Chapter 6
HJB Equations Through Backward Stochastic Differential Equations

Marco Fuhrman and Gianmario Tessitore

This last chapter of the book completes the picture of the main methods used to study second-order HJB equations in Hilbert spaces and related optimal control problems by presenting a survey of results that can be achieved with the techniques of Backward SDEs in infinite dimension.

The chapter has been written independently and autonomously. In order to maintain some coherence with the notation used in the Backward SDE literature, the notation used in this chapter is not always identical to that in the rest of the book. This is explained in Sects. 6.1.1 and 6.1.2.

The chapter has the following structure.

- Section 6.1 explains the basic notation and collects some useful results about generalized gradients and SDEs which are needed in the rest of the chapter.
- Section 6.2 provides results about regular dependence of solutions of SDEs on the data.
- Section 6.3 presents results about well-posedness and regular dependence on the data for Backward SDEs (BSDEs from now on) and Forward–Backward systems (FBSDEs) in Hilbert spaces.
- In Sect. 6.4 existence and uniqueness of mild solutions of HJB equations through FBSDEs are discussed.
- Section 6.5 gives applications of the results of Sect. 6.4 to optimal control problems. An example of a control problem with delay is studied in Sect. 6.6.

M. Fuhrman
Dipartimento di Matematica, Università degli studi di Milano, Milano, Italy

G. Tessitore (✉)
Dipartimento di Matematica e Applicazioni, Università degli studi di Milano-Bicocca, Milano, Italy
e-mail: gianmario.tessitore@unimib.it

G. Fabbri et al., *Stochastic Optimal Control in Infinite Dimension*,
Probability Theory and Stochastic Modelling 82,
DOI 10.1007/978-3-319-53067-3_6

- Sections 6.7–6.10 develop the same program for elliptic HJB equations and infinite horizon control problems. An application to an infinite horizon optimal control problem driven by a heat equation with additive noise is discussed in Sect. 6.11.
- Results for elliptic HJB equations with non-constant second-order coefficients and some applications are collected in Sect. 6.12.

6.1 Complements on Forward Equations with Multiplicative Noise

6.1.1 Notation on Vector Spaces and Stochastic Processes

The notation for Banach spaces and linear operators between them is the same as that used in the other parts of the book, see, for instance, Appendix A.1.

In this chapter the letters Ξ, H, K will always denote Hilbert spaces. The scalar product is denoted, as usual, by $\langle \cdot, \cdot \rangle$, with a subscript to specify the space, if necessary. All Hilbert spaces are assumed to be real and separable.

We only consider stochastic differential equations driven by *cylindrical* Wiener processes W. By a cylindrical Wiener process with values in a Hilbert space Ξ, defined on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we mean a family $W(t)$, $t \geq 0$, of linear mappings $\Xi \to L^2(\Omega)$ such that

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

Recall that, in this case, when the noise space Ξ has finite dimension d the Wiener process can be naturally identified with a d-dimensional standard Wiener process $(\beta_1, \ldots, \beta_d)$, where $\beta_i(t) = W(t)e_i$ and $(e_1, \ldots, e_d)$ denotes an orthonormal basis of Ξ. In other parts of the book Q-Wiener processes and in particular cylindrical Wiener processes are introduced in a slightly different (but equivalent) way, see Sect. 1.2.4 and in particular Remark 1.89.

Unless stated otherwise, $\{\mathscr{F}_t\}_{t\geq 0}$ will denote the natural filtration of W, augmented by the family $\mathcal{N}$ of $\mathbb{P}$-null sets of $\mathscr{F}$:

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

The filtration $\mathscr{F}_t$ satisfies the usual conditions. All the concepts of measurability for stochastic processes (e.g. adaptedness, predictability etc.) refer to this filtration. By $\mathcal{P}$ we denote the predictable σ-field on $\Omega \times [0, \infty)$ or (by abuse of notation) its trace on $\Omega \times [0, T]$.

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

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

To denote the value of a process X at time s, sometimes instead of $X(s)$ the shortened notation X_s will be used, especially in proofs. The short-hand "a.a. (a.e.)" means "almost all (almost everywhere) with respect to the Lebesgue measure".

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

- $L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$ denotes the space of equivalence classes of processes $Y \in L^2(\Omega \times [0, T]; K)$, admitting a predictable version. $L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$ is endowed with the norm
$$|Y|^2 = \mathbb{E}\int_0^T |Y(s)|^2 ds.$$
- $L^p_{\mathcal{P}}(\Omega; L^2([0, T]; K))$ denotes the space of equivalence classes of processes Y such that the norm
$$|Y|^p = \mathbb{E}\left(\int_0^T |Y(s)|^2 ds\right)^{p/2}$$
is finite, and Y admits a predictable version.
- $C_{\mathcal{P}}([0, T], L^2(\Omega; K))$ denotes the space of K-valued processes Y such that $Y : [0, T] \to L^2(\Omega; K)$ is continuous and Y has a predictable modification, endowed with the norm
$$|Y|^2 = \sup_{s\in[0,T]} \mathbb{E}\,|Y(s)|^2.$$
Elements of $C_{\mathcal{P}}([0, T], L^2(\Omega; K))$ are identified up to modification.
- $L^p_{\mathcal{P}}(\Omega; C([0, T], K))$ denotes the space of predictable processes Y with continuous paths in K, such that the norm
$$|Y|^p = \mathbb{E} \sup_{s\in[0,T]} |Y(s)|^p$$
is finite. Elements of $L^p_{\mathcal{P}}(\Omega; C([0, T], K))$ are identified up to indistinguishability.

Recall that, for a given element Ψ of $L^2_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$, the Itô stochastic integral $\int_0^t \Psi(s)\, dW(s)$, $t \in [0, T]$, is a K-valued martingale belonging to $L^2_{\mathcal{P}}(\Omega; C([0, T], K))$.

If Ψ belongs to $L^2_{\mathcal{P}}(\Omega \times [0, T]; \Xi)$, the real-valued Itô stochastic integral $\int_0^t \langle\Psi(s), dW(s)\rangle_\Xi$ is by definition the integral $\int_0^t \Psi(s)^*\, dW(s)$, where $\Psi(\omega, s)^* \in \Xi^*$ denotes the element corresponding to $\Psi(\omega, s) \in \Xi$ by the Riesz isometry (i.e., $\Psi(\omega, s)^* h = \langle\Psi(\omega, s), h\rangle_\Xi$, $h \in \Xi$).

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

6.1.2 The Class $\mathcal{G}$

In this section we introduce a class of maps acting among Banach spaces, possessing suitable continuity and differentiability properties. Many assumptions in the following sections will be stated in terms of membership in this class.

The class we are going to introduce has several useful properties. First, membership in this class is often easy to verify: see Lemmas 6.4 and 6.6 below. Next, it is a well-behaved class as far as chain rules are concerned. Finally, it is sufficiently large to include operators commonly arising in applications to stochastic partial differential equations, such as Nemytskii (evaluation) operators; it is well known that the Nemytskii operators are not Fréchet differentiable except in trivial cases.

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

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

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

Remark 6.1 When $V = \mathbb{R}$ the Gâteaux derivative $\nabla F(x)$ belongs to $\mathcal{L}(X, \mathbb{R}) = X^*$, the dual space of X. If, in addition, X is a Hilbert space then it can be identified canonically with X^* and the Gâteaux derivative of F at x can be thought of as an element of X that we denote by $DF(x)$. Thus, $DF(x)$ is the unique element of X such that $\nabla F(x; h) = \nabla F(x)h = \langle DF(x), h\rangle_X$ for every $h \in X$. Similarly, in the same circumstances, the second Gâteaux derivative will be identified with a (symmetric) element of $\mathcal{L}(X)$, denoted by $D^2F(x)$. This convention is a little different from the rest of the book, where the notation $DF(x)$ is employed for the Fréchet derivative of F at x. ■

Definition 6.2 We say that a mapping $F : X \to V$ belongs to the class $\mathcal{G}^1(X, V)$ if it is continuous, Gâteaux differentiable on X, and $\nabla F : X \to \mathcal{L}(X, V)$ is strongly continuous.

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

Lemma 6.3 *Suppose $F \in \mathcal{G}^1(X, V)$. Then*

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

Proof (i) Let $x_n \to x$ and $h_n \to h$ in X. By the Banach–Steinhaus theorem we have $|\nabla F(x_n)|_{\mathcal{L}(X,V)} < L$ for every n and for a suitable constant L. Therefore

$$|\nabla F(x_n)h_n - \nabla F(x)h| \leq L|h - h_n| + |\nabla F(x_n)h - \nabla F(x)h|$$

and the claim follows immediately.

(ii) First we notice that for all $x,\ y \in H$:

$$F(x+y) = F(x) + \int_0^1 \nabla F(x+ry)y\,dr. \tag{6.1}$$

Therefore, given $x, h \in X$, $s \in (0,1]$, repeated application of (6.1) yields

$$G(F(x+sh)) - G(F(x)) \\ = \int_0^1 \left[\nabla G\left(F(x) + \sigma \int_0^1 \nabla F(x+srh)shdr\right) \int_0^1 \nabla F(x+srh)shdr\right] d\sigma.$$

Let $g(s) = \int_0^1 \nabla F(x+srh)hdr$, $K = \{\nabla F(x+rh)h\ :\ r \in [0,1]\}$ and $\hat{K}$ be the closed convex hull of K. Clearly K, and hence $\hat{K}$, are compact subsets of V and $g(s) \in \hat{K}$ for all $s \in [0,1]$. Moreover,

$$\{F(x) + \sigma s \int_0^1 \nabla F(x+srh)dr : \sigma \in [0,1], s \in [0,1]\} \\ \subset \hat{K}_1 := \{F(x) + \sigma k : \sigma \in [0,1], k \in \hat{K}\},$$

which is itself compact. By the dominated convergence theorem $\lim_{s\to 0^+} g(s) = \nabla F(x)h$ and since, by the continuity of ∇G, $\sup_{z\in\hat{K}_1, k\in\hat{K}} |\nabla G(z)k| < +\infty$, applying again the dominated convergence theorem we can conclude that

$$\lim_{s\to 0^+} \frac{G(F(x+sh)) - G(F(x))}{s} \\ = \int_0^1 \lim_{s\to 0^+} [\nabla G(F(x) + \sigma s g(s))g(s)]\,d\sigma = \nabla G(F(x))\nabla F(x)h.$$

The proof that the map $x \to \nabla G(F(x))\nabla F(x)$ is strongly continuous is identical to the proof of point (i). □

In addition to the ordinary chain rule in point (ii) above, a chain rule for the Malliavin derivative operator holds: see Sect. 6.2.2. Membership of a map in $\mathcal{G}^1(X,V)$ may be conveniently checked as shown in the following lemma.

Lemma 6.4 *A map $F : X \to V$ belongs to $\mathcal{G}^1(X,V)$ provided the following conditions hold:*

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

Proof We have to show that F is continuous and the map $h \to \nabla F(x; h)$, where $\nabla F(x; h)$ denotes the directional derivative of F at a fixed point $x \in X$ in the direction $h \in X$, is linear. To start, we notice that a version of formula (6.1) still holds under the present assumptions, namely: $F(x + y) = F(x) + \int_0^1 \nabla F(x + ry; y)\, dr$ for all $x, y \in X$.

First we show linearity. By definition of the directional derivative it is obvious that for all $\rho \geq 0$ and all $x, h \in X$: $\nabla F(x, \rho h) = \rho \nabla F(x, h)$. Since, for fixed $h, k \in X$,

$$\frac{F(x + s(h + k)) - F(x)}{s} = \frac{F(x + s(h + k)) - F(x + sh)}{s} + \frac{F(x + sh) - F(x)}{s},$$

we have, by (6.1),

$$\nabla F(x; h + k) = \lim_{s \to 0^+} \int_0^1 \nabla F(x + sh + rsk; k)\, dr + \nabla F(x; h),$$

provided the limit exists. The continuity of $\nabla F(\cdot; k)$ implies that we can pass to the limit under the integral, by a dominated convergence argument, obtaining $\nabla F(x; h + k) = \nabla F(x; k) + \nabla F(x; h)$. It follows, in particular, that $\nabla F(x; -h) = -\nabla F(x; h)$ and so $\nabla F(x, \rho h) = \rho \nabla F(x, h)$ for all $\rho \in \mathbb{R}$ and all $x, h \in X$. Linearity is proved. From now on, we denote the directional derivative $\nabla F(x; k)$ by $\nabla F(x)k$.

Now we come to the continuity of F. Let $y_n \to 0$ in X and fix $x \in X$. By (6.1) we have: $F(x + y_n) - F(x) = \int_0^1 \nabla F(x + ry_n)y_n\, dr$. We see that the set $\{x + ry_n : r \in [0, 1], n \in \mathbb{N}\}$ is a compact subset of X. Therefore (using again the Banach–Steinhaus theorem) $\sup_{r \in [0,1], n \in \mathbb{N}} |\nabla F(x + ry_n)|_{\mathcal{L}(X,V)} < +\infty$ and we can apply the dominated convergence theorem to conclude that $F(x + y_n) - F(x) \to 0$. □

We need to generalize these definitions to functions depending on several variables. For a function $F : X \times Y \to V$ the partial directional and Gâteaux derivatives with respect to the first argument, at point (x, y) and in the direction $h \in X$, are denoted $\nabla_x F(x, y; h)$ and $\nabla_x F(x, y)$, respectively, their definitions being obvious.

Definition 6.5 We say that a mapping $F : X \times Y \to V$ belongs to the class $\mathcal{G}^{1,0}(X \times Y, V)$ if it is continuous, Gâteaux differentiable with respect to x on $X \times Y$, and $\nabla_x F : X \times Y \to \mathcal{L}(X, V)$ is strongly continuous.

As in Lemma 6.3 one can prove that for $F \in \mathcal{G}^{1,0}(X \times Y, V)$ the mapping $(x, y, h) \to \nabla_x F(x, y)h$ is continuous from $X \times Y \times X$ to V, and analogues of the previously stated chain rules hold. The following result is proved in the same way as Lemma 6.4 (but note that continuity is explicitly required).

Lemma 6.6 *A continuous map* $F : X \times Y \to V$ *belongs to* $\mathcal{G}^{1,0}(X \times Y, V)$ *provided the following conditions hold:*

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

The previous definitions and properties have obvious generalizations to slightly different situations, provided obvious changes are made. For instance, the space Y might be replaced by an interval $[0, T]$ or $[0, \infty)$. Another situation occurs when F depends on additional arguments: for instance, we say that $F : X \times Y \times Z \to V$ belongs to $\mathcal{G}^{1,1,0}(X \times Y \times Z, V)$ if it is continuous, Gâteaux differentiable with respect to x and y on $X \times Y \times Z$, and $\nabla_x F : X \times Y \times Z \to \mathcal{L}(X, V)$ and $\nabla_y F : X \times Y \times Z \to \mathcal{L}(Y, V)$ are strongly continuous.

We will make systematic use of a parameter-dependent contraction principle, stated below as Proposition 6.7. It will be used to study regular dependence of solutions to stochastic equations on their initial data, which is crucial to the investigation of regularity properties of the nonlinear Kolmogorov equation which is the object of this Chapter. The first part of the following proposition is proved in [582], Theorems 10.1, 10.2 (see also [106] Appendix C). The second part is an immediate corollary.

Proposition 6.7 (Parameter-dependent contraction principle) *Let* $F : X \times Y \to X$ *be a continuous mapping satisfying*

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

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

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

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

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

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

$$\nabla_Y \phi(y, z) = \nabla_X F(\phi(y, z), y, z) \nabla_Y \phi(y, z) + \nabla_Y F(\phi(y, z), y, z), \qquad y \in Y, \ z \in Z. \tag{6.2}$$

6.1.3 The Forward Equation: Existence, Uniqueness and Regularity

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

$$\begin{cases} dX(s) = AX(s)\, ds + b(s, X(s))\, ds + \sigma(s, X(s))\, dW(s), & s \in [t, T], \\ X(t) = x \in H. \end{cases} \tag{6.3}$$

The precise notion of solution will be given next. For the moment we emphasize the fact that W will only denote a cylindrical Wiener process. Other cases can be reduced to this one by standard reformulations; for instance, the case of a finite-dimensional driving Brownian motion corresponds to the case where Ξ has finite dimension.

We assume the following:

Hypothesis 6.8 (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 $b : [0, T] \times H \to H$ is measurable and satisfies, for some constant $L > 0$,

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

(iii) σ is a mapping $[0, T] \times H \to \mathcal{L}(\Xi, H)$ such that for every $v \in \Xi$ the map $\sigma v : [0, T] \times H \to H$ is measurable, $e^{sA}\sigma(t, x) \in \mathcal{L}_2(\Xi, H)$ for every $s > 0$, $t \in [0, T]$ and $x \in H$. Moreover, for every $s > 0$, $t \in [0, T]$, $x, y \in H$,

$$\begin{aligned} &|e^{sA}\sigma(t, x)|_{\mathcal{L}_2(\Xi,H)} \leq L\, s^{-\gamma}(1 + |x|), \\ &|e^{sA}\sigma(t, x) - e^{sA}\sigma(t, y)|_{\mathcal{L}_2(\Xi,H)} \leq L\, s^{-\gamma}|x - y| \end{aligned} \tag{6.4}$$

and

$$|\sigma(t, x)|_{\mathcal{L}(\Xi,H)} \leq L\, (1 + |x|), \tag{6.5}$$

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

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

$$b(t, \cdot) \in \mathcal{G}^1(H, H), \qquad e^{sA}\sigma(t, \cdot) \in \mathcal{G}^1(H, \mathcal{L}_2(\Xi, H)).$$

By a solution to Eq. (6.3) we mean an $\mathscr{F}_t$-adapted process $X(s)$, $s \in [t, T]$, with continuous paths in H, such that, $\mathbb{P}$-a.s.

$$X(s) = e^{(s-t)A}x + \int_t^s e^{(s-r)A} b(r, X(r))\, dr + \int_t^s e^{(s-r)A}\sigma(r, X(r))\, dW(r), \quad s \in [t, T]. \tag{6.6}$$

To shorten the notation slightly, we will often write X_s and W_s instead of $X(s)$, $W(s)$. We note that X is clearly a predictable process in H and that the measurability assumption in Hypothesis 6.8-(iii) is needed to ensure that the integrand process $e^{(s-r)A}G(r, X(r))$, $r \in [s, t]$, is a predictable process with values in $\mathcal{L}_2(\Xi, H)$ (endowed with the Borel σ-field). To stress dependence on the initial data we denote the solution by $X(s; t, x)$. Note that $X(s; t, x)$ is $\mathscr{F}_{[t,T]}$-measurable, hence independent of $\mathscr{F}_t$.

The inequality (6.5) and Hypothesis 6.8-(iv) are needed to have additional regularity for the process X, but they are not used in Proposition 6.9 below. It is a consequence of our assumptions that for every $s > 0$, $t \in [0, T]$, $x, h \in H$,

$$|\nabla_x b(t, x)h| \leq L\ |h|, \qquad |\nabla_x (e^{sA}\sigma(t, x))h|_{\mathcal{L}_2(\Xi,H)} \leq L\ s^{-\gamma}|h|. \tag{6.7}$$

Proposition 6.9 *Under the assumptions of Hypothesis 6.8-(i)-(ii)-(iii), for every $p \in [2, \infty)$ there exists a unique process $X \in L^p_{\mathcal{P}}(\Omega; C([t, T], H))$ which is a solution to (6.6). Moreover,*

$$\mathbb{E} \sup_{s\in[t,T]} |X(s; t, x)|^p \leq C(1 + |x|)^p, \tag{6.8}$$

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

Proof The result is well known, see e.g. [177], Theorem 5.3.1. We include the proof for completeness and because it will be useful in the following. We often write X_s for $X(s)$ and similar conventions are used for other stochastic processes. The argument is as follows: we define a mapping Φ from $L^p_{\mathcal{P}}(\Omega; C([t, T], H))$ to itself by the formula

$$\Phi(X)_s = e^{(s-t)A}x + \int_t^s e^{(s-r)A}b(r, X_r)\,dr + \int_t^s e^{(s-r)A}\sigma(r, X_r)\,dW_r, \qquad s \in [t, T],$$

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

For simplicity, we set $t = 0$ and we treat only the case $b = 0$, the general case being handled in a similar way. Let us introduce the norm $\|X\|^p = \mathbb{E} \sup_{s\in[0,T]} e^{-\beta sp}|X_s|^p$, where $\beta > 0$ will be chosen later. In the space $L^p(\Omega; C([0, T], H))$ this norm is equivalent to the original one. We will use the so-called factorization method, see [177], Theorem 5.2.5. Let us take $p > 2$ and $\alpha \in (0, 1)$ such that

$$\frac{1}{p} < \alpha < \frac{1}{2} - \gamma, \qquad \text{and let} \qquad c_\alpha^{-1} = \int_r^s (s-u)^{\alpha-1}(u-r)^{-\alpha}du.$$

Then, by the stochastic Fubini theorem,

$$\Phi(X)_s = e^{sA}x + c_\alpha \int_0^s \int_r^s (s-u)^{\alpha-1}(u-r)^{-\alpha} e^{(s-u)A} e^{(u-r)A}\, du\, \sigma(r, X_r)\, dW_r$$
$$= e^{sA}x + c_\alpha \int_0^s (s-u)^{\alpha-1} e^{(s-u)A} Y_u\, du,$$

where

$$Y_u = \int_0^u (u-r)^{-\alpha} e^{(u-r)A} \sigma(r, X_r)\, dW_r.$$

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

$$e^{-\beta s} \left| \int_0^s (s-u)^{\alpha-1} e^{(s-u)A} Y_u\, du \right| \le \left(\int_0^s e^{-p'\beta(s-u)} (s-u)^{(\alpha-1)p'}\, ds \right)^{\frac{1}{p'}} \cdot$$
$$\cdot \left(\int_0^s e^{-p\beta u} |e^{(s-u)A} Y_u|^p\, du \right)^{\frac{1}{p}}$$
$$\le M \left(\int_0^T e^{-p'\beta u} u^{(\alpha-1)p'}\, du \right)^{\frac{1}{p'}} \left(\int_0^T e^{-p\beta u} |Y_u|^p\, du \right)^{\frac{1}{p}}, \quad (6.9)$$

and we obtain

$$\|\Phi(X)\| \le M|x| + Mc_\alpha \left(\int_0^T e^{-p'\beta u} u^{(\alpha-1)p'}\, du \right)^{\frac{1}{p'}} \left(\mathbb{E} \int_0^T e^{-p\beta u} |Y_u|^p\, du \right)^{\frac{1}{p}}.$$

By the Burkholder–Davis–Gundy inequalities, taking into account the assumption (6.4), we have, for some constant c_p depending only on p,

$$\mathbb{E}\, |Y_u|^p \le c_p \mathbb{E} \left(\int_0^u (u-r)^{-2\alpha} |e^{(u-r)A} \sigma(r, X_r)|^2_{L_2(\Xi,H)}\, dr \right)^{\frac{p}{2}}$$
$$\le L^p c_p \mathbb{E} \left(\int_0^u (u-r)^{-2\alpha-2\gamma} (1 + |X_r|)^2\, dr \right)^{\frac{p}{2}}$$
$$\le L^p c_p \mathbb{E} \sup_{r\in[0,u]} [(1+|X_r|)^p e^{-p\beta r}] \left(\int_0^u (u-r)^{-2\alpha-2\gamma} e^{2\beta r}\, dr \right)^{\frac{p}{2}},$$

which implies

$$e^{-p\beta u} \mathbb{E}\, |Y_u|^p \le L^p c_p (1 + \|X\|^p) \left(\int_0^u (u-r)^{-2\alpha-2\gamma} e^{-2\beta(u-r)}\, dr \right)^{\frac{p}{2}}$$
$$\le L^p c_p (1 + \|X\|^p) \left(\int_0^T r^{-2\alpha-2\gamma} e^{-2\beta r}\, dr \right)^{\frac{p}{2}}.$$

We conclude that

$$\|\Phi(X)\| \le M|x| + MLc_\alpha \left(Tc_p(1+\|X\|^p)\right)^{\frac{1}{p}} \cdot \\ \cdot \left(\int_0^T e^{-p'\beta u} u^{(\alpha-1)p'}\, du\right)^{\frac{1}{p'}} \left(\int_0^T r^{-2\alpha-2\gamma} e^{-2\beta r}\, dr\right)^{\frac{1}{2}}.$$

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

$$\|\Phi(X) - \Phi(X^1)\| \le MLc_\alpha \left(Tc_p\right)^{\frac{1}{p}} \|X - X^1\| \cdot \\ \cdot \left(\int_0^T e^{-p'\beta u} u^{(\alpha-1)p'}\, du\right)^{\frac{1}{p'}} \left(\int_0^T r^{-2\alpha-2\gamma} e^{-2\beta r}\, dr\right)^{\frac{1}{2}},$$

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

In particular, we obtain $\|X\| \le C(1+|x|)$, which proves the estimate (6.8). □

6.2 Regular Dependence on Data

6.2.1 Differentiability

For further developments we need to investigate the dependence of the solution $X(s;t,x)$ on the initial data x and t. We first reformulate Eq. (6.6) as an equation on $[0,T]$. We set

$$S(s) = e^{sA} \quad \text{for} \quad s \ge 0, \qquad S(s) = I \quad \text{for} \quad s < 0, \tag{6.10}$$

and we consider the equation

$$X(s) = S(s-t)x + \int_0^s 1_{[t,T]}(r) S(s-r) b(r, X(r))\, dr \\ + \int_0^s 1_{[t,T]}(r) S(s-r) \sigma(r, X(r))\, dW(r), \tag{6.11}$$

for the unknown process $X(s)$, $s \in [0,T]$. Under the assumptions of Hypothesis 6.8, Eq. (6.11) has a unique solution $X \in L^p_{\mathcal{P}}(\Omega; C([0,T], H))$ for every $p \in [2,\infty)$. It clearly satisfies $X(s) = x$ for $s \in [0,t)$, and its restriction to the time interval $[t,T]$ is the unique solution to (6.6).

From now on we denote by $X(s;t,x)$, $s \in [0,T]$, the solution to (6.11).

Proposition 6.10 *Assume Hypothesis 6.8. Then, for every $p \in [2,\infty)$, the following hold.*

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

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

$$\begin{cases} \nabla_x X(s; t, x)h = e^{(s-t)A}h + \int_t^s e^{(s-r)A} \nabla_x b(r, X(r; t, x)) \nabla_x X(r; t, x)h \, dr \\ \qquad\qquad + \int_t^s \nabla_x (e^{(s-r)A} \sigma(r, X(r; t, x))) \nabla_x X(r; t, x)h \, dW(r), \quad s \in [t, T], \\ \nabla_x X(s; t, x)h = h, \quad s \in [0, t). \end{cases} \tag{6.12}$$

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

Proof Let us consider again the map Φ defined in the proof of Proposition 6.9. In our present notation, Φ can be seen as a mapping from $L^p_{\mathcal{P}}(\Omega; C([0, T], H)) \times [0, T] \times H$ to $L^p_{\mathcal{P}}(\Omega; C([0, T], H))$:

$$\Phi(X, t, x)_s = S(s - t)x + \int_0^s 1_{[t,T]}(r) S(s - r) b(r, X_r)\, dr \\ + \int_0^s 1_{[t,T]}(r) S(s - r) \sigma(r, X_r)\, dW_r,$$

for $s \in [0, T]$. By the arguments of the proof of Proposition 6.9, $\Phi(\cdot, t, x)$ is a contraction in $L^p_{\mathcal{P}}(\Omega; C([0, T], H))$, under an equivalent norm, uniformly with respect to t, x. The process $X(\cdot; t, x)$ is the unique fixed point of $\Phi(\cdot, t, x)$. So, by the parameter-dependent contraction principle (Proposition 6.7), it suffices to show that

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

By an obvious extension of Lemma 6.6, the proof is concluded by the following steps.

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

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

$$\nabla_X \Phi(X, t, x; N)_s = \int_t^s e^{(s-r)A} \nabla_x b(r, X_r) N_r\, dr \\ + \int_t^s \nabla_x (e^{(s-r)A} \sigma(r, X_r)) N_r\, dW_r, \quad s \in [t, T], \\ \nabla_X \Phi(X, t, x; N)_s = 0, \quad s \in [0, t);$$

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

We limit ourselves to proving this claim in the special case $b = 0$, the general case being a straightforward extension. For fixed $t \in [0, T]$ and $x \in H$, for all $s \in [t, T]$:

$$
\begin{aligned}
I_s^\varepsilon &:= \frac{1}{\varepsilon}\Phi(X + \varepsilon N, t, x)_s - \frac{1}{\varepsilon}\Phi(X, t, x)_s - \int_t^s \nabla_x (e^{(s-r)A}\sigma(r, X_r))N_r dW_r \\
&= \int_t^s \left(\int_0^1 \left(\nabla_x (e^{(s-r)A}\sigma(r, X_r + \zeta\varepsilon N_r))N_r - \nabla_x (e^{(s-r)A}\sigma(r, X_r))N_r \right) d\zeta \right) dW_{r.}
\end{aligned}
$$

Proceeding as in the proof of Proposition 6.9 (with $\beta = 0$) we get for $1/p < \alpha < 1/2 - \gamma$ and for a suitable constant c_p:

$$
|I^\varepsilon|^p_{L^p_{\mathcal{P}}(\Omega; C([0,T],H))} \le c_p \mathbb{E} \int_t^T |Y_u^\varepsilon|^p du,
$$

where

$$
\begin{aligned}
Y_u^\varepsilon = \int_t^u (u - r)^{-\alpha} \bigg(\int_0^1 \Big(& \nabla_x (e^{(u-r)A}\sigma(r, X_r + \zeta\varepsilon N_r))N_r \\
& - \nabla_x (e^{(u-r)A}\sigma(r, X_r))N_r \Big) d\zeta \bigg) dW_r.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mathbb{E}|Y_u^\varepsilon|^p \le c\mathbb{E}\bigg(\int_t^u (u - r)^{-2\alpha} \bigg| \int_0^1 \Big(& \nabla_x (e^{(u-r)A}\sigma(r, X_r + \zeta\varepsilon N_r))N_r \\
& - \nabla_x (e^{(u-r)A}\sigma(r, X_r))N_r \Big) d\zeta \bigg|^2_{L_2(\Xi,H)} dr \bigg)^{p/2}
\end{aligned}
$$

for a suitable constant c. Since for all ε

$$
\left| \int_0^1 \nabla_x (e^{(u-r)A}\sigma(r, X_r + \zeta\varepsilon N_r))N_r d\zeta \right|_{L_2(\Xi,H)} \le L(u - r)^{-\gamma}|N|_{C([0,T],H)}
$$

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

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

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

$$
\begin{aligned}
&\nabla_x \Phi(X, t, x; h)_s = e^{(s-t)A}h, \quad s \in [t, T], \\
&\nabla_x \Phi(X, t, x; h)_s = h, \quad s \in [0, t),
\end{aligned}
$$

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

To complete the proof we observe that the Eq. (6.12) is just a re-writing of (6.2) and that the estimate in (iii) is a trivial consequence of Eq. (6.12) and the fact that $|\nabla_X \Phi|$ is uniformly bounded by a constant <1, by the contraction property of Φ. □

6.2.2 Differentiability in the Sense of Malliavin

In order to proceed further in the study of the properties of the solution to the forward equation we need to introduce basic notions and tools of the Malliavin calculus. We refer the reader to the book [468] for a detailed exposition; the paper [328] treats the extensions to Hilbert space-valued random variables and processes. We will report without proofs only the results that will be used in the sequel. This digression on the Malliavin calculus ends after Lemma 6.12, when we come back to the forward equation.

We also inform the reader that the aim of this entire section is just to prove Proposition 6.17, whose statement can be understood after reading a few introductory lines preceding it, and that no reference to the Malliavin calculus will be made in the sections that follow.

Our starting point will be a cylindrical Wiener process $\{W_t\}_{t\geq 0}$ on a real separable Hilbert space Ξ. For every (deterministic) function $h \in L^2([0, T]; \Xi)$ the integral $\int_0^T h(t)^* \, dW_t$ will be denoted by $W(h)$, where $h(t)^* \in \Xi^*$ denotes the image of $h(t) \in \Xi$ under the Riesz isometry. We will also use the notation $W(h) = \int_0^T \langle h(t), dW_t \rangle_\Xi$. Given a Hilbert space K, let S_K be the set of K-valued random variables F of the form

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

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

$$D^{\mathcal{M}}_\eta F = \sum_{j=1}^{m} \sum_{k=1}^{n} \partial_k f_j(W(h_1), \ldots, W(h_n))e_j \otimes h_k(\eta),$$

with values in $\mathcal{L}_2(\Xi, K)$; by ∂_k we denote the partial derivatives with respect to the k-th variable and by $e_j \otimes h_k(\eta)$ the operator $u \rightarrow e_j \, \langle h_k(\eta), u\rangle_\Xi$. It is known that the operator $D^{\mathcal{M}} : S_K \subset L^2(\Omega; K) \to L^2(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$ is closable. We denote by $\mathbb{D}^{1,2}(K)$ the domain of its closure, and use the same letter to denote $D^{\mathcal{M}}$ and its closure:

$$D^{\mathcal{M}} : \mathbb{D}^{1,2}(K) \subset L^2(\Omega; K) \to L^2(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K)).$$

The adjoint operator of $D^{\mathcal{M}}$,

$$\delta : \text{ dom }(\delta) \subset L^2(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K)) \to L^2(\Omega; K),$$

is called the Skorohod integral. Thus, δ acts on a certain subset of square-integrable stochastic processes u_η, $\eta \in [0, T]$, with values in $\mathcal{L}_2(\Xi, K)$ (more precisely, on equivalence classes up to the product measure $\mathbb{P} \otimes d\eta$) and its value at u is a square-integrable random variable with values in K (more precisely, a $\mathbb{P}$-equivalence class), that will be denoted $\delta(u)$ or $\int_0^T u_\eta \, \hat{d} W_\eta$, because of its close connections with the Itô integral (see, for instance, Proposition 6.11 below). We also need to introduce the space $\mathbb{L}^{1,2}(\mathcal{L}_2(\Xi, K))$ of processes $u \in L^2(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$ such that $u_r \in \mathbb{D}^{1,2}(\mathcal{L}_2(\Xi, K))$ for a.e. $r \in [0, T]$, and there exists a measurable version of $D^{\mathcal{M}}_\eta u_r$ satisfying

$$\|u\|^2_{\mathbb{L}^{1,2}(\mathcal{L}_2(\Xi,K))} = \|u\|^2_{L^2(\Omega\times[0,T];\mathcal{L}_2(\Xi,K))} + \mathbb{E}\int_0^T\int_0^T \|D^{\mathcal{M}}_\eta u_r\|^2_{\mathcal{L}_2(\Xi,\mathcal{L}_2(\Xi,K))}\, dr\, d\eta < \infty.$$

The definition of $\mathbb{L}^{1,2}(K)$ for an arbitrary Hilbert space K (instead of $\mathcal{L}_2(\Xi, K)$) is entirely analogous.

