

# Differentiability of backward stochastic differential equations in Hilbert spaces with monotone generators.

Philippe Briand

*IRMAR, Université Rennes 1, 35042 Rennes Cedex, FRANCE*

`philippe.briand@univ-rennes1.fr`

Fulvia Confortola

*Dipartimento di Matematica, Politecnico di Milano*

*piazza Leonardo da Vinci 32, 20133 Milano, Italy*

`confortola@mate.polimi.it`

March 17, 2006

## Abstract

The aim of the present paper is to study the regularity properties of the solution of a backward stochastic differential equation with a monotone generator in infinite dimension. We show some applications to the nonlinear Kolmogorov equation and to stochastic optimal control.

## 1 Introduction

In this paper we are concerned mainly with backward stochastic differential equations (BSDEs for short in the remaining of the paper) in infinite dimension in the markovian framework. More precisely, we consider the following backward stochastic evolution equation:

$$\begin{cases} dY_\tau = Z_\tau dW_\tau - \psi(\tau, X_\tau, Y_\tau, Z_\tau) d\tau, & \tau \in [t, T], \\ Y_T = \phi(X_T), \end{cases} \quad (1.1)$$

where  $W$  is a cylindrical Wiener process in some Hilbert space  $\Xi$ . The unknowns are the processes  $Y$  and  $Z$ ;  $Y$  takes its values in a Hilbert space  $K$  and  $Z$  belongs to  $L_2(\Xi, K)$ , the space of Hilbert-Schmidt operators from  $\Xi$  to  $K$ . The process  $X$  takes its values in a Hilbert space  $H$  and is the solution of the forward equation

$$\begin{cases} dX_\tau = AX_\tau d\tau + F(\tau, X_\tau) d\tau + G(\tau, X_\tau) dW_\tau, & \tau \in [t, T], \\ X_t = x \in H, \end{cases} \quad (1.2)$$

where  $A$  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.

The above equations (1.1) and (1.2) form a so called forward-backward system. Under suitable assumptions on  $F$ ,  $G$  and  $\psi$ , there exists a unique adapted process  $(X, Y, Z)$  in the space  $H \times K \times L_2(\Xi, K)$  solution to this system. The processes  $X, Y, Z$  depend on the values

of  $x$  and  $t$ , occurring as initial conditions in (1.2): we may denote them by  $X(\tau, t, x)$ ,  $Y(\tau, t, x)$ ,  $Z(\tau, t, x)$ ,  $\tau \in [t, T]$ .

The goal of this work is the study of the continuity and differentiability with respect to the parameters  $t$  and  $x$  of the process  $Y$  solution to the BSDE (1.1) when the coefficient  $\psi$  in (1.1) is a monotone operator (see Theorem 4.3).

We show, moreover, some interesting applications of this result to the solvability of nonlinear stochastic Kolmogorov equations and to stochastic optimal control problems.

BSDEs in finite and infinite dimensions have been intensively studied in recent years, starting from the paper by E. Pardoux and S. Peng [15]: we refer the reader to [10], [6] and [14] for an exposition of this subject and to [12] for coupled forward-backward systems.

The problem of regular dependence of the solution of a stochastic forward-backward system has been studied in finite dimension by Pardoux, Peng [16] and by El Karoui, Peng and Quenez [10], and, in infinite dimension, by Fuhrman and Tessitore in [7], [8]. In both cases,  $\psi$  is assumed to be Lipschitz continuous with respect to  $y$  and  $z$ . We will assume on  $\psi$  a weaker condition:  $\psi$  will be Lipschitz continuous only with respect to  $z$  and monotone with respect to  $y$  in the spirit of the works [17], [13] and more recently [3].

We should point out that, since we are working in infinite dimensional spaces, the derivatives are understood in the Gâteaux sense: for instance Nemytskii (evaluation) operators are not Fréchet differentiable, except in trivial cases.

This result, beside its intrinsic interest in the framework of the theory of BSDEs, allows us to treat nonlinear partial differential equations as the well known Kolmogorov equation: for  $t \in [0, T]$  and  $x \in H$ ,

$$\partial_t v(t, x) + \mathcal{L}_t[v(t, \cdot)](x) + \psi(t, x, v(t, x), G(t, x)^* \nabla_x v(t, x)) = 0, \quad v(T, x) = \phi(x), \quad (1.3)$$

where  $\mathcal{L}_t$  is the operator associated with the coefficients  $A$ ,  $F$  and  $G$  in (1.2), namely

$$\mathcal{L}_t \phi(x) = \frac{1}{2} \text{Trace} (G(t, x) G(t, x)^* \nabla^2 \phi(x)) + \langle x, A^* \nabla \phi(x) \rangle_H + \langle F(t, x), \nabla \phi(x) \rangle_H. \quad (1.4)$$

Important results concerning connections between BSDEs and PDEs have been stated by Pardoux and Peng [16] and Peng [17], [18], [19], [20] in the Markovian case: these Markovian BSDEs give a Feynman-Kac representation formula for nonlinear parabolic partial differential equations. Conversely, under smoothness assumptions, the solution of a BSDE corresponds to the solution of a system of quasilinear parabolic PDEs.

One of the main results of this paper, Theorem 5.3, specifies conditions for unique solvability of equation (1.3). As usual, if we define the function  $v$  by the formula  $v(t, x) = Y(t, t, x)$ , where  $Y(\cdot, t, x)$  is the solution to BSDE (1.1), then it turns out that the function  $v$  is deterministic and it is a solution of the nonlinear Kolmogorov equation. Various concepts of solutions are known for (linear and) nonlinear parabolic equations in finite and infinite dimensions. Many investigations have been carried out in connection with the Hamilton Jacobi Bellman equation arising in optimal control for nonlinear infinite dimensional stochastic systems. One possibility to deal with equation (1.3) is to look for classical solutions, see e.g. [16] or [18], i.e. 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 (1.3) holds. This requires heavy assumptions on the functions  $\psi$  and  $\phi$ , 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 in the finite dimensional case, see e.g. [13]. However, the theory of viscosity solutions is much harder in the infinite dimensional case. Moreover, in view of applications

to optimal control theory, it is important to show the existence of  $\nabla_x v$ , since this allows to characterize the optimal control by feedback laws. In this paper we will consider solutions in the so called mild sense (already considered in the literature, see [4], [11] and references within, and, in connection with the backward stochastic equations approach, [7], [8]). Namely a mild solution  $v$  of equation (1.3) will satisfy the equality, for  $t \in [0, T]$  and  $x \in H$ ,

$$v(t, x) = \int_t^T P_{t,\tau} [\psi(\tau, \cdot, v(\tau, \cdot), G(\tau, \cdot)^* \nabla_x v(\tau, \cdot))] (x) d\tau + P_{t,T} [\phi](x),$$

which arises formally from (1.3) as the variation of parameters formula. We notice that this formula is meaningful provided  $v$  is only once differentiable with respect to  $x$  and, of course, provided  $\psi$ ,  $v$  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 solution  $v$  by requiring existence and boundedness (or growth conditions) of first derivatives of  $\psi$  and  $\phi$ : compare Theorem 5.3.

As mentioned above, results on system (1.1)-(1.2) are suitable for applications to problems of nonlinear stochastic optimal control. Let us consider a controlled Markov process  $X^u$  in  $H$ , on a time interval  $[t, T] \subset [0, T]$ , described by an Itô stochastic differential equation of the form:

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

where the control process  $u$  takes values in a given subset  $\mathcal{U} \subset U$ . The aim is to choose a control process  $u$ , within a set of admissible controls, in such way to minimize a cost functional of the form:

$$J(t, x, u) = \mathbb{E} \left[ \int_t^T \exp \left( \int_t^s \lambda(r, X_r^u, u_r) dr \right) l(s, X_s^u, u_s) ds + \exp \left( \int_t^T \lambda(r, X_r^u, u_r) dr \right) \phi(X_T^u) \right],$$

where  $\lambda, l, \phi$  are real functions.  $\lambda$  may be called the discount function. The Hamilton Jacobi Bellman equation for the value function is of the form (1.3), provided we set

$$\psi(t, x, y, z) = \inf_{u \in \mathcal{U}} [l(t, x, u) + \langle r(t, x, u), z \rangle + \lambda(t, x, u)y].$$

The control problem is understood in the usual weak sense (see [9] and Subsection 5.2 below). The distinctive feature of our case is the fact that  $v$  occurs explicitly in the Hamilton Jacobi Bellman equation (not only through its derivatives) in a nonlinear way whenever  $\lambda$  effectively depends on  $u$  (i.e. we have the control on the discount). In particular, if we take  $r$  bounded and  $\lambda$  non positive, we obtain a coefficient  $\psi$  monotone with respect to  $y$  and Lipschitz with respect to  $z$ . Remarks on this special case of Hamilton Jacobi Bellman equation are in the classical books by Bensoussan [1] and Bensoussan-Lions [2], but they require  $\lambda$  to be bounded. Fuhrman and Tessitore [7], [8] deal with a similar optimal control problem in infinite dimension, but in their applications the Hamilton Jacobi Bellman equation depends at most linearly on  $v$ . Under suitable conditions, if we let  $v$  denote the unique solution of the Hamilton Jacobi Bellman equation, then we have  $J(t, x, u) \geq v(t, x)$  and the equality holds if and only if  $u$  and  $X^u$  satisfy a suitable feedback law. Thus, the optimal control  $u$  is related to the corresponding optimal trajectory  $X^u$  by a feedback law involving  $\nabla_x v$ . One of the main results of this paper is the characterization in this case of the optimal control by a feedback law (Theorem 5.6). To this end we use the previous results to guarantee existence, uniqueness and regularity of the solution of Hamilton Jacobi Bellman equation.

