
\begin{aligned}
\nabla_x Y(\tau, t, x)h &+ \int_\tau^T \nabla_x Z(\sigma, t, x)h \, dW_\sigma \\
&= - \int_\tau^T \nabla_x \psi\big(\sigma, X(\sigma, t, x), Y(\sigma, t, x), Z(\sigma, t, x)\big) \nabla_x X(\sigma, t, x)h \, d\sigma \\
&\quad - \int_\tau^T \nabla_y \psi\big(\sigma, X(\sigma, t, x), Y(\sigma, t, x), Z(\sigma, t, x)\big) \nabla_x Y(\sigma, t, x)h \, d\sigma \\
&\quad - \int_\tau^T \nabla_z \psi\big(\sigma, X(\sigma, t, x), Y(\sigma, t, x), Z(\sigma, t, x)\big) \nabla_x Z(\sigma, t, x)h \, d\sigma \\
&\quad + \nabla\phi\big(X(T, t, x)\big) \nabla_x X(T, t, x)h, \qquad \tau \in [0, T].
\end{aligned}
$$

(5.4)

Finally the following estimate holds:

$$
\begin{aligned}
\Big[\mathbb{E} \sup_{\tau \in [0,T]} |\nabla_x Y(\tau, t, x)h|^p \Big]^{1/p} &+ \Big[\mathbb{E} \Big(\int_0^T |\nabla_x Z(\sigma, t, x)h|^2 \, d\sigma \Big)^{p/2} \Big]^{1/p} \\
&\leq c|h|(1 + |x|^{(m+1)^2}).
\end{aligned}
$$

(5.5)

PROOF.   On the first two statements we have already commented. The final estimate follows from (4.18) applied with

$$
\begin{aligned}
X &= X(\cdot, t, x), & N &= \nabla_x X(\cdot, t, x)h, \\
\eta &= \phi\big(X(T, t, x)\big), & \zeta &= \nabla\phi\big(X(T, t, x)\big) \nabla_x X(T, t, x)h,
\end{aligned}
$$

taking into account that by Propositions 3.2 and 3.3 we have

$$|N|_{L^p_{\mathcal{P}}(\Omega; C([0,T]; H))} \leq c|h|, \qquad |X|_{L^p_{\mathcal{P}}(\Omega; C([0,T]; H))} \leq c(1 + |x|),$$

and, by Hypothesis 5.1, we also obtain $|\eta|_{L^r(\Omega)} \leq c(1 + |x|)$, $|\zeta|_{L^p(\Omega)} \leq c|h|$ for a suitable constant c. $\square$

REMARK 5.3. If Hypothesis 4.6(iii) is replaced by (iii-bis) in Remark 4.9 then applying (4.21) instead of (4.18) we obtain, instead of (5.5) the stronger estimate,

$$(5.6) \quad \left[\mathbb{E} \sup_{\tau \in [0,T]} |\nabla_x Y(\tau, t, x)h|^p \right]^{1/p} + \left[\mathbb{E} \left(\int_0^T |\nabla_x Z(\sigma, t, x)h|^2 \, d\sigma \right)^{p/2} \right]^{1/p}$$
$$\leq c|h|.$$

5.2. *Regularity in the Malliavin spaces.* It has been proved in Proposition 3.5 that X belongs to the Malliavin space $\mathbb{L}^{1,2}(H)$ and DX has additional smoothness properties. Similar results hold for (Y, Z).

PROPOSITION 5.4. *Assume Hypotheses* 3.1, 4.6 *and* 5.1. *Then the following properties hold*:

(i) $Y \in \mathbb{L}^{1,2}(\mathbb{R})$, $Z \in \mathbb{L}^{1,2}(L_2(\Xi, \mathbb{R}))$.

(ii) *There exists a version of* (DY, DZ) *such that for a.a.* $s \in [t, T)$, *the process* $\{(D_s Y_\tau, D_s Z_\tau), \ \tau \in [s, T]\}$ *belongs to* $L_{\mathcal{P}}^2(\Omega; C([s, T]; L_2(\Xi, \mathbb{R}))) \times L^2(\Omega \times [s, T]; L_2(\Xi, L_2(\Xi, \mathbb{R})))$ *and satisfies*, $\mathbb{P}$-*a.s.*,

$$
\begin{aligned}
(5.7) \quad D_s Y_\tau + \int_\tau^T D_s Z_\sigma \, dW_\sigma &= -\int_\tau^T \nabla_x \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) D_s X_\sigma \, d\sigma \\
&\quad - \int_\tau^T \nabla_y \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) D_s Y_\sigma \, d\sigma \\
&\quad - \int_\tau^T \nabla_z \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) D_s Z_\sigma \, d\sigma \\
&\quad + \nabla \phi(X_T) D_s X_T, \qquad \tau \in [s, T].
\end{aligned}
$$

(iii) *For a.a.* $s \in [t, T)$, *we have*

$$(5.8) \qquad\qquad Z_s = \lim_{\tau \downarrow s} D_s Y_\tau$$

in the norm of $L^2(\Xi; \mathbb{R})$.

In the following lemma we collect two facts needed for the proof. Part (i) is a simple consequence of the definition, while a proof of part (ii) can be found in [36], Lemma 2.3, for the finite-dimensional case; the extension to the present case follows from the results in [20].

LEMMA 5.5. (i) *If* $u \in L^2(\Omega \times [0, T]; \mathbb{R})$, $u_r \in \mathbb{D}^{1,2}(\mathbb{R})$ *for a.a.* r *and there exists a version of* Du *such that* $\mathbb{E} \int_0^T (\int_0^T |D_s u_r| dr)^2 ds < \infty$, *then*

$$\int_0^T u_r \, dr \in \mathbb{D}^{1,2}(\mathbb{R}) \quad and \quad D_s \int_0^T u_r \, dr = \int_0^T D_s u_r \, dr.$$

(ii) *Suppose* $Z \in L^2_{\mathcal{P}}(\Omega \times [0,T]; L_2(\Xi, \mathbb{R}))$ *and* $\int_0^T Z_\sigma \, dW_\sigma \in \mathbb{D}^{1,2}(\mathbb{R})$. *Then the process* $\{Z_\sigma, \sigma \in [0,T]\}$, *belongs to* $\mathbb{L}^{1,2}(L_2(\Xi, \mathbb{R}))$ *and for a.a.* $s \in [0,T]$

$$D_s \int_0^T Z_\sigma \, dW_\sigma = Z_s + \int_0^T D_s Z_\sigma \, dW_\sigma.$$

PROOF OF PROPOSITION 5.4.   We extend the domain of the process X setting $X_\tau = x$ for $\tau \in [0,t)$. This way the backward equation in system (5.1) has a solution defined on $[0,T]$.

We define a sequence (Y^n, Z^n) setting $Y^0 = 0$, $Z^0 = 0$ and letting (Y^{n+1}, Z^{n+1}) be the pair such that, for $\tau \in [0,T]$,

$$Y_\tau^{n+1} + \int_\tau^T Z_\sigma^{n+1} \, dW_\sigma = -\int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) \, d\sigma + \phi(X_T).$$

It has been shown in the proof of Proposition 4.3 (see also the remark following that proposition) that the map $(Y^n, Z^n) \mapsto (Y^{n+1}, Z^{n+1})$ is a contraction in the space $L^2_{\mathcal{P}}(\Omega \times [0,T]; \mathbb{R}) \times L^2_{\mathcal{P}}(\Omega \times [0,T]; L_2(\Xi, \mathbb{R}))$, so that the sequence converges to the solution (Y,Z) in the norm of this space. By the closedness of the operator D, in order to prove point (i) of Proposition 5.4 it suffices to show that $Y^n \in \mathbb{L}^{1,2}(\mathbb{R})$ and $Z^n \in \mathbb{L}^{1,2}(L_2(\Xi, \mathbb{R}))$ for every n and that (DY^n, DZ^n) is a bounded sequence in $L^2_{\mathcal{P}}(\Omega \times [0,T] \times [0,T]; L_2(\Xi, \mathbb{R})) \times L^2_{\mathcal{P}}(\Omega \times [0,T] \times [0,T]; L_2(\Xi, L_2(\Xi, \mathbb{R})))$.

Proceeding by induction, assume that $Y^n \in \mathbb{L}^{1,2}(\mathbb{R})$ and $Z^n \in \mathbb{L}^{1,2}(L_2(\Xi, \mathbb{R}))$ for some n. We claim that $\int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) \, d\sigma \in \mathbb{D}^{1,2}(\mathbb{R})$ for every τ and for a.a. s,

$$(5.9) \qquad D_s \int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) \, d\sigma = \int_\tau^T f^n(s, \sigma) \, d\sigma,$$

where for brevity we set

$$f^n(s,\sigma) = \nabla_x \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) D_s X_\sigma + \nabla_y \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) D_s Y_\sigma^n$$
$$+ \nabla_z \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) D_s Z_\sigma^n.$$

By Hypothesis 4.6 and the chain rule for the Malliavin derivative [Lemma 3.4(i)], for a.a. σ we have $\psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) \in \mathbb{D}^{1,2}_{loc}(\mathbb{R})$ and $D\psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) = f^n(\cdot, \sigma)$. If we can show that

$$(5.10) \qquad \mathbb{E} \int_0^T \int_0^T |f^n(s,\sigma)|^2 \, ds \, d\sigma < \infty,$$

then, by Lemma 3.4(ii), $\psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) \in \mathbb{D}^{1,2}(\mathbb{R})$ for a.a. σ and the claim follows from Lemma 5.5(i). Next we prove (5.10). By the assumptions on ψ,

$$(5.11) \qquad \begin{aligned} |f^n(s,\sigma)| &\leq L \, (1 + |Z_\sigma^n|)(1 + |X_\sigma| + |Y_\sigma^n|)^m |D_s X_\sigma| \\ &\quad + L \, |D_s Y_\sigma^n| + L \, |D_s Z_\sigma^n|. \end{aligned}$$

So setting $I^n = \int_0^T \int_s^T (1 + |Z_\sigma^n|)^2 (1 + |X_\sigma| + |Y_\sigma^n|)^{2m} |D_s X_\sigma|^2 \, d\sigma \, ds$, it suffices to show that $\mathbb{E} \, I^n < \infty$. We will even show that

$$(5.12) \quad \sup_n \mathbb{E} \int_0^T \int_s^T (1 + |Z_\sigma^n|)^2 (1 + |X_\sigma| + |Y_\sigma^n|)^{2m} |D_s X_\sigma|^2 \, d\sigma \, ds < \infty.$$