In the following proposition we summarize all the properties that we need in the sequel concerning the objects introduced above. We omit the proofs, which can be found in [328] or, after appropriate reformulation, in [468] or [469]. In particular, point 4 is proved in [328], Proposition 3.4. Point 5 can be found in [469], Theorem 3.2, or [328], Proposition 2.11.

Proposition 6.11 *With the previous notation, the following holds.*

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

(2) If u is an (adapted) process belonging to $L^2_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$ then $u \in$ dom(δ) and the Skorohod integral $\delta(u)$ coincides with the Itô integral, i.e.,

$$\int_0^T u_\eta \, \hat{d} W_\eta = \int_0^T u_\eta \, d W_\eta.$$

(3) If $u \in \mathbb{L}^{1,2}(\mathcal{L}_2(\Xi, K))$ then $u \in$ dom(δ) and $\|\delta(u)\|^2_{L^2(\Omega;K)} \le \|u\|^2_{\mathbb{L}^{1,2}(\mathcal{L}_2(\Xi,K))}$. In particular, the Skorohod integral δ is a continuous linear operator from $\mathbb{L}^{1,2}(\mathcal{L}_2(\Xi, K))$ to $L^2(\Omega; K)$.

(4) If $u \in \mathbb{L}^{1,2}(\mathcal{L}_2(\Xi, K))$, and for a.a. η the process $\{D^{\mathcal{M}}_\eta u_r\}_{r\in[0,T]}$ belongs to dom(δ), and the map $\eta \to \delta(D^{\mathcal{M}}_\eta u)$ belongs to $L^2(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$, then $\delta(u) \in \mathbb{D}^{1,2}(K)$ and $D^{\mathcal{M}}_\eta \delta(u) = u_\eta + \delta(D^{\mathcal{M}}_\eta u)$, i.e.,

$$D^{\mathcal{M}}_\eta \int_0^T u_r \, \hat{d} W_r = u_\eta + \int_0^T D^{\mathcal{M}}_\eta u_r \, \hat{d} W_r.$$

(5) If $F \in \mathbb{D}^{1,2}(\mathbb{R})$*,* $u \in L^2(\Omega \times [0,T]; \mathcal{L}_2(\Xi, \mathbb{R})) \simeq L^2(\Omega \times [0,T]; \Xi^*)$ *belongs to* $dom(\delta)$ *and* $Fu \in L^2(\Omega \times [0,T]; \Xi^*)$*, then* $Fu \in dom(\delta)$ *and* $\delta(Fu) = F\delta(u) - \langle D^{\mathcal{M}} F, u \rangle$*, which means*

$$\int_0^T F u_\eta \, \hat{d} W_\eta = F \int_0^T u_\eta \, \hat{d} W_\eta - \int_0^T \left\langle D_\eta^{\mathcal{M}} F, u_\eta \right\rangle_{\mathcal{L}_2(\Xi, K)} d\eta,$$

provided the right-hand side belongs to $L^2(\Omega; \mathbb{R})$.

In particular, if $0 \le a \le b \le T$*,* $\xi \in \Xi$*, and upon taking* $u_\eta = \xi^* 1_{[a,b]}(\eta)$*, we have* $F\xi^* 1_{[a,b]} \in dom(\delta)$ *and*

$$\int_a^b F\,\xi^* \hat{d} W_\eta = F \int_a^b \xi^* \hat{d} W_\eta - \int_a^b D_\eta^{\mathcal{M}} F \xi \, d\eta = F(W_b \xi - W_a \xi) - \int_a^b D_\eta^{\mathcal{M}} F \xi \, d\eta, \tag{6.13}$$

provided $F \in \mathbb{D}^{1,2}(\mathbb{R})$ *and the right-hand side of (6.13) belongs to* $L^2(\Omega; \mathbb{R})$.

Finally, we need to define the space $\mathbb{D}^{1,2}_{loc}(K)$. If $F \in \mathbb{D}^{1,2}(K)$ and $F = 0$ on a measurable subset $A \subset \Omega$ then $1_A D^{\mathcal{M}} F = 0$; this follows immediately from the corresponding result for $K = \mathbb{R}^d$ ([469], Lemma 2.6). Therefore the following definition is meaningful: we say that a random variable $F : \Omega \to K$ belongs to the space $\mathbb{D}^{1,2}_{loc}(K)$ if there exists an increasing sequence of measurable subsets $\Omega_k \subset \Omega$ and elements $F_k \in \mathbb{D}^{1,2}(K)$ such that $\cup_k \Omega_k = \Omega$ $\mathbb{P}$-a.s. and $1_{\Omega_k} F = 1_{\Omega_k} F_k$. $D^{\mathcal{M}} F : \Omega \times [0,T] \to \mathcal{L}_2(\Xi, K)$ is then defined by requiring $1_{\Omega_k} D^{\mathcal{M}} F = 1_{\Omega_k} D^{\mathcal{M}} F_k$. The following chain rule holds; the proof consists in standard approximation arguments and is left to the reader.

Lemma 6.12 *Suppose* K, H *are Hilbert spaces,* $\psi \in \mathcal{G}^1(K, H)$ *and*

$$\sup_{|x| \le n} |\nabla \psi(x)|_{\mathcal{L}(K,H)} < \infty, \qquad n = 1, 2, \ldots. \tag{6.14}$$

(i) If $F \in \mathbb{D}^{1,2}_{loc}(K)$ *then* $\psi(F) \in \mathbb{D}^{1,2}_{loc}(H)$.
(ii) If $F \in \mathbb{D}^{1,2}(K)$ *and* $\sup_{x \in K} |\nabla \psi(x)|_{\mathcal{L}(K,H)} < \infty$ *then* $\psi(F) \in \mathbb{D}^{1,2}(H)$.
(iii) More generally, if $F \in \mathbb{D}^{1,2}(K)$*, (6.14) holds and*

$$\mathbb{E}\, |\psi(F)|_H^2 < \infty, \qquad \mathbb{E} \int_0^T |\nabla \psi(F) D_\eta^{\mathcal{M}} F|^2_{\mathcal{L}_2(K,H)} d\eta < \infty,$$

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

In any of the cases (i)–(iii) we have $D^{\mathcal{M}} \psi(F) = \nabla \psi(F) D^{\mathcal{M}} F$.

After this digression on general Malliavin calculus we come back to the properties of the forward equation and consider again the solution $X = \{X(s; t, x)\}_{s \in [t,T]}$ to (6.6) with (t, x) fixed, denoted simply by (X_s). We set as before $X_s = x$, $s \in [0, t)$.

We will soon prove that X belongs to $\mathbb{L}^{1,2}(H)$. Then it is clear that the equality $D^{\mathcal{M}}_\eta X_s = 0$ $\mathbb{P}$-a.s. holds for a.a. η, t, s if $s < t$ or $\eta > s$.

Proposition 6.13 *Assume Hypothesis 6.8. Then the following properties hold.*

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

(ii) *There exists a version of $D^{\mathcal{M}} X$ such that for every $\eta \in [0, T)$, $\{D^{\mathcal{M}}_\eta X_s\}_{s\in(s,T]}$ is a predictable process in $\mathcal{L}_2(\Xi, H)$ with continuous paths satisfying, for every $p \in [2, \infty)$,*

$$\sup_{\eta\in[0,T]} \mathbb{E}\left(\sup_{s\in(\eta,T]} (s-\eta)^{p\gamma} |D^{\mathcal{M}}_\eta X_s|^p_{\mathcal{L}_2(\Xi,H)} \right) \le c, \tag{6.15}$$

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

$$\begin{aligned} D^{\mathcal{M}}_\eta X_s = e^{(s-\eta)A}\sigma(\eta, X_\eta) + \int_\eta^s e^{(s-r)A} \nabla_x b(r, X_r) D^{\mathcal{M}}_\eta X_r \, dr \\ + \int_\eta^s \nabla_x (e^{(s-r)A}\sigma(r, X_r)) D^{\mathcal{M}}_\eta X_r \, dW_r, \quad s \in (\eta, T]. \end{aligned} \tag{6.16}$$

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

(iii) *Given any element v of Ξ, the process $Q_{\eta s} = D^{\mathcal{M}}_\eta X_s v$ is a solution to the equation:*

$$\begin{aligned} Q_{\eta s} = e^{(s-\eta)A}\sigma(\eta, X_\eta)v + \int_\eta^s e^{(s-r)A} \nabla_x b(r, X_r) Q_{\eta r} \, dr \\ + \int_\eta^s \nabla_x (e^{(s-r)A}\sigma(r, X_r)) Q_{\eta r} \, dW_r, \quad \mathbb{P}\text{-}a.s. \end{aligned} \tag{6.17}$$

for a.a. η, s with $t \le \eta \le s \le T$. It is unique in the sense that if $\{Q_{\eta s}, t \le \eta \le s \le T\}$ is another process with values in H such that $\{Q_{\eta s}\}_{s\in[\eta,T]}$ is predictable for every $\eta \in [t, T]$ and $\mathbb{E} \int_t^T \int_\eta^T |Q_{\eta s}|^2 ds d\eta < \infty$ then, for a.a. η, s, we have $Q_{\eta s} = D^{\mathcal{M}}_\eta X_s v$ $\mathbb{P}$-a.s.

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

Lemma 6.14 *If $X \in \mathbb{L}^{1,2}(H)$ then the random processes*

$$\int_0^s e^{(s-r)A} b(r, X_r) \, dr, \qquad \int_0^s e^{(s-r)A}\sigma(r, X_r) \, dW_r, \qquad s \in [0, T],$$

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

$$D_\eta^{\mathcal{M}} \int_0^s e^{(s-r)A} b(r, X_r)\, dr = \int_\eta^s e^{(s-r)A} \nabla_x b(r, X_r) D_\eta^{\mathcal{M}} X_r\, dr,$$

$$D_\eta^{\mathcal{M}} \int_0^s e^{(s-r)A} \sigma(r, X_r)\, dW_r = e^{(s-\eta)A}\sigma(s, X_s) + \int_\eta^s \nabla_x (e^{(s-r)A}\sigma(r, X_r)) D_\eta^{\mathcal{M}} X_r\, dW_r. \tag{6.18}$$

Proof We will prove only (6.18). Recall that, by Proposition 6.11-4, if $u \in \mathbb{L}^{1,2}(\mathcal{L}_2(\Xi, H))$, and for a.a. η the process $\{D_\eta^{\mathcal{M}} u_r\}_{r\in[0,T]}$ belongs to dom(δ), and the map $\eta \to \delta(D_\eta^{\mathcal{M}} u)$ belongs to $L^2(\Omega \times [0, T]; \mathcal{L}_2(\Xi, H))$, then $\delta(u) \in \mathbb{D}^{1,2}(H)$ and $D_\eta^{\mathcal{M}} \delta(u) = u_\eta + \delta(D_\eta^{\mathcal{M}} u)$.

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

$$\mathbb{E}\int_0^T |u_r|^2\, dr = \mathbb{E}\int_0^s |e^{(s-r)A}\sigma(r, X_r)|^2_{\mathcal{L}_2(\Xi,H)}\, dr \\ \leq L^2 \mathbb{E}\int_0^s (s-r)^{-2\gamma}(1 + |X_r|)^2\, dr.$$

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

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

since $X \in \mathbb{L}^{1,2}(H) \subset L^2(\Omega \times [0, T]; H)$. Next, for every r, by the chain rule for the Malliavin derivative (Lemma 6.12-(ii)), $D_\eta^{\mathcal{M}} u_r = \nabla_x (e^{(s-r)A}\sigma(r, X_r)) D_\eta^{\mathcal{M}} X_r$ for a.a. $\eta < r$, whereas $D_\eta^{\mathcal{M}} u_r = 0$ for a.a. $\eta > r$, by adaptedness. Next, recalling (6.7),

$$\mathbb{E}\int_0^T |D_\eta^{\mathcal{M}} u_r|^2\, dr = \mathbb{E}\int_\eta^s |\nabla_x (e^{(s-r)A}\sigma(r, X_r)) D_\eta^{\mathcal{M}} X_r|^2_{\mathcal{L}_2(\Xi,\mathcal{L}_2(\Xi,H))}\, dr \\ \leq L^2 \mathbb{E}\int_\eta^s (s-r)^{-2\gamma} |D_\eta^{\mathcal{M}} X_r|^2_{\mathcal{L}_2(\Xi,H)}\, dr,$$

so that

$$\mathbb{E}\int_0^T \int_0^T |D_\eta^{\mathcal{M}} u_r|^2\, dr\, d\eta \leq L^2 \mathbb{E}\int_0^s \int_\eta^s (s-r)^{-2\gamma} |D_\eta^{\mathcal{M}} X_r|^2_{\mathcal{L}_2(\Xi,H)}\, dr\, d\eta \\ = L^2 \int_0^s (s-r)^{-2\gamma} \int_0^r \mathbb{E}\, |D_\eta^{\mathcal{M}} X_r|^2_{\mathcal{L}_2(\Xi,H)}\, d\eta\, dr.$$

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

$$\int_0^T \left(\int_0^s (s-r)^{-2\gamma} \int_0^r \mathbb{E}\,|D_\eta^{\mathcal{M}} X_r|^2_{\mathcal{L}_2(\Xi,H)}\, d\eta\, dr\right) ds$$
$$\leq \int_0^T r^{-2\gamma}\, dr \int_0^T \int_0^r \mathbb{E}\,|D_\eta^{\mathcal{M}} X_r|^2_{\mathcal{L}_2(\Xi,H)}\, d\eta\, dr$$
$$= \int_0^T r^{-2\gamma}\, dr |D^{\mathcal{M}} X|^2_{L^2(\Omega\times[0,T]\times[0,T];\mathcal{L}_2(\Xi,H))} < \infty,$$

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

$$\int_0^T \mathbb{E}|\delta(D_\eta^{\mathcal{M}} u)|^2\, d\eta = \int_0^T \mathbb{E}\left|\int_0^T D_\eta^{\mathcal{M}} u_r\, dW_r\right|^2 d\eta = \mathbb{E}\int_0^T \int_0^T |D_\eta^{\mathcal{M}} u_r|^2\, dr\, d\eta < \infty.$$

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

$$\delta(u) = \int_0^s e^{(s-r)A}\sigma(r, X_r)\, dW_r, \qquad \delta(D_\eta^{\mathcal{M}} u) = \int_\eta^s \nabla_x (e^{(s-r)A}\sigma(r, X_r)) D_\eta^{\mathcal{M}} X_r\, dW_r,$$

formula (6.18) is proved. The estimate

$$\int_0^T \int_0^s \mathbb{E}\left| D_\eta^{\mathcal{M}} \int_0^s e^{(s-r)A}\sigma(r, X_r)\, dW_r \right|^2 d\eta\, ds$$
$$\leq 2\int_0^T \int_0^s \mathbb{E}|e^{(s-\eta)A}\sigma(\eta, X_\eta)|^2_{\mathcal{L}_2(\Xi,H)}\, d\eta\, ds$$
$$+2\int_0^T \int_0^s \mathbb{E}\int_\eta^s |\nabla_x (e^{(s-r)A}\sigma(r, X_r)) D_\eta^{\mathcal{M}} X_r|^2_{\mathcal{L}_2(\Xi,\mathcal{L}_2(\Xi,H))}\, dr\, d\eta\, ds$$
$$\leq 2L^2 \int_0^T r^{-2\gamma}\, dr \int_0^T \mathbb{E}\,(1+|X_r|)^2\, dr$$
$$+2L^2 \int_0^T r^{-2\gamma}\, dr |D^{\mathcal{M}} X|^2_{L^2(\Omega\times[0,T]\times[0,T];\mathcal{L}_2(\Xi,H))} < \infty,$$

is a consequence of the previous passages, and shows that the process $\int_0^s e^{(s-r)A}$ $\sigma(r, X_r)\, dW_r$, $s \in [0, T]$, belongs to $\mathbb{L}^{1,2}(H)$. □

For $\eta \in [0, T)$ and for arbitrary predictable processes X_s, Q_s, $s \in [\eta, T]$, with values in H and $\mathcal{L}_2(\Xi, H)$ respectively, we define, for $s \in [\eta, T]$,

$$\Gamma_1(X, Q)_{\eta s} = \int_\eta^s e^{(s-r)A}\nabla_x b(r, X_r) Q_r\, dr,$$

$$\Gamma_2(X, Q)_{\eta s} = \int_\eta^s \nabla_x (e^{(s-r)A}\sigma(r, X_r)) Q_r\, dW_r.$$

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

Proof of Proposition 6.13. We fix $t \in [0, T)$. Let us consider the sequence X^n defined as follows: $X^0 = 0$,

$$X_s^{n+1} = e^{(s-t)A}x + \int_t^s e^{(s-r)A} b(r, X_r^n)\, dr + \int_t^s e^{(s-r)A} \sigma(r, X_r^n)\, dW_r, \quad s \in [t, T],$$

and $X_s^n = x$ for $s < t$. It follows from the proof of Proposition 6.9 that X^n converges to the solution X of Eq. (6.6) in the space $L^p_{\mathcal{P}}(\Omega; C([0, T], H))$ hence, in particular, in the space $L^2(\Omega \times [0, T]; H)$. By Lemma 6.14, $X^n \in \mathbb{L}^{1,2}(H)$ and, for a.a. η and s with $\eta < s$,

$$\begin{aligned} D_\eta^{\mathcal{M}} X_s^{n+1} &= e^{(s-\eta)A}\sigma(\eta, X_\eta^n) + \int_\eta^s e^{(s-r)A} \nabla_x b(r, X_r^n) D_\eta^{\mathcal{M}} X_r^n\, dr \\ &\quad + \int_\eta^s \nabla_x (e^{(s-r)A}\sigma(r, X_r^n)) D_\eta^{\mathcal{M}} X_r^n\, dW_r. \end{aligned} \tag{6.19}$$

Setting $I(X^n)_{\eta s} = e^{(s-\eta)A}\sigma(\eta, X_\eta^n)$ for $s > \eta$ and $I(X^n)_{\eta s} = 0$ for $s < \eta$, and recalling the operators introduced above, we may write equality (6.19) as

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

We note that $I(X^n)$ is a bounded sequence in $L^2(\Omega \times [0, T] \times [0, T]; \mathcal{L}_2(\Xi, H))$, since

$$\begin{aligned} \mathbb{E} \int_0^T \int_0^s &|e^{(s-\eta)A}\sigma(\eta, X_\eta^n)|^2_{\mathcal{L}_2(\Xi, H)}\, d\eta\, ds \\ &\le L^2 \mathbb{E} \int_0^T \int_0^s (s-\eta)^{-2\gamma}(1 + |X_\eta^n|)^2\, d\eta\, ds \\ &\qquad \le L^2 \int_0^T s^{-2\gamma}\, ds \int_0^T \mathbb{E}\,(1 + |X_\eta^n|)^2\, d\eta, \end{aligned}$$

and X^n is a bounded sequence in $L^2(\Omega \times [0, T]; H)$. Next we show that there exists an equivalent norm $\|\cdot\|$ in $L^2(\Omega \times [0, T] \times [0, T]; \mathcal{L}_2(\Xi, H))$ such that

$$\|\Gamma_1(X^n, D^{\mathcal{M}} X^n)\| + \|\Gamma_2(X^n, D^{\mathcal{M}} X^n)\| \le \alpha \|D^{\mathcal{M}} X^n\|, \tag{6.20}$$

for some $\alpha \in [0, 1)$ independent of n. For simplicity we only consider the operator Γ_2. For a process $(Z_{\eta s}) \in L^2(\Omega \times [0, T] \times [0, T]; \mathcal{L}_2(\Xi, H))$ we introduce the norm

$$\|Z\|^2 = \int_0^T \int_0^T \mathbb{E}\, |Z_{\eta s}|^2_{\mathcal{L}_2(\Xi, H)} e^{-\beta(s-\eta)} ds\, d\eta,$$

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

$$
\begin{aligned}
&\int_\eta^T \mathbb{E}|\Gamma_2(X^n, D^{\mathcal{M}} X^n)_{\eta s}|^2_{\mathcal{L}_2(\Xi,H)} e^{-\beta(s-\eta)}\, ds \\
&= \int_\eta^T \int_\eta^s \mathbb{E}\,|\nabla_x (e^{(s-r)A}\sigma(r, X^n_r)) D^{\mathcal{M}}_\eta X^n_r|^2_{\mathcal{L}_2(\Xi,\mathcal{L}_2(\Xi,H))}\, dr\; e^{-\beta(s-\eta)}\, ds \\
&\le L^2 \int_\eta^T \int_\eta^s (s-r)^{-2\gamma} \mathbb{E}\,|D^{\mathcal{M}}_\eta X^n_r|^2_{\mathcal{L}_2(\Xi,H)}\, dr\; e^{-\beta(s-\eta)}\, ds \\
&= L^2 \int_\eta^T e^{-\beta(r-\eta)} \mathbb{E}\,|D^{\mathcal{M}}_\eta X^n_r|^2_{\mathcal{L}_2(\Xi,H)} \int_r^T (s-r)^{-2\gamma}\, e^{-\beta(s-r)}\, ds\, dr \\
&\le L^2 \int_\eta^T e^{-\beta(r-\eta)} \mathbb{E}\,|D^{\mathcal{M}}_\eta X^n_r|^2_{\mathcal{L}_2(\Xi,H)}\, dr \left(\sup_{r\in[\eta,T]} \int_r^T (s-r)^{-2\gamma}\, e^{-\beta(s-r)}\, ds \right).
\end{aligned}
$$

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

$$
\|\Gamma_2(X^n, D^{\mathcal{M}} X^n)\|^2 \le L^2 \int_0^T r^{-2\gamma}\, e^{-\beta r}\, dr \|D^{\mathcal{M}} X^n\|^2.
$$

Now to prove (6.20) it suffices to take β sufficiently large.

From (6.20) and from the fact that $I(X^n)$ is bounded in $L^2(\Omega \times [0,T] \times [0,T]; \mathcal{L}_2(\Xi, H))$, it follows easily that the sequence $D^{\mathcal{M}} X^n$ is also bounded in this space. Since, as mentioned before, X^n converges to X in $L^2(\Omega \times [0,T]; H)$, it follows from the closedness of the operator $D^{\mathcal{M}}$ that X belongs to $\mathbb{L}^{1,2}(H)$. Point (i) of Proposition 6.13 is now proved.

By Lemma 6.14, we can compute the Malliavin derivative of both sides of (6.6) and we obtain, for a.a. η and s with $\eta < s$,

$$
D^{\mathcal{M}}_\eta X_s = I(X)_{\eta s} + \Gamma_1(X, D^{\mathcal{M}} X)_{\eta s} + \Gamma_2(X, D^{\mathcal{M}} X)_{\eta s}, \qquad \mathbb{P}\text{-a.s.}, \tag{6.21}
$$

where

$$
I(X)_{\eta s} = e^{(s-\eta)A}\sigma(\eta, X_\eta). \tag{6.22}
$$

Let us introduce the space $\mathcal{K}$ of processes $Q_{\eta s}$, $0 \le \eta < s \le T$, such that for every $\eta \in [t, T)$, $\{Q_{\eta s}\}_{s\in(\eta,T]}$ is a predictable process in $\mathcal{L}_2(\Xi, H)$ with continuous paths, and such that

$$
\sup_{\eta\in[0,T]} \mathbb{E}\left(\sup_{s\in(\eta,T]} e^{-\beta p(s-\eta)} (s-\eta)^{p\gamma} |Q_{\eta s}|^p_{\mathcal{L}_2(\Xi,H)} \right) < \infty. \tag{6.23}
$$

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

$$
Q_{\eta s} = I(X)_{\eta s} + \Gamma_1(X, Q)_{\eta s} + \Gamma_2(X, Q)_{\eta s}, \qquad s \in (\eta, T]. \tag{6.24}
$$

We are going to prove that there exists a unique solution $Q \in \mathcal{K}$ of this equation. Assume this for a moment. Then, subtracting (6.24) from (6.21), we obtain for a.a. η and s with $\eta < s$

$$D_\eta^{\mathcal{M}} X_s - Q_{\eta s} = \Gamma_1(X, D^{\mathcal{M}}X - Q)_{\eta s} + \Gamma_2(X, D^{\mathcal{M}}X - Q)_{\eta s}, \qquad \mathbb{P}\text{-a.s.}$$

Repeating the passages that led to (6.20) we obtain

$$\|\Gamma_1(X, D^{\mathcal{M}}X - Q)\| + \|\Gamma_2(X, D^{\mathcal{M}}X - Q)\| \le \alpha \|D^{\mathcal{M}}X - Q\|,$$

for some $\alpha \in [0, 1)$. This proves that Q is a version of $D^{\mathcal{M}}X$. Then equality (6.24) coincides with (6.16), and this proves point (ii) of the Proposition, except for the last assertion.

Now we prove unique solvability of (6.24) in the space $\mathcal{K}$. It suffices to show that $I(X) \in \mathcal{K}$ and that $\Gamma_1(X, \cdot) + \Gamma_2(X, \cdot)$ is a contraction in $\mathcal{K}$. Since, for $s > \eta$,

$$|e^{(s-\eta)A}\sigma(\eta, X_\eta)|_{\mathcal{L}_2(\Xi,H)} \le L(s-\eta)^{-\gamma}(1 + |X_\eta|),$$

we have

$$\sup_{\eta\in[0,T]} \mathbb{E} \sup_{s\in(\eta,T]} (s-\eta)^{p\gamma}|e^{(s-\eta)A}\sigma(\eta, X_\eta)|^p_{\mathcal{L}_2(\Xi,H)} \le L^p \sup_{\eta\in[0,T]} \mathbb{E}\,(1 + |X_\eta|)^p,$$

which is finite, since $X \in L^p_{\mathcal{P}}(\Omega; C([0, T], H))$. This shows that $I(X) \in \mathcal{K}$; the contraction property for $\Gamma_1(X, \cdot) + \Gamma_2(X, \cdot)$ requires a longer argument, and it is postponed to Lemma 6.15 below.

The last assertion of point (ii) is clear for $s \in [0, t]$, since $X_s = x$. For $s \in (t, T]$ we take a sequence $s_n \uparrow s$ such that $X_{s_n} \in \mathbb{D}^{1,2}(H)$ and we note that by (6.15) the sequence $\mathbb{E} \int_0^T |D_\eta^{\mathcal{M}} X_{s_n}|^2 d\eta$ is bounded by a constant independent of n; since $X_{s_n} \to X_s$ in $L^2(\Omega; H)$, it follows from the closedness of the operator $D^{\mathcal{M}}$ that $X_s \in \mathbb{D}^{1,2}(H)$.

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

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

is finite, where $\beta > 0$ is a parameter to be chosen later. Since $I(X)$ (defined in (6.22)) belongs to the space $\mathcal{K}$ introduced above, $I(X)v$ belongs to $\mathcal{S}$ and the equality (6.17) is equivalent to the equality in the space $\mathcal{S}$:

$$Q = I(X)v + \Gamma_1(X, Q) + \Gamma_2(X, Q). \tag{6.25}$$

It turns out that this equation has a unique solution in $\mathcal{S}$: indeed, $\Gamma_1(X, \cdot) + \Gamma_2(X, \cdot)$ is a contraction in the space $\mathcal{S}$ if β is chosen sufficiently large, as it can be proved by passages almost identical to those leading to (6.20). Finally, $D^{\mathcal{M}} X v$ belongs to $\mathcal{S}$ since $D^{\mathcal{M}} X \in L^2(\Omega \times [0, T] \times [0, T]; \mathcal{L}_2(\Xi, H))$, and applying both sides of (6.16) to v we check that $D^{\mathcal{M}} X v = I(X)v + \Gamma_1(X, D^{\mathcal{M}} X v) + \Gamma_2(X, D^{\mathcal{M}} X v)$. Point (iii) of the proposition is now proved. □

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

Lemma 6.15 *For $\eta \in [0, T)$, let X_s, $s \in [\eta, T]$, be a predictable process in H and let Q_s, $s \in (\eta, T]$, be an $\mathcal{L}_2(\Xi, H)$-valued continuous adapted process.*

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

$$\mathbb{E}\left(\sup_{s\in[\eta,T]} (s-\eta)^{\gamma p} e^{-\beta p(s-\eta)} \left(|\Gamma_1(X, Q)_{\eta s}|^p_{\mathcal{L}_2(\Xi,H)} + |\Gamma_2(X, Q)_{\eta s}|^p_{\mathcal{L}_2(\Xi,H)}\right)\right)$$
$$\le C(\beta)\mathbb{E}\left(\sup_{s\in[\eta,T]} (s-\eta)^{\gamma p} e^{-\beta p(s-\eta)} |Q_s|^p_{\mathcal{L}_2(\Xi,H)}\right),$$

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

Proof For simplicity, we only consider the operator Γ_2. Fixing $\eta \in [0, T)$ we introduce the space of $\mathcal{L}_2(\Xi, H)$-valued continuous adapted processes Q_s, $s \in (\eta, T]$ such that the norm

$$\|Q\|^p_\eta := \mathbb{E} \sup_{s\in[\eta,T]} (s-\eta)^{\gamma p} e^{-\beta p(s-\eta)} |Q_s|^p_{\mathcal{L}_2(\Xi,H)}$$

is finite. We use the factorization method, see [177], Theorem 5.2.5. Let us take $p > 2$ and $\alpha \in (0, 1)$ such that

$$\frac{1}{p} < \alpha < \frac{1}{2} - \gamma, \qquad \text{and let} \qquad c_\alpha^{-1} = \int_r^s (s-u)^{\alpha-1}(u-r)^{-\alpha} du.$$

Then, by the stochastic Fubini theorem,

$$\begin{aligned}\Gamma_2(X, Q)_{\eta s} &= c_\alpha \int_\eta^s \int_r^s (s-u)^{\alpha-1}(u-r)^{-\alpha}\, e^{(s-u)A} \nabla_x (e^{(u-r)A}\sigma(r, X_r)) Q_r \, du \, dW_r \\ &= c_\alpha \int_\eta^s (s-u)^{\alpha-1} e^{(s-u)A} V_u \, du,\end{aligned}$$

where

$$V_u = \int_\eta^u (u-r)^{-\alpha} \nabla_x (e^{(u-r)A}\sigma(r, X_r)) Q_r \, dW_r.$$

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

$$\left|\Gamma_2(X,Q)_{\eta s}\right| \leq c_\alpha M \int_\eta^s (s-u)^{\alpha-1}|V_u|\,du$$
$$\leq c_\alpha M \left(\int_\eta^s e^{-p\beta(u-\eta)}(u-\eta)^{\gamma p}|V_u|^p\,du\right)^{\frac{1}{p}}$$
$$\cdot \left(\int_\eta^s e^{p'\beta(u-\eta)}(u-\eta)^{-\gamma p'}(s-u)^{(\alpha-1)p'}\,du\right)^{\frac{1}{p'}}.$$

$$\|\Gamma_2(X,Q)\|_\eta^p \leq c_\alpha^p M^p \int_\eta^T e^{-p\beta(u-\eta)}(u-\eta)^{\gamma p}\mathbb{E}\,|V_u|^p\,du$$
$$\cdot \sup_{s\in(\eta,T]} (s-\eta)^{\gamma p} e^{-\beta p(s-\eta)} \left(\int_\eta^s e^{p'\beta(u-\eta)}(u-\eta)^{-\gamma p'}(s-u)^{(\alpha-1)p'}\,du\right)^{\frac{p}{p'}}.$$

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

$$\sup_{s\in(\eta,T]} (s-\eta)^{p\alpha-1} e^{-\beta p(s-\eta)} \left(\int_0^1 e^{p'\beta u(s-\eta)}u^{-\gamma p'}(1-u)^{(\alpha-1)p'}\,du\right)^{\frac{p}{p'}} \leq a(\beta)^p,$$

where we set

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

So we arrive at

$$\|\Gamma_2(X,Q)\|_\eta \leq c_\alpha M a(\beta) \left(\int_\eta^T e^{-p\beta(u-\eta)}(u-\eta)^{\gamma p}\mathbb{E}\,|V_u|^p\,du\right)^{\frac{1}{p}}.$$

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

$$\mathbb{E}\,|V_u|^p \leq c_p \mathbb{E}\left(\int_\eta^u (u-r)^{-2\alpha}|\nabla_x(e^{(u-r)A}\sigma(r,X_r))Q_r|^2_{\mathcal{L}_2(\Xi,\mathcal{L}_2(\Xi,H))}\,dr\right)^{\frac{p}{2}}$$
$$\leq L^p c_p \mathbb{E}\left(\int_\eta^u (u-r)^{-2\alpha-2\gamma}|Q_r|^2_{\mathcal{L}_2(\Xi,H)}\,dr\right)^{\frac{p}{2}}$$
$$\leq L^p c_p \|Q\|_s^p \left(\int_\eta^u (u-r)^{-2\alpha-2\gamma}(r-\eta)^{-2\gamma}e^{2\beta(r-\eta)}\,dr\right)^{\frac{p}{2}}.$$

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

$$
\begin{aligned}
(u-\eta)^{\gamma p} e^{-p\beta(u-\eta)}\mathbb{E}\,|V_u|^p &\le L^p c_p \|Q\|_\eta^p (u-\eta)^{p(-\alpha-\gamma+1/2)} \\
&\quad \cdot \left(\int_0^1 (1-r)^{-2\alpha-2\gamma} r^{-2\gamma} e^{-2\beta(1-r)(r-\eta)}\, dr \right)^{\frac{p}{2}} \\
&\le L^p c_p \|Q\|_\eta^p T^{p(\frac{1}{2}-\alpha-\gamma)} \left(\int_0^1 (1-r)^{-2\alpha-2\gamma} r^{-2\gamma}\, dr \right)^{\frac{p}{2}}.
\end{aligned}
$$

We conclude that

$$
\|\Gamma_2(X, Q)\|_\eta \le c_\alpha M L c_p^{\frac{1}{p}} a(\beta) T^{\frac{1}{2}-\alpha-\gamma+\frac{1}{p}} \left(\int_0^1 (1-r)^{-2\alpha-2\gamma} r^{-2\gamma}\, dr \right)^{\frac{1}{2}} \|Q\|_\eta.
$$

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

The following result relates the Malliavin derivative of the process X with $\nabla_x X(s; t, x)$, the partial Gâteaux derivative with respect to x (compare Proposition 6.10).