The plan of the paper is as follows: Section 2 is devoted to notations. In Section 3, we state our main result concerning the differentiability of the solution to a BSDE with respect

to the data. In Section 4, we apply the previous results to study the regularity of the map  $(t, x) \mapsto (Y(\cdot, t, x), Z(\cdot, t, x))$  solution of the forward-backward system (1.1)–(1.2). The last section contains the applications to PDEs: in particular, we study the nonlinear Kolmogorov equation and we give some applications to optimal control.

## 2 Notations

### 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  $\Xi, H, K$  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  $\Xi$  to  $K$  endowed with the Hilbert-Schmidt norm.

By a cylindrical Wiener process with values in a Hilbert space  $\Xi$ , defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we mean a family  $W(t)$ ,  $t \geq 0$ , of linear mappings from  $\Xi$  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  $[0, T]$ .  $\{\mathcal{F}_t\}_{t \in [0, T]}$ , will denote, except in Subsection 5.2, the natural filtration of  $W$ , augmented with the family  $\mathcal{N}$  of  $\mathbb{P}$ -null sets of  $\mathcal{F}_T$ :

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

The filtration  $\{\mathcal{F}_t\}_{t \in [0, T]}$  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

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

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

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

$$|Y|^2 = \mathbb{E} \left[ \int_0^T |Y_\tau|^2 d\tau \right].$$

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

$$\|Z\|_{2,p} = \mathbb{E} \left[ \left( \int_0^T |Z_\tau|^2 d\tau \right)^{p/2} \right]^{1/p}$$

is finite, and  $Z$  admits a predictable version.

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

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

$$\|Y\|_{\infty, p} = \mathbb{E} \left[ \sup_{\tau \in [0, T]} |Y_{\tau}|^p \right]^{1/p}$$

is finite. Elements of  $\mathcal{S}^p(K)$  are identified up to indistinguishability.

If  $Y$  is a process in  $K$ , we will use the following notations:

$$|Y|_1 = \int_0^T |Y_s| ds, \quad |Y|_{\infty} = \sup_{t \in [0, T]} |Y_t|, \quad |Y|_2 = \left( \int_0^T |Y_s|^2 ds \right)^{1/2}, \dots$$

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

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

## 2.2 The class $\mathcal{G}$

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

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

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 this point 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  $F : X \rightarrow V$  belongs to the class  $\mathcal{G}^1(X; V)$  if it is continuous, Gâteaux differentiable on  $X$ , and  $\nabla F : X \rightarrow L(X, V)$  is strongly continuous.

The last requirement of the definition means that for every  $h \in X$  the map  $\nabla F(\cdot)h : X \rightarrow V$  is continuous. Note that  $\nabla F : X \rightarrow L(X, V)$  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 2.2.** 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  $\nabla(G(F))(x) = \nabla G(F(x))\nabla F(x)$ .

**Lemma 2.3.** A map  $F : X \rightarrow V$  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  $\nabla F(\cdot; h) : X \rightarrow V$  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 in [7]. We need to generalize these definitions to functions depending on several variables. For a function  $F : X \times Y \rightarrow V$  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 2.4.** We say that a mapping  $F : X \times Y \rightarrow V$  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  $\nabla_x F : X \times Y \rightarrow L(X, V)$  is strongly continuous.

As in Lemma 2.2 for  $F \in \mathcal{G}^{1,0}(X \times Y, V)$  the map  $(x, y, h) \mapsto \nabla_x F(x, y)h$  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 2.3 (but note that continuity is explicitly required).

**Lemma 2.5.** A continuous map  $F : X \times Y \rightarrow V$  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  $\nabla F(\cdot, \cdot; h) : X \times Y \rightarrow V$  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  $F : X \times Y \times Z \rightarrow V$  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  $\nabla_x F : X \times Y \times Z \rightarrow L(X, V)$  and  $\nabla_y F : X \times Y \times Z \rightarrow L(Y, V)$  are strongly continuous.

### 3 Differentiability of BSDEs

In this section we want to study the BSDE

$$Y_t = \xi + \int_t^T \psi(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s dW_s, \quad 0 \leq t \leq T \quad (3.1)$$

where  $\xi$  is a given  $\mathcal{F}_T$ -measurable random variable with values in  $K$  and  $X_\tau$ ,  $\tau \in [0, T]$ , is a given continuous predictable process with values in  $H$ . To stress the dependence of the solution to the BSDE with respect to the terminal condition  $\xi$  and the process  $X$ , we will denote  $(Y(X, \xi), Z(X, \xi))$  the solution to (3.1). The main point is to study the regularity of the map  $(X, \xi) \mapsto (Y(X, \xi), Z(X, \xi))$ .

$\psi : [0, T] \times H \times K \times L_2(\Xi, K) \rightarrow K$  is assumed to be a Borel-measurable function which satisfies moreover the following assumption:

**Hypothesis A.1.** There exist some constants  $m \geq 0$ ,  $n \geq 0$ ,  $c \geq 0$ ,  $L \geq 0$  and  $\mu \in \mathbb{R}$  such that

- (i) for every  $\sigma \in [0, T]$ ,  $x \in H$ ,  $y \in K$ ,  $z_1, z_2 \in L_2(\Xi, K)$ ,

$$|\psi(\sigma, x, y, z_1) - \psi(\sigma, x, y, z_2)| \leq L |z_1 - z_2|;$$

- (ii) for every  $\sigma \in [0, T]$ ,  $x \in H$ ,  $z \in L_2(\Xi, K)$ ,  $y \mapsto \psi(\sigma, x, y, z)$  is  $\mu$ -monotone meaning that

$$\forall (y_1, y_2) \in K^2, \quad \langle y_1 - y_2, \psi(\sigma, x, y_1, z) - \psi(\sigma, x, y_2, z) \rangle \leq \mu |y_1 - y_2|^2 ;$$

(iii) for every  $\sigma \in [0, T]$ ,  $\psi(\sigma, \cdot, \cdot, \cdot) \in \mathcal{G}^{1,1,1}(H \times K \times L_2(\Xi, K))$ .

(iv) for every  $\sigma \in [0, T]$ ,  $x \in H$ ,  $y \in K$ ,  $z \in L_2(\Xi, K)$ ,

$$|\nabla_x \psi(\sigma, x, y, z)| + |\nabla_y \psi(\sigma, x, y, z)| \leq q(\sigma) + c(|x|^m + |y|^n + |z|^2)$$

where  $q \in L^1(0, T)$ .

(v)  $\psi(s) := \psi(s, 0, 0, 0)$  belongs to  $L^1(0, T)$ .

*Remark 3.1.* It follows from Hypothesis A.1 i), iii) and iv) that

$$\begin{aligned} & |\psi(s, x_1, y_1, z_1) - \psi(s, x_2, y_2, z_2)| \\ & \leq L|z_1 - z_2| + (|x_1 - x_2| + |y_1 - y_2|)(q(s) + C(|x_1|^m + |x_2|^m + |y_1|^n + |y_2|^n + |z_1|^2)). \end{aligned}$$

Consequently, we have

$$|\psi(s, x, y, z)| \leq L|z| + (|x| + |y|)(q(s) + C(|x|^m + |y|^n)) + |\psi(s, 0, 0, 0)|. \quad (3.2)$$

For  $p \geq 1$  we denote  $\mathcal{K}^p$  the Banach space

$$\mathcal{K}^p = \mathcal{S}^p(K) \times \mathcal{M}^p(L_2(\Xi, K))$$

endowed with the norm

$$\|(Y, Z)\|_p = \left( \|Y\|_{\infty, p}^p + \|Z\|_{2, p}^p \right)^{1/p} = \mathbb{E} \left[ \sup_{t \in [0, T]} |Y_t|^p + \left( \int_0^T \|Z_t\|^2 dt \right)^{p/2} \right]^{1/p}.$$

**Proposition 3.2.** *Let the assumption A.1 hold.*

*Let  $p > 1$  and let  $\xi \in L^p(\Omega; K)$ ,  $X \in \mathcal{S}^{p(m+1)}(H)$ . The BSDE (3.1) has a unique solution in  $\mathcal{K}^p$ ,  $(Y(X, \xi), Z(X, \xi))$ , which satisfies*

$$\|(Y(X, \xi), Z(X, \xi))\|_p \leq C \left( 1 + \|\xi\|_p + \|X\|_{\infty, p(m+1)}^{m+1} \right) \quad (3.3)$$

for a suitable constant  $C$  depending only on  $p$ ,  $T$ ,  $L$  and  $\mu$ .

*Proof.* The proof relies heavily on the results given in [3] in the finite dimensional case. The generalization of these results to the case of a cylindrical Wiener process taking its values in an Hilbert space is immediate.

Let  $X \in \mathcal{S}^{p(m+1)}(H)$  and  $\xi \in L^p(\Omega; K)$ . Since the function  $\psi$  is Lipschitz with respect to  $z$  and  $\mu$ -monotone with respect to  $y$ , Theorem 4.2 in [3] shows that the previous BSDE has a unique solution,  $(Y, Z)$  in  $\mathcal{K}^p$ . Moreover, it follows from [3][Proposition 3.2] that

$$\mathbb{E} \left[ \sup_{\tau \in [0, T]} |Y_\tau|^p + \left( \int_0^T |Z_\sigma|^2 d\sigma \right)^{p/2} \right] \leq K \mathbb{E} \left[ |\xi|^p + \left( \int_0^T |\psi(\sigma, X_\sigma, 0, 0)| d\sigma \right)^p \right].$$

Hence, taking into account (3.2) we have,

$$\|(Y, Z)\|_p \leq K \left( 1 + \|\xi\|_p + \|X\|_{\infty, (m+1)p}^{m+1} \right)$$

which is exactly (3.3). □

Let us introduce some further notations. For any  $p > 1$ , we denote by  $\mathcal{X}^p$  the product space  $\mathcal{S}^{p(m+1)}(H) \times L^p(\Omega; K)$  and we consider, according the previous proposition, the map  $\Phi$  from  $\mathcal{X}^p$  to  $\mathcal{K}^p$  defined by  $\Phi(X, \xi) = (Y(X, \xi), Z(X, \xi))$  where  $(Y(X, \xi), Z(X, \xi))$  stands for the solution to the BSDE (3.1).