By the Hölder inequality, for every $p > 1$, setting $p' = p/(p-1)$,

$$I^n \leq \left(\sup_{\sigma \in [0,T]} (1 + |X_\sigma| + |Y_\sigma^n|)^{2m} \right) \left(\int_0^T \sup_{\sigma \in [s,T]} (\sigma - s)^{2\gamma p'} |D_s X_\sigma|^{2p'} \, ds \right)^{1/p'}$$

$$\times \left(\int_0^T \left(\int_s^T (\sigma - s)^{-2\gamma} (1 + |Z_\sigma^n|)^2 \, d\sigma \right)^p d\sigma \right)^{1/p} ds.$$

By the Minkowsky inequality,

$$\left(\int_0^T \left(\int_s^T (\sigma - s)^{-2\gamma} (1 + |Z_\sigma^n|)^2 \, d\sigma \right)^p d\sigma \right)^{1/p} ds$$

$$\leq \int_0^T (1 + |Z_\sigma^n|)^2 \left(\int_0^\sigma (\sigma - s)^{-2\gamma p} \, ds \right)^{1/p} d\sigma$$

$$\leq \int_0^T (1 + |Z_\sigma^n|)^2 d\sigma \left(\int_0^T s^{-2\gamma p} ds \right)^{1/p} = c \int_0^T (1 + |Z_\sigma^n|)^2 \, d\sigma,$$

provided p is so small that $2\gamma p < 1$. Taking expectation, and using the Hölder inequality again,

$$\mathbb{E} \, I^n \leq c \left\{ \mathbb{E} \sup_{\sigma \in [0,T]} (1 + |X_\sigma| + |Y_\sigma^n|)^{4mp} \right\}^{1/(2p)} \left\{ \mathbb{E} \left(\int_0^T (1 + |Z_\sigma^n|)^2 \, d\sigma \right)^{2p} \right\}^{1/(2p)}$$

$$\times \left\{ \int_0^T \mathbb{E} \sup_{\sigma \in [s,T]} (\sigma - s)^{2\gamma p'} |D_s X_\sigma|^{2p'} \, ds \right\}^{1/p'}.$$

Now (5.12) follows from (3.11) and Lemma 4.5. We have therefore completed the proof that $\int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) \, d\sigma \in \mathbb{D}^{1,2}(\mathbb{R})$ for every τ and for a.a. s and that (5.9) holds.

By the chain rule, $\phi(X_T) \in \mathbb{D}^{1,2}(\mathbb{R})$ and $D_s \phi(X_T) = \nabla \phi(X_T) D_s X_T$ for a.a. s. Since

$$Y_\tau^{n+1} = \mathbb{E}^{\mathcal{F}_\tau} \left(- \int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n) \, d\sigma + \phi(X_T) \right),$$

we also have $Y_\tau^{n+1} \in \mathbb{D}^{1,2}(\mathbb{R})$ for every τ. This implies that

$$\int_\tau^T Z_\sigma^{n+1} \, dW_\sigma = \int_0^T \mathbb{1}_{[\tau,T]}(\sigma) Z_\sigma^{n+1} \, dW_\sigma \in \mathbb{D}^{1,2}(\mathbb{R}).$$

It follows from Lemma 5.5(ii) that $\mathbb{1}_{[\tau,T]}Z^{n+1} \in \mathbb{L}^{1,2}(L_2(\Xi,\mathbb{R}))$ and that for a.a. $s \in [0,T]$,

$$
\begin{aligned}
D_s Y_\tau^{n+1} &+ \mathbb{1}_{[\tau,T]}(s)Z_s^{n+1} + \int_\tau^T D_s Z_\sigma^{n+1}\, dW_\sigma \\
&= -\int_\tau^T \nabla_x \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n)D_s X_\sigma\, d\sigma \\
&\quad -\int_\tau^T \nabla_y \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n)D_s Y_\sigma^n d\sigma \\
&\quad -\int_\tau^T \nabla_z \psi(\sigma, X_\sigma, Y_\sigma^n, Z_\sigma^n)D_s Z_\sigma^n\, d\sigma + \nabla\phi(X_T)D_s X_T.
\end{aligned}
$$
(5.13)

Let us take $s < \tau$ in equality (5.13), so that $\mathbb{1}_{[\tau,T]}(s)Z_s^{n+1} = 0$; then for a.a. s we can apply the estimates (4.2) to the pair $(D_s Y_\tau, D_s Z_\tau)$, $\tau \in [s, T]$; using again the estimate (5.11) we obtain

$$
\begin{aligned}
\mathbb{E}\int_s^T & e^{2\beta\sigma}(|D_s Y_\sigma^{n+1}|^2 + |D_s Z_\sigma^{n+1}|^2)\, d\sigma \\
&\leq (T-s)\mathbb{E}\sup_{\tau\in[s,T]} e^{2\beta\tau}|D_s Y_\tau|^2 + \mathbb{E}\int_s^T e^{2\beta\sigma}|D_s Z_\sigma|^2\, d\sigma \\
&\leq \frac{8(1+T)}{\beta}\,\mathbb{E}\int_s^T e^{2\beta\sigma}|f^n(s,\sigma)|^2\, d\sigma + 16(1+T)\,e^{2\beta T}\mathbb{E}\,|\nabla\phi(X_T)D_s X_T|^2 \\
&\leq \frac{24L^2(1+T)}{\beta}\,\mathbb{E}\int_s^T e^{2\beta\sigma}(|D_s Y_\sigma^n|^2 + |D_s Z_\sigma^n|^2)\, d\sigma \\
&\quad + 16(1+T)L^2\, e^{2\beta T}\mathbb{E}\,|D_s X_T|^2 \\
&\quad + \frac{24L^2(1+T)}{\beta}\,\mathbb{E}\int_s^T e^{2\beta\sigma}(1+|Z_\sigma^n|)^2(1+|X_\sigma|+|Y_\sigma^n|)^{2m}|D_s X_\sigma|^2\, d\sigma.
\end{aligned}
$$

Integrating over $[0,T]$ with respect to s and choosing β so large that $\alpha := 24L^2(1+T)\beta^{-1} < 1$ we obtain, setting $b_n = \mathbb{E}\int_0^T\int_s^T e^{2\beta\sigma}(|D_s Y_\sigma^n|^2 + |D_s Z_\sigma^n|^2)$ $\times\, d\sigma\, ds$,

$$
\begin{aligned}
b_{n+1} \leq \alpha\, b_n &+ 16(1+T)L^2\, e^{2\beta T}\mathbb{E}\int_0^T |D_s X_T|^2\, ds \\
&+ \frac{24L^2(1+T)}{\beta} \\
&\quad \times \mathbb{E}\int_0^T\int_s^T e^{2\beta\sigma}(1+|Z_\sigma^n|)^2(1+|X_\sigma|+|Y_\sigma^n|)^{2m}|D_s X_\sigma|^2\, d\sigma\, ds.
\end{aligned}
$$

It follows from (5.12) that $\sup_n b_n < \infty$. The required boundedness property of (DY^n, DZ^n) is verified. Now we have shown that $Y \in \mathbb{L}^{1,2}(\mathbb{R})$, $Z \in \mathbb{L}^{1,2}(L_2(\Xi, \mathbb{R}))$ and point (i) of the proposition is proved.

We proceed to point (ii). Repeating the arguments that led to (5.13), we conclude that for a.a. $s \in [0, T]$,

$$
(5.14) \quad
\begin{aligned}
D_s Y_\tau &+ \mathbb{1}_{[\tau,T]}(s)Z_s + \int_\tau^T D_s Z_\sigma \, dW_\sigma \\
&= - \int_\tau^T \nabla_x \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) D_s X_\sigma \, d\sigma \\
&\quad - \int_\tau^T \nabla_y \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) D_s Y_\sigma d\sigma \\
&\quad - \int_\tau^T \nabla_z \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) D_s Z_\sigma \, d\sigma + \nabla \phi(X_T) D_s X_T.
\end{aligned}
$$

This shows that equality (5.7) holds $\mathbb{P}$-a.s. for a.a. τ and s with $\tau > s$.

Now let us fix s and define

$$
\begin{aligned}
f(s, \sigma, U, V) &= \nabla_x \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) D_s X_\sigma + \nabla_y \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) U \\
&\quad + \nabla_z \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) V.
\end{aligned}
$$

Then (5.7) can be written as a backward equation: for a.a. $\tau \in [s, T]$,

$$
(5.15) \quad
\begin{aligned}
D_s Y_\tau + \int_\tau^T D_s Z_\sigma \, dW_\sigma &= - \int_\tau^T f(s, \sigma, D_s Y_\tau, D_s Z_\tau) \, d\sigma \\
&\quad + \nabla \phi(X_T) D_s X_T.
\end{aligned}
$$

Let us verify that the assumptions of Proposition 4.3 hold for this equation. The Lipschitz condition for the map $(U, V) \mapsto f(s, \sigma, U, V)$ follows from the assumption $|\nabla_y \psi| \leq L$, $|\nabla_z \psi| \leq L$. The requirement that $\mathbb{E} \int_s^T |f(s, \sigma, 0, 0)|^2 \, d\sigma < \infty$ can be verified as follows: first, by the assumptions on ψ we have the estimate

$$
\int_s^T |f(s, \sigma, 0, 0)|^2 \, d\sigma \leq L^2 \int_s^T (1 + |Z_\sigma|)^2 (1 + |X_\sigma| + |Y_\sigma|)^{2m} |D_s X_\sigma|^2 \, d\sigma.
$$

By the same arguments that led to (5.12) we obtain

$$
\begin{aligned}
\mathbb{E} \int_0^T \int_s^T &|f(s, \sigma, 0, 0)|^2 \, d\sigma \, ds \\
&\leq L^2 \mathbb{E} \int_0^T \int_s^T (1 + |Z_\sigma|)^2 (1 + |X_\sigma| + |Y_\sigma|)^{2m} |D_s X_\sigma|^2 \, d\sigma \, ds < \infty,
\end{aligned}
$$

which shows that $\mathbb{E} \int_s^T |f(s, \sigma, 0, 0)|^2 d\sigma < \infty$ holds for a.a. s. Finally,

$$
\mathbb{E} \int_0^T |\nabla_x \phi(X_T) D_s X_T|^2 \, ds \leq L^2 \mathbb{E} \int_0^T |D_s X_T|^2 \, ds < \infty,
$$

which implies $\mathbb{E}|\nabla_x\phi(X_T)D_sX_T|^2 < \infty$ for a.a. s. By Proposition 4.3, for a.a. s the solution (D_sY_τ, D_sZ_τ), $\tau \in [s, T]$, belongs to the space $L^2_{\mathcal{P}}(\Omega; C([s, T]; L_2(\Xi, \mathbb{R}))) \times L^2(\Omega \times [s, T]; L_2(\Xi, L_2(\Xi, \mathbb{R})))$. It also follows that (5.15), hence (5.7), is verified for all $\tau \in [s, T]$. Point (ii) of Proposition 5.4 is now proved.

