

# Generalized directional gradients, backward stochastic differential equations and mild solutions of semilinear parabolic equations

Marco Fuhrman \*

Dipartimento di Matematica, Politecnico di Milano  
piazza Leonardo da Vinci 32, 20133 Milano, Italy  
e-mail: marco.fuhrman@polimi.it

Gianmario Tessitore<sup>†</sup>

Dipartimento di Matematica, Università di Parma  
via d'Azeglio 85, 43100 Parma, Italy  
e-mail: gianmario.tessitore@unipr.it

## Abstract

We study a forward-backward system of stochastic differential equations in an infinite dimensional framework and its relationships with a semilinear parabolic differential equation on a Hilbert space, in the spirit of the approach of Pardoux-Peng. We prove that the stochastic system allows to construct a unique solution of the parabolic equation in a suitable class of locally Lipschitz real functions. The parabolic equation is understood in a mild sense which requires the notion of a generalized directional gradient, that we introduce by a probabilistic approach and prove to exist for locally Lipschitz functions. The use of the generalized directional gradient allows to cover various applications to option pricing problems and to optimal stochastic control problems (including control of delay equations and reaction-diffusion equations), where the lack of differentiability of the coefficients precludes differentiability of solutions to the associated parabolic equations of Black-Scholes or Hamilton-Jacobi-Bellman type.

## 1 Introduction

Consider the following system of stochastic differential equations with unknowns  $(X, Y, Z)$

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

where  $H$  is a Hilbert space,  $A$  the generator of a strongly continuous semigroup in  $H$ ,  $W$  a cylindrical Wiener process and  $F, G, \psi, \phi$  are given functions. The first equation is a *forward*

---

\*Partially supported by the European Community's Human Potential Programme under contract HPRN-CT-2002-00279, QP-Applications.

<sup>†</sup>Partially supported by the European Community's Human Potential Programme under contract HPRN-CT-2002-00281, Evolution Equations

It is an equation for a Hilbert-valued process  $X$  while the second one is a one-dimensional *backward* stochastic differential equation (BSDE) in the sense of [29] (that is the final condition  $Y_T = \phi(X_T)$  is imposed and  $Y, Z$  are required to be adapted).

One of the main issues of the recent theory of BSDEs is to relate (1.1) to the following (formally written) semilinear parabolic partial differential equation

$$\begin{cases} \frac{\partial u(t, x)}{\partial t} + \mathcal{L}_t[u(t, \cdot)](x) = \psi(t, x, u(t, x), \nabla_x u(t, x)G(t, x)), & t \in [0, T], x \in H, \\ u(T, x) = \phi(x), \end{cases} \quad (1.2)$$

where  $(\mathcal{L}_t)_{t \in [0, T]}$  is the generator of the Markov semigroup corresponding to the forward equation in (1.1); that is, at least for regular enough  $u : H \rightarrow \mathbb{R}$ ,

$$\mathcal{L}_t[u](x) = \frac{1}{2} \text{Tr} \left( G(t, x) G^*(t, x) \nabla_x^2 u(x) \right) + \langle Ax, \nabla_x u(x) \rangle_H + \langle F(t, x), \nabla_x u(x) \rangle_H.$$

More precisely (see for instance [16], [30], [28] or [22] and references within) if we define

$$u(t, x) = Y_t \quad (1.3)$$

then  $u$  is a candidate solution for equation (1.2). This is sometimes called the non-linear Feynman-Kac formula.

In the finite dimensional case (that is when  $H$  is finite dimensional,  $A = 0$  and  $W$  has values in a finite dimensional space) different notions of solution to equation (1.2) have been checked upon  $u$ . Namely  $u$  turns out to be a viscosity solution of (1.2) under very general assumptions on  $F, G, \psi$  and  $\phi$  (see for instance [28] or [22] and references within). On the other side when  $F, G, \psi$  and  $\phi$  are very regular it is proved in [30] that  $u$  is the classical solution of equation (1.2). Moreover, see [3], under suitable conditions,  $u$  is a weak solution of (1.2) in the Sobolev sense. Finally when  $\mathcal{L}_t$  is in divergence form and corresponds to a closable coercive symmetric Dirichlet form then  $u$  is a weak solution of (1.2) in the sense of Dirichlet forms (see [2]). It is of special interest for our purposes to notice that in this last paper an important role is played by a generalized notion of directional gradient that is naturally related to the Dirichlet form.

In the general infinite dimensional case the authors showed in [18] that a suitable notion of solution for BSDE approach to semilinear PDEs is the one of *mild solution*. More precisely assuming that  $F, G, \phi$  and  $\psi$  are once Gâteaux differentiable in  $x$ , with continuous first order directional derivatives having polynomial growth, then the function  $u : [0, T] \times H \rightarrow \mathbb{R}$  given by (1.3) has the same regularity properties and, within this function class, it is the unique element verifying for all  $0 \leq t \leq T$  and  $x \in H$  the variation of constants formula:

$$u(t, x) = P_{t, T}[\phi](x) - \int_t^T P_{t, \tau} \left[ \psi(\tau, \cdot, u(\tau, \cdot), \nabla_x u(\tau, \cdot)G(\tau, \cdot)) \right](x) d\tau, \quad (1.4)$$

$(P_{t, \tau})_{0 \leq t \leq \tau \leq T}$  being the Markov transition semigroup of the process  $X$ , i.e. corresponding to the operators  $(\mathcal{L}_t)_{t \in [0, T]}$ .

Equation (1.4), as it is written, is meaningful for a function  $u$  differentiable in  $x$ . If we drop differentiability assumptions for the coefficients  $F, G, \phi$  and  $\psi$  but retain suitable Lipschitz (or local Lipschitz) conditions we can still prove existence and uniqueness of the solution to the forward-backward system (1.1) and consequently construct the candidate solution  $u$ . However, since we do not assume any non-degeneracy on  $G$  nor require any smoothing property of  $P_{t, \tau}$ , there is no reason to expect solutions to equation (1.2) to be differentiable. This difficulty arises

in many concrete examples from mathematical finance and optimal control theory where it is natural to consider coefficients which are Lipschitz but not differentiable (see Sections 6, 7.1 and 7.2). Thus, assuming that  $F$ ,  $G$ ,  $\phi$ ,  $\psi$  are only Lipschitz continuous, we are faced with the following problems: in which sense is the function  $u$  defined in (1.3) a solution to equation (1.2)? Is it the unique solution in an appropriate function class?

First we notice that the function  $u$  in (1.3) is locally lipschitz in  $x$  with polynomial growth, more precisely there exist  $C > 0$  and  $r \geq 0$  such that

$$|u(t, x) - u(t, y)| \leq C|x - y|(1 + |x| + |y|)^r, \quad |u(t, 0)| \leq C, \quad t \in [0, T], \quad x, y \in H.$$

Let us denote by  $\mathcal{K}$  the class of real measurable functions on  $[0, T] \times H$  with this property. We give a complete answer to the problems formulated above in the following two steps. First, for each element  $v$  of  $\mathcal{K}$  we define a new notion of generalized gradient that coincides with  $\nabla_x v G$  when  $v$  is smooth; in particular the right-hand side of (1.4) is meaningful for each element of  $\mathcal{K}$ , provided the term  $\nabla_x u G$  is replaced by the generalized gradient, and consequently the notion of mild solution makes sense for the class  $\mathcal{K}$ . Second, we show that the function  $u$  in (1.3) is the unique mild solution of (1.2) in the class  $\mathcal{K}$ .

In order to define a generalized gradient for the class  $\mathcal{K}$  we proceed as follows. Given  $v \in \mathcal{K}$  and an interval  $[t, \tau]$ , we compute the joint quadratic variation, denoted  $\langle v(\cdot, X_\cdot), W^\xi \rangle_{[t, \tau]}$ , between the process  $v(\sigma, X_\sigma)$  and an arbitrary projection of the noise  $W^\xi_\sigma := \langle W_\sigma, \xi \rangle$ ,  $\xi \in \Xi$ . We prove that the quadratic variation exists and admits the following expression:

$$\langle v(\cdot, X_\cdot), W^\xi \rangle_{[t, \tau]} = \int_t^\tau \zeta(s, X_s) \xi \, ds \quad (1.5)$$

where  $\zeta : [0, T] \times H \rightarrow \Xi^*$  is a measurable map with suitable integrability properties.

When  $v$  is regular enough it is an obvious consequence of Ito rule that

$$\langle v(\cdot, X_\cdot), W^\xi \rangle_{[t, \tau]} = \int_t^\tau \nabla_x v(s, X_s) G(s, X_s) \xi \, ds.$$

In [19] this formula has been extended to cases when  $v$  is only once Gâteaux differentiable in  $x$ . Thus it seems reasonable to take (1.5) as the definition of the *generalized directional gradient* of  $v$  along the directions of  $G$ : we say that map  $\zeta : [0, T] \times H \rightarrow \Xi^*$  belongs to the generalized directional gradient of  $v$  (in symbols  $\zeta \in \widetilde{\nabla}^G v$ ) if (1.5) holds. Although our definition does not guarantee uniqueness of the map  $\zeta$ , the law of  $\zeta(\cdot, X_\cdot)$  is uniquely determined: as it will be clear in the following this is all we need for applications to semilinear PDEs, finance and optimal control.

The existence of the joint quadratic variation for every  $v \in \mathcal{K}$  and the representation formula (1.5) is the main technical result in this paper. To prove it we have found it convenient to use an extension of the classical definition of joint quadratic variation due to F. Russo and P. Vallois (see [36] and Section 2.2 here) who exploited it in [37] to obtain a generalized Ito formula in the finite dimensional case. We can not apply their results directly because we are treating infinite dimensional processes and, more important, we are dealing with non differentiable functions  $v$ , so we need a specific new proof. The first step in our argument is an extension to general Hilbert space valued diffusions of a classical result in Malliavin calculus (see e.g. [25]), stating that if  $F$  and  $G$  are only Lipschitz continuous then the solution  $X$  to the forward equation in (1.1) has the Malliavin gradient (see Section 4.1). Then we exploit a relationship between existence of Malliavin gradient and joint quadratic variation with a Wiener process originated in [26], already used in [19] in connection with forward-backward systems, and extended here to the

non smooth case (see Proposition 4.4). Finally we conclude by adapting general arguments on the representability of additive functionals of Markov processes to the present case (see [39]).

For locally Lipschitz functions an analytic notion of gradient is well established in the literature on non-smooth analysis, namely the notion of Clarke's gradient, see [4], [11] and [12]. The relationship between our definition and Clarke's gradients for the functions of the class  $\mathcal{K}$  is discussed in Appendix A.

Once a suitable definition of generalized directional gradient is introduced we can extend the definition of mild solution in a straightforward way. Namely we say that  $u \in \mathcal{K}$  is a mild solution of (1.2) in the sense of the generalized gradient if for all  $\zeta \in \tilde{\nabla}^G u$  the following holds:

$$u(t, x) = P_{t,T}[\phi](x) - \int_t^T P_{t,\tau} \left[ \psi(\tau, \cdot, u(\tau, \cdot), \zeta(\tau, \cdot)) \right](x) d\tau, \quad t \in [0, T], x \in H. \quad (1.6)$$

With such a definition we are able to prove, see Theorem 5.1, that the function  $u$  given by the non-linear Feynman-Kac formula (1.3) is the unique mild solution to equation (1.2) in the class  $\mathcal{K}$ , in the sense of the generalized gradient. In this kind of result, when dealing with a differentiable solution  $u$ , an essential step is to identify the process  $Z$  in (1.1) with the process  $\nabla_x u(\cdot, X)G(\cdot, X)$ . In our general framework we need to identify  $Z$  with  $\zeta(\cdot, X)$  for  $\zeta \in \tilde{\nabla}^G u$ . This follows immediately from our definition of generalized directional gradient and leads to a very direct proof of both existence and uniqueness.

As far as we know, without any smoothing requirement on the semigroup  $P_{t,\tau}$  the treatment of equation (1.2) is out of the scope of all the existing analytic techniques.

In the last two sections we show how our general results can be applied to the problem of option pricing in mathematical finance, and to optimal control problems.

In Section 6 we consider a generalized Black and Scholes model of market taking into account possible *memory effects* and we consider the problem of hedging a *path dependent* claim, see for instance [8], [24] and [40]. Following [15], we also take into account the fact that an investor is allowed to borrow money at interest rate  $R$  larger then the bond rate  $r$  (see also [13]). Our general results allow us to prove that, under suitable assumptions, the value  $u$  of the hedging portfolio is the unique mild solution in the sense of the generalized directional gradient of the following infinite-dimensional generalization of the Black and Scholes parabolic equation

$$\left\{ \begin{array}{l} \frac{\partial}{\partial t} u(t, s, \nu) + \frac{1}{2} \sum_{i,j=1}^d s^i s^j a_{ij}(t, s, \nu) \frac{\partial^2 u}{\partial s^i \partial s^j}(t, s, \nu) + \int_{-T}^0 \left\langle \frac{d\nu}{d\theta}(\theta), \frac{\partial u}{\partial \nu}(t, s, \nu)(\theta) \right\rangle_{\mathbb{R}^d} d\theta \\ \\ = (r(t, s, \nu) - R(t, s, \nu)) \left( u(t, s, \nu) - \left\langle \frac{\partial}{\partial s} u(t, s, \nu), s \right\rangle_{\mathbb{R}^d} \right)^- \\ \\ + r(t, s, \nu) \left( u(t, s, \nu) - \left\langle \frac{\partial}{\partial s} u(t, s, \nu), s \right\rangle_{\mathbb{R}^d} \right), \\ \\ u(T, s, \nu) = \varphi(s, \nu). \end{array} \right. \quad (1.7)$$

where  $a = (a_{ij}) = \sigma \sigma^*$ ,  $\sigma$  is the volatility,  $\varphi$  is the contingent claim,  $s = (s^i) \in \mathbb{R}^d$  and  $\nu \in L^2((0, T); \mathbb{R}^d)$ . Moreover the hedging strategy is given by

$$\pi_t^i = S_t^i \frac{\partial u}{\partial s^i}(t, S_t, S_t(\cdot)), \quad i = 1, \dots, d.$$

We notice that in equation (1.7) the non linear term is locally Lipschitz under reasonable assumptions, but it is never differentiable. Moreover even in very classical examples (e.g. European call and put options) the claim  $\varphi$  is not differentiable. We notice that here we are able to cover a large class of lookback type claims, see Remark 6.3, but, due to the Hilbertian framework, they have to include a smoothing procedure.

Finally in Section 7 we show that the notion of solution that we have introduced for equation (1.2) is suitable for applications to optimal control problems. Namely we show that the Hamilton Jacobi Bellman equation corresponding to a large class of optimal control problems for infinite dimensional stochastic differential equations has the form (1.2) and satisfies the conditions of Theorem 5.1 that guarantee existence and uniqueness of a mild solution in the sense of the generalized gradient. Then in Theorem 7.2 we prove that the solution equals the optimal cost (i.e.  $u$  coincides with the value function) and its generalized directional gradient allows to construct an optimal feedback law.

As it is usual in the BSDE approach to optimal control we can only treat controlled state equations that satisfy a structural requirement. Roughly speaking we have to impose that the image of the control is included in the image of the noise (see the form of equation (7.1)). In Sections 7.1 and 7.2 we show that in many concrete situations such a requirement is not really restrictive. In [23] other non-academic examples can be found. In Section 7.1 we consider a simple (finite dimensional) stochastic differential equation with delay. Following [9] we settle it in an infinite dimensional framework by introducing a suitable delay semigroup. Even in this simple case the Hamilton Jacobi Bellman equation is highly degenerate while the structural requirement we need is satisfied. Thus, under general Lipschitz assumptions on the coefficients of the state equation, Theorem 7.2 can be used to find an optimal control. Finally in Section 7.2 we show that similar results hold for a controlled stochastic heat equation with delay driven by an infinite-dimensional noise.

## 2 Preliminaries

### 2.1 General Notations

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  $(\Omega, \mathcal{F}, \mathbb{P})$  we always denote a complete probability space and by  $\{W_t, t \geq 0\}$  a cylindrical Wiener process defined on  $[0, T] \times \Omega$  with values in a Hilbert space  $\Xi$ . That is  $W(t)$ ,  $t \geq 0$ , is a family of linear mappings  $\Xi \rightarrow L^2(\Omega)$  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  $t \geq 0$ ,  $\mathbb{E} (W(t)u \cdot W(t)v) = \langle u, v \rangle_\Xi \cdot t$ .

For  $0 \leq t \leq \tau$  we let  $\mathcal{F}_{[t, \tau]}$  denote the  $\sigma$ -algebra generated by  $\{W_\sigma, \sigma \in [t, \tau]\}$  and augmented by the sets of  $\mathcal{F}$  with  $\mathbb{P}$ -measure zero. We set  $\mathcal{F}_\tau = \mathcal{F}_{[0, \tau]}$  (except when it is differently specified as in Section 7).

$L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$  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

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

$L^p_{\mathcal{P}}(\Omega; L^2([0, T]; K))$ ,  $p \in [1, +\infty)$ , denotes the space of equivalence classes of processes  $Y : \Omega \times [0, T] \rightarrow K$  such that the norm

$$|Y|_{L^p_{\mathcal{P}}(\Omega \times [0, T]; K)}^p := \mathbb{E} \left( \int_0^T |Y_\tau|^2 d\tau \right)^{p/2}$$

is finite, and  $Y$  admits an  $(\mathcal{F}_t)_{t \geq 0}$ -predictable version.

$L^p_{\mathcal{P}}(\Omega; C([0, T]; K))$ ,  $p \in [2, +\infty)$ , denotes the space of predictable processes  $Y$  with continuous paths in  $K$ , such that the norm

$$|Y|_{L^p_{\mathcal{P}}(\Omega; C([0, T]; K))}^p = \mathbb{E} \sup_{\tau \in [0, T]} |Y_\tau|^p$$

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

## 2.2 Joint quadratic variations

We say that a pair of real stochastic processes  $(X_t, Y_t)$ ,  $t \geq 0$ , admits a joint quadratic variation on the interval  $[0, T]$  if setting

$$C_{[0, T]}^\epsilon(X, Y) = \frac{1}{\epsilon} \int_0^T (X_{t+\epsilon} - X_t)(Y_{t+\epsilon} - Y_t) dt, \quad \epsilon > 0,$$

the limit  $\lim_{\epsilon \rightarrow 0} C_{[0, T]}^\epsilon(X, Y)$  exists in probability. The limit will be denoted  $\langle X, Y \rangle_{[0, T]}$ .

This definition is taken from [36], except that we do not require that the convergence in probability holds uniformly with respect to time. In [36] the process  $\langle X, Y \rangle$  is called generalized covariation process; several properties are investigated in [37], [38], often in connection with the stochastic calculus introduced in [35]. With respect to the classical definition, the present one has some technical advantages that are useful when dealing with convergence issues (compare for instance the proof of Proposition 4.4 below).

In the following we will use obvious modifications of this definition, for instance the interval  $[0, T_1]$  with  $T_1 > T$ , or  $[0, T]$  may be replaced by another interval  $[t, T]$ .

It is easy to show that if  $X$  has paths with finite variation and  $Y$  has continuous paths then  $\langle X, Y \rangle_{[0, T]} = 0$ .

The following result shows that if  $X$  and  $Y$  are stochastic integrals with respect to  $W$  then the joint quadratic variation as defined above coincides with the classical one. A similar conclusion holds for general semimartingales: see [36], Proposition 1.1, where only the finite-dimensional case is addressed.

**Proposition 2.1** *Let  $\{U_t, V_t, : t \geq 0\}$  be predictable processes in  $\Xi^*$  satisfying  $\int_0^T [|U_t|^2 + |V_t|^2] dt < \infty$ ,  $\mathbb{P}$ -a.s. for every  $T > 0$ . Then the processes  $X_t = \int_0^t U_s dW_s$ ,  $Y_t = \int_0^t V_s dW_s$ ,  $t \geq 0$ , admit a joint quadratic variation on every interval  $[0, T]$  and  $\langle X, Y \rangle_{[0, T]} = \int_0^T U_s V_s^* ds$ .*

In the proposition,  $V_s(\omega)^*$  is the element of  $\Xi$  that corresponds to  $V_s(\omega) \in \Xi^*$  by the Riesz isometry ( $s \geq 0$ ,  $\omega \in \Omega$ ).

**Proof.** We first assume that

$$\int_0^T [|U_t|^2 + |V_t|^2] dt \leq C, \quad \mathbb{P} - a.s. \quad (2.1)$$

for some constant  $C$ . We have

$$C_{[0,T]}^\epsilon(X, Y) = \frac{1}{\epsilon} \int_0^T [X_{t+\epsilon} Y_{t+\epsilon} - X_t Y_t - (Y_{t+\epsilon} - Y_t) X_t - (X_{t+\epsilon} - X_t) Y_t] dt.$$

Applying the Ito formula:  $d(X_t Y_t) = X_t V_t dW_t + Y_t U_t dW_t + U_t V_t^* dt$ , we obtain

$$\begin{aligned} C_{[0,T]}^\epsilon(X, Y) &= \frac{1}{\epsilon} \int_0^T \int_t^{t+\epsilon} U_s V_s^* ds dt + \frac{1}{\epsilon} \int_0^T \int_t^{t+\epsilon} (X_s - X_t) V_s dW_s dt \\ &\quad + \frac{1}{\epsilon} \int_0^T \int_t^{t+\epsilon} (Y_s - Y_t) U_s dW_s dt \\ &=: I_1^\epsilon + I_2^\epsilon + I_3^\epsilon. \end{aligned}$$

We have

$$\left| I_1^\epsilon - \int_0^T U_t V_t^* dt \right| \leq \int_0^T \left| \frac{1}{\epsilon} \int_t^{t+\epsilon} U_s V_s^* ds - U_t V_t^* \right| dt \rightarrow 0, \quad \mathbb{P} - a.s.$$

Indeed,  $\mathbb{P}$ -a.s., the function  $t \rightarrow U_t V_t^*$  belongs to  $L^1([0, T_1])$ , for any  $T_1 > T$ , and the result follows easily by approximation with a continuous function with respect to the  $L^1([0, T_1])$ -norm.

By the assumption (2.1) we can apply the stochastic Fubini theorem and we obtain

$$I_2^\epsilon = \int_0^{T+\epsilon} \left[ \frac{1}{\epsilon} \int_{(s-\epsilon)^+}^{s \wedge T} (X_s - X_t) dt V_s \right] dW_s.$$

We note that

$$\int_0^{T+\epsilon} \left| \frac{1}{\epsilon} \int_{(s-\epsilon)^+}^{s \wedge T} (X_s - X_t) dt V_s \right|^2 ds \leq \sup_{s \in [0, T]} \left| \frac{1}{\epsilon} \int_{(s-\epsilon)^+}^{s \wedge T} (X_s - X_t) dt \right|^2 \int_0^{T+\epsilon} |V_s|^2 ds,$$

and,  $\mathbb{P}$ -a.s., the right-hand side tends to 0 by the continuity of the process  $X$  and by (2.1). The well-known inequality

$$\mathbb{P}(|I_2^\epsilon| > \delta) \leq \frac{\eta}{\delta^2} + \mathbb{P} \left( \int_0^{T+\epsilon} \left| \frac{1}{\epsilon} \int_{(s-\epsilon)^+}^{s \wedge T} (X_s - X_t) dt V_s \right|^2 ds > \eta \right), \quad \delta > 0, \eta > 0,$$

then implies that  $I_2^\epsilon \rightarrow 0$  in probability as  $\epsilon \rightarrow 0$ .

In a similar way we show that  $I_3^\epsilon \rightarrow 0$  in probability.

Finally, the restriction (2.1) can be removed in a standard way by introducing the stopping times  $T_n = \inf\{t \geq 0 : \int_0^t [|U_s|^2 + |V_s|^2] ds \geq n\}$  and the stopped processes  $\{X_{t \wedge T_n}, t \geq 0\}$ .  $\square$

### 2.3 The forward equation.

We fix  $T > 0$  and an interval  $[t, T] \subset [0, T]$  and we consider the following stochastic differential equation for an unknown process  $\{X_\tau, \tau \in [t, T]\}$  taking values in a Hilbert space  $H$ :

$$\begin{cases} dX_\tau = AX_\tau d\tau + F(\tau, X_\tau) d\tau + G(\tau, X_\tau) dW_\tau, & \tau \in [t, T], \\ X_t = x \in H. \end{cases} \quad (2.2)$$

We will refer to equation (2.2) as the forward equation. We assume the following.