**Proposition 3.3.** *Let the assumption A.1 hold and let  $p_* = \max(2, n) + \frac{1}{m+1}$ .*

(a) *If  $p > p_*$ , the map  $\Phi$  is continuous from  $\mathcal{X}^p$  to  $\mathcal{K}^r$  where  $r = p/p_*$ .*

(b) *If  $p > p_* + \max(np_*, 2)$ ,  $\Phi$  belongs to  $\mathcal{G}^1(\mathcal{X}^p, \mathcal{K}^\rho)$  with  $r = p/p_*$  and  $\rho = r \min\left(\frac{1}{n+1}, \frac{p_*}{2+p_*}\right)$ .*

*Moreover, for all  $(X, \xi) \in \mathcal{X}^p$ , the directional derivative of  $\Phi$  in the direction  $(N, \zeta) \in \mathcal{X}^p$ ,  $\nabla_{X, \xi} \Phi(X, \xi)(N, \zeta)$  is the unique solution  $(G, H)$  in  $\mathcal{K}^r$  to :*

$$\begin{aligned} G_t + \int_t^T H_\sigma dW_\sigma &= \zeta + \int_t^T \nabla_x \psi(\sigma, X_\sigma, Y_\sigma(X, \xi), Z_\sigma(X, \xi)) N_\sigma d\sigma \\ &+ \int_\tau^T \nabla_y \psi(\sigma, X_\sigma, Y_\sigma(X, \xi), Z_\sigma(X, \xi)) G_\sigma d\sigma \\ &+ \int_\tau^T \nabla_z \psi(\sigma, X_\sigma, Y_\sigma(X, \xi), Z_\sigma(X, \xi)) H_\sigma d\sigma. \end{aligned} \quad (3.4)$$

*Finally the following estimate holds:*

$$\|\nabla_{X, \xi} \Phi(X, \xi)(N, \zeta)\|_r \leq K \left( \|\zeta\|_p + \|N\|_{\infty, p(m+1)} \left( 1 + \|\xi\|_p^{n\vee 2} + \|X\|_{\infty, p(m+1)}^{(m+1)(n\vee 2)} \right) \right). \quad (3.5)$$

*Proof.* (a) Let us assume that  $p > p_*$  and let us set  $r = p/p_* > 1$ . Using Proposition 3.2 in [3], we have, if  $(Y, Z) = \Phi(X, \xi)$  and  $(Y', Z') = \Phi(X', \xi')$  where  $(X, \xi)$  and  $(X', \xi')$  belongs to  $\mathcal{X}^p$ ,

$$\begin{aligned} \|(Y, Z) - (Y', Z')\|_r^r &\leq C \mathbb{E} \left[ |\xi - \xi'|^r + \left( \int_0^T |\psi(s, X_s, Y_s, Z_s) - \psi(s, X'_s, Y_s, Z_s)| ds \right)^r \right] \\ &\leq C \mathbb{E} \left[ |\xi - \xi'|^r + |X - X'|_\infty^r \left( |q|_1^r + |X|_\infty^{mr} + |X'|_\infty^{mr} + |Y|_\infty^{nr} + |Z|_2^{2r} \right) \right] \\ &\leq C \mathbb{E} \left[ |\xi - \xi'|^r + |X - X'|_\infty^r \left( 1 + |X|_\infty^{mr} + |X'|_\infty^{mr} + |Y|_\infty^{nr} + |Z|_2^{2r} \right) \right]. \end{aligned}$$

Using Hölder's inequality, we deduce that

$$\|(Y, Z) - (Y', Z')\|_r \leq C \left( \|\xi - \xi'\|_r + \|X - X'\|_{\infty, p(m+1)} \times A \right)$$

with  $A$  given by

$$A = 1 + \|X\|_{\infty, u(m, r)}^m + \|X'\|_{\infty, u(m, r)}^m + \|Y\|_{\infty, u(n, r)}^n + \|Z\|_{2, u(2, r)}^2$$

where, for any  $a > 0$ ,  $u(a, r) = a \frac{rp(m+1)}{p(m+1)-r} = a \frac{p}{2\vee n}$ . It follows that  $\|(Y, Z) - (Y', Z')\|_r$  is bounded from above by, up to a multiplicative constant  $C$ ,

$$\|\xi - \xi'\|_p + \|X - X'\|_{\infty, p(m+1)} \left( 1 + \|X\|_{\infty, p(m+1)}^m + \|X'\|_{\infty, p(m+1)}^m + \|Y\|_{\infty, p}^n + \|Z\|_{2, p}^2 \right).$$

This gives the continuity of  $\Phi$ .

(b) Let us assume now that  $p > p_* + \max(np_*, 2)$  and let us define  $r = p/p_*$  together with  $\rho = r \min\left(\frac{1}{n+1}, \frac{p_*}{2+p_*}\right) = \min\left(\frac{p}{p_*(n+1)}, \frac{p}{2+p_*}\right) > 1$ .

Let us pick  $(X, \xi)$  and  $(N, \zeta)$  in  $\mathcal{X}^p$  and, for simplicity, let us denote by  $T$  the triple  $(X, Y, Z) = (X, \Phi(X, \xi))$ . We consider the BSDE

$$G_t = \zeta + \int_t^T \nabla_x \psi(s, T_s) N_s ds + \int_t^T \nabla_y \psi(s, T_s) G_s ds + \int_t^T \nabla_z \psi(s, T_s) H_s ds - \int_t^T H_s dB_s.$$

Let us observe that

$$\int_0^T |\nabla_x \psi(s, T_s) N_s| ds \leq K \|N\|_\infty (|q|_1 + |X|_\infty^m + |Y|_\infty^n + |Z|_2^2)$$

and arguing as before with Hölder inequality we deduce that

$$\left[ \mathbb{E} \left( \int_0^T |\nabla_x \psi(s, T_s) N_s| ds \right)^r \right]^{1/r} \leq K \|N\|_{\infty, p(m+1)} \left( 1 + \|X\|_{\infty, p(m+1)}^m + \|Y\|_{\infty, p}^n + \|Z\|_{2, p}^2 \right).$$

Moreover, we have, since  $\psi$  is Lipschitz with respect to  $z$  and  $\mu$ -monotone with respect to  $y$ ,

$$|\nabla_z \psi(s, T_s) (h - h')| \leq L |h - h'|, \quad (g - g', \nabla_y \psi(s, T_s) (g - g'))_K \leq \mu |g - g'|^2.$$

It follows from [3] that the previous BSDE has a unique solution  $(G, H)$  in  $\mathcal{K}^r$  such that

$$\|(G, H)\|_r \leq K \left( \|\zeta\|_r + \|N\|_{\infty, p(m+1)} \left( 1 + \|X\|_{\infty, p(m+1)}^m + \|Y\|_{\infty, p}^n + \|Z\|_{2, p}^2 \right) \right).$$

Taking into account the inequality (3.3), we deduce that

$$\|(G, H)\|_r \leq K \left( \|\zeta\|_p + \|N\|_{\infty, p(m+1)} \left( 1 + \|\xi\|_p^{n \vee 2} + \|X\|_{\infty, p(m+1)}^{(m+1)(n \vee 2)} \right) \right). \quad (3.6)$$

It remains to prove that the directional derivative of  $\Phi$  at  $(X, \xi) \in \mathcal{X}^p$  in the direction  $(N, \zeta) \in \mathcal{X}^p$  is given by  $(G, H)$ . For  $\varepsilon > 0$ , we set  $X^\varepsilon = X + \varepsilon N$  and we consider  $(Y^\varepsilon, Z^\varepsilon)$  the solution in  $\mathcal{K}^p$  to the BSDE

$$Y_t^\varepsilon = \xi + \varepsilon \zeta + \int_t^T \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds - \int_t^T Z_s^\varepsilon dB_s.$$

Before going further, let us introduce the following notations:  $U_t^\varepsilon = \varepsilon^{-1} (Y_t^\varepsilon - Y_t) - G_t$  and  $V_t^\varepsilon = \varepsilon^{-1} (Z_t^\varepsilon - Z_t) - H_t$ . We have,

$$\begin{aligned} U_t^\varepsilon &= \frac{1}{\varepsilon} \int_t^T (\psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) - \psi(s, X_s, Y_s, Z_s)) ds - \int_t^T V_s^\varepsilon dB_s \\ &\quad - \int_t^T \nabla_x \psi(s, T_s) N_s ds - \int_t^T \nabla_y \psi(s, T_s) G_s ds - \int_t^T \nabla_z \psi(s, T_s) H_s ds. \end{aligned}$$

Using the fact that  $\psi(s, \cdot, \cdot, \cdot)$  belongs to  $\mathcal{G}^{1,1,1}$ , we can write

$$\frac{1}{\varepsilon} (\psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) - \psi(s, X_s, Y_s, Z_s)) = A_s^\varepsilon \frac{Y_s^\varepsilon - Y_s}{\varepsilon} + B_s^\varepsilon \frac{Z_s^\varepsilon - Z_s}{\varepsilon} + C_s^\varepsilon$$

where  $A_s^\varepsilon \in L(K, K)$  and  $B_s^\varepsilon \in L(L_2(\Xi, K), K)$  are defined by

$$\forall y \in K, \quad A_s^\varepsilon y = \int_0^1 \nabla_y \psi(s, X_s^\varepsilon, Y_s + \alpha(Y_s^\varepsilon - Y_s), Z_s) y d\alpha,$$

$$\forall z \in L_2(\Xi, K), \quad B_s^\varepsilon z = \int_0^1 \nabla_z \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s + \alpha(Z_s^\varepsilon - Z_s)) z \, d\alpha$$

and where

$$C_s^\varepsilon = \frac{1}{\varepsilon} (\psi(s, X_s^\varepsilon, Y_s, Z_s) - \psi(s, X_s, Y_s, Z_s)) = \int_0^1 \nabla_x \psi(x, X_s + \alpha \varepsilon N_s, Y_s, Z_s) N_s \, d\alpha.$$