To prove point (iii), we start from (5.15) and take the limit as $\tau \downarrow s$ obtaining, for a.a. s,

$$\lim_{\tau\downarrow s} D_sY_\tau = -\int_s^T D_sZ_\sigma\,dW_\sigma - \int_s^T f(s, \sigma, D_sY_\tau, D_sZ_\tau)\,d\sigma + \nabla\phi(X_T)D_sX_T.$$

On the other hand, we may write (5.14) for a.a. τ and s with $\tau < s$: recalling that $D_sY_\tau = 0$, we have

$$Z_s + \int_\tau^T D_sZ_\sigma\,dW_\sigma = -\int_\tau^T f(s, \sigma, D_sY_\tau, D_sZ_\tau)\,d\sigma + \nabla\phi(X_T)D_sX_T.$$

Taking the limit as $\tau \uparrow s$ we obtain, for a.a. s,

$$Z_s = -\int_s^T D_sZ_\sigma\,dW_\sigma - \int_s^T f(s, \sigma, D_sY_\tau, D_sZ_\tau)\,d\sigma + \nabla\phi(X_T)D_sX_T.$$

This completes the proof of Proposition 5.4.   □

The following result relates the Malliavin derivatives DY, DZ with the partial Gâteaux derivatives ∇_xY, ∇_xZ introduced in Proposition 5.2.

PROPOSITION 5.6.   *Assume Hypotheses* 3.1, 4.6 *and* 5.1. *Then for a.a.* s, τ *such that* $t \le s \le \tau \le T$ *we have*

(5.16) $D_sY(\tau, t, x) = \nabla_xY(\tau, s, X(s, t, x))G(s, X(s, t, x))$, $\mathbb{P}$-*a.s.*,

(5.17) $D_sZ(\tau, t, x) = \nabla_xZ(\tau, s, X(s, t, x))G(s, X(s, t, x))$, $\mathbb{P}$-*a.s.*

Moreover, for a.a. $s \in [t, T]$,

(5.18) $Z(s, t, x) = \nabla_xY(s, s, X(s, t, x))G(s, X(s, t, x))$, $\mathbb{P}$-*a.s.*

PROOF.   Proposition 5.2 states that for every $s \in [0, T]$ and every direction $h \in H$ the directional derivative process $(\nabla_xY(\tau, s, x)h, \nabla_xZ(\tau, s, x)h)$, $\tau \in [s, T]$ solves the equation, $\mathbb{P}$-a.s.,

$$\nabla_xY(\tau, s, x)h + \int_\tau^T \nabla_xZ(\sigma, s, x)h\,dW_\sigma$$

$$= -\int_\tau^T \nabla_x\psi(\sigma, X(\sigma, s, x), Y(\sigma, s, x), Z(\sigma, s, x))\nabla_xX(\sigma, s, x)h\,d\sigma$$

$$\qquad -\int_\tau^T \nabla_y\psi(\sigma, X(\sigma, s, x), Y(\sigma, s, x), Z(\sigma, s, x))\nabla_xY(\sigma, s, x)h\,d\sigma$$

$$-\int_\tau^T \nabla_z \psi\big(\sigma, X(\sigma, s, x), Y(\sigma, s, x), Z(\sigma, s, x)\big)\nabla_x Z(\sigma, s, x)h\, d\sigma$$
$$+\nabla\phi(X(T, s, x))\nabla_x X(T, s, x)h, \qquad \tau \in [s, T].$$

Given $v \in \Xi$ and $t \in [0, s]$, we can replace x by $X(s, t, x)$ and h by $G(s, X(s, t, x))v$ in this equation, since $X(s, t, x)$ is $\mathcal{F}_s$-measurable. Recalling Proposition 3.8 and taking into account the equalities (5.2) and (5.3) we obtain, $\mathbb{P}$-a.s.,

$$\nabla_x Y\big(\tau, s, X(s, t, x)\big)G\big(s, X(s, t, x)\big)v$$
$$+\int_\tau^T \nabla_x Z\big(\sigma, s, X(s, t, x)\big)G\big(s, X(s, t, x)\big)v\, dW_\sigma$$
$$=-\int_\tau^T \nabla_x \psi\big(\sigma, X(\sigma, t, x), Y(\sigma, t, x), Z(\sigma, t, x)\big)D_s X(\sigma, t, x)v\, d\sigma$$
$$-\int_\tau^T \nabla_y \psi\big(\sigma, X(\sigma, t, x), Y(\sigma, t, x), Z(\sigma, t, x)\big)\nabla_x Y\big(\sigma, s, X(s, t, x)\big)$$
$$\times G\big(s, X(s, t, x)\big)v\, d\sigma$$
$$-\int_\tau^T \nabla_z \psi\big(\sigma, X(\sigma, t, x), Y(\sigma, t, x), Z(\sigma, t, x)\big)\nabla_x Z\big(\sigma, s, X(s, t, x)\big)$$
$$\times G\big(s, X(s, t, x)\big)v\, d\sigma$$
$$+\nabla\phi\big(X(T, t, x)\big)\nabla_x X\big(T, s, X(s, t, x)\big)G\big(s, X(s, t, x)\big)v, \qquad \tau \in [s, T].$$

Comparing with (5.7) we conclude that the pairs $(D_s Y(\tau, t, x)v, D_s Z(\tau, t, x)v)$ and

$$\big(\nabla_x Y\big(\tau, s, X(s, t, x)\big)G\big(s, X(s, t, x)\big)v, \nabla_x Z\big(\tau, s, X(s, t, x)\big)G\big(s, X(s, t, x)\big)v\big)$$

solve the same backward equation for $\tau \in [s, T]$. Then the assertion follows from the uniqueness property (compare Remark 4.4).

Finally, equality (5.18) follows immediately from (5.16) and (5.8). $\quad\square$

PROPOSITION 5.7. *Assume Hypotheses* 3.1, 4.6 *and* 5.1. *Then the function* $u(t, x) = Y(t, t, x)$ *has the following properties*:

  (i) $u \in \mathcal{G}^{0,1}([0, T] \times H, \mathbb{R})$;
  (ii) *there exists* $C > 0$ *such that* $|\nabla_x u(t, x)h| \leq C|h|(1 + |x|^{(m+1)^2})$ *for all* $t \in [0, T], x \in H, h \in H$;

(iii) *if, in addition,*

$$\sup_{\sigma \in [0,T], x \in H} |\psi(\sigma, x, 0, 0)| < \infty, \qquad \sup_{x \in H} |\phi(x)| < \infty,$$

then $\sup_{t \in [0,T], x \in H} |u(t,x)| < \infty$;

(iv) *similarly if Hypothesis* 4.6(iii) *is replaced by* (iii-bis) *in Remark* 4.9 *then* $|\nabla_x u(t,x)h| \le c|h|$ *for a suitable constant c and all* $x, h \in H, t \in [0,T]$.

PROOF.   (i) Since $Y(t,t,x)$ is deterministic, we have $u(t,x) = \mathbb{E} Y(t,t,x)$. So the map $(t,x) \to u(t,x)$ can be written as a composition letting $u(t,x) = \Gamma_3(\Gamma_2(t, \Gamma_1(t,x)))$ with

$$\Gamma_1 \colon [0,T] \times H \to L^p_{\mathcal{P}}(\Omega; C([0,T]; \mathbb{R})), \qquad \Gamma_1(t,x) = Y(\cdot, t, x),$$

$$\Gamma_2 \colon [0,T] \times L^p_{\mathcal{P}}(\Omega; C([0,T]; \mathbb{R})) \to L^p(\Omega; \mathbb{R}), \qquad \Gamma_2(t,U) = U(t),$$

$$\Gamma_3 \colon L^p(\Omega; \mathbb{R}) \to \mathbb{R}, \qquad \Gamma_3 \zeta = \mathbb{E}\zeta.$$

By Proposition 5.2, $\Gamma_1 \in \mathcal{G}^{0,1}$. The inequality

$$|U(t) - V(s)|_{L^p(\Omega;\mathbb{R})} \le |U(t) - U(s)|_{L^p(\Omega;\mathbb{R})} + |U - V|_{L^p_{\mathcal{P}}(\Omega; C([0,T];\mathbb{R}))}$$

shows that Γ_2 is continuous; moreover Γ_2 is clearly linear in the second variable. Finally Γ_3 is a bounded linear operator. Then the assertion follows from the chain rule.

(ii) This is an immediate consequence on the estimate in Proposition 5.2: indeed,

$$|u(t,x)|^2 = |Y(t,t,x)|^2 = \mathbb{E}|Y(t,t,x)|^2 \le \sup_{\tau \in [t,T]} \mathbb{E}|Y(\tau,t,x)|^2.$$

(iii) Since (Y,Z) is a solution of the backward equation, the estimate in Proposition 4.3 yields

$$\sup_{\tau \in [t,T]} \mathbb{E}|Y(\tau,t,x)|^2 \le c\, \mathbb{E} \int_0^T |\psi(\sigma, X(\sigma,t,x), 0, 0)|^2 \, d\sigma$$

$$+ c\, \mathbb{E} |\phi(X(\sigma,t,x))|^2 \le c.$$

(iv) This follows immediately by Remarks 4.9 and 5.3.   □

6. Mild solutions of the Kolmogorov nonlinear equation.   We denote by $\mathcal{B}_p(H)$ the set of measurable functions $\phi \colon H \to \mathbb{R}$ with polynomial growth, that is, such that $\sup_{x \in H} |\phi(x)|(1+|x|^a)^{-1} < \infty$ for some $a > 0$.