Proposition 6.16 *Assume Hypothesis 6.8. Then for a.a. η, s such that $t \le \eta \le s \le T$ we have*

$$
D_\eta^{\mathcal{M}} X(s; t, x) = \nabla_x X(s; \eta, X(\eta; t, x))\sigma(\eta, X(\eta; t, x)), \qquad \mathbb{P}\text{-}a.s. \tag{6.26}
$$

Moreover, $D_\eta^{\mathcal{M}} X(T; t, x) = \nabla_x X(T; \eta, X(\eta; t, x))\sigma(\eta, X(\eta; t, x))$, $\mathbb{P}$-a.s. for a.a. η.

Proof Proposition 6.10 states that for every $\eta \in [0, T]$ and every direction $h \in H$ the directional derivative process $\nabla_x X(s; \eta, x)h$, $s \in [\eta, T]$, solves the equation: $\mathbb{P}$-a.s.,

$$
\begin{aligned}
\nabla_x X(s; \eta, x)h = e^{(s-\eta)A} h &+ \int_\eta^s e^{(s-r)A} \nabla_x b(r, X(r; \eta, x)) \nabla_x X(r; \eta, x) h\, dr \\
&+ \int_\eta^s \nabla_x (e^{(s-r)A} \sigma(r, X(r; \eta, x)) \nabla_x X(r; \eta, x) h\, dW_r, \quad s \in [\eta, T].
\end{aligned}
$$

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

$$
X(r; \eta, X(\eta; t, x)) = X(r; t, x), \qquad r \in [\eta, T],
$$

which is a consequence of the uniqueness of the solution to (6.6), and we obtain: $\mathbb{P}$-a.s.,

$$\nabla_x X(s;\eta,X(\eta;t,x))\sigma(\eta,X(\eta;t,x))v = e^{(s-\eta)A}\sigma(\eta,X(\eta;t,x))v$$
$$+\int_\eta^s e^{(s-r)A}\nabla_x b(r,X(r;t,x))\nabla_x X(r;\eta,X(\eta;t,x))\sigma(\eta,X(\eta;t,x))v\,dr$$
$$+\int_\eta^s \nabla_x(e^{(s-r)A}\sigma(r,X(r;t,x))\nabla_x X(r;\eta,X(\eta;t,x))\sigma(\eta,X(\eta;t,x))v\,dW_r,\ \ s\in[\eta,T].$$

This shows that the process $\{\nabla_x X(s;t,X(\eta;t,x))\sigma(\eta,X(\eta;t,x))v : t\le\eta\le s\le T\}$ is a solution to Eq. (6.17). Then (6.26) follows from the uniqueness property.

To prove the last assertion, it suffices to take a sequence $s_n \uparrow T$ such that (6.26) holds for s_n and let $n\to\infty$. The conclusion follows from the regularity properties of $D^{\mathcal{M}}X$ and $\nabla_x X$ stated above, as well as the closedness of the operator $D^{\mathcal{M}}$. □

Now, for $\xi\in\Xi$, recall that $W\xi = \{W(\tau)\xi\}_{\tau\ge 0}$ is a real Wiener process. Also fix $t\in[0,T]$ and $x\in H$ and set $X_\tau = X(\tau;t,x)$, $\tau\in[t,T]$, for simplicity. Given a function $u:[0,T]\times H\to\mathbb{R}$, we investigate the existence of the joint quadratic variation of the process $\{u(\tau,X_\tau)\}_{\tau\in[t,T]}$ with $W\xi$. As usual, this is defined for every $\tau\in[t,T]$ as the limit in probability of

$$\sum_{i=1}^n (u(\tau_i,X_{\tau_i})-u(\tau_{i-1},X_{\tau_{i-1}}))(W(\tau_i)\xi - W(\tau_{i-1})\xi),$$

where $\{\tau_i\}$, $t=\tau_0<\tau_1<\cdots<\tau_n=\tau$, is an arbitrary subdivision of $[t,\tau]$ whose mesh tends to 0. The existence of the joint quadratic variation is not trivial. Indeed, due to the occurrence of convolution type integrals in the definition of a mild solution, it is not obvious that the process X is a semimartingale. Moreover, even in this case, the process $u(\cdot,X_\cdot)$ might fail to be a semimartingale if u is not regular enough. Nevertheless, the following result holds true. Its proof could be deduced from the generalization of some results obtained in [469] to the infinite-dimensional case, but we prefer to give a simpler direct proof.

Proposition 6.17 *Assume Hypothesis 6.8, let u be a function in $\mathcal{G}^{0,1}([0,T]\times H,\mathbb{R})$ having polynomial growth together with its derivative $\nabla_x u$. Then the process $\{u(\tau,X_\tau)\}_{\tau\in[t,T]}$ admits a joint quadratic variation process V with $W\xi$, given by*

$$V_\tau = \int_t^\tau \nabla_x u(s,X_s)\,\sigma(s,X_s)\xi\,ds,\qquad \tau\in[t,T].$$

Proof Let us write $\bar u_\tau = u(\tau,X_\tau)$, $\tau\in[t,T]$, for simplicity. By Proposition 6.13 and the assumptions on u we can apply the chain rule for the Malliavin derivative operator presented in Lemma 6.12 and conclude that, for every $\tau\in[t,T]$, we have $\bar u_\tau\in\mathbb{D}^{1,2}(\mathbb{R})$ and $D^{\mathcal{M}}\bar u_\tau = \nabla_x u(\tau,X_\tau)D^{\mathcal{M}}X_\tau$. Taking into account (6.26), for a.e. $s\in[0,\tau]$ we obtain

$$D_s^{\mathcal{M}}\bar u_\tau\xi = \nabla_x u(\tau,X_\tau)\ \nabla_x X(\tau;s,X_s)\ \sigma(s,X_s)\ \xi,\qquad \mathbb{P}\text{-a.s.} \tag{6.27}$$

whereas $D_s^{\mathcal{M}} \bar{u}_\tau \xi = 0$ $\mathbb{P}$-a.s., for a.e. $s \in (\tau, T]$.

Let us now compute the joint quadratic variation of $\bar{u}$ and $W\xi$. Let $t = \tau_0 < \tau_1 < \cdots < \tau_n = \tau$ be a subdivision of $[t, \tau] \subset [0, T]$. We use formula (6.13) in Proposition 6.11 with $[a, b] = [\tau_{i-1}, \tau_i]$ and $F = \bar{u}_{\tau_i} - \bar{u}_{\tau_{i-1}}$ and obtain

$$(\bar{u}_{\tau_i} - \bar{u}_{\tau_{i-1}})(W(\tau_i)\xi - W(\tau_{i-1})\xi) = \int_{\tau_{i-1}}^{\tau_i} (\bar{u}_{\tau_i} - \bar{u}_{\tau_{i-1}})\xi^* \,\hat{d}W_s + \int_{\tau_{i-1}}^{\tau_i} D_s^{\mathcal{M}} (\bar{u}_{\tau_i} - \bar{u}_{\tau_{i-1}})\xi \, ds,$$

where as usual we use the symbol $\hat{d}W$ to denote the Skorohod integral. We note that $D_s^{\mathcal{M}} \bar{u}_{\tau_{i-1}} = 0$ for $s > \tau_{i-1}$, so recalling (6.27) and setting $U_n(s) = \sum_{i=1}^n (\bar{u}_{\tau_i} - \bar{u}_{\tau_{i-1}})\, 1_{(\tau_{i-1}, \tau_i]}(s)$ we obtain

$$\sum_{i=1}^n (\bar{u}_{\tau_i} - \bar{u}_{\tau_{i-1}})(W^\xi_{\tau_i} - W^\xi_{\tau_{i-1}}) = \int_t^\tau U_n(s)\,\xi^* \,\hat{d}W_s + \sum_{i=1}^n \int_{\tau_{i-1}}^{\tau_i} \nabla_x u(\tau_i, X_{\tau_i})\, \nabla_x X(\tau_i; s, X_s) \sigma(s, X_s)\xi \, ds.$$

By (6.27) and the continuity properties asserted in Proposition 6.10, it is easily verified that the maps $\tau \to \bar{u}_\tau$ and $\tau \to D^{\mathcal{M}} \bar{u}_\tau \xi$ are continuous on $[0, T]$ with values in $L^2(\Omega; \mathbb{R})$ and $L^2(\Omega \times [0, T]; \mathbb{R})$, respectively. In particular, $U_n \to 0$ in $\mathbb{L}^{1,2}(\mathbb{R})$, which implies that the Skorohod integral in the last equation tends to zero in $L^2(\Omega; \mathbb{R})$. Letting the mesh of the subdivision tend to 0 and using the continuity properties of $\nabla_x u$, X, $\nabla_x X$, we obtain

$$\sum_{i=1}^n (\bar{u}_{\tau_i} - \bar{u}_{\tau_{i-1}})(W(\tau_i)\xi - W(\tau_{i-1})\xi) \to V_\tau,$$

in probability, which finishes the proof of the proposition. □

6.3 Backward Stochastic Differential Equations (BSDEs)

6.3.1 Well-Posedness

Some of the basic results on backward equations rely on the following well-known representation theorem (see e.g. [350]). Recall that $(\mathscr{F}_t)$ is the filtration generated by the cylindrical Wiener process W, augmented in the usual way. We denote by $\mathbb{E}^{\mathscr{F}_s}$ the conditional expectation with respect to $\mathscr{F}_s$.

Proposition 6.18 *Let K be a Hilbert space and* $T > 0$. *For arbitrary* $\mathscr{F}_T$*-measurable* $\xi \in L^2(\Omega; K)$ *there exists a* $V \in L^2_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$ *such that* $\xi = \mathbb{E}\,\xi + \int_0^T V(r)\, dW(r)$, $\mathbb{P}$*-a.s. Equivalently, for every* $s \in [0, T]$,

$$\mathbb{E}^{\mathscr{F}_s}\xi = \xi - \int_s^T V(r)\, dW(r), \qquad \mathbb{P}\text{-}a.s.$$

Lemma 6.19 *Assume* $\eta \in L^2(\Omega; K)$ *is* $\mathscr{F}_T$*-measurable and* $f \in L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$. *Then there exists a unique pair of processes* $Y(s)$, $Z(s)$, $s \in [0, T]$, *such that*

(i) $Y \in L^2_{\mathcal{P}}(\Omega \times [0, T]; K)$, $Z \in L^2_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$;
(ii) *for a.a.* $s \in [0, T]$, $\mathbb{P}$*-a.s.,*

$$Y(s) + \int_s^T Z(r)\, dW(r) = \int_s^T f(r)\, dr + \eta. \tag{6.28}$$

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

$$\begin{aligned} \mathbb{E} \int_0^T e^{2\beta r} |Z(r)|^2 dr &\le \frac{4}{\beta}\, \mathbb{E} \int_0^T e^{2\beta r} |f(r)|^2 dr + 8\, e^{2\beta T}\, \mathbb{E}\, |\eta|^2, \\ \mathbb{E} \sup_{s\in[0,T]} e^{2\beta s} |Y(s)|^2 &\le \frac{4}{\beta}\, \mathbb{E} \int_0^T e^{2\beta r} |f(r)|^2 dr + 8\, e^{2\beta T}\, \mathbb{E}\, |\eta|^2. \end{aligned} \tag{6.29}$$

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

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

$$\mathbb{E} \left(\int_0^T |f(r)|^2 dr \right)^{p/2} < \infty, \qquad \mathbb{E}\, |\eta|^p < \infty,$$

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

$$\mathbb{E} \sup_{s\in[T-\delta,T]} |Y(s)|^p + \mathbb{E} \left(\int_{T-\delta}^T |Z(r)|^2 dr \right)^{p/2} \le c_p \delta^{p/2} \mathbb{E} \left(\int_{T-\delta}^T |f(r)|^2 dr \right)^{p/2} + c_p \mathbb{E}\, |\eta|^p, \tag{6.30}$$

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

Proof We modify the argument in [350]. We write Y_s instead of $Y(s)$ etc. to shorten notation.

Uniqueness. Assume that (6.28) holds. Then, taking conditional expectation with respect to $\mathscr{F}_s$ we obtain, for a.e. s,

$$Y_s = \mathbb{E}^{\mathscr{F}_s}\eta + \int_s^T \mathbb{E}^{\mathscr{F}_s} f_r\, dr. \tag{6.31}$$

If $\eta = 0$ and $f = 0$ this equality implies that $Y = 0$; from (6.28) it follows that $\int_s^T Z_r \, dW_r = 0$, which implies $Z = 0$ as well.

Existence. Define $\xi = \eta + \int_0^T f_r \, dr$. Since $\xi \in L^2(\Omega; K)$ is $\mathscr{F}_T$-measurable, by Proposition 6.18 there exists a $Z \in L^2_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$ such that

$$\mathbb{E}^{\mathscr{F}_s} \xi = \xi - \int_s^T Z_r \, dW_r,$$

for every $s \in [0, T]$. Now it suffices to define $Y_s = \mathbb{E}^{\mathscr{F}_s} \xi - \int_0^s f_r \, dr$ and Eq. (6.28) is satisfied. The existence of a continuous version is immediate, since (6.28) implies

$$Y_s - Y_0 = \int_0^s Z_r \, dW_r - \int_0^s f_r \, dr.$$

Estimates (6.29). Since $\eta \in L^2(\Omega; K)$ is $\mathscr{F}_T$-measurable, by Proposition 6.18 there exists an $L \in L^2_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$ such that

$$\mathbb{E}^{\mathscr{F}_s} \eta = \eta - \int_s^T L_\theta \, dW_\theta, \tag{6.32}$$

for every $s \in [0, T]$. Similarly, for a.a. r there exists a predictable process $\{K(\theta, r)\}_{\theta \in [0,r]}$ in $L^2_{\mathcal{P}}(\Omega \times [0, r]; \mathcal{L}_2(\Xi, K))$ such that

$$\mathbb{E}^{\mathscr{F}_s} f_r = f_r - \int_s^r K(\theta, r) \, dW_\theta, \tag{6.33}$$

for $s \in [0, r]$. We set $K(\theta, r) = 0$ for $\theta \in (r, T]$ and we can verify that the map $K : \Omega \times [0, T] \times [0, T] \to \mathcal{L}_2(\Xi, K)$ can be taken to be $\mathcal{P} \times \mathcal{B}([0, T])$-measurable, where $\mathcal{P}$ is the predictable σ-field on $\Omega \times [0, T]$ and $\mathcal{B}([0, T])$ denotes the Borel subsets of $[0, T]$; the existence of such a version of K can be proved by approximating f by simple processes and by a monotone class argument (or one can argue as in [350], proof of Lemma 2.1). Substituting into (6.31) and applying the stochastic Fubini theorem gives

$$\begin{aligned} Y_s &= \eta - \int_s^T L_\theta \, dW_\theta + \int_s^T \left(f_r - \int_s^r K(\theta, r) \, dW_\theta \right) dr \\ &= \eta + \int_s^T f_r \, dr - \int_s^T L_\theta \, dW_\theta - \int_s^T \left(\int_\theta^T K(\theta, r) \, dr \right) dW_\theta. \end{aligned}$$

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

$$Z_\theta = L_\theta + \int_\theta^T K(\theta, r) \, dr.$$

Now let $\beta \neq 0$.

From (6.32) we deduce that

$$\mathbb{E}\int_0^T e^{2\beta\theta}|L_\theta|^2\,d\theta \le e^{2\beta T}\mathbb{E}\left|\int_0^T L_\theta\,dW_\theta\right|^2 = e^{2\beta T}\mathbb{E}\left|\eta - \mathbb{E}^{\mathscr{F}_0}\eta\right|^2$$
$$\le 2e^{2\beta T}\mathbb{E}\,|\eta|^2 + 2e^{2\beta T}\mathbb{E}\,|\mathbb{E}^{\mathscr{F}_0}\eta|^2 \le 4e^{2\beta T}\mathbb{E}\,|\eta|^2.$$

Next note that

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

so that

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

Since (6.33) yields

$$\mathbb{E}\int_0^r |K(\theta,r)|^2\,d\theta = \mathbb{E}\left|\int_0^r K(\theta,r)\,dW_\theta\right|^2 = \mathbb{E}\left|f_r - \mathbb{E}^{\mathscr{F}_0}f_r\right|^2$$
$$\le 2\mathbb{E}|f_r|^2 + 2\mathbb{E}\left|\mathbb{E}^{\mathscr{F}_s}f_r\right|^2 \le 4\mathbb{E}|f_r|^2,$$

the proof of the first inequality in (6.29) is finished. Now we prove the second one, estimating separately the two terms on the right-hand side of (6.31). By the Doob inequality for martingales,

$$\mathbb{E}\sup_{s\in[0,T]} e^{2\beta s}|\mathbb{E}^{\mathscr{F}_s}\eta|^2 \le e^{2\beta T}4\,\mathbb{E}\,|\eta|^2.$$

Next, since

$$\left(\int_s^T |f_r|\,dr\right)^2 \le \int_s^T e^{-2\beta r}\,dr\int_s^T e^{2\beta r}|f_r|^2\,dr \le \frac{e^{-2\beta s}}{2\beta}\int_s^T e^{2\beta r}|f_r|^2\,dr,$$

we obtain

$$e^{\beta s}\left|\int_s^T \mathbb{E}^{\mathscr{F}_s}f_r\,dr\right| \le \mathbb{E}^{\mathscr{F}_s}\left(e^{\beta s}\int_s^T |f_r|\,dr\right) \le \frac{1}{\sqrt{2\beta}}\mathbb{E}^{\mathscr{F}_s}\left(\int_s^T e^{2\beta r}|f_r|^2\,dr\right)^{1/2}$$

and by the Doob inequality,

$$\mathbb{E} \sup_{s\in[0,T]} e^{2\beta s} \left| \int_s^T \mathbb{E}^{\mathscr{F}_s} f_r\, dr \right|^2 \le \frac{4}{2\beta} \mathbb{E} \int_0^T e^{2\beta r} |f_r|^2\, dr.$$

Estimates (6.30). Since, for $s \in [T-\delta, T]$,

$$\int_s^T |f_r|\, dr \le \left(\int_s^T |f_r|^2\, dr \right)^{1/2} (T-s)^{1/2} \le \left(\int_s^T f_r\, dr \right)^{1/2} \delta^{1/2},$$

it follows from (6.31) that

$$\begin{aligned} \mathbb{E} \sup_{s\in[T-\delta,T]} |Y_s|^p \le c_p \mathbb{E} \sup_{s\in[T-\delta,T]} |\mathbb{E}^{\mathscr{F}_s}\eta|^p \\ + c_p \delta^{p/2} \mathbb{E} \sup_{s\in[T-\delta,T]} \left| \mathbb{E}^{\mathscr{F}_s} \left(\int_s^T |f_r|^2\, dr \right)^{1/2} \right|^p \\ \le c_p \mathbb{E}\, |\eta|^p + c_p \delta^{p/2} \mathbb{E} \left(\int_{T-\delta}^T |f_r|^2\, dr \right)^{p/2}, \end{aligned}$$

which proves the desired inequality on the process Y. To obtain a similar estimate on Z we first set $Z^1_\theta = \int_\theta^T K(\theta, r)\, dr$, so that $Z_\theta = L_\theta + Z^1_\theta$.

From (6.32) it follows that $\mathbb{E}^{\mathscr{F}_s}\eta - \mathbb{E}^{\mathscr{F}_{T-\delta}}\eta = \int_{T-\delta}^s L_\theta\, dW_\theta$, so by the Burkholder–Davis–Gundy and the Doob inequalities,

$$\begin{aligned} \mathbb{E} \left(\int_{T-\delta}^T |L_\theta|^2 d\theta \right)^{\frac{p}{2}} \le c_p\, \mathbb{E} \sup_{s\in[T-\delta,T]} \left| \int_{T-\delta}^s L_\theta\, dW_\theta \right|^p \\ = c_p\, \mathbb{E} \sup_{s\in[T-\delta,T]} |\mathbb{E}^{\mathscr{F}_s}\eta - \mathbb{E}^{\mathscr{F}_{T-\delta}}\eta|^p \le c_p\, \mathbb{E}\, |\eta|^p. \end{aligned}$$

In order to prove a similar estimate for Z^1 we first note that, setting $Y^1_s = \int_s^T \mathbb{E}^{\mathscr{F}_s} f_r\, dr$, the pair (Y^1, Z^1) is the solution corresponding to $\eta = 0$. Therefore

$$Y^1_s - Y^1_{T-\delta} = \int_{T-\delta}^s Z^1_r\, dW_r - \int_{T-\delta}^s f_r\, dr.$$

So we obtain

$$\begin{aligned} \mathbb{E} \left(\int_{T-\delta}^T |Z^1_r|^2 dr \right)^{\frac{p}{2}} \le c_p\, \mathbb{E} \sup_{s\in[T-\delta,T]} \left| \int_{T-\delta}^s Z^1_r\, dW_r \right|^p \\ \le c_p\, \mathbb{E} \sup_{s\in[T-\delta,T]} |Y^1_s|^p + c_p\, \mathbb{E} \left(\int_{T-\delta}^T |f_r|\, dr \right)^p. \end{aligned}$$

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

$$\mathbb{E} \sup_{s\in[T-\delta,T]} |\mathbb{E}^{\mathscr{F}_s} Y_s^1|^p \le c_p \delta^{p/2} \mathbb{E} \left(\int_{T-\delta}^T |f_r|^2\, dr \right)^{p/2}.$$

Finally, the required estimate follows from

$$\int_{T-\delta}^T |f_r|\, dr \le \left(\int_{T-\delta}^T |f_r|^2\, dr \right)^{1/2} \delta^{1/2}.$$

□

Now we are concerned with the equation

$$Y_s + \int_s^T Z_r\, dW_r = \int_s^T f(r, Y_r, Z_r)\, dr + \eta. \tag{6.34}$$

In the following Proposition K is a Hilbert space, the mapping $f : \Omega \times [0, T] \times K \times \mathcal{L}_2(\Xi, K) \to K$ is assumed to be measurable with respect to $\mathcal{P} \times \mathcal{B}([0, T] \times K \times \mathcal{L}_2(\Xi, K))$ and $\mathcal{B}(K)$, respectively (we recall that by $\mathcal{P}$ we denote the predictable σ-field on $\Omega \times [0, T]$ and by $\mathcal{B}(\Lambda)$ the Borel σ-field of any topological space Λ). $\eta : \Omega \to K$ is assumed to be $\mathscr{F}_T$-measurable.

Proposition 6.20 *Assume that*

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

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

$\mathbb{P}$-*a.s. for every* $t \in [0, T]$, $y_1, y_2 \in K$, $z_1, z_2 \in \mathcal{L}_2(\Xi, K)$;

(ii) $\mathbb{E} \int_0^T |f(r, 0, 0)|^2 dr < \infty, \qquad \mathbb{E}\, |\eta|^2 < \infty.$

Then there exists a unique pair of processes $Y(s)$, $Z(s)$, $s \in [0, T]$, *such that*

$$Y \in C_{\mathcal{P}}([0, T], L^2(\Omega; K)), \qquad Z \in L^2_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$$

and (6.34) holds for $s \in [0, T]$. *Moreover, Y has a continuous version and* $\mathbb{E} \sup_{s\in[0,T]} |Y(s)|^2 < \infty$.

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

$$\mathbb{E} \left(\int_0^T |f(r, 0, 0)|^2 dr \right)^{p/2} < \infty, \qquad \mathbb{E}\, |\eta|^p < \infty, \tag{6.35}$$

then we have $Y \in L^p_{\mathcal{P}}(\Omega; C([0, T], K))$, $Z \in L^p_{\mathcal{P}}(\Omega \times [0, T]; \mathcal{L}_2(\Xi, K))$ *and*

$$\mathbb{E}\sup_{s\in[0,T]}|Y(s)|^p+\mathbb{E}\left(\int_0^T|Z(r)|^2dr\right)^{p/2}\le c\,\mathbb{E}\left(\int_0^T|f(r,0,0)|^2dr\right)^{p/2}+c\,\mathbb{E}\,|\eta|^p,\tag{6.36}$$

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

Finally assume that, for all λ in a metric space Λ, a function f_λ is given satisfying (6.35) and assumption i) with L independent of λ. Also assume that, as $\lambda\to\lambda_0$,

$$\mathbb{E}\left(\int_0^T|f_\lambda(r,Y,Z)-f_{\lambda_0}(r,Y,Z)|^2dr\right)^{p/2}\to 0\tag{6.37}$$

for all $Y\in L^p_{\mathcal{P}}(\Omega;C([0,T],K))$, $Z\in L^p_{\mathcal{P}}(\Omega;L^2([0,T];\mathcal{L}_2(\Xi,K)))$.

If we denote by $(Y(\lambda,\eta),Z(\lambda,\eta))$ the solution to (6.34) corresponding to $f=f_\lambda$ and to the final data $\eta\in L^p(\Omega,\mathbb{R})$ then the map $(\lambda,\eta)\to(Y(\lambda,\eta),Z(\lambda,\eta))$ is continuous from $\Lambda\times L^p(\Omega;\mathbb{R})$ to $L^p_{\mathcal{P}}(\Omega;C([0,T],K))\times L^p_{\mathcal{P}}(\Omega;L^2([0,T];\mathcal{L}_2(\Xi,K)))$.

Proof We let $\mathcal{K}=C_{\mathcal{P}}([0,T],L^2(\Omega;K))\times L^2_{\mathcal{P}}(\Omega\times[0,T];\mathcal{L}_2(\Xi,K))$ and we define a mapping $\Gamma:\mathcal{K}\to\mathcal{K}$ by setting $(Y,Z)=\Gamma(U,V)$ if (Y,Z) is the pair satisfying

$$Y_s+\int_s^T Z_r\,dW_r=\int_s^T f(r,U_r,V_r)\,dr+\eta,\tag{6.38}$$

compare Lemma 6.19. The estimates (6.29) show that Γ is well defined, and it is a contraction if $\mathcal{K}$ is endowed with the norm

$$|(Y,Z)|^2_{\mathcal{K}}=\mathbb{E}\int_0^T e^{2\beta r}\left(|Y_r|^2+|Z_r|^2\right)dr,$$

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

$$\overline{Y}_s+\int_s^T\overline{Z}_r\,dW_r=\int_s^T\overline{f}_r\,dW_r,\tag{6.39}$$

so that by (6.29),

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

Now we prove the estimate (6.36). We let $\mathcal{K}_{p,\delta}=L^p(\Omega;C([T-\delta,T],\mathbb{R}))\times L^p(\Omega;L^2([T-\delta,T];\mathcal{L}_2(\Xi,\mathbb{R})))$ and define $\Gamma:\mathcal{K}_{p,\delta}\to\mathcal{K}_{p,\delta}$, setting $(Y,Z)=\Gamma(U,V)$ if (Y,Z) is the pair satisfying Eq. (6.38) for $s\in[T-\delta,T]$. It is easily verified that Γ is well defined and it is a contraction in $\mathcal{K}_{p,\delta}$, provided $\delta > 0$ is chosen sufficiently small; indeed, arguing as before, we deduce from (6.39) and from (6.30) the inequalities

$$
\begin{aligned}
|(\overline{Y}, \overline{Z})|_{\mathcal{K}}^p &= \mathbb{E} \sup_{s\in[T-\delta,T]} |\overline{Y}_s|^p + \mathbb{E}\left(\int_{T-\delta}^T |\overline{Z}_r|^2\, dr\right)^{\frac{p}{2}} \\
&\le c_p \delta^{p/2} L^p\, \mathbb{E}\left(\int_{T-\delta}^T (|\overline{U}_r| + |\overline{V}_r|)^2 dr\right)^{\frac{p}{2}} \\
&\le c_p 2^{p/2}\delta^p L^p \delta\, \mathbb{E} \sup_{s\in[T-\delta,T]} |\overline{U}_s|^p + c_p (2\delta)^{p/2} L^p\, \mathbb{E}\left(\int_{T-\delta}^T |\overline{V}_r|^2 dr\right)^{\frac{p}{2}} \\
&\le c_p (2\delta)^{p/2} L^p (1+\delta^{p/2})\, |(\overline{U}, \overline{V})|_{\mathcal{K}}^p,
\end{aligned}
$$

and the contraction property holds provided $c_p(2\delta)^{p/2}L^p(1+\delta^{p/2}) < 1$. Repeating this argument on intervals $[T-\delta, T-2\delta]$, $[T-2\delta, T-3\delta]$ etc. shows that $Y \in L^p(\Omega; C([0,T],\mathbb{R}))$ and $Z \in L^p(\Omega; L^2([0,T]; \mathcal{L}_2(\Xi,\mathbb{R})))$.

Next note that it follows from our assumptions that

$$|f(r,x,y)| \le |f(r,0,0)| + L(|x| + |y|).$$

Applying the estimate (6.30) to Eq. (6.34) we obtain

$$
\begin{aligned}
&\mathbb{E} \sup_{s\in[T-\delta,T]} |Y_s|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_r|^2 dr\right)^{p/2} \\
&\quad\le c_p \delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |f(r,Y_r,Z_r)|^2 dr\right)^{p/2} + c_p \mathbb{E}\,|\eta|^p \\
&\quad\le c_p \mathbb{E}\,|\eta|^p + c_p 3^{p-1}\delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |f(r,0,0)|^2 dr\right)^{p/2} \\
&\quad + c_p 3^{p-1} L^p \delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |Y_r|^2 dr\right)^{p/2} + c_p 3^{p-1} L^p \delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |Z_r|^2 dr\right)^{p/2} \\
&\quad\le c_p \mathbb{E}\,|\eta|^p + c_p 3^{p-1}\delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |f(r,0,0)|^2 dr\right)^{p/2} \\
&\quad + c_p 3^{p-1} L^p \delta^p \mathbb{E} \sup_{s\in[T-\delta,T]} |Y_s|^p + c_p 3^{p-1} L^p \delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |Z_r|^2 dr\right)^{p/2}.
\end{aligned}
\tag{6.40}
$$

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

$$\mathbb{E} \sup_{s\in[T-\delta,T]} |Y_s|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_r|^2 dr\right)^{p/2}$$
$$\leq c_p \mathbb{E}\,|\eta|^p + c_p 3^{p-1}\delta^{p/2}\mathbb{E}\left(\int_{T-\delta}^T |f(r,0,0)|^2 dr\right)^{p/2}$$
$$+ \alpha\left[\mathbb{E} \sup_{s\in[T-\delta,T]} |Y_s|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_r|^2 dr\right)^{p/2}\right], \tag{6.41}$$

and it follows that

$$\mathbb{E} \sup_{s\in[T-\delta,T]} |Y_s|^p + \mathbb{E}\left(\int_{T-\delta}^T |Z_r|^2 dr\right)^{p/2} \leq c\mathbb{E}\,|\eta|^p + c\mathbb{E}\left(\int_{T-\delta}^T |f(r,0,0)|^2 dr\right)^{p/2},$$

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

$$Y_s + \int_s^{T-\delta} Z_r\, dW_r = \int_s^{T-\delta} f(r, Y_r, Z_r)\, dr + Y_{T-\delta},$$

and proceeding as before we obtain

$$\mathbb{E} \sup_{s\in[T-2\delta,T-\delta]} |Y_s|^p + \mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |Z_r|^2 dr\right)^{p/2} \leq c\mathbb{E}\,|Y_{T-\delta}|^p + c\mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |f(r,0,0)|^2 dr\right)^{p/2},$$

with the same choice of δ and the same value of c. After a finite number of steps we arrive at (6.36).

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

Remark 6.21 The mapping Γ defined in the previous proof was shown to be a contraction in the space $\mathcal{K} = C_{\mathcal{P}}([0,T], L^2(\Omega; K)) \times L^2_{\mathcal{P}}(\Omega \times [0,T]; \mathcal{L}_2(\Xi, K))$. In a similar way, the estimates (6.29) allow us to show that Γ is well defined and it is a contraction in the space $L^2_{\mathcal{P}}(\Omega; C([0,T], K)) \times L^2_{\mathcal{P}}(\Omega \times [0,T]; \mathcal{L}_2(\Xi, K))$ as well as in the space $L^2_{\mathcal{P}}(\Omega \times [0,T]; K) \times L^2_{\mathcal{P}}(\Omega \times [0,T]; \mathcal{L}_2(\Xi, K))$. In particular, uniqueness holds for Eq. (6.34) in the latter space, too. ■

6.3.2 *Regular Dependence on Data*

Now we are dealing with the backward equation

$$Y(s) + \int_s^T Z(r)\, dW(r) = \int_s^T F(r, X(r), Y(r), Z(r))\, dr + \eta, \tag{6.42}$$

on the time interval $[0, T]$, where η is a given $\mathscr{F}_T$-measurable real random variable and $X(s)$, $s \in [0, T]$, is a given predictable process. The mapping $F : [0, T] \times H \times K \times \mathcal{L}_2(\Xi, K) \to K$ is assumed to be Borel measurable. The solution we are looking for is a pair of predictable processes $Y(s)$, $Z(s)$, $s \in [0, T]$, with values in K and $\mathcal{L}_2(\Xi, K)$, respectively.

We fix the following assumptions on F.

Hypothesis 6.22 (i) There exists an $L > 0$ such that

$$|F(t, x, y_1, z_1) - F(t, x, y_2, z_2)| \leq L(|y_1 - y_2| + |z_1 - z_2|),$$

for every $t \in [0, T]$, $x \in H$, $y_1, y_2 \in K$, $z_1, z_2 \in \mathcal{L}_2(\Xi, K)$.

(ii) For every $t \in [0, T]$, $F(t, \cdot, \cdot, \cdot) \in \mathcal{G}^{1,1,1}(H \times K \times \mathcal{L}_2(\Xi, K), K)$.

(iii) There exist $L > 0$ and $m \geq 0$ such that

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

for every $t \in [0, T]$, $x, h \in H$, $y \in K$, $z \in \mathcal{L}_2(\Xi, K)$.