Then  $(U^\varepsilon, V^\varepsilon)$  solves the following BSDE

$$U_t^\varepsilon = \int_t^T (A_s^\varepsilon U_s^\varepsilon + B_s^\varepsilon V_s^\varepsilon) \, ds + \int_t^T (P^\varepsilon(s) + Q^\varepsilon(s) + R^\varepsilon(s)) \, ds - \int_t^T V_s^\varepsilon \, dB_s$$

where we have set

$$P^\varepsilon(s) = (A_s^\varepsilon - \nabla_y \psi(s, T_s)) G_s, \quad Q^\varepsilon(s) = (B_s^\varepsilon - \nabla_z \psi(s, T_s)) H_s,$$

together with

$$R^\varepsilon(s) = \int_0^1 \nabla_x \psi(x, X_s + \alpha \varepsilon N_s, Y_s, Z_s) N_s \, d\alpha - \nabla_x \psi(s, T_s) N_s.$$

Since  $(y, A_s^\varepsilon y) \leq \mu |y|^2$  and  $|B_s^\varepsilon z| \leq L |z|$ , we can apply Proposition 3.2 in [3] to learn that,

$$\|(U^\varepsilon, V^\varepsilon)\|_\rho \leq K \left( \|P^\varepsilon\|_1 + \|Q^\varepsilon\|_1 + \|R^\varepsilon\|_1 \right).$$

for a suitable constant  $K$  depending on  $L, \mu, T$  and  $\rho$  and where we use the notation

$$|P^\varepsilon|_1 = |P^\varepsilon(\cdot)|_1 = \int_0^T |P^\varepsilon(s)| \, ds.$$

It remains to check that the right hand side goes to zero as  $\varepsilon$  tends to 0.

Let us recall that

$$\begin{aligned} P^\varepsilon(s) &= \int_0^1 \nabla_y \psi(s, X_s^\varepsilon, Y_s + \alpha(Y_s^\varepsilon - Y_s), Z_s) G_s \, d\alpha - \nabla_y \psi(s, T_s) G_s \\ &= \int_0^1 (\nabla_y \psi(s, X_s^\varepsilon, Y_s + \alpha(Y_s^\varepsilon - Y_s), Z_s) - \nabla_y \psi(s, T_s)) G_s \, d\alpha, \end{aligned}$$

so that

$$|P^\varepsilon(s)| \leq \int_0^1 |u^\varepsilon(\alpha, s)| \, d\alpha$$

with  $u^\varepsilon(\alpha, s) = (\nabla_y \psi(s, X_s^\varepsilon, Y_s + \alpha(Y_s^\varepsilon - Y_s), Z_s) - \nabla_y \psi(s, T_s)) G_s$ .

We consider the product space  $[0, 1] \times [0, T] \times \Omega$  with the measure  $\lambda \otimes \lambda \otimes \mathbb{P}$  where  $\lambda$  is the Lebesgue measure. Since  $\psi(s, \cdot, \cdot, \cdot)$  belongs to  $\mathcal{G}^{1,1,1}$ ,  $u^\varepsilon(s, \alpha)$  goes to zero in  $\lambda \otimes \lambda \otimes \mathbb{P}$ -measure by continuity of  $\Phi$  and we have,

$$\begin{aligned} |u^\varepsilon(\alpha, s)| &\leq K (q(s) + |Z_s|^2 + |X_s|^m + |X_s^\varepsilon|^m + |Y_s|^n + |Y_s^\varepsilon|^n) |G_s| \\ &\leq K (q(s) + |Z_s|^2 + |X|_\infty^m + |X^\varepsilon|_\infty^m + |Y|_\infty^n + |Y^\varepsilon|_\infty^n) |G|_\infty. \end{aligned}$$

First of all, let us recall that  $|G|_\infty$  belongs to  $L^r(\Omega)$  and that  $q \in L^1(0, T)$  and  $|Z|_2^2$  belongs to  $L^{p/2}$ . Moreover,  $Y^\varepsilon$  converges to  $Y$  in  $\mathcal{S}^r(K)$ ,  $X^\varepsilon$  converges to  $X$  in  $\mathcal{S}^{p(m+1)}(H)$  so that  $\{|Y^\varepsilon|_\infty^n\}_{\varepsilon>0}$  is uniformly integrable in  $L^{r/n}$  and  $\{|X^\varepsilon|_\infty^m\}_{\varepsilon>0}$  is uniformly in  $L^{p(m+1)/m}$ . Since we have  $\frac{1}{r} + \frac{2}{p} = \frac{2+p_*}{p} \leq 1$ ,  $\frac{n}{r} + \frac{1}{r} = \frac{(n+1)p_*}{p} \leq 1$  and  $\frac{1}{r} + \frac{m}{p(m+1)} = \frac{\max(2,n)+1}{p} \leq 1$ , the right

hand side of the previous inequality is uniformly integrable as a function of the three variables  $(\alpha, s, \omega)$ . In particular,  $|P^\varepsilon|_1$  converges to 0 in probability as  $\varepsilon$  goes to zero. Moreover we have

$$|P^\varepsilon|_1 \leq K \left(1 + |Z|_2^2 + |X|_\infty^m + |X^\varepsilon|_\infty^m + |Y|_\infty^n + |Y^\varepsilon|_\infty^n\right) |G|_\infty.$$

Since  $|G|_\infty$  is in  $L^r$  we have only to check that the RHS of the previous inequality is uniformly integrable in  $L^{\rho\sigma}$  where  $\sigma = r/(r - \rho)$  is the conjugate exponent of  $r/\rho$  to prove that  $|P^\varepsilon|_1$  tends to 0 in  $L^\rho$  as  $\varepsilon$  goes to 0. Let us observe that

$$\rho\sigma = r \frac{\min\left(\frac{1}{n+1}, \frac{p_*}{2+p_*}\right)}{1 - \min\left(\frac{1}{n+1}, \frac{p_*}{2+p_*}\right)} = r \min\left(\frac{1}{n}, \frac{p_*}{2}\right) = \min\left(\frac{p}{np_*}, \frac{p}{2}\right).$$

As we said before  $|Z|_2^2 \in L^{p/2} \subset L^{\rho\sigma}$ ,  $|Y^\varepsilon|_\infty^n$  is uniformly integrable in  $L^{r/n}$  and  $r/n = p/np_* \geq \rho\sigma$  and  $|X^\varepsilon|_\infty^m$  is uniformly integrable in  $L^{p(m+1)/m}$  and  $\rho\sigma \leq p/2 \leq p(1 + 1/m)$ . Thus  $|P^\varepsilon|_1$  goes to 0 in  $L^\rho$ .

The term  $R^\varepsilon$  can be treated exactly in the same way as we did for  $P^\varepsilon$ . Indeed, we have

$$\begin{aligned} R^\varepsilon(s) &= \int_0^1 \nabla_x \psi(s, X_s + \alpha \varepsilon N_s, Y_s, Z_s) N_s d\alpha - \nabla_x \psi(s, T_s) N_s \\ &= \int_0^1 (\nabla_x \psi(s, X_s + \alpha \varepsilon N_s, Y_s, Z_s) - \nabla_x \psi(s, T_s)) N_s d\alpha, \end{aligned}$$

$\nabla_x \psi$  satisfies the same growth condition as  $\nabla_y \psi$  and  $N$  belongs to  $\mathcal{S}^{p(m+1)}(H)$  which is a better situation than the previous one where we have only  $G \in \mathcal{S}^r(K)$ .

Let us see that  $|Q^\varepsilon|_1$  goes to 0 in  $L^\rho$  as  $\varepsilon$  goes to 0. We write

$$\begin{aligned} Q^\varepsilon(s) &= \int_0^1 \nabla_z \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s + \alpha(Z_s^\varepsilon - Z_s)) H_s d\alpha - \nabla_z \psi(s, T_s) H_s \\ &= \int_0^1 (\nabla_z \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s + \alpha(Z_s^\varepsilon - Z_s)) - \nabla_z \psi(s, T_s)) H_s d\alpha. \end{aligned}$$

Since  $\psi(s, \cdot, \cdot, \cdot)$  is  $\mathcal{G}^{1,1,1}$ ,  $v^\varepsilon(s, \alpha) := (\nabla_z \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s + \alpha(Z_s^\varepsilon - Z_s)) - \nabla_z \psi(s, T_s)) H_s$  goes to zero in  $\lambda \otimes \lambda \otimes \mathbb{P}$ -measure and since  $\psi$  is Lipschitz with respect to  $z$ , we have  $|v^\varepsilon(s, \alpha)| \leq 2L |H_s|$ . Taking into account the fact that  $H$  belongs to the space  $\mathcal{M}^r(L_2(\Xi, K))$ , we deduce that  $v^\varepsilon$  goes to 0 in  $L^1((0, 1) \times (0, T) \times \Omega)$ . In particular,  $|Q^\varepsilon|_1$  goes to 0 in probability. We have also  $|Q^\varepsilon|_1 \leq 2L\sqrt{T} |H|_2$  and since  $|H|_2$  belongs to  $L^r \subset L^\rho$ , the bounded convergence theorem gives  $\lim_{\varepsilon \rightarrow 0} \|Q^\varepsilon\|_\rho = 0$ .

Thus we have proved that  $(G, H) = \nabla_{(X, \xi)} \Phi(X, \xi)(N, \zeta)$  in  $\mathcal{K}^\rho$ .