**Hypothesis 2.2** (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 Borel measurable and satisfies, for some constant  $L > 0$ ,

$$\begin{aligned} |F(t, 0)| &\leq L, \\ |F(t, x_1) - F(t, x_2)| &\leq L |x_1 - x_2|, \quad t \in [0, T], \quad x_1, x_2 \in H. \end{aligned}$$

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

$$\begin{aligned} |e^{sA}G(t, 0)|_{L_2(\Xi, H)} &\leq L s^{-\gamma}, \\ |e^{sA}G(t, x_1) - e^{sA}G(t, x_2)|_{L_2(\Xi, H)} &\leq L s^{-\gamma} |x_1 - x_2|, \quad s > 0, t \in [0, T], x_1, x_2 \in H, \end{aligned} \quad (2.3)$$

$$|G(t, x)|_{L(\Xi, H)} \leq L (1 + |x|), \quad t \in [0, T], x \in H, \quad (2.4)$$

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

By a solution of the forward equation 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.

$$X_\tau = e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-\sigma)A}F(\sigma, X_\sigma) d\sigma + \int_t^\tau e^{(\tau-\sigma)A}G(\sigma, X_\sigma) dW_\sigma, \quad \tau \in [t, T]. \quad (2.5)$$

To stress dependence on initial data  $t$  and  $x$ , sometimes we denote the solution by  $X(\cdot, t, x)$ , we set  $X(\tau, t, x) = x$  for  $\tau \in [0, t]$  and we treat  $X(\cdot, t, x)$  as a process on  $[0, T]$ .

The following result is well known, see e.g. [14], Theorem 5.3.1. or [18] Propositions 3.2 and 3.3.

**Proposition 2.3** Under the assumptions of Hypothesis 2.2, for every  $p \in [2, \infty)$  there exists a unique process  $X \in L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$  solution of (2.5).

Denoting the solution by  $X(\cdot, t, x)$ , we have that the map  $(t, x) \rightarrow X(\cdot, t, x)$  is continuous  $[0, T] \times H \rightarrow L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$ . Moreover:

$$\mathbb{E} \sup_{\tau \in [t, T]} |X(\tau, t, x)|^p \leq C(1 + |x|)^p, \quad t \in [0, T], x \in H, \quad (2.6)$$

$$\mathbb{E} \sup_{\tau \in [t, T]} |X(\tau, t, x_1) - X(\tau, t, x_2)|^p \leq C|x_1 - x_2|^p, \quad t \in [0, T], x_1, x_2 \in H, \quad (2.7)$$

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

Finally, for all  $\tau \in [t, T]$ ,  $X(\tau, t, x)$  is  $\mathcal{F}_{[t, T]}$ -measurable (hence, in particular, independent of  $\mathcal{F}_t$ ).

## 2.4 The forward-backward system.

Let us denote by  $\mathcal{L}_t$  the generator of the Markov process  $X$  associated to equation (2.2). We are interested in finding a solution  $u : [0, T] \times H \rightarrow \mathbb{R}$  of the nonlinear parabolic equation

$$\begin{cases} \frac{\partial u(t, x)}{\partial t} + \mathcal{L}_t[u(t, \cdot)](x) = \psi(t, x, u(t, x), \nabla u(t, x)G(t, x)), & t \in [0, T], x \in H, \\ u(T, x) = \phi(x), \end{cases} \quad (2.8)$$

where  $\psi : [0, T] \times H \times \mathbb{R} \times \Xi^* \rightarrow \mathbb{R}$  and  $\phi : H \rightarrow \mathbb{R}$  are given functions.



The solution of the above parabolic equation is related to the following markovian forward-backward system of stochastic differential equations on an interval  $[t, T] \subset [0, T]$ :

$$\begin{cases} dX_\tau = AX_\tau d\tau + F(\tau, X_\tau) d\tau + G(\tau, X_\tau) dW_\tau, & \tau \in [t, T], \\ dY_\tau = \psi(\tau, X_\tau, Y_\tau, Z_\tau) d\tau + Z_\tau dW_\tau, & \tau \in [t, T], \\ X_t = x \in H, \\ Y_T = \phi(X_T). \end{cases} \quad (2.9)$$

We have a triple of unknown processes  $(X_\tau, Y_\tau, Z_\tau)$ ,  $\tau \in [t, T]$ , where  $X$  takes values in the Hilbert space  $H$ ,  $Y$  is real valued and  $Z$  takes values in the dual  $\Xi^*$  of  $\Xi$ . The first equation in (2.9) is equation (2.2) and will be referred to as the forward equation; the other one will be referred to as the backward equation. On the functions  $\phi$  and  $\psi$  we make the following assumptions.

**Hypothesis 2.4** (i) *The mapping  $\psi : [0, T] \times H \times \mathbb{R} \times \Xi^* \rightarrow \mathbb{R}$  is measurable and there exist  $L > 0$ ,  $m \geq 0$  such that*

$$\begin{aligned} |\psi(t, x, y_1, z_1) - \psi(t, x, y_2, z_2)| &\leq L(|y_1 - y_2| + |z_1 - z_2|), \\ |\psi(t, x_1, y, z) - \psi(t, x_2, y, z)| &\leq L(1 + |z|)(1 + |x_1| + |x_2| + |y|)^m |x_2 - x_1| \\ |\psi(t, 0, 0, 0)| &\leq L \end{aligned}$$

for every  $t \in [0, T]$ ,  $x, x_1, x_2 \in H$ ,  $y, y_1, y_2 \in \mathbb{R}$ ,  $z, z_1, z_2 \in \Xi^*$ .

(ii) *The mapping  $\phi : H \rightarrow \mathbb{R}$  satisfies*

$$|\phi(x_1) - \phi(x_2)| \leq L(1 + |x_1| + |x_2|)^m |x_1 - x_2|, \quad (2.10)$$

for suitable constants  $L > 0$ ,  $m \geq 0$  and every  $x_1, x_2 \in H$ .

We note that the assumptions on  $\phi$  and  $\psi$  imply that there exists a constant  $c > 0$  such that

$$|\psi(t, x, y, z)| \leq c(1 + |x|^{m+1} + |y| + |z|), \quad |\phi(x)| \leq c(1 + |x|^{m+1}),$$

for every  $t \in [0, T]$ ,  $x \in H$ ,  $y \in \mathbb{R}$ ,  $z \in \Xi^*$ .

By a solution of the system (2.9) we mean a triple of  $(\mathcal{F}_t)$ -predictable process  $\{X_\tau, Y_\tau, Z_\tau, \tau \in [t, T]\}$ , ( $X$  with values in  $H$  and continuous paths,  $Y$  with values in  $\mathbb{R}$  and continuous paths,  $Z$  with values in  $\Xi^*$  and  $\mathbb{P}\{\int_t^T |Z_\tau|^2 d\tau < +\infty\} = 1$ ) verifying,  $\mathbb{P}$ -a.s., the following equalities for all  $\tau \in [t, T]$ :

$$\begin{aligned} X_\tau &= e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-\sigma)A}F(\sigma, X_\sigma) d\sigma + \int_t^\tau e^{(\tau-\sigma)A}G(\sigma, X_\sigma) dW_\sigma, \\ Y_\tau &= \phi(X_T) - \int_\tau^T Z_\sigma dW_\sigma - \int_\tau^T \psi(\sigma, X_\sigma, Y_\sigma, Z_\sigma) d\sigma. \end{aligned} \quad (2.11)$$

If we set  $X(\tau, t, x) = x$  for  $\tau \in [0, t]$  as before, then the backward equation can be solved as an equation on  $[0, T]$  and we obtain a triple of processes on  $[0, T]$  that we will sometimes denote  $\{X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x) : \tau \in [0, T]\}$  to stress dependence on the initial data  $t$  and  $x$ .

**Theorem 2.5** *Under the assumptions of Hypotheses 2.2 and 2.4, for every  $p \in [2, \infty)$  there exists a unique solution  $(X, Y, Z)$  of (2.11) with  $X \in L^p_{\mathcal{P}}(\Omega; C([0, T]; H))$ ,  $Y \in L^p_{\mathcal{P}}(\Omega; C([0, T]; \mathbb{R}))$ ,  $Z \in L^p_{\mathcal{P}}(\Omega; L^2([0, T]; \Xi^*))$ . Moreover denoting the solution by  $(X(\cdot, t, x), Y(\cdot, t, x), Z(\cdot, t, x))$  then*

the map  $(t, x) \rightarrow (X(\cdot, t, x), Y(\cdot, t, x), Z(\cdot, t, x))$  is continuous with values in the above spaces. Finally:

$$\begin{aligned} \mathbb{E} \sup_{\tau \in [t, T]} |X(\tau, t, x_1) - X(\tau, t, x_2)|^p &\leq C|x_1 - x_2|^p \\ \mathbb{E} \sup_{\tau \in [t, T]} |Y(\tau, t, x_1) - Y(\tau, t, x_2)|^p &\leq C(1 + |x_1| + |x_2|)^{(m+1)^2 p} |x_2 - x_1|^p \\ \mathbb{E} \left[ \int_t^T |Z(\tau, t, x_1) - Z(\tau, t, x_2)|^2 d\tau \right]^{p/2} &\leq C(1 + |x_1| + |x_2|)^{(m+1)^2 p} |x_2 - x_1|^p \end{aligned} \quad (2.12)$$

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

**Proof.** The statements concerning  $X$  follow from Proposition 2.3. Moreover existence and uniqueness of the solution  $(Y(\cdot, t, x), Z(\cdot, t, x))$  of the backward equation and its continuous dependence on  $(t, x)$  is proved exactly in [18] (see Propositions 4.3, 4.8 and 5.2 in that paper and notice that, if we are only interested in existence and uniqueness of the solution and on its continuous dependence on  $(t, x)$ , the differentiability of the coefficients required there is irrelevant).

It remains to prove the Lipschitz property. In the following by  $c$  we denote a constant depending only on  $p, \gamma, T, L, m$  and  $M$ . Its value can change from line to line.

We start noticing that Hypothesis 2.4 implies that letting  $f(\sigma, y, z) = \psi(\sigma, X_\sigma, y, z)$  then the assumptions of Proposition 4.3 in [18] hold. Thus by inequality (4.9) in [18]:

$$\mathbb{E} \sup_{\tau \in [t, T]} |Y(\tau, t, x)|^p + \mathbb{E} \left( \int_t^T |Z(\tau, t, x)|^2 d\tau \right)^{p/2} \leq c(1 + |x|)^{p(m+1)}, \quad x \in H, t \in [0, T]. \quad (2.13)$$

We define  $\mathcal{K} \subset [0, T]$  as the set of all  $s$  for which there exists a constant  $k_s$  such that

$$\begin{aligned} \mathbb{E} \sup_{\tau \in [s \vee t, T]} |Y(\tau, t, x_1) - Y(\tau, t, x_2)|^p + \mathbb{E} \left[ \int_{(s \vee t)}^T |Z(\tau, t, x_1) - Z(\tau, t, x_2)|^2 d\tau \right]^{p/2} \\ \leq k_s(1 + |x_1| + |x_2|)^{p(m+1)^2} |x_2 - x_1|^p, \quad x_1, x_2 \in H, t \in [0, T]. \end{aligned} \quad (2.14)$$

By (2.10), (2.6) and (2.7) we immediately get that  $T \in \mathcal{K}$ . We claim that there exists a constant  $\delta$  depending only on  $p, \gamma, T, L, m$  and  $M$  such that the following holds

$$s \in \mathcal{K} \quad \Rightarrow \quad [(s - \delta) \vee 0, s] \subset \mathcal{K}.$$

Clearly the above would immediately imply that  $\mathcal{K} = [0, T]$  and complete the proof.

We proceed now to the proof of the claim. We fix  $s \in \mathcal{K}$  and notice that if  $t > s$  there is nothing to prove. Thus we consider an arbitrary  $t \in [0, s]$  and  $x_1, x_2 \in H$ . We introduce the notation  $X_\tau^i = X(\tau, t, x_i)$ ,  $Y_\tau^i = Y(\tau, t, x_i)$ ,  $Z_\tau^i = Z(\tau, t, x_i)$ ,  $i = 1, 2$ ;  $\widehat{Y}_\tau = Y_\tau^2 - Y_\tau^1$ ,  $\widehat{Z}_\tau = Z_\tau^2 - Z_\tau^1$ .

For all  $\tau \in [t, s]$  we have

$$\widehat{Y}_\tau + \int_\tau^s \widehat{Z}_\sigma dW_\sigma = \widehat{Y}_s - \int_\tau^s [\psi(\sigma, X_\sigma^2, Y_\sigma^2, Z_\sigma^2) - \psi(\sigma, X_\sigma^1, Y_\sigma^1, Z_\sigma^1)] d\sigma.$$

Thus, by (4.3) in [18] we have, for all  $\delta > 0$ ,

$$\begin{aligned} \mathbb{E} \sup_{\tau \in [(s-\delta) \vee t, s]} |\widehat{Y}_\tau|^p + \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |\widehat{Z}_\tau|^2 d\tau \right)^{p/2} \\ \leq c\mathbb{E} |\widehat{Y}_s|^p + c\delta^{p/2} \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |\psi(\sigma, X_\sigma^2, Y_\sigma^2, Z_\sigma^2) - \psi(\sigma, X_\sigma^1, Y_\sigma^1, Z_\sigma^1)|^2 d\tau \right)^{p/2}, \end{aligned}$$

and by Hypothesis 2.4

$$\begin{aligned}
& \mathbb{E} \sup_{\tau \in [(s-\delta) \vee t, s]} |\widehat{Y}_\tau|^p + \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |\widehat{Z}_\tau|^2 d\tau \right)^{p/2} \\
& \leq c \mathbb{E} |\widehat{Y}_s|^p + c \delta^{p/2} \mathbb{E} \sup_{\tau \in [(s-\delta) \vee t, s]} |\widehat{Y}_\tau|^p + c \delta^{p/2} \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |\widehat{Z}_\tau|^2 d\tau \right)^{p/2} \\
& + c \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |X_\sigma^2 - X_\sigma^1|^2 (1 + |Z_\sigma^1|)^2 (1 + |X_\sigma^2| + |X_\sigma^1| + |Y_\sigma^1|)^{2m} d\tau \right)^{p/2}.
\end{aligned}$$

If we choose  $\delta$  such that  $c \delta^{p/2} \leq 1/2$  (notice that  $c$  only depends on  $p, \gamma, T, L, m$  and  $M$ )

$$\begin{aligned}
& \mathbb{E} \sup_{\tau \in [(s-\delta) \vee t, s]} |\widehat{Y}_\tau|^p + \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |\widehat{Z}_\tau|^2 d\tau \right)^{p/2} \\
& \leq c \mathbb{E} |\widehat{Y}_s|^p + c \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |X_\sigma^2 - X_\sigma^1|^2 (1 + |Z_\sigma^1|)^2 (1 + |X_\sigma^2| + |X_\sigma^1| + |Y_\sigma^1|)^{2m} d\tau \right)^{p/2} \\
& \leq c \mathbb{E} |\widehat{Y}_s|^p + c \mathbb{E} \left[ \left( 1 + \int_{(s-\delta) \vee t}^s |Z_\sigma^1|^2 d\sigma \right)^{p/2} \sup_{(s-\delta) \vee t \leq \sigma \leq s} \left( |X_\sigma^2 - X_\sigma^1|^p (1 + |X_\sigma^2| + |X_\sigma^1| + |Y_\sigma^1|)^{mp} \right) \right] \\
& \leq c \mathbb{E} |\widehat{Y}_s|^p + c \left( 1 + \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |Z_\sigma^1|^2 d\sigma \right)^{r_1 p/2} \right)^{1/r_1} \left( \mathbb{E} \sup_{(s-\delta) \vee t \leq \sigma \leq s} |X_\sigma^2 - X_\sigma^1|^{pr_2} \right)^{1/r_2} \\
& \quad \times \left( \mathbb{E} \sup_{(s-\delta) \vee t \leq \sigma \leq s} (1 + |X_\sigma^2| + |X_\sigma^1| + |Y_\sigma^1|)^{mpr_3} \right)^{1/r_3},
\end{aligned}$$

where  $r_1, r_2, r_3 \in (1, \infty)$  satisfy  $1/r_1 + 1/r_2 + 1/r_3 = 1$ .

Consequently by (2.7) and (2.13) we obtain

$$\begin{aligned}
& \mathbb{E} \sup_{\tau \in [(s-\delta) \vee t, s]} |\widehat{Y}_\tau|^p + \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |\widehat{Z}_\tau|^2 d\tau \right)^{p/2} \\
& \leq c \mathbb{E} |\widehat{Y}_s|^p + c \left( 1 + |x_1|^{(m+1)} \right)^p \left( 1 + |x_1|^m + |x_2|^m + |x_1|^{(m+1)m} \right)^p |x_2 - x_1|^p \\
& \leq c \mathbb{E} |\widehat{Y}_s|^p + c (1 + |x_1| + |x_2|)^{p(m+1)^2} |x_2 - x_1|^p.
\end{aligned}$$

Since by construction  $s \in \mathcal{K}$  we also have

$$\mathbb{E} |\widehat{Y}_s|^p \leq k_s (1 + |x_1| + |x_2|)^{p(m+1)^2} |x_2 - x_1|^p$$

for a suitable constant  $k_s$ , thus we can conclude

$$\mathbb{E} \sup_{\tau \in [(s-\delta) \vee t, s]} |\widehat{Y}_\tau|^p + \mathbb{E} \left( \int_{(s-\delta) \vee t}^s |\widehat{Z}_\tau|^2 d\tau \right)^{p/2} \leq k_{s-\delta} (1 + |x_1| + |x_2|)^{p(m+1)^2} |x_2 - x_1|^p$$

for a suitable constant  $k_{s-\delta}$ . The proof is therefore completed.  $\square$

**Remark 2.6** By the uniqueness of the solution of the forward equation we deduce that for  $0 \leq t \leq s \leq T$  we have,  $\mathbb{P}$ -a.s.,

$$X(\tau, s, X(s, t, x)) = X(\tau, t, x), \quad \tau \in [s, T]. \quad (2.15)$$

We note that 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. One consequence is that for  $0 \leq t \leq s \leq T$  we have,  $\mathbb{P}$ -a.s.,

$$Y(\tau, s, X(s, t, x)) = Y(\tau, t, x), \quad \text{for } \tau \in [s, T], \quad Z(\tau, s, X(s, t, x)) = Z(\tau, t, x) \quad \text{for a.a. } \tau \in [s, T]. \quad (2.16)$$

Another consequence is that, since the process  $\{X(\tau, t, x), \tau \in [s, T]\}$  is measurable with respect to  $\mathcal{F}_{[t, T]}$ , so is the process  $\{Y(\tau, t, x), \tau \in [s, T]\}$ . In particular,  $Y(t, t, x)$  is  $\mathcal{F}_{[t, T]}$ -measurable and, by adaptedness, it is also  $\mathcal{F}_t$ -measurable. Since these  $\sigma$ -algebras are independent we conclude that  $Y(t, t, x)$  is deterministic.

### 3 The generalized directional gradient

Let us recall our general framework.  $(\Omega, \mathcal{F}, \mathbb{P})$  is a complete probability space and  $\{W_t, t \geq 0\}$  is a cylindrical Wiener process with values in a Hilbert space  $\Xi$ . For  $0 \leq t \leq \tau$   $\mathcal{F}_{[t, \tau]}$  is the  $\sigma$ -algebra generated by  $\{W_\sigma - W_t, \sigma \in [t, \tau]\}$  and augmented with the sets of  $\mathcal{F}$  with  $\mathbb{P}$ -measure zero. We set  $\mathcal{F}_\tau = \mathcal{F}_{[0, \tau]}$ .

For every  $t \in [0, T]$  and  $x \in H$ , let  $\{X(\tau, t, x), \tau \in [t, T]\}$  be the solution of the forward equation (2.2); in particular it is an  $H$ -valued process with continuous paths and adapted to the filtration  $\{\mathcal{F}_{[t, \tau]}, \tau \in [t, T]\}$ .

We let  $X = \{X(\tau, t, x), 0 \leq t \leq \tau \leq T, x \in H\}$  denote the family of solutions of the forward equation.  $X$  is a process with values in  $H$ , measurable with respect to  $\mathcal{F} \times \mathcal{B}(\bar{\Delta}) \times \mathcal{B}(H)$  and  $\mathcal{B}(H)$  respectively (here by  $\bar{\Delta}$  we denote the set  $\{(t, \tau), 0 \leq t \leq \tau \leq T\}$  and by  $\mathcal{B}(\Lambda)$  the Borel  $\sigma$ -algebra of any topological space  $\Lambda$ ).

Let  $u : [0, T] \times H \rightarrow \mathbb{R}$  be a Borel measurable function satisfying

$$|u(t, x) - u(t, y)| \leq C|x - y|(1 + |x| + |y|)^r, \quad |u(t, 0)| \leq C, \quad (3.1)$$

for some  $C > 0, r \geq 0$  and for every  $t \in [0, T], x, y \in H$ .

It follows from this assumption that  $u(t, \cdot)$  is Lipschitz continuous on bounded sets of  $H$ , uniformly with respect to  $t \in [0, T]$ , and there exist  $c > 0$  such that  $|u(t, x)| \leq c(1 + |x|)^{r+1}$  for all  $t \in [0, T], x \in H$ .

We recall the notation  $W^\xi$  for the real Wiener process corresponding to  $\xi \in \Xi$ .

**Theorem 3.1** *Assume that Hypothesis 2.2 holds and that  $u : [0, T] \times H \rightarrow \mathbb{R}$  is a Borel measurable function satisfying (3.1). Then there exists a Borel measurable function  $\zeta : [0, T] \times H \rightarrow \Xi^*$  with the following properties.*

(i) *For every  $t \in [0, T], x \in H$  and  $p \in [2, \infty)$ ,*

$$\mathbb{E} \int_t^T |\zeta(\tau, X(\tau, t, x))|^p d\tau < +\infty. \quad (3.2)$$

(ii) *For  $\xi \in \Xi, x \in H$  and  $0 \leq t \leq T' < T$  the processes  $\{u(\tau, X(\tau, t, x)), \tau \in [t, T']\}$  and  $W^\xi$  admit a joint quadratic variation on the interval  $[t, T']$  and*

$$\langle u(\cdot, X(\cdot, t, x)), W^\xi \rangle_{[t, T']} = \int_t^{T'} \zeta(\tau, X(\tau, t, x)) \xi d\tau, \quad \mathbb{P} - a.s.$$

(iii) Moreover there exists a Borel measurable function  $\rho : [0, T] \times H \rightarrow H^*$  such that for all  $t \in [0, T]$  and all  $x \in H$

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

**Remark 3.2** Concerning uniqueness of  $\zeta$  in Theorem 3.1, we can only assert the following: if  $\hat{\zeta}$  is another function with the stated properties then for  $t \in [0, T]$  and  $x \in H$  we have

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

**Remark 3.3** Under appropriate assumptions, in particular if  $u$  is twice continuously differentiable in  $x$  and the coefficients of the forward equation are sufficiently regular, the Ito formula implies

$$\begin{aligned} u(\tau, X(\tau, t, x)) - u(t, x) \\ = \int_t^\tau \mathcal{L}_\sigma u(\sigma, X(\sigma, t, x)) d\sigma + \int_t^\tau \nabla u(\sigma, X(\sigma, t, x))G(\sigma, X(\sigma, t, x)) dW_\sigma, \end{aligned}$$

where  $\mathcal{L}_\sigma$  denotes the generator of the (nonhomogeneous) Markov process  $X$ . By Proposition 2.1 we obtain, for every  $\xi \in \Xi$ , and every  $T' \in [t, T]$ :

$$\langle u(\cdot, X(\cdot, t, x)), W^\xi \rangle_{[t, T']} = \int_t^{T'} \nabla u(\sigma, X(\sigma, t, x))G(\sigma, X(\sigma, t, x))\xi d\sigma. \quad (3.3)$$

Comparing with the Theorem 3.1 we conclude that in this case the function  $\zeta$  can be chosen equal to

$$\zeta(t, x) = \nabla u(t, x)G(t, x), \quad t \in [0, T], x \in H. \quad (3.4)$$

**Remark 3.4** In (3.3) second derivatives of  $u$  do not appear and one may conjecture that this formula may still hold if  $u$  is once continuous differentiable. This kind of properties have in fact been proved for the generalized covariation process introduced in [36] and ultimately lead to generalized Ito formulas, see [37]. In the present paper a different proof is needed, since  $X$  is an infinite dimensional process and  $u$  is only assumed to be locally Lipschitz continuous.

Remark 3.3, in particular formula (3.4), shows that we may regard  $\zeta$  as a generalized gradient “in the direction  $G$ ”, and motivates the following definition. In Appendix A we clarify the relations between the above definition and another notion of gradient, that can be defined for locally lipschitz maps, namely the Clarke gradient.

**Definition 3.1** Let  $u : [0, T] \times H \rightarrow \mathbb{R}$  be a Borel measurable function satisfying (3.1). The family of all measurable functions  $\zeta : [0, T] \times H \rightarrow \Xi^*$  satisfying properties (i) and (ii) in Theorem 3.1 will be called the generalized directional gradient of  $u$  and denoted by  $\tilde{\nabla}^G u$ .

**Remark 3.5** The map  $\rho$  introduced in point (iii) in Theorem 3.1 plays the role of the gradient of  $u$ : compare (3.4). We notice however that  $\rho$  depends on the coefficients  $A$ ,  $F$  and  $G$  occurring in equation (2.2).