(iv) There exists an $L > 0$ such that $|F(t, 0, 0, 0)| \leq L$ for every $t \in [0, T]$.

Conditions (i) and (ii) imply that the Gâteaux derivatives of F with respect to y and z are uniformly bounded: for every point (x, y, z) and all directions $k \in K$, $v \in \mathcal{L}_2(\Xi, K)$,

$$|\nabla_y F(t, x, y, z)k| \leq L\ |k|, \quad |\nabla_z F(t, x, y, z)v| \leq L\ |v|.$$

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

$$|F(t, x, y, z)| \leq L(1 + |x|^{m+1} + |z| + |y|). \tag{6.43}$$

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

$$|F(t, x_1, y, z) - F(t, x_2, y, z)| \leq L(1 + |z|)(1 + |x_1|^m + |x_2|^m + |y|^m)|x_2 - x_1|. \tag{6.44}$$

Remark 6.23 Instead of condition (iii), in some of the statements below we will assume that the stronger condition holds: there exists $L > 0$ such that

$$|\nabla_x F(t, x, y, z)h| \leq L|h|, \qquad t \in [0, T],\ x, h \in H,\ y \in K,\ z \in \mathcal{L}_2(\Xi, K). \tag{6.45}$$

Whenever (6.45) is assumed to hold, this will be explicitly mentioned. ■

To start we need the following general lemma that generalizes the classical result on continuity of evaluation operators, see e.g. [10].

Lemma 6.24 *Let K_1, K_2 and K_3 be Banach spaces and $\ell : [0,T] \times K_1 \times K_2 \to K_3$ be a measurable map such that, for all $t \in [0,T]$, $\ell(t,\cdot) : K_1 \times K_2 \to K_3$ is continuous*

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

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

For all $U \in L^{r_1}_{\mathcal{P}}(\Omega; C([0,T], K_1))$, $V \in L^{r_2}_{\mathcal{P}}(\Omega; L^2([0,T]; K_2))$ with $r_1, r_2 \geq 1$, let us define in the natural way the evaluation operator $\ell(U,V)(t,\omega) = \ell(t, U(t,\omega), V(t,\omega))$.
If $\mu/r_1 + 1/r_2 = 1/r_3$ and $r_1 \geq \mu$ then the evaluation operator is continuous from $L^{r_1}_{\mathcal{P}}(\Omega; C([0,T], K_1)) \times L^{r_2}_{\mathcal{P}}(\Omega; L^2([0,T]; K_2))$ to $L^{r_3}_{\mathcal{P}}(\Omega; L^2([0,T]; K_3))$.

(ii) Similarly, if

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

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

Proof We prove only (i), the proof of (ii) being identical.

Step 1. Firstly we consider only dependence on t. Define the evaluation operator (denoted again by ℓ by abuse of language): $\ell(\mathcal{U}, \mathcal{V})(t) = \ell(t, \mathcal{U}(t), \mathcal{V}(t))$ with $\mathcal{U} \in C([0,T], K_1)$, $\mathcal{V} \in L^2([0,T]; K_2)$. We claim that ℓ is continuous from $C([0,T], K_1) \times L^2([0,T]; K_2)$ to $L^2([0,T]; K_3)$. It is enough to prove that

$$\int_0^T |\ell(t, \mathcal{U}_n(t), \mathcal{V}_n(t)) - \ell(t, \mathcal{U}(t), \mathcal{V}(t))|^2 \, dt \to 0$$

for each pair of sequences $\mathcal{U}_n$, $\mathcal{V}_n$ with $\mathcal{U}_n \to \mathcal{U}$ in $C([0,T], K_1)$ and $\mathcal{V}_n \to \mathcal{V}$ in $L^2([0,T]; K_2)$. Extracting a subsequence, if necessary, we can always assume that $\sum_{n=1}^{\infty} |\mathcal{V}_n - \mathcal{V}|_{L^2([0,T];K_2)} < +\infty$ and $\mathcal{V}_n(t) \to \mathcal{V}(t)$ for a.a. $t \in [0,T]$. Let $\mathcal{V}^*(t) = \sum_{n=1}^{\infty} |\mathcal{V}_n(t) - \mathcal{V}(t)|_{K_2}$. By construction $\mathcal{V}^* \in L^2([0,T]; \mathbb{R})$ and $|\mathcal{V}_n(t)|_{K_2} \leq |\mathcal{V}(t)|_{K_2} + \mathcal{V}^*(t)$. Therefore

$$|\ell(t, \mathcal{U}_n(t), \mathcal{V}_n(t)) - \ell(t, \mathcal{U}(t), \mathcal{V}(t))|^2 \leq L \left(1 + \sup_n |\mathcal{U}_n|^{\mu}_{C([0,T],K_1)}\right)^2 \left(1 + |\mathcal{V}(t)|_{K_2} + \mathcal{V}^*(t)\right)^2,$$

for a suitable constant L. Since the right-hand term is a fixed summable function of $t \in [0,T]$ the claim follows from the dominated convergence theorem. Finally, we observe that

$$|\ell(\mathcal{U},\mathcal{V})|_{L^2([0,T];K_3)} \le L\left(1+|\mathcal{U}|^{\mu}_{C([0,T],K_1)}\right)\left(1+|\mathcal{V}|_{L^2([0,T];K_2)}\right)$$

for a suitable constant L.

Step 2. Now we consider dependence on ω. Let $\hat{\ell}$ be a continuous map $\hat{K}_1\times\hat{K}_2\to\hat{K}_3$, with $\hat{K}_i$ Banach spaces, $i=1,2,3$, and $|\hat{\ell}(u,v)|_{\hat{K}_3}\le L(1+|u|^{\mu}_{\hat{K}_1})(1+|v|_{\hat{K}_2})$. For $U\in L^{r_1}(\Omega;\hat{K}_1)$, $V\in L^{r_2}(\Omega;\hat{K}_2)$ with $\mu/r_1+1/r_2=1/r_3$, we define the evaluation operator $\hat{\ell}(U,V)(\omega)=\hat{\ell}(U(\omega),V(\omega))$ and claim that it is continuous from $L^{r_1}(\Omega;\hat{K}_1)\times L^{r_2}(\Omega;\hat{K}_2)$ to $L^{r_3}(\Omega;\hat{K}_3)$. Before proving the claim we notice that it completes the proof of Lemma 6.24: indeed, it suffices to apply it to $\hat{K}_1=C([0,T],K_1)$, $\hat{K}_2=L^2([0,T];K_2)$, $\hat{K}_3=L^2([0,T];K_3)$ and to the evaluation operator introduced in Step 1.

The proof of the claim is similar to that of Step 1. It is enough to show that:

$$\mathbb{E}\left(\left|\hat{\ell}(U_n,V_n)-\hat{\ell}(U,V)\right|^{r_3}_{\hat{K}_3}\right)\to 0$$

for each pair of sequences U_n in $L^{r_1}(\Omega;\hat{K}_1)$ and V_n in $L^{r_2}(\Omega;\hat{K}_2)$ with $U_n\to U$ in $L^{r_1}(\Omega;\hat{K}_1)$ and $V_n\to V$ in $L^{r_2}(\Omega;\hat{K}_2)$. Extracting a subsequence, if necessary, we can assume that $U_n\to U$ and $V_n\to V$ $\mathbb{P}$-a.s., and

$$\sum_{n=1}^{\infty}|U_n-U|_{L^{r_1}(\Omega;\hat{K}_1)}<+\infty,\qquad \sum_{n=1}^{\infty}|V_n-V|_{L^{r_2}(\Omega;\hat{K}_2)}<+\infty.$$

Let:

$$U^*=\sum_{n=1}^{\infty}|U_n-U|_{\hat{K}_1},\qquad V^*=\sum_{n=1}^{\infty}|V_n-V|_{\hat{K}_2}.$$

By construction $U^*\in L^{r_1}(\Omega;\mathbb{R})$ and $V^*\in L^{r_2}(\Omega;\mathbb{R})$. Moreover:

$$|U_n(\omega)|_{\hat{K}_1}\le|U(\omega)|_{\hat{K}_1}+U^*(\omega),\qquad |V_n(\omega)|_{\hat{K}_2}\le|V(\omega)|_{\hat{K}_2}+V^*(\omega),\qquad \mathbb{P}\text{-a.s.}$$

Therefore

$$\begin{aligned}\left|\hat{\ell}(U_n(\omega),V_n(\omega))-\hat{\ell}(U(\omega),V(\omega))\right|^{r_3}_{\hat{K}_3}&\le L\left(1+|U(\omega)|^{\mu r_3}_{\hat{K}_1}+(U^*(\omega))^{\mu r_3}\right)\\&\quad\cdot\left(1+|V(\omega)|^{r_3}_{\hat{K}_2}+(V^*(\omega))^{r_3}\right),\qquad \mathbb{P}\text{-a.s.,}\end{aligned}$$

for a suitable constant L. Since $(\mu r_3)/r_1+r_3/r_2=1$ the left-hand term has finite mean and the claim follows from the dominated convergence theorem. □

We are now in a position to show the existence and uniqueness and regular dependence on data of the solution to Eq. (6.42). For $p\ge 2$ we define:

$$\mathcal{K}_p = L^p_{\mathcal{P}}(\Omega; C([0, T], K)) \times L^p_{\mathcal{P}}(\Omega; L^2([0, T]; \mathcal{L}_2(\Xi, K))),$$

endowed with the natural norm.

Proposition 6.25 *Assume Hypotheses 6.8 and 6.22.*

(i) If $X \in L^\rho_{\mathcal{P}}(\Omega; C([0, T], H))$, $\eta \in L^r(\Omega; K)$ with $\rho = r(m+1)$, $r \geq 2$ then there exists a unique solution in $\mathcal{K}_r$ of Eq. (6.42), which we will denote by $(Y(\cdot, X, \eta), Z(\cdot, X, \eta))$.

(ii) The following estimate holds:

$$\mathbb{E} \sup_{s\in[0,T]} |Y(s, X, \eta)|^r + \left(\mathbb{E} \int_0^T |Z(s, X, \eta)|^2 ds\right)^{r/2} \leq c \left(1 + |X|^\rho_{L^\rho_{\mathcal{P}}(\Omega;C([0,T],H))}\right) + c\mathbb{E}|\eta|^r \tag{6.46}$$

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

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

(iv) The map $(X, \eta) \to (Y(\cdot, X, \eta), Z(\cdot, X, \eta))$ is in $\mathcal{G}^{1,1}(L^\rho_{\mathcal{P}}(\Omega; C([0, T], H)) \times L^r(\Omega; \mathbb{R}), \mathcal{K}_p)$ with $r = (m+2)p$, $p \geq 2$ (consequently $\rho = p(m+1)(m+2)$).
Moreover, for all $X \in L^\rho_{\mathcal{P}}(\Omega; C([0, T], H))$, $\eta \in L^r(\Omega; K)$ the directional derivative in the direction (N, ζ) with $N \in L^\rho_{\mathcal{P}}(\Omega; C([0, T], H))$ and $\zeta \in L^r(\Omega; K)$, which we will denote by $(\nabla_{X,\eta} Y(\cdot, X, \eta)(N, \zeta), \nabla_{X,\eta} Z(\cdot, X, \eta)(N, \zeta))$, is the unique solution in $\mathcal{K}_p$ of:

$$\begin{aligned}
&\nabla_{X,\eta} Y(s, X, \eta)(N, \zeta) + \int_s^T \nabla_{X,\eta} Z(r, X, \eta)(N, \zeta) dW_r \\
&\quad = \int_s^T \nabla_x F(r, X_r, Y_r(X, \eta), Z_r(X, \eta)) N_r dr \\
&\qquad + \int_s^T \nabla_y F(r, X_r, Y_r(X, \eta), Z_r(X, \eta)) \nabla_{X,\eta} Y(r, X, \eta)(N, \zeta) dr \\
&\qquad + \int_s^T \nabla_z F(r, X_r, Y_r(X, \eta), Z_r(X, \eta)) \nabla_{X,\eta} Z(r, X, \eta)(N, \zeta) dr + \zeta.
\end{aligned}$$

(v) Finally, the following estimate holds:

$$\begin{aligned}
&\mathbb{E} \sup_{s\in[0,T]} |\nabla_{X,\eta} Y(s, X, \eta)(N, \zeta)|^p + \mathbb{E} \left(\int_0^T |\nabla_{X,\eta} Z(s, X, \eta)(N, \zeta)|^2 ds\right)^{p/2} \\
&\leq c|N|^p_{L^r_{\mathcal{P}}(\Omega;C([0,T],H))} \left(1 + |X|^{(m+1)^2}_{L^\rho_{\mathcal{P}}(\Omega;C([0,T],H))} + |\eta|^{m+1}_{L^r(\Omega;K)}\right)^p + c|\zeta|^p_{L^p(\Omega;K)}.
\end{aligned} \tag{6.47}$$

(vi) If, in addition, there exists an $L > 0$ such that

$$|\nabla_x F(t, x, y, z)h| \le L|h|, \qquad t \in [0, T],\ x, h \in H,\ y \in K,\ z \in \mathcal{L}_2(\Xi, K),$$

then the following estimate (stronger than (6.47)) holds:

$$\mathbb{E} \sup_{s\in[0,T]} |\nabla_{X,\eta} Y(s, X, \eta)(N, \zeta)|^p + \mathbb{E}\left(\int_0^T |\nabla_{X,\eta} Z(s, X, \eta)(N, \zeta)|^2 ds\right)^{p/2} \le c|N|^p_{L^p_{\mathcal{P}}(\Omega; C([0,T],H))} + c|\zeta|^p_{L^p(\Omega;K)}. \tag{6.48}$$

Proof Let $\Lambda = L^\rho_{\mathcal{P}}(\Omega; C([0, T], H))$ and, for every $X \in \Lambda$,

$$f_X(s, y, z) = F(s, X_s, y, z).$$

By (6.43) and Lemma 6.24-(ii) applied with $K_1 = H$, $K_2 = K \times \mathcal{L}_2(\Xi, K)$, $U = X$, $V = (Y, Z)$ we obtain that for all $(Y, Z) \in \mathcal{K}_r$ the map $X \to f_X(Y, Z)$ is continuous from Λ to $L^r_{\mathcal{P}}(\Omega; L^2([0, T]; K))$ and

$$\mathbb{E}\left(\int_0^T |f_X(s, 0, 0)|^2 ds\right)^{r/2} \le c\left(1 + \mathbb{E}(\sup_{s\in[0,T]} |X_s|^{r(m+1)})\right).$$

Therefore points (i)–(iii) of the claim follow immediately from Proposition 6.20.

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

$$\begin{aligned} \widehat{Y}_s + \int_s^T \widehat{Z}_r dW_r = \int_s^T \nabla_x F(r, X_r, Y_r, Z_r) N_r dr \\ \int_s^T \nabla_y F(r, X_r, Y_r, Z_r)\widehat{Y}_r dr + \int_s^T \nabla_z F(r, X_r, Y_r, Z_r)\widehat{Z}_r dr + \zeta. \end{aligned} \tag{6.49}$$

By Hypothesis 6.22-(iii) we have

$$\begin{aligned} &\mathbb{E}\left(\int_0^T |\nabla_x F(r, X_r, Y_r, Z_r)N_r|^2 dr\right)^{p/2} \\ &\quad \le L|N|^p_{L^r_{\mathcal{P}}(\Omega; C([0,T],H))}\left(1 + |Z|_{L^r_{\mathcal{P}}(\Omega; L^2([0,T]; \mathcal{L}_2(\Xi,K)))}\right)^p \\ &\qquad \cdot \left(1 + |X|^m_{L^r_{\mathcal{P}}(\Omega; C([0,T],H))} + |Y|^m_{L^r_{\mathcal{P}}(\Omega; C([0,T],H))}\right)^p \end{aligned}$$

for a suitable constant L. Since $\nabla_y F$ and $\nabla_z F$ are bounded, by Proposition 6.20 the Eq. (6.49) admits a unique solution in $\mathcal{K}_p$. Moreover, by Lemma 6.24-(i), the map $(X, N, Y, Z) \to \nabla_x F(\cdot, X_{(\cdot)}, Y_{(\cdot)}, Z_{(\cdot)})N_{(\cdot)}$ is continuous from the space

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

to $L^p_{\mathcal{P}}(\Omega; L^2([0, T]; K))$. Therefore, taking into account once more the boundedness of $\nabla_y F$ and $\nabla_z F$, we can apply the final statement of Proposition 6.20 with $\Lambda = K^{\#}$ and conclude that the map $(X, N, Y, Z, \zeta) \to (\widehat{Y}(X, N, Y, Z, \zeta), \widehat{Z}(X, N, Y, Z, \zeta))$ is continuous from $K^{\#} \times L^p(\Omega; K)$ to $\mathcal{K}_p$ and the estimate

$$\begin{aligned} \mathbb{E}(\sup_{s\in[0,T]} |\widehat{Y}_s|^p) + \mathbb{E}\left(\int_0^T |\widehat{Z}_r|^2 dr\right)^{p/2} &\le c|N|^p_{L^r_{\mathcal{P}}(\Omega;C([0,T],H))}\left(1 + |Z|_{L^r_{\mathcal{P}}(\Omega;L^2([0,T];\mathcal{L}_2(\Xi,K)))}\right)^p \\ &\cdot \left(1 + |X|^m_{L^r_{\mathcal{P}}(\Omega;C([0,T],H))} + |Y|^m_{L^r_{\mathcal{P}}(\Omega;C([0,T],H))}\right)^p + c\mathbb{E}|\zeta|^p \end{aligned} \tag{6.50}$$

holds for some constant $c > 0$.

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

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

Let us define

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

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

For $\varepsilon \to 0$ we show that $\overline{Y}^\varepsilon \to 0$ in $L^p_{\mathcal{P}}(\Omega; C([0, T], K))$ and $\overline{Z}^\varepsilon \to 0$ in $L^p_{\mathcal{P}}(\Omega; L^2([0, T]; \mathcal{L}_2(\Xi, K)))$. For short we let $Y = Y(X, \eta)$, $Z = Z(X, \eta)$, $Y^\varepsilon = Y(X + \varepsilon N, \eta + \varepsilon\zeta)$, $Z^\varepsilon = Z(X + \varepsilon N, \eta + \varepsilon\zeta)$, $\widehat{Y} = \widehat{Y}(X, N, Y(X, \eta), Z(X, \eta), \zeta)$, and $\widehat{Z} = \widehat{Z}(X, N, Y(X, \eta), Z(X, \eta), \zeta)$.

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

$$\overline{Y}^\varepsilon_s + \int_s^{T-\delta} \overline{Z}^\varepsilon_r\, dr = \int_s^{T-\delta} \nu^\varepsilon(r)dr + \overline{Y}^\varepsilon_{T-\delta},$$

where $\nu^\varepsilon = \nu_1^\varepsilon + \nu_2^\varepsilon$ and:

$$\nu_1^\varepsilon(r) = \frac{1}{\varepsilon}\left[F(r, X_r + \varepsilon N_r, Y_r^\varepsilon, Z_r^\varepsilon) - F(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon)\right] - \nabla_x F(r, X_r, Y_r, Z_r)N_r,$$

$$\nu_2^\varepsilon(r) = \frac{1}{\varepsilon}\left[F(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon) - F(r, X_r, Y_r, Z_r)\right] - \nabla_y F(r, X_r, Y_r, Z_r)\widehat{Y}_r - \nabla_z F(r, X_r, Y_r, Z_r)\widehat{Z}_r.$$

By Proposition 6.20 we have:

$$\mathbb{E} \sup_{s\in[T-2\delta, T-\delta]} |\overline{Y}_s^\varepsilon|^p + \mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |\overline{Z}_r^\varepsilon|^2 dr\right)^{p/2} \le c_p \delta^{p/2} \sum_{i=1}^{2} \mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |\nu_i^\varepsilon(r)|^2 dr\right)^{p/2} + c_p \mathbb{E}|\overline{Y}_{T-\delta}^\varepsilon|^p$$

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

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

$$\nu_1^\varepsilon(r) = \int_0^1 \nabla_x F(r, X_r + \varepsilon\tau N_r, Y_r^\varepsilon, Z_r^\varepsilon)N_r d\tau - \int_0^1 \nabla_x F(r, X_r, Y_r, Z_r)N_r d\tau.$$

For all $x, g, n \in H$, $y \in K$, $z \in \mathcal{L}_2(\Xi, K)$ let $\chi(x, g, n, y, z) = \int_0^1 \nabla_x F(x + \tau g, y, z) n d\tau$, so that $\nu_1^\varepsilon(r) = \chi(X_r, \varepsilon N_r, N_r, Y_r^\varepsilon, Z_r^\varepsilon) - \chi(X_r, 0, N_r, Y_r, Z_r)$. Moreover, $|\chi(x, g, n, y, z)| \le L|n|(1 + |z|)(1 + |x|^m + |g|^m + |y|^m)$ and χ is a continuous map. Applying Lemma 6.24-(i) with $K_1 = H^{\times 3} \times K$ $K_2 = \mathcal{L}_2(\Xi, K)$, $r_1 = r_2 = r$, $\mu = m + 1$ and taking into account that $(X, \varepsilon N, N, Y^\varepsilon) \to (X, 0, N, Y)$ in $L^r_{\mathcal{P}}(\Omega, C([T - 2\delta, T - \delta], K_1))$ and $Z^\varepsilon \to Z$ in $L^r_{\mathcal{P}}(\Omega, L^2([T - 2\delta, T - \delta], K_2))$ we immediately obtain $\mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |\nu_1^\varepsilon(r)|^2 dr\right)^{p/2} \to 0$.

Dealing now with ν_2^ε we can rewrite $\nu_2^\varepsilon = \nu_{2.1}^\varepsilon + \nu_{2.2}^\varepsilon$ where:

$$\begin{aligned}\nu_{2.1}^\varepsilon(r) = \int_0^1 \Big(&\nabla_y F(r, X_r, Y_r + \tau(Y_r^\varepsilon - Y_r), Z_r + \tau(Z_r^\varepsilon - Z_r))\widehat{Y}_r \\ &- \nabla_y F(r, X_r, Y_r, Z_r)\widehat{Y}_r\Big) d\tau \\ + \int_0^1 \Big(&\nabla_z F(r, X_r, Y_r + \tau(Y_r^\varepsilon - Y_r), Z_r + \tau(Z_r^\varepsilon - Z_r))\widehat{Z}_r \\ &- \nabla_z F(r, X_r, Y_r, Z_r)\widehat{Z}_r\Big) d\tau,\end{aligned}$$

$$\nu^{\varepsilon}_{2.2}(r) = \int_0^1 \nabla_y F(r, X_r, Y_r + \tau(Y^{\varepsilon}_r - Y_r), Z_r + \tau(Z^{\varepsilon}_r - Z_r))\overline{Y}^{\varepsilon}_r d\tau + \int_0^1 \nabla_z F(r, X_r, Y_r + \tau(Y^{\varepsilon}_r - Y_r), Z_r + \tau(Z^{\varepsilon}_r - Z_r))\overline{Z}^{\varepsilon}_r d\tau.$$

Since $\nabla_y F$ and $\nabla_z F$ are bounded, by the dominated convergence theorem we immediately obtain $\mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |\nu^{\varepsilon}_{2.1}(r)|^2 dr\right)^{p/2} \to 0$. Moreover,

$$\mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |\nu^{\varepsilon}_{2.2}(r)|^2 dr\right)^{p/2} \leq \overline{c}\left(\mathbb{E}\sup_{\tau\in[T-2\delta,T-\delta]} |\overline{Y}^{\varepsilon}_{\tau}|^p + \mathbb{E}\left(\int_{T-2\delta}^{T-\delta} |\overline{Z}^{\varepsilon}_r|^2 dr\right)^{p/2}\right)$$

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

Finally, (6.47) follows plugging (6.46) into (6.50), and (6.48) is proved in the same way, taking into account the additional assumption. □

6.3.3 Forward–Backward Systems

In this subsection we consider the system of stochastic differential equations

$$\begin{cases} X(s) = e^{(s-t)A}x + \int_t^s e^{(s-r)A} b(r, X(r))\, dr + \int_t^s e^{(s-r)A}\sigma(r, X(r))\, dW(r), \\ Y(s) + \int_s^T Z(r) dW(r) = \int_s^T F(r, X(r), Y(r), Z(r)) dr + g(X(T)), \end{cases} \tag{6.51}$$

for s varying on the time interval $[t, T] \subset [0, T]$. As in Sect. 6.2 we extend the domain of the solution setting $X(s) = x$ for $s \in [0, t)$. We assume that $F : [0, T] \times H \times \mathbb{R} \times \mathcal{L}_2(\Xi, \mathbb{R}) \to \mathbb{R}$ satisfies Hypothesis 6.22 with $K = \mathbb{R}$. On the function $g : H \to \mathbb{R}$ we make the following assumptions:

Hypothesis 6.26 (i) $g \in \mathcal{G}^1(H, \mathbb{R})$;
(ii) There exist $L > 0$ and $m \geq 0$ such that, for every $x, h \in H$,

$$|\nabla g(x)h| \leq L\, |h|\, (1 + |x|)^m.$$

For simplicity, and without any real loss of generality, we suppose that m is the same as in Hypothesis 6.22. Notice that Hypothesis 6.26 implies that

$$|g(x)| \leq c(1 + |x|^{m+1}).$$

In some of the statements below we will assume the stronger condition: $|\nabla g(x)h| \leq L|h|$, for every $x, h \in H$. Whenever this is the case, this requirement will be explicitly mentioned.

We note that the system (6.51) is decoupled, i.e., the first equation does not contain the solution (Y, Z) of the second one. Therefore, under the assumptions of Hypotheses 6.8, 6.22 and 6.26 by Propositions 6.9 and 6.25 there exists a unique solution to (6.51). We remark that the process X is $\mathscr{F}_{[t,T]}$-measurable, so that Y_t is measurable both with respect to $\mathscr{F}_{[t,T]}$ and $\mathscr{F}_t$; it follows that Y_t is indeed deterministic (see also [207]).

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

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

$$X(s;r,X(r;t,x)) = X(s;t,x), \qquad s \in [r,T], \tag{6.52}$$

is a consequence of the uniqueness of the solution to (6.6). Since the solution to the backward equation is uniquely determined on an interval $[r,T]$ by the values of the process X on the same interval, for $t \leq r \leq T$ we have, $\mathbb{P}$-a.s.,

$$\begin{aligned} Y(s;r,X(r;t,x)) &= Y(s;t,x), \quad \text{for } s \in [r,T], \\ Z(s;r,X(r;t,x)) &= Z(s;t,x) \quad \text{for a.a. } s \in [r,T]. \end{aligned} \tag{6.53}$$

Next we proceed to investigate regularity properties of the dependence on t and x. To this end we first notice that with the notation of Propositions 6.10 and 6.25:

$$Y(s;t,x) = Y(s;X(\cdot;t,x), g(X(T;t,x))),$$

$$Z(s;t,x) = Z(s;X(\cdot;t,x), g(X(T;t,x))).$$

Moreover, as a consequence of Hypothesis 6.26, it can be easily proved that the map $\eta \to g(\eta)$ belongs to the space $\mathcal{G}^1(L^p(\Omega;H), L^q(\Omega;\mathbb{R}))$, for every $p \in [2,\infty)$ and for all q sufficiently large (depending on p and m). The following Proposition is then an immediate consequence of Propositions 6.9, 6.10 and 6.25, and the chain rule for the class $\mathcal{G}$, stated in Lemma 6.3.

Proposition 6.27 *Assume Hypotheses 6.8, 6.22 and 6.26. Recall the notation:*

$$\mathcal{K}_p = L^p_{\mathcal{P}}(\Omega; C([0,T],\mathbb{R})) \times L^p_{\mathcal{P}}(\Omega; L^2([0,T];\mathcal{L}_2(\Xi,\mathbb{R}))).$$

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

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

$$
\begin{aligned}
&\nabla_x Y(s;t,x)h + \int_s^T \nabla_x Z(r;t,x)h\, dW_r \\
&\quad = \int_s^T \nabla_x F(r, X(r;t,x), Y(r;t,x), Z(r;t,x))\nabla_x X(r;t,x)h\, dr \\
&\quad \int_s^T \nabla_y F(r, X(r;t,x), Y(r;t,x), Z(r;t,x))\nabla_x Y(r;t,x)h\, dr \\
&\quad \int_s^T \nabla_z F(r, X(r;t,x), Y(r;t,x), Z(r;t,x))\nabla_x Z(r;t,x)h\, dr \\
&\quad +\nabla g(X(T;t,x))\nabla_x X(T;t,x)h, \qquad s \in [0,T].
\end{aligned}
\tag{6.54}
$$

Finally, the following estimate holds:

$$
\left[\mathbb{E}\sup_{s\in[0,T]} |\nabla_x Y(s;t,x)h|^p\right]^{\frac{1}{p}} + \left[\mathbb{E}\left(\int_0^T |\nabla_x Z(r;t,x)h|^2 dr\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \le c|h|(1+|x|^{(m+1)^2}).
\tag{6.55}
$$

If, in addition, there exists an $L > 0$ *such that*

$$
|\nabla_x F(t,x,y,z)h| \le L|h|, \qquad |\nabla g(x)h| \le L|h|,
$$

for every $t \in [0,T]$, $x, h \in H$, $y \in \mathbb{R}$, $z \in \mathcal{L}_2(\Xi, \mathbb{R})$*, then the following stronger estimate holds:*

$$
\left[\mathbb{E}\sup_{s\in[0,T]} |\nabla_x Y(s;t,x)h|^p\right]^{1/p} + \left[\mathbb{E}\left(\int_0^T |\nabla_x Z(r;t,x)h|^2 dr\right)^{p/2}\right]^{1/p} \le c|h|.
\tag{6.56}
$$

Proof We have already commented on the first two statements. The estimate (6.55) follows from (6.47) applied with

$$
X = X(\cdot;t,x),\ N = \nabla_x X(\cdot;t,x)h,\ \eta = g(X(T;t,x)),\ \zeta = \nabla g(X(T;t,x))\nabla_x X(T;t,x)h,
$$

taking into account that by Propositions 6.9 and 6.10 we have

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

and, by Hypothesis 6.26, we also obtain $|\eta|_{L^r(\Omega)} \le c(1+|x|)^{m+1}$, $|\zeta|_{L^p(\Omega)} \le c|h|(1+|x|)^m$ for a suitable constant c.

The estimate (6.56) is proved in a similar way, applying (6.48) instead of (6.47) and taking into account that under the additional assumption we have $|\eta|_{L^r(\Omega)} \le c(1+|x|)$, $|\zeta|_{L^p(\Omega)} \le c|h|$ for a suitable constant c. □

Proposition 6.28 *Assume Hypotheses 6.8, 6.22 and 6.26. Then the function* $u(t,x) = Y(t,t,x)$ *has the following properties:*

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

(ii) *there exists a* $C > 0$ *such that* $|\nabla_x u(t,x)h| \le C|h|(1 + |x|^{(m+1)^2})$ *for all* $t \in [0,T]$, $x \in H$, $h \in H$;

(iii) *if, in addition,*

$$\sup_{t\in[0,T],x\in H} |F(t,x,0,0)| < \infty, \qquad \sup_{x\in H} |g(x)| < \infty,$$

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

(iv) *similarly, if there exists an* $L > 0$ *such that*

$$|\nabla_x F(t,x,y,z)h| \le L|h|, \qquad |\nabla g(x)h| \le L|h|,$$

for every $t \in [0,T]$, $x, h \in H$, $y \in \mathbb{R}$, $z \in \mathcal{L}_2(\Xi, \mathbb{R})$, *then*

$$|\nabla_x u(t,x)h| \le c|h|$$

for a suitable constant c and all $x, h \in H$.

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

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

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

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

By Proposition 6.27, $\Gamma_1 \in \mathcal{G}^{0,1}$. The inequality

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

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

(ii) is an immediate consequence of the estimate in Proposition 6.27-(iii): indeed,

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

(iii) Since (Y,Z) is a solution to the backward equation, the estimate in Proposition 6.20 yields

$$\sup_{s\in[t,T]} \mathbb{E}\,|Y(s;t,x)|^2 \le c\,\mathbb{E}\int_0^T |F(r, X(r;t,x), 0, 0)|^2 dr + c\,\mathbb{E}\,|g(X(T;t,x))|^2 \le c.$$

(iv) follows immediately from (6.56). □

Corollary 6.29 *For every $t \in [0, T]$, $x \in H$ we have*

$$Y(s; t, x) = u(s, X(s; t, x)), \text{ for } s \in [u, T], \tag{6.57}$$

$$Z(s; t, x) = \nabla_x u(s, X(s; t, x))\, \sigma(s, X(s; t, x)), \text{ for a.a. } s \in [u, T]. \tag{6.58}$$

Proof Setting $s = r$ in the first equality of (6.53) we obtain (6.57).

To prove (6.58) we first write the backward equation in system (6.51) as

$$Y_s = Y_t + \int_t^s Z_r\, dW_r - \int_t^s F(r, X_r, Y_r, Z_r)\, dr, \qquad s \in [t, T]$$

and by (6.57) this can be written

$$u(s, X(s; t, x)) = u(t, x) + \int_t^s Z_r\, dW_r - \int_t^s F(r, X_r, Y_r, Z_r)\, dr, \qquad s \in [t, T]. \tag{6.59}$$

Now we fix an arbitrary $\xi \in \Xi$ and take the joint quadratic variation of both sides of (6.59) with the Wiener process $W\xi$. The joint quadratic variation of the left-hand side is

$$\int_t^s \nabla_x u(r, X(r; t, x))\sigma(r, X(r; t, x))\xi\, dr, \quad s \in [t, T], \tag{6.60}$$

by Proposition 6.17. Since the ordinary integral in (6.59) is a finite variation process, the joint quadratic variation of $W\xi$ and the right-hand side of (6.59) is

$$\int_t^s Z_r \xi\, dr, \quad s \in [t, T]. \tag{6.61}$$

Equating (6.60) and (6.61) we obtain (6.58). □

6.4 BSDEs and Mild Solutions to HJB

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

Let $X(s; t, x)$, $s \in [t, T]$, denote the solution to the stochastic equation

$$X(s) = e^{(s-t)A}x + \int_t^s e^{(s-r)A} b(r, X(r))\, dr + \int_t^s e^{(s-r)A} \sigma(r, X(r))\, dW(r),$$

where A, b, σ, satisfy the assumptions in Hypothesis 6.8. The transition semigroup $P_{t,s}$ is defined for arbitrary $\phi \in \mathcal{B}_p(H)$ and for $0 \le t \le s \le T$ by the formula

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

The estimate $\mathbb{E}\sup_{s\in[t,T]}|X(s;t,x)|^p \leq C(1+|x|)^p$, see (6.8), shows that $P_{t,s}$ is well defined as a linear operator $\mathcal{B}_p(H) \to \mathcal{B}_p(H)$; the semigroup property $P_{t,u}P_{u,s} = P_{t,s}$, $t \leq u \leq s$, is well known.