By linearity, it follows directly from (3.6), that  $(N, \zeta) \longrightarrow (G, H)$  is continuous from  $\mathcal{X}^p$  to  $\mathcal{K}^r$ . Let us prove that  $(X, \xi) \longrightarrow (G, H)$  is continuous from  $\mathcal{X}^p$  to  $\mathcal{K}^\rho$ . Let us consider  $(G, H) \in \mathcal{K}^r$  and  $(G', H') \in \mathcal{K}^r$  the solutions to (3.4) associated to  $(X, \xi) \in \mathcal{X}^p$  and  $(X', \xi') \in \mathcal{X}^p$  ( $(N, \zeta)$  is fixed in  $\mathcal{X}^p$ ). Once again, using the a priori estimates in [3], we have

$$\|(G, H) - (G', H')\|_\rho \leq C \left( \|P\|_1 + \|Q\|_1 + \|R\|_1 \right) \quad (3.7)$$

for a suitable constant  $C$  where we have set

$$\begin{aligned} P(s) &= \{\nabla_y \psi(s, X'_s, Y'_s, Z'_s) - \nabla_y \psi(s, X_s, Y_s, Z_s)\} G_s, \\ Q(s) &= \{\nabla_z \psi(s, X'_s, Y'_s, Z'_s) - \nabla_z \psi(s, X_s, Y_s, Z_s)\} H_s, \\ R(s) &= \{\nabla_x \psi(s, X'_s, Y'_s, Z'_s) - \nabla_x \psi(s, X_s, Y_s, Z_s)\} N_s. \end{aligned}$$

If  $(X', \xi') \longrightarrow (X, \xi)$  in  $\mathcal{X}^p$  then  $(Y', Z') \longrightarrow (Y, Z)$  in  $\mathcal{K}^r$  and, since  $\psi(s, \cdot, \cdot, \cdot)$  belongs to  $\mathcal{G}^{1,1,1}$ , arguing exactly as we did to show the Gâteaux-differentiability of  $\Phi$ , the right hand side of (3.7) tends to 0.

In conclusion,  $\Phi$  belongs to  $\mathcal{G}^1(\mathcal{X}^p, \mathcal{K}^p)$ .  $\square$

*Remark 3.4.* This result can be rewritten in the following way.

Let  $\rho > 1$  and let us define  $p = \rho \max((n+1)p_*, p_* + 2)$  where  $p_* = \max(2, n) + \frac{1}{m+1}$ . Then the map  $(X, \xi) \longrightarrow (Y(X, \xi), Z(X, \xi))$  from  $\mathcal{X}^p$  to  $\mathcal{K}^\rho$  is in  $\mathcal{G}^1$ .

## 4 The forward-backward system

In this section, we apply the previous results on the differentiability of BSDEs to study the differentiability of the solution to the forward-backward system (1.1)–(1.2).

We start by recalling some results on the solution to the forward equation.

### 4.1 The forward equation

Let us recall that  $\{W_t\}_{t \in [0, T]}$  is a cylindrical Wiener process with values in a Hilbert space  $\Xi$ , 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:

$$\begin{cases} dX_\tau = AX_\tau d\tau + F(\tau, X_\tau) d\tau + G(\tau, X_\tau) dW_\tau, & \tau \in [t, T], \\ X_t = x \in H. \end{cases} \quad (4.1)$$

We assume the following:

**Hypothesis A.2. (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  $F : [0, T] \times H \rightarrow H$  is measurable and satisfies, for some constant  $L > 0$ ,

$$\begin{aligned} |F(t, x) - F(t, y)| &\leq L|x - y|, & t \in [0, T], x, y \in H, \\ |F(t, x)| &\leq L(1 + |x|), & t \in [0, T], x \in H. \end{aligned}$$

**(iii)**  $G : [0, T] \times H \longrightarrow L(\Xi, H)$  is such that, for every  $v \in \Xi$ , the map  $Gv : [0, T] \times H \rightarrow H$  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

$$\begin{aligned} |e^{sA}G(t, x)|_{L_2(\Xi, H)} &\leq L s^{-\gamma}(1 + |x|), \\ |e^{sA}G(t, x) - e^{sA}G(t, y)|_{L_2(\Xi, H)} &\leq L s^{-\gamma}|x - y|, \\ |G(t, x)|_{L(\Xi, H)} &\leq L(1 + |x|), \end{aligned}$$

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

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

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

By a solution of equation (4.1) we mean an  $(\mathcal{F}_t)$ -predictable process  $X_\tau$ ,  $\tau \in [t, T]$ , with continuous paths in  $H$ , such that,  $\mathbb{P}$ -a.s.

$$X_\tau = e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-\sigma)A}F(\sigma, X_\sigma) d\sigma + \int_t^\tau e^{(\tau-\sigma)A}G(\sigma, X_\sigma) dW_\sigma, \quad \tau \in [t, T]. \quad (4.2)$$

To stress the 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  $\mathcal{F}_t$ . A consequence of the previous assumptions is that, for every  $s > 0$ ,  $t \in [0, T]$ ,  $x, h \in H$ ,

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

The following results are proved by Fuhrman and Tessitore in [7].

**Proposition 4.1.** *Under the assumptions of Hypothesis A.2-(i)-(ii)-(iii), (4.1) has a unique solution  $X$ . For every  $p \geq 1$ ,  $X$  belongs to  $\mathcal{S}^p(H)$  and*

$$\mathbb{E} \left[ \sup_{\tau \in [t, T]} |X_\tau|^p \right] \leq C(1 + |x|)^p, \quad (4.3)$$

for some constant  $C$  depending only on  $p, \gamma, T, L$  and  $M := \sup_{\tau \in [0, T]} |e^{\tau A}|$ .

To go further, we need to investigate the dependence of the solution  $X(\tau, t, x)$  on the initial data  $x$  and  $t$ . We first reformulate equation (4.2) as an equation on  $[0, T]$ . We set

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

and we consider the equation

$$X_\tau = S(\tau - t)x + \int_0^\tau 1_{[t, T]}(\sigma) S(\tau - \sigma) F(\sigma, X_\sigma) d\sigma + \int_0^\tau 1_{[t, T]}(\sigma) S(\tau - \sigma) G(\sigma, X_\sigma) dW_\sigma, \quad (4.4)$$

for the unknown process  $X_\tau$ ,  $\tau \in [0, T]$ . Under the assumptions of Hypothesis A.2, equation (4.4) has a unique solution  $X \in \mathcal{S}^2(H)$  which belongs to  $\mathcal{S}^p(H)$  for every  $p \in [1, \infty)$ . 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 (4.2). From now on we denote by  $X(\tau, t, x)$ ,  $\tau \in [0, T]$ , the solution of (4.4).

**Proposition 4.2.** *Assume Hypothesis A.2. Then, for every  $p \geq 1$ , the following hold.*

- (i) *The map  $(t, x) \mapsto X(\cdot, t, x)$  belongs to  $\mathcal{G}^{0,1}([0, T] \times H, \mathcal{S}^p(H))$ .*
- (ii) *For every  $h \in H$ , the directional derivative process  $\nabla_x X(\tau, t, x)h$ ,  $\tau \in [0, T]$ , solves the equation:*

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

- (iii) *Finally  $\|\nabla_x X(\tau, t, x)h\|_{\infty, p} \leq c |h|$  for some constant  $c$ .*

## 4.2 The backward equation

With this result in hands, we are now in position to study the regularity with respect to  $(t, x)$  of the solution  $(Y(\tau, t, x), Z(\tau, t, x))$  to the BSDE

$$Y_\tau + \int_\tau^T Z_\sigma dW_\sigma = \phi(X(T, t, x)) + \int_\tau^T \psi(\sigma, X(\sigma, t, x), Y_\sigma, Z_\sigma) d\sigma. \quad (4.5)$$

We assume that  $\psi : [0, T] \times H \times K \times L_2(\Xi, K) \rightarrow K$  verifies Hypothesis A.1. On the function  $\phi : H \rightarrow K$  we make the following assumptions:

**Hypothesis A.3. (i)**  $\phi \in \mathcal{G}^1(H, K)$ .

**(ii)** There exists  $C > 0$  and  $m \geq 0$  such that,

$$\forall x \in H, \quad \forall h \in H, \quad |\nabla_x \phi(x)h| \leq C |h| (1 + |x|^m).$$

Under the assumptions of Hypotheses A.2 and A.1, A.3 by Propositions 4.1 and 3.2 there exists a unique solution of (4.5) that we denote by  $(Y(\tau, t, x), Z(\tau, t, x))$ ,  $\tau \in [0, T]$ . This solution belongs to  $\mathcal{K}^p$  for each  $p > 1$ .

Let us recall that the  $X(\cdot, t, x)$   $\mathcal{F}_{[t, T]}$ -measurable, so that  $Y(t, t, x)$  is measurable with respect to  $\mathcal{F}_{[t, T]}$  and  $\mathcal{F}_t$ ; it follows that  $Y(t, t, x)$  is indeed deterministic (see also [6]).

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

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

is a consequence of the uniqueness of the solution of (4.2). 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  $t \leq s \leq T$  we have,  $\mathbb{P}$ -a.s.,

$$Y(\tau, s, X(s, t, x)) = Y(\tau, t, x), \quad \text{for } \tau \in [s, T], \quad (4.6)$$

together with

$$Z(\tau, s, X(s, t, x)) = Z(\tau, t, x) \quad \text{for a.e. } \tau \in [s, T].$$

To investigate regularity properties of the dependence on  $t$  and  $x$ , we notice that the solution  $(Y(\sigma, t, x), Z(\sigma, t, x))$  of the backward equation in (4.5) can be written, with the notation of Propositions 4.2 and 3.2, as

$$(Y(\cdot, t, x), Z(\cdot, t, x)) = \Phi(X(\cdot, t, x), \phi(X(T, t, x))).$$

Moreover, as a consequence of Hypothesis A.3, it can be easily proved that  $\xi \mapsto \phi(\xi)$  belongs to the space  $\mathcal{G}^1(L^{(m+1)p}(\Omega; H), L^p(\Omega; K))$ , for every  $p \in [1, \infty)$ . The following Proposition is then an immediate consequence of Propositions 4.1, 4.2 and 3.2, and the chain rule for the class  $\mathcal{G}$ , stated in Lemma 2.2.