Let $X(\tau,t,x)$, $\tau \in [t,T]$, denote the solution of the stochastic equation

$$X_\tau = e^{(\tau-t)A} x + \int_t^\tau e^{(\tau-\sigma)A} F(\sigma, X_\sigma)\, d\sigma + \int_t^\tau e^{(\tau-\sigma)A} G(\sigma, X_\sigma)\, dW_\sigma,$$

where A, F, G, satisfy the assumptions in Hypothesis 3.1. The transition semigroup $P_{t,\tau}$ is defined for arbitrary $\phi \in \mathcal{B}_p(H)$ by the formula

$$P_{t,\tau}[\phi](x) = \mathbb{E}\,\phi\big(X(\tau, t, x)\big), \qquad x \in H.$$

The estimate $\mathbb{E}\sup_{\tau \in [t,T]} |X_\tau|^p \leq C(1 + |x|)^p$ [see (3.6)] shows that $P_{t,\tau}$ is well defined as a linear operator $\mathcal{B}_p(H) \to \mathcal{B}_p(H)$; the semigroup property $P_{t,s} P_{s,\tau} = P_{t,\tau}$, $t \leq s \leq \tau$, is well known.

Let us denote by $\mathcal{L}_t$ the generator of $P_{t,\tau}$:

$$\mathcal{L}_t[\phi](x) = \tfrac{1}{2}\,\text{Trace}\big(G(t, x)G(t, x)^* \nabla^2 \phi(x)\big) + \langle Ax + F(t, x), \nabla\phi(x)\rangle,$$

where $\nabla\phi$ and $\nabla^2\phi$ are first and second Gâteaux derivatives of ϕ [identified with elements of H and $L(H)$ respectively]. This definition is formal, since the domain of $\mathcal{L}_t$ is not specified; however, if ϕ is sufficiently regular, the function $v(t, x) = P_{t,T}[\phi](x)$, is a classical solution of the backward Kolmogorov equation:

$$\begin{cases} \dfrac{\partial v(t, x)}{\partial t} + \mathcal{L}_t[v(t, \cdot)](x) = 0, & t \in [0, T],\ x \in H, \\ v(T, x) = \phi(x). \end{cases}$$

We refer to [9] and [41] for a detailed exposition. When ϕ is not regular, the function $v(t, x) = P_{t,T}[\phi](x)$ can be considered as a generalized solution of the backward Kolmogorov equation.

Here we are interested in a generalization of this equation, written formally as

(6.1)
$$\begin{cases} \dfrac{\partial u(t, x)}{\partial t} + \mathcal{L}_t[u(t, \cdot)](x) = \psi\big(t, x, u(t, x), G(t, x)^* \nabla_x u(t, x)\big), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t \in [0, T],\ x \in H, \\ u(T, x) = \phi(x). \end{cases}$$

We will refer to this equation as the nonlinear Kolmogorov equation. $\psi \colon [0, T] \times H \times \mathbb{R} \times \Xi \to \mathbb{R}$ is a given function verifying (4.6) and $\nabla_x u(t, x)$ is the Gâteaux derivative of $u(t, x)$ with respect to x: it is identified with an element of H, so that $G(t, x)^* \nabla_x u(t, x) \in \Xi$.

Now we define the notion of solution of the nonlinear Kolmogorov equation. We consider the variation of constants formula for (6.1):

(6.2)
$$\begin{aligned} u(t, x) = &-\int_t^T P_{t,\tau}\big[\psi\big(\tau, \cdot, u(\tau, \cdot), G(\tau, \cdot)^* \nabla_x u(\tau, \cdot)\big)\big](x)\, d\tau \\ &+ P_{t,T}[\phi](x), \qquad t \in [0, T],\ x \in H, \end{aligned}$$

and we notice that formula (6.2) is meaningful, provided $\psi(t, \cdot, \cdot, \cdot)$, $u(t, \cdot)$ and $\nabla_x u(t, \cdot)$ have polynomial growth (and, of course, provided they satisfy appropriate measurability assumptions). We use this formula as a definition for the solution of (6.1):

DEFINITION 6.1.   We say that a function $u \colon [0, T] \times H \to \mathbb{R}$ is a mild solution of the non linear Kolmogorov equation (6.1) if the following conditions hold:

(i) $u \in \mathcal{G}^{0,1}([0, T] \times H, \mathbb{R})$;

(ii) there exists $C > 0$ and $d \in \mathbb{N}$ such that $|\nabla_x u(t, x)h| \leq C|h|(1 + |x|^d)$ for all $t \in [0, T]$, $x \in H$, $h \in H$;

(iii) equality (6.2) holds.

REMARK 6.1.   An equivalent formulation of (6.1) or (6.2) would be the following: we consider the Gâteaux derivative $\nabla_x u(t, x)$ as an element of $\Xi^* = L(\Xi, \mathbb{R}) = L_2(\Xi, \mathbb{R})$, we take a function $\psi : [0, T] \times H \times \mathbb{R} \times L_2(\Xi, \mathbb{R}) \to \mathbb{R}$ and we write the equation in the form

$$\frac{\partial u(t, x)}{\partial t} + \mathcal{L}_t[u(t, \cdot)](x) = \psi\big(t, x, u(t, x), \nabla_x u(t, x)G(t, x)\big).$$

The two forms are clearly equivalent provided we identify $\Xi^* = L_2(\Xi, \mathbb{R})$ with Ξ by the Riesz isometry. This will be done in the sequel. In particular, although we keep the notation in (6.1), we will sometimes consider ψ as a real valued function defined on $[0, T] \times H \times \mathbb{R} \times L_2(\Xi, \mathbb{R})$, satisfying Hypothesis 4.6 with $K = \mathbb{R}$.

We are now ready to state the main result of this paper.

THEOREM 6.2.   *Assume that Hypothesis* 3.1 *holds*, *and let* ψ, ϕ *be functions satisfying the assumptions in Hypotheses* 4.6 (*with* $K = \mathbb{R}$) *and* 5.1. *Then there exists a unique mild solution of the nonlinear Kolmogorov equation* (6.1).

The solution u is given by the formula

$$(6.3) \qquad\qquad\qquad u(t, x) = Y(t, t, x),$$

where (X, Y, Z) *is the solution of the backward–forward system* (5.1).

If, *in addition*, $\sup_{t \in [0,T], x \in H} |\psi(t, x, 0, 0)| < \infty$ *and* ϕ *is bounded then u is also bounded*.

Similarly if $|\nabla_x \psi|$ *is bounded then* $|\nabla_x u|$ *is also bounded*.

PROOF (Existence).   By Proposition 5.7, the solution u has the regularity properties stated in Definition 6.1 and the last two statements of the claim hold. It remains to verify that equality (6.2) holds. To this end we first fix $t \in [0, T]$ and $x \in H$. We note that

$$\psi\big(\tau, \cdot, u(\tau, \cdot), G(\tau, \cdot)^* \nabla_x u(\tau, \cdot)\big) = \psi\big(\tau, \cdot, Y(\tau, \tau, \cdot), G(\tau, \cdot)^* \nabla_x Y(\tau, \tau, \cdot)\big)$$

so that

$$P_{t,\tau}\big[\psi\big(\tau, \cdot, u(\tau, \cdot), G(\tau, \cdot)^* \nabla_x u(\tau, \cdot)\big)\big](x)$$
$$= \mathbb{E}\psi\big(\tau, X(\tau, t, x), Y\big(\tau, \tau, X(\tau, t, x)\big), G\big(\tau, X(\tau, t, x)\big)^*$$
$$\times \nabla_x Y\big(\tau, \tau, X(\tau, t, x)\big)\big).$$

We recall the identity, $\mathbb{P}$-a.s., $Y(\tau, \tau, X(\tau, t, x)) = Y(\tau, t, x)$ for $\tau \in [t, T]$ [compare (5.3)], and equality (5.18), which can be written in the present notation as

$$Z(\tau, t, x) = G(\tau, X(\tau, t, x))^* \nabla_x Y(\tau, \tau, X(\tau, t, x)),$$

(6.4)

$$\mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T].$$

It follows that

(6.5)
$$\int_t^T P_{t,\tau} [\psi(\tau, \cdot, u(\tau, \cdot), G(\tau, \cdot)^* \nabla_x u(\tau, \cdot))](x) \, d\tau$$

$$= \mathbb{E} \int_t^T \psi(\tau, X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x)) \, d\tau.$$

The backward equation of system (5.1) for $\tau = t$ is

$$Y(t, t, x) + \int_t^T Z_\tau dW_\tau$$

$$= -\int_t^T \psi(\tau, X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x)) \, d\tau + \phi(X(T, t, x)).$$

Taking expectation we obtain

$$u(t, x) = -\mathbb{E} \int_t^T \psi(\tau, X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x)) \, d\tau + P_{t,T}[\phi](x).$$

Comparing with (6.5) gives the required equality (6.2).

(Uniqueness.)   Let u be a mild solution. We look for a convenient expression for the process $u(s, X(s, t, x))$, $s \in [t, T]$. By (6.2) and the definition of $P_{t,\tau}$, for every $s \in [t, T]$ and $x \in H$,

$$u(s, x) = \mathbb{E}[\phi(X(T, s, x))]$$

$$- \mathbb{E}\left[\int_s^T \psi(\tau, X(\tau, s, x), u(\tau, X(\tau, s, x)),\right.$$

$$\left. G(\tau, X(\tau, s, x))^* \nabla_x u(\tau, X(\tau, s, x))) \, d\tau\right].$$

Since $X(\tau, s, x)$ is $\mathcal{F}_s$-measurable, we can replace the expectation by the conditional expectation given $\mathcal{F}_s$:

$$u(s, x) = \mathbb{E}^{\mathcal{F}_s}[\phi(X(T, s, x))] - \mathbb{E}^{\mathcal{F}_s}\left[\int_s^T \psi(\tau, X(\tau, s, x), u(\tau, X(\tau, s, x)),\right.$$

$$\left. G(\tau, X(\tau, s, x))^* \nabla_x u(\tau, X(\tau, s, x))) \, d\tau\right].$$

For the same reason, we can replace x by $X(s, t, x)$ and use the equality

$$X(\tau, s, X(s, t, x)) = X(\tau, t, x), \qquad \mathbb{P}\text{-a.s. for } \tau \in [s, T].$$