Let us denote by $\mathcal{A}(t)$ the (formal) generator of $P_{t,s}$:

$$\mathcal{A}(t)[\phi](x) = \frac{1}{2}\mathrm{Tr}\left(\sigma(t,x)\sigma(t,x)^* D^2\phi(x)\right) + \langle Ax + b(t,x), D\phi(x)\rangle\,,$$

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

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

We refer to [179, 180, 582] for a detailed exposition. When g is not regular, the function $v(t,x) = P_{t,T}[g](x)$ can be considered as a generalized solution to the backward Kolmogorov equation.

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

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t} + \mathcal{A}(t)[u(t,\cdot)](x) + F(t,x,u(t,x),\nabla_x u(t,x)\sigma(t,x)) = 0, & t \in [0,T],\ x \in H, \\ u(T,x) = g(x). \end{cases} \tag{6.62}$$

We will refer to this equation as the nonlinear Kolmogorov equation. In the sequel we will be mostly concerned with the case when F is a Hamiltonian function related to an optimal control problem and in this case Eq. (6.62) is the Hamilton–Jacobi–Bellman equation for the corresponding value function. However, the results given in this section are more general, they do not rely on a control-theoretic interpretation and may be of independent interest.

In (6.62) $F : [0,T] \times H \times \mathbb{R} \times \Xi^* \to \mathbb{R}$ is a given function satisfying Hypothesis 6.22. Note that $\nabla_x u(t,x)$, the Gâteaux derivative of $u(t,x)$ with respect to x, is an element of H^*, so that the composition $\nabla_x u(t,x)\sigma(t,x)$ belongs to $\Xi^* = \mathcal{L}(\Xi,\mathbb{R}) = \mathcal{L}_2(\Xi,\mathbb{R})$. Thus, we are in the framework of Hypothesis 6.22 with $K = \mathbb{R}$.

Remark 6.30 A different formulation of Eq. (6.62) is possible, which differs only notationally. We could start with a real-valued function F defined on $[0,T] \times H \times \mathbb{R} \times \Xi$ and write the first equality in (6.62) as

$$\frac{\partial u(t,x)}{\partial t} + \mathcal{A}(t)[u(t,\cdot)](x) + F(t,x,u(t,x),\sigma(t,x)^* D_x u(t,x)) = 0,$$

where $\sigma(t,x)^* \in \mathcal{L}(H,\Xi)$ denotes the Hilbert space adjoint of $\sigma(t,x) \in \mathcal{L}(\Xi,H)$. We recall that D_x denotes the Gâteaux derivative identified with an element of H, so that $\nabla_x u(t,x)h = \langle D_x u(t,x), h\rangle_H$ for every $h \in H$. Of course, identifying Ξ with Ξ^* by the Riesz isometry, one checks immediately the equivalence of the two formulations. ■

Now we define the notion of solution to the nonlinear Kolmogorov equation. We consider the variation of constants formula for (6.62):

$$u(t,x) = \int_t^T P_{t,s}[F(s,\cdot,u(s,\cdot),\nabla_x u(s,\cdot)\sigma(s,\cdot))](x)\,ds + P_{t,T}[g](x), \quad (6.63)$$

for $t \in [0,T]$ and $x \in H$, and we see that formula (6.63) is meaningful, provided $F(t,\cdot,\cdot,\cdot)$, $u(t,\cdot)$ and $\nabla_x u(t,\cdot)$ have polynomial growth (and, of course, provided they satisfy appropriate measurability assumptions). We use this formula as a definition for the solution to (6.62):

Definition 6.31 We say that a function $u : [0,T] \times H \to \mathbb{R}$ is a mild solution to the nonlinear Kolmogorov equation (6.62) if the following conditions hold:

(i) $u \in \mathcal{G}^{0,1}([0,T] \times H, \mathbb{R})$;
(ii) there exist $C > 0$ and $d \in \mathbb{N}$ such that $|\nabla_x u(t,x)h| \le C|h|(1+|x|^d)$ for all $t \in [0,T]$, $x \in H$, $h \in H$;
(iii) equality (6.63) holds.

Note that the specific form of the operator $\mathcal{A}(t)$ plays no role in this definition. We are now ready to state the main result of this section.

Theorem 6.32 *Assume that Hypothesis 6.8 holds, and let F, g be functions satisfying the assumptions in Hypotheses 6.22 (with $K = \mathbb{R}$) and 6.26. Then there exists a unique mild solution to the nonlinear Kolmogorov equation (6.62).*

The solution u is given by the formula

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

where (X,Y,Z) is the solution to the forward–backward system (6.51).

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

Similarly, if $|\nabla_x F|$ is uniformly bounded then $|\nabla_x u|$ is also uniformly bounded.

Proof Existence. By Proposition 6.28, the proposed solution u has the regularity properties stated in Definition 6.31 and the last two statements of the claim hold. It remains to verify that equality (6.63) holds. To this purpose we first fix $t \in [0,T]$ and $x \in H$ and write the backward equation of system (6.51) for $s = t$:

$$Y(t;t,x)+\int_t^T Z(s;t,x)\,dW_s$$
$$=\int_t^T F\Big(s,X(s;t,x),Y(s;t,x),Z(s;t,x)\Big)\,ds+g(X(T;t,x)).$$

Taking the expectation we obtain

$$u(t,x)=\mathbb{E}\int_t^T F\Big(s,X(s;t,x),Y(s;t,x),Z(s;t,x)\Big)\,ds+P_{t,T}[g](x).$$

By (6.57), (6.58) we have

$$u(t,x)=\mathbb{E}\int_t^T F\Big(s,X(s;t,x),u(s,X(s;t,x)),\nabla_x u(s,X(s;t,x))\,\sigma(s,X(s;t,x))\Big)\,ds$$
$$+P_{t,T}[g](x)$$

and equality (6.63) follows.

Uniqueness. Let u be a mild solution. We look for a convenient expression for the process $u(r,X(r;t,x))$, $r\in[t,T]$. By (6.63) and the definition of $P_{r,s}$, for every $r\in[t,T]$ and $x\in H$,

$$u(r,x)=\mathbb{E}\,[g(X(T;r,x))]$$
$$+\mathbb{E}\left[\int_r^T F\Big(s,X(s;r,x),u(s,X(s;r,x)),\nabla_x u(s,X(s;r,x))\,\sigma(s,X(s;r,x))\Big)\,ds\right].$$

Since $X(s;r,x)$ is $\mathscr{F}_r$-independent, we can replace the expectation by the conditional expectation given $\mathscr{F}_r$:

$$u(r,x)=\mathbb{E}^{\mathscr{F}_r}\,[g(X(T;r,x))]$$
$$+\mathbb{E}^{\mathscr{F}_r}\left[\int_r^T F\Big(s,X(s;r,x),u(s,X(s;r,x)),\nabla_x u(s,X(s;r,x))\,\sigma(s,X(s;r,x))\Big)\,ds\right].$$

For the same reason, we can replace x by $X(r;t,x)$ and use the equality: $\mathbb{P}$-a.s.

$$X(s;r,X(r;t,x))=X(s;t,x),\qquad \text{for } s\in[r,T].$$

We arrive at

$$u(r,X(r;t,x))=\mathbb{E}^{\mathscr{F}_r}\,[g(X(T;t,x))]$$
$$+\mathbb{E}^{\mathscr{F}_r}\left[\int_r^T F\Big(s,X(s;t,x),u(s,X(s;t,x)),\nabla_x u(s,X(s;t,x)\sigma(s,X(s;t,x))\Big)\,ds\right]$$
$$=\mathbb{E}^{\mathscr{F}_r}\,[\xi]$$
$$-\int_t^r F\Big(s,X(s;t,x),u(s,X(s;t,x)),\nabla_x u(s,X(s;t,x))\,\sigma(s,X(s;t,x))\Big)ds,$$

where we have defined

$$\xi = g(X(T;t,x)) + \int_t^T F\Big(s, X(s;t,x), u(s, X(s;t,x))\nabla_x u(s, X(s;t,x))\,\sigma(s, X(s;t,x))\Big)\,ds.$$

We note that $\mathbb{E}^{\mathscr{F}_t}[\xi] = u(t,x)$. Since $\xi \in L^2(\Omega;\mathbb{R})$ is $\mathscr{F}_T$-measurable, by the representation theorem recalled in Proposition 6.18, there exists a $\widetilde{Z} \in L^2_{\mathcal{P}}(\Omega \times [t,T]; \mathcal{L}_2(\Xi,\mathbb{R}))$ such that $\mathbb{E}^{\mathscr{F}_r}[\xi] = \int_t^r \widetilde{Z}_s\,dW_s + u(t,x)$. We conclude that the process $u(r, X(r;t,x)), r \in [t,T]$, is a (real) continuous semimartingale with canonical decomposition

$$u(r, X(r;t,x)) = \int_t^r \widetilde{Z}_s\,dW_s + u(t,x) - \int_t^r F\Big(s, X(s;t,x), u(s, X(s;t,x)), \nabla_x u(s, X(s;t,x))\,\sigma(s, X(s;t,x))\Big)\,ds \tag{6.64}$$

into its continuous martingale part and continuous finite variation part. Let $\xi \in \Xi$. By Proposition 6.17, the joint quadratic variation process of $u(r, X(r;t,x))$ and $W(r)\xi$, $r \in [t,T]$, is

$$\int_t^r \nabla_x u(s, X(s;t,x))\sigma(s, X(s;t,x))\xi\,ds, \quad r \in [t,T]. \tag{6.65}$$

Taking into account the canonical decomposition (6.64), we note that the process (6.65) can also be obtained as the joint quadratic variation process between $W(r)\xi$, $r \in [t,T]$, and the process $\int_t^r \widetilde{Z}_s\,dW_s$. This yields the identity

$$\int_t^r \nabla_x u(s, X(s;t,x))\,\sigma(s, X(s;t,x))\xi\,ds = \int_t^s \widetilde{Z}_s\,\xi\,ds. \quad r \in [t,T].$$

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

$$\nabla_x u(s, X(s;t,x))\,\sigma(s, X(s;t,x)) = \widetilde{Z}_s.$$

Substituting into (6.64) we obtain

$$u(r, X(r;t,x)) = \int_t^r \nabla_x u(s, X(s;t,x))\,\sigma(s, X(s;t,x))\,dW_s + u(t,x) + \int_t^r F\Big(s, X(s;t,x), u(s, X(s;t,x)), \nabla_x u(s, X(s;t,x))\,\sigma(s, X(s;t,x))\Big)\,ds,$$

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

$$u(r, X(r; t, x)) + \int_r^T \nabla_x u(s, X(s; t, x))\, \sigma(s, X(s; t, x))\, dW_s = g(X(T; t, x))$$
$$+ \int_r^T F\Big(s, X(s; t, x), u(s, X(s; t, x)), \nabla_x u(s, X(s; t, x))\, \sigma(s, X(s; t, x))\Big)\, ds,$$

for $r \in [t, T]$. Comparing with the backward equation in (6.51) we note that the pairs

$$\Big(Y(r; t, x), Z(r; t, x)\Big) \quad \text{and} \quad \Big(u(r, X(r; t, x)), \nabla_x u(r, X(r; t, x))\, \sigma(r, X(r; t, x))\Big),$$

for $r \in [t, T]$, solve the same equation. By uniqueness, we have in particular $Y(r; t, x) = u(r, X(r; t, x))$, $r \in [t, T]$. Setting $r = t$ we obtain $Y(t; t, x) = u(t, x)$. □

6.5 Applications to Optimal Control Problems

We wish to apply the above results to perform the synthesis of the optimal control for a general nonlinear control system. We will see that this approach allows great generality, particularly with respect to degeneracy of the noise. To be able to use non-smooth feedbacks we settle the problem in the framework of optimal control problems formulated in the extended weak formulation, but we will present results on the extended strong formulation as well.

Let again H, Ξ, denote real separable Hilbert spaces (the state space and the noise space, respectively) and let Λ be a Polish space (the control space). For $t \in [0, T]$ a *generalized reference probability space* is given by $\mu = (\Omega, \mathscr{F}, \mathscr{F}_s^t, \mathbb{P}, W)$, where

- $(\Omega, \mathscr{F}, \mathbb{P})$ is a complete probability space;
- $\{\mathscr{F}_s^t\}_{s \geq t}$ is a filtration in it, satisfying the usual conditions;
- $(W(s))_{s \geq t}$ is a cylindrical $\mathbb{P}$-Wiener process in Ξ, with respect to the filtration $\mathscr{F}_s^t$, starting from $W(t) = 0$.

Given such μ, for every starting point $x \in H$ we will consider the following controlled state equation

$$\begin{cases} dX(s) = \Big(AX(s) + b(s, X(s)) + \sigma(s, X(s)) R(s, X(s), a(s))\Big)\, ds \\ \qquad\qquad\qquad\qquad\qquad\qquad + \sigma(s, X(s))\, dW(s), \;\; s \in [t, T], \\ X(t) = x \in H. \end{cases} \tag{6.66}$$

In (6.66), and below in this section, the equation is understood in the mild sense. $a(\cdot) : \Omega \times [t, T] \to \Lambda$ is the control process, which is always assumed to be progressively measurable with respect to $\{\mathscr{F}_s^t\}_{s \geq t}$. On the coefficients A, b, σ, R precise assumptions will be formulated in Hypothesis 6.33 below. In particular, to allow more generality, on the coefficient R we will only impose measurability and boundedness assumptions, so that, in particular, we cannot guarantee the existence or uniqueness of the solution to the state equation for an arbitrary control process $a(\cdot)$. Therefore the formulations of the control problems require some slight changes with respect to the previous sections and is given as follows (the word *extended* is used to distinguish such formulations, see Remark 2.6). We call $(a(\cdot), X(\cdot))$ an *admissible control pair* if $a(\cdot)$ is an $\mathscr{F}_s^t$-progressively measurable process with values in Λ and $X(\cdot)$ is a mild solution to (6.66) corresponding to $a(\cdot)$. To every admissible control pair we associate the cost:

$$J^\mu(t, x; a(\cdot), X(\cdot)) = \mathbb{E} \int_t^T l(s, X(s), a(s))\, ds + \mathbb{E}\, g(X(T)),$$

where l, g are suitable real functions. The optimal control problem in the extended strong formulation consists in minimizing the functional $J^\mu(t, x; a(\cdot), X(\cdot))$ over all admissible control pairs $(a(\cdot), X(\cdot))$, and characterizing the value function

$$V_t^\mu(x) = \inf_{(a(\cdot), X(\cdot))} J^\mu(t, x; a(\cdot), X(\cdot)).$$

We will also address the optimal control problem in the extended weak formulation, which consists in further minimizing with respect to all generalized reference probability spaces, i.e., in characterizing the value function

$$\overline{V}(t, x) = \inf_\mu V_t^\mu(x).$$

Notice the occurrence of the operator σ in the control term of (6.66): this special structure of the state equation is imposed by our techniques and seems to be essential in different contexts as well (see [298]). The corresponding Hamiltonian function is defined for all $t \in [0, T]$, $x \in H$, $z \in \Xi^*$ setting

$$F_0(t, x, z) = \inf_{a \in \Lambda} \{l(t, x, a) + z\, R(t, x, a)\}. \tag{6.67}$$

Note that this differs from the Hamiltonian as introduced in the previous chapters. In particular, the third argument z ranges over Ξ^* instead of H.

We make the following assumptions:

Hypothesis 6.33 The following holds:

(1) A, b and σ satisfy Hypothesis 6.8.
(2) $R : [0, T] \times H \times \Lambda \to \Xi$ is Borel measurable and $|R(t, x, a)|_\Xi \le L$ for a suitable constant $L > 0$ and all $t \in [0, T]$, $x \in H$, $a \in \Lambda$.
(3) $l : [0, T] \times H \times \Lambda \to \mathbb{R}$ is continuous and $|l(t, x, a)| \le L(1 + |x|^m)$ for suitable constants $L > 0$, $m \ge 0$ and all $t \in [0, T]$, $x \in H$, $a \in \Lambda$.
(4) g satisfies Hypothesis 6.26.
(5) Taking $K = \mathbb{R}$ (and noting that $\mathcal{L}_2(\Xi, \mathbb{R}) = \Xi^*$) the function $F_0 : [0, T] \times H \times \Xi^* \to \mathbb{R}$ satisfies Hypothesis 6.22.

(6) For all $t \in [0, T]$, $x \in H$ and $z \in \Xi^*$ we denote by $\Gamma(t, x, z) \subset \Lambda$ the set of elements $a \in \Lambda$ such that the infimum in (6.67) is attained and we assume that $\Gamma(t, x, z)$ is non-empty. We will denote by γ a measurable selection of Γ, i.e., a measurable function $\gamma : [0, T] \times H \times \Xi^* \to \Lambda$ such that $\gamma(t, x, z) \in \Gamma(t, x, z)$ for every $t \in [0, T]$, $x \in H$and $z \in \Xi^*$. γ is not always assumed to exist.

The Hamilton–Jacobi–Bellman equation relative to the above stated problem is written formally:

$$\begin{cases} \dfrac{\partial v(t,x)}{\partial t} + \mathcal{A}(t)[v(t,\cdot)](x) + F_0(t, x, \nabla_x v(t,x)\sigma(t,x)) = 0, & t \in [0,T],\ x \in H, \\ v(T,x) = g(x). \end{cases} \tag{6.68}$$

Notice the special form of this equation where the nonlinear term depends on $\nabla_x v$ only via the composition $\nabla_x v\,\sigma$: this is consistent with the definition of F_0 given above.

The Hamilton–Jacobi–Bellman equation takes the form of a nonlinear Kolmogorov equation as considered in the previous sections. In particular, under our assumptions, it admits a unique mild solution in the sense specified by Theorem 6.32.

In the proof of our main results, Theorems 6.35 and 6.36 below, we will make use of a classical tool in stochastic analysis, namely the Girsanov Theorem. We recall its statement, in a form suitable for our purposes. Its infinite-dimensional version, which we are about to state, can be found, for example, in [180].

Theorem 6.34 *Let $\mu = (\Omega, \mathscr{F}, \mathscr{F}_s^t, \mathbb{P}, W)$ be a generalized reference probability space, let $R(r)$, $r \in [t, T]$, be an $\mathscr{F}_s^t$-progressively measurable process with values in Ξ such that $\int_t^T |R(r)|_\Xi^2 dr < \infty$ $\mathbb{P}$-a.s., and define*

$$\rho^t(s) = \exp\left(-\int_t^s \langle R(r), dW(r)\rangle_\Xi - \frac{1}{2}\int_t^s |R(r)|_\Xi^2\, dr\right), \qquad s \in [t, T].$$

Then the following holds:

(1) $\rho^t(\cdot)$ is a $\mathbb{P}$-supermartingale;

(2) if

$$\mathbb{E}\,[\rho^t(T)] = 1 \tag{6.69}$$

then $\rho^t(\cdot)$ is a $\mathbb{P}$-martingale and we can define a probability $\widetilde{\mathbb{P}}$ setting $\widetilde{\mathbb{P}}(A) = \mathbb{E}\,[1_A \rho^t(T)]$, $A \in \mathscr{F}$;

(3) the process $\widetilde{W}$ defined by

$$\widetilde{W}(s) = W(s) - W(t) + \int_t^s R(r)\, dr, \qquad s \in [t, T], \tag{6.70}$$

is a cylindrical Wiener process in Ξ with respect to $\mathscr{F}_s^t$ and $\widetilde{\mathbb{P}}$;

(4) finally, if R is bounded in Ξ then (6.69) holds and for every $p \in [1, \infty)$ we have

$$\mathbb{E}\,[(\rho^t(T))^p] < \infty, \qquad \widetilde{\mathbb{E}}\,[(\rho^t(T))^{-p}] < \infty, \tag{6.71}$$

where $\widetilde{\mathbb{E}}$ denotes expectation with respect to $\widetilde{\mathbb{P}}$.

Note that (6.70) does not make sense as it is written since W, being a cylindrical Wiener process, is not a genuine stochastic process taking values in Ξ. (6.70) should be understood as the equality $\widetilde{W}(s)h = W(s)h - W(t)h + \int_t^s \langle R(r), h\rangle_\Xi\, dr$ for any $h \in \Xi$. Nevertheless, in the following we will use a shortened notation as in (6.70).

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

Theorem 6.35 *Assume Hypothesis 6.33 and let $t \in [0, T]$, $x \in H$.*

(1) For all generalized reference probability spaces μ and all admissible control pairs (a, X) we have $J^\mu(t, x; a(\cdot), X(\cdot)) \geq v(t, x)$.
It follows that $V_t^\mu(x) \geq v(t, x)$ for every μ, and so $\overline{V}(t, x) \geq v(t, x)$.

(2) For all μ and all admissible control pairs (a, X), the equality $J^\mu(t, x; a(\cdot), X(\cdot)) = v(t, x)$ holds if and only if the following feedback law is satisfied:

$$a(s) \in \Gamma(s, X(s), \nabla_x v(s, X(s))\,\sigma(s, X(s))), \quad \mathbb{P}\text{-a.s. for a.a. } s \in [t, T]. \tag{6.72}$$

Therefore, (6.72) implies the optimality of an admissible control pair in the extended strong formulation with respect to a given generalized reference probability space μ. If such a control pair exists then $V_t^\mu(x) = v(t, x)$.

Proof For all $\mu = (\Omega, \mathscr{F}, \mathscr{F}_s^t, \mathbb{P}, W)$ and admissible control pairs $(a(\cdot), X(\cdot))$, using the boundedness of R, the Girsanov theorem ensures that there exists a probability measure $\widetilde{\mathbb{P}}$ on Ω such that

$$\widetilde{W}_s := W_s - W_t + \int_t^s R(r, X_r, a(r))\, dr, \qquad s \in [t, T],$$

is a $\widetilde{\mathbb{P}}$-Wiener process (note that $\widetilde{\mathbb{P}}$ and $\widetilde{W}$ depend on (a, X), but we neglect this dependence in the notation). Equation (6.66) can be rewritten as:

$$\begin{cases} dX_s = AX_s\, ds + b(s, X_s)\, ds + \sigma(s, X_s)\, d\widetilde{W}_s, & s \in [t, T], \\ X_t = x \in H, \end{cases} \tag{6.73}$$

which, as usual, is to be understood in the mild sense. The process X turns out to be adapted to the filtration, denoted $(\widetilde{\mathscr{F}}_s^t)_{s\in[t,T]}$, generated by $\widetilde{W}$ and completed in the usual way by means of null sets. In the filtered probability space $(\Omega, \mathcal{F}, \widetilde{\mathscr{F}}_s^t, \widetilde{\mathbb{P}})$ we can consider the system of forward–backward equations on $[t, T]$:

$$\begin{cases} \widetilde{X}(s;t,x) = e^{(s-t)A}x + \int_t^s e^{(s-r)A} b(r, \widetilde{X}(r;t,x))\, dr + \int_t^s e^{(s-r)A}\sigma(r, \widetilde{X}(r;t,x))\, d\widetilde{W}_r, \\ \widetilde{Y}(s;t,x) + \int_s^T \widetilde{Z}(r;t,x) d\widetilde{W}_r = \int_s^T F_0(r, \widetilde{X}(r;t,x), \widetilde{Z}(r;t,x)) dr + g(\widetilde{X}(T;t,x)). \end{cases} \tag{6.74}$$

We notice that $\widetilde{X}(s;t,x) = X_s$. Writing the backward equation in (6.74) for $s = t$ and with respect to the original process W we get:

$$\begin{aligned} &\widetilde{Y}(t;t,x) + \int_t^T \widetilde{Z}(r;t,x)\, dW_r \\ &\quad = \int_t^T \big[F_0(r, X_r, \widetilde{Z}(r;t,x)) - \widetilde{Z}(r;t,x) R(r, X_r, a(r))\big]\, dr + g(X_T). \end{aligned} \tag{6.75}$$

We note that

$$\begin{aligned} \mathbb{E}\left[\left(\int_t^T |\widetilde{Z}(r;t,x)|^2\, dr\right)^{1/2}\right] &= \tilde{\mathbb{E}}\left[(\rho^t(T))^{-1}\left(\int_t^T |\widetilde{Z}(r;t,x)|^2\, dr\right)^{1/2}\right] \\ &\le \left(\tilde{\mathbb{E}}\,[(\rho^t(T))^{-2}]\right)^{1/2}\left(\tilde{\mathbb{E}}\int_t^T |\widetilde{Z}(r;t,x)|^2\, dr\right)^{1/2} < \infty \end{aligned}$$

by (6.71). Therefore, by the Burkholder–Davis–Gundy inequalities, the stochastic integral $\int_t^T \widetilde{Z}(r;t,x)\, dW_r$ has finite $\mathbb{P}$-expectation, equal to zero. Now we recall the equalities (6.57) and (6.58) which imply in the present notation that $\widetilde{Y}(t;t,x) = v(t,x)$ and

$$\widetilde{Z}(s;t,x) = \nabla_x v(s, \widetilde{X}(s;t,x))\, \sigma(s, \widetilde{X}(s;t,x)) = \nabla_x v(s, X_s)\, \sigma(s, X_s).$$

Taking expectation with respect to the original probability $\mathbb{P}$ in (6.75) we obtain:

$$\begin{aligned} \mathbb{E}\, g(X_T) - v(t,x) = &- \mathbb{E}\int_t^T F_0(r, X_r, \nabla_x v(r, X_r)\sigma(r, X_r))\, dr \\ &+ \mathbb{E}\int_t^T \nabla_x v(r, X_r)\sigma(r, X_r) R(r, X_r, a(r))\, dr. \end{aligned}$$

Adding and subtracting $\mathbb{E}\int_t^T l(r, X_r, a(r))\, dr$ we conclude that:

$$\begin{aligned} J^{\mu}(t,x; a(\cdot), X(\cdot)) = v(t,x) + \mathbb{E}\int_t^T \Big[&- F_0(r, X_r, \nabla_x v(r, X_r)\sigma(r, X_r)) \\ &+ \nabla_x v(r, X_r)\sigma(r, X_r) R(r, X_r, a(r)) + l(r, X_r, a(r))\Big]\, dr. \end{aligned} \tag{6.76}$$

The above equality is known as the *fundamental identity*. By the definition of F_0 and Γ it implies immediately that $v(t,x) \le J^{\mu}(t,x;a(\cdot),X(\cdot))$ and that equality holds if and only if (6.72) holds. This proves all the conclusions of the theorem. □

Theorem 6.36 *Assume Hypothesis 6.33, assume in addition that* Γ *admits a measurable selection* γ*, and let* $t \in [0,T]$*,* $x \in H$*. Then there exists at least one generalized reference probability space* $\overline{\mu} = (\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathscr{F}}^t_s, \overline{\mathbb{P}}, \overline{W})$ *and an admissible control pair* $(\overline{a}(\cdot), \overline{X}(\cdot))$ *for which the analogue of (6.72) holds. In particular, it follows that* $V_t^{\overline{\mu}}(x) = v(t,x)$ *and so* $\overline{V}(t,x) = v(t,x)$*. In the space* $\overline{\mu}$ *the process* $\overline{X}$ *is a mild solution to the closed loop equation:*

$$\begin{cases} d\overline{X}(s) = A\overline{X}(s)\,ds + \sigma(s,\overline{X}(s))\,R\Big(s,\overline{X}(s),\ \gamma(s,\overline{X}(s),\nabla_x v(s,\overline{X}(s))\,\sigma(s,\overline{X}(s)))\Big)\,ds \\ \qquad\qquad + b(s,\overline{X}(s))\,ds + \sigma(s,\overline{X}(s))\,d\overline{W}(s), \qquad s \in [t,T], \\ \overline{X}(t) = x \in H, \end{cases} \tag{6.77}$$

the feedback law takes the form

$$\overline{a}(s) = \gamma(s,\overline{X}(s),\nabla_x v(s,\overline{X}(s))\,\sigma(s,\overline{X}(s))), \qquad \overline{\mathbb{P}}\text{-a.s. for a.a. } s \in [t,T],$$

and the pair $(\overline{a}(\cdot), \overline{X}(\cdot))$ *is optimal for the control problem in the extended weak formulation.*

Proof We start by showing the existence of a extended weak solution to Eq. (6.77), again by an application of the Girsanov theorem. We take an arbitrary generalized reference probability space $(\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathscr{F}}^t_s, \mathbb{P}, W)$ and denote by $\overline{X}$ the mild solution on $[t,T]$ of the (uncontrolled) equation

$$\begin{cases} d\overline{X}_s = A\overline{X}_s dt + b(s,\overline{X}_s)ds + \sigma(s,\overline{X}_s)dW_s, \\ \overline{X}_t = x. \end{cases}$$

Recalling the boundedness assumption on R, we see that the Girsanov Theorem provides a probability $\overline{\mathbb{P}}$ on $\overline{\Omega}$ under which the process

$$\overline{W}_s := -\int_t^s R(r,\overline{X}_r,\gamma(r,\overline{X}_r,\nabla_x v(r,\overline{X}_r)\sigma(r,\overline{X}_r)))\,dr + W_s - W_t, \qquad s \in [t,T],$$

is a Wiener process. Then $\overline{X}$ is the mild solution to Eq. (6.77) relative to the generalized reference probability space $\overline{\mu} := (\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathscr{F}}^t_s, \overline{\mathbb{P}}, \overline{W})$. Setting $\overline{a}(s) := \gamma(s,\overline{X}_s,\nabla_x v(s,\overline{X}_s)\,\sigma(s,\overline{X}_s))$, the feedback inclusion (6.72) holds by definition of γ and all the required conclusions follow from Theorem 6.35. □

Remark 6.37 Slight changes in the arguments of Theorem 6.35 allow us to prove an existence result for the control problem in the extended strong formulation, under additional assumptions. More precisely, assume Hypothesis 6.33 and, in addition, that the following holds:

(i) $|\nabla_x F_0(t,x,z)h| \le L|h|$ for a suitable constant L and all $t \in [0,T]$, $x, h \in H$ and $z \in \Xi^*$.
(ii) $\sup_{t\in[0,T],x\in H} |\sigma(t,x)|_{\mathcal{L}(\Xi,H)} < \infty$.
(iii) Γ admits a measurable selection γ; in addition the functions $R(t,\cdot,a) : H \to \Xi$, $\gamma(t,\cdot,\cdot) : H \times \Xi^* \to \Lambda$ and $\nabla_x v(t,\cdot) : H \to H$ are globally Lipschitz, uniformly with respect to $t \in [0,T]$, $a \in \Lambda$ (Lipschitzianity of γ is understood with respect to the metric defined in Λ).

Notice that, by the last statement in Theorem 6.32, (i) implies that $|\nabla_x v|$ is uniformly bounded.