When  $H$  is finite dimensional and  $u$  is of the form

$$u(t, x) = \mathbb{E} \left( \phi(X(T, t, x)) + \int_t^T f(s, X(s, t, x)) ds \right)$$

for suitable  $\phi$  and  $f$ , a factorization result similar to point (iii) in Theorem 3.1 is proved in [15] by earlier general results (see [10]).

## 4 Existence of generalized directional gradients.

We proceed to the proof of Theorem 3.1. The argument consists of several steps which require intermediate results.

### 4.1 Regularity of the forward equation in the Malliavin spaces.

In this section we show that the solution of the forward equation (2.5) is differentiable in the sense of the Malliavin calculus. Now we recall some basic definitions from the Malliavin calculus. We refer the reader to the book [25] 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  $\int_0^T h(t)^* dW(t)$ , where  $h(t)^*$  denotes the element of  $\Xi^*$  that corresponds to  $h(t) \in \Xi$  by the Riesz isometry. Given a Hilbert space  $K$ , let  $S_K$  be the set of  $K$ -valued random variables  $F$  of the form

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

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

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

with values in  $L_2(\Xi, K)$ ; by  $\partial_k$  we denote the partial derivatives with respect to the  $k$ -th 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  $D : S_K \subset L^2(\Omega; K) \rightarrow L^2(\Omega \times [0, T]; L_2(\Xi; K))$  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:

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

The adjoint operator of  $D$ ,

$$\delta : \text{dom}(\delta) \subset L^2(\Omega \times [0, T]; L_2(\Xi; K)) \rightarrow L^2(\Omega; K),$$

is called Skorohod integral. For a process  $u \in \text{dom}(\delta)$  we will also use the notation

$$\delta(u) = \int_0^T u_s dW_s.$$

It is known that  $\text{dom}(\delta)$  contains every  $(\mathcal{F}_t)$ -predictable process in  $L^2(\Omega \times [0, T]; L_2(\Xi; K))$  and for such processes the Skorohod integral coincides with the Itô integral;  $\text{dom}(\delta)$  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,  $\|\delta(u)\|_{L^2(\Omega; K)}^2 \leq \|u\|_{\mathbb{L}^{1,2}(L_2(\Xi; K))}^2$ . The definition of  $\mathbb{L}^{1,2}(K)$  for an arbitrary Hilbert space  $K$  is entirely analogous. We note that the space  $\mathbb{L}^{1,2}(K)$  is isometrically isomorphic to

$L^2([0, T]; \mathbb{D}^{1,2}(K))$ . Finally we recall that if  $F \in \mathbb{D}^{1,2}(K)$  is measurable with respect to  $\mathcal{F}_t$  then  $DF = 0$  a.s. on  $\Omega \times (t, T]$ .

Now let us consider again the process  $\{X(\tau, t, x), \tau \in [t, T]\}$ , solution of the forward equation.

When  $(t, x)$  is fixed, we denote this process simply by  $(X_\tau)$ , and we set  $X_\tau = x$  for  $\tau \in [0, t]$ . 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.  $s, t, \tau$  if  $\tau < t$  or  $s > \tau$ .

**Proposition 4.1** *Assume Hypothesis 2.2. For  $t \in [0, T)$ ,  $x \in H$ , denote  $X_\tau = X(\tau, t, x)$ . Then the following properties hold.*

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

(ii) For every  $q \in [1, \infty)$  we have

$$\mathbb{E} \left( \int_0^T \int_0^T |D_s X_\tau|_{L_2(\Xi, H)}^2 ds d\tau \right)^q < \infty. \quad (4.1)$$

(iii) For every  $\xi \in \Xi$ ,  $T' \in [t, T)$ , we have, as  $\epsilon \rightarrow 0$ ,

$$\mathbb{E} \int_t^{T'} \left| \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi ds - G(\tau, X_\tau) \xi \right|^2 d\tau \rightarrow 0. \quad (4.2)$$

**Remark 4.2** 1. We recall that the coefficients  $F, G$  of the forward equations are required to satisfy some Lipschitz conditions, but no differentiability conditions. In the finite dimensional case it is known that the solution  $X$  is differentiable in the sense of the Malliavin calculus, see e.g. [25] Section 2.2, but we are not aware of any extension to the infinite dimensional case.

2. In [26] a pair of operators, denoted  $D^-$  and  $D^+$ , have been introduced for a subclass of  $\mathbb{L}^{1,2}$ , in the finite dimensional case. Formula (4.2) shows that  $G(\tau, X_\tau)$  coincides with  $D_\tau^- X_\tau$ , whereas  $D_\tau^+ X_\tau = 0$  since  $X$  is adapted. The relevance of these notions with the joint quadratic variation in Theorem 3.1 is made clear by [26] Theorem 7.6 and its infinite dimensional analogues [18] Lemmas 6.3 and 6.4, and [19] Proposition 4.5.

**Proof.** We claim that there exist mappings  $F_n : [0, T] \times H \rightarrow H$ ,  $G_n : [0, T] \times H \rightarrow L(\Xi, H)$ ,  $n = 1, 2, \dots$ , such that

(i)  $F_n$  and  $G_n$  satisfy Hypothesis 2.2 with constants  $L, \gamma$  independent of  $n$ .

(ii) For  $s > 0$ ,  $t \in [0, T]$ , the mappings  $F_n(t, \cdot) : H \rightarrow H$ ,  $e^{sA} G_n(t, \cdot) : H \rightarrow L_2(\Xi, H)$ , are continuously Fréchet differentiable. We denote by  $\nabla F_n(t, x) \in L(H)$ ,  $\nabla(e^{sA} G_n(t, x)) \in L(H, L_2(\Xi, H))$  their Fréchet differentials at point  $x \in H$ . They satisfy the inequalities

$$|\nabla F_n(t, x)|_{L(H)} \leq L, \quad |\nabla(e^{sA} G_n(t, x))|_{L(H, L_2(\Xi, H))} \leq Ls^{-\gamma}. \quad (4.3)$$

(iii) For  $x \in H$ ,  $t \in [0, T]$ ,  $s > 0$ ,

$$F_n(t, x) \rightarrow F(t, x), \quad e^{sA} G_n(t, x) \rightarrow e^{sA} G(t, x)$$

in  $H$  and  $L_2(\Xi, H)$  respectively.

Such functions  $F_n, G_n$  can be constructed as follows by the same procedure as in [27], to which we refer the reader for details. We choose nonnegative functions  $\rho_n \in C^1(\mathbb{R}^n)$  with support in  $\{\eta \in \mathbb{R}^n : |\eta|_{\mathbb{R}^n} \leq 1/n\}$  such that  $\int_{\mathbb{R}^n} \rho_n(\eta) d\eta = 1$ . Then we set

$$F_n(t, x) = \int_{\mathbb{R}^n} F\left(t, \sum_{i=1}^n \eta_i e_i\right) \rho_n(\langle x, e_1 \rangle - \eta_1, \dots, \langle x, e_n \rangle - \eta_n) d\eta_1 \dots d\eta_n,$$

$$G_n(t, x)\xi = \int_{\mathbb{R}^n} G\left(t, \sum_{i=1}^n \eta_i e_i\right) \xi \rho_n(\langle x, e_1 \rangle - \eta_1, \dots, \langle x, e_n \rangle - \eta_n) d\eta_1 \dots d\eta_n, \quad \xi \in \Xi,$$

where  $\{e_k\}$  is an orthonormal basis of  $H$ . Then it is easy to verify that  $F_n$  and  $G_n$  satisfy the required conditions. Now let us denote by  $X^n$  the mild solution of the equation

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

Then standard estimates and the Gronwall Lemma imply that  $\sup_{\tau \in [t, T]} \mathbb{E}|X_\tau^n - X_\tau|^2 \rightarrow 0$ , for  $n \rightarrow \infty$ . Let us set  $X_\tau^n = x$  for  $\tau \in [0, t)$ . By Proposition 3.5 in [18],  $X^n$  belongs to  $\mathbb{L}^{1,2}(H)$  and there exists a version of  $DX^n$  such that for every  $s \in [0, T)$ ,  $\{D_s X_\tau^n, \tau \in (s, T]\}$  is a predictable process in  $L_2(\Xi, H)$  with continuous paths satisfying,  $\mathbb{P}$ -a.s.

$$\begin{aligned} D_s X_\tau^n &= e^{(\tau-s)A}G_n(s, X_s^n) + \int_s^\tau e^{(\tau-\sigma)A}\nabla F_n(\sigma, X_\sigma^n)D_s X_\sigma^n d\sigma \\ &\quad + \int_s^\tau \nabla(e^{(\tau-\sigma)A}G_n(\sigma, X_\sigma^n))D_s X_\sigma^n dW_\sigma, \quad \tau \in (s, T], \end{aligned} \quad (4.4)$$

and, for every  $p \in [2, \infty)$ ,

$$\sup_{s \in [0, T]} \mathbb{E} \left( \sup_{\tau \in (s, T]} (\tau - s)^{p\gamma} |D_s X_\tau^n|_{L_2(\Xi, H)}^p \right) \leq c, \quad (4.5)$$

where  $c > 0$  depends only on  $p, L, T, \gamma$  and  $M = \sup_{\tau \in [0, T]} |e^{\tau A}|$ . We will use only one consequence of (4.5), namely that it implies

$$\mathbb{E} \left( \int_0^T \int_0^T |D_s X_\tau^n|_{L_2(\Xi, H)}^2 ds d\tau \right)^q \leq c_q, \quad (4.6)$$

for every  $q \in [1, \infty)$  and some constant  $c_q$  independent of  $n$ . It follows in particular that the norm of  $DX^n$  in the space  $L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$  is bounded by a constant independent of  $n$ . Since  $X^n$  converges to  $X$  in  $L^2(\Omega \times [0, T]; H)$ , it follows from the closedness of the operator  $D$  that  $X$  belongs to  $\mathbb{L}^{1,2}(H)$  and that  $X^n$  converges to  $X$  weakly in  $\mathbb{L}^{1,2}(H)$ . Point (i) of the Proposition 4.1 is now proved.

The estimate (4.6) also shows that the sequence  $(DX^n)$  is bounded by the constant  $c_q$  in the space  $L^{2q}(\Omega; L^2([0, T] \times [0, T]; L_2(\Xi, H)))$ . By the weak lower semicontinuity of the norm, any weak limit point also satisfies the same inequality. But clearly any weak limit point in  $L^{2q}(\Omega; L^2([0, T] \times [0, T]; L_2(\Xi, H)))$  is also a weak limit point in the space  $L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$  and must therefore coincide with  $DX$ . This proves point (ii) of the Proposition.

Applying both sides of (4.4) to  $\xi \in \Xi$  and recalling (4.3) we obtain

$$\mathbb{E} |D_s X_\tau^n \xi|^2 \leq c(1 + \mathbb{E} |X_s^n|)^2 |\xi|^2 + c \int_s^\tau \mathbb{E} |D_s X_\sigma^n \xi|^2 d\sigma + c \int_s^\tau (\tau - \sigma)^{-2\gamma} \mathbb{E} |D_s X_\sigma^n \xi|^2 d\sigma,$$



and since  $\mathbb{E}|X_s^n|^2$  is bounded uniformly with respect to  $s$  and  $n$  by Proposition 2.3, we conclude by the Gronwall Lemma that

$$\mathbb{E}|D_s X_\tau^n \xi|^2 \leq c|\xi|^2, \quad 0 \leq s \leq \tau \leq T, \quad \xi \in \Xi, \quad (4.7)$$

with  $c$  independent of  $n$ .

Next we fix  $T' \in [t, T)$  and consider  $\epsilon > 0$  so small that  $T' + \epsilon \leq T$ . From (4.4) we obtain, for  $\tau \in [t, T']$ ,

$$\begin{aligned} \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} D_s X_{\tau+\epsilon}^n \xi \, ds &= \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} e^{(\tau+\epsilon-s)A} G_n(s, X_s^n) \xi \, ds \\ &\quad + \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} \int_s^{\tau+\epsilon} e^{(\tau+\epsilon-\sigma)A} \nabla F_n(\sigma, X_\sigma^n) D_s X_\sigma^n \xi \, d\sigma \, ds \\ &\quad + \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} \int_s^{\tau+\epsilon} \nabla(e^{(\tau+\epsilon-\sigma)A} G_n(\sigma, X_\sigma^n)) D_s X_\sigma^n \xi \, dW_\sigma \, ds, \\ &=: J_\tau^{1,\epsilon,n} + J_\tau^{2,\epsilon,n} + J_\tau^{3,\epsilon,n}, \quad \tau \in [t, T']. \end{aligned} \quad (4.8)$$

We note that

$$|J_\tau^{3,\epsilon,n}|^2 \leq \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} \left| \int_s^{\tau+\epsilon} \nabla(e^{(\tau+\epsilon-\sigma)A} G_n(\sigma, X_\sigma^n)) D_s X_\sigma^n \xi \, dW_\sigma \right|^2 ds,$$

and recalling (4.3) and (4.7) we obtain

$$\begin{aligned} \mathbb{E}|J_\tau^{3,\epsilon,n}|^2 &\leq \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} \int_s^{\tau+\epsilon} \mathbb{E} |\nabla(e^{(\tau+\epsilon-\sigma)A} G_n(\sigma, X_\sigma^n)) D_s X_\sigma^n \xi|^2 \, d\sigma \, ds \\ &\leq \frac{c|\xi|^2}{\epsilon} \int_\tau^{\tau+\epsilon} \int_s^{\tau+\epsilon} (\tau + \epsilon - \sigma)^{-2\gamma} \, d\sigma \, ds \\ &\leq \frac{c|\xi|^2}{\epsilon} \epsilon^{1-2\gamma} \end{aligned}$$

with  $c$  independent of  $n$  and  $\tau$ .

Similarly we obtain  $\mathbb{E}|J_\tau^{2,\epsilon,n}|^2 \leq c|\xi|^2 \epsilon^2$ .

Since  $\sup_{\tau \in [t, T]} \mathbb{E}|X_\tau^n - X_\tau|^2 \rightarrow 0$  for  $n \rightarrow \infty$  it follows easily that for every  $\tau \in [t, T']$

$$J_\tau^{1,\epsilon,n} \rightarrow \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} e^{(\tau+\epsilon-s)A} G(s, X_s) \xi \, ds$$

in  $L^2(\Omega; H)$ . The convergence also takes place in  $L^2(\Omega \times [t, T']; H)$ , by the dominated convergence theorem.

Finally, since  $DX^n$  converges weakly to  $DX$  in  $L^2(\Omega \times [0, T] \times [0, T]; L_2(\Xi, H))$ , the process  $\{\int_\tau^{\tau+\epsilon} D_s X_{\tau+\epsilon}^n \xi \, ds, \tau \in [t, T']\}$  converges weakly in  $L^2(\Omega \times [t, T']; L_2(\Xi, H))$  to the process  $\{\int_\tau^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi \, ds, \tau \in [t, T']\}$ : this follows at once from the identity

$$\begin{aligned} \mathbb{E} \int_t^{T'} \left\langle \int_\tau^{\tau+\epsilon} D_s X_{\tau+\epsilon}^n \xi \, ds, Y_\tau \right\rangle_{L_2(\Xi, H)} d\tau \\ = \mathbb{E} \int_0^T \int_0^T \langle D_s X_\tau^n \xi, Y_{\tau-\epsilon} 1_{(t+\epsilon, T'+\epsilon)}(\tau) 1_{(\tau-\epsilon, \tau)}(s) \rangle_{L_2(\Xi, H)} d\tau \, ds, \end{aligned}$$

which holds, by the Fubini theorem, for every  $Y \in L^2(\Omega \times [t, T']; L_2(\Xi, H))$ .

Taking into account the lower semicontinuity of the norm in the space  $L^2(\Omega \times [t, T']; L_2(\Xi, H))$  and letting  $n \rightarrow \infty$  in (4.8) we obtain

$$\mathbb{E} \int_t^{T'} \left| \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi \, ds - \frac{1}{\epsilon} \int_\tau^{\tau+\epsilon} e^{(\tau+\epsilon-s)A} G(s, X_s) \xi \, ds \right|^2 d\tau \leq c|\xi|^2 (\epsilon^{1-2\gamma} + \epsilon^2).$$

Next we note that, for  $\epsilon \rightarrow 0$ ,

$$\begin{aligned} \frac{1}{\epsilon} \int_{\tau}^{\tau+\epsilon} e^{(\tau+\epsilon-s)A} G(s, X_s) \xi \, ds - G(\tau, X_{\tau}) \xi &= \frac{1}{\epsilon} \int_{\tau}^{\tau+\epsilon} (e^{(\tau+\epsilon-s)A} - I) G(\tau, X_{\tau}) \xi \, ds \\ &+ \frac{1}{\epsilon} \int_{\tau}^{\tau+\epsilon} e^{(\tau+\epsilon-s)A} (G(s, X_s) \xi - G(\tau, X_{\tau}) \xi) \, ds \end{aligned}$$

tends to 0 for a.e.  $\tau$ ,  $\mathbb{P}$ -a.s., by the strong continuity of the semigroup and the Lebesgue differentiation theorem. Convergence also takes place in  $L^2(\Omega \times [t, T']; H)$ , by the dominated convergence theorem. Thus (4.2) follows and the proof of Proposition 4.1 is completed.  $\square$

## 4.2 Computing a joint quadratic variation.

We need an easy approximation lemma.

**Lemma 4.3** *If  $u : [0, T] \times H \rightarrow \mathbb{R}$  is a measurable function satisfying (3.1), there exist measurable mappings  $u_n : [0, T] \times H \rightarrow \mathbb{R}$ ,  $n = 1, 2, \dots$ , such that*

(i) *For some  $C > 0, r \geq 0$  and for every  $t \in [0, T]$ ,  $x, y \in H$  and  $n = 1, 2, \dots$*

$$|u_n(t, x) - u_n(t, y)| \leq C|x - y|(1 + |x| + |y|)^r, \quad |u_n(t, 0)| \leq C.$$

(ii) *For  $t \in [0, T]$ , the mappings  $u_n(t, \cdot) : H \rightarrow \mathbb{R}$  are continuously Fréchet differentiable, bounded together with their derivative, uniformly with respect to  $t$ . We denote by  $\nabla u_n(t, x)$  the Fréchet derivative at point  $x \in H$ . We will consider  $\nabla u_n(t, x)$  as an element of  $H^*$ .*

(iii) *For  $x \in H$ ,  $t \in [0, T]$ , we have  $u_n(t, x) \rightarrow u(t, x)$ .*

**Proof.** Just notice that functions  $u_n$  can be constructed by the same procedure used to define the approximating sequences  $F_n, G_n$  in the proof of Proposition 4.1 (and multiplied by appropriate cut-off functions).  $\square$

**Proposition 4.4** *Assume that Hypothesis 2.2 holds and that  $u : [0, T] \times H \rightarrow \mathbb{R}$  is a measurable function satisfying (3.1). For every  $t \in [0, T]$ ,  $x \in H$  the following holds.*

(i) *For  $t \leq T' < T$  and  $\xi \in \Xi$  the processes  $\{u(\tau, X(\tau, t, x)), \tau \in [t, T]\}$  and  $W^\xi$  admit a joint quadratic variation on the interval  $[t, T']$ .*

(ii) *There exists a process  $\{U_\tau^{t,x}, \tau \in [t, T]\}$ , with values in  $\Xi^*$ , predictable with respect to the filtration  $\{\mathcal{F}_{[t,\tau]}, \tau \in [t, T]\}$ , satisfying  $\mathbb{E} \int_t^{T'} |U_\tau^{t,x}|^p \, d\tau < \infty$  for every  $p \in [2, \infty)$  and such that, for  $t \leq T' < T$  and  $\xi \in \Xi$ ,*

$$\langle u(\cdot, X(\cdot, t, x)), W^\xi \rangle_{[t, T']} = \int_t^{T'} U_\tau^{t,x} G(\tau, X(\tau, t, x)) \xi \, d\tau, \quad \mathbb{P} - a.s.$$

**Proof.** As a consequence of point(i) in Lemma 4.3 we obtain the following inequalities:

$$|u_n(t, x)| \leq c(1 + |x|)^{r+1}, \quad |\nabla u_n(t, x)|_{H^*} \leq c(1 + |x|)^r, \quad t \in [0, T], \, x \in H, \quad (4.9)$$

with  $c$  independent of  $n$ .

Now we fix  $t \in [0, T]$ ,  $x \in H$  and we denote  $X(\tau, t, x)$  by  $X_\tau$  for simplicity. We also set  $X_\tau = x$  for  $\tau \in [0, t]$ . It follows from (4.9) and Proposition 2.3 that

$$\sup_n \mathbb{E} \sup_{\tau \in [0, T]} (|u_n(\tau, X_\tau)|^p + |\nabla u_n(\tau, X_\tau)|_{H^*}^p) < \infty \quad (4.10)$$

for every  $p \in [2, \infty)$ . By the chain rule for the Malliavin derivative,  $u_n(\cdot, X_\cdot) \in \mathbb{L}^{1,2}(H)$  and  $D_s(u_n(\tau, X_\tau)) = \nabla u_n(\tau, X_\tau) D_s X_\tau$ . By (4.1), (4.10) and Hölder's inequality, the sequence  $u_n(\cdot, X_\cdot)$  is bounded in  $\mathbb{L}^{1,2}(\mathbb{R})$ . Moreover it converges to  $u(\cdot, X_\cdot)$  in  $L^2(\Omega \times [0, T]; \mathbb{R})$ . It follows from the closedness of the operator  $D$  that  $u(\cdot, X_\cdot)$  belongs to  $\mathbb{L}^{1,2}(\mathbb{R})$  and  $u_n(\cdot, X_\cdot)$  converges to  $u(\cdot, X_\cdot)$  weakly in  $\mathbb{L}^{1,2}(\mathbb{R})$ .

After these preliminaries we also fix  $\xi \in \Xi$  and  $T' \in [t, T)$  and we prove existence of the joint quadratic variation of  $u(\cdot, X_\cdot)$  and  $W^\xi$  on the interval  $[t, T']$ . We first re-write in an appropriate way the expression

$$C_{[t, T']}^\epsilon(u_n(\cdot, X_\cdot), W^\xi) = \frac{1}{\epsilon} \int_t^{T'} (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))(W_{\tau+\epsilon}^\xi - W_\tau^\xi) d\tau,$$

with  $\epsilon > 0$  so small that  $T' + \epsilon \leq T$ . We first explain our argument by writing down some informal passages: by the rules of Malliavin calculus we have, for a.a.  $\tau \in [t, T']$ ,

$$\begin{aligned} (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))(W_{\tau+\epsilon}^\xi - W_\tau^\xi) &= (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))\xi^* \int_\tau^{\tau+\epsilon} dW_s \\ &= \int_\tau^{\tau+\epsilon} D_s(u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))\xi ds + \int_\tau^{\tau+\epsilon} (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))\xi^* \hat{d}W_s, \end{aligned} \quad (4.11)$$

where the symbol  $\hat{d}W$  denotes the Skorohod integral, and by  $\xi^*$  we denote the element in  $\Xi^*$  corresponding to  $\xi$  by the Riesz isometry. Integrating over  $[t, T']$  with respect to  $\tau$  and interchanging integrals gives

$$\begin{aligned} \epsilon C_{[t, T']}^\epsilon(u_n(\cdot, X_\cdot), W^\xi) &= \int_t^{T'} \int_\tau^{\tau+\epsilon} D_s(u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))\xi ds d\tau \\ &\quad + \int_t^{T'} \int_{(s-\epsilon) \vee t}^{s \wedge T'} (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))\xi^* d\tau \hat{d}W_s. \end{aligned} \quad (4.12)$$

To justify (4.11) and (4.12) rigorously we proceed as follows. To shorten notation we define

$$y_\tau = (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau)) 1_{[0, T']}(\tau), \quad \tau \in [0, T],$$

$$A^\epsilon = \{(\tau, s) \in [0, T] \times [0, T] : t \leq \tau \leq T', \tau \leq s \leq \tau + \epsilon\}.$$

We note that  $y$  is a process in  $\mathbb{L}^{1,2}(\mathbb{R})$  and it is bounded; consequently, for a.a.  $\tau$ ,  $y_\tau \xi^*$  belongs to  $\mathbb{D}^{1,2}(\Xi^*)$  and the process  $y_\tau \xi^* 1_{A^\epsilon}(\tau, \cdot)$  belongs to  $L^2(\Omega \times [0, T]; \Xi^*)$ . By [20] Proposition 2.11 (which extends the finite-dimensional result of [26] Theorem 3.2, see also [25] Section 1.3.1 (2)) we conclude that  $y_\tau \xi^* 1_{A^\epsilon}(\tau, \cdot)$  is Skorohod integrable and the formula

$$\int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, s) \hat{d}W_s = y_\tau \xi^* \int_0^T 1_{A^\epsilon}(\tau, s) \hat{d}W_s - \int_0^T D_s y_\tau \xi 1_{A^\epsilon}(\tau, s) ds =: z_\tau, \quad (4.13)$$

holds provided  $z_\tau$  belongs to  $L^2(\Omega)$  for a.a.  $\tau$ . Since  $\xi^* \int_0^T 1_{A^\epsilon}(\tau, s) \hat{d}W_s$  coincides with the Ito integral  $\int_0^T 1_{A^\epsilon}(\tau, s) dW_s^\xi = (W_{\tau+\epsilon}^\xi - W_\tau^\xi) 1_{[t, T']}(\tau)$ , it is in fact easy to verify that we even have  $z \in L^2(\Omega \times [0, T])$ ; thus (4.13) holds. We note that (4.13) yields (4.11).