We arrive at

$$u\big(s, X(s, t, x)\big)$$

$$= \mathbb{E}^{\mathcal{F}_s}\big[\phi\big(X(T, t, x)\big)\big] - \mathbb{E}^{\mathcal{F}_s}\bigg[\int_s^T \psi\big(\tau, X(\tau, t, x), u\big(\tau, X(\tau, t, x)\big),$$

$$G\big(\tau, X(\tau, t, x)\big)^* \nabla_x u\big(\tau, X(\tau, t, x)\big)\big) \, d\tau\bigg]$$

$$= \mathbb{E}^{\mathcal{F}_s}[\xi] + \int_t^s \psi\big(\tau, X(\tau, t, x), u\big(\tau, X(\tau, t, x)\big),$$

$$G\big(\tau, X(\tau, t, x)\big)^* \nabla_x u\big(\tau, X(\tau, t, x)\big)\big) \, d\tau,$$

where we have defined

$$\xi = \phi\big(X(T, t, x)\big) - \int_t^T \psi\big(\tau, X(\tau, t, x), u\big(\tau, X(\tau, t, x)\big), G\big(\tau, X(\tau, t, x)\big)^*$$

$$\times \nabla_x u\big(\tau, X(\tau, t, x)\big)\big) \, d\tau.$$

We note that $\mathbb{E}^{\mathcal{F}_t}[\xi] = \mathbb{E}^{\mathcal{F}_t} u(s, X(s, t, x)) = u(t, x)$. Since $\xi \in L^2(\Omega; \mathbb{R})$ is $\mathcal{F}_T$-measurable, by the representation theorem recalled in Proposition 4.1, there exists $\widetilde{Z} \in L^2_{\mathcal{P}}(\Omega \times [t, T]; L_2(\Xi, \mathbb{R}))$ such that $\mathbb{E}^{\mathcal{F}_s}[\xi] = \int_t^s \widetilde{Z}_\tau \, dW_\tau + u(t, x)$. We conclude that the process $u(s, X(s, t, x))$, $s \in [t, T]$ is a (real) continuous semimartingale with canonical decomposition

$$\tag{6.6}
\begin{aligned}
u(s, X(s, t, x)) = &\int_t^s \widetilde{Z}_\tau \, dW_\tau + u(t, x) \\
&+ \int_t^s \psi\big(\tau, X(\tau, t, x), u\big(\tau, X(\tau, t, x)\big), \\
&\qquad G\big(\tau, X(\tau, t, x)\big)^* \nabla_x u\big(\tau, X(\tau, t, x)\big)\big) \, d\tau,
\end{aligned}$$

into its continuous martingale part and continuous finite variation part. Let $\{e_i\}$ denote a basis of the space Ξ and consider the standard real Wiener process $W_\tau^i = \int_0^\tau \langle e_i, dW_\sigma \rangle$, $\tau \geq 0$. Now we need the following lemma:

LEMMA 6.3. *For every i, the joint quadratic variation process of $u(s, X(s, t, x))$ and W_s^i, $s \in [t, T]$, is*

$$\tag{6.7}
\int_t^s \nabla_x u\big(\tau, X(\tau, t, x)\big) G\big(\tau, X(\tau, t, x)\big) e_i \, d\tau, \qquad s \in [t, T].$$

We assume the lemma for the moment. Taking into account the canonical decomposition (6.6), we note that the process (6.7) can also be obtained as the joint quadratic variation process between W_s^i, $s \in [t, T]$, and the process $\int_t^s \widetilde{Z}_\tau \, dW_\tau$. This yields the identity

$$\int_t^s \nabla_x u(\tau, X(\tau, t, x)) G(\tau, X(\tau, t, x)) e_i \, d\tau = \int_t^s \langle \widetilde{Z}_\tau, e_i \rangle \, d\tau, \qquad s \in [t, T].$$

Therefore, for a.a. $\tau \in [t, T]$, we have, $\mathbb{P}$-a.s.,

$$\nabla_x u(\tau, X(\tau, t, x)) G(\tau, X(\tau, t, x)) e_i = \langle \widetilde{Z}_\tau, e_i \rangle,$$

for every i. Identifying $\nabla_x u(t, x)$ with an element of Ξ, we conclude that for a.a. $\tau \in [t, T]$,

$$G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x)) = \widetilde{Z}_\tau.$$

Substituting into (6.6) we obtain

$$
\begin{aligned}
u(s, X(s, t, x)) = &\int_t^s G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x)) \, dW_\tau + u(t, x) \\
& + \int_t^s \psi(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \\
& \qquad G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x))) \, d\tau,
\end{aligned}
$$

(6.8)

for $s \in [t, T]$. Since $u(T, X(T, t, x)) = \phi(X(T, t, x))$, we also have

$$
\begin{aligned}
u(s, X(s, t, x)) &+ \int_s^T G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x)) \, dW_\tau \\
&= \phi(X(T, t, x)) - \int_s^T \psi(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \\
& \qquad G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x))) \, d\tau,
\end{aligned}
$$

for $s \in [t, T]$. Comparing with the backward equation in (5.1) we note that the pairs

$$(Y(s, t, x), Z(s, t, x))$$

and

$$(u(s, X(s, t, x)), G(s, X(s, t, x))^* \nabla_x u(s, X(s, t, x))), \qquad s \in [t, T],$$

solve the same equation. By uniqueness, we have in particular $Y(s, t, x) = u(s, X(s, t, x))$, $s \in [t, T]$. Setting $s = t$ we obtain $Y(t, t, x) = u(t, x)$. $\quad\square$

It remains to prove Lemma 6.3. We need to introduce the following class of processes.

DEFINITION 6.2. Let $\{e_i\}$ be a basis of Ξ. We say that a process q belongs to the class $\mathbb{L}_c$ if $q \in C_{\mathcal{P}}([0, T]; L^2(\Omega, \mathbb{R})) \cap \mathbb{L}^{1,2}(\mathbb{R})$ and for every i there exists a version of Dq such that the map $(\tau, s) \mapsto D_s q_\tau e_i$, defined for $t \le s \le \tau \le T$, is continuous (hence uniformly continuous and bounded) with values in $L^2(\Omega, \mathbb{R})$.

For $q \in \mathbb{L}_c$, we define

$$D_s^+ q_s e_i = \lim_{\tau \downarrow s} D_s q_\tau e_i,$$

in the norm of $L^2(\Omega; \mathbb{R})$.

This definition is modeled after [33], Definition 7.2. Note that we need only consider adapted processes, so that in particular $D_s q_\tau = 0$, $\mathbb{P}$-a.s. for a.a. $s > \tau$.

The following result can be proved as an easy adaptation of Theorem 7.6 in [33], taking into account the extensions to the case of infinite-dimensional Wiener process W in [20], Définition 5.2 and Théorème 5.3, so we omit the proof. Recall that we set $W_\tau^i = \int_0^\tau \langle e_i, dW_\sigma \rangle$, $\tau \geq 0$.

LEMMA 6.4. *If $q \in \mathbb{L}_c$ then, for every i, the joint quadratic variation process of q_τ and W_τ^i, $\tau \in [0, T]$, is*

(6.9)
$$\int_0^\tau D_s^+ q_s e_i \, ds, \qquad \tau \in [0, T].$$

Now we can prove Lemma 6.3. We define $q_\tau = u(\tau, X(\tau, t, x))$ for $\tau \in [t, T]$ and $q_\tau = u(t, x)$ for $\tau \in [0, t)$. Since $X \in L_\mathcal{P}^r(\Omega; C([0, T]; H))$ for every $r \geq 2$, u is continuous and $u(t, \cdot)$ has polynomial growth (uniformly in t), it is easily proved that $q \in C_\mathcal{P}([0, T]; L^2(\Omega; \mathbb{R}))$. Next, since $u \in \mathcal{G}^{0,1}([0, T] \times H, \mathbb{R})$, by the chain rule stated in Lemma 3.4(i) we have $q_\tau \in \mathbb{D}_{\text{loc}}^{1,2}(\mathbb{R})$ for a.a. $\tau \in [t, T]$ and

$$D_s q_\tau = \nabla_x u(\tau, X(\tau, t, x)) D_s X(\tau, t, x), \qquad \mathbb{P}\text{-a.s. for a.a. } s.$$

Next we have, for some constant c,

$$\int_0^T |D_s q_\tau|_{L_2(\Xi, \mathbb{R})}^2 \, ds$$

$$= \int_t^\tau \left| \nabla_x u(\tau, X(\tau, t, x)) D_s X(\tau, t, x) \right|_{L_2(\Xi, \mathbb{R})}^2 \, ds$$

$$\leq c \left(1 + |X(\tau, t, x)|^{2d}\right) \int_t^\tau |D_s X(\tau, t, x)|_{L_2(\Xi, H)}^2 \, ds$$

$$\leq c \left(1 + |X(\tau, t, x)|^{2d}\right)$$

$$\times \int_t^\tau (\tau - s)^{-2\gamma} \left[\sup_{\sigma \in [s, T]} (\sigma - s)^{2\gamma} |D_s X(\sigma, t, x)|_{L_2(\Xi, H)}^2 \right] ds$$

$$\leq c \left(1 + |X(\tau, t, x)|^{2d}\right) \left(\int_t^\tau (\tau - s)^{-2\gamma p} \right)^{1/p}$$

$$\times \left(\int_t^\tau \left[\sup_{\sigma \in [s, T]} (\sigma - s)^{2\gamma p'} |D_s X(\sigma, t, x)|_{L_2(\Xi, H)}^{2p'} \right] ds \right)^{1/p'}$$

where $p > 1$ is so small that $2\gamma p < 1$ and $p' := p/(p-1)$. Taking expectation and using the Hölder inequality we obtain

$$\mathbb{E} \int_0^T |D_s q_\tau|^2 \, ds \leq c \left(1 + |X(\cdot, t, x)|^{2d}_{L_{\mathcal{P}}^{2dp}(\Omega; C([0,T]; H))} \right)$$

$$\times \left(\int_t^\tau \mathbb{E} \left[\sup_{\sigma \in [s,T]} (\sigma - s)^{2\gamma p'} |D_s X(\sigma, t, x)|^{2p'}_{L_2(\Xi, H)} \right] ds \right)^{1/p'}.$$

From (3.11) it follows that $\mathbb{E} \int_0^T |D_s q_\tau|^2 \, ds \leq c$ for some constant c independent of τ; by Lemma 3.4(iii) we have $q_\tau \in \mathbb{D}^{1,2}(\mathbb{R})$ and we even conclude that $q \in \mathbb{L}^{1,2}(\mathbb{R})$. By Proposition 3.5(iv), the map $(s, \tau) \mapsto D_s X_\tau e_i$, defined for $t \leq s \leq \tau \leq T$, is continuous (hence uniformly continuous and bounded) with values in $L^p(\Omega; H)$, for every $p \geq 2$. Using the strong continuity and polynomial growth of $\nabla_x u$ it is easy to conclude that Dqe_i has the properties stated in Definition 6.2 and taking the limit as $\tau \downarrow s$,

$$D_s^+ q_s e_i = \nabla_x u\big(s, X(s, t, x)\big) G\big(s, X(s, t, x)\big) e_i,$$

for a.a. $s \in [t, T]$. Clearly, $D_s^+ q_s e_i = 0$ for a.a. $s \in [0, t]$, by the definition of q. Now the equality (6.7) follows from an application of Lemma 6.4.