Now, given $t \in [0,T]$ and $x \in H$, fix an *arbitrary* generalized reference probability space $\overline{\mu} = (\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathscr{F}}_s^t, \overline{\mathbb{P}}, \overline{W})$. Then Eq. (6.77) admits a unique mild solution $\overline{X}$, since it has globally Lipschitz coefficients. If we define the control process $\overline{a}(s) = \gamma(s, \overline{X}(s), \nabla_x v(s, \overline{X}(s))\,\sigma(s, \overline{X}(s)))$ we see that the pair $(\overline{a}(\cdot), \overline{X}(\cdot))$ is optimal for the control problem in the extended strong formulation corresponding to $\overline{\mu}$, namely

$$J^{\overline{\mu}}(t,x;\overline{a}(\cdot),\overline{X}(\cdot)) = V_t^{\overline{\mu}}(x).$$

Also note that under the additional assumptions the state equation admits a unique mild solution for an arbitrary control process, so the optimal control problem could also be formulated in a more standard way as in the previous chapters, i.e., as a minimization problem over a class of control processes. ■

6.6 Application: Controlled Stochastic Equation with Delay

In this section we show how the previous results can be applied to perform the synthesis of an optimal control for a stochastic differential equation in $\mathbb{R}^n$ with unit delay:

$$\begin{cases} dx(s) = \left[\int_{-1}^{0} x(s+\theta)\,\alpha(d\theta) + f(s,x(s)) + r(s,x(s),a(s))\right] ds \\ \qquad\qquad + \sigma_0(s,x(s))dW(s), \quad s \in [t,T], \\ x(t) = y, \qquad x(t+\theta) = \beta(\theta), \qquad \text{for } \theta \in (-1,0), \end{cases} \tag{6.78}$$

and a cost functional of the form

$$J^{\mu}(t,y,\beta;a(\cdot),x(\cdot)) = \mathbb{E}\int_t^T h(s,x(s),a(s))\,ds + \mathbb{E}\,k(x(T)).$$

Here $\mu = (\Omega, \mathscr{F}, \mathscr{F}_s^t, \mathbb{P}, W)$ denotes a generalized reference probability space as defined at the beginning of Sect. 6.5 and $(a(\cdot), x(\cdot))$ is an admissible control pair, i.e., the control process $a(\cdot)$ is $\left\{\mathscr{F}_s^t\right\}_{s\ge t}$ progressive with values in $\Lambda \subset \mathbb{R}^N$ and $x(\cdot)$

is a corresponding solution to Eq. (6.78). We will address the optimal control problem in the extended weak formulation, which consists in minimizing the functional $J^{\mu}(t, y, \beta; a(\cdot), x(\cdot))$ over all triples $(\mu, a(\cdot), X(\cdot))$, and characterizing the value function

$$\overline{V}(t, y, \beta) = \inf_{(\mu, a(\cdot), x(\cdot))} J^{\mu}(t, y, \beta; a(\cdot), x(\cdot)).$$

We assume the following (other assumptions are needed and will be stated below):

- $y \in \mathbb{R}^n$, $\beta \in L^2((-1, 0); \mathbb{R}^n)$;
- Λ is a Borel subset of $\mathbb{R}^N$;
- α is an $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$-valued finite measure on $[-1, 0]$;
- $f : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ is measurable, $f(s, \cdot) \in C^1(\mathbb{R}^n)$ and there exists a constant $C > 0$ such that

$$|f(s, 0)| \leq C, \quad |\nabla_x f(s, x)| \leq C, \qquad s \in [0, T], \ x \in \mathbb{R}^n;$$

- $\sigma_0 : [0, T] \times \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is measurable and, for $t \in [0, T], x \in \mathbb{R}^n, \sigma_0(s, x)$ is invertible, we have $\sigma_0(s, \cdot) \in C^1(\mathbb{R}^n)$ and

$$|\sigma_0(s, 0)| \leq C, \qquad |\nabla_x \sigma_0(s, x)| \leq C, \qquad |\sigma_0^{-1}(s, x)| \leq C;$$

- $r : [0, T] \times \mathbb{R}^n \times \Lambda \to \mathbb{R}^n$ is measurable, $r(s, \cdot, a) \in C^1(\mathbb{R}^n)$ and, for some constant $m \geq 0$ and every $s \in [0, T], a \in \Lambda, x \in \mathbb{R}^n$,

$$|r(s, x, a)| \leq C, \qquad |\nabla_x r(s, x, a)| \leq C(1 + |x|)^m.$$

- $h : [0, T] \times \mathbb{R}^n \times \Lambda \to \mathbb{R}$ is continuous, $h(s, \cdot, a) \in C^1(\mathbb{R}^n)$ and, for every $s \in [0, T], a \in \Lambda, x \in \mathbb{R}^n$,

$$|h(s, x, a)| + |\nabla_x h(s, x, a)| \leq C(1 + |x|)^m.$$

- $k : \mathbb{R}^n \to \mathbb{R}$ belongs to $C^1(\mathbb{R}^n)$ and satisfies

$$|\nabla_x k(x)| \leq C(1 + |x|)^m, \qquad x \in \mathbb{R}^n.$$

We set $H = \mathbb{R}^n \times L^2((-1, 0); \mathbb{R}^n)$, $\Xi = \mathbb{R}^n$,

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

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

Then A generates a strongly continuous semigroup in H. Moreover, if we set, for $t \in [0, T]$, $y \in \mathbb{R}^n$, $\beta \in L^2((-1, 0); \mathbb{R}^n)$, $a \in \Lambda$,

$$x = \begin{pmatrix} y \\ \beta \end{pmatrix}, \quad b\left(t, \begin{pmatrix} y \\ \beta \end{pmatrix}\right) = \begin{pmatrix} f(t, y) \\ 0 \end{pmatrix}, \quad \sigma\left(t, \begin{pmatrix} y \\ \beta \end{pmatrix}\right) = \begin{pmatrix} \sigma_0(t, y) \\ 0 \end{pmatrix},$$

$$R\left(t, \begin{pmatrix} y \\ \beta \end{pmatrix}, a\right) = \sigma_0^{-1}(t, y) r(t, y, a),$$

$$l\left(t, \begin{pmatrix} y \\ \beta \end{pmatrix}, a\right) = h(t, y, a), \qquad g\begin{pmatrix} y \\ \beta \end{pmatrix} = k(y),$$

then Eq. (6.78) is reformulated as

$$\begin{cases} dX(s) = \Big(AX(s) + b(s, X(s)) + \sigma(s, X(s)) R(s, X(s), a(s))\Big)\, ds \\ \qquad\qquad\qquad\qquad + \sigma(s, X(s))\, dW(s), \ \ s \in [t, T], \\ X(t) = x. \end{cases}$$

Noting the product form of the state space H, we will write $X(s) = (x(s), x(s + \cdot))$ when we need to distinguish the two components of the solution process. The functional to be minimized can be rewritten as

$$\mathbb{E} \int_t^T l(s, X(s), a(s))\, ds + \mathbb{E}\, g(X(T)).$$

Remark 6.38 We see that the special form of the infinite-dimensional controlled equation (6.66) arises naturally from the finite-dimensional equation (6.78) of general form. ■

Taking into account that Ξ is finite-dimensional, it is easy to check that the assumptions of Hypothesis 6.8 are satisfied. In particular, we may take $\gamma = 0$ in Hypothesis 6.8-(iii).

Next we define, for $s \in [0, T]$, $y \in \mathbb{R}^n$, $\beta \in L^2((-1, 0); \mathbb{R}^n)$, $z \in (\mathbb{R}^n)^*$ (this notation means that z is considered as a row vector),

$$F_0\left(s, \begin{pmatrix} y \\ \beta \end{pmatrix}, z\right) = F_{00}(s, y, z) := \inf_{a \in \Lambda} \left\{h(s, y, a) + z\sigma_0^{-1}(s, y) r(s, y, a)\right\}, \tag{6.79}$$

$$\Gamma\left(s, \begin{pmatrix} y \\ \beta \end{pmatrix}, z\right) = \Gamma_0(s, y, z) := \{a \in \Lambda : F_{00}(s, y, a) = h(s, y, a) + z\sigma_0^{-1}(s, y) r(s, y, a)\}. \tag{6.80}$$

We notice that F_0 and Γ only depend on the finite-dimensional coordinate in H.

The (linear) function $z \to z\sigma_0^{-1}(s, y)r(s, y, a)$ has a Lipschitz constant that only depends on the uniform bounds imposed on r and σ_0^{-1}. It follows that $F_{00}(s, \cdot, a)$ is Lipschitz on $\mathbb{R}^n$ with a Lipschitz constant that does not depend on (s, a).

Moreover, taking into account the growth conditions on the gradients of h, σ, r, it is easy to prove an estimate of the form

$$|\nabla_y \Big[h(s, y, a) + z\sigma_0^{-1}(s, y)r(s, y, a) \Big])| \le C(1 + |z|)(1 + |y|)^m,$$

which implies a local Lipschitz estimate on the function in square parentheses and hence on F_{00}:

$$|F_{00}(s, y, z) - F_{00}(s, y', z)| \le C(1 + |z|)(1 + |y| + |y'|)^m |y - y'|, \tag{6.81}$$

for $s \in [0, T]$, $z \in (\mathbb{R}^n)^*$ and $y, y' \in \mathbb{R}^n$.

To proceed further we also need the following assumptions.

- F_{00} is Borel measurable and, for every $s \in [0, T]$, $F_{00}(s, \cdot, \cdot)$ is of class C^1.
- We assume that $\Gamma_0(s, y, z) \neq \emptyset$ and that there exists a measurable selection γ_0 of Γ_0, i.e., a measurable function $\gamma_0 : [0, T] \times \mathbb{R}^n \times (\mathbb{R}^n)^* \to \Lambda$ such that $\gamma_0(s, y, z) \in \Gamma_0(s, y, z)$ for every $s \in [0, T]$, $y \in \mathbb{R}^n$ and $z \in (\mathbb{R}^n)^*$. It follows that $\gamma(s, (y, \beta), z) := \gamma_0(s, y, z)$, defined on $[0, T] \times H \times (\mathbb{R}^n)^*$, is a measurable selection of Γ.

We note that the local Lipschitz estimate (6.81) implies

$$|\nabla_y F_{00}(s, y, z)| \le C(1 + |z|)(1 + |y|)^m$$

for $s \in [0, T]$, $z \in (\mathbb{R}^n)^*$ and $y \in \mathbb{R}^n$. Now it is easy to see that the conditions required in Hypothesis 6.22 (in the case $K = \mathbb{R}$) are all satisfied by F_0 and that Hypothesis 6.33 holds.

As a consequence of Theorem 6.36 we have the following result.

Theorem 6.39 *Under the previous assumptions there exists at least one generalized reference probability space* $\overline{\mu} = (\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathscr{F}}_s^t, \overline{\mathbb{P}}, \overline{W})$ *and an admissible control pair* $(\overline{a}(\cdot), \overline{x}(\cdot))$ *for which*

$$\overline{V}(t, y, \beta) = J^{\overline{\mu}}(t, y, \beta; \overline{a}(\cdot), \overline{x}(\cdot)), \qquad t \in [0, T], y \in \mathbb{R}^n, \beta \in L^2((-1, 0); \mathbb{R}^n).$$

In particular, the triple $(\overline{\mu}, \overline{a}(\cdot), \overline{x}(\cdot))$ *is optimal.*

The value function $\overline{V}(t, y, \beta) = \overline{V}(t, x)$ *coincides with the function* $v(t, x)$ *which is the unique mild solution to the Hamilton–Jacobi–Bellman equation* (6.68) *in the sense specified by Theorem 6.32.*

In the space $\overline{\mu}$ *the process* $\overline{X}$ *given by* $\overline{X}(s) = (\overline{x}(s), \overline{x}(s + \cdot))$ *is a mild solution to the closed loop equation* (6.77) *and the optimal pair* $(\overline{a}(\cdot), \overline{X}(\cdot))$ *satisfies the feedback law equality*

$$\begin{aligned}\overline{a}(s) &= \gamma(s, \overline{X}(s), \nabla_x v(s, \overline{X}(s))\, \sigma(s, \overline{X}(s))) \\ &= \gamma_0(s, \overline{x}(s), \nabla_\mu \overline{V}(s, \overline{x}(s), \overline{x}(s+\cdot))\, \sigma_0(s, \overline{x}(s))) \quad \mathbb{P}\text{-}a.s. \text{ for a.a. } s \in [t, T].\end{aligned}$$

6.7 Elliptic HJB Equation with Arbitrarily Growing Hamiltonian

In this section we address the solvability of the nonlinear stationary Kolmogorov equation:

$$\mathcal{A}u(x) - \lambda\, u(x) + F(x, u(x), Du(x)\, \sigma) = 0, \qquad x \in H. \tag{6.82}$$

We recall that, formally, the generator $\mathcal{A}$ of (P_t) is the operator

$$\mathcal{A}\phi(x) = \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^* D^2\phi(x)\right) + \langle Ax + b(x), D\phi(x)\rangle\,.$$

Our purpose is to extend the probabilistic techniques and BSDE representation to cover elliptic equations such as (6.82). We consider a general nonlinearity F that will only be assumed to be locally Lipschitz (with arbitrary growth) and no limitations are made on the size of λ. On the other hand, we assume that F is bounded with respect to x and that the noise is additive (that is, σ is independent of x).

We add the following standard piece of notation. If K is a Hilbert space, by $L^p_{\mathcal{P},\mathrm{loc}}(\Omega; L^2([0,\infty); K))$ we denote the space of processes $Y : \Omega \times [0,\infty) \to K$ such that Y restricted to $[0, T]$ is in $L^p_{\mathcal{P}}(\Omega; L^2([0, T]; K))$, $T > 0$.

An analogous definition is given for $L^p_{\mathcal{P},\mathrm{loc}}(\Omega, C([0, +\infty), H))$.

The standing assumptions will be (as far as the linear part of the HJB equation, or, equivalently the forward equation, is concerned):

Hypothesis 6.40 (i) The operator A is the generator of a strongly continuous semigroup e^{tA}, $t \geq 0$, in the Hilbert space H.

(ii) σ does not depend on x (that is, $\sigma \in \mathcal{L}(\Xi, H)$). Moreover, $|e^{tA}\sigma|_{\mathcal{L}_2(\Xi,H)} \leq Lt^{-\gamma}e^{at}$, for a suitable $\gamma \in [0, 1/2)$).

(iii) $b(\cdot) \in \mathcal{G}^1(H, H)$ and $|\nabla b(x)|_{\mathcal{L}(H)} \leq L$.

(iv) The operators $A + \nabla b(x)$ are dissipative (that is, $\langle Ay, y\rangle + \langle \nabla b(x)y, y\rangle \leq 0$ for all $x \in H$ and $y \in D(A)$).

(v) $\lambda > 0$,

and as far as the nonlinear part is concerned:

Hypothesis 6.41 (i) F is locally Lipschitz in z and y, that is, for all $R > 0$ there exists a K_R such that $|F(x, y, z) - F(x, y', z')| \leq K_R(|z - z'| + |y - y'|)$, $\forall x \in H, \forall y, y' \in H, \forall z, z' \in \Xi^*$ with $|z| \leq R$, $|z'| \leq R$, $|y| \leq R$, $|y'| \leq R$.

(ii) The map $x \to F(x, y, z)$ is continuous for all $z \in \Xi^*$, $y \in \mathbb{R}$.

(iii) $\sup_{x\in H} |F(x, 0, 0)| := M < +\infty$.
(iv) $F(\cdot, \cdot, \cdot) \in \mathcal{G}^1(H \times \mathbb{R} \times \Xi^*, \mathbb{R})$ and $|\nabla_x F(x, y, z)|_{H^*} \le c$, for a suitable constant $c > 0$ and all $x \in H$, $y \in \mathbb{R}$, $z \in \Xi^*$.
(v) F is dissipative with respect to y, that is, $\nabla_y F(x, y, z) \le 0$ for all $x \in H$, $y \in \mathbb{R}$, $z \in \Xi^*$

We will also need to add the following Lipschitzianity assumption, which we will eventually remove

Hypothesis 6.42 F is Lipschitz in z and y with constant κ:

$$|F(x, y, z) - F(x, y', z')| \le \kappa(|z - z'| + |y - y'|), \forall x \in H, \forall y, y' \in H, \forall z, z' \in \Xi^*.$$

6.8 The Associated Forward–Backward System

We start from a known result on bounded solutions of Lipschitz BSDEs on an infinite horizon, i.e., the following type of BSDE:

$$Y(\tau) = Y(T) + \int_\tau^T (f(\zeta, Y(\zeta), Z(\zeta)) - \lambda Y(\zeta))d\zeta - \int_\tau^T Z(\zeta)dW(\zeta), \quad 0 \le \tau \le T < \infty, \tag{6.83}$$

where $f : \Omega \times [0, \infty) \times \mathbb{R} \times \Xi^* \to \mathbb{R}$ is such that the process $(f(t, z))_{t\ge 0}$ is progressively measurable for all $z \in \Xi^*$. We suppose the following:

Hypothesis 6.43 (i) f is uniformly Lipschitz in z with Lipschitz constant K:

$$\forall t \ge 0, \forall y \in \mathbb{R}, \forall z, z' \in \Xi^*, \quad |f(t, y, z) - f(t, y, z')| \le K|z - z'|, \quad \mathbb{P}\text{-a.s.}$$

(ii) f is uniformly Lipschitz in y with Lipschitz constant k:

$$\forall t \ge 0, \forall y, y' \in \mathbb{R}, \forall z \in \Xi^*, \quad |f(t, y, z) - f(t, y', z)| \le k|y - y'|, \quad \mathbb{P}\text{-a.s.}$$

(iii) f is dissipative with respect to y that is

$$\forall t \ge 0, \forall y, y' \in \mathbb{R}, \forall z, z' \in \Xi^*, \quad (f(t, y, z) - f(t, y', z))(y - y') \le 0, \quad \mathbb{P}\text{-a.s.}$$

(iv) There exists a constant M such that $\forall t \ge 0$, $|f(t, 0, 0)| \le M$, $\mathbb{P}$-a.s.
We denote $\sup_{t\ge 0} |f(t, 0, 0)|$ by M.

We now turn to the existence and uniqueness of solution to (6.83) under Hypothesis 6.43.

Lemma 6.44 *Let us suppose that Hypothesis 6.43 holds. Then we have:*

(i) There exists a solution (Y, Z) to the BSDE (6.83) such that Y is a continuous process bounded by $\frac{M}{\lambda}$, and $Z \in L^2_{\mathcal{P},\mathrm{loc}}(\Omega; L^2([0, \infty); \Xi))$ with $\mathbb{E}\int_0^\infty e^{-2\varepsilon s}$

$|Z_s|^2 ds < \infty$ for all $\varepsilon > 0$. Moreover, the solution is unique in the class of processes (Y, Z) such that Y is continuous and uniformly bounded, and Z belongs to $L^2_{\mathcal{P},\text{loc}}(\Omega; L^2([0,\infty); \Xi))$.

(ii) Denoting by (Y^n, Z^n) the unique solution to the following finite horizon BSDE:

$$Y^n(\tau) = \int_\tau^n (f(\zeta, Y^n(\zeta), Z^n(\zeta)) - \lambda Y^n(\zeta))d\zeta - \int_\tau^n Z^n(\zeta)dW(\zeta), \quad t \in [0, T], \tag{6.84}$$

we have $|Y^n(\tau)| \leq \frac{M}{\lambda}$ and the following convergence rate holds:

$$|Y^n(\tau) - Y(\tau)| \leq \frac{M}{\lambda} \exp\left(-\lambda(n-\tau)\right). \tag{6.85}$$

Moreover, $\forall \varepsilon > 0$

$$\mathbb{E}\int_0^{+\infty} e^{-2\varepsilon\zeta}|Z^n(\zeta) - Z(\zeta)|^2 d\zeta \to 0. \tag{6.86}$$

Proof The result is contained in [79] and, under more general assumptions, in [518]. For the reader's convenience we report the proof here.

We start from a priori estimates. Fixing T, suppose that (Y, Z) with $Y \in L^2_{\mathcal{P}}(\Omega; C([0,T], K))$ and $Z \in L^2_{\mathcal{P},}(\Omega; L^2([0,T]; \Xi))$ satisfy

$$Y(\tau) = Y(T) + \int_\tau^T (f(\zeta, Y(\zeta), Z(\zeta)) - \lambda Y(\zeta))d\zeta - \int_\tau^T Z(\zeta)dW(\zeta), \quad 0 \leq \tau \leq T. \tag{6.87}$$

Applying Itô's rule to $e^{-\lambda(s-t)}Y_s$, $s \geq t$, we get

$$-d_s\left(e^{-\lambda(s-t)}Y_s\right) = e^{-\lambda(s-t)}f(s, Y_s, 0)ds - e^{-\lambda(s-t)}Z_s(-\theta_s ds + dW(s)),$$

where

$$\theta_s = [f(s, Y_s, Z_s) - f(s, Y_s, 0)]\,|Z_s|^{-2}Z_s^*$$

is a bounded process. Thus by Girsanov's Theorem there exists a probability $\tilde{\mathbb{P}}$ (mean value $\tilde{\mathbb{E}}$) under which $\tilde{W}(t) = -\int_t^s \theta_r dr + W(s)$ is an Ξ-valued Wiener process.

With respect to $(\tilde{W}(t))$ the above equation reads:

$$-d_s\left(e^{-\lambda(s-t)}Y_s\right) = e^{-\lambda(s-t)}f(s, Y_s, 0)ds - e^{-\lambda(s-t)}Z_s d\tilde{W}(s).$$

So applying Itô's rule to $\left(\varepsilon + e^{-2\lambda(s-t)}|Y_s|^s\right)^{1/2} := \mathcal{Y}_s$, $s \geq t$, we obtain

$$\begin{aligned} d_s\mathcal{Y}_s = {}& \mathcal{Y}_s e^{-2\lambda(s-t)}\left[-\langle Y_s, f(s,0,0)\rangle - \langle Y_s, f(s, Y_s, 0) - f(s,0,0)\rangle\right]ds \\ & + \mathcal{Y}_s e^{-2\lambda(s-t)}\langle Y_s, Z_s\rangle\, d\tilde{W}(s) \\ & + \frac{1}{2}\mathcal{Y}_s e^{-2\lambda(s-t)}\left[|Z_s|^2 - \mathcal{Y}_s^{-2}e^{-2\lambda(s-t)}\langle Y_s, Z_s\rangle^2\right]. \end{aligned} \tag{6.88}$$

Taking into account the dissipativity of f with respect to Y and the fact that, by construction, $\mathcal{Y}_s^{-1} e^{-\lambda(s-t)} |Y_s| \leq 1$ we obtain, integrating in $[t, T]$ and then computing the conditional expectation with respect to $\tilde{\mathbb{P}}$:

$$\sqrt{|Y_t|^2 + \varepsilon} \leq \tilde{\mathbb{E}} \left(\sqrt{e^{-2\lambda(T-t)} |Y_T|^2 + \varepsilon} \,\Big|\, \mathscr{F}_t \right) + \tilde{\mathbb{E}} \left(\int_t^T e^{-\lambda(s-t)} |f(s, 0, 0)| ds \,\Big|\, \mathscr{F}_t \right)$$

and by dominated convergence, recalling that $|f(s, 0, 0)| \leq M$:

$$|Y_t| \leq e^{-\lambda(T-t)} \tilde{\mathbb{E}} \left(|Y_T| \,\Big|\, \mathscr{F}_t \right) + M/\lambda.$$

In particular, if (Y^n, Z^n) is a solution to (6.84) then $|Y_t^n| \leq M/\lambda$ for all $t \leq n$.

Moreover, if (Y, Z) is a solution in the whole $[0, \infty)$ with $Z \in L^2_{\mathcal{P},\mathrm{loc}}(\Omega; L^2([0, \infty); \Xi))$ and Y bounded then, letting $T \to \infty$, we get again: $|Y_t| \leq M/\lambda$.

If now $(Y^{(i)}, Z^{(i)})$, $i = 1, 2$, with $Y^{(i)} \in L^2_{\mathcal{P}}(\Omega; C([0, T], K))$ and $Z^{(i)} \in L^2_{\mathcal{P},}(\Omega; L^2([0, T]; \Xi))$ are both solutions to Eq. (6.87) then, by the above computations, applied this time to $(Y_t^{(2)} - Y_t^{(1)}, Z^{(2)} - Z^{(1)})$ we get

$$|Y_T^{(2)} - Y_T^{(1)}| \leq e^{-\lambda(T-t)} \tilde{\mathbb{E}} \left(|Y_T^{(2)} - Y_T^{(1)}| \,\Big|\, \mathscr{F}_t \right), \quad \forall t \in [0, T].$$

Consequently, if $m > n$ and (Y^n, Z^n) and (Y^m, Z^m) satisfy Eq. (6.84) then

$$|Y_t^n - Y_t^m| \leq e^{-\lambda(n-t)} \tilde{\mathbb{E}} \left(|Y_n^m| \,\Big|\, \mathscr{F}_t \right) \leq e^{-\lambda(n-t)} M/\lambda \quad \forall t \in [0, T]. \tag{6.89}$$

In the same way, if (Y, Z) is a solution of (6.83) on the whole $[0, \infty)$ with $Z \in L^2_{\mathcal{P},\mathrm{loc}}(\Omega; L^2([0, \infty); \Xi))$ and we know that (Y_s) is bounded, we get

$$|Y_t - Y_t^n| \leq e^{-\lambda(n-t)} M/\lambda.$$

We notice that the above relation immediately yields that if $(Y^{(i)}, Z^{(i)})$, $i = 1, 2$, are both solutions to Eq. (6.83) on the whole $[0, \infty)$ and we a priori know that both $(Y_t^{(1)})$ and $(Y_t^{(2)})$ are bounded, then $Y_t^{(1)} = Y_t^{(2)}$, $\mathbb{P}$-a.s., for all $t \in [0, T)$.

Concerning the estimate of the Z term we again fix T. If (Y, Z) with $Y \in L^2_{\mathcal{P}}(\Omega; C([0, T], K))$ and $Z \in L^2_{\mathcal{P},}(\Omega; L^2([0, T]; \Xi))$ satisfy (6.87) then, applying Itô's rule to $e^{-2\varepsilon}|Y_s|^2$, (with $0 < \varepsilon < \lambda$) and integrating between 0 and T, we get:

$$\int_0^T e^{-2\varepsilon s}|Z_s|^2 ds + |Y_0|^2 = 2e^{-2\varepsilon T}|Y_T|^2 + 2\int_0^T e^{-2\varepsilon s}\left[\langle f(s, Y_s, Z_s), Y_s\rangle - (\lambda - \varepsilon)|Y_s|^2\right] ds - \int_0^T e^{-2\varepsilon s}\langle Y_s, Z_s dW(s)\rangle .$$

Since $\mathbb{E}\left(\int_0^T e^{-2\varepsilon s}\langle Y_s, Z_s\rangle^2 ds\right)^{1/2} \le \mathbb{E}\left[(\sup_{t\in[0,T]}|Y_t|)\left(\int_0^T |Z_s|^2 ds\right)^{1/2}\right] < \infty$, the stochastic integral in the above formula is a martingale. Thus, computing the expectation, taking into account the Lipschitzianity of f with respect to Z and its dissipativity with respect to Y, we get

$$\mathbb{E}\int_0^T e^{-2\varepsilon s}|Z_s|^2 ds \le ce^{-2\varepsilon T}\mathbb{E}|Y_T|^2 + c\mathbb{E}\int_0^T e^{-2\varepsilon s}|Y_s|^2 ds + c\mathbb{E}\int_0^T e^{-2\varepsilon s}|f(s, 0, 0)|^2 ds,$$

where c is a constant depending only on f and ε.

In particular, if (Y, Z) is a solution on the whole $[0, \infty)$ with $Z \in L^2_{\mathcal{P},\mathrm{loc}}(\Omega; L^2([0, \infty); \Xi))$ and Y is bounded, then:

$$\mathbb{E}\int_0^\infty e^{-2\varepsilon s}|Z_s|^2 ds < +\infty.$$

Similarly, if $(Y^{(i)}, Z^{(i)})$, $i = 1, 2$, with $Y^{(i)} \in L^2_{\mathcal{P}}(\Omega; C([0, T], K))$ and $Z^{(i)} \in L^2_{\mathcal{P},}(\Omega; L^2([0, T]; \Xi))$ are solutions to Eq. (6.87), then:

$$\int_0^T e^{-2\varepsilon s}|Z_s^{(2)} - Z_s^{(1)}|^2 ds \le ce^{-2\varepsilon T}\mathbb{E}|Y_T^{(2)} - Y_T^{(1)}|^2 + c\mathbb{E}\int_0^T e^{-2\varepsilon s}|Y_s^{(2)} - Y_s^{(1)}|^2 ds.$$

In particular, if $m > n$ and (Y^n, Z^n) and (Y^m, Z^m) satisfy Eq. (6.84) then, exploiting the estimates on Y^n and Y^m, we get, for all $T < n$

$$\mathbb{E}\int_0^T e^{-2\varepsilon t}|Z_t^n - Z_t^m| \le ce^{-\lambda(n-T)},$$

and if (Y, Z) is a solution on the whole $[0, \infty)$ with Y bounded, then

$$\mathbb{E}\int_0^T e^{-2\varepsilon t}|Z_t^n - Z_t| \le ce^{-\lambda(n-T)}. \tag{6.90}$$

Thus we have proved that, if a solution of Eq. (6.83) with (Y) bounded on the whole $[0, +\infty)$ exists, then it is unique and it satisfies estimates (6.85) and (6.86).

We now need to prove the existence of a bounded solution. By (6.89), fixing an arbitrary $T > 0$, the sequence of continuous functions $[0, T] \ni t \to Y_t^n$ is, $\mathbb{P}$ almost surely, a Cauchy sequence in $C([0, T])$. Thus there exists an adapted process with continuous trajectories such that, for any $T > 0$:

$$\sup_{t \in [0,T]} |Y_t^n - Y_t| \to 0, \qquad \mathbb{P}\text{-a.s.}$$

Notice that $|Y_t| \le M/\lambda$.

Moreover, by (6.90), for any $T > 0$, the sequence (Z^n) is Cauchy in $L^2_{\mathcal{P}}(\Omega; L^2([0, T]; \Xi))$, so there exists a $Z \in L^2_{\mathcal{P},\text{loc}}(\Omega; L^2([0, \infty); \Xi))$ such that

$$\mathbb{E} \int_0^T |Z_t - Z_t^n|^2 dt \to 0.$$

To prove that (Y, Z) is the desired solution to Eq. (6.83) it is enough to observe that, for any fixed $0 < t < T < n$, we have

$$Y^n(\tau) = Y^n(T) + \int_\tau^T (f(\zeta, Y^n(\zeta), Z^n(\zeta)) - \lambda Y^n(\zeta))d\zeta - \int_\tau^T Z^n(\zeta)dW_\zeta.$$

The claim then follows just by letting $n \to \infty$ in the above formula. □

Now we come to the actual (Markovian) forward backward system. As far as the forward equation is concerned we consider the following special case of (6.6):

$$X(s; x) = e^{sA}x + \int_0^s e^{(s-\zeta)A} b(X(\zeta; x))d\zeta + \int_0^s e^{(s-\zeta)A} \sigma dW(\zeta), \ s \ge 0. \quad (6.91)$$

We know that for every $p \in [2, \infty)$ and $T > 0$ there exists a unique process $X(\cdot; x) \in L^p_{\mathcal{P}}(\Omega; C([0, T], H))$ which is a solution to (6.91). Moreover, for all fixed $T > 0$, the map $x \to X(\cdot; x)$ is continuous from H to $L^p_{\mathcal{P}}(\Omega; C([0, T], H))$.

$$\mathbb{E} \sup_{\tau \in [0,T]} |X(\tau; x)|^p \le C(1 + |x|)^p, \quad (6.92)$$

for some constant C depending only on T and pm.

We then consider the infinite horizon BSDE under the extra assumption (which will be removed later) that F is Lipschitz with respect to z. Namely, we deal with the equation (for $0 \le \tau \le T < \infty$)

$$\begin{aligned} Y(\tau; x) = Y(T; x) &+ \int_\tau^T (F(X(\zeta; x), Y(\zeta; x), Z(\zeta; x)) - \lambda Y(\zeta; x))d\zeta \\ &- \int_\tau^T Z(\zeta; x)dW(\zeta). \end{aligned} \quad (6.93)$$

Here $X(\cdot; x)$ is the unique mild solution to (6.91) starting with $X(0; x) = x$.

Applying Lemma 6.44, we obtain:

Proposition 6.45 *Let us suppose that Hypotheses 6.40–6.42 hold. Then we have:*

(i) *For any $x \in H$, there exists a solution $(Y(\cdot;x), Z(\cdot;x))$ to the BSDE (6.93) such that $Y(\cdot;x)$ is a continuous process bounded by M/λ, and $Z \in L^2_{\mathcal{P},\mathrm{loc}}(\Omega; L^2([0,\infty);\Xi))$ with $\mathbb{E}\int_0^\infty e^{-2\lambda s}|Z(s;x)|^2 ds < \infty$. The solution is unique in the class of processes (Y, Z) such that Y is continuous and bounded, and Z belongs to $L^2_{\mathcal{P},\mathrm{loc}}(\Omega; L^2([0,\infty);\Xi))$.*

(ii) *Denoting by $(Y^n(\cdot;x), Z^n(\cdot;x))$ the unique solution of the following BSDE (with finite horizon):*

$$Y^n(\tau;x) = \int_\tau^n (F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x)) - \lambda Y^n(\zeta;x))d\zeta - \int_\tau^n Z^n(\zeta;x)dW(\zeta), \tag{6.94}$$

we have $|Y^n(\zeta;x)| \le \frac{M}{\lambda}$ and the following convergence rate holds:

$$|Y^n(\tau;x) - Y(\tau;x)| \le \frac{M}{\lambda}\exp\left(-\lambda(n-\tau)\right). \tag{6.95}$$

Moreover,

$$\mathbb{E}\int_0^{+\infty} e^{-2\lambda\zeta}|Z^n(\zeta;x) - Z(\zeta;x)|^2 d\zeta \to 0. \tag{6.96}$$

(iii) *For all $T > 0$ and $p \ge 1$, the map $x \to (Y(\cdot;x)\big|_{[0,T]}, Z(\cdot;x)\big|_{[0,T]})$ is continuous from H to the space $L^p_{\mathcal{P}}(\Omega; C([0,T],\mathbb{R})) \times L^p_{\mathcal{P}}(\Omega; L^2([0,T];\Xi))$.*

Proof Statements (i) and (ii) are immediate consequences of Lemma 6.44. Let us prove (iii). If $x'_m \to x$ as $m \to +\infty$ then

$$\begin{aligned}|Y(T;x'_m) - Y(T;x)| &\le |Y(T;x'_m) - Y^n(T;x'_m)| + |Y^n(T;x'_m) - Y^n(T;x)| \\ &\quad + |Y^n(T;x) - Y(T;x)| \\ &\le 2\frac{M}{\lambda}\exp\left(-\lambda(n-T)\right) + |Y^n(T;x'_m) - Y^n(T;x)|.\end{aligned}$$

Moreover, for fixed n, $Y^n(\cdot;x'_m) \to Y^n(\cdot;x)$ in $L^p(\Omega,\mathscr{F}_T,\mathbb{P};\mathbb{R})$ (see Proposition 6.27) and notice that we are now dealing with a finite horizon BSDE. Thus $Y(T;x'_m) \to Y(T;x)$ in $L^p(\Omega,\mathscr{F}_T,\mathbb{P};\mathbb{R})$.

Now we can see that $(Y(\cdot;x)\big|_{[0,T]}, Z(\cdot;x)\big|_{[0,T]})$ is the unique solution of the following BSDE (with finite horizon):

$$Y(\tau; x) = Y(T; x) + \int_\tau^T (F(X(\zeta; x), Y(\zeta; x), Z(\zeta; x)) - \lambda Y(\zeta; x))d\zeta$$
$$- \int_\tau^T Z(\zeta; x)dW(\zeta),$$

and the same holds for $(Y(\cdot; x'_m)\big|_{[0,T]}, Z(\cdot; x'_m)\big|_{[0,T]})$. So it is enough to apply again the continuity result in Proposition 6.27 to conclude that $(Y(\cdot; x'_m)\big|_{[0,T]}, Z(\cdot; x'_m)\big|_{[0,T]})$ converges to $(Y(\cdot; x)\big|_{[0,T]}, Z(\cdot; x)\big|_{[0,T]})$ in $L^p_{\mathcal{P}}(\Omega; C([0, T], \mathbb{R})) \times L^p_{\mathcal{P}}(\Omega; L^2([0, T]; \Xi))$. □

Remark 6.46 We stress the fact that the uniform bound of Y does not depend on the Lipschitz constant κ of F with respect to y and z (provided that F is dissipative with respect to y). ■

6.8.1 Differentiability of the BSDE and a Priori Estimate on the Gradient

We need to study the regularity of $Y(\cdot, x)$. More precisely, we would like to show that $Y(0, x)$ belongs to $\mathcal{G}^1(H, \mathbb{R})$. Moreover, we will obtain a crucial a priori bound on the derivative $\nabla Y(0; x)$ independent of the Lipschitz constant of F with respect to z.