Next we wish to show that the process  $\int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, \cdot) d\tau$  is Skorohod integrable and to compute its integral, which occurs in the right-hand side of (4.12). For arbitrary  $G \in \mathbb{D}^{1,2}(\mathbb{R})$ ,

by the definition of the Skorohod integral and by (4.13),

$$\begin{aligned}
\mathbb{E} \int_0^T \left\langle \int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, s) d\tau, D_s G \right\rangle_{\Xi^*} ds &= \int_0^T \mathbb{E} \int_0^T \langle y_\tau \xi^* 1_{A^\epsilon}(\tau, s), D_s G \rangle_{\Xi^*} ds d\tau \\
&= \int_0^T \mathbb{E} \left[ G \int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, s) d\hat{W}_s \right] d\tau \\
&= \mathbb{E} \left[ G \int_0^T z_\tau d\tau \right].
\end{aligned}$$

This shows, by definition, that  $\int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, \cdot) d\tau$  is Skorohod integrable and

$$\int_0^T \int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, s) d\tau d\hat{W}_s = \int_0^T z_\tau d\tau = \int_0^T \int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, s) d\hat{W}_s d\tau.$$

Recalling (4.13) we obtain

$$\int_0^T \int_0^T y_\tau \xi^* 1_{A^\epsilon}(\tau, s) d\tau d\hat{W}_s = \int_0^T y_\tau (W_{\tau+\epsilon}^\xi - W_\tau^\xi) 1_{[t, T']}(s) d\tau - \int_0^T \int_0^T D_s y_\tau \xi 1_{A^\epsilon}(\tau, s) ds d\tau,$$

and (4.12) is proved.

Recalling the definition of  $y$ , noting that  $D_s(u_n(\tau, X_\tau)) = 0$  for  $s > \tau$  by adaptedness, and using the chain rule for the Malliavin derivative we have, for a.a.  $s, \tau$  with  $s \in [\tau, \tau + \epsilon]$ ,

$$D_s y_\tau \xi = D_s(u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau))\xi = D_s(u_n(\tau + \epsilon, X_{\tau+\epsilon}))\xi = \nabla u_n(\tau + \epsilon, X_{\tau+\epsilon}) D_s X_{\tau+\epsilon} \xi$$

and we obtain

$$\begin{aligned}
&\int_0^T \int_0^T (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau)) \xi^* 1_{A^\epsilon}(\tau, s) d\tau d\hat{W}_s \\
&= \int_t^{T'} (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau)) (W_{\tau+\epsilon}^\xi - W_\tau^\xi) d\tau \\
&\quad - \int_0^T \int_0^T \nabla u_n(\tau + \epsilon, X_{\tau+\epsilon}) D_s X_{\tau+\epsilon} \xi 1_{A^\epsilon}(\tau, s) ds d\tau. \\
&= \epsilon C_{[t, T']}^\epsilon(u_n(\cdot, X_\cdot), W^\xi) - \int_t^{T'} \nabla u_n(\tau + \epsilon, X_{\tau+\epsilon}) \int_\tau^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi ds d\tau.
\end{aligned} \tag{4.14}$$

Now we let  $n \rightarrow \infty$ . Recalling (4.1) and (4.10) it is easy to prove that the right-hand side of (4.14) is bounded in  $L^2(\Omega)$  by a constant independent of  $n$ . Thus the Skorohod integral of the process  $\int_0^T (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau)) \xi^* 1_{A^\epsilon}(\tau, \cdot) d\tau$  is bounded uniformly with respect to  $n$ . Since this process also converges to the process  $\int_0^T (u(\tau + \epsilon, X_{\tau+\epsilon}) - u(\tau, X_\tau)) \xi^* 1_{A^\epsilon}(\tau, \cdot) d\tau$ , in  $L^2(\Omega \times [0, T]; \Xi^*)$ , it follows from the closedness of the Skorohod integral operator that

$$\begin{aligned}
&\int_0^T \int_0^T (u_n(\tau + \epsilon, X_{\tau+\epsilon}) - u_n(\tau, X_\tau)) \xi^* 1_{A^\epsilon}(\tau, s) d\tau d\hat{W}_s \\
&\rightarrow \int_0^T \int_0^T (u(\tau + \epsilon, X_{\tau+\epsilon}) - u(\tau, X_\tau)) \xi^* 1_{A^\epsilon}(\tau, s) d\tau d\hat{W}_s
\end{aligned}$$

weakly in  $L^2(\Omega)$ .

Next we note that the convergence  $u_n(\cdot, X_\cdot) \rightarrow u(\cdot, X_\cdot)$  in  $L^2(\Omega \times [0, T]; \mathbb{R})$  easily implies  $C_{[t, T]}^\epsilon(u_n(\cdot, X_\cdot), W^\xi) \rightarrow C_{[t, T]}^\epsilon(u(\cdot, X_\cdot), W^\xi)$  in  $L^1(\Omega)$ .

Finally by (4.10) the sequence  $\nabla u_n(\cdot, X_\cdot)$  is bounded in  $L^p(\Omega \times [0, T]; H^*)$  for every  $p \in [2, \infty)$ , so that an appropriate subsequence converges weakly in this space to a limit that we

denote by  $U^{t,x}$ . It follows that  $\nabla u_n(\cdot + \epsilon, X_{\cdot+\epsilon}) \rightarrow U^{t,x}_{\cdot+\epsilon}$  weakly in  $L^p(\Omega \times [t, T']; H^*)$ . Since  $DX\xi \in L^2(\Omega \times [0, T] \times [0, T]; H)$ , therefore the process  $\{\int_{\tau}^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi ds, \tau \in [t, T']\}$  belongs to the space  $L^{p'}(\Omega \times [t, T']; H)$  (where  $1/p' = 1 - 1/p$ ) and consequently

$$\int_t^{T'} \nabla u_n(\tau + \epsilon, X_{\tau+\epsilon}) \int_{\tau}^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi ds d\tau \rightarrow \int_t^{T'} U^{t,x}_{\tau+\epsilon} \int_{\tau}^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi ds d\tau$$

weakly in  $L^1(\Omega)$ . Letting  $n \rightarrow \infty$  in (4.14) along a subsequence we obtain

$$\begin{aligned} C_{[t, T']}^{\epsilon}(u(\cdot, X_{\cdot}), W^{\xi}) &= \int_t^{T'} U^{t,x}_{\tau+\epsilon} \frac{1}{\epsilon} \int_{\tau}^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi ds d\tau \\ &+ \int_0^T \frac{1}{\epsilon} \int_0^T (u(\tau + \epsilon, X_{\tau+\epsilon}) - u(\tau, X_{\tau})) \xi^* 1_{A^{\epsilon}}(\tau, s) d\tau \hat{d}W_s. \end{aligned}$$

Now we let  $\epsilon \rightarrow 0$ . First we note that  $U^{t,x}_{\cdot+\epsilon}$  converges to  $U^{t,x}$  in  $L^2(\Omega \times [t, T']; H^*)$  and it follows from Proposition 4.1 that

$$\int_t^{T'} U^{t,x}_{\tau+\epsilon} \frac{1}{\epsilon} \int_{\tau}^{\tau+\epsilon} D_s X_{\tau+\epsilon} \xi ds d\tau \rightarrow \int_t^{T'} U^{t,x}_{\tau} G(\tau, X_{\tau}) \xi d\tau$$

in  $L^1(\Omega)$ .

Next we claim that the process  $\frac{1}{\epsilon} \int_0^T (u(\tau + \epsilon, X_{\tau+\epsilon}) - u(\tau, X_{\tau})) 1_{A^{\epsilon}}(\tau, \cdot) d\tau$  converges to 0 in  $\mathbb{L}^{1,2}(\mathbb{R})$ .

Admitting the claim for a moment, since the Skorohod integral is a bounded linear operator from  $\mathbb{L}^{1,2}(\Xi^*)$  to  $L^2(\Omega)$ , it follows that

$$\int_0^T \frac{1}{\epsilon} \int_0^T (u(\tau + \epsilon, X_{\tau+\epsilon}) - u(\tau, X_{\tau})) \xi^* 1_{A^{\epsilon}}(\tau, s) d\tau \hat{d}W_s \rightarrow 0$$

in  $L^2(\Omega)$ . This shows that  $C_{[t, T']}^{\epsilon}(u(\cdot, X_{\cdot}), W^{\xi})$  converges in probability and

$$\langle u(\cdot, X(\cdot, t, x)), W^{\xi} \rangle_{[t, T']} = \int_t^{T'} U^{t,x}_{\tau} G(\tau, X_{\tau}) \xi d\tau, \quad \mathbb{P} - a.s. \quad (4.15)$$

The required summability condition  $\mathbb{E} \int_t^T |U^{t,x}_{\tau}|^p d\tau < \infty$  is immediate since by definition  $U^{t,x}$  belongs to  $L^p(\Omega \times [0, T]; H^*)$  and  $p$  is arbitrary. To prove predictability of  $U^{t,x}$  with respect to the filtration  $\{\mathcal{F}_{[t, \tau]}, \tau \in [t, T]\}$  we first note that the restriction of  $U^{t,x}$  to  $\Omega \times [t, T]$  is a weak limit point of  $\nabla u_n(\cdot, X(\cdot, t, x))$  in the space  $L^p(\Omega \times [t, T]; H^*)$ . Next we denote by  $l$  the Lebesgue measure, we consider the finite measure space  $(\Omega \times [t, T], \mathcal{F} \times \mathcal{B}([t, T]), \mathbb{P} \times l)$  and we denote by  $\mathcal{P}_{tT} \subset \mathcal{F} \times \mathcal{B}([t, T])$  the predictable  $\sigma$ -algebra relative to the filtration  $\{\mathcal{F}_{[t, \tau]}, \tau \in [t, T]\}$ , and by  $\mathbb{E}^{\mathcal{P}_{tT}}$  the corresponding conditional expectation. Since  $\nabla u_n(\cdot, X(\cdot, t, x))$  is  $\mathcal{P}_{tT}$ -measurable we have  $\mathbb{E}^{\mathcal{P}_{tT}} \nabla u_n(\cdot, X(\cdot, t, x)) = \nabla u_n(\cdot, X(\cdot, t, x))$ ,  $\mathbb{P} \times l$ -a.s. Since  $\mathbb{E}^{\mathcal{P}_{tT}}$  is a bounded linear operator in  $L^p(\Omega \times [t, T]; H^*)$ , it is also weakly continuous, and therefore  $\mathbb{E}^{\mathcal{P}_{tT}} U^{t,x} = U^{t,x}$ ,  $\mathbb{P} \times l$ -a.s. Thus  $U^{t,x}$  can be made predictable after modification on a set of zero  $\mathbb{P} \times l$ -measure and this does not affect equality (4.15).

To conclude the proof of Proposition 4.4 it remains to show that the process  $\frac{1}{\epsilon} \int_0^T (u(\tau + \epsilon, X_{\tau+\epsilon}) - u(\tau, X_{\tau})) 1_{A^{\epsilon}}(\tau, \cdot) d\tau$  converges to 0 in  $\mathbb{L}^{1,2}(\mathbb{R})$ . We prove, more generally, that for an arbitrary element  $y \in \mathbb{L}^{1,2}(\mathbb{R})$ , if we set

$$T^{\epsilon}(y)_s = \frac{1}{\epsilon} \int_0^T (y_{\tau+\epsilon} - y_{\tau}) 1_{A^{\epsilon}}(\tau, s) d\tau = \frac{1}{\epsilon} \int_{(s-\epsilon) \vee t}^{s \wedge T'} (y_{\tau+\epsilon} - y_{\tau}) d\tau, \quad s \in [0, T],$$

then the process  $T^\epsilon(y)$  converges to 0 in  $\mathbb{L}^{1,2}(\mathbb{R})$ . Let us recall that  $\mathbb{L}^{1,2}(\mathbb{R})$  is isomorphic to  $L^2([0, T]; \mathbb{D}^{1,2}(\mathbb{R}))$ . It is clear that  $T^\epsilon(y) \rightarrow 0$  if  $y$  belongs to  $C([0, T]; \mathbb{D}^{1,2}(\mathbb{R}))$ , a dense subspace of  $L^2([0, T]; \mathbb{D}^{1,2}(\mathbb{R}))$ . So to prove the claim it is enough to show that the norm of  $T^\epsilon$ , as an operator on  $\mathbb{L}^{1,2}(\mathbb{R})$ , is bounded uniformly with respect to  $\epsilon$ . We have

$$\begin{aligned} |T^\epsilon(y)_s|_{\mathbb{D}^{1,2}(\mathbb{R})}^2 &\leq \frac{1}{\epsilon^2} \int_0^T 1_{A^\epsilon}(\tau, s) d\tau \int_0^T |y_{\tau+\epsilon} - y_\tau|_{\mathbb{D}^{1,2}(\mathbb{R})}^2 1_{A^\epsilon}(\tau, s) d\tau \\ &\leq \frac{1}{\epsilon} \int_0^T |y_{\tau+\epsilon} - y_\tau|_{\mathbb{D}^{1,2}(\mathbb{R})}^2 1_{A^\epsilon}(\tau, s) d\tau, \\ |T^\epsilon(y)|_{\mathbb{L}^{1,2}(\mathbb{R})}^2 &= \int_0^T |T^\epsilon(y)_s|_{\mathbb{D}^{1,2}(\mathbb{R})}^2 ds \\ &\leq \frac{1}{\epsilon} \int_0^T |y_{\tau+\epsilon} - y_\tau|_{\mathbb{D}^{1,2}(\mathbb{R})}^2 \int_0^T 1_{A^\epsilon}(\tau, s) ds d\tau \\ &\leq \int_0^{T'} |y_{\tau+\epsilon} - y_\tau|_{\mathbb{D}^{1,2}(\mathbb{R})}^2 d\tau \\ &\leq 2|y|_{\mathbb{L}^{1,2}(\mathbb{R})}^2. \end{aligned}$$

This shows the required bound. The claim is proved, and so is the proposition.  $\square$

In the following lemmas we keep the same assumptions as in the previous proposition. We also set  $\Delta = \{(t, \tau) : 0 \leq t \leq \tau < T\}$  and by  $\mathcal{B}(\Lambda)$  we denote the Borel  $\sigma$ -algebra of any topological space  $\Lambda$ .

**Lemma 4.5** *For every  $\xi \in \Xi$ , the process  $\{\langle u(\cdot, X(\cdot, t, x)), W^\xi \rangle_{[t, \tau]}, (t, \tau) \in \Delta, x \in H\}$  has a modification, denoted  $\{A^\xi(\tau, t, x), (t, \tau) \in \Delta, x \in H\}$ , with the following properties:*

- (i)  $A^\xi$  is measurable with respect to  $\mathcal{F} \times \mathcal{B}(\Delta) \times \mathcal{B}(H)$ .
- (ii) for  $(t, \tau) \in \Delta$ , the map  $(x, \omega) \rightarrow A^\xi(\tau, t, x, \omega)$  is measurable with respect to  $\mathcal{B}(H) \times \mathcal{F}_{[t, T]}$ .

**Proof.** We adapt a classical argument to prove joint measurability of limits in probability (see, for instance, [32], chapter IV, exercise (5.17)). We recall the definition

$$C_{[t, \tau]}^\epsilon(u(\cdot, X(\cdot, t, x)), W^\xi) = \frac{1}{\epsilon} \int_t^\tau [u(\sigma + \epsilon, X(\sigma + \epsilon, t, x)) - u(\sigma, X(\sigma, t, x))] [W_{\sigma+\epsilon}^\xi - W_\sigma^\xi] d\sigma,$$

and we define, for  $0 \leq t \leq \tau < T$ ,  $x \in H$ ,  $\omega \in \Omega$ ,

$$f_n(\tau, t, x, \omega) = C_{[t, \tau]}^{1/n}(u(\cdot, X(\cdot, t, x)), W^\xi)(\omega).$$

We construct inductively an integer-valued sequence  $(n_k)$  setting  $n_1 = 1$ ,

$$n_{k+1} = \min\{n > n_k : \mathbb{P}(|f_m(\tau, t, x) - f_l(\tau, t, x)| > 2^{-k}) < 2^{-k}, \text{ for every } m \geq n, l \geq n\}.$$

Each  $n_k$  depends in fact on  $\tau, t, x$  and we will write  $n_k(\tau, t, x)$  to stress this dependence. We define

$$A^\xi(\tau, t, x, \omega) = \lim_{k \rightarrow \infty} f_{n_k(\tau, t, x)}(\tau, t, x, \omega)$$

on the set of points  $(\tau, t, x, \omega)$  where the limit exists, and  $A^\xi(\tau, t, x, \omega) = 0$  otherwise. We note that the limit exists  $\mathbb{P}$ -a.s. for any choice of  $\tau, t, x$ , by the definition of  $(n_k)$  and the Borel-Cantelli lemma. Since  $f_n(\tau, t, x)$  tends in probability to  $\langle u(\cdot, X(\cdot, t, x)), W^\xi \rangle_{[t, \tau]}$ , it follows that for  $0 \leq t \leq \tau < T$  and  $x \in H$ ,

$$A^\xi(\tau, t, x) = \langle u(\cdot, X(\cdot, t, x)), W^\xi \rangle_{[t, \tau]}, \quad \mathbb{P} - a.s.$$

This proves that  $A^\xi$  is a modification of the quadratic variation process.

The formula for  $C_{[t,\tau]}^\epsilon(u(\cdot, X(\cdot, t, x)), W^\xi)(\omega)$  shows that the maps  $(t, \tau, x, \omega) \rightarrow f_n(t, \tau, x, \omega)$  are measurable with respect to  $\mathcal{B}(\Delta) \times \mathcal{B}(H) \times \mathcal{F}$ . It is easy to show that this entails that  $(t, \tau, x) \rightarrow n_k(t, \tau, x)$  are measurable with respect to  $\mathcal{B}(\Delta) \times \mathcal{B}(H)$ . We conclude that  $A^\xi$  is also measurable with respect to  $\mathcal{B}(\Delta) \times \mathcal{B}(H) \times \mathcal{F}$  as a pointwise limit of measurable functions.

Now we fix  $(t, \tau) \in \Delta$ . The formula for  $C_{[t,\tau]}^\epsilon(u(\cdot, X(\cdot, t, x)), W^\xi)(\omega)$  shows that the maps  $(x, \omega) \rightarrow f_n(t, \tau, x, \omega)$  are measurable with respect to  $\mathcal{B}(H) \times \mathcal{F}_{[t,T]}$ . Since  $x \rightarrow n_k(t, \tau, x)$  are measurable with respect to  $\mathcal{B}(H)$  we conclude that  $(x, \omega) \rightarrow A^\xi(t, \tau, x, \omega)$  is also measurable with respect to  $\mathcal{B}(H) \times \mathcal{F}_{[t,T]}$  as a pointwise limit of measurable functions.  $\square$

The process  $A^\xi$  is an additive functional of the (nonhomogeneous) Markov process  $X$ , in the sense specified by the following result.

**Lemma 4.6** *For  $\xi \in \Xi$ ,  $0 \leq t \leq s \leq \tau < T$ ,  $x \in H$  we have*

$$A^\xi(\tau, t, x) = A^\xi(s, t, x) + A^\xi(\tau, s, X(s, t, x)), \quad \mathbb{P} - a.s.$$

**Proof.** Let us define, for  $y \in H$  and  $\omega \in \Omega$ ,

$$q^\epsilon(y, \omega) = |A^\xi(\tau, s, y, \omega) - C_{[s,\tau]}^\epsilon(u(\cdot, X(\cdot, s, y)), W^\xi)(\omega)|.$$

Then by the definition of  $A^\xi$  we have  $q^\epsilon(y, \cdot) \rightarrow 0$  in probability as  $\epsilon \rightarrow 0$ , or equivalently

$$\mathbb{E}(q^\epsilon(y, \cdot) \wedge 1) \rightarrow 0, \quad y \in H.$$

By Lemma 4.5  $(y, \omega) \rightarrow q^\epsilon(y, \omega)$  is measurable with respect to  $\mathcal{B}(H) \times \mathcal{F}_{[s,T]}$  and the  $\sigma$ -algebras  $\mathcal{F}_{[s,T]}$  and  $\mathcal{F}_s$  are independent. Thus for arbitrary,  $\mathcal{F}_s$ -measurable  $\eta : \Omega \rightarrow H$ , denoting by  $P_\eta$  the law of  $\eta$ , we have

$$\mathbb{E}(q^\epsilon(\eta(\cdot), \cdot) \wedge 1) = \int_H \mathbb{E}(q^\epsilon(y, \cdot) \wedge 1) P_\eta(dy) \rightarrow 0.$$

Thus  $A^\xi(\tau, s, \eta) = \lim_{\epsilon \rightarrow 0} C_{[s,\tau]}^\epsilon(u(\cdot, X(\cdot, s, \eta)), W^\xi)$  in probability.

Now we choose  $\eta = X(s, t, x)$ . Since by (2.15)  $X(\cdot, s, \eta)$  is indistinguishable from  $X(\cdot, t, x)$  we clearly have

$$\begin{aligned} C_{[s,\tau]}^\epsilon(u(\cdot, X(\cdot, s, \eta)), W^\xi) &= C_{[s,\tau]}^\epsilon(u(\cdot, X(\cdot, t, x)), W^\xi) \\ &= C_{[t,\tau]}^\epsilon(u(\cdot, X(\cdot, t, x)), W^\xi) - C_{[t,s]}^\epsilon(u(\cdot, X(\cdot, t, x)), W^\xi) \end{aligned}$$

and this converges in probability to  $A^\xi(\tau, t, x) - A^\xi(s, t, x)$ .  $\square$

We are now ready to conclude the proof. In the time-homogeneous, finite-dimensional case some of the arguments may be shortened taking into account some general properties of additive functionals of Markov processes (see [39] Theorem 66.2). We report below a full proof for our case.

### 4.3 Conclusion of the proof of Theorem 3.1.

To shorten notation let  $v_\tau^{t,x} = U_\tau^{t,x} G(\tau, X(\tau, t, x))$ .

First we note that Proposition 4.4 and Lemma 4.5 imply that for  $\xi \in \Xi$ ,  $0 \leq t \leq \tau < T$ ,  $x \in H$ ,

$$A^\xi(\tau, t, x) = \int_t^\tau v_\sigma^{t,x} \xi \, d\sigma, \quad \mathbb{P} - a.s. \quad (4.16)$$

Next, for  $N > 0$ , we introduce the truncation operator  $T_N(r) = (r \wedge N) \vee (-N)$ ,  $r \in \mathbb{R}$ , we take an orthonormal basis  $\{\xi_i\}$  of  $\Xi$  and we define

$$\zeta^{i,N}(t, x) = \liminf_{n \rightarrow \infty} \mathbb{E} T_N \left( n A^{\xi_i} \left( t + \frac{1}{n}, t, x \right) \right), \quad t \in [0, T], x \in H,$$

and  $\zeta^{i,N}(T, x) = 0$ ,  $x \in H$ . Clearly  $\zeta^{i,N} : [0, T] \times H \rightarrow \mathbb{R}$  is a Borel function.