7. Application to optimal control. We wish to apply the previous results to perform the synthesis of the optimal control for a general nonlinear control system. We will show (see Example 7.3.1) that the generality of the model that our approach allows (particularly in the direction of the degeneracy of the noise) is essential to treat models of great importance in mathematical finance.

To be able to use nonsmooth feedbacks we settle the problem in the framework of weak control problems (see [12]).

Again H, Ξ, U denote Hilbert spaces. Fixed $t_0 \geq 0$ and $x_0 \in H$ an *admissible control system* (a.c.s.) is given by $(\Omega, \mathcal{E}, \mathcal{F}_t, \mathbb{P}, W_t, u)$ where:

- $(\Omega, \mathcal{E}, \mathbb{P})$ is a probability space,
- $\{\mathcal{F}_t : t \geq 0\}$ is a filtration in it, satisfying the usual conditions,
- $\{W_t : t \geq 0\}$ is a cylindrical $\mathbb{P}$-Wiener process with values in Ξ and adapted to the filtration $\{\mathcal{F}_t\}$,
- $u \in L_{\mathcal{P}}^2(\Omega \times [t_0, T]; U)$ satisfies the constraint: $u(t) \in \mathcal{U}$, $\mathbb{P}$-a.s. for a.a. $t \in [t_0, T]$, where $\mathcal{U}$ is a fixed bounded subset of U.

To each a.c.s. we associate the mild solution $X^u \in C_{\mathcal{P}}([t_0, T]; L^2(\Omega; H))$ of the state equation

$$(7.1) \qquad \begin{cases} dX_\tau^u = \big(AX_\tau^u + F(\tau, X_\tau^u) + G(\tau, X_\tau^u) R(\tau, X_\tau^u) u(\tau)\big) \, d\tau \\ \qquad\quad + G(\tau, X_\tau) \, dW_\tau, \qquad \tau \in [t_0, T], \\ X_{t_0} = x_0 \in H, \end{cases}$$

and the cost

$$(7.2) \qquad J(t_0, x_0, u) = \mathbb{E} \int_{t_0}^{T} g\big(\sigma, X_\sigma^u, u(\sigma)\big) \, d\sigma + \mathbb{E} \, \phi(X_T^u).$$

Our purpose is to minimize the functional J over all a.c.s. Notice the occurrence of the operator G in the control term: this special structure of the state equation is imposed by our techniques and seems to be essential in different contexts as well (see [15]). On the contrary the presence of the operator R allows more generality.

We define in a classical way the Hamiltonian function relative to the above problem: for all $t \in [0, T]$, $x \in H$, $p \in U$,

$$(7.3) \qquad \psi_0(t, x, p) = \inf\{g(t, x, u) + \langle p, u \rangle : u \in \mathcal{U}\}.$$

We make the following assumption.

HYPOTHESIS 7.1. *The following hold*:

1. *A, F and G verify Hypothesis* 3.1.
2. R: $[0, T] \times H \to L(U, \Xi)$ *enjoys the following*: R^*z *is measurable* $[0, T] \times H \to U$ *for every* $z \in \Xi$; $R^*(t, \cdot)z$ *belongs to* $\mathcal{G}^1(H, U)$ *for every* $t \in [0, T]$, $z \in \Xi$; *finally* $|R(t, x)|_{L(\Xi, U)} \leq L$ *and* $|\nabla_x (R(t, x)^*z)h| \leq L |z| |h|$ *for a suitable constant* $L > 0$ *and all* $t \in [0, T]$, $z \in \Xi$, $x, h \in H$.
3. $g \in C([0, T] \times H \times U; \mathbb{R})$ *with* $|g(\sigma, x, u)| \leq L(1 + |x|^m)$ *for suitable constants* $L > 0$, $m \geq 0$ *and all* $t \in [0, T]$, $x \in H$, $u \in \mathcal{U}$.
4. ϕ *satisfies Hypothesis* 5.1.
5. ψ_0: $[0, T] \times H \times U \to \mathbb{R}$ *is a measurable map*; *moreover for every* $t \in [0, T]$, $\psi_0(t, \cdot, \cdot)$ *belongs to* $\mathcal{G}^{0,1,1}([0, T] \times H \times U, \mathbb{R})$ *with* $|\psi_0(t, 0, 0)| \leq L$, $|\nabla_p \psi_0(t, x, p)v| \leq L|v|$, $|\nabla_x \psi_0(t, x, p)h| \leq L|h|(1 + |p|)(1 + |x|^m)$ *for suitable constants* $L > 0$, $m \geq 0$ *and all* $t \in [0, T]$, $x, h \in H$ *and* $p, v \in U$.
6. *For all* $t \in [0, T]$, $x \in H$ *and* $p \in U$ *there exists a unique* $\Gamma(t, x, p) \in \mathcal{U}$ *that realizes the minimum in* (7.3). *Namely*

$$g\big(t, x, \Gamma(t, x, p)\big) + \langle p, \Gamma(t, x, p) \rangle = \psi_0(t, x, p).$$

Moreover, $\Gamma \in C([0, T] \times H \times U; \mathcal{U})$.

Finally, we define

$$\psi(t, x, z) = -\psi_0\big(t, x, R(t, x)^*z\big), \qquad t \in [0, T], x \in H, z \in \Xi.$$

EXAMPLE 7.1.1. If $\mathcal{U}$ is the ball $\{v \in U : |v| \leq r\}$ for some fixed $r > 0$, and $g(t, x, u) = g_0(|u|^\alpha) + g_1(t, x)$ with $g_0 \in C^1(\mathbb{R}^+; \mathbb{R}^+)$ convex, $g_0'(0) > 0$, $\alpha > 1$, $g_1 \in \mathcal{G}^{0,1}([0, T] \times H, \mathbb{R})$ with $|g_1(t, 0)| \leq L$ and $|\nabla_x g_1(t, x)h| \leq L|h|(1 + |x|^m)$ for suitable constants $L > 0$, $m \geq 0$ and all $t \in [0, T]$, $x, h \in H$ then Hypothesis 7.1 holds. Moreover $\psi_0(t, x, p)$ is Fréchet differentiable with respect to p and $\Gamma(t, x, p) = \nabla_p \psi_0(t, x, p)$ turns out to be a function of p only.

Notice that, identifying Ξ with $L_2(\Xi, \mathbb{R})$, the function $\psi \colon [0, T] \times H \times L_2(\Xi, \mathbb{R}) \to \mathbb{R}$ verifies Hypothesis 4.6 with $K = \mathbb{R}$. Therefore by Theorem 6.2 the Hamilton–Jacobi–Bellman equation relative to the above stated problem, namely,

(7.4)
$$\begin{cases} \dfrac{\partial v(t, x)}{\partial t} + \mathcal{L}_t[v(t, \cdot)](x) \\ \qquad = \psi\big(t, x, v(t, x), G(t, x)^* \nabla_x v(t, x)\big), \qquad t \in [0, T], \ x \in H, \\ v(T, x) = \phi(x), \end{cases}$$

admits a unique mild solution.

We are in a position to prove the main result of this section:

THEOREM 7.2. *Assume Hypothesis* 7.1. *For all a.c.s. we have* $J(t_0, x_0, u) \geq v(t_0, x_0)$ *and the equality holds if and only if the following feedback law is verified by* u *and* X^u:

(7.5)
$$\begin{aligned} u(\sigma) &= \Gamma(\sigma, X_\sigma^u, R(\sigma, X_\sigma^u)^* G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u)), \\ &\quad \mathbb{P}\text{-}a.s. \text{ for } a.a. \ \sigma \in [t_0, T]. \end{aligned}$$

Finally there exists at least an a.c.s. for which (7.5) *holds. In such a system the closed loop equation*

(7.6)
$$\begin{cases} d\overline{X}_\tau = A\overline{X}_\tau \, d\tau \\ \qquad + G(\tau, \overline{X}_\tau) R(\tau, \overline{X}_\tau) \Gamma\big(\tau, \overline{X}_\tau, R(\tau, \overline{X}_\tau)^* G(\tau, \overline{X}_\tau)^* \nabla_x v(\tau, \overline{X}_\tau)\big) \, d\tau \\ \qquad + F(\tau, \overline{X}_\tau) \, d\tau + G(\tau, \overline{X}_\tau) \, dW_\tau, \qquad \tau \in [t_0, T], \\ \overline{X}_{t_0} = x_0 \in H, \end{cases}$$

admits a solution and if $\overline{u}(\sigma) = \Gamma(\sigma, \overline{X}_\sigma, R(\sigma, \overline{X}_\sigma)^* G(\sigma, \overline{X}_\sigma)^* \nabla_x v(\sigma, \overline{X}_\sigma))$ *then the couple* $(\overline{u}, \overline{X})$ *is optimal for the control problem.*

PROOF. For all a.c.s., setting $u(s) = 0$ for $s < t_0$, the Girsanov theorem ensures that there exists a probability measure $\widetilde{\mathbb{P}}$ on Ω such that

$$\widetilde{W}_t := W_t + \int_0^t R(\sigma, X_\sigma^u) u(\sigma) \, d\sigma$$

is a $\widetilde{\mathbb{P}}$-wiener process (notice that u is bounded, since it takes values in $\mathcal{U}$).