Lemma 6.47 *Under Hypothesis 6.40 the map $x \to X(\cdot, x)$ is Gâteaux differentiable (that is, it belongs to $\mathcal{G}(H, L^p_{\mathcal{P}}(\Omega, C([0, T], H)))$). Moreover, denoting by $\nabla X(\cdot, x)$ the partial Gâteaux derivative, for every direction $h \in H$, the directional derivative process $\nabla X(\cdot, x)h$, $\tau \in \mathbb{R}$, solves, $\mathbb{P}$-a.s., the equation*

$$\nabla X(\tau; x)h = e^{\tau A}h + \int_0^\tau e^{\zeta A}\nabla b(X(\zeta; x))\nabla X(\zeta; x)h\, d\zeta, \quad \tau \in \mathbb{R}^+. \tag{6.97}$$

Finally, $\mathbb{P}$-a.s., $|\nabla X(\tau; x)h| \leq |h|$, for all $\tau > 0$.

Proof The first assertion and relation (6.97) is a special case of Proposition 6.10. To prove the last assertion we proceed by a classical approximation argument (notice that the equation for ∇X has no stochastic integral term). Let $J_n := n(nI - A)^{-1}$ be the Yosida approximation for n large enough. As is well known (see also Appendix B.4.2) $J_n \in \mathcal{L}(H, D(A))$, $J_n x \to x$ for all $x \in H$. Let $L^n_t = J_n\nabla X(t; x)h$, then, for all $T > 0$, $L^n \in L^p_{\mathcal{P}}(\Omega; C([0, T], D(A)))$ and satisfies

$$(L^n_t)' = AL^n_t + J_n\nabla b(X(t; x))\nabla X(t; x)h.$$

Computing $\frac{d}{dt}|L_t^n|^2$, by Hypothesis 6.40 (iv) we get:

$$\frac{d}{dt}|L_t^n|^2 \le 2\langle L_t^n, \big(J_n\nabla b(X_t^x)\nabla X(t;x)h - \nabla b(X(t;x))J_n\nabla_x X(t;x)h\big)\rangle$$

and

$$|L_t^n|^2 \le |J_n h|^2 + \\ +2\int_0^t \langle L_s^n, (J_n\nabla b(X(s;x))\nabla X(s;x)h - \nabla b(X(s;x))J_n\nabla_x X(s;x)h)\rangle ds$$

and the claim follows by passing to the limit as $n \to \infty$. □

The following is the main technical result of this section.

Theorem 6.48 *Under Hypotheses 6.40–6.42 the map $x \to Y(0;x)$ belongs to $\mathcal{G}^1(H,\mathbb{R})$. Moreover, $|Y(0;x)| + |\nabla Y(0;x)| \le c$, for a suitable constant c. We notice that the constant c does not depend on the Lipschitz constant κ of F with respect to y and z*

Proof The uniform bound on $|Y(0;x)|$ is an immediate consequence of Proposition 6.45.

Coming now to differentiability, fix $n \ge 1$, and let us consider the solution $(Y^n(\cdot;x), Z^n(\cdot;x))$ of (6.94). Then, see Proposition 6.27, the map $x \to (Y^n(\cdot;x), Z^n(\cdot;x))$ is Gâteaux differentiable from H to $L^p_{\mathcal{P}}(\Omega; C([0,T],\mathbb{R})) \times L^p_{\mathcal{P}}(\Omega; L^2([0,T];\Xi^*))$, $\forall p \in [2,\infty)$. Denoting by $\nabla Y^n(\cdot;x)h, \nabla Z^n(\cdot;x)h$ the partial Gâteaux derivatives with respect to x in the direction $h \in H$, the processes

$$\{\nabla Y^n(\tau;x)h\}_{\tau\in[0,n]}, \qquad \{\nabla Z^n(\tau;x)h\}_{\tau\in[0,n]}$$

solve the following equation, $\mathbb{P}$-a.s.,

$$\begin{aligned}\nabla Y^n(\tau;x)h = &\int_\tau^n \nabla_x F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x))\nabla X(\zeta;x)h\,d\zeta \\ &+\int_\tau^n (-\lambda + \nabla_y F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x)))\nabla Y^n(\zeta;x)h\,d\zeta \\ &+\int_\tau^n \nabla_z F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x)\nabla Z^n(\zeta;x)h\,d\zeta \\ &-\int_\tau^n \nabla Z^n(\zeta;x)h\,dW(\zeta).\end{aligned} \tag{6.98}$$

We see that in the above formula, we are considering that $Z^n(\cdot;x)$, $\nabla Z(\cdot;x)$ have values in Ξ^* and $\nabla_z F$ has values in Ξ^{**}. So if we identify Ξ^{**} and Ξ we can assume that $\nabla_z F$ has values in Ξ and Eq. (6.98) can be rewritten as:

$$\nabla Y^n(\tau;x)h = \int_\tau^n \nabla_x F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x))\nabla X(\zeta;x)h\, d\zeta$$
$$+ \int_\tau^n (-\lambda + \nabla_y F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x)))\nabla Y^n(\zeta;x)h\, d\zeta$$
$$+ \int_\tau^n (\nabla Z^n(\zeta;x)h)\left(\nabla_z F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x))\, d\zeta - dW(\zeta)\right).$$

By Hypotheses 6.41 and Lemma 6.47, we have that for all $x, h \in H$ the following holds $\mathbb{P}$-a.s. for all $n \in \mathbb{N}$ and all $\zeta \in [0, n]$:

$$\left|\nabla_x F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x))\nabla_x X(\zeta;x)h\right| \le c|h|,$$
$$\nabla_y F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x)) \le 0, \qquad \left|\nabla_z F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x))\right|_\Xi \le \hat{c}.$$

Therefore, by Lemma 6.44, we obtain:

$$\sup_{\tau\in[0,n]} |\nabla Y^n(\tau;x)| \le C|h|, \quad \mathbb{P}\text{-a.s.}, \tag{6.99}$$

where C does not depend on $\hat{c}$. Applying Itô's formula to $e^{-2\lambda t}|\nabla Y^n(\cdot;x)_t h|^2$, we get:

$$\mathbb{E}\int_0^\infty e^{-2\lambda t}(|\nabla Y^n(t;x)h|^2 + |\nabla_x Z^n(t;x)h|^2)dt \le C|h|^2. \tag{6.100}$$

Let now $\mathcal{M}^{2,-2\lambda}$ be the Hilbert space of all pairs of $\{\mathscr{F}_t\}_{t\ge 0}$-adapted and measurable processes (y, z), where y has values in $\mathbb{R}$ and z in Ξ^*, such that

$$|(y,z)|^2_{\mathcal{M}^{2,-2\lambda}} := \mathbb{E}\int_0^\infty e^{-2\lambda t}(|y_t|^2 + |z_t|^2)dt < +\infty.$$

Fix $x, h \in H$, then there exists a subsequence of $\Big(\nabla Y^n(\cdot;x)h, \nabla Z^n(\cdot;x)h, \nabla Y^n(0;x)h\Big)_{n\in\mathbb{N}}$ which we still denote by itself, such that $(\nabla_x Y^n(\cdot;x)h, \nabla Z^n(\cdot;x)h)$ converges weakly to $(U^1(\cdot;x,h), V^1(\cdot;x,h))$ in $\mathcal{M}^{2,-2\lambda}$ and $\nabla_x Y^n(0;x)h$ converges to $\xi(x,h) \in \mathbb{R}$.

We define now

$$\begin{aligned} U^2(\tau;x,h) = \xi(x,h) &- \int_0^\tau \nabla_x F(X(\zeta;x), Y(\zeta;x), Z(\zeta;x))\nabla X(\zeta;x)\, hd\zeta \\ &- \int_0^\tau (-\lambda + \nabla_y F(X(\zeta;x), Y(\zeta;x), Z(\zeta;x))U^1(\zeta;x,h)d\zeta \\ &- \int_0^\tau \nabla_z F(X(\zeta;x), Y(\zeta;x), Z(\zeta;x))V^1(\zeta;x,h)d\zeta \\ &+ \int_0^\tau V^1(\zeta;x,h)dW(\zeta), \end{aligned} \tag{6.101}$$

where $(Y(\cdot;x), Z(\cdot;x))$ is the unique bounded solution to the backward equation (6.93), see Proposition 6.45. Moreover, we rewrite (6.98) as follows:

$$\begin{aligned}\nabla Y^n(\tau;x)h = \nabla Y^n(0;x)h &- \int_0^\tau \nabla_x F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x))\nabla X(\zeta;x)hd\zeta \\ &+ \int_0^\tau (\lambda - \nabla_y F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x)))\nabla Y^n(\zeta;x)hd\zeta \\ &- \int_0^\tau \nabla_z F(X(\zeta;x), Y^n(\zeta;x), Z^n(\zeta;x))\nabla Z^n(\zeta;x)hd\zeta \\ &+ \int_0^\tau \nabla Z^n(\zeta;x)hdW(\zeta).\end{aligned} \tag{6.102}$$

Since, in particular, $(Y^n(\cdot;x), Z^n(\cdot;x)) \to (Y(\cdot;x), Z(\cdot;x))$ in measure $\mathbb{P} \times dt$; $\nabla_x F$, $\nabla_y F$, $\nabla_z F$ are bounded and finally $(\nabla Y^n(\cdot;x)h, \nabla Z^n(\cdot;x)h) \rightharpoonup (Y(\cdot;x), Z(\cdot;x))$ weakly in $\mathcal{M}^{2,-2\lambda}$, it is easy to show that $\nabla Y^n(\cdot;x)h$ converges to $U^2(\cdot;x,h)$ weakly in $L^2_{\mathcal{P}}(\Omega \times [0,T];\mathbb{R})$ for all $T > 0$. Thus $U^2(t;x,h) = U^1(t;x,h)$, $\mathbb{P}$-a.s. for a.e. $t \in \mathbb{R}^+$ and $|U^2(t;x,h)| \le c|h|$, $\mathbb{P}$-a.s. for all $t \in \mathbb{R}^+$ (this last assertion follows from continuity of the trajectories of $U^2(\cdot;x,h)$ and from the fact that $|U^1(t;x,h)| \le c|h|$ $\mathbb{P}$-a.s. for almost every $t \in \mathbb{R}^+$). Therefore, coming back to Eq. (6.101), we have that $(U^2(\cdot;x,h), V^1(\cdot;x,h))$ is the unique bounded solution in $\mathbb{R}^+$ of the equation

$$\begin{aligned}U(\tau,x,h) = U(0,x,h) &- \int_0^\tau \nabla_x F(X(\zeta;x), Y(\zeta;x), Z(\zeta;x))\nabla X(\zeta;x)hd\zeta \\ &- \int_0^\tau (-\lambda + \nabla_y F(X(\zeta;x), Y(\zeta;x), Z(\zeta;x)))U(\tau,x,h)d\zeta \\ &- \int_0^\tau \nabla_z F(X(\zeta;x), Y(\zeta;x), Z(\zeta;x))V(\zeta,x,h)d\zeta \\ &+ \int_0^\tau V(\zeta,x,h)dW(\zeta).\end{aligned} \tag{6.103}$$

Notice that in particular $U(0,x,h) = \xi(x,h)$ is the limit of $\nabla Y^n(\cdot;x)_0 h$ (along the chosen subsequence). The uniqueness of the solution to (6.103) (see Lemma 6.44) implies that in reality $U(0,x,h) = \lim_{n\to\infty} \nabla Y^n(\cdot;x)_0 h$ along the original sequence.

Now let $x'_m \to x$. By (6.85), proceeding as in the proof of point (iii) in Proposition 6.45,

$$|U(0,x,h) - U(0,x'_m,h)| \le \frac{2c}{\lambda} e^{-\lambda n}|h| + |U_n(0,x,h) - U_n(0,x'_m,h)|, \tag{6.104}$$

where $(U_n(\cdot,x,h), V_n(\cdot,x,h)) \in L^p_{\mathcal{P}}(\Omega; C([0,T],\mathbb{R})) \times L^p_{\mathcal{P}}(\Omega; L^2_{\mathcal{P}}([0,T];\Xi))$ is the unique solution of the finite horizon BSDE:

$$
\begin{aligned}
U_n(\tau, x, h) = &\int_\tau^n \nabla_x F(X(\zeta; x), Y(\zeta; x), Z(\zeta; x)) \nabla X(\zeta; x) h d\zeta \\
&+ \int_\tau^n (-\lambda + \nabla_y F(X(\zeta; x), Y(\zeta; x), Z(\zeta; x))) U_n(\tau, x, h) d\zeta \\
&+ \int_\tau^n \nabla_z F(X(\zeta; x), Y(\zeta; x), Z(\zeta; x)) V_n(\zeta, x, h) d\zeta \\
&- \int_\tau^n V_n(\zeta, x, h) dW(\zeta),
\end{aligned} \tag{6.105}
$$

and similarly for $(U_n(\cdot, x'_m, h), V_n(\cdot, x'_m, h))$. We now see that $\nabla_x F$, $\nabla_y F$, $\nabla_z F$ are, by assumptions, continuous and bounded. Moreover, the following statements on continuous dependence on x hold:

the maps $x \to X^x$, $x \to \nabla X^x h$ are continuous from H to $L^p_{\mathcal{P}}(\Omega; C([0, T], H))$ (see Proposition 6.10);

the map $x \to Y^x\big|_{[0,T]}$ is continuous from H to $L^p_{\mathcal{P}}(\Omega; C([0, T], \mathbb{R}))$ (see Proposition 6.45);

the map $x \to Z^x\big|_{[0,T]}$ is continuous from H to $L^p_{\mathcal{P}}(\Omega; L^2([0, T]; \Xi))$ (see Proposition 6.45).

We can therefore apply to (6.105) the continuous dependence on data result for finite horizon BSDEs (see Proposition 6.20) to obtain in particular that $U_n(0, x'_m, h) \to U_n(0, x, h)$ for all fixed n as $m \to \infty$. And by (6.104) we can conclude that $U(0, x'_m, h) \to U(0, x, h)$ as $m \to \infty$.

Summarizing, $U(0, x, h) = \lim_{n\to\infty} \nabla Y^n(\cdot; x)_0 h$ exists, moreover it is clearly linear in h and satisfies $|U(0, x, h)| \le C|h|$. Finally, it is continuous in x for every fixed h.

Lastly, for $t > 0$,

$$
\begin{aligned}
\lim_{t\searrow 0} \frac{1}{t}[Y(0; x + th) - Y(0; x)] &= \lim_{t\searrow 0} \frac{1}{t} \lim_{n\to+\infty} [Y^n(0; x + th) - Y^n(0; x)] \\
&= \lim_{t\searrow 0} \lim_{n\to+\infty} \int_0^1 \nabla Y^n(0; x + th) h d\theta \\
&= \lim_{t\searrow 0} \int_0^1 U(0, x + \theta th) h d\theta = U(0, x) h
\end{aligned}
$$

and the claim is proved. □

6.9 Mild Solution of the Elliptic PDE

Assuming that Hypothesis 6.40 holds, we define in the usual way the transition semigroup $(P_t)_{t\ge 0}$ associated to the process X:

$$
P_t[\phi](x) = \mathbb{E}\, \phi(X(t; 0, x), \qquad x \in H, \tag{6.106}
$$

for every bounded measurable function $g : H \to \mathbb{R}$. Formally, the generator $\mathcal{A}$ of (P_t) is the operator

$$\mathcal{A}\phi(x) = \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^* D^2\phi(x)\right) + \langle Ax + b(x), D\phi(x)\rangle .$$

In this section we address the solvability of the nonlinear stationary Kolmogorov equation:

$$\mathcal{A}u(x) - \lambda\, u(x) + F(x, u(x), \nabla u(x)\,\sigma) = 0, \qquad x \in H. \tag{6.107}$$

Definition 6.49 We say that a function $u : H \to \mathbb{R}$ is a mild solution of the nonlinear stationary Kolmogorov equation (6.107) if the following conditions hold:

(i) $u \in \mathcal{G}^1(H, \mathbb{R})$ and $\exists\, C > 0$ such that $|u(x)| \le C$, $|\nabla u(x)h| \le C\,|h|$, for all $x, h \in H$;
(ii) the following equality holds, for every $x \in H$ and $T \ge 0$:

$$u(x) = e^{-\lambda T}\ P_T[u](x) + \int_0^T e^{-\lambda\tau}\ P_\tau\Big[F\Big(\cdot, u(\cdot), \nabla u(\cdot)\,\sigma\Big)\Big](x)\ d\tau. \tag{6.108}$$

Remark 6.50 In order to motivate this definition one may consider the equation $\mathcal{A}u - \lambda u = -F$, where u, F are elements of a Banach space and $\mathcal{A}$ is a generator of a strongly continuous semigroup of bounded linear operators $(P_t)_{t\ge 0}$: if λ is sufficiently large, then

$$u = \int_0^\infty e^{-\lambda\tau} P_\tau F\ d\tau,$$

and, for arbitrary $T \ge 0$, by a change of variable,

$$e^{-\lambda T} P_T u = \int_T^\infty e^{-\lambda\tau} P_\tau F\ d\tau = u - \int_0^T e^{-\lambda\tau} P_\tau F\ d\tau.$$

■

Theorem 6.51 *Assume that Hypothesis 6.40 and 6.41 hold, then Eq. (6.107) has a unique mild solution given by the formula*

$$u(x) = Y(0; x). \tag{6.109}$$

Moreover, the following holds:

$$Y(\tau; x) = u(X(\tau; x)), \quad Z(\tau; x) = \nabla u(X(\tau; x))\,\sigma. \tag{6.110}$$

Proof We initially assume that in addition F is Lipschitz with respect to z, uniformly in x and y.

We introduce the following equation, slightly more general than (6.91) since we consider a general initial time $t \geq 0$:

$$X(\tau) = e^{(\tau-t)A}x + \int_t^\tau e^{(\tau-\zeta)A} b(X(\zeta))\, d\zeta, + \int_t^\tau e^{(\tau-\zeta)A} \sigma\, dW(\zeta), \qquad (6.111)$$

for τ varying on an arbitrary time interval $[t, \infty) \subset [0, \infty)$. We set $X(\tau) = x$ for $\tau \in [0, t)$ and we denote by $\{X(\tau; t, x)\}_{\tau \geq 0}$ the solution, to indicate dependence on x and t. By an obvious extension of the results in the previous sections, we can solve the backward equation (6.93) with X given by (6.111); we denote the corresponding solution (Y, Z) by $\{(Y(\tau; t, x), Z(\tau; t, x))\}_{\tau \geq 0}$.

Thus, $\{(X(\tau; 0, x), Y(\tau; 0, x), Z(\tau; 0, x))\}_{\tau \geq 0}$ coincides with the process $\{X(\tau; x), Y(\tau; x), Z(\tau; x),\ \tau \geq 0\}$ occurring in relations (6.91) and (6.93). Note that, for bounded measurable $\phi : H \to \mathbb{R}$, we have

$$P_{\tau-t}[\phi](x) = \mathbb{E}\, \phi(X(\tau; t, x)), \qquad x \in H,\ 0 \leq t \leq \tau,$$

since the coefficients of Eq. (6.111) do not depend on time.

We first prove that u, given by (6.109), is a solution. The solutions of (6.111) satisfy the well-known property: for $0 \leq t \leq s$, $\mathbb{P}$-a.s.,

$$X(\tau; s, X(s; t, x)) = X(\tau; t, x), \qquad \text{for } \tau \in [s, \infty).$$

Since the solution of the backward equation is uniquely determined on an interval $[s, \infty)$ by the values of the process X on the same interval, for $0 \leq t \leq s$ we have, $\mathbb{P}$-a.s.,

$$\begin{aligned} Y(\tau; s, X(s; t, x)) &= Y(\tau; t, x), \ \text{for } \tau \in [s, \infty), \\ Z(\tau; s, X(s; t, x)) &= Z(\tau; t, x) \text{ for a.a. } \tau \in [s, \infty). \end{aligned} \qquad (6.112)$$

In particular, for every $\tau \geq 0$,

$$Y(\tau; \tau, X(\tau; 0, x)) = Y(\tau; 0, x), \qquad \mathbb{P}\text{-a.s.} \qquad (6.113)$$

Since the coefficients of Eq. (6.111) do not depend on time, we have

$$X(\cdot; 0, x) \overset{(d)}{=} X(\cdot + t; t, x), \qquad t \geq 0,$$

where $\overset{(d)}{=}$ denotes equality in distribution (both sides of the equality are viewed as random elements with values in the space $C(\mathbb{R}^+, H)$). As a consequence we obtain

$$(Y(\cdot; 0, x), Z(\cdot; 0, x)) \overset{(d)}{=} (Y(\cdot + t; t, x), Z(\cdot + t; t, x)), \qquad t \geq 0,$$

where both sides of the equality are viewed as random elements with values in the space $C(\mathbb{R}^+, \mathbb{R}) \times L^2_{\mathrm{loc}}(\mathbb{R}^+; \Xi^*)$. In particular, $Y(0; 0, x) \overset{(d)}{=} Y(t; t, x)$, and since they are both deterministic we have

$$u(x) = Y(0; 0, x) = Y(t; t, x), \qquad x \in H,\ t \geq 0.$$

Denoting for simplicity

$$(X(\tau), Y(\tau), Z(\tau)) = (X(\tau, 0, x), Y(\tau, 0, x), Z(\tau, 0, x)), \qquad \tau \geq 0,$$

it follows from (6.113) and path continuity that, $\mathbb{P}$-a.s.,

$$u(X(\tau)) = Y(\tau), \qquad \tau \geq 0.$$

It follows that, for all $0 < t < T$,

$$Y(t) = u(X(T)) - \int_t^T Z(\zeta)\, dW(\zeta) + \lambda \int_t^T Y(\zeta)\, d\zeta + \int_t^T F(X(\zeta), Y(\zeta), Z(\zeta))\, d\zeta. \tag{6.114}$$

Thus by Corollary 6.29, considering the above equation as a BSDE on the finite horizon $[0, T]$ with final condition, it follows that, $\mathbb{P}$-a.s. for a.a. $\tau \geq 0$,

$$Z(\tau) = \nabla u(X(\tau))\, \sigma.$$

We see that by Theorem 6.48 ∇u and consequently Z is bounded by a constant that does not depend on the Lipschitz constant of F with respect to z.

Applying the Itô formula to the equation solved by (Y, Z) we get

$$\begin{aligned} e^{-\lambda\tau} Y(\tau) - e^{-\lambda T} Y(T) + \int_\tau^T e^{-\lambda\zeta} Z(\zeta)\, dW(\zeta) \\ = \int_\tau^T e^{-\lambda\zeta} F(X(\zeta), Y(\zeta), Z(\zeta))\, d\zeta,\ 0 \leq \tau \leq T < \infty, \end{aligned}$$

and it follows that

$$\begin{aligned} &\int_0^T e^{-\lambda\tau} P_\tau\Big[F\Big(\cdot, u(\cdot), \nabla u(\cdot)\Big)\Big](x)\, d\tau \\ &\qquad = \mathbb{E}\int_0^T e^{-\lambda\tau} F(X(\tau), u(X(\tau)), \nabla u(X(\tau))\,\sigma)\, d\tau \\ &\qquad = \mathbb{E}\int_0^T e^{-\lambda\tau} F(X(\tau), Y(\tau), Z(\tau))\, d\tau \\ &\qquad = \mathbb{E}\left[Y(0) - e^{-\lambda T} Y(T) + \int_0^T e^{-\lambda\tau} Z(\tau)\, dW(\tau)\right] \end{aligned}$$

$$= u(x) - e^{-\lambda T}\ \mathbb{E}\,[u(X_T)] = u(x) - e^{-\lambda T}\ P_T[u](x).$$

This completes the proof of the existence part.

Now we prove the uniqueness of the solution. Assume that u is a solution. For any $y \in H$, $0 \le \tau \le T$ we have

$$u(y) = e^{-\lambda(T-\tau)}\ P_{T-\tau}[u](y) + \int_0^{T-\tau} e^{-\lambda t}\ P_t\Big[F\Big(\cdot, u(\cdot), \nabla u(\cdot)\,\sigma\Big)\Big](y)\, dt.$$

Set $y = X(\tau, 0, x)$, which we denote by $X(\tau)$ for simplicity. By the Markov property of X, denoting by $\mathbb{E}^{\mathscr{F}_\tau}$ the conditional expectation with respect to $\mathscr{F}_\tau$, we obtain

$$\begin{aligned} u(X(\tau)) = e^{-\lambda(T-\tau)}\ &\mathbb{E}^{\mathscr{F}_\tau} u(X_T) \\ &+ \int_0^{T-\tau} e^{-\lambda t}\ \mathbb{E}^{\mathscr{F}_\tau} F\Big(X(t+\tau), u(X(t+\tau)), \nabla u(X(t+\tau))\,\sigma\Big)\, dt \end{aligned}$$

and, by a change of variable,

$$e^{-\lambda\tau} u(X(\tau)) = e^{-\lambda T}\ \mathbb{E}^{\mathscr{F}_\tau} u(X_T) + \int_\tau^T e^{-\lambda\zeta}\ \mathbb{E}^{\mathscr{F}_\tau} F\Big(X(\zeta), u(X(\zeta)), \nabla u(X(\zeta))\,\sigma\Big)\, d\zeta.$$

Now let $T > 0$ be fixed and let us define

$$F_\zeta = F(X(\zeta), u(X(\zeta)), \nabla u(X(\zeta))\,\sigma), \quad \zeta \in [0, T],$$

$$\xi = e^{-\lambda T} u(X_T) + \int_0^T e^{-\lambda\zeta} F_\zeta\, d\zeta.$$

Then we obtain

$$e^{-\lambda\tau} u(X(\tau)) = \mathbb{E}^{\mathscr{F}_\tau}\xi + \mathbb{E}^{\mathscr{F}_\tau}\int_0^\tau e^{-\lambda\zeta} F_\zeta\, d\zeta = \mathbb{E}^{\mathscr{F}_\tau}\xi + \int_0^\tau e^{-\lambda\zeta} F_\zeta\, d\zeta,$$

where the last equality holds since $\int_0^\tau e^{-\lambda\zeta} F_\zeta\, d\zeta$ is $\mathscr{F}_\tau$-adapted. Notice that ξ is square-integrable. Since $\mathscr{F}_t$ is generated by the Wiener process W, it follows that there exists a square-integrable, $\mathscr{F}_t$-predictable process $\widetilde{Z}(\tau)$, $\tau \in [0, T]$, with values in Ξ^*, such that, $\mathbb{P}$-a.s.,

$$\mathbb{E}^{\mathscr{F}_\tau}\xi = \mathbb{E}\,\xi + \int_0^\tau \widetilde{Z}(\zeta)\, dW(\zeta), \qquad \tau \in [0, T].$$

An application of the Itô formula gives

$$u(X(\tau)) = \mathbb{E}\,\xi + \int_0^\tau e^{\lambda\zeta}\widetilde{Z}(\zeta)\,dW(\zeta) + \lambda\int_0^\tau u(X(\zeta))\,d\zeta + \int_0^\tau F(\zeta)\,d\zeta. \tag{6.115}$$

This shows that $u(X(\tau))$, $\tau \in [0, T]$, is a semimartingale. For $\xi \in \Xi$, we denote again by W^ξ the real Wiener process $W^\xi(\tau) := \langle \xi, W(\tau)\rangle$, $\tau \geq 0$. Let us consider the joint quadratic variation process of W^ξ with both sides of (6.115). Applying Proposition 6.17 (recall that u is by definition differentiable) we obtain, $\mathbb{P}$-a.s.,

$$\int_0^\tau \nabla u(X(\zeta))\,\sigma\xi\,d\zeta = \int_0^\tau e^{\lambda\zeta}\widetilde{Z}(\zeta)\xi\,d\zeta, \qquad \tau \in [0, T],\ \xi \in \Xi,$$

and we deduce that $\nabla u(X(\tau))\,\zeta = e^{\lambda\tau}\widetilde{Z}(\tau)$, $\mathbb{P}$-a.s. for almost all $\tau \in [0, T]$. Now setting

$$Y'(\tau) = u(X(\tau)), \qquad Z'(\tau) = e^{\lambda\tau}\nabla u(X(\tau))\sigma, \qquad \tau \geq 0,$$

it follows from (6.115) that, $\mathbb{P}$-a.s.,

$$Y(0) = Y'(\tau) + \int_0^\tau Z'(\zeta)\,dW(\zeta) + \lambda\int_0^\tau Y'(\zeta)\,d\zeta + \int_0^\tau F(X(\zeta), Y'(\zeta), Z'(\zeta))\,d\zeta,$$

for $\tau \in [0, T]$. Since T is arbitrary, we conclude that the process (Y', Z') is a solution of the backward equation, so that, by uniqueness, it must coincide with (Y, Z). In particular,

$$u(x) = u(X_0) = Y'(0) = Y(0).$$

This concludes the proof of the theorem. □

6.10 Application to Optimal Control in an Infinite Horizon

We wish to apply the above results to perform the synthesis of the optimal control for a general nonlinear control system on an infinite time horizon. To be able to use non-smooth feedbacks we settle the problem in the framework of weak control problems.

As above, by H, Ξ we denote separable real Hilbert spaces.

Moreover, a *generalized reference probability space* is given by $\mu = (\Omega, \mathscr{F}, \mathscr{F}_s, \mathbb{P}, W)$, where

- $(\Omega, \mathscr{F}, \mathbb{P})$ is a complete probability space;
- $\{\mathscr{F}_s\}_{s\geq 0}$ is a filtration in it, satisfying the usual conditions;
- $(W(s))_{s\geq 0}$ is a cylindrical $\mathbb{P}$-Wiener process in Ξ, with respect to the filtration $\mathscr{F}_s$

(notice that, since our problem is homogeneous in time, we always choose the initial time $t = 0$).

Given such μ, we call an *admissible control pair* the pair $(a(\cdot), X(\cdot))$ of progressively measurable processes with respect to $\{\mathscr{F}_s\}_{s\geq 0}$ such that: a is defined on $\Omega \times [0, \infty)$ and takes its values in a fixed closed subset (not necessarily bounded) Λ of a Banach space E. Moreover, a is uniformly bounded, that is belongs to $L^\infty(\Omega \times [0, \infty), \mathbb{P} \otimes dt; E)$. Finally, X is the mild solution (on the whole $[0, \infty)$) of the following state equation:

$$\begin{cases} dX(\tau) = (AX(\tau) + b(X(\tau) + \sigma R(a(\tau))) \ d\tau + \sigma \ dW(\tau), & \tau \geq 0, \\ X^a(0) = x \in H. \end{cases} \tag{6.116}$$

Notice that in the present case the assumptions on R will guarantee the existence and uniqueness of the mild solution X given and control a satisfying the above, so we work in the framework of the weak and strong formulations in the sense of Sect. 2.1.

To each admissible control pair we associate the cost:

$$J^\mu(x; a(\cdot), X(\cdot)) = \mathbb{E} \int_0^{+\infty} e^{-\lambda\zeta}[l(X(\zeta)) + |a(\zeta)|_E^2] \ d\zeta, \tag{6.117}$$

where $l : H \times \Lambda \to \mathbb{R}$. As in the finite horizon case we minimize the functional $J^\mu(x; a(\cdot))$ over all admissible controls $a(\cdot)$ and characterize the value function

$$V^\mu(x) = \inf_a J^\mu(x; a(\cdot), X(\cdot)).$$

We will also address the optimal control problem in the weak formulation, which consists in further minimizing with respect to all generalized reference probability spaces, i.e., in characterizing the value function

$$\overline{V}(x) = \inf_\mu V^\mu(x).$$

Notice the occurrence of the operator σ in the control term: this special structure of the state equation is imposed by our techniques. Also notice that in contrast to what happens in the previous sections of this Chapter we now restrict ourselves to R that does not depend on x. This also ensures that for all $a(\cdot) \in \overline{\mathcal{U}}^\mu$ and $x \in H$ Eq. (6.116) admits a unique mild solution

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

$$\begin{aligned} F_0(x, z) &= l(x) + \inf\{|a|_E^2 + zR(a) : a \in \Lambda\} \\ \Gamma(z) &= \{a \in \Lambda : |a|_E^2 + zR(a) = \inf_{a\in\Lambda}\{|a|_E^2 + zR(a)\}\}. \end{aligned} \tag{6.118}$$

We will work in the following general setting:

Hypothesis 6.52 The following holds:

(1) A, b, σ and satisfy Hypothesis 6.40.

(2) $R : \Lambda \to \Xi$ is Lipschitz.
(3) $l : H \to \mathbb{R}$ is uniformly Lipschitz, bounded and of class $\mathcal{G}^1(H, \mathbb{R})$.
(4) F_0 is of class $\mathcal{G}^1(\Xi^*, \mathbb{R})$.

Remark 6.53 Since R is Lipschitz $\inf\{|a|_E^2 + zR(a) : a \in \Lambda\}$ is always a real number. Moreover, there exists a constant c_R such that

$$\inf\{|a|_E^2 + zR(a) : a \in \Lambda\} = \inf\{|a|_E^2 + zR(a) : a \in \Lambda \cap B_E(0, c|z|)\}.$$

This immediately implies that $\Gamma(z) \subset B(0, c_R|z|)$ and that

$$\left|\inf\{|a|_E^2 + zR(a) : a \in \Lambda\} - \inf\{|a|_E^2 + z'R(a) : a \in \Lambda\}\right| \leq c_{1,R}(|z| + |z'|)|z - z'|.$$

So Hypothesis 6.41 holds true. ■

We see that for all $\lambda > 0$ the cost functional is well defined and $J^\mu(x; a(\cdot), X(\cdot)) < \infty$ for all $x \in H$, all admissible control systems μ and all admissible control pairs (a, X).

By Theorem 6.51, for all $\lambda > 0$ the stationary Hamilton–Jacobi–Bellman equation relative to the above stated problem, namely:

$$\mathcal{A}v(x) = \lambda v(x) - F_0(x, \nabla v(x)\sigma), \qquad x \in H, \tag{6.119}$$

admits a unique mild solution, in the sense of Definition 6.49.

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

Theorem 6.54 *Assume Hypothesis 6.52 and suppose that $\lambda > 0$. Then the following holds*

(1) For all generalized reference probability spaces and admissible pairs (a, X) we have $J^\mu(x; a(\cdot), X(\cdot)) \geq v(x)$. Therefore $V^\mu(x) \leq v(x)$.
(2) The equality $J^\mu(x; a(\cdot), X(\cdot)) = v(x)$ holds if and only if the following feedback law is satisfied:

$$a(\tau) \in \Gamma(\nabla v(X(\tau))\sigma), \quad \mathbb{P}\text{-a.s. for a.e. } \tau \geq 0. \tag{6.120}$$

Notice that since ∇v is bounded, if (6.120) holds then the control a is uniformly bounded.