We fix  $x \in H$  and  $0 \leq t \leq s < T$ . We wish to compute

$$\zeta^{i,N}(s, X(s, t, x)) = \liminf_{n \rightarrow \infty} \left\{ \mathbb{E} T_N \left( n A^{\xi_i} \left( s + \frac{1}{n}, s, y \right) \right) \Big|_{y=X(s,t,x)} \right\},$$

where we denote  $\mathbb{E} T_N (n A^{\xi_i}(s + 1/n, s, y)) \Big|_{y=X(s,t,x)}$  the random variable obtained by composing  $X(s, t, x)$  with the map  $y \mapsto \mathbb{E} T_N (n A^{\xi_i}(s + 1/n, s, y))$ . Since, by Lemma 4.5,  $(y, \omega) \rightarrow T_N (n A^{\xi_i}(s + 1/n, s, y, \omega))$  is  $\mathcal{B}(H) \times \mathcal{F}_{[s,T]}$ -measurable and  $X(s, t, x)$  is  $\mathcal{F}_s$ -measurable, and the  $\sigma$ -algebras  $\mathcal{F}_{[s,T]}$  and  $\mathcal{F}_s$  are independent, we obtain

$$\mathbb{E} T_N \left( n A^{\xi_i} \left( s + \frac{1}{n}, s, y \right) \right) \Big|_{y=X(s,t,x)} = \mathbb{E}^{\mathcal{F}_s} T_N \left( n A^{\xi_i} \left( s + \frac{1}{n}, s, X(s, t, x) \right) \right), \quad \mathbb{P} - a.s.$$

and by Lemma 4.6 and (4.16) it follows that

$$\begin{aligned} \mathbb{E} T_N \left( n A^{\xi_i} \left( s + \frac{1}{n}, s, x \right) \right) \Big|_{y=X(s,t,x)} &= \mathbb{E}^{\mathcal{F}_s} T_N \left( n A^{\xi_i} \left( s + \frac{1}{n}, t, x \right) - n A^{\xi_i}(s, t, x) \right) \\ &= \mathbb{E}^{\mathcal{F}_s} T_N \left( n \int_s^{s+1/n} v_\sigma^{t,x} \xi_i \, d\sigma \right), \quad \mathbb{P} - a.s. \end{aligned}$$

We conclude that

$$\zeta^{i,N}(s, X(s, t, x)) = \liminf_{n \rightarrow \infty} \mathbb{E}^{\mathcal{F}_s} T_N \left( n \int_s^{s+1/n} v_\sigma^{t,x} \xi_i \, d\sigma \right), \quad \mathbb{P} - a.s.$$

Now we fix  $x$  and  $t$ . We note that  $\mathbb{P}$ -a.s. the equality

$$\lim_{n \rightarrow \infty} n \int_s^{s+1/n} v_\sigma^{t,x} \xi_i \, d\sigma = v_s^{t,x} \xi_i \quad (4.17)$$

holds for almost all  $s \in [t, T]$ , by the Lebesgue differentiation theorem. Thus, for almost all  $s \in [t, T]$ , (4.17) holds  $\mathbb{P}$ -a.s. and therefore

$$\lim_{n \rightarrow \infty} T_N \left( n \int_s^{s+1/n} v_\sigma^{t,x} \xi_i \, d\sigma \right) = T_N (v_s^{t,x} \xi_i)$$

in the  $L^1(\Omega)$ -norm, by the dominated convergence theorem. It follows that

$$\zeta^{i,N}(s, X(s, t, x)) = \mathbb{E}^{\mathcal{F}_s} T_N (v_s^{t,x} \xi_i) = T_N (v_s^{t,x} \xi_i), \quad \mathbb{P} - a.s. \text{ for a.a. } s \in [t, T],$$



the last equality holding because  $v^{t,x}$  is an adapted process.

So far we have proved that for  $x \in H$ ,  $0 \leq t < T$ ,

$$\zeta^{i,N}(\tau, X(\tau, t, x)) = T_N(v_\tau^{t,x} \xi_i), \quad \mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T], \quad (4.18)$$

for every  $i, N$ . Now let  $C \subset [0, T] \times H$  denote the set of pairs  $(t, x)$  such that  $\lim_{N \rightarrow \infty} \zeta^{i,N}(t, x)$  exists and the series  $\sum_{i=1}^{\infty} (\lim_{N \rightarrow \infty} \zeta^{i,N}(t, x)) \xi_i^*$  converges in  $\Xi^*$  (as usual,  $\xi_i^* \in \Xi^*$  is the element that corresponds to  $\xi_i \in \Xi$  under the Riesz isometry). Let us define

$$\zeta(t, x) = \sum_{i=1}^{\infty} \left( \lim_{N \rightarrow \infty} \zeta^{i,N}(t, x) \right) \xi_i^*, \quad (t, x) \in C, \quad \zeta(t, x) = 0, \quad (t, x) \notin C.$$

Clearly,  $\zeta : [0, T] \times H \rightarrow \Xi^*$  is Borel measurable. Since the process  $v$  takes values in  $\Xi^*$ , it satisfies

$$v_\tau^{t,x}(\omega) = \sum_{i=1}^{\infty} \left( \lim_{N \rightarrow \infty} T_N(v_\tau^{t,x}(\omega) \xi_i) \right) \xi_i^*,$$

for every  $t, x, \tau, \omega$ , so it follows from (4.18) that, for every  $t \in [0, T]$  and  $x \in H$ , we have  $(\tau, X(\tau, t, x)) \in C$ ,  $\mathbb{P}$ -a.s. for almost all  $\tau \in [t, T]$ , and

$$v_\tau^{t,x} = \zeta(\tau, X(\tau, t, x)), \quad \mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T]. \quad (4.19)$$

Assertions (i) and (ii) in Theorem 3.1 now follow from Proposition 4.4.

To prove point (iii) of Theorem 3.1 we notice that the definition of  $v^{t,x}$  and (4.19) yield that for every  $t \in [0, T]$  and  $x \in H$

$$U_\tau^{t,x} G(\tau, X(\tau, t, x)) = \zeta(\tau, X(\tau, t, x)), \quad \mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T]. \quad (4.20)$$

Let us consider, for every  $t, x$ , the equation

$$G(t, x)^* \beta = \zeta(t, x)^*, \quad (4.21)$$

in the unknown  $\beta \in H$ . Proposition B.1 in Appendix B shows that there exist a Borel set  $A_0 \subset [0, T] \times H$  and a measurable function  $\beta : [0, T] \times H \rightarrow H$  such that for  $(t, x) \notin A_0$  there is no solution of equation (4.21) and for  $(t, x) \in A_0$  a solution of (4.21) is given by  $\beta(t, x)$ . Now we define  $\rho(t, x) = \beta(t, x)^*$  and we claim that for every  $t \in [0, T]$  and  $x \in H$  we have

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

Indeed, (4.20) shows that a solution  $\beta$  to the equation  $G(\tau, X(\tau, t, x))^* \beta = \zeta(\tau, X(\tau, t, x))^*$  is given by  $(U_\tau^{t,x})^*$  and consequently  $(\tau, X(\tau, t, x)) \in A_0$ ,  $\mathbb{P}$ -a.s. for a.e.  $\tau \in [t, T]$ . The claim now follows from the definition of  $\rho$ .  $\square$

**Remark 4.7** The proof of Theorem 3.1 we have just completed shows that  $\zeta$  depends on the law of  $X$  rather than on  $X$  itself. Thus  $\zeta$  is a functional of  $u$  and the coefficients  $A, F, G$  of the forward equation.

**Remark 4.8** As a byproduct of the above proof we get the following characterization of the elements of  $\tilde{\nabla}^G u$  along the trajectories of  $X$ . For any sequence  $\{u_n : n = 1, 2, \dots\}$  verifying (i), (ii) and (iii) in Lemma 4.3, for any  $t \in [0, T]$  and  $x \in H$ , if  $\nabla u_n(\cdot, X(\cdot, t, x)) \rightharpoonup U^{t,x}$  weakly in  $L^2(\Omega \times [t, T]; H^*)$  then:

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

where  $\zeta$  is any element in  $\tilde{\nabla}^G u$ .

## 5 Mild solutions of semilinear PDEs in the sense of the generalized directional gradient

In this section we study the equation

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

We assume that Hypotheses 2.2 and 2.4 hold.

**Definition 5.1** *We say that a Borel measurable function  $u : [0, T] \times H \rightarrow \mathbb{R}$  is a mild solution of equation (5.1) in the sense of the generalized directional gradient if the following holds:*

1. *for some  $C > 0, r \geq 0$  and for every  $t \in [0, T], x, y \in H$*

$$|u(t, x) - u(t, y)| \leq C|x - y|(1 + |x| + |y|)^r, \quad |u(t, 0)| \leq C;$$

2. *for all  $0 \leq t \leq T$  and  $x \in H$*

$$u(t, x) = P_{t,T}[\phi](x) - \int_t^T P_{t,\tau} \left[ \psi(\tau, \cdot, u(\tau, \cdot), \zeta(\tau, \cdot)) \right](x) d\tau, \quad (5.2)$$

where  $\zeta$  is an arbitrary element of the generalized gradient  $\tilde{\nabla}^G u$ .

The generalized gradient has been introduced in Definition 3.1. We note that equality (5.2) is a weak form of the variation of constants formula for the equation (5.1). We also note that (5.2) can be written

$$u(t, x) = \mathbb{E} \phi(X(T, t, x)) - \mathbb{E} \int_t^T \psi \left( \tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x)) \right) d\tau. \quad (5.3)$$

Since  $u, \psi, \phi$  have polynomial growth, (2.6) and (3.2) yield

$$\phi(X(T, t, x)) - \int_t^T \psi \left( \tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x)) \right) d\tau \in L^p(\Omega), \quad p \in [2, \infty).$$

Moreover, although  $\zeta$  is not uniquely determined, the process  $\{\zeta(\tau, X(\tau, t, x)), \tau \in [t, T]\}$  is unique up to modification: compare Remark 3.2. Thus equality (5.3) and consequently Definition 5.1 is meaningful.

**Theorem 5.1** *Under Hypotheses 2.2 and 2.4 there exists a unique mild solution of equation (5.1) in the sense of the generalized directional gradient. Moreover  $u$  is connected to the forward-backward system (2.9) by the following equalities: for every  $t \in [0, T]$  and  $x \in H$*

$$(i) \quad u(t, x) = Y(t, t, x);$$

$$(ii) \quad \mathbb{P}\text{-a.s.}, \quad u(\tau, X(\tau, t, x)) = Y(\tau, t, x) \text{ for all } \tau \in [t, T];$$

$$(iii) \quad \text{if } \zeta \in \tilde{\nabla}^G u \text{ then, } \mathbb{P}\text{-a.s.}, \quad \zeta(\tau, X(\tau, t, x)) = Z(\tau, t, x) \text{ for almost every } \tau \in [t, T].$$

Finally,  $u : [0, T] \times H \rightarrow \mathbb{R}$  is continuous.

**Proof.** *Uniqueness.* Let  $u$  be a solution and  $\zeta \in \tilde{\nabla}^G u$ . In particular,

$$u(s, y) = P_{s,T}[\phi](y) - \int_s^T P_{s,\tau} \left[ \psi(\tau, \cdot, u(\tau, \cdot), \zeta(\tau, \cdot)) \right](y) d\tau, \quad s \in [0, T], y \in H. \quad (5.4)$$

By the Markov property, for  $0 \leq t \leq s \leq \tau \leq T$  and  $x \in H$ , we have,  $\mathbb{P}$ -a.s.,

$$\mathbb{E}^{\mathcal{F}_s} \psi \left( \tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x)) \right) = P_{s,\tau} \left[ \psi(\tau, \cdot, u(\tau, \cdot), \zeta(\tau, \cdot)) \right](X(s, t, x)),$$

and similarly  $\mathbb{E}^{\mathcal{F}_s} \phi(X(\tau, t, x)) = P_{s,T}[\phi](X(s, t, x))$ . Substituting  $y$  with  $X(s, t, x)$  in (5.4) we obtain

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

where we have defined

$$\eta = \phi(X(T, t, x)) - \int_t^T \psi \left( \tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x)) \right) d\tau.$$

We note that  $\mathbb{E}^{\mathcal{F}_t}[\eta] = u(t, X(t, t, x)) = u(t, x)$ . As noticed earlier,  $\eta$  belongs to  $L^2(\Omega; \mathbb{R})$  and since it is  $\mathcal{F}_T$ -measurable, by the martingale representation theorem, there exists  $\tilde{Z} \in L^2_{\mathcal{P}}(\Omega \times [t, T]; L_2(\Xi, \mathbb{R}))$  such that  $\mathbb{E}^{\mathcal{F}_s}[\eta] = \int_t^s \tilde{Z}_\tau dW_\tau + u(t, x)$ . We conclude that the process  $\{u(s, X(s, t, x)), s \in [t, T]\}$  is a (real) continuous semimartingale with canonical decomposition

$$\begin{aligned} u(s, X(s, t, x)) &= \int_t^s \tilde{Z}_\tau dW_\tau \\ &\quad + u(t, x) + \int_t^s \psi \left( \tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x)) \right) d\tau. \end{aligned} \quad (5.5)$$

Fixing  $\xi \in \Xi$  and computing the joint quadratic variation of both sides of (5.5) with  $W^\xi$  we get by Proposition 2.1

$$\langle u(\cdot, X(\cdot, t, x)), W^\xi \rangle_{[t,s]} = \int_t^s \tilde{Z}_\tau \xi d\tau.$$

Thus by Definition 3.1 for a.a.  $\tau \in [t, T]$ , we have  $\mathbb{P}$ -a.s.  $\zeta(\tau, X(\tau, t, x)) = \tilde{Z}_\tau$ . Substituting into (5.5) we obtain

$$\begin{aligned} u(s, X(s, t, x)) &= \int_t^s \zeta(\tau, X(\tau, t, x)) dW_\tau \\ &\quad + u(t, x) + \int_t^s \psi \left( \tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x)) \right) d\tau, \end{aligned}$$

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

$$\begin{aligned} u(s, X(s, t, x)) &+ \int_s^T \zeta(\tau, X(\tau, t, x)) dW_\tau = \phi(X(T, t, x)) \\ &- \int_s^T \psi \left( \tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x)) \right) d\tau, \end{aligned}$$

for  $s \in [t, T]$ . Comparing with the backward equation in (2.9) we note that the pairs

$$\left( Y(s, t, x), Z(s, t, x) \right) \quad \text{and} \quad \left( u(s, X(s, t, x)), \zeta(s, X(s, t, x)) \right), \quad s \in [t, T],$$

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  $Y(t, t, x) = u(t, x)$ .

*Existence.* Let us define  $u(t, x) = Y(t, t, x)$ ,  $t \in [0, T]$ ,  $x \in H$ . Then point 1. in Definition 5.1 is satisfied by (2.12); let  $\zeta$  be an element of the generalized gradient of  $u$ . Let us fix  $t \in [0, T]$  and  $x \in H$ . From (2.16) it follows that,  $\mathbb{P}$ -a.s.,  $u(\tau, X(\tau, t, x)) = Y(\tau, t, x)$  for all  $\tau \in [t, T]$ . The backward equation can be written,  $\mathbb{P}$ -a.s.,

$$u(\tau, X(\tau, t, x)) = u(t, x) + \int_t^\tau Z(\sigma, t, x) dW_\sigma + \int_t^\tau \psi(\sigma, X(\sigma, t, x), Y(\sigma, t, x), Z(\sigma, t, x)) d\sigma,$$

for every  $\tau \in [t, T]$ . For  $\xi \in \Xi$ , let us consider the joint quadratic variation of both sides with  $W^\xi$ . Proposition 2.1 and Theorem 3.1 yield, for  $t \leq \tau < T$  and  $\zeta \in \widetilde{\nabla}^G u$ ,

$$\int_t^\tau \zeta(\sigma, X(\sigma, t, x)) \xi d\sigma = \int_t^\tau Z(\sigma, t, x) \xi d\sigma, \quad (5.6)$$

$\mathbb{P}$ -a.s. Since both sides of (5.6) are continuous with respect to  $\tau$ , it follows that,  $\mathbb{P}$ -a.s., they coincide for all  $\tau \in [t, T]$ . This implies that  $\zeta(\tau, s, X(s, t, x)) = Z(\tau, t, x)$ ,  $\mathbb{P}$ -a.s. for a.a.  $\tau \in [s, T]$ . Now we take expectation in the backward equation and we obtain

$$\begin{aligned} u(t, x) &= -\mathbb{E} \int_t^T \psi\left(\tau, X(\tau, t, x), Y(\tau, t, x), Z(\tau, t, x)\right) d\tau + \mathbb{E} \phi(X(T, t, x)) \\ &= -\mathbb{E} \int_t^T \psi\left(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), \zeta(\tau, X(\tau, t, x))\right) d\tau + \mathbb{E} \phi(X(T, t, x)), \end{aligned}$$

which is the required equality.

Finally, the continuity of  $u$  follows from the definition  $u(t, x) = Y(t, t, x)$  and the continuity statement in Theorem 2.5.  $\square$

**Remark 5.2** By Remark 4.8 we get a different characterization of mild solutions in the sense of the generalized directional gradient.

Namely if  $u$  is a measurable function  $[0, T] \times H \rightarrow \mathbb{R}$  verifying point 1 in the Definition 5.1, then  $u$  is a mild solutions in the sense of the generalized directional gradient of equation (5.1) if and only if the following is verified:

for every  $t \in [0, T]$  and  $x \in H$ , every sequence  $\{u_n : n = 1, 2, \dots\}$  verifying (i), (ii) and (iii) in Lemma 4.3 and every  $U^{t,x} \in L^2(\Omega \times [t, T]; H^*)$  with  $\nabla u_n(\cdot, X(\cdot, t, x)) \rightharpoonup U^{t,x}$  weakly in  $L^2(\Omega \times [t, T]; H^*)$  it holds

$$u(t, x) = \mathbb{E} \phi(X(T, t, x)) - \mathbb{E} \int_t^T \psi\left(\tau, X(\tau, t, x), u(\tau, X(\tau, t, x)), U_\tau G(\tau, X(\tau, t, x))\right) d\tau.$$

## 6 Application: Black and Scholes PDE for path dependent options

In this section we assume that the reader has some acquaintance of some basic concepts in mathematical finance, in particular concerning option pricing.

We consider a generalized Black and Scholes market with  $d$  risky assets, whose prices at time  $t$  are denoted by  $S_t^i$  ( $i = 1, \dots, d$ ), and one non risky asset, whose price is denoted by  $B_t$ . We assume the following prices evolution:

$$\begin{cases} dS_t = \text{diag}[S_t] \sigma(t, S_t, S_t(\cdot)) dW_t, & t \in [0, T], \\ S_0 = s_0 \in \mathbb{R}_+^d, \\ S_\theta = \nu_0(\theta), & \theta \in (-T, 0), \\ dB_t = r(t, S_t, S_t(\cdot)) B_t dt, & t \in [0, T], \\ B_0 = 1. \end{cases} \quad (6.1)$$

where  $S_t = \begin{pmatrix} S_t^1 \\ \vdots \\ S_t^d \end{pmatrix}$  and, for all  $s \in \mathbb{R}^d$ ,  $\text{diag}[s] = \begin{pmatrix} s^1 & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & s^d \end{pmatrix}$ .

In the above equation the bond rate  $r$  is real-valued and the volatility matrix  $\sigma$  takes values in  $\mathbb{R}^{d \times d}$ . We will assume that  $W$  is a Wiener process in  $\mathbb{R}^d$ , with respect to a suitable probability.

$S_t(\cdot)$  denotes the whole trajectory  $S_t(\theta) = S_{t+\theta}$ ,  $\theta \in [-T, 0]$ .  $\nu_0$  is a given function defined in  $(-T, 0)$ . Thus the model described in (6.1) takes into account *memory effects* in the prices evolution.

Moreover we consider a contingent claim of the form

$$\varphi(S_T, S_T(\cdot))$$

that also depends on the whole evolution in time of the prices of the shares: see [8], [24] or [40] and references within for a general discussion on such kind of options, usually referred to as *path-dependent*.

Finally, following [15], we take into account the fact that an investor is allowed to borrow money at interest rate  $R_t > r_t$  different from the bond rate (see also [13]). Again we allow  $R_t = R(t, S_t, S_t(\cdot))$  to depend on the whole path of  $S$ .

We denote by  $\pi_t^i$  the value of the investor's portfolio invested in the  $i$ -th asset ( $i = 1, \dots, d$ ) at time  $t$ , and will only consider square-integrable, predictable investment strategies  $\pi \in L_{\mathcal{P}}^2([0, T] \times \Omega; \mathbb{R}^d)$ . As in [15] we notice that the value  $V_t$  of the corresponding self financing portfolio satisfies the equation:

$$\begin{cases} dV_t = r(t, S_t, S_t(\cdot)) (V_t - \pi_t^* \mathbf{1}) dt + (r(t, S_t, S_t(\cdot)) - R(t, S_t, S_t(\cdot))) (V_t - \pi_t^* \mathbf{1})^- dt \\ \quad + \pi_t^* \sigma(t, S_t, S_t(\cdot)) dW_t, \\ V_0 = v \in \mathbb{R}_+, \end{cases} \quad (6.2)$$

where  $\mathbf{1}$  denotes the vector in  $\mathbb{R}^d$  having all coordinates equal to 1.

We assume the following

1.  $\{W_t : t \geq 0\}$  is a  $d$ -dimensional Wiener process defined in a complete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and  $\{\mathcal{F}_t : t \geq 0\}$  is the filtration generated by  $W$  augmented with null probability sets.
2.  $\sigma : [0, T] \times \mathbb{R}^d \times L^2((-T, 0); \mathbb{R}^d) \rightarrow \mathbb{R}^{d \times d}$  is measurable. Moreover

$$|\sigma^{-1}(t, s, \nu)| \leq c$$

$$|\sigma(t, s_1, \nu_1) - \sigma(t, s_2, \nu_2)|_{\mathbb{R}^d} \leq c \left( 1 + |s_1|_{\mathbb{R}^d} + |\nu_1|_{L^2((-T, 0); \mathbb{R}^d)} + |s_2|_{\mathbb{R}^d} + |\nu_2|_{L^2((-T, 0); \mathbb{R}^d)} \right)^m \times \left( |s_2 - s_1|_{\mathbb{R}^d} + |\nu_2 - \nu_1|_{L^2((-T, 0); \mathbb{R}^d)} \right) \quad (6.3)$$

$$|\text{diag}[s_1]\sigma(t, s_1, \nu_1) - \text{diag}[s_2]\sigma(t, s_2, \nu_2)|_{\mathbb{R}^d} \leq c \left( |s_2 - s_1|_{\mathbb{R}^d} + |\nu_2 - \nu_1|_{L^2((-T, 0); \mathbb{R}^d)} \right) \quad (6.4)$$

for a suitable  $c \in \mathbb{R}_+$  and for all  $t \in [0, T]$ ,  $s, s_1, s_2 \in \mathbb{R}^d$ ,  $\nu, \nu_1, \nu_2 \in L^2((-T, 0); \mathbb{R}^d)$ .

3.  $r$  and  $R$  are real measurable functions defined on  $[0, T] \times \mathbb{R}^d \times L^2((-T, 0); \mathbb{R}^d)$  and  $\varphi$  is a real function defined on  $\mathbb{R}^d \times L^2((-T, 0); \mathbb{R}^d)$ . Moreover, for suitable  $c, m$  in  $\mathbb{R}_+$ , they satisfy

$$\begin{aligned} |r(t, s, \nu)| + |R(t, s, \nu)| &\leq c, \\ |r(t, s_1, \nu_1) - r(t, s_2, \nu_2)| + |R(t, s_1, \nu_1) - R(t, s_2, \nu_2)| + |\varphi(s_1, \nu_1) - \varphi(s_2, \nu_2)| \\ &\leq c \left( 1 + |s_1|_{\mathbb{R}^d} + |\nu_1|_{L^2((-T, 0); \mathbb{R}^d)} + |s_2|_{\mathbb{R}^d} + |\nu_2|_{L^2((-T, 0); \mathbb{R}^d)} \right)^m \cdot \left( |s_2 - s_1|_{\mathbb{R}^d} + |\nu_2 - \nu_1|_{L^2((-T, 0); \mathbb{R}^d)} \right), \end{aligned} \quad (6.5)$$

for all  $t \in [0, T]$ ,  $s, s_1, s_2 \in \mathbb{R}^d$ ,  $\nu, \nu_1, \nu_2 \in L^2((-T, 0); \mathbb{R}^d)$ .

The idea is now to set the problem in an infinite dimensional framework by considering equation (6.1) as a delay equation. Following [9] we set  $H = \mathbb{R}^d \times L^2((-T, 0); \mathbb{R}^d)$

$$\mathcal{D}(A) = \left\{ \begin{pmatrix} s \\ \nu \end{pmatrix} : \nu \in W^{1,2}((-T, 0), \mathbb{R}^d) \text{ and } \nu(0) = s \right\}, \quad A \begin{pmatrix} s \\ \nu \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{d\nu}{d\theta} \end{pmatrix},$$

$$G \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) = \begin{pmatrix} \text{diag}[s]\sigma(t, s, \nu) \\ 0 \end{pmatrix}, \quad x_0 = \begin{pmatrix} s_0 \\ \nu_0 \end{pmatrix}.$$

We consider the forward SDE in the Hilbert space  $H$

$$\begin{cases} dX_t = AX_t dt + G(t, X_t) dW_t, & t \in [0, T], \\ X_0 = x_0. \end{cases} \quad (6.6)$$

In [9] it is proved that equation (6.1) is equivalent to equation (6.6). In particular if  $S$  is the unique solution to (6.1) then  $X_t = \begin{pmatrix} S_t \\ S_t(\cdot) \end{pmatrix}$  is the solution to (6.6).

Then we let  $\phi \begin{pmatrix} s \\ \nu \end{pmatrix} = \varphi(s, \nu)$  and

$$\psi \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix}, y, z \right) = r(t, s, \nu) (y - z\sigma^{-1}(t, s, \nu)\mathbf{1}) + (r(t, s, \nu) - R(t, s, \nu)) (y - z\sigma^{-1}(t, s, \nu)\mathbf{1})^-$$

for  $t \in [0, T]$ ,  $\begin{pmatrix} s \\ \nu \end{pmatrix} \in H$ ,  $y \in \mathbb{R}$ ,  $z \in (\mathbb{R}^d)^*$  and consider the backward stochastic differential equation

$$\begin{cases} dY_t = \psi(t, X_t, Y_t, Z_t) dt + Z_t dW_t, & t \in [0, T], \\ Y_T = \phi(X_T), \end{cases} \quad (6.7)$$

where  $Y$  takes values in  $\mathbb{R}$  and  $Z$  in  $(\mathbb{R}^d)^*$ .