Relatively to $\widetilde{W}$ equation (7.1) can be rewritten

(7.7)
$$\begin{cases} dX_\tau^u = AX_\tau^u \, d\tau + F(\tau, X_\tau^u) \, d\tau + G(\tau, X_\tau) \, d\widetilde{W}_\tau, \qquad \tau \in [t_0, T], \\ X_{t_0} = x_0 \in H. \end{cases}$$

The process X^u turns out to be adapted to the filtration $\{\widetilde{\mathcal{F}}_t\}$ generated by $\widetilde{W}$ and completed in the usual way. In the space $(\Omega, \mathcal{E}, \{\widetilde{\mathcal{F}}_t\}, \widetilde{\mathbb{P}})$, for arbitrary t, x, we can

consider the system of forward–backward equations

$$
(7.8) \quad
\begin{cases}
\widetilde{X}(\tau, t, x) = e^{(\tau - t)A} x + \int_t^\tau e^{(\tau - \sigma)A} F\big(\sigma, \widetilde{X}(\sigma, t, x)\big)\, d\sigma \\
\qquad\qquad + \int_t^\tau e^{(\tau - \sigma)A} G\big(\sigma, \widetilde{X}(\sigma, t, x)\big)\, d\widetilde{W}_\sigma, \\
\widetilde{Y}(\tau, t, x) + \int_\tau^T \widetilde{Z}(\sigma, t, x)\, d\widetilde{W}_\sigma \\
\qquad = -\int_\tau^T \psi\big(\sigma, \widetilde{X}(\sigma, t, x), \widetilde{Z}(\sigma, t, x)\big)\, d\sigma + \phi\big(\widetilde{X}(T, t, x)\big).
\end{cases}
$$

We notice that $\widetilde{X}(\sigma, t_0, x_0) = X_\sigma^u$. Writing the backward equation in (7.8) for $t = t_0$, $x = x_0$ and with respect to the original process W we get

$$
\begin{aligned}
\widetilde{Y}(\tau, t_0, x_0) &+ \int_\tau^T \widetilde{Z}(\sigma, t_0, x_0)\, dW_\sigma \\
(7.9) \quad &= -\int_\tau^T \big[\psi\big(\sigma, X_\sigma^u, \widetilde{Z}(\sigma, t_0, x_0)\big) + \widetilde{Z}(\sigma, t_0, x_0) R(\sigma, X_\sigma^u) u(\sigma)\big]\, d\sigma \\
&\quad + \phi(X_T^u).
\end{aligned}
$$

Now we identify $L_2(\Xi, \mathbb{R})$ with Ξ (and $\nabla_x v$ with an element of H) and we recall equalities (6.3) and (6.4) which can be written in the present notation as $\widetilde{Y}(t_0, t_0, x_0) = v(t_0, x_0)$,

$$
\widetilde{Z}(\tau, t_0, x_0) = G\big(\tau, \widetilde{X}(\tau, t_0, x_0)\big)^* \nabla_x v\big(\tau, \widetilde{X}(\tau, t_0, x_0)\big) = G(\tau, X_\tau^u)^* \nabla_x v(\tau, X_\tau^u).
$$

Taking expectation with respect to the original probability $\mathbb{P}$ in (7.9) and taking into account the definition of ψ we obtain

$$
\begin{aligned}
\mathbb{E}\,\phi(X_T^u) - v(t_0, x_0) &= -\mathbb{E} \int_{t_0}^T \psi_0\big(\sigma, X_\sigma^u, R(\sigma, X_\sigma^u)^* G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u)\big)\, d\sigma \\
&\quad + \mathbb{E} \int_{t_0}^T \langle R(\sigma, X_\sigma^u)^* G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u), u(\sigma) \rangle\, d\sigma.
\end{aligned}
$$

Adding and subtracting $\mathbb{E} \int_{t_0}^T g(\sigma, X_\sigma^u, u(\sigma))\, d\sigma$ we conclude

$$
\begin{aligned}
(7.10) \quad J(t_0, x_0, u) &= v(t_0, x_0) \\
&\quad + \mathbb{E} \int_{t_0}^T \big[-\psi_0\big(\sigma, X_\sigma^u, R(\sigma, X_\sigma^u)^* G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u)\big) \\
&\qquad\qquad + \langle R(\sigma, X_\sigma^u)^* G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u), u(\sigma) \rangle \\
&\qquad\qquad\qquad + g\big(\sigma, X_\sigma^u, u(\sigma)\big) \big]\, d\sigma.
\end{aligned}
$$

The above equality is known as the *fundamental relation* and immediately implies that $v(t_0, x_0) \leq J(t_0, x_0, u)$ and that the equality holds if and only if (7.5) holds.

Finally the existence of a weak solution to equation (7.6) is again a consequence of the Girsanov theorem. Namely let $X \in C_{\mathscr{P}}([t_0, T]; L^2(\Omega; H))$ be the mild solution of

$$\begin{cases} dX_\tau = AX_\tau \, d\tau + F(\tau, X_\tau) \, d\tau + G(\tau, X_\tau) \, dW_\tau, \\ X_{t_0} = x_0, \end{cases}$$

and let $\widehat{P}$ be the probability on Ω under which

$$\widehat{W}_t := -\int_0^t R(\sigma, X_\sigma)\Gamma(\sigma, X_\sigma, R(\sigma, X_\sigma)^*G(\sigma, X_\sigma)^*\nabla_x v(\sigma, X_\sigma)) \, d\sigma + W_t$$

is a Wiener process (notice that Γ is bounded, since it takes values in $\mathcal{U}$). Then X is the mild solution of equation (7.6) relatively to the probability $\widehat{P}$ and the Wiener process $\widehat{W}$. $\square$

REMARK 7.3. Assume Hypothesis 7.1 and, in addition, that the following hold:

(i) $|\nabla_x \psi(t, x, p)h| \le L|h|$ for a suitable constant L and all $t \in [0, T]$, $x, h \in H$ and $p \in U$;

(ii) $\sup_{t \in [0,T], x \in H} |G(t, x)|_{L(\Xi, H)} < \infty$;

(iii) $\Gamma(t, \cdot, \cdot)$: $H \times U \to U$ and $\nabla_x v(t, \cdot)$: $H \to H$ are globally Lipschitz, uniformly with respect to $t \in [0, T]$.

Notice that for (i) to hold it is sufficient that R is independent of x and $|\nabla_x \psi_0|$ is uniformly bounded. Moreover, by the last statement in Theorem 6.2, (i) implies that $|\nabla_x v|$ is uniformly bounded.

Now let $(\Omega, \mathcal{E}, \mathbb{P})$ be a fixed in advance probability space with a given filtration $\{\mathcal{F}_t: t \ge 0\}$ satisfying the usual conditions, and let W be a given cylindrical $\mathbb{P}$-Wiener process in Ξ, adapted to $\{\mathcal{F}_t\}$. Then equation (7.6) admits a unique mild solution $\overline{X} \in C_{\mathscr{P}}([t_0, T]; L^2(\Omega; H))$, since it has globally Lipschitz coefficients. Therefore the control

$$\overline{u}(\sigma) = \Gamma(\sigma, \overline{X}_\sigma, R(\sigma, \overline{X}_\sigma)^*G(\sigma, \overline{X}_\sigma)^*\nabla_x v(\sigma, \overline{X}_\sigma))$$

is the unique optimal control for the control problem in the strong formulation, namely $J(t_0, x_0, \overline{u}) = \inf J(t_0, x_0, u)$ where the infimum is taken over all $u \in L^2_{\mathscr{P}}(\Omega \times [t_0, T]; U)$ satisfying $u(t) \in \mathcal{U}$, $\mathbb{P}$-a.s. for a.a. $t \in [t_0, T]$.

In the following example we show that our results can be applied to a model of great interest in mathematical finance, where absence of nondegeneracy assumptions reveals to be essential. A similar problem is studied in [15] by analytic techniques; in that paper, however, weak nondegeneracy assumptions have still to be required.

EXAMPLE 7.3.1. We consider the so called Musiela parameterization of the Heath–Jarrow–Morton model for the forward rate curve f of a zero coupon bond: see [31]; here we follow [42], where an infinite-dimensional formulation is introduced. $f(t, \xi)$, $t \geq 0$, $\xi \geq 0$, is a real valued process satisfying

$$
\begin{cases}
d_t f(t, \xi) = \left(\dfrac{\partial f}{\partial \xi}(t, \xi) + \displaystyle\sum_{j=1}^{d} \sigma_j\big(\xi, f(t, \xi)\big) \int_0^\xi \sigma_j\big(\eta, f(t, \eta)\big) d\eta \right) dt \\
\qquad\qquad + \displaystyle\sum_{j=1}^{d} \sigma_j\big(\xi, f(t, \xi)\big) \, d\beta_j(t), \\
f(0, \xi) = x(\xi), \qquad x \in L^2_\rho(0, +\infty),
\end{cases}
$$

where $\rho > 0$ is a given parameter and $L^2_\rho(0, +\infty)$ is the space of measurable functions on $[0, +\infty)$ such that $\int_0^{+\infty} e^{-\rho\xi} x^2(\xi) \, d\xi < +\infty$, endowed with the natural norm. σ_j, $j = 1, \ldots, d$, are given functions $\mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$, continuous, bounded and differentiable in the second variable with uniformly bounded derivative. Finally, β_j, $j = 1, \ldots, d$, are independent real Wiener processes. In the above equation the state space $L^2_\rho(0, +\infty)$ is infinite dimensional while the noise is finite dimensional. Consequently the associated Kolmogorov and Hamilton–Jacobi–Bellman equations are highly degenerate.

By our general results we can treat for instance the control problem for the forward rate given by the state equation

$$
\begin{cases}
d_t y(t, \xi) = \left(\dfrac{\partial y}{\partial \xi}(t, \xi) + \displaystyle\sum_{j=1}^{d} \sigma_j\big(\xi, y(t, \xi)\big) u_j(t, \xi) \right. \\
\qquad\qquad \left. + \displaystyle\sum_{j=1}^{d} \sigma_j\big(\xi, y(t, \xi)\big) \int_0^\xi \sigma_j\big(\eta, y(t, \eta)\big) d\eta \right) dt \\
\qquad\qquad + \displaystyle\sum_{j=1}^{d} \sigma_j\big(\xi, y(t, \xi)\big) \, d\beta_j(t), \\
y(t, \xi) = x(\xi), \qquad x \in L^2_\rho(0, +\infty),
\end{cases}
$$

and the nonlinear cost

$$
\begin{aligned}
J(t, x, u) = {} & \mathbb{E} \int_t^T g_0\big(|u(s)|^\alpha_{\mathbb{R}^d}\big) \, ds \\
& + \mathbb{E} \int_t^T \int_0^{+\infty} e^{-\rho\xi} g_1\big(s, y(s, \xi)\big) \, d\xi \, ds \\
& + \mathbb{E} \int_0^{+\infty} e^{-\rho\xi} \phi\big(y(T, \xi)\big) \, d\xi,
\end{aligned}
$$

that we wish to minimize over all controls $u = (u_1, \ldots, u_d) \in L^2_{\mathcal{P}}([t, T] \times \Omega; \mathbb{R}^d)$ satisfying the constraint $|u(s)|_{\mathbb{R}^d} \leq r$, $\mathbb{P}$-a.s., for a fixed $r > 0$ and a.a. $s \in [t, T]$. We assume that $g_0 \in C^1(\mathbb{R}^+; \mathbb{R}^+)$ is convex with $g'_0(0) > 0$ and $\phi(y)$, $g_1(t, y)$, $y \in \mathbb{R}$, $t \geq 0$ are continuous real functions differentiable with respect to y with bounded and continuous first derivative. $\alpha > 1$ is a given constant.