**Proposition 4.3.** *Assume Hypotheses A.2, A.1 and A.3.*

*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  $p \in (1, \infty)$ .*

*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., for  $\tau \in [0, T]$ ,*

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

Finally the following estimate holds, for each  $p > 1$  :

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

for a constant  $C$  depending on  $p, \mu, T$  and  $L$ .

*Proof.* The first two statements comes from the chain rule for the class  $\mathcal{G}$ . The final estimate follows from (3.5) applied with

$$X = X(\cdot, t, x), \quad N = \nabla_x X(\cdot, t, x)h, \quad \xi = \phi(X(T, t, x)), \quad \zeta = \nabla \phi(X(T, t, x))\nabla_x X(T, t, x)h,$$

taking into account that by Propositions 4.1 and 4.2 we have

$$\|N\|_{\infty, p} \leq c|h|, \quad \|X\|_{\infty, p(m+1)} \leq c(1 + |x|),$$

and, by Hypothesis A.3, we also obtain  $\|\xi\|_p \leq c(1 + |x|)^{m+1}$ ,  $\|\zeta\|_p \leq c|h|(1 + |x|)^m$  for a suitable constant  $c$ .  $\square$

## 5 Application to nonlinear PDEs

### 5.1 The nonlinear Kolmogorov equation

We denote by  $\mathcal{B}_p(H)$  the set of measurable functions  $\phi : H \rightarrow \mathbb{R}$  with polynomial growth, i.e. 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 A.2. The transition semigroup  $P_{t,\tau}$  is defined for arbitrary  $\phi \in \mathcal{B}_p(H)$  by the formula

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

The estimate  $\mathbb{E} \sup_{\tau \in [t, T]} |X_\tau|^p \leq C(1 + |x|)^p$ , see (4.3), shows that  $P_{t,\tau}$  is well defined as a linear operator from  $\mathcal{B}_p(H)$  into itself; 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  $P_{t,\tau}$ :

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

where  $\nabla \phi$  and  $\nabla^2 \phi$  are the first and the second Gâteaux derivatives of  $\phi$  (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  $\phi$  is sufficiently regular, the function  $v(t, x) = P_{t,T}[\phi](x)$ , is a classical solution of the backward Kolmogorov equation:

$$\partial_t v(t, x) + \mathcal{L}_t[v(t, \cdot)](x) = 0, \quad t \in [0, T], \quad x \in H, \quad v(T, x) = \phi(x), \quad x \in H.$$

We refer to [5] and [21] for a detailed exposition. When  $\phi$  is not regular, the function  $v$  defined by the formula  $v(t, x) = P_{t,T}[\phi](x)$  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

$$\begin{cases} \partial_t u(t, x) + \mathcal{L}_t[u(t, \cdot)](x) + \psi(t, x, u(t, x), G(t, x)^* \nabla_x u(t, x)) = 0, & t \in [0, T], x \in H, \\ u(T, x) = \phi(x). \end{cases} \quad (5.1)$$

We will refer to this equation as the nonlinear Kolmogorov equation.  $\psi : [0, T] \times H \times \mathbb{R} \times \Xi \rightarrow \mathbb{R}$  is a given function verifying (A.1) 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 (5.1):

$$u(t, x) = \int_t^T P_{t, \tau} [\psi(\tau, \cdot, u(\tau, \cdot), G(\tau, \cdot)^* \nabla_x u(\tau, \cdot))](x) d\tau + P_{t, T}[\phi](x), \quad t \in [0, T], x \in H, \quad (5.2)$$

and we notice that this formula is meaningful, provided  $\psi(t, \cdot, \cdot, \cdot)$ ,  $u(t, \cdot)$  and  $\nabla_x u(t, \cdot)$  have polynomial growth. We use this formula as a definition for the solution of (5.1):

**Definition 5.1.** We say that a function  $u : [0, T] \times H \rightarrow \mathbb{R}$  is a mild solution of the nonlinear Kolmogorov equation (5.1) 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 (5.2) holds.

*Remark 5.2.* An equivalent formulation of (5.1) or (5.2) 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  $\psi : [0, T] \times H \times \mathbb{R} \times L_2(\Xi, \mathbb{R}) \rightarrow \mathbb{R}$  and we write the equation in the form

$$\partial_t u(t, x) + \mathcal{L}_t[u(t, \cdot)](x) + \psi(t, x, u(t, x), \nabla_x u(t, x)G(t, x)) = 0.$$

The two forms are clearly equivalent provided we identify  $\Xi^* = L_2(\Xi, \mathbb{R})$  with  $\Xi$  by the Riesz isometry. This will be done in the sequel. In particular, although we keep the notation in (5.1), we will sometimes consider  $\psi$  as a real valued function defined on  $[0, T] \times H \times \mathbb{R} \times L_2(\Xi, \mathbb{R})$ , satisfying Hypothesis A.1 with  $K = \mathbb{R}$ .

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

**Theorem 5.3.** Assume that Hypothesis A.2, A.1 (with  $K = \mathbb{R}$ ) and A.3 hold.

The nonlinear Kolmogorov equation (5.1) has a unique mild solution  $u$  given by the formula

$$u(t, x) = Y(t, t, x), \quad (t, x) \in [0, T] \times H$$

where  $(X, Y, Z)$  is the solution of the backward-forward system (4.5). Moreover, we have,  $\mathbb{P}$ -a.s.

$$Y(s, t, x) = u(s, X(s, t, x)), \quad Z(s, t, x) = G(s, X(s, t, x))^* \nabla_x u(s, t, X(s, t, x)).$$

*Proof.* Let us first recall a result of [7, Lemma 6.3]. Let  $\{e_i\}$  be a basis of  $\Xi$  and let us consider the standard real Wiener process  $W_\tau^i = \int_0^\tau \langle e_i, dW_\sigma \rangle$ ,  $\tau \geq 0$ .

If  $v \in \mathcal{G}^{0,1}([0, T] \times H, \mathbb{R})$ , for every  $i$ , the quadratic variation of  $u(s, X(s, t, x))$  and  $W_s^i$  is given by

$$[v(\cdot, X(\cdot, t, x)), W^i]_s = \int_t^s \nabla_x v(\tau, X(\tau, t, x)) G(\tau, X(\tau, t, x)) e_i d\tau, \quad s \in [t, T]. \quad (5.3)$$

(a) *Existence.* As we pointed out before, for  $s \in [t, T]$ ,  $Y(s, t, x)$  is  $\mathcal{F}_{[t, s]}$ -measurable so that  $Y(t, t, x)$  is deterministic. Moreover, as a byproduct of Proposition 4.3, the function  $u$  defined by the formula  $u(t, x) = Y(t, t, x)$  has the regularity properties stated in Definition 5.1. It remains to verify that equality (5.2) holds true for  $u$ .

To this purpose we first fix  $t \in [0, T]$  and  $x \in H$ . Since  $(Y(\cdot, t, x), Z(\cdot, t, x))$  solves the BSDE (4.5), we have, for  $s \in [t, T]$ ,

$$Y(s, t, x) + \int_s^T Z(\tau, t, x) dW_\tau = \phi(X(T, t, x)) + \int_s^T \psi\left(\tau, X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x)\right) d\tau,$$

and, taking expectation for  $s = t$  we obtain, coming back to the definition of  $u$  and  $P_{t, T}$ ,

$$u(t, x) = P_{t, T}[\phi](x) + \mathbb{E} \left[ \int_t^T \psi\left(\tau, X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x)\right) d\tau \right]. \quad (5.4)$$

Moreover, we have, for each  $i$ ,

$$[Y(\cdot, t, x), W^i]_s = \int_t^s \langle Z_\tau, e_i \rangle d\tau, \quad s \in [t, T].$$

Now let us observe that from the Markov property (4.6) we have,  $\mathbb{P}$ -a.s.,

$$u(\tau, X(\tau, t, x)) = Y(\tau, t, x), \quad \tau \in [t, T].$$

It follows from (5.3) that, for each  $i$ ,

$$[Y(\cdot, t, x), W^i]_s = \int_t^s \nabla_x u(\tau, X(\tau, t, x)) G(\tau, X(\tau, t, x)) e_i d\tau, \quad s \in [t, T].$$

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

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

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

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

Thus,  $\psi(\tau, X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x))$  can be rewritten as

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

and (5.4) leads to

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

which is (5.2).

(b) *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 (5.2) and the definition of  $P_{t, \tau}$ , for every  $s \in [t, T]$  and  $x \in H$ ,

$$\begin{aligned} u(s, x) &= \mathbb{E}[\phi(X(T, s, x))] \\ &+ \mathbb{E} \left[ \int_s^T \psi\left(\tau, X(\tau, s, x), u(\tau, X(\tau, s, x)), G(\tau, X(\tau, s, x))^* \nabla_x u(\tau, X(\tau, s, x))\right) d\tau \right]. \end{aligned}$$

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

$$\begin{aligned} u(s, x) &= \mathbb{E}^{\mathcal{F}_s} [\phi(X(T, s, x))] \\ &\quad + \mathbb{E}^{\mathcal{F}_s} \left[ \int_s^T \psi(\tau, X(\tau, s, x), u(\tau, X(\tau, s, x)), G(\tau, X(\tau, s, x))^* \nabla_x u(\tau, X(\tau, s, x))) d\tau \right]. \end{aligned}$$

Taking into account the Markov property of  $X$ ,  $\mathbb{P}$ -a.s.