It is easy to verify that under the present assumptions  $G, \phi$  and  $\psi$  verify Hypotheses 2.2 and 2.4 (notice that since  $\Xi$  is finite dimensional  $L_2(\Xi, H) = L(\Xi, H)$ ). Moreover comparing (6.7) and (6.2) we deduce that an admissible strategy  $\pi$  replicates the claim if and only if  $\pi_t^* = Z_t \sigma_t^{-1}$ .

Thus the results of Theorem 5.1 can be applied to the forward-backward system given by (6.6) and (6.7) and to the corresponding parabolic PDE. In this specific case we obtain the following infinite-dimensional extension of the celebrated Black and Scholes result:

**Theorem 6.1** *Under the previous assumptions there exists a unique mild solution  $u : [0, T] \times H \rightarrow \mathbb{R}$ , in the sense of the generalized directional gradient, of the degenerate parabolic equation:*

$$\begin{cases} \frac{\partial}{\partial t} u \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) + \mathcal{L}_t[u(t, \cdot)] \left( \begin{pmatrix} s \\ \nu \end{pmatrix} \right) \\ = \psi \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix}, u \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right), \nabla u \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) G \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) \right), \\ u \left( T, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) = \phi \left( \begin{pmatrix} s \\ \nu \end{pmatrix} \right), \end{cases} \quad (6.8)$$

where  $t \in [0, T]$ ,  $s \in \mathbb{R}^d$  and  $\nu \in L^2([-T, 0], \mathbb{R}^d)$ . Moreover for  $v : H \rightarrow \mathbb{R}$  regular enough:

$$\mathcal{L}_t[v] \left( \begin{pmatrix} s \\ \nu \end{pmatrix} \right) = \frac{1}{2} \text{Tr}_H \left[ G \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) G^* \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) \nabla^2 v \left( \begin{pmatrix} s \\ \nu \end{pmatrix} \right) \right] + \nabla v \left( \begin{pmatrix} s \\ \nu \end{pmatrix} \right) A \left( \begin{pmatrix} s \\ \nu \end{pmatrix} \right)$$

(in the above formula, as usual, we do not identify  $H$  and  $H^*$  and we consider  $\nabla v$  as an element of  $H^*$  and  $\nabla^2 v$  as an element of  $L(H)$ ).

Moreover the hedging strategy is given by  $\pi_t^* = \zeta \left( t, \begin{pmatrix} S_t \\ S_t(\cdot) \end{pmatrix} \right) \sigma^{-1}(t, S_t, S_t(\cdot))$  where  $\zeta$  is any element in  $\widetilde{\nabla}^G u$ .

Finally the value of the hedging portfolio is given by  $V_t = u \left( t, \begin{pmatrix} S_t \\ S_t(\cdot) \end{pmatrix} \right)$ , in particular the non-arbitrage price of the option is  $V_0 = u \left( 0, \begin{pmatrix} s_0 \\ \nu_0 \end{pmatrix} \right)$ .

Equation (6.8) has a precise meaning given by Definition 5.2. To give (6.8) a less abstract appearance that can be more easily interpreted we notice that if  $u : [0, T] \times H = [0, T] \times \mathbb{R}^d \times L^2((-T, 0); \mathbb{R}^d) \rightarrow \mathbb{R}$  is regular enough then

$$\begin{aligned} \text{Tr}_H \left[ G \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) G^* \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) \nabla^2 u \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) \right] \\ = \text{Tr}_{\mathbb{R}^d} \left[ \text{diag}[s] \sigma(t, s, \nu) \sigma^*(t, s, \nu) \text{diag}[s] \frac{\partial^2 u}{\partial s^2}(t, s, \nu) \right] \\ = \sum_{i,j=1}^d s^i s^j a_{ij}(t, s, \nu) \frac{\partial^2 u}{\partial s^i \partial s^j}(t, s, \nu), \end{aligned}$$

where  $a_{ij}(t, s, \nu)$  are the elements of the matrix  $a(t, s, \nu) = \sigma(t, s, \nu) \sigma(t, s, \nu)^*$ ,

$$\begin{aligned} \nabla u \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) A \left( \begin{pmatrix} s \\ \nu \end{pmatrix} \right) &= \int_{-T}^0 \left\langle \frac{d\nu}{d\theta}(\theta), \frac{\partial u}{\partial \nu}(t, s, \nu)(\theta) \right\rangle_{\mathbb{R}^d} d\theta, \\ \nabla u \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) G \left( t, \begin{pmatrix} s \\ \nu \end{pmatrix} \right) &= \frac{\partial u}{\partial s}(t, s, \nu)^* \text{diag}[s] \sigma(t, (s, \nu)), \end{aligned}$$

where for notational simplicity we have identified  $L^2((-T, 0); \mathbb{R}^d)$  with its dual, and consequently considered  $\frac{\partial}{\partial \nu} u$  as an element of  $L^2((-T, 0); \mathbb{R}^d)$ , while  $\frac{\partial}{\partial s} u$  is a (column) vector in  $\mathbb{R}^d$  and  $\frac{\partial^2}{\partial s^2} u$  is a matrix in  $\mathbb{R}^{d \times d}$ . Moreover we have dropped vectorial notation in the right hand of equalities. We conclude that equation (6.8) is the abstract version of the following concrete degenerate parabolic PDE:

$$\left\{ \begin{array}{l} \frac{\partial}{\partial t} u(t, s, \nu) + \frac{1}{2} \sum_{i,j=1}^d s^i s^j a_{ij}(t, s, \nu) \frac{\partial^2 u}{\partial s^i \partial s^j}(t, s, \nu) + \int_{-T}^0 \left\langle \frac{d\nu}{d\theta}(\theta), \frac{\partial u}{\partial \nu}(t, s, \nu)(\theta) \right\rangle_{\mathbb{R}^d} d\theta \\ \\ = (r(t, s, \nu) - R(t, s, \nu)) \left( u(t, s, \nu) - \left\langle \frac{\partial}{\partial s} u(t, s, \nu), s \right\rangle_{\mathbb{R}^d} \right)^- \\ \\ + r(t, s, \nu) \left( u(t, s, \nu) - \left\langle \frac{\partial}{\partial s} u(t, s, \nu), s \right\rangle_{\mathbb{R}^d} \right), \\ \\ u(T, s, \nu) = \varphi(s, \nu). \end{array} \right. \quad (6.9)$$

Moreover the hedging strategy is given by

$$\pi_t^* = \frac{\partial u}{\partial s}(t, S_t, S_t(\cdot))^* \text{diag}[S_t], \quad \text{i.e.} \quad \pi_t^i = S_t^i \frac{\partial u}{\partial s^i}(t, S_t, S_t(\cdot)), \quad i = 1, \dots, d.$$

**Remark 6.2** We notice that in equation (6.8) (see also equation (6.9)) the nonlinearity  $\psi$  is Lipschitz but not differentiable. Moreover as it is well known in the most common examples of claims (even in classical non path-dependent cases, as standard european call options)  $\phi$  is lipschitz but not differentiable. So differentiability assumptions on the coefficients would not be natural for the model treated here.

**Remark 6.3** The requirement on  $\varphi$  imposed by (6.5) is restrictive, mainly because it requires  $\varphi$  to be locally lipschitz in the variable  $\nu$  with respect to the norm of  $L^2((-T, 0); \mathbb{R}^d)$ . Nevertheless it is satisfied, for instance, by functionals of the form:

$$\varphi(s, \nu) = \varphi_0 \left( s, \sup_{\theta \in [0, T]} \nu^\epsilon(\theta) \right) \quad \text{where} \quad \nu^\epsilon(\theta) = \frac{1}{\epsilon} \int_{(\theta-\epsilon) \vee 0}^{\theta} a(\zeta - T) \nu(\zeta) d\zeta,$$

where  $a \in C([0, T]; \mathbb{R}^{d \times d})$ , the ‘sup’ is intended coordinatewise, and  $\varphi_0 : \mathbb{R}^{2d} \rightarrow \mathbb{R}$  verifies

$$|\varphi_0(s_1, \mu_1) - \varphi_0(s_2, \mu_2)| \leq c(1 + |s_1|_{\mathbb{R}^d} + |\mu_1|_{\mathbb{R}^d} + |s_2|_{\mathbb{R}^d} + |\mu_2|_{\mathbb{R}^d})^m (|s_2 - s_1|_{\mathbb{R}^d} + |\mu_2 - \mu_1|_{\mathbb{R}^d}),$$

for suitable  $c, m$  in  $\mathbb{R}_+$  and for all  $s_1, s_2, \mu_1, \mu_2 \in \mathbb{R}^d$ .

Thus our abstract results allow us to treat, for instance, all contingent claims of the type

$$\varphi(S_T, S_T(\cdot)) = \varphi_0 \left( S_T, \sup_{\theta \in [0, T]} S^\epsilon(\theta) \right) \quad \text{where} \quad S^\epsilon(\theta) = \frac{1}{\epsilon} \int_{(\theta-\epsilon) \vee 0}^{\theta} a(\zeta) S(\zeta) d\zeta. \quad (6.10)$$

Since the solutions to equation (6.1) (or equivalently (6.6)) have continuous paths, see Proposition 6.6, it is clear that for  $\epsilon$  small the above functionals are good approximations of the following general type of *look-back* claim (see [8], [24] or [40]):

$$\varphi(S_T, S_T(\cdot)) = \varphi_0 \left( S_T, \sup_{\theta \in [0, T]} a(\theta) S_\theta \right).$$



**Remark 6.4** In the special case in which  $\sigma$ ,  $r$  and  $R$  do not depend on the past values  $S(\cdot)$  of the prices (thus when only the contingent claim  $\varphi(S_T, S_T(\cdot))$  is ‘path dependent’) we expect the price of the option to be independent on  $\nu_0$ . This can be easily verified noticing that if  $X$  is the solution to (6.6) then  $X_t$  only depends on  $\nu_0(\theta)$ ,  $\theta \in (-T + t, 0)$ .

In this particular case, and for specific classes of path dependent options, in [5] the authors proposed to relate the price of the option and the hedging strategy to the solution of a finite dimensional degenerate parabolic PDE. Such an idea was exploited in [6] to compute the price of asian options. Nevertheless this approach seems to be inapplicable to lookback options.

**Remark 6.5** Assumptions (6.3) and (6.4) are satisfied if, for instance,

$$\sigma(t, s, \nu) = \sigma_1(t) + \sigma_2(t, s, \bar{\nu}) \quad \text{with} \quad \bar{\nu} = \int_{-T}^0 \bar{a}(\theta) \nu(\theta) d\theta,$$

where  $\sigma_1 : [0, T] \rightarrow \mathbb{R}^{d \times d}$  is bounded measurable,  $\bar{a} \in L^2((-T, 0); \mathbb{R}^{d \times d})$  and  $\sigma_2 : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  is bounded measurable, Lipschitz in  $(s, \nu)$  uniformly with respect to  $t$ , and verifies  $\sigma_2(\cdot, s, \cdot) = 0$  if  $|s| \geq \rho$  for a suitable  $\rho$ .

In a similar way the requirements on  $r$  imposed by (6.5) are satisfied if, for instance,

$$r(t, s, \nu) = r_0(t, s, \hat{\nu}) \quad \text{with} \quad \hat{\nu} = \int_{-T}^0 \hat{a}(\theta) \nu(\theta) d\theta,$$

where  $\hat{a} \in L^2((-T, 0); \mathbb{R}^{d \times d})$  and  $r_0 : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  is bounded measurable and verifies

$$|r_0(t, s_1, \mu_1) - r_0(t, s_2, \mu_2)| \leq c (1 + |s_1|_{\mathbb{R}^d} + |\mu_1|_{\mathbb{R}^d} + |s_2|_{\mathbb{R}^d} + |\mu_2|_{\mathbb{R}^d})^m \cdot (|s_2 - s_1|_{\mathbb{R}^d} + |\mu_2 - \mu_1|_{\mathbb{R}^d}),$$

for suitable  $c, m$  in  $\mathbb{R}_+$ , and for all  $t \in [0, T]$ ,  $s_1, s_2, \mu_1, \mu_2 \in \mathbb{R}^d$ .

Similar considerations apply to  $R$ .

## 7 Applications to optimal control

We wish to apply the above results to perform the synthesis of the optimal control for a general nonlinear control system. To be able to use non-smooth feedbacks we settle the problem in the framework of weak control problems (see e.g. [17]).

Again  $H, \Xi$  denote Hilbert spaces,  $U$  is a general set endowed with a  $\sigma$ -field  $\mathcal{E}_U$ . For fixed  $T > 0$ ,  $t_0 \in [0, T]$  and  $x_0 \in H$ , an *admissible control system* (a.c.s)  $\mathbb{U}$  is given by  $\mathbb{U} = (\Omega, \mathcal{E}, (\mathcal{F}_t), \mathbb{P}, W, u, X^{\mathbb{U}})$  where

- $(\Omega, \mathcal{E}, \mathbb{P})$  is a complete probability space, with a filtration  $(\mathcal{F}_t)_{t \geq 0}$  satisfying the usual conditions;
- $\{W_t : t \geq 0\}$  is a cylindrical Wiener process in  $\Xi$ , with respect to  $\mathbb{P}$  and  $(\mathcal{F}_t)$ ;
- $u : \Omega \times [t_0, T] \rightarrow U$  is an  $(\mathcal{F}_t)$ -predictable process;
- $X^{\mathbb{U}} \in L^p_{\mathcal{P}}(\Omega; C([t_0, T]; H))$ , for all  $p \in [1, +\infty)$ , is a mild solution of the state equation:

$$\begin{cases} dX^{\mathbb{U}}_{\tau} = \left( AX^{\mathbb{U}}_{\tau} + F(\tau, X^{\mathbb{U}}_{\tau}) + G(\tau, X^{\mathbb{U}}_{\tau})R(\tau, X^{\mathbb{U}}_{\tau}, u_{\tau}) \right) d\tau + G(\tau, X^{\mathbb{U}}_{\tau}) dW_{\tau}, \tau \in [t_0, T], \\ X^{\mathbb{U}}_{t_0} = x_0 \in H. \end{cases} \quad (7.1)$$

To each a.c.s. we associate the cost:

$$J(t_0, x_0, \mathbb{U}) = \mathbb{E} \int_{t_0}^T \ell(\tau, X_\tau^u, u_\tau) d\tau + \mathbb{E} \phi(X_T^u). \quad (7.2)$$

We make the following assumption.

**Hypothesis 7.1** *The following holds:*

1.  $A, F, G$  verify the assumptions in Hypothesis 2.2
2.  $\phi$  satisfies (2.10).
3.  $R : [0, T] \times H \times U \rightarrow \Xi$  is measurable and moreover

$$|R(t, x_1, u) - R(t, x_2, u)| \leq C (1 + |x_1| + |x_2|)^m |x_2 - x_1|, \quad |R(t, x, u)| \leq C,$$

for suitable constants  $C > 0$ ,  $m \geq 0$  and all  $t \in [0, T]$ ,  $u \in U$ ,  $x, x_1, x_2 \in H$ .

4.  $\ell : [0, T] \times H \times U \rightarrow \mathbb{R} \cup \{+\infty\}$  is measurable and

$$\begin{aligned} \ell(t, x_1, u) &\leq C (1 + |x_1| + |x_2|)^m |x_2 - x_1| + \ell(t, x_2, u), \\ \ell(t, 0, u) &\geq -C, \quad \inf_{v \in U} \ell(t, 0, v) \leq C, \end{aligned}$$

for suitable constants  $C > 0$ ,  $m \geq 0$  and all  $t \in [0, T]$ ,  $u \in U$ ,  $x_1, x_2 \in H$ .

Our purpose is to minimize the functional  $J$  over all a.c.s  $\mathbb{U}$ .

Notice the occurrences of the operator  $G$  in the state equation (7.1): this special structure of the state equation is imposed by our techniques. Nevertheless in many situations it appears to be natural, see Examples 7.1 and 7.2 here and [23].

We note that the assumptions on  $\ell$  imply that, for given  $t$  and  $u$ , the function  $\ell(t, \cdot, u)$  is either finite or identically equal to  $+\infty$ .

Under the above assumptions an a.c.s. can easily be constructed as follows. We arbitrarily choose  $(\Omega, \mathcal{E}, \mathbb{P})$ ,  $(\mathcal{F}_t)$  and  $W$  as above and we consider the (uncontrolled) equation

$$\begin{cases} dX_\tau = AX_\tau + F(\tau, X_\tau)d\tau + G(\tau, X_\tau) dW_\tau, & \tau \in [t_0, T], \\ X_{t_0} = x_0 \in H. \end{cases}$$

By Proposition 2.3 this equation has a unique mild solution  $X \in L_{\mathcal{P}}^p(\Omega; C([t_0, T]; H))$  for all  $p \in [1, +\infty)$ . Moreover since  $R$  is bounded, by the Girsanov Theorem, for any fixed  $\eta \in U$  there exists a probability  $\mathbb{P}^{(\eta)}$  such that

$$W_\tau^{(\eta)} := W_\tau - \int_{t_0 \wedge \tau}^{\tau \wedge T} R(s, X_s, \eta) ds, \quad \tau \geq 0,$$

is a Wiener process. It is clear that setting  $u_\tau = \eta$ ,  $\tau \in [t_0, T]$  then  $(\Omega, \mathcal{E}, (\mathcal{F}_t), \mathbb{P}^{(\eta)}, W^{(\eta)}, u, X)$  is an a.c.s.

Moreover we note that for any a.c.s. the cost  $J(t_0, x_0, \mathbb{U})$  is well defined (possibly equal to  $+\infty$ ).

We define in a classical way the Hamiltonian function relative to the above problem: for all  $t \in [0, T]$ ,  $x \in H$ ,  $z \in \Xi^*$ ,

$$\begin{aligned}\psi(t, x, z) &= - \inf_{u \in U} \{ \ell(t, x, u) + zR(t, x, u) \}, \\ \Gamma(t, x, z) &= \{ u \in U : \psi(t, x, z) + \ell(t, x, u) + zR(t, x, u) = 0 \}.\end{aligned}\tag{7.3}$$

Thus  $\Gamma(t, x, z)$  is a (possibly empty) subset of  $U$ . Notice that under the above assumptions it holds:

$$\begin{aligned}|\psi(t, 0, 0)| &\leq C, \\ |\psi(t, x_1, z) - \psi(t, x_2, z)| &\leq C (1 + |z|) (1 + |x_1| + |x_2|)^m |x_2 - x_1|, \\ |\psi(t, x, z_1) - \psi(t, x, z_2)| &\leq C |z_2 - z_1|,\end{aligned}$$

for all  $t \in [0, T]$ ,  $x, x_1, x_2 \in H$ ,  $z, z_1, z_2 \in \Xi^*$ .

Therefore by Theorem 5.1 the Hamilton Jacobi Bellman equation relative to the above stated problem, namely:

$$\begin{cases} \frac{\partial v(t, x)}{\partial t} + \mathcal{L}_t[v(t, \cdot)](x) = \psi(t, x, \nabla v(t, x)G(t, x)), & t \in [0, T], x \in H, \\ v(T, x) = \phi(x). \end{cases}\tag{7.4}$$

admits a unique mild solution in the sense of the generalized directional gradient. As a consequence, we can perform the synthesis of the optimal control. The following theorem is proved exactly as in [18] Theorem 7.2.

**Theorem 7.2** *Assume Hypothesis 7.1. Let  $v$  be the mild solution of equation (7.4) in the sense of the generalized directional gradient and  $\zeta$  an arbitrary element of  $\widetilde{\nabla}^G v$ .*

*For all a.c.s. we have  $J(t_0, x_0, \mathbb{U}) \geq v(t_0, x_0)$ , and the equality  $J(t_0, x_0, \mathbb{U}) = v(t_0, x_0)$  holds if and only if*

$$u_\tau \in \Gamma(\tau, X_\tau^\mathbb{U}, \zeta(\tau, X_\tau^\mathbb{U})), \quad \mathbb{P} - \text{a.s. for a.a. } \tau \in [t_0, T].$$

*If there exists a measurable function  $\gamma : [0, T] \times H \times \Xi^* \rightarrow U$  with*

$$\gamma(t, x, z) \in \Gamma(t, x, z), \quad t \in [0, T], x \in H, z \in \Xi^*,$$

*then there exists at least an a.c.s.  $\bar{\mathbb{U}} = (\Omega, \mathcal{E}, (\mathcal{F}_t), \mathbb{P}, \bar{u}, X^{\bar{\mathbb{U}}})$  for which*

$$\bar{u}_\tau = \gamma(\tau, X_\tau^{\bar{\mathbb{U}}}, \zeta(\tau, X_\tau^{\bar{\mathbb{U}}})), \quad \mathbb{P} - \text{a.s. for a.a. } \tau \in [t_0, T].\tag{7.5}$$

*Thus  $J(t_0, x_0, \bar{\mathbb{U}}) = v(t_0, x_0)$  and  $\bar{\mathbb{U}}$  is optimal. Finally,  $X^{\bar{\mathbb{U}}}$  is a mild solution of the equation*

$$\begin{cases} dX_\tau^{\bar{\mathbb{U}}} = AX_\tau^{\bar{\mathbb{U}}} d\tau + F(\tau, X_\tau^{\bar{\mathbb{U}}}) d\tau + G(\tau, X_\tau^{\bar{\mathbb{U}}}) dW_\tau \\ \quad + G(\tau, X_\tau^{\bar{\mathbb{U}}}) R\left(\tau, X_\tau^{\bar{\mathbb{U}}}, \gamma(\tau, X_\tau^{\bar{\mathbb{U}}}, \zeta(\tau, X_\tau^{\bar{\mathbb{U}}}))\right) d\tau, & \tau \in [t_0, T], \\ X_{t_0}^{\bar{\mathbb{U}}} = x_0. \end{cases}\tag{7.6}$$

If the function  $\gamma$  above exists then it is called a measurable selection of  $\Gamma$ , and the function  $(t, x) \rightarrow \gamma(t, x, \zeta(t, x))$  is called the optimal feedback law. Assuming that a measurable selection exists, the theorem states in particular that an optimal a.c.s  $\bar{\mathbb{U}}$  is characterized by the equality (7.5) and the corresponding optimal trajectory is a solution of the so-called closed loop equation (7.6). Thus we have shown that the generalized gradient  $\zeta$  of the mild solution of the Hamilton Jacobi Bellman equation (7.4) allows to construct optimal feedback laws and to identify optimal

a.c.s. Finally, if a measurable selection exists, then the mild solution  $v$  of equation (7.4) in the sense of the generalized directional gradient coincides with the so called value function, i.e.

$$v(t, x) = \inf_{\mathbb{U}} J(t, x, \mathbb{U}), \quad t \in [0, T], \quad x \in H,$$

where the infimum is taken with respect to all a.c.s.  $\mathbb{U}$ , and it is achieved for at least one a.c.s.

In the following sections we present two examples of infinite dimensional systems satisfying our assumptions. In particular, the special structure of the state equation arises naturally in some applications, both in finite and infinite dimensional situations.

## 7.1 Example: controlled stochastic equation with delay

We consider the controlled stochastic differential equation in  $\mathbb{R}^n$ :

$$\begin{cases} dx(\tau) = \left[ \int_{-1}^0 x(\tau + \theta) a(d\theta) + f(\tau, x(\tau)) + r(\tau, x(\tau), u(\tau)) \right] d\tau + \sigma(\tau, x(\tau)) dW_\tau, & \tau \in [t_0, T] \\ x(t_0) = \mu_0, & x(t_0 + \theta) = \nu_0(\theta), \quad \text{for a.e. } \theta \in (-1, 0), \end{cases} \quad (7.7)$$

and a cost functional of the form

$$J(t_0, \mu_0, \nu_0, u) = \mathbb{E} \int_{t_0}^T h(\tau, x(\tau), u(\tau)) d\tau + \mathbb{E} k(x(T)),$$

that we minimize over all predictable controls  $u$  with values in  $U \subset \mathbb{R}^N$ .