The above problem falls under the scope of the general results proved in this section letting $U = \Xi = \mathbb{R}^d$ and $W_t = (\beta_1(t), \ldots, \beta_d(t))$, $t \geq 0$. We also define $H = L^2_\rho(0, +\infty)$ and e^{tA} the shift operators: $(e^{tA}x)(\xi) = x(t + \xi)$, $x \in H$, $t, \xi \geq 0$. Next we define coefficients G and F,

$$G(y)v(\xi) = \sum_{j=1}^d v_j \sigma_j(\xi, y(\xi)), \qquad v \in \mathbb{R}^d,$$

$$F(y)(\xi) = \sum_{j=1}^d \sigma_j(\xi, y(\xi)) \int_0^\xi \sigma_j(\eta, y(\eta)) \, d\eta,$$

and we take R to be the identity operator. Finally, the functions ϕ, g_1 give rise to the Nemytskii operators $H \to H$, $x(\cdot) \mapsto \phi(x(\cdot))$, and $[0, T] \times H \to H$, $(t, x(\cdot)) \mapsto g_1(t, x(\cdot))$; notice that they belong to the required classes $\mathcal{G}$ of Gâteaux differentiable functions even if they are not Fréchet differentiable on H.

Then Theorem 7.2 can be applied and we obtain a characterization of the optimal control by a feedback law.

REFERENCES

[1] AMBROSETTI, A. and PRODI, G. (1995). *A Primer of Nonlinear Analysis*. Cambridge Univ. Press.

[2] BARBU, V. and DA PRATO, G. (1983). *Hamilton–Jacobi Equations in Hilbert Spaces*. Longman, Essex.

[3] BONACCORSI, S. (1998). Some applications in Malliavin calculus. Ph.D. thesis, Dept. Mathematics, Univ. Trento.

[4] BONACCORSI, S. and FUHRMAN, M. (1999). Regularity results for infinite dimensional diffusions. A Malliavin calculus approach. *Rend. Mat. Acc. Lincei* **10** 35–45.

[5] CANNARSA, P. and DA PRATO, G. (1991). Second-order Hamilton–Jacobi equations in infinite dimensions. *SIAM J. Control Optim.* **29** 474–492.

[6] CANNARSA, P. and DA PRATO, G. (1992). Direct solution of a second-order Hamilton–Jacobi equations in Hilbert spaces. In *Stochastic Partial Differential Equations and Applications* (G. Da Prato and L. Tubaro, eds.) 72–85. Longman, Essex.

 [7] CRANDALL, M. G., ISHII, H. and LIONS, P. L. (1992). User's guide to viscosity solutions of
        second order partial differential equations. *Bull. Amer. Math. Soc.* **27** 1–67.

 [8] CRANDALL, M. G., KOCAN, M. and ŚWIĘCH, A. (1993/94). On partial sup-convolutions,
        a lemma of P. L. Lions and viscosity solutions in Hilbert spaces. *Adv. Math. Sci. Appl.* **3**
        (Special Issue) 1–15.

 [9] DA PRATO, G. and ZABCZYK, J. (1992). Stochastic equations in infinite dimensions. In
        Encyclopedia of Mathematics and Its Applications **44**. Cambridge Univ. Press.

[10] DA PRATO, G. and ZABCZYK, J. (1996). *Ergodicity for Infinite-Dimensional Systems.*
        Cambridge Univ. Press.

[11] EL KAROUI, N. (1997). Backward stochastic differential equations a general introduction. In
        Backward Stochastic Differential Equations (N. El Karoui and L. Mazliak, eds.) 7–26.
        Longman, Essex.

[12] FLEMING, W. H. and SONER, H. M. (1993). *Controlled Markov Processes and Viscosity
        Solutions.* Springer, New York.

[13] FUHRMAN, M. (1996). Smoothing properties of transition semigroups in Hilbert spaces.
        NoDEA **3** 445–464.

[14] FUHRMAN, M. and TESSITORE, G. (2001). The Bismut–Elworthy formula for Backward
        SDE's and applications to nonlinear Kolmogorov equations and control in infinite
        dimensional spaces. Preprint, Dipartimento di Matematica Politecnico di Milano, 470/P
        1–25.

[15] GOŁDYS, B. and GOZZI, F. Second order Hamilton–Jacobi equations in Hilbert spaces and
        stochastic control: an approach via invariant measures. Unpublished manuscript.

[16] GOZZI, F. (1995). Regularity of solutions of second order Hamilton–Jacobi equations and
        application to a control problem. *Comm. Partial Differential Equations* **20** 775–826.

[17] GOZZI, F. (1996). Global regular solutions of second order Hamilton–Jacobi equations in
        Hilbert spaces with locally Lipschitz nonlinearities. *J. Math. Anal. Appl.* **198** 399–443.

[18] GOZZI, F., ROUY, E. and ŚWIĘCH, A. (2000). Second order Hamilton–Jacobi equations in
        Hilbert spaces and stochastic boundary control. *SIAM J. Control Optim.* **38** 400–430.

[19] GOZZI, F. and ŚWIĘCH, A. (2000). Hamilton–Jacobi–Bellman equations for the optimal
        control of the Duncan–Mortensen–Zakai equation. *J. Funct. Anal.* **172** 466–510.

[20] GRORUD, A. and PARDOUX, E. (1992). Intégrales Hilbertiennes anticipantes par rapport à un
        processus de Wiener cylindrique et calcul stochastique associé. *Appl. Math. Optim.* **25**
        31–49.

[21] KOCAN, M. and ŚWIĘCH, A. (1995). Second order unbounded parabolic equations in separated
        form. *Studia Math.* **115** 291–310.

[22] HENRY, D. (1981). *Geometric Theory of Semilinear Parabolic Equations. Lecture Notes in
        Math.* **840**. Springer, Berlin.

[23] HU, Y. and PENG, S. (1991). Adapted solution of a backward semilinear stochastic evolution
        equation. *Stochastic Anal. Appl.* **9** 445–459.

[24] LEÓN, J. A. and NUALART, D. (1998). Stochastic evolution equations with random generators.
        Ann. Probab. **26** 149–186.

[25] LIONS, P. L. (1988). Viscosity solutions of fully nonlinear second-order equations and optimal
        stochastic control in infinite dimensions. I. The case of bounded stochastic evolutions.
        Acta Math. **161** 243–278.

[26] LIONS, P. L. (1989). Viscosity solutions of fully nonlinear second order equations and
        optimal stochastic control in infinite dimensions. II. Optimal control of Zakai's equation.
        Stochastic Partial Differential Equations and Applications II. Lecture Notes in Math. **1390**
        147–170. Springer, Berlin.

[27] LIONS, P. L. (1989). Viscosity solutions of fully nonlinear second-order equations and optimal stochastic control in infinite dimensions. III. Uniqueness of viscosity solutions for general second-order equations. *J. Funct. Anal.* **86** 1–18.

[28] MA, J. and YONG, J. (1997). Adapted solution of a degenerate backward SPDE with applications. *Stochastic Process. Appl.* **70** 59–84.

[29] MA, J. and YONG, J. (1999). *Forward-Backward Stochastic Differential Equations and Their Applications*. *Lecture Notes in Math.* **1702**. Springer, Berlin.

[30] MA, Z. M. and RÖCKNER, M. (1992). *Introduction to the Theory of* (*Non-Symmetric*) *Dirichlet Forms*. Springer, Berlin.

[31] MUSIELA, M. (1993). Stochastic PDEs and term structure models. *J. Internationale de Finance*. IGR-AFFI, La Baule.

[32] NUALART, D. (1995). The Malliavin calculus and related topics. In *Probability and Its Applications*. Springer, Berlin.

[33] NUALART, D. and PARDOUX, E. (1988). Stochastic calculus with anticipative integrands. *Probab. Theory Related Fields* **78** 535–581.

[34] PARDOUX, E. (1998). BSDEs and viscosity solutions of a system of semilinear parabolic and elliptic PDEs of second order. In *Stochastic Analysis and Related Topics*: *The Geilo Workshop 1996* (L. Decreusefond, J. Gjerde, B. Øksendal and A. S. Üstünel, eds.) 76–127. Birkhäuser, Berlin.

[35] PARDOUX, E. and PENG, S. (1990). Adapted solution of a backward stochastic differential equation. *Systems and Control Lett.* **14** 55–61.

[36] PARDOUX, E. and PENG, S. (1992). Backward stochastic differential equations and quasilinear parabolic partial differential equations. *Stochastic Partial Differential Equations and Their Applications*. *Lecture Notes in Control Inf. Sci.* **176** 200–217. Springer, Berlin.

[37] RÖCKNER, M. (1999). L^p-analysis of finite and infinite dimensional diffusion operators. *Stochastic PDE's and Kolmogorov Equations in Infinite Dimensions*. *Lecture Notes in Math.* **1715** 65–116. Springer, Berlin.

[38] ŚWIĘCH, A. (1993). Viscosity solutions of fully nonlinear partial differential equations with "unbounded" terms in infinite dimensions. Ph.D. dissertation, Univ. California, Santa Barbara.

[39] ŚWIĘCH, A. (1994). "Unbounded" second order partial differential equations in infinite-dimensional Hilbert spaces. *Comm. Partial Differential Equations* **19** 1999–2036.

[40] TESSITORE, G. (1996). Existence, uniqueness and space regularity of the adapted solutions of a backward SPDE. *Stochastic Anal. Appl.* **14** 461–486.

[41] ZABCZYK, J. (1999). Parabolic equations on Hilbert spaces. *Stochastic PDE's and Kolmogorov Equations in Infinite Dimensions*. *Lecture Notes in Math.* **1715** 117–213. Springer, Berlin.

[42] ZABCZYK, J. (2000). Stochastic invariance and consistency of financial models. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei* (*9*) *Mat. Appl.* **11** 67–80.

DIPARTIMENTO DI MATEMATICA
POLITECNICO DI MILANO
PIAZZA LEONARDO DA VINCI, 32
20133 MILANO
ITALY
E-MAIL: marco.fuhrman@polimi.it

DIPARTIMENTO DI MATEMATICA
UNIVERSITÁ DI PARMA
VIA D'AZEGLIO, 85
43100 PARMA
ITALY
E-MAIL: gianmario.tessitore@unipr.it