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

we have

$$\begin{aligned} u(s, X(s, t, x)) &= \mathbb{E}^{\mathcal{F}_s} [\phi(X(T, t, x))] \\ &\quad + \mathbb{E}^{\mathcal{F}_s} \left[ \int_s^T \psi(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x))) d\tau \right]. \end{aligned}$$

If we set

$$\xi = \phi(X(T, t, x)) + \int_t^T \psi(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x))) d\tau$$

the previous equality leads to

$$\begin{aligned} u(s, X(s, t, x)) &= \mathbb{E}^{\mathcal{F}_s} [\xi] - \int_t^s \psi(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x))) d\tau. \end{aligned}$$

Let us observe that  $\mathbb{E}^{\mathcal{F}_t}[\xi] = u(t, x)$ . Since  $\xi \in L^2(\Omega; \mathbb{R})$  is  $\mathcal{F}_{[t, T]}$ -measurable, by the representation theorem, there exists  $\tilde{Z} \in L^2_{\mathcal{P}}(\Omega \times [t, T]; L_2(\Xi, \mathbb{R}))$  such that

$$\mathbb{E}^{\mathcal{F}_s}[\xi] = u(t, x) + \int_t^s \tilde{Z}_\tau dW_\tau, \quad s \in [t, T].$$

We conclude that the process  $u(s, X(s, t, x))$ ,  $s \in [t, T]$  is a (real) continuous semimartingale with canonical decomposition

$$\begin{aligned} u(s, X(s, t, x)) &= u(t, x) + \int_t^s \tilde{Z}_\tau dW_\tau \\ &\quad - \int_t^s \psi(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x))) d\tau. \end{aligned} \tag{5.5}$$

Using (5.3) and arguing as in the proof of existence, we deduce that for a.a.  $\tau \in [t, T]$ ,  $\mathbb{P}$ -a.s.

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

Substituting into (5.5) we obtain

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

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

$$\left\{ (u(s, X(s, t, x)), G(\tau, X(\tau, t, x))^* \nabla_x u(\tau, X(\tau, t, x))) \right\}_{s \in [t, T]}$$

solves the backward equation (4.5). By uniqueness, we have  $Y(s, t, x) = u(s, X(s, t, x))$ , for each  $s \in [t, T]$  and in particular, for  $s = t$ ,  $u(t, x) = Y(t, t, x)$ .

□

## 5.2 Application to Optimal Control

We wish to apply the above results to perform the synthesis of the optimal control for a general nonlinear control system.

Fixed  $t \in [0, T]$  and  $x \in H$  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  $\Xi$  and adapted to the filtration  $\{\mathcal{F}_t\}$ ,
- $u \in L^2_{\mathcal{P}}(\Omega \times [t, T]; U)$  satisfies the constraint:  $u_t \in \mathcal{U}$   $\mathbb{P}$ -a.s. for a.a.  $t \in [t, T]$ , where  $U$  is a separable real Hilbert space and  $\mathcal{U}$  is a fixed subset of  $U$ .

To each a.c.s. we associate the mild solution  $X^u \in C_{\mathcal{P}}([t, T]; L^2(\Omega; H))$  of the state equation:

$$\begin{cases} dX^u_{\tau} = AX^u_{\tau} d\tau + F(\tau, X^u_{\tau}) d\tau + G(\tau, X^u_{\tau})[r(\tau, X^u_{\tau}, u_{\tau}) d\tau + dW_{\tau}], & \tau \in [t, T], \\ X_t = x \in H. \end{cases} \quad (5.6)$$

and the cost:

$$\begin{aligned} J(t, x, u) &= \mathbb{E} \left[ \int_t^T \exp \left( \int_t^s \lambda(r, X^u_r, u_r) dr \right) l(s, X^u_s, u(s)) ds + \exp \left( \int_t^T \lambda(r, X^u_r, u_r) dr \right) \phi(X^u_T) \right] \end{aligned}$$

where  $\lambda, l$  and  $\Phi$  are real functions.  $\lambda$  may be called the discount function.

Our purpose is to minimize the functional  $J$  over all a.c.s. Notice that in this problem the discount  $\lambda$  depends on the control  $u$ . We define the Hamiltonian function relative to the above problem: for all  $t \in [0, T]$ ,  $x \in H$ ,  $y \in \mathbb{R}$  and  $z \in \Xi$

$$\psi(t, x, y, z) = \inf \{ l(t, x, u) + \langle z, r(t, x, u) \rangle_{\Xi} + \lambda(t, x, u)y : u \in \mathcal{U} \}. \quad (5.7)$$

We make the following assumption.

**Hypothesis A.4.** The following holds:

1.  $A, F$  and  $G$  verify Hypothesis A.2.
2.  $r : [0, T] \times H \times \mathcal{U} \rightarrow \Xi$  and there exists a constant  $C > 0$  such that

$$|r(t, x, u)|_{\Xi} \leq C$$

3.  $l : [0, T] \times H \times \mathcal{U} \rightarrow \mathbb{R}$  satisfies the following condition

$$0 \leq l(t, 0, u) \leq q_1(t) + h(u),$$

for some functions  $q_1 \in L^1(0, T)$  and  $h : \mathcal{U} \rightarrow \mathbb{R}_+$ ;

4.  $\lambda : [0, T] \times H \times \mathcal{U} \rightarrow \mathbb{R}$  is a non positive function;
5.  $\phi$  satisfies Hypothesis A.3.

6.  $\psi : [0, T] \times H \times \mathbb{R} \times \Xi \rightarrow \mathbb{R}$  is measurable and, for  $\sigma \in [0, T]$ ,  $\psi(\sigma, \cdot, \cdot, \cdot) \in \mathcal{G}^{1,1,1}(H \times \mathbb{R} \times \Xi)$  with

$$|\nabla_x \psi(\sigma, x, y, z)| + |\nabla_y \psi(\sigma, x, y, z)| \leq q(\sigma) + c(|x|^m + |y|^n + |z|^2)$$

where  $q$  is in  $L^1(0, T)$ .

7. For all  $t \in [0, T]$ ,  $x \in H$ ,  $y \in \mathbb{R}$  and  $z \in \Xi$  there exists a unique  $\Gamma(t, x, y, z) \in \mathcal{U}$  that realizes the minimum in (5.7). Namely

$$\psi(t, x, y, z) = l(t, x, \Gamma(t, x, y, z)) + \langle z, r(t, x, \Gamma(t, x, y, z)) \rangle + \lambda(t, x, \Gamma(t, x, y, z))y.$$

Moreover the map  $\Gamma : [0, T] \times H \times \mathbb{R} \times \Xi \rightarrow \mathcal{U}$  is measurable.

*Remark 5.4.* The function  $\psi : [t, T] \times H \times \mathbb{R} \times \Xi \rightarrow \mathbb{R}$  verifies Hypothesis A.1. In particular, the Hypothesis A.4-2. implies that  $\psi$  is Lipschitz with respect to  $z$ . The condition A.4-3. implies the Hypothesis A.1 (v). Finally, it follows from A.4-4. that  $\psi$  is nonincreasing so that we can take  $\mu = 0$  in Hypothesis A.1 (ii).

Thus, if we identify  $\Xi$  with  $L_2(\Xi, \mathbb{R})$ , the real function  $\psi$  defined on  $[t, T] \times H \times \mathbb{R} \times L_2(\Xi, \mathbb{R})$  verifies Hypothesis A.1 with  $K = \mathbb{R}$ . Therefore by Theorem 5.3 the Hamilton Jacobi Bellman equation

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

admits a unique mild solution  $v \in \mathcal{G}^{0,1}([0, T] \times H)$ .

**Example 5.5.** Let us consider the following situation:  $U = \mathbb{R}$ ,  $\mathcal{U} = \mathbb{R}_+$  and  $r(t, x, u) = 0$ ,  $l(t, x, u) = u^2/2$ ,  $\lambda(t, x, u) = -u$ . Then we have

$$\psi(t, x, y, z) = -\frac{1}{2}y_+^2, \quad \Gamma(t, x, y, z) = y_+.$$

We see on this simple example that the Hamiltonian function  $\psi$  is not Lipschitz with respect to  $y$  but nonincreasing with a polynomial growth order.

We are in a position to prove the main result of this section:

**Theorem 5.6.** *Assume Hypothesis A.4. For all a.c.s. 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  $X^u$ :*

$$u(\sigma) = \Gamma(\sigma, X_\sigma^u, v(\sigma, X_\sigma^u), G(\sigma, X_\sigma^u) \nabla_x v(\sigma, X_\sigma^u)), \quad \mathbb{P} - \text{a.s. for a.a. } \sigma \in [t, T]. \quad (5.8)$$

*Finally there exists at least an a.c.s. for which (5.8) holds. In such a system the closed loop equation:*

$$\begin{cases} d\bar{X}_\tau = A\bar{X}_\tau d\tau + F(\tau, \bar{X}_\tau) d\tau + G(\tau, \bar{X}_\tau) dW_\tau \\ \quad + G(\tau, \bar{X}_\tau) r(\tau, \bar{X}_\tau, \Gamma(\tau, \bar{X}_\tau, v(\tau, \bar{X}_\tau), G(\tau, \bar{X}_\tau) \nabla_x v(\tau, \bar{X}_\tau))) d\tau \\ \bar{X}_t = x \in H. \end{cases} \quad (5.9)$$

*admits a solution and if  $\bar{u}(\sigma) = \Gamma(\sigma, \bar{X}_\sigma, v(\sigma, \bar{X}_\sigma), G(\sigma, \bar{X}_\sigma) \nabla_x v(\sigma, \bar{X}_\sigma))$  then the couple  $(\bar{u}, \bar{X})$  is optimal for the control problem.*

*Proof.* For all a.c.s., setting  $u(s) = 0$  for  $s < t$ , the Girsanov theorem ensures that there exists a probability measure  $\mathbb{P}$  on  $\Omega$  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 the function  $r$  is bounded). Relatively to  $\widetilde{W}$  equation (5.6) can be rewritten:

$$dX_\tau^u = AX_\tau^u d\tau + F(\tau, X_\tau^u) d\tau + G(\tau, X_\tau^u) d\widetilde{W}_\tau, \quad \tau \in [t, T], \quad X_t^u = x \in H.$$