We assume the following:

- $\mu_0 \in \mathbb{R}^n$ ,  $\nu_0 \in L^2((-1, 0); \mathbb{R}^n)$ ;
- $(\Omega, \mathcal{E}, \mathbb{P})$  is a complete probability space, with a filtration  $(\mathcal{F}_t)_{t \geq 0}$  satisfying the usual conditions and  $\{W_t : t \geq 0\}$  is a cylindrical Wiener process in  $\mathbb{R}^n$ , with respect to  $\mathbb{P}$  and  $(\mathcal{F}_t)$ ;
- $U$  is a Borel subset of  $\mathbb{R}^N$  and  $u$  is a  $(\mathcal{F}_t)_{t \geq 0}$ -predictable process with values in  $U$ ;
- $a$  is a  $L(\mathbb{R}^n, \mathbb{R}^n)$ -valued finite measure on  $[-1, 0]$ ;
- $f : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  is measurable and there exists a constant  $C > 0$  such that

$$|f(t, 0)| \leq C, \quad |f(t, x_1) - f(t, x_2)| \leq C|x_1 - x_2|, \quad t \in [0, T], \quad x_1, x_2 \in \mathbb{R}^n;$$

- $\sigma : [0, T] \times \mathbb{R}^n \rightarrow L(\mathbb{R}^n, \mathbb{R}^n)$  is measurable and for  $t \in [0, T]$ ,  $x, x_1, x_2 \in \mathbb{R}^n$ ,

$$|\sigma(t, 0)| \leq C, \quad |\sigma(t, x_1) - \sigma(t, x_2)| \leq C|x_1 - x_2|, \quad |\sigma^{-1}(t, x)| \leq C;$$

- $r : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}^n$  is measurable and

$$\begin{aligned} |r(t, x, u)| &\leq C, \quad t \in [0, T], \quad u \in U, \quad x \in \mathbb{R}^n, \\ |r(t, x_1, u) - r(t, x_2, u)| &\leq C(1 + |x_1| + |x_2|)^m |x_1 - x_2|, \quad t \in [0, T], \quad u \in U, \quad x_1, x_2 \in \mathbb{R}^n; \end{aligned}$$

- $h : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R} \cup \{+\infty\}$  is measurable and, for some constant  $m \geq 0$ ,

$$\begin{aligned} h(t, 0, u) &\geq -C, & \inf_{v \in U} h(t, 0, v) &\leq C, & t &\in [0, T], \quad u \in U, \\ h(t, x_1, u) &\leq C(1 + |x_1| + |x_2|)^m |x_1 - x_2| + h(t, x_2, u), & t &\in [0, T], \quad u \in U, \quad x_1, x_2 \in \mathbb{R}^n; \end{aligned}$$

- $k : \mathbb{R}^n \rightarrow \mathbb{R}$  verifies

$$|k(x_1) - k(x_2)| \leq C(1 + |x_1| + |x_2|)^m |x_1 - x_2|, \quad x_1, x_2 \in \mathbb{R}^n.$$

Following [9] and [14] we set  $H = \mathbb{R}^n \times L^2((-1, 0); \mathbb{R}^n)$ ,

$$\mathcal{D}(A) = \left\{ \begin{pmatrix} \mu \\ \nu \end{pmatrix} \in H : \nu \in W^{1,2}((-1, 0); \mathbb{R}^n) \text{ and } \nu(0) = \mu \right\},$$

$$A \begin{pmatrix} \mu \\ \nu \end{pmatrix} = \begin{pmatrix} \int_{-1}^0 \nu(\theta) a(d\theta) \\ \frac{d\nu}{d\theta} \end{pmatrix}.$$

It is proved in [21], among other places, that  $A$  generates a strongly continuous semigroup in  $H$  (see also [14]). Moreover if we set, for  $t \in [0, T]$ ,  $\mu \in \mathbb{R}^n$ ,  $\nu \in L^2((-1, 0); \mathbb{R}^n)$ ,  $u \in U$ ,

$$x_0 = \begin{pmatrix} \mu_0 \\ \nu_0 \end{pmatrix}, \quad F\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}\right) = \begin{pmatrix} f(t, \mu) \\ 0 \end{pmatrix}, \quad G\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}\right) = \begin{pmatrix} \sigma(t, \mu) \\ 0 \end{pmatrix},$$

$$R\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, u\right) = \sigma^{-1}(t, \mu) r(t, \mu, u), \quad \ell\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, u\right) = h(t, \mu, u), \quad \phi\left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}\right) = k(\mu),$$

then equation (7.7) is equivalent (see [9] and [14]) to

$$\begin{cases} dX_\tau = (AX_\tau + F(\tau, X_\tau) + G(\tau, X_\tau)R(\tau, X_\tau, u_\tau)) d\tau + G(\tau, X_\tau) dW_\tau, & \tau \in [t_0, T], \\ X_{t_0} = x_0. \end{cases}$$

Moreover it is easy to verify that Hypothesis 7.1 holds (as in Section 6  $\Xi$  is finite dimensional). Thus Theorem 7.2 can be applied to obtain the synthesis of the optimal control. The optimal control is given in terms of a feedback law that involves the weak directional gradient of the mild solution to the Hamilton Jacobi Bellman equation (7.4).

We notice that  $\psi$  and  $\Gamma$  only depend on the finite dimensional coordinate in  $H$  namely, for  $t \in [0, T]$ ,  $\mu \in \mathbb{R}^n$ ,  $\nu \in L^2((-1, 0); \mathbb{R}^n)$ ,  $z \in (\mathbb{R}^n)^*$

$$\psi\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, z\right) = \psi_0(t, \mu, z) := - \inf_{u \in U} \{h(t, \mu, u) + z\sigma^{-1}(t, \mu)r(t, \mu, u)\} \quad (7.8)$$

$$\Gamma\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, z\right) = \Gamma_0(t, \mu, z) := \left\{u \in U : \psi_0(t, \mu, u) + h(t, \mu, u) + z\sigma^{-1}(t, \mu)r(t, \mu, u) = 0\right\} \quad (7.9)$$

Moreover the Hamilton Jacobi Bellman equation (7.4) is, in the present case, the abstract version of the following equation:

$$\left\{ \begin{array}{l} \frac{\partial v}{\partial t} \left( t, \left( \begin{array}{c} \mu \\ \nu \end{array} \right) \right) + \frac{1}{2} \text{Tr}_{\mathbb{R}^n} \left[ \sigma(t, \mu) \sigma^*(t, \mu) \frac{\partial^2 v}{\partial \mu^2} \left( t, \left( \begin{array}{c} \mu \\ \nu \end{array} \right) \right) \right] \\ + \left\langle \frac{\partial v}{\partial \mu} \left( t, \left( \begin{array}{c} \mu \\ \nu \end{array} \right) \right), \int_{-1}^0 \nu(\theta) a(d\theta) \right\rangle_{\mathbb{R}^n} + \int_{-1}^0 \left\langle \frac{d\nu}{d\theta}(\theta), \frac{\partial v}{\partial \nu} \left( t, \left( \begin{array}{c} \mu \\ \nu \end{array} \right) \right) (\theta) \right\rangle_{\mathbb{R}^n} d\theta \\ + \left\langle f(t, \mu), \frac{\partial v}{\partial \mu} \left( t, \left( \begin{array}{c} \mu \\ \nu \end{array} \right) \right) \right\rangle_{\mathbb{R}^n} = \psi_0 \left( t, \mu, \frac{\partial v}{\partial \mu} \left( t, \left( \begin{array}{c} \mu \\ \nu \end{array} \right) \right)^* \sigma(t, \mu) \right), \\ v \left( T, \left( \begin{array}{c} \mu \\ \nu \end{array} \right) \right) = \phi(\mu), \end{array} \right.$$

where  $t \in [0, T]$ ,  $\mu \in \mathbb{R}^n$ ,  $\nu \in L^2([-1, 0], \mathbb{R}^n)$ . In the above equation we have considered  $\frac{\partial v}{\partial \mu}$  as a vector in  $\mathbb{R}^n$ ,  $\frac{\partial^2 v}{\partial \mu^2}$  as a matrix in  $\mathbb{R}^{n \times n}$ ,  $\frac{\partial v}{\partial \nu}$  as a function in  $L^2((-1, 0); \mathbb{R}^n)$ .

Finally an admissible control system  $\mathbb{U}$  is optimal if and only if

$$u_\tau \in \Gamma_0(\tau, X_\tau^\mathbb{U}, \zeta(\tau, X_\tau^\mathbb{U})) \quad \mathbb{P}\text{-a.s. for a.e. } \tau \in [t_0, T],$$

where  $\zeta : [0, T] \times H \rightarrow (\mathbb{R}^n)^*$  is any element in  $\tilde{\nabla}^G v$ .

**Remark 7.3** The interest of the present example is to show that in a very natural and general finite dimensional framework the introduction of a delay leads to an infinite dimensional system that has the same structure as equation (7.1). For other cases in which the same structure arises see [23].

## 7.2 Example: controlled stochastic reaction diffusion PDE with delay

We consider a controlled stochastic parabolic partial differential equation with delay:

$$\left\{ \begin{array}{l} d_\tau x(\tau, \xi) = \left[ \frac{\partial^2}{\partial \xi^2} x(\tau, \xi) + \int_{-1}^0 x(\tau + \theta, \xi) a(d\theta) + f(\tau, \xi, x(\tau, \xi)) \right] d\tau \\ + r(\tau, \xi, x(\tau, \xi), u(\tau, \xi)) d\tau + \sigma(\tau, \xi, x(\tau, \xi)) dW(\tau, \xi) \quad \tau \in [t_0, T], \xi \in [0, 1], \\ x(\tau, 0) = x(\tau, 1) = 0, \quad \tau \in [t_0, T], \\ x(t_0, \xi) = \mu_0(\xi), \quad x(t_0 + \theta, \xi) = \nu_0(\theta, \xi), \quad \xi \in [0, 1], \theta \in (-1, 0), \end{array} \right. \quad (7.10)$$

and a cost functional of the form

$$J(t_0, \mu_0, \nu_0, u) = \mathbb{E} \int_{t_0}^T \int_0^1 h(\tau, \xi, x(\tau, \xi), u(\tau, \xi)) d\xi d\tau + \mathbb{E} \int_0^1 k(\xi, x(T, \xi)) d\xi$$

that we minimize over a suitable set of predictable controls  $u : \Omega \times [t_0, T] \times [0, 1] \rightarrow \mathcal{U} \subset \mathbb{R}$ .

We assume the following

- $\mu_0 \in L^2([0, 1])$ ,  $\nu_0 \in L^2([-1, 0] \times [0, 1])$ .
- $(\Omega, \mathcal{E}, \mathbb{P})$  is a complete probability space, with a filtration  $(\mathcal{F}_t)_{t \geq 0}$  satisfying the usual conditions and  $\{W_t : t \geq 0\}$  is a cylindrical Wiener process in  $L^2([0, 1])$  relatively to  $\mathbb{P}$  and  $(\mathcal{F}_t)$ ;
- $\mathcal{U}$  is a Borel subset of  $\mathbb{R}$ . Moreover by  $U$  we denote the set of all measurable functions  $[0, 1] \rightarrow \mathcal{U}$  endowed with the Borel the  $\sigma$ -field corresponding to convergence in measure (the Lebesgue measure on  $[0, 1]$ ).
- $a$  is a signed finite measure on  $[-1, 0]$ .
- $f : [0, T] \times [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$  is measurable. Moreover there exists  $C > 0$  such that

$$\begin{aligned} \int_0^1 f(t, \xi, 0)^2 d\xi &\leq C, \quad t \in [0, T], \\ |f(t, \xi, x_1) - f(t, \xi, x_2)| &\leq C|x_1 - x_2|, \quad t \in [0, T], \xi \in [0, 1], x_1, x_2 \in \mathbb{R}. \end{aligned}$$

- $\sigma : [0, T] \times [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$  is measurable. Moreover

$$C^{-1} \leq |\sigma(t, \xi, x)| \leq C, \quad |\sigma(t, \xi, x_1) - \sigma(t, \xi, x_2)| \leq C|x_1 - x_2|,$$

for  $t \in [0, T]$ ,  $\xi \in [0, 1]$ ,  $x, x_1, x_2 \in \mathbb{R}$ .

- $r : [0, T] \times [0, 1] \times \mathbb{R} \times \mathcal{U} \rightarrow \mathbb{R}^n$  is measurable and

$$|r(t, \xi, x, u)| \leq C, \quad |r(t, \xi, x_1, u) - r(t, \xi, x_2, u)| \leq C|x_1 - x_2|,$$

for  $t \in [0, T]$ ,  $u \in \mathcal{U}$ ,  $\xi \in [0, 1]$ ,  $x, x_1, x_2 \in \mathbb{R}$ .

- $h : [0, T] \times [0, 1] \times \mathbb{R} \times \mathcal{U} \rightarrow \mathbb{R}$  is measurable and moreover

$$\begin{aligned} |h(t, \xi, x_1, u) - h(t, \xi, x_2, u)| &\leq C(1 + |x_1| + |x_2|)|x_1 - x_2|, \\ h(t, \xi, 0, u) &\geq -C, \quad \inf_{v \in U} \int_0^1 h(t, \eta, 0, v(\eta)) d\eta \leq C, \end{aligned}$$

for  $t \in [0, T]$ ,  $u \in \mathcal{U}$ ,  $\xi \in [0, 1]$ ,  $x_1, x_2 \in \mathbb{R}$ .

- $k : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$  verifies

$$\begin{aligned} |k(\xi, x_1) - k(\xi, x_2)| &\leq C(1 + |x_1| + |x_2|)|x_1 - x_2|, \quad \xi \in [0, 1], x_1, x_2 \in \mathbb{R}, \\ \int_0^1 |k(\xi, 0)| d\xi &< +\infty. \end{aligned}$$

We set  $\mathcal{H} = L^2([0, 1])$  and

$$\mathcal{D}(\mathcal{A}) = H^2([0, 1]) \cap H_0^1([0, 1]), \quad \mathcal{A}\mu = \frac{\partial^2}{\partial \xi^2} \mu, \quad \forall \mu \in \mathcal{D}(\mathcal{A}).$$

Moreover  $H = \mathcal{H} \times L^2([-1, 0]; \mathcal{H})$  and

$$\begin{aligned} \mathcal{D}(A) &= \left\{ \begin{pmatrix} \mu \\ \nu \end{pmatrix} \in H : \mu \in \mathcal{D}(\mathcal{A}), \nu \in W^{1,2}([-1, 0]; \mathcal{H}) \text{ and } \nu(0) = \mu \right\} \\ A \begin{pmatrix} \mu \\ \nu \end{pmatrix} &= \begin{pmatrix} \mathcal{A}\mu + \int_{-1}^0 \nu(\theta, \cdot) a(d\theta) \\ \frac{d\nu}{d\theta} \end{pmatrix} \end{aligned}$$

It is well known that  $\mathcal{A}$  generates a strongly continuous semigroup  $\{e^{t\mathcal{A}} : t \geq 0\}$  in  $\mathcal{H}$ . Moreover in [7], see Theorem 3.28, it is proved that  $A$  generates a strongly continuous semigroup  $\{e^{tA} : t \geq 0\}$  in  $H$ . Thus point (i) in Hypothesis 2.2 is satisfied.

As in the finite dimensional case, if we define

- $\Xi = L^2([0, 1])$ , so that  $\{W_t : t \geq 0\}$  is a cylindrical Wiener process in  $\Xi$ ,
- $F\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}\right) = \begin{pmatrix} f(t, \cdot, \mu(\cdot)) \\ 0 \end{pmatrix}$ ,
- $G\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}\right)w = \begin{pmatrix} \sigma(t, \cdot, \mu(\cdot))w(\cdot) \\ 0 \end{pmatrix}$ ,  $w \in L^2([0, 1])$ ,
- $R\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, u\right) = \sigma^{-1}(t, \cdot, \mu(\cdot))r(t, \cdot, \mu(\cdot), u(\cdot))$ ,
- $x_0 = \begin{pmatrix} \mu_0(\cdot) \\ \nu_0(\cdot, \cdot) \end{pmatrix}$ ,

then, (see [7]) equation (7.10) is equivalent to:

$$\begin{cases} dX_\tau = (AX_\tau + F(\tau, X_\tau) + G(\tau, X_\tau)R(\tau, X_\tau, u_\tau)) d\tau + G(\tau, X_\tau) dW_\tau, & \tau \in [t_0, T], \\ X_{t_0} = x_0 \in H. \end{cases}$$

Point (ii) in Hypothesis 2.2 and relation (2.4) can be easily checked. Thus to ensure that  $A$ ,  $F$  and  $G$  verify Hypothesis 2.2 it remains to show that (2.3) holds.

In order to do this we start noticing that  $G\left(\tau, \begin{pmatrix} \mu \\ \nu \end{pmatrix}\right)$  is independent on  $\nu$ .

Then, fixed  $\mu, \mu' \in \mathcal{H}$  and  $\tau \in [0, T]$  we write for  $t \geq 0$ ,

$$e^{tA}G\left(\tau, \begin{pmatrix} \mu \\ 0 \end{pmatrix}\right) - e^{tA}G\left(\tau, \begin{pmatrix} \mu' \\ 0 \end{pmatrix}\right) = \begin{pmatrix} V_t \\ U_t \end{pmatrix}$$

For all  $t \in [0, T]$ ,  $V_t \in L(\Xi, \mathcal{H})$  and, for all  $\xi \in \Xi$ ,  $V\xi \in C([0, T], \mathcal{H})$ ; in particular, by the Banach-Steinhaus theorem,  $\sup_{t \in [0, T]} \|V_t\|_{L(\Xi, \mathcal{H})} < \infty$ . Moreover by [7], see Theorems 1.37, 3.25, 3.28, we deduce the following variation of constants formula for  $V$ :

$$V_t\xi = e^{tA}G^0\xi + \int_0^t e^{(t-s)A} \int_{-1}^0 V_{\theta+s}\xi \mathbf{1}_{\theta+s \geq 0} a(d\theta) ds, \quad t \geq 0, \xi \in \Xi, \quad (7.11)$$

where  $G^0\xi = \sigma(\tau, \cdot, \mu(\cdot))\xi(\cdot) - \sigma(\tau, \cdot, \mu'(\cdot))\xi(\cdot)$ . Equality (7.11) expresses the fact that  $V\xi$  is the solution to a delay equation that is written formally as

$$\begin{cases} \frac{d}{dt}V_t\xi = AV_t\xi + \int_{-1}^0 V_{\theta+t}\xi a(d\theta), & t \geq 0, \xi \in \Xi, \\ V_0\xi = G^0\xi, & V_\theta\xi = 0, \quad \theta \in (-1, 0). \end{cases}$$

In [14] Section 11.2.1 it is proved that

$$\|e^{tA}\|_{L_2(\mathcal{H}, \mathcal{H})} \leq Lt^{-\gamma}, \quad \|e^{tA}G^0\|_{L_2(\Xi, \mathcal{H})} \leq Lt^{-\gamma}\|\mu - \mu'\|_{\mathcal{H}}, \quad t > 0, \quad (7.12)$$

where  $\gamma = 1/4$  and the constant  $L \geq 0$  does not depend on  $\tau, \mu, \mu'$ .



Using this estimate in (7.11) we get, for  $\sigma > 0$ ,  $t \geq 0$ ,

$$\begin{aligned} |e^{\sigma\mathcal{A}}V_t|_{L_2(\Xi,\mathcal{H})} &\leq L(t+\sigma)^{-\gamma}|\mu - \mu'|_{\mathcal{H}} \\ &\quad + |a|([-1,0]) \sup_{\theta \in [-1,0]} \int_0^t |e^{(\sigma+t-s)\mathcal{A}}V_{\theta+s}|_{L_2(\Xi,\mathcal{H})} 1_{\theta+s \geq 0} ds. \end{aligned}$$

Fixing  $T_0 \in [0, T]$  we obtain, for  $t + \sigma \leq T_0$ ,

$$\begin{aligned} \sup_{t+\sigma \leq T_0} \sigma^\gamma |e^{\sigma\mathcal{A}}V_t|_{L_2(\Xi,\mathcal{H})} &\leq L|\mu - \mu'|_{\mathcal{H}} + |a|([-1,0]) \sup_{t+\sigma \leq T_0} \sigma^\gamma |e^{\sigma\mathcal{A}}V_t|_{L_2(\Xi,\mathcal{H})} \\ &\quad \times \sup_{t+\sigma \leq T_0} \int_0^t (t-s)^{-\gamma} ds. \end{aligned}$$

We note that both sides are finite, by the first inequality in (7.12) and since  $V$  is bounded in  $L(\Xi, \mathcal{H})$  on  $[0, T]$ . Thus we can choose  $T_0$  independently on  $\tau, \mu, \mu'$  (for simplicity  $T_0 \leq 1$ ) such that

$$\sup_{t+\sigma \leq T_0} \sigma^\gamma |e^{\sigma\mathcal{A}}V_t|_{L_2(\Xi,\mathcal{H})} \leq 2L|\mu - \mu'|_{\mathcal{H}}.$$

Coming back to equation (7.11) and using the last estimate we get for all  $t \in (0, T_0]$

$$|V_t|_{L_2(\Xi,\mathcal{H})} \leq Lt^{-\gamma}|\mu - \mu'|_{\mathcal{H}} + 2L|a|([-1,0]) \left( \int_0^t (t-s)^{-\gamma} ds \right) |\mu - \mu'|_{\mathcal{H}}.$$

Consequently

$$|V_t|_{L_2(\Xi,\mathcal{H})} \leq \widehat{L}t^{-\gamma}|\mu - \mu'|_{\mathcal{H}}, \quad t \in (0, T_0] \quad (7.13)$$

for a suitable constant  $\widehat{L}$  independent on  $\tau, \mu, \mu'$ .

Moreover since  $U_t(\theta) = \begin{cases} V_{t+\theta} & \text{if } t+\theta \geq 0 \\ 0 & \text{if } t+\theta < 0 \end{cases}$  inequality (7.13) immediately implies that

$$|U_t|_{L_2(\Xi, L^2([-1,0]; \mathcal{H}))} \leq \widehat{L}(1-2\gamma)^{-1/2}|\mu - \mu'|_{\mathcal{H}}, \quad t \in (0, T_0].$$

Thus

$$\left| e^{tA}G\left(\tau, \begin{pmatrix} \mu \\ 0 \end{pmatrix}\right) - e^{tA}G\left(\tau, \begin{pmatrix} \mu' \\ 0 \end{pmatrix}\right) \right|_{L_2(\Xi, H)} \leq \widetilde{L}t^{-\gamma}|\mu - \mu'|_{\mathcal{H}}, \quad t \in (0, T_0], \tau \in [0, T]$$

for a suitable constant  $\widetilde{L}$ . Clearly by the semigroup property of  $(e^{tA})_{t \geq 0}$  the same inequality holds for all  $t \in [0, T]$  (changing the value of the constant  $\widetilde{L}$  if it is necessary).

By similar arguments we can show that

$$\left| e^{tA}G\left(\tau, \begin{pmatrix} 0 \\ \mu \end{pmatrix}\right) \right|_{L_2(\Xi, H)} \leq \widetilde{L}t^{-\gamma}, \quad t \in (0, T], \tau \in [0, T],$$

and conclude that Hypothesis 2.2 holds.

Consequently point 1 in Hypothesis 7.1 is verified. Moreover it is immediate to verify that  $R$  satisfies point 3.

Let now for  $t \in [0, T]$ ,  $\begin{pmatrix} \mu \\ \nu \end{pmatrix} \in H$ ,  $u \in U$ ,

$$\ell\left(t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, u\right) = \int_0^1 h(t, \xi, \mu(\xi), u(\xi)) d\xi, \quad \phi\left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}\right) = \int_0^1 k(\xi, \mu(\xi)) d\xi.$$

Again it is easy to verify that points 2 and 4 in Hypothesis 7.1 are verified by  $\phi$  and  $\ell$  respectively (both with  $m = 1$ ).

Summarizing, Hypothesis 7.1 holds and we can apply Theorem 7.2 to obtain the synthesis of the optimal control.

Moreover if we define  $\psi_0 : [0, T] \times [0, 1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\Gamma_0 : [0, T] \times [0, 1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathcal{P}(\mathcal{U})$  by

$$\begin{aligned}\psi_0(t, \xi, x, \chi) &= - \inf_{v \in \mathcal{U}} \left\{ \ell(t, \xi, x, v) + \chi \sigma^{-1}(t, \xi, x) r(t, \xi, x, v) \right\}, \\ \Gamma_0(t, \xi, x, \chi) &= \left\{ v \in \mathcal{U} : \psi_0(t, \xi, x, \chi) + \ell(t, \xi, x, v) + \chi \sigma^{-1}(t, \xi, x) r(t, \xi, x, v) = 0 \right\},\end{aligned}$$

and we assume that  $\Gamma_0(t, \xi, x, \chi)$  is non empty  $\forall t \in [0, T], \forall \xi \in [0, 1], \forall x, \chi \in \mathbb{R}$  and there exists a measurable selection  $\gamma_0 : [0, T] \times [0, 1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathcal{U}$  with  $\gamma_0(t, \xi, x, \chi) \in \Gamma_0(t, \xi, x, \chi)$  then (identifying  $L^2([0, 1])$  with its dual):

$$\psi \left( t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, z \right) = \psi_0 \left( t, \cdot, \mu(\cdot), z(\cdot) \right), \quad \Gamma \left( t, \begin{pmatrix} \mu \\ \nu \end{pmatrix}, z \right) = \Gamma_0 \left( t, \cdot, \mu(\cdot), z(\cdot) \right),$$

for all  $t \in [0, T], \mu \in L^2([0, 1]), \nu \in L^2([-1, 0] \times [0, 1]), z \in L^2([0, 1])$ .

Moreover the Hamilton Jacobi Bellman equation (7.4) is, in the present case, the abstract version of the following equation:

$$\left\{ \begin{aligned} & \frac{\partial v}{\partial t}(t, \mu, \nu)(\xi) + \frac{1}{2} \text{Tr}_{L^2([0, 1])} \left[ \sigma(t, \cdot, \mu(\cdot))^2 \frac{\partial^2 v}{\partial \mu^2}(t, \mu, \nu) \right] \\ & + \int_0^1 \int_{-1}^0 \frac{\partial v}{\partial \mu}(t, \mu, \nu)(\xi) \nu(\theta, \xi) a(d\theta) d\xi + \int_0^1 \int_{-1}^0 \frac{\partial v}{\partial \theta}(\theta, \xi) \frac{\partial v}{\partial \nu}(t, \mu, \nu)(\theta, \xi) d\theta d\xi \\ & + \int_0^1 \left( \frac{\partial^2 \mu}{\partial \xi^2}(\xi) + f(t, \xi, \mu(\xi)) \right) \frac{\partial v}{\partial \mu}(t, \mu, \nu)(\xi) d\xi = \psi_0 \left( t, \xi, \mu(\xi), \frac{\partial v}{\partial \mu}(t, \mu, \nu)(\xi) \sigma(t, \xi, \mu(\xi)) \right), \\ & v(T, \mu, \nu) = \int_0^1 k(\xi, \mu(\xi)) d\xi, \end{aligned} \right.$$

where we have identified  $\frac{\partial v}{\partial \mu}$  with an element of  $L^2([0, 1])$ ,  $\frac{\partial v}{\partial \nu}$  with an element of  $L^2([-1, 0] \times [0, 1])$ ,  $\frac{\partial^2 v}{\partial \mu^2}$  with a bounded linear operator from  $L^2([0, 1])$  to itself and  $\sigma(t, \cdot, \mu(\cdot))^2 \frac{\partial^2 v}{\partial \mu^2}$  denotes the composition of  $\frac{\partial^2 v}{\partial \mu^2}$  and multiplication by  $\sigma(t, \cdot, \mu(\cdot))^2$ .

Finally an admissible control system  $\mathbb{U}$  is optimal if and only if

$$u(\tau, \xi) \in \Gamma_0(\tau, \xi, X_\tau^\mathbb{U}(\xi), \zeta(\tau, X_\tau^\mathbb{U})(\xi)), \quad \mathbb{P}\text{-a.s. for a.e. } \xi \in [0, 1], \tau \in [t_0, T].$$

where  $\zeta : [0, T] \times H \rightarrow L^2([0, 1])$  is any element in  $\widetilde{\nabla}^G v$  (here we have identified  $L^2([0, 1])$  with its dual) and there exists an optimal admissible control system  $\overline{\mathbb{U}}$  for which:

$$u(\tau, \xi) = \gamma_0(\tau, \xi, X_\tau^{\overline{\mathbb{U}}}(\xi), \zeta(\tau, X_\tau^{\overline{\mathbb{U}}}(\xi))), \quad \mathbb{P}\text{-a.s. for a.e. } \xi \in [0, 1], \tau \in [t_0, T].$$

## A Appendix: on the connections with the Clarke gradient

Several notions of gradients have been introduced in convex and non-smooth analysis: see for instance [1], [4], [11], [12] and [33]. It turns out that the generalized directional gradient introduced in Section 3 for a function  $u$  which is only assumed to be locally Lipschitz continuous is strictly related with the Clarke gradient (see [4], [11] and [12]) whose definition we are now going to recall. We consider a Borel measurable function  $u : [0, T] \times H \rightarrow \mathbb{R}$  such that for every integer  $n$  there exists  $L_n > 0$  such that

$$|u(t, x) - u(t, y)| \leq L_n |x - y|, \quad \text{for} \quad |x| \leq n, |y| \leq n, t \in [0, T],$$

and we will consider the gradient which corresponds to increments with respect to the variable  $x \in H$  only; thus  $t$  is fixed. For arbitrary direction  $h \in H$  one first defines

$$u^0(t, x; h) = \limsup_{\substack{y \rightarrow x \\ \lambda \downarrow 0}} \frac{u(t, y + \lambda h) - u(t, x)}{\lambda}.$$

Then the Clarke gradient at  $x$  is defined by

$$\partial u(t, x) = \{\eta \in H^* : \eta h \leq u^0(t, x; h) \text{ for all } h \in H\}.$$

For later use we recall that the mapping  $(x, h) \rightarrow u^0(t, x; h)$  is upper semicontinuous (see [12, Chapter 2, Proposition 1.1]).

It is convenient for our purposes to extend this definition slightly. Given  $G$  satisfying Hypothesis 2.2-(iii), we introduce the directional Clarke gradient

$$\partial^G u(t, x) = \{\zeta \in \Xi^* : \zeta \xi \leq u^0(t, x; G(t, x)\xi) \text{ for all } \xi \in \Xi\}.$$

**Remark A.1** Clearly, if  $\eta \in \partial u(t, x)$  then  $\eta G(t, x) \in \partial^G u(t, x)$ .

Now let us recall that for any Borel measurable function  $u : [0, T] \times H \rightarrow \mathbb{R}$  satisfying, for some  $C > 0$  and  $r \geq 0$ ,

$$|u(t, x) - u(t, y)| \leq C|x - y|(1 + |x| + |y|)^r, \quad |u(t, 0)| \leq C, \quad t \in [0, T], x, y \in H, \quad (\text{A.1})$$

we have defined the generalized directional gradient  $\tilde{\nabla}^G u$ . The elements of  $\tilde{\nabla}^G u$  are functions  $\zeta : [0, T] \times H \rightarrow \Xi^*$  with the properties stated in Theorem 3.1. The question arises what is the connection between  $\tilde{\nabla}^G u$  and the directional Clarke gradient  $\partial^G u$ . Since the function  $\partial^G u$  is in general multivalued, one could expect the inclusion  $\zeta(t, x) \in \partial^G u(t, x)$  to hold. We can show that such an inclusion holds if the gradients are computed along the trajectories of the Markov process  $X$ , in the sense specified by the following proposition.

**Proposition A.2** *Assume that Hypothesis 2.2 holds and that  $u : [0, T] \times H \rightarrow \mathbb{R}$  is a measurable function satisfying (A.1). Then for every  $\zeta \in \tilde{\nabla}^G u$  and for every  $t \in [0, T]$ ,  $x \in H$ ,*

$$\zeta(\tau, X(\tau, t, x)) \in \partial^G u(\tau, X(\tau, t, x)), \quad \mathbb{P}\text{-a.s. for a.e. } \tau \in [t, T]$$

**Proof.** The argument partially follows the proof of Proposition 2.15 in [4]. Our starting point is formula (4.20), which states that for every  $t \in [0, T]$  and  $x \in H$

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

where  $\{U_\tau^{t,x}, \tau \in [t, T]\}$  is the process constructed in Proposition 4.4. By Remark A.1 it suffices to show that

$$U_\tau^{t,x} \in \partial u(\tau, X(\tau, t, x)), \quad \mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T]. \quad (\text{A.2})$$

Let us recall how  $U_\tau^{t,x}$  was constructed in the proof of Proposition 4.4. Let  $u_n : [0, T] \times H \rightarrow \mathbb{R}$ ,  $n = 1, 2, \dots$ , be mappings satisfying the conditions stated in Lemma 4.3. Then the sequence  $\nabla u_n(\cdot, X(\cdot))$  is bounded in  $L^p(\Omega \times [t, T]; H^*)$  for every  $p \in [2, \infty)$ . Choosing any subsequence that converges weakly in this space  $U^{t,x}$  is defined as its weak limit (more precisely,  $U^{t,x}$  is taken as a predictable modification of the limit).

In order to prove (A.2) we need the following explicit construction of the functions  $u_n$ , slightly different from the one indicated in the proof of Lemma 4.3. Let  $\rho_n : \mathbb{R}^n \rightarrow \mathbb{R}$  be infinitely differentiable nonnegative functions, with support in the unit ball and satisfying  $\int_{\mathbb{R}^n} \rho_n(\eta) d\eta = 1$ . Let  $\{e_k\}$  be an orthonormal basis of  $H$ ; we define  $\pi_n$  to be the projection in  $H$  onto the space spanned by  $\{e_1, \dots, e_n\}$  and the mapping  $I_n : \mathbb{R}^n \rightarrow H$  given by

$$I_n(\eta_1, \dots, \eta_n) = \sum_{i=1}^n \eta_i e_i, \quad \eta = (\eta_1, \dots, \eta_n) \in \mathbb{R}^n.$$

We set

$$\bar{u}_n(t, x) = \int_{\mathbb{R}^n} u \left( t, \pi_n x - \frac{1}{n} I_n(\eta) \right) \rho_n(\eta) d\eta.$$

We take an infinitely differentiable cut-off function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\phi(\eta) = 1$  for  $\eta \in [0, 1]$  and  $\phi(\eta) = 0$  for  $\eta > 2$  and we finally define

$$u_n(t, x) = \bar{u}_n(t, x) \phi(|x|/n), \quad t \in [0, T], x \in H. \quad (\text{A.3})$$

Then for  $\lambda > 0$ ,  $h \in H$ ,

$$\begin{aligned} & [\bar{u}_n(t, x + \lambda h) - \bar{u}_n(t, x)] / \lambda \\ &= \frac{1}{\lambda} \int_{\mathbb{R}^n} \left[ \bar{u} \left( t, \pi_n x + \lambda \pi_n h - \frac{1}{n} I_n(\eta) \right) - \bar{u} \left( t, \pi_n x - \frac{1}{n} I_n(\eta) \right) \right] \rho_n(\eta) d\eta. \end{aligned}$$

Letting  $\lambda \rightarrow 0$  and using the Fatou lemma we obtain

$$\nabla \bar{u}_n(t, x) h \leq \int_{\mathbb{R}^n} u^0 \left( t, \pi_n x - \frac{1}{n} I_n(\eta); \pi_n h \right) \rho_n(\eta) d\eta,$$

with  $u^0$  defined above. So we obtain, for an arbitrary bounded measurable mapping  $h : \Omega \times [t, T] \rightarrow H$ ,

$$\mathbb{E} \int_t^T \nabla \bar{u}_n(\tau, X(\tau, t, x)) h_\tau d\tau \leq \mathbb{E} \int_t^T \int_{\mathbb{R}^n} u^0 \left( \tau, \pi_n X(\tau, t, x) - \frac{1}{n} I_n(\eta); \pi_n h_\tau \right) \rho_n(\eta) d\eta d\tau. \quad (\text{A.4})$$

Carrying out differentiation in (A.3) it is easy to show that  $U^{t,x}$  is also a weak limit in  $L^p(\Omega \times [t, T]; H^*)$  of an appropriate subsequence of  $\nabla \bar{u}_n(\cdot, X(\cdot, t, x))$ . So the left-hand side of (A.4) converges to  $\mathbb{E} \int_t^T U_\tau^{t,x} h_\tau d\tau$ . By the upper semicontinuity of  $u^0(t, \cdot; \cdot)$  the lim sup of the right-hand side does not exceed  $\mathbb{E} \int_t^T u^0(\tau, X(\tau, t, x); h_\tau) d\tau$ . So we have proved the inequality

$$\mathbb{E} \int_t^T U_\tau^{t,x} h_\tau d\tau \leq \mathbb{E} \int_t^T u^0(\tau, X(\tau, t, x); h_\tau) d\tau. \quad (\text{A.5})$$

We finally show that (A.5) implies (A.2).

We first note that (A.5) implies that for every  $h \in H$  we have

$$\mathbb{P} \otimes l \{U^{t,x}h > u^0(\cdot, X(\cdot, t, x); h)\} = 0, \quad (\text{A.6})$$

where  $\mathbb{P} \otimes l$  denotes the product of the probability measure on  $\Omega$  and the Lebesgue measure on  $[t, T]$ . Indeed suppose that  $\mathbb{P} \otimes l \{U^{t,x}\bar{h} > u^0(\cdot, X(\cdot, t, x); \bar{h})\} > 0$  for some  $\bar{h} \in H$ ; then denoting by  $A$  the set in curly brackets and setting  $h_\tau(\omega) = \bar{h}1_A(\tau, \omega)$  we obtain a contradiction with (A.5).

Let us fix a dense sequence  $(h_n)$  in  $H$ . Then by (A.6),

$$\mathbb{P} \otimes l \{\exists n \in \mathbb{N} : U^{t,x}h_n > u^0(\cdot, X(\cdot, t, x); h_n)\} = 0.$$

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

$$U_\tau^{t,x}h_n \leq u^0(\tau, X(\tau, t, x); h_n), \quad \text{for all } n.$$

Approximating an arbitrary element of  $H$  by an appropriate subsequence of  $(h_n)$  and using the upper semicontinuity of  $u^0$  it follows that

$$U_\tau^{t,x}h \leq u^0(\tau, X(\tau, t, x); h), \quad \mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T],$$

which is the same as (A.2).  $\square$

**Remark A.3** Several interesting questions arise at this stage. For instance, fixed  $\zeta \in \tilde{\nabla}^G u$  we could ask how large is the set of  $(t, x)$  for which  $\zeta(t, x) \in \partial^G u(t, x)$ . Moreover it is natural to ask whether any measurable map  $\ell : [0, T] \times H \rightarrow \Xi^*$  with  $\ell(t, x) \in \partial^G u(t, x)$ ,  $t \in [0, T]$ ,  $x \in H$  belongs to  $\tilde{\nabla}^G u$ . Giving an answer to such questions is anyway beyond the scope of this paper.

## B Appendix: a result on measurability of pseudo-inverses

The following proposition was applied in the proof of Theorem 3.1 to show that solutions of (4.21) can be chosen to depend measurably on  $(t, x)$ .

**Proposition B.1** *Let  $(A, \mathcal{A})$  be a measurable space,  $X$  and  $Y$  two real separable Hilbert spaces,  $y : A \rightarrow Y$  a measurable function and  $T : A \rightarrow L(X, Y)$  a function such that  $Tx : A \rightarrow Y$  is measurable for every  $x \in X$ .*

*Then there exist  $A_0 \in \mathcal{A}$  and a measurable function  $x : A \rightarrow X$  such that:*

- i) for  $\alpha \notin A_0$  there is no  $x \in X$  such that  $T_\alpha x = y_\alpha$ ;*
- ii) we have  $T_\alpha x_\alpha = y_\alpha$  for every  $\alpha \in A_0$ .*

**Proof.** First we note that the assumption on  $T$  is equivalent to the measurability of the map  $T : A \rightarrow L(X, Y)$  when the space  $L(X, Y)$  is endowed with the Borel  $\sigma$ -algebra of the strong operator topology; we say briefly that the map  $T$  is strongly measurable. It is easy to verify that the map  $T^* : A \rightarrow L(Y, X)$  is also strongly measurable.

For any  $\alpha \in A$ , the element  $y_\alpha$  is in the image of  $T_\alpha$  if and only if there exists a constant  $c_\alpha$  such that  $|\langle y_\alpha, x \rangle| \leq c_\alpha |T_\alpha^* x|$  for every  $x \in X$ . This inequality holds for every  $x$  if and only if it holds for  $x$  in a fixed countable dense subset  $D$  of  $X$ . So the equation  $T_\alpha x = y_\alpha$  has a solution if and only if

$$\sup_{x \in D} \frac{|\langle y_\alpha, x \rangle|}{|T_\alpha^* x|} < \infty, \quad (\text{B.1})$$

with the obvious conventions when the denominator is zero. If we define  $A_0$  as the set of all  $\alpha$  for which the left-hand side of (B.1) is infinite, property *i*) follows immediately.

For  $\alpha \notin A_0$  we define  $x_\alpha = 0$  and from now on we consider  $\alpha \in A_0$ .

We first assume that  $\sup_{\alpha \in A_0} |T_\alpha|_{L(X,Y)} < \sqrt{M}$  for some  $M > 0$ .

Since  $y_\alpha$  is in the image of  $T_\alpha$ , we can define  $x_\alpha$  as the unique element satisfying  $T_\alpha x_\alpha = y_\alpha$  and orthogonal to the kernel of  $T_\alpha$ . To show that  $x_\alpha$  depends measurably on  $\alpha$  we introduce the self-adjoint operators  $Q_\alpha = T_\alpha^* T_\alpha$  and consider their spectral decomposition: see e.g. [34] or [31]. Since the spectrum of  $Q_\alpha$  is contained in  $[0, M]$  we have

$$Q_\alpha = \int_{[0,M]} \lambda P_\alpha(d\lambda)$$

for a projection-valued measure  $P_\alpha(d\lambda)$ . We introduce the operators

$$Q_\alpha^{-1} = \int_{(0,M]} \lambda^{-1} P_\alpha(d\lambda)$$

and noting that  $Q_\alpha x_\alpha = T_\alpha^* y_\alpha$  we can easily prove, using symbolic calculus, that  $Q_\alpha x_\alpha$  belongs to the domain of  $Q_\alpha^{-1}$  and

$$Q_\alpha^{-1} T_\alpha^* y_\alpha = x_\alpha - P_\alpha(\{0\})x_\alpha = x_\alpha,$$

the last equality holding because the kernels of  $Q_\alpha$  and  $T_\alpha$  coincide and  $x_\alpha$  was chosen orthogonal to them. For integer  $N > 0$  we define  $\lambda_i^{(N)} = iM2^{-N}$  and note that, for every  $x \in X$ ,

$$\langle x_\alpha, x \rangle = \int_{(0,M]} \lambda^{-1} \langle P_\alpha(d\lambda) T_\alpha^* y_\alpha, x \rangle = \lim_{N \rightarrow \infty} \sum_{i=1}^{2^N} (\lambda_i^{(N)})^{-1} \langle P_\alpha((\lambda_{i-1}^{(N)}, \lambda_i^{(N)}]) T_\alpha^* y_\alpha, x \rangle.$$

This formula shows the measurability of the map  $\alpha \rightarrow x_\alpha$  provided we can prove that, for fixed  $\lambda \in [0, M]$ , the map  $\alpha \rightarrow P_\alpha([0, \lambda])$  is strongly measurable. To this purpose we fix a sequence of polynomials  $p_k(t)$ ,  $t \in \mathbb{R}$ , uniformly bounded on  $[0, M]$ , such that  $p_k(t) \rightarrow 1_{[0, \lambda]}(t)$  for every  $t \in [0, M]$ . By the spectral theorem, for every  $x \in X$ ,

$$|P_\alpha([0, \lambda])x - p_k(Q_\alpha)x|^2 = \int_{[0,M]} |1_{[0, \lambda]}(t) - p_k(t)|^2 \langle P_\alpha(dt)x, x \rangle \rightarrow 0,$$

as  $k \rightarrow \infty$ . This shows that  $P_\alpha([0, \lambda])$  is the strong limit of  $p_k(Q_\alpha)$  and the required measurability follows.

Finally, in the case  $\sup_{\alpha \in A_0} |T_\alpha| = \infty$ , we subdivide  $A_0$  into the measurable sets  $A_n = \{\alpha \in A_0 : n-1 \leq |T_\alpha|_{L(X,Y)} < n\}$ ,  $n = 1, 2, \dots$ , and define  $x_\alpha$  as before on each  $A_n$ . This gives the required function everywhere defined on  $A$ .  $\square$

## Acknowledgements

We wish to express our gratitude to Viorel Barbu for attracting our attention to the relations between our notion of generalized gradient and Clarke's definition. We also thank Fabio Antonelli for useful discussions on the infinite dimensional nature of path dependent options.

## References

- [1] J.P. Aubin, H. Frankowska, Set valued analysis, Birkhäuser, Boston, 1990.
- [2] V. Bally, E. Pardoux, L. Stoica, Backward stochastic differential equations associated to a symmetric markov process, Preprint n 4454, INRIA Rocquencourt, 2002.
- [3] V. Bally, A. Matoussi, Weak solutions for SPDEs and backward doubly stochastic differential equations, J. Theoret. Probab., 14(2001), no. 1, 125–164.
- [4] V. Barbu, Analysis and control of nonlinear infinite dimensional systems, Academic Press Inc., San Diego, 1993.
- [5] J. Barraquand, T. Pudet, Pricing of American path-dependent contingent claims, Mathematical Finance 6 (1996), 17–51.
- [6] E. Barucci, S. Polidoro, V. Vespri, Some results on partial differential equations and Asian options. Math. Models Methods Appl. Sci., 11 (2001), no. 3, 475–497
- [7] A. Bátkai, S. Piazzera, Semigroups for delay equations in  $L^p$ -phase spaces, to appear, Springer Verlag.
- [8] T. Björk, Arbitrage theory in continuous time. Oxford University Press, Oxford 1998.
- [9] A. Chojnowska-Michalik, Representation theorem for general stochastic delay equations, Bull. Acad. Pol. Sci. Ser. Sci. Math. 26 (1978), no. 7, 634–641.
- [10] E. Cinlar, J. Jacod, P. Protter, M.J. Sharpe, Semimartingale and Markov processes, Z. Wahrsch., 54 (1980), 161–219.
- [11] F.H. Clarke, Optimization and nonsmooth analysis, Classics in Applied Mathematics n.5, SIAM, Philadelphia, 1989.
- [12] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, P.R. Wolenski, Nonsmooth analysis and control theory, Graduate Texts in Mathematics n.178, Springer Verlag, 1998.
- [13] J. Cvitanic, I. Karatzas, Hedging contingent claims with constrained portfolios, Ann. Appl. Probab. 3 (1993), 652–681.
- [14] G. Da Prato, J. Zabczyk, Ergodicity for infinite-dimensional systems. London Mathematical Society Lecture Note Series, 229. Cambridge University Press, Cambridge, 1996.
- [15] N. El Karoui, S. Peng, M. C. Quenez, Backward stochastic differential equations in finance. Mathematical Finance 7 (1997), no. 1, 1-71.
- [16] N. El Karoui, L. Mazliak ed.: Backward Stochastic Differential Equations, Pitman Research Notes in Mathematics Series 364, Longman, 1997.
- [17] W. H. Fleming, H. M. Soner, Controlled Markov processes and viscosity solutions. Applications of Mathematics 25. Springer-Verlag, 1993.
- [18] M. Fuhrman, G. Tessitore, Nonlinear 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.

- [19] M. Fuhrman, G. Tessitore, Infinite horizon backward stochastic differential equations and elliptic equations in Hilbert spaces. *Ann. Probab.* 32 (2004), no. 1B, 607–660.
- [20] A. Grorud, E. Pardoux, Intégrales Hilbertiennes anticipantes par rapport à un processus de Wiener cylindrique et calcul stochastique associé. *Appl. Math. Optim.* 25 (1992), 31-49.
- [21] J. Hale, Theory of functional differential equations, *Appl. Math. Sci.*, vol. 3, Springer Verlag, 1971.
- [22] J. Ma, J. Yong, Forward-backward stochastic differential equations and their applications, *Lecture Notes in Mathematics* 1702, Springer-Verlag, 1999.
- [23] F. Masiero, Semilinear Kolmogorov equations and applications to stochastic optimal control. To appear on *Appl. Math. Optim.*
- [24] M. Musiela, M. Rutkowski, Martingale methods in financial modelling, Springer Verlag 1997.
- [25] D. Nualart, The Malliavin calculus and related topics. *Probability and its Applications*. Springer-Verlag, New York, 1995.
- [26] D. Nualart, E. Pardoux, Stochastic calculus with anticipative integrands. *Probab. Th. Rel. Fields* 78 (1988), 535-581.
- [27] S. Peszat, J. Zabczyk, Strong Feller property and irreducibility for diffusions on Hilbert spaces. *Ann. Probab.* 23 (1995), no. 1, 157-172.
- [28] E. Pardoux, BSDE's, weak convergence and homogeneization of semilinear PDE's. In: *Non-linear analysis, differential equations and control*, eds. F.H. Clarke, R.J. Stern, 503-549, Kluwer, 1999.
- [29] E. Pardoux, S. Peng, Adapted solution of a backward stochastic differential equation. *Systems and Control Lett.* 14, 1990, 55-61.
- [30] 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.
- [31] M. Reed, B. Simon, *Methods of modern mathematical physics*, Academic Press 1972-1979.
- [32] D. Revuz, M. Yor, *Continuous martingales and brownian motion*, Springer Verlag 1991.
- [33] R.T. Rockafellar, *Convex analysis*. Princeton Mathematical Series, vol. 28, Princeton University Press, Princeton, New York, 1970.
- [34] W. Rudin, *Functional analysis*, McGrah-Hill, 1973.
- [35] F. Russo, P. Vallois, Forward, backward and symmetric stochastic integration. *Probab. Theory Related Fields* 97 (1993), no. 3, 403–421.
- [36] F. Russo, P. Vallois, The generalized covariation process and Ito formula. *Stochastic Process. Appl.* 59 (1995), no. 1, 81–104.
- [37] F. Russo, P. Vallois, Ito formula for  $C^1$ -functions of semimartingales. *Probab. Theory Related Fields* 104 (1996), no. 1, 27–41.



- [38] F. Russo, P. Vallois, Stochastic calculus with respect to continuous finite quadratic variation processes. *Stochastics Stochastics Rep.* 70 (2000), no. 1-2, 1–40.
- [39] M. Sharpe, General theory of Markov processes. *Pure and Applied Mathematics*, 133. Academic Press, Inc., Boston, MA, 1988.
- [40] P. Willmott, J. Dewynne, S. Howison, *Option pricing*, Oxford Financial Press, Oxford, 1993.