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  $(\Omega, \mathcal{E}, \{\widetilde{\mathcal{F}}_t\}, \widetilde{\mathbb{P}})$ , we can solve the SDE

$$\widetilde{X}(\tau, t, x) = e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-\sigma)A}F(\sigma, \widetilde{X}(\sigma, t, x)) d\sigma + \int_t^\tau e^{(\tau-\sigma)A}G(\sigma, \widetilde{X}(\sigma, t, x)) d\widetilde{W}_\sigma$$

and then the BSDE

$$\widetilde{Y}(\tau, t, x) = \phi(\widetilde{X}(T, t, x)) + \int_\tau^T \psi(\sigma, \widetilde{X}(\sigma, t, x), \widetilde{Y}(\sigma, t, x), \widetilde{Z}(\sigma, t, x)) d\sigma - \int_\tau^T \widetilde{Z}(\sigma, t, x) d\widetilde{W}_\sigma.$$

Actually, this construction can be done for arbitrary  $t \in [0, T]$  and  $x \in H$ . Let us set

$$\forall \tau \in [t, T], \quad D_\tau = \exp \left( \int_t^\tau \lambda(r, \widetilde{X}(r, t, x), u_r) dr \right).$$

We have

$$\begin{aligned} dD_\tau \widetilde{Y}(\tau, t, x) &= D_\tau \left\{ \lambda(\tau, \widetilde{X}(\tau, t, x), u_\tau) \widetilde{Y}(\tau, t, x) - \psi(\tau, \widetilde{X}(\tau, t, x), \widetilde{Y}(\tau, t, x), \widetilde{Z}(\tau, t, x)) \right\} d\tau \\ &\quad + D_\tau \widetilde{Z}(\tau, t, x) d\widetilde{W}_\tau \end{aligned}$$

and thus

$$\begin{aligned} D_T \phi(X_T) &= \widetilde{Y}(t, t, x) + \int_t^T D_\tau \widetilde{Z}(\tau, t, x) d\widetilde{W}_\tau \\ &\quad + \int_t^T D_\tau \left\{ \lambda(\tau, \widetilde{X}(\tau, t, x), u_\tau) \widetilde{Y}(\tau, t, x) - \psi(\tau, \widetilde{X}(\tau, t, x), \widetilde{Y}(\tau, t, x), \widetilde{Z}(\tau, t, x)) \right\} d\tau \end{aligned}$$

We notice that  $\widetilde{X}(\sigma, t, x) = X_\sigma^u$  and coming back to the original Wiener process we get:

$$\begin{aligned} D_T \phi(X_T^u) &= \widetilde{Y}(t, t, x) + \int_t^T D_\tau \left\{ \widetilde{Z}(\tau, t, x) r(\tau, X_\tau^u, u_\tau) d\tau + \widetilde{Z}(\tau, t, x) dW_\tau \right\} \\ &\quad + \int_t^T D_\tau \left\{ \lambda(\tau, X_\tau^u, u_\tau) \widetilde{Y}(\tau, t, x) - \psi(\tau, X_\tau^u, \widetilde{Y}(\tau, t, x), \widetilde{Z}(\tau, t, x)) \right\} d\tau. \end{aligned}$$

Now, from Proposition 5.3, we have  $\widetilde{Y}(\tau, t, x) = v(t, X_\tau^u)$  and

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

Taking expectation with respect to the original probability  $\mathbb{P}$  in the previous relation we obtain:

$$\begin{aligned} &\mathbb{E} [D_T \phi(X_T^u)] - v(t, x) \\ &= \mathbb{E} \left[ \int_t^T D_\sigma \left\{ \lambda(\sigma, X_\sigma^u, u_\sigma) v(\sigma, X_\sigma^u) - \psi(\sigma, X_\sigma^u, v(\sigma, X_\sigma^u), G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u)) \right\} d\sigma \right] \\ &\quad + \mathbb{E} \left[ \int_t^T D_\sigma G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u) r(\sigma, X_\sigma^u, u_\sigma) d\sigma \right]. \end{aligned}$$

Adding to both hand side the term

$$\mathbb{E} \left[ \int_t^T D_\sigma l(\sigma, X_\sigma^u, u_\sigma) d\sigma \right]$$

we get to the following expression

$$J(t, x, u) - v(t, x) = \mathbb{E} \left[ \int_t^T D_\sigma H(\sigma, X_\sigma^u, v(\sigma, X_\sigma^u), G(\sigma, X_\sigma^u)^* \nabla_x v(\sigma, X_\sigma^u), u_\sigma) d\sigma \right],$$

where we have set

$$H(\sigma, x, y, z, u) = -\psi(\sigma, x, y, z) + \lambda(\sigma, x, u)y + zr(\sigma, x, u) + l(\sigma, x, u)$$

The above equality is known as the *fundamental relation* and immediately implies, by definition of  $\psi$  that  $v(t, x) \leq J(t, x, u)$  and that the equality holds if and only if (5.8) holds.

Finally the existence of a weak solution to equation (5.9) is again a consequence of the Girsanov theorem. Namely let  $X \in C_{\mathcal{P}}([t, T]; L^2(\Omega; H))$  be the mild solution of

$$\begin{cases} dX_\tau = AX_\tau d\tau + F(\tau, X_\tau) d\tau + G(\tau, X_\tau) dW_\tau, \\ X_t = x, \end{cases}$$

and let  $\widehat{P}$  be the probability on  $\Omega$  under which

$$\widehat{W}_t := - \int_0^t r(\sigma, X_\sigma, \Gamma(\sigma, X_\sigma, r(\sigma, X_\sigma, G(\sigma, X_\sigma)^* \nabla_x v(\sigma, X_\sigma^u)) d\sigma + W_t$$

is a Wiener process ( $r$  is bounded). Then  $X$  is the mild solution of equation (5.9) relatively to the probability  $\widehat{P}$  and the Wiener process  $\widehat{W}$ .  $\square$

## References

- [1] Bensoussan, A.; *Stochastic control by functional analysis methods*. Studies in Mathematics and its Applications, 11. North-Holland Publishing Co., Amsterdam-New York, 1982.
- [2] Bensoussan, A.; Lions, J.L.; *Application of variational inequalities in stochastic control*. Studies in Mathematics and its Applications, 12. North-Holland Publishing Co., Amsterdam-New York, 1982.
- [3] Briand, Ph.; Delyon, B.; Hu, Y.; Pardoux, E.; Stoica, L.  *$L^p$  solutions of backward stochastic differential equations*. Stochastic Process. Appl. 108 (2003), no. 1, 109-129.
- [4] Cerrai, S.; *Second order PDE's in finite and infinite dimensions. A probabilistic approach*. Probab. Theory Relat. Fields, 115, 1999, 383-399.
- [5] G. Da Prato, J. Zabczyk, **Stochastic equations in infinite dimensions**, Encyclopedia of Mathematics and its Applications, 44, Cambridge University Press, 1992.
- [6] N. El Karoui, *Backward stochastic differential equations a general introduction*, in: **Backward Stochastic Differential Equations**, ed. N. El Karoui, L. Mazliak 7-26. Pitman Research Notes in Mathematics Series 364, Longman, 1997.
- [7] Fuhrman, M.; Tessitore, G.; *Non linear Kolmogorov equations in infinite dimensional spaces: the backward stochastic differential equations approach and applications to optimal control*. Ann. Probab. 30 (2002), no. 3, 1397-1465.

- [8] Fuhrman, M.; Tessitore, G.; *Infinite horizon backward stochastic differential equations and elliptic equations in Hilbert spaces*. Ann. Probab. 32 (2004), no. 1B, 607–660.
- [9] W. H. Fleming, H. M. Soner, **Controlled Markov processes and viscosity solutions**, Springer-Verlag, 1993.
- [10] N. El Karoui; S. Peng; M. C. Quenez, *Backward stochastic differential equations in finance*. Math. Finance 7 (1997), no. 1, 1–71.
- [11] F. Gozzi, E. Rouy, *Regular solutions of second-order stationary Hamilton-Jacobi equations*. (English. English summary) J. Differential Equations 130 (1996), no. 1, 201–234.
- [12] J. Ma, J. Yong, **Forward-backward stochastic differential equations and their applications**, Lecture Notes in Mathematics 1702, Springer, 1999.
- [13] Pardoux, E.; *Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order*. Stochastic analysis and related topics, VI (Geilo, 1996), 79–127, Progr. Probab., 42, Birkhuser Boston, Boston, MA, 1998.
- [14] Pardoux, E.; *BSDEs, weak convergence and homogenization of semilinear PDEs*. Nonlinear analysis, differential equations and control (Montreal, QC, 1998), 503–549, NATO Sci. Ser. C Math. Phys. Sci., 528, Kluwer Acad. Publ., Dordrecht, 1999.
- [15] E. Pardoux, S. Peng, *Adapted solution of a backward stochastic differential equation*, Systems and Control Lett. **14**, 1990, 55–61.
- [16] E. Pardoux, S. Peng, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, in: **Stochastic partial differential equations and their applications**, eds. B.L. Rozowski, R.B. Sowers, 200–217, Lecture Notes in Control Inf. Sci. 176, Springer, 1992.
- [17] Peng, S. *Probabilistic interpretation for Systems of Quasilinear Parabolic Partial Differential Equations* Stochastics, 1991, 37, 61–74.
- [18] Peng, S. *Stochastic Hamilton-Jacobi-Bellman equations*. SIAM J. Control Optim., 30, 284–304, 1992.
- [19] Peng, S. *A generalized dynamic programming principle and Hamilton-Jacobi-Bellman Equation* Stochastics, 38, 119–134, 1992.
- [20] Peng, S. *A nonlinear Feynman-Kac formula and applications* in Proceedings of Symposium of System Sciences and Control Theory, ed Chen and Yong. Singapore: Word Scientific, 173–184.
- [21] J. Zabczyk, *Parabolic equations on Hilbert spaces*, in: **Stochastic PDE's and Kolmogorov Equations in Infinite Dimensions**, ed. G. Da Prato, 117–213. Lecture Notes in Mathematics 1715, Springer, 1999.