Proof Choose any generalized reference probability space μ and denote by $\rho(T)$ the Girsanov density

$$\rho(T) = \exp\left(-\int_0^T \langle R(a(\zeta)), dW(\zeta)\rangle_\Xi - \frac{1}{2}\int_0^T |R(a(\zeta))|_\Xi^2\, d\zeta\right), \tag{6.121}$$

Let $\widetilde{\mathbb{P}}_T$ be the probability measure on $\mathscr{F}_T$ defined by $\widetilde{\mathbb{P}}_T = \rho(T)\,\mathbb{P}\Big|_{\mathscr{F}_T}$ and let $\tilde{\mathbb{E}}_T$ be the corresponding expectation. By Girsanov's Theorem (see Theorem 6.34) under $\widetilde{\mathbb{P}}_T$ the process

$$\widetilde{W}_\tau := \int_0^\tau R(a(\zeta))\, d\zeta + W_\tau, \qquad 0 \le \tau \le T, \tag{6.122}$$

is a cylindrical Wiener process. Equation (6.116) can be written:

$$\begin{cases} dX(\tau) = AX(\tau)\, d\tau + b(X(\tau))\, d\tau + \sigma\, d\widetilde{W}_\tau, & \tau \ge 0, \\ X_0 = x. \end{cases} \tag{6.123}$$

Let v be the unique mild solution of Eq. (6.119). Consider the following finite horizon Markovian forward–backward system (with respect to probability $\widetilde{\mathbb{P}}_T$ and to the filtration generated by $\{\widetilde{W}_\tau\}_{\tau\in[0,T]}$).

$$\begin{cases} \widetilde{X}(\tau; x) = e^{\tau A}x + \int_0^\tau e^{(\tau-\zeta)A} b(\widetilde{X}(\zeta; x))\, d\zeta + \int_0^\tau e^{(\tau-\zeta)A}\sigma\, d\widetilde{W}_\zeta, \quad \tau \ge 0, \\ \widetilde{Y}(\tau; x) - v(\widetilde{X}(T; x)) + \int_\tau^T \widetilde{Z}(\zeta; x)\, d\widetilde{W}_\zeta + \lambda \int_\tau^T \widetilde{Y}(\zeta; x)\, d\zeta \\ \qquad\qquad = \int_\tau^T F_0(\widetilde{X}(\zeta; x), \widetilde{Z}(\zeta; x))\, d\zeta,\ 0 \le \tau \le T, \end{cases} \tag{6.124}$$

and let $(\widetilde{X}(x), \widetilde{Y}(x), \widetilde{Z}(x))$ be its unique solution with the three processes predictable relative to the filtration generated by $\{\widetilde{W}_\tau\}_{\tau\in[0,T]}$ and: $\tilde{\mathbb{E}}_T \sup_{t\in[0,T]} |\widetilde{X}(t; x)|^2 < +\infty$, $\widetilde{Y}(x)$ bounded and continuous, $\tilde{\mathbb{E}}_T \int_0^T |\widetilde{Z}(t; x)|^2 dt < +\infty$.

Moreover, Theorem 6.51 and uniqueness of the solution of system (6.124) together with Theorem 6.32 yields

$$\widetilde{Y}(\tau; x) = v(\widetilde{X}(\tau; x)), \qquad \widetilde{Z}(\tau; x) = \nabla v(\widetilde{X}(\tau; x))\sigma. \tag{6.125}$$

Comparing the forward equation in (6.124) with the state equation, rewritten as (6.123), we get $\widetilde{X}(t; x) = X_t$, $t \in [0, T]$, $\mathbb{P}$-a.s. Applying the Itô formula to $e^{-\lambda\tau}\widetilde{Y}(\tau; x)$, and restoring the original noise W, we get

$$\begin{aligned} &\widetilde{Y}(0; x) + \int_0^T e^{-\lambda\zeta}\widetilde{Z}(\zeta; x)\, dW_\zeta \\ &\quad = \int_0^T e^{-\lambda\zeta}\left[F_0(X(\zeta), \widetilde{Z}(\zeta; x)) - \widetilde{Z}(\zeta; x)R(a(\zeta))\right] d\zeta + e^{-\lambda T} v(X(T)). \end{aligned} \tag{6.126}$$

Using the identification in (6.125) and taking expectation with respect to $\mathbb{P}$, (6.126) yields

$$e^{-\lambda T}\mathbb{E}v(\widetilde{X}(T;x)) - v(x) = -\mathbb{E}\int_0^T e^{-\lambda\zeta}F_0(X(\zeta),\nabla v(X(\zeta))\sigma)\,d\zeta \\ +\mathbb{E}\int_0^T e^{-\lambda\zeta}\nabla v(X(\zeta))\zeta R(a(\zeta))\,d\zeta.$$

Recalling that v is bounded, letting $T\to\infty$, we conclude that

$$J^{\mu}(x;a(\cdot),X(\cdot)) = v(x) - \mathbb{E}\int_0^\infty e^{-\lambda\zeta}F_0(X(\zeta),\nabla v(X(\zeta))\sigma)d\zeta \\ -\mathbb{E}\int_0^\infty e^{-\lambda\zeta}\Big[\nabla_x v(X(\zeta))\sigma R(a(\zeta)) - l(X(\zeta),a(\zeta))\Big]\,d\zeta.$$

The above equality is known as the *fundamental relation* and immediately implies that $v(x)\le J^{\mu}(x;a(\cdot))$ and that equality holds if and only if (6.120) holds. □

Theorem 6.55 *Assume Hypothesis 6.52 and that $\lambda>0$. If $\Gamma(x,z)$ is non-empty for all $x\in H$ and $z\in\Xi^*$ and $\gamma:\Xi^*\to\Lambda$ is a measurable selection of Γ (which exists, see Theorem 8.2.10, in [20]) then there exists a generalized reference probability space $\bar{\mu}$ in which the closed loop equation*

$$\begin{cases} d\overline{X}(\tau) = A\overline{X}(\tau)\,d\tau + \sigma R(\gamma(\nabla v(\overline{X}(\tau))\sigma)\,d\tau + b(\overline{X}(\tau))\,d\tau + \sigma\,dW(\tau), \ \tau\ge 0,\\ \overline{X}_0 = x_0\in H, \end{cases} \tag{6.127}$$

admits a solution. Moreover, setting $\overline{a}(\tau)=\gamma(\nabla v(\overline{X}(\tau))\sigma)$, the pair $(\overline{a}(\cdot),\overline{X}(\cdot))$ is admissible and optimal for the control problem in the sense that

$$J^{\bar{\mu}}(x;\bar{a}(\cdot),\bar{X}(\cdot)) = v(x).$$

Consequently, we have $v(x)=\bar{V}(x)$.

Proof The point here is to prove the existence of a weak (in the probabilistic sense) solution to Eq. (6.127) in the whole $[0,+\infty)$, see also Sect. 4 in [274]. In order to do this we realize a "canonical"-Ξ-valued Wiener process. We choose a larger Hilbert space $\Xi'\supset\Xi$ in such a way that Ξ is continuously and densely embedded in Ξ' with Hilbert–Schmidt inclusion operator $\mathcal{J}$. By Ω we denote the space $C([0,\infty),\Xi')$ of continuous functions $\omega:[0,\infty)\to\Xi'$ endowed with the standard locally convex topology and by $\mathcal{B}$ its Borel σ-field. Since $\mathcal{J}\mathcal{J}^*$ is nuclear on Ξ' we know (see [180]) that there exists a probability $\mathbb{P}$ on $\mathcal{B}$ such that $W_t'(\omega):=\omega(t)$ is a $\mathcal{J}\mathcal{J}^*$-Wiener process in Ξ' (that is, $t\to\langle W_t',\xi'\rangle_{\Xi'}$ is a real-valued Wiener process for all $\xi'\in\Xi'$ and $\mathbb{E}[\langle W_t',\xi'\rangle_{\Xi'}\langle W_s',\eta'\rangle_{\Xi'}]=\langle\mathcal{J}\mathcal{J}^*\xi',\eta'\rangle_{\Xi'}(t\wedge s)$ for all $\xi',\eta'\in\Xi'$, $t,s\in[0,\infty)$). We denote by $\mathcal{E}$ the $\mathbb{P}$-completion of $\mathcal{B}$ and by $\mathscr{F}_t$, $t\ge0$, the $\mathbb{P}$-completion of $\mathcal{B}_t=\sigma\left(W_s':s\in[0,t]\right)$.

The Ξ-valued cylindrical Wiener process $\{W_t^{\xi}:t\ge0,\xi\in\Xi\}$ can now be defined as follows. For ξ in the image of $\mathcal{J}^*\mathcal{J}$ we take η such that $\xi=\mathcal{J}^*\mathcal{J}\eta$ and define $W_s^{\xi}=\langle W_s',\mathcal{J}\eta\rangle_{\Xi'}$. Then we observe that $\mathbb{E}|W_t^{\xi}|^2=t|\mathcal{J}\eta|^2_{\Xi'}=t|\xi|^2_{\Xi}$ and that $\mathcal{J}^*\mathcal{J}\Xi$

is dense in Ξ to deduce that the linear continuous mapping $\xi \to W_s^\xi$ (with values in $L^2(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R})$) can be extended by continuity to the whole Ξ. An appropriate modification of $\{W_t^\xi : t \geq 0, \xi \in \Xi\}$ gives the required cylindrical Wiener process.

Now let $X \in L^p_{\mathcal{P},\mathrm{loc}}(\Omega, C([0, +\infty), H))$ be the mild solution of

$$\begin{cases} dX(\tau) = AX(\tau)\, d\tau + b(X(\tau))\, d\tau + \sigma\, dW(\tau) \\ X(0) = x \end{cases} \tag{6.128}$$

and let, $\forall T > 0$

$$\rho(T) = \exp\left(-\int_0^T \langle R(\gamma(\nabla v(X(\zeta))\sigma), dW(\zeta)\rangle_\Xi - \frac{1}{2}\int_0^T |R(\gamma(\nabla v(X(\zeta))\sigma)|^2_\Xi\, d\zeta \right). \tag{6.129}$$

Recall that ∇v is bounded. Thus let $\widehat{\mathbb{P}}_T$ be the probability on $\mathscr{F}_T$ admitting $\rho(T)$ as a density with respect to $\mathbb{P}$. Since Ξ' is a Polish space and $\widehat{\mathbb{P}}_{T+h}$ coincides with $\widehat{\mathbb{P}}_T$ on $\mathcal{B}_T$, $T, h \geq 0$, by known results (see [508], Chap. VIII, Sect. 1, Proposition 1.13) there exists a probability $\widehat{\mathbb{P}}$ on $\mathcal{B}$ such that the restriction on $\mathcal{B}_T$ of $\widehat{\mathbb{P}}_T$ and that of $\widehat{\mathbb{P}}$ coincide, $T \geq 0$. Let $\widehat{\mathcal{E}}$ be the $\widehat{\mathbb{P}}$-completion of $\mathcal{B}$ and $\widehat{\mathscr{F}}_T$ be the $\widehat{\mathbb{P}}$-completion of $\mathcal{B}_T$. Moreover, let

$$\widehat{W}(t) := -\int_0^t R(\gamma(\nabla v(X(\zeta))\sigma)\, d\zeta + W(t).$$

Since, for all $T > 0$, $\{\widehat{W}_t\}_{t\in[0,T]}$ is a Ξ-valued cylindrical Wiener process under $\widehat{\mathbb{P}}_T$ (see again Theorem 6.34) and the restriction of $\widehat{\mathbb{P}}_T$ and of $\widehat{\mathbb{P}}$ coincide on $\mathcal{B}_T$, modifying $\{\widehat{W}_t\}_{t\geq 0}$ in a suitable way on a $\widehat{\mathbb{P}}$-null probability set we can conclude that $\bar{\mu} = (\Omega, \widehat{\mathcal{E}}, \{\widehat{\mathscr{F}}_t\}_{t\geq 0}, \widehat{\mathbb{P}}, \{\widehat{W}_t\}_{t\geq 0})$ is a generalized reference probability space and that if we set $\bar{a}(\tau) = \gamma(\nabla v(X(\tau))\sigma)$ then $(\bar{a}(\cdot), X(\cdot))$ is an admissible pair and (6.127) is satisfied. Indeed, if we rewrite (6.128) in terms of $\{\widehat{W}_t\}_{t\geq 0}$ we get

$$\begin{cases} dX(\tau) = AX(\tau)\, d\tau + b(X(\tau))\, d\tau + G\, [R(\gamma(\nabla v(X(\tau))\sigma)) + d\widehat{W}(\tau)], \\ X_0 = x \end{cases}$$

and this concludes the proof. □

6.11 Application: The Heat Equation with Additive Noise

We show here how the previous results can be applied to a stochastic heat equation with additive white noise in dimension 1. Let, for $t \geq 0$, $\xi \in [0, 1]$:

$$\begin{cases} \frac{\partial}{\partial t}x(t,\xi) = \frac{\partial^2}{\partial \xi^2}x(t,\xi) + f_0(\xi, x(t,\xi)) + \sigma_0(\xi)r(\xi)\,a(t,\xi) + \sigma_0(\xi)\,\frac{\partial}{\partial t}\mathcal{W}(t,\xi), \\ x(t,0) = x(t,1) = 0, \\ x(0,\xi) = x_0(\xi), \end{cases} \tag{6.130}$$

where $\frac{\partial}{\partial t}\mathcal{W}$ is a space-time white noise on $\mathbb{R}^+ \times [0,1]$. Moreover, we introduce the cost functional:

$$J(x_0, a(\cdot), x(\cdot)) = \mathbb{E}\int_0^\infty \int_0^1 e^{-\lambda t}\left[\ell_0(\xi, x(t,\xi)) + |a(t,\xi)|^2\right] d\xi\, dt \tag{6.131}$$

which we minimize over all progressive controls $a : [0,\infty) \times [0,1] \to \mathbb{R}$ bounded in $L^2([0,1])$. By this we mean that there exists a suitable constant c_a (depending on the control a) such that:

$$\int_0^1 a^2(t,\xi)d\xi \le c_a, \quad \mathbb{P}\otimes dt\text{-a.s.}$$

To fit the assumptions of our abstract results we will suppose that the functions f_0, σ_0, r, ℓ_0 are all measurable and real-valued and moreover:

(1) f_0 is defined on $[0,1]\times\mathbb{R}$ and $\int_0^1 f_0^2(\xi,0)d\xi < +\infty$.
Moreover, for a.a. $\xi \in [0,1]$, we require that $f_0(\xi,\cdot) \in C^1(\mathbb{R})$ and

$$-L_f \le \frac{\partial}{\partial \eta} f(\xi,\eta) \le 0$$

for a suitable constant $L_f > 0$, almost all $\xi \in [0,1]$, and all $\eta \in \mathbb{R}$.

(2) σ_0 and r are bounded measurable functions from $[0,1]$ to $\mathbb{R}$.

(3) ℓ_0 is defined on $[0,1]\times\mathbb{R}$ and, for a.a. $\xi \in [0,1]$, the map $\ell_0(\xi,\cdot)$ is in $C^1(\mathbb{R},\mathbb{R})$.
Moreover:

$$|\ell_0(\xi,\eta)| \le c_0(\xi), \quad \left|\frac{\partial}{\partial \eta}\ell_0(\xi,\eta)\right| \le c_1(\xi), \quad \text{with} \int_0^1 \left(c_0(\xi) + c_1^2(\xi)\right) d\xi < +\infty. \tag{6.132}$$

(4) $x_0 \in L^2([0,1])$.

To rewrite the above problem in the abstract way we set (with the notation of Sect. 6.10): $H = \Xi = \Lambda = L^2([0,1])$. By $\{W(t)\}_{t\ge 0}$ we denote a cylindrical Wiener process in $L^2([0,1])$. Moreover, we define the operator A with domain $D(A)$ by:

$$D(A) = W^{2,2}([0,1]) \cap W_0^{1,2}([0,1]), \qquad (Ay)(\xi) = \frac{\partial^2}{\partial \xi^2} y(\xi), \quad \forall y \in D(A),$$

where $W^{2,2}([0,1])$ and $W_0^{1,2}([0,1])$ are the usual Sobolev spaces, and we set

$$b(x)(\xi) = f_0(\xi, x(\xi)), \quad (\sigma z)(\xi) = \sigma_0(\xi) z(\xi), \quad R(a)(\xi) = (Ra)(\xi) = r(\xi)a(\xi),$$
$$l(x, a) = \int_0^1 \left[|a(\xi)|^2 + \ell_0(\xi, x(\xi)) \right] d\xi$$

for all $x, z \in L^2([0,1])$ $a \in L^\infty([0,1])$ and a.a. $\xi \in [0,1]$.

Under the previous assumptions we know, see [177] Sect. 11.2.1, that A, b and σ satisfy Hypothesis 6.40. Moreover, R is a bounded linear operator on $L^2([0,1])$ and

$$\nabla_x l(x, a) h = \int_0^1 \frac{\partial}{\partial \eta} \ell_0(\xi, x(\xi)) h(\xi) d\xi.$$

Hence points 2 and 3 in Hypothesis 6.52 are satisfied.

We also notice that

$$\inf_{a \in H} (|a|_H^2 + z(Ra)) = \inf_{a \in H} (|a|_H^2 + (R^* z)a) = -\frac{1}{4} |R^* z|_{H^*}^2 = -\frac{1}{4} \int_0^1 r^2(\xi) z^2(\xi) d\xi.$$

So $F_0(x, z) = l(x) - \frac{1}{4}|R^* z|^2$ and, taking into account the regularity of ℓ_0, it is immediate to see that point 4 in Hypothesis 6.40 is satisfied. In addition, $\inf_{a \in L^2([0,1])} (|a|_H^2 + z(Ra))$ is a minimum achieved for $a = -\frac{1}{2} r z$.

As a consequence of Theorems 6.54 and 6.55 we have the following result.

Theorem 6.56 *Under the previous assumptions, fixing $\lambda > 0$, there exists at least one generalized reference probability space $\overline{\mu} = (\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathscr{F}}_s, \overline{\mathbb{P}}, \overline{W})$ and an admissible control pair $(\overline{a}(\cdot), \overline{x}(\cdot))$ for which*

$$\overline{V}(x_0) = J^{\overline{\mu}}(x_0; \overline{a}(\cdot), \overline{x}(\cdot)), \qquad x_0 \in L^2([0,1]).$$

In particular, the triple $(\overline{\mu}, \overline{a}(\cdot), \overline{x}(\cdot))$ is optimal.

The value function $\overline{V}(x_0)$ coincides with the function $v(x_0)$, which is the unique mild solution to the Hamilton–Jacobi–Bellman equation (6.119) in the sense specified by Definition 6.49 (see Theorem 6.51) where (with the standard identifications)

$$F_0(x, \nabla v(x)\sigma) = l(x) - \frac{1}{4} |R^* \nabla v(x) \sigma|_{H^*}^2$$
$$= \int_0^1 \ell_0(\xi, x(\xi)) d\xi - \frac{1}{4} \int_0^1 r^2(\xi) \sigma_0^2(\xi) (\nabla v(x)(\xi))^2 d\xi.$$

In the space $\overline{\mu}$ the process $(\overline{x}(s, \cdot))_{s \geq 0}$ is a mild solution to the closed loop equation

$$\begin{cases} \frac{\partial}{\partial t} \bar{x}(t, \xi) = \frac{\partial^2}{\partial \xi^2} \bar{x}(t, \xi) + f_0(\xi, \bar{x}(t, \xi)) - \frac{1}{2} \sigma_0^2(\xi) r^2(\xi) \nabla_x v(t, \bar{x}(t, \cdot))(\xi) + \sigma_0(\xi) \frac{\partial}{\partial t} \mathcal{W}(t, \xi), \\ \bar{x}(t, 0) = \bar{x}(t, 1) = 0, \\ \bar{x}(0, \xi) = x_0(\xi), \end{cases}$$

and the optimal pair $(\overline{a}(t, \cdot), \overline{x}(t, \cdot))$ satisfies the feedback law equality

$$\overline{a}(s,\xi) = -\frac{1}{2}\sigma_0(\xi)r(\xi)\nabla_x v(t,x(t,\cdot))(\xi).$$

6.12 Elliptic HJB Equations with Non-constant Diffusion

In this section we wish to briefly expose the results on the probabilistic representation of the solution to an elliptic HJB equation when the second-order operator $\mathrm{Tr}\left(\sigma(x)\sigma(x)^* D^2\phi(x)\right)$ depends on x. Namely, we will address the resolvability of the following equation:

$$\mathcal{A}u(x) - \lambda\, u(x) = F(x, u(x), \nabla u(x)\, \sigma(x)), \qquad x \in H,$$

where

$$\mathcal{A}\phi(x) = \frac{1}{2}\mathrm{Tr}\left(\sigma(x)\sigma(x)^* D^2\phi(x)\right) + \langle Ax + b(x), D\phi(x)\rangle\,.$$

The price to pay to allow σ to depend on x is that we will have to assume λ to be large enough.

The detailed proofs of the results reported below can be founded in [285].

Our analysis here will be done on the weighted (in time) spaces that we introduce below.

- $L^p_{\mathcal{P}}(\Omega; L^q_\beta(K))$, defined for $\beta \in \mathbb{R}$ and $p, q \in [1, \infty)$, denotes the space of equivalence classes of processes $\{Y(t)\}_{t\geq 0}$, with values in K, such that the norm

$$|Y|^p_{L^p_{\mathcal{P}}(\Omega; L^q_\beta(K))} = \mathbb{E}\left(\int_0^\infty e^{q\beta s}|Y(s)|^q_K\, ds\right)^{p/q}$$

  is finite, and Y admits a predictable version.
- $\mathcal{K}^p_\beta$ denotes the space $L^p_{\mathcal{P}}(\Omega; L^2_\beta(K)) \times L^p_{\mathcal{P}}(\Omega; L^2_\beta(\mathcal{L}_2(\Xi, K)))$. The norm of an element $(Y, Z) \in \mathcal{K}^p_\beta$ is $|(Y, Z)|_{\mathcal{K}^p_\beta} = |Y|_{L^p_{\mathcal{P}}(\Omega; L^2_\beta(K))} + |Z|_{L^p_{\mathcal{P}}(\Omega; L^2_\beta(\mathcal{L}_2(\Xi, K)))}$.
- $L^q_{\mathcal{P}}(\Omega; C_\eta(K))$, defined for $\eta \in \mathbb{R}$ and $q \in [1, \infty)$, denotes the space of predictable processes $\{Y(t)\}_{t\geq 0}$ with continuous paths in K, such that the norm

$$|Y|^q_{L^q_{\mathcal{P}}(\Omega; C_\eta(K))} = \mathbb{E}\,\sup_{\tau\geq 0} e^{\eta q\tau}|Y(\tau)|^q_K$$

  is finite. Elements of $L^q_{\mathcal{P}}(\Omega; C_\eta(K))$ are identified up to indistinguishability.
- Finally, for $\eta \in \mathbb{R}$ and $q \in [1, \infty)$, we define $\mathcal{H}^q_\eta$ as the space $L^q_{\mathcal{P}}(\Omega; L^q_\eta(K)) \cap L^q_{\mathcal{P}}(\Omega; C_\eta(K))$, endowed with the norm

$$|Y|_{\mathcal{H}^q_\eta} = |Y|_{L^q_{\mathcal{P}}(\Omega; L^q_\eta(K))} + |Y|_{L^q_{\mathcal{P}}(\Omega; C_\eta(K))}.$$

Clearly, similar definitions and notations also apply to processes with values in other Hilbert spaces, different from K.

As in the previous sections, we denote by $\{W(\tau)\}_{\tau\geq 0}$ a cylindrical Wiener process with values in a Hilbert space Ξ, defined on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Now we consider the Itô stochastic differential equation for an unknown process $\{X(\tau; x)\}_{\tau\geq 0}$ with values in a Hilbert space H:

$$X(\tau; x) = e^{\tau A}x + \int_0^\tau e^{(\tau-s)A} b(X(s; x))\, ds + \int_0^\tau e^{(\tau-s)A}\sigma(X(s; x))\, dW(s), \qquad \tau \geq 0. \tag{6.133}$$

Hypothesis 6.57 (i) The operator A is the generator of a strongly continuous semigroup e^{tA}, $t \geq 0$, in the Hilbert space H. We denote by M and a two constants such that $|e^{tA}| \leq Me^{at}$ for $t \geq 0$.

(ii) The mapping $b : H \to H$ satisfies, for some constant $L > 0$,

$$|b(x) - b(y)| \leq L\, |x - y|, \qquad x, y \in H.$$

(iii) σ is a mapping from H to $\mathcal{L}(\Xi, H)$ such that for every $\xi \in \Xi$ the map $\sigma(\cdot)\xi : H \to H$ is measurable, $e^{tA}\sigma(x) \in \mathcal{L}_2(\Xi, H)$ for every $t > 0$ and $x \in H$, and

$$\begin{aligned} &|e^{tA}\sigma(x)|_{\mathcal{L}_2(\Xi,H)} \leq L\, t^{-\gamma} e^{at}(1 + |x|),\\ &|e^{tA}\sigma(x) - e^{tA}\sigma(y)|_{\mathcal{L}_2(\Xi,H)} \leq L\, t^{-\gamma} e^{at}|x - y|, \qquad t > 0,\ x, y \in H, \end{aligned} \tag{6.134}$$

$$|\sigma(x)|_{\mathcal{L}(\Xi,H)} \leq L\, (1 + |x|), \qquad x \in H, \tag{6.135}$$

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

(iv) For every $t > 0$, we have $b(\cdot) \in \mathcal{G}^1(H, H)$ and $e^{tA}\sigma(\cdot) \in \mathcal{G}^1(H, \mathcal{L}_2(\Xi, H))$.

Proposition 6.58 *Assume that Hypothesis 6.57 holds. Then for all $q \in [1, \infty)$ there exists a constant $\eta(q)$, depending also on γ, L, a, M, with the following properties:*

(i) For all $x \in H$ the process $X(\cdot; x)$, a solution of (6.133), is in $\mathcal{H}^q_{\eta(q)}$ (here $K = H$).

(ii) For a suitable constant $C > 0$ we have

$$\mathbb{E} \sup_{\tau\geq 0} e^{\eta(q)q\tau}|X(\tau; x)|^q + \mathbb{E}\int_0^\infty e^{\eta(q)qs}\, |X(s; x)|^q\, ds \leq C(1 + |x|)^q. \tag{6.136}$$

(iii) The map $x \to X(\cdot; x)$ belongs to $\mathcal{G}^1(H, \mathcal{H}^q_{\eta(q)})$ and its derivative is uniformly bounded:

$$|\nabla X(\cdot; x)h|_{\mathcal{H}^q_{\eta(q)}} \leq C\, |h|, \qquad x, h \in H, \tag{6.137}$$

for a suitable constant C.

Let us now denote by $\mathscr{F}_\tau$ the natural filtration of $\{W(\tau)\}_{\tau\geq 0}$ augmented in the usual way. We again consider the system of stochastic differential equations: $\mathbb{P}$-a.s., for $0 \leq \tau \leq T < \infty$

$$\begin{cases} X(\tau; x) = e^{\tau A}x + \int_0^\tau e^{(\tau-s)A} b(X(s;x))\,ds + \int_0^\tau e^{(\tau-s)A}\sigma(X(s;x))\,dW(s), \\ Y(\tau;x) + \int_\tau^T Z(s;x)\,dW(s) + \lambda \int_\tau^T Y(s;x)\,ds \\ \qquad\qquad = \int_\tau^T F(X(s;x), Y(s;x), Z(s;x))\,ds. \end{cases} \tag{6.138}$$

Y is real-valued and Z takes values in Ξ^*, $F : H \times \mathbb{R} \times \Xi^* \to \mathbb{R}$ is a given measurable function, x is in H and λ is a real number.

For any $q \in [1, \infty)$ we choose $\eta(q)$ as in Proposition 6.58. Then, we know that for every $x \in H$, there exists a unique solution $\{X(\tau;x)\}_{\tau\geq 0}$ in $\mathcal{H}^q_{\eta(q)}$ of the forward equation and the map $x \to X(\cdot;x)$ belongs to $\mathcal{G}^1(H, \mathcal{H}^q_{\eta(q)})$.

Then we fix $p > 2$ and choose q and β satisfying

$$q \geq p(m+1)(m+2), \qquad \beta < \eta(q)(m+1)(m+2), \qquad \beta < 0. \tag{6.139}$$

On F we shall ask the following

Hypothesis 6.59 (i) There exist $\mu \in \mathbb{R}$ and nonnegative constants L_y, L_z such that

$$|F(x, y_1, z_1) - F(x, y_2, z_2)| \leq L_y|y_1 - y_2| + L_z|z_1 - z_2|,$$
$$\langle F(x, y_1, z) - F(x, y_2, z), y_1 - y_2\rangle_K \leq -\mu|y_1 - y_2|^2,$$

for every $x \in H$, $y_1, y_2 \in \mathbb{R}$, $z, z_1, z_2 \in \Xi^*$.

(ii) $F \in \mathcal{G}^1(H \times \mathbb{R} \times \Xi^*, K)$.

(iii) There exist $L > 0$ and $m \geq 0$ such that

$$|\nabla_x F(x, y, z)h| \leq L_x|h|(1 + |z|)(1 + |x| + |y|)^m,$$

for every $x, h \in H$, $y \in \mathbb{R}$, $z \in \Xi^*$.

We have the following existence and uniqueness result (in the weighted spaces introduced above).

Proposition 6.60 *Assume that Hypothesis 6.57 holds and that F satisfies the conditions in Hypothesis 6.59. For $p > 2$, β and q satisfying (6.139), and for every $\lambda > \widehat{\lambda} = -(\beta + \mu - L_z^2/2)$, the following holds.*

(i) For every $x \in H$ there exists a unique solution $(X(\cdot;x), Y(\cdot;x), Z(\cdot;x))$ of the forward–backward system (6.138) such that $X(\cdot;x) \in \mathcal{H}^q_{\eta(q)}$ and $(Y(\cdot;x), Z(\cdot;x)) \in \mathcal{K}^p_\beta$ (here $K = \mathbb{R}$ and consequently $\mathcal{L}_2(\Xi, K)$ is Ξ^*). Moreover, $Y(\cdot;x) \in L^p_{\mathcal{P}}(\Omega; C_\beta(\mathbb{R}))$.

(ii) *The maps* $x \to X(\cdot; x)$, $x \to (Y(\cdot; x), Z(\cdot; x))$, $x \to Y(\cdot; x)$ *belong to the spaces* $\mathcal{G}^1(H, \mathcal{H}^q_{\eta(q)})$, $\mathcal{G}^1(H, \mathcal{K}^p_\beta)$ *and* $\mathcal{G}^1(H, L^p_{\mathcal{P}}(\Omega; C_\beta(\mathbb{R})))$, *respectively.*

(iii) *Setting* $u(x) = Y(0; x)$, *we have* $u \in \mathcal{G}^1(H, \mathbb{R})$, *and* u *and* ∇u *have polynomial growth. More precisely, there exists a constant* $C > 0$ *such that*

$$|u(x)| \le C\,(1 + |x|)^{m+1}, \qquad |\nabla u(x)h| \le C\,|h|(1 + |x|)^{[(m+1)^2]}, \qquad x, h \in H.$$

Remark 6.61 Notice that we have shown that the system (6.138) admits a unique solution (in suitable spaces $\mathcal{H}^q_{\eta(q)}, \mathcal{K}^p_\beta$ with parameters satisfying $p > 2$ and condition (6.139)) for all $\lambda > \widehat{\lambda}$ where

$$\widehat{\lambda} = -\mu + L_z^2/2 - \sup\{\eta(q)(m+1)(m+2) \wedge 0 : q > 2(m+1)(m+2)\}. \tag{6.140}$$

■

Remark 6.62 If, in addition to Hypothesis 6.59, we suppose that $F(\cdot, 0, 0)$ is bounded and satisfies Hypothesis 6.59 with $m = 0$, then the above results can be improved in the following way. Instead of asking (6.139) it is enough to require: $q > p > 2$ and $\beta < \eta(q) \wedge 0$. Then the conclusions of Proposition 6.60 still hold for $\lambda > -(\beta + \mu - L_z^2/2)$. Thus instead of (6.140) we have

$$\widehat{\lambda} = -\mu + L_z^2/2 - \sup\{\eta(q) \wedge 0 : q > 2\}. \tag{6.141}$$

Moreover, we have $|u(x)| \le C$ and $|\nabla_x u(x)h| \le C|h|$ for all $x, h \in H$. ■

Assuming that Hypothesis 6.57 holds and denoting by $(X(\tau; x))_{\tau \ge 0}$ the solution of Eq. (6.133), we define in the usual way the transition semigroup $(P_t)_{t \ge 0}$, associated to the process X:

$$P_t[\phi](x) = \mathbb{E}\,\phi(X(t; x)), \qquad x \in H, \tag{6.142}$$

for every bounded measurable function $\phi : H \to \mathbb{R}$. By Proposition 6.57, ϕ can be taken unbounded, with polynomial growth. Formally, the generator $\mathcal{A}$ of (P_t) is the operator

$$\mathcal{A}\phi(x) = \frac{1}{2}\mathrm{Tr}\left(\sigma(x)\sigma(x)^* D^2\phi(x)\right) + \langle Ax + b(x), D\phi(x)\rangle\,.$$

We consider now the solvability of the nonlinear stationary Kolmogorov equation:

$$\mathcal{A}u(x) - \lambda\, u(x) = F(x, u(x), \nabla u(x)\,\sigma(x)), \qquad x \in H, \tag{6.143}$$

where the function $F : H \times \mathbb{R} \times \Xi^* \to \mathbb{R}$ satisfies the conditions in Hypothesis 6.59 (with $K = \mathbb{R}$) and λ is a given number (that will eventually be assumed to be large enough). Note that, for $x \in H$, $\nabla u(x)$ belongs to H^*, so that $\nabla u(x)\,\sigma(x)$ is in Ξ^*.

The definition of a mild solution has to be slightly modified in order to take into account the polynomial growth:

Definition 6.63 We say that a function $u : H \to \mathbb{R}$ is a mild solution of the nonlinear stationary Kolmogorov equation (6.143) if the following conditions hold:

(i) $u \in \mathcal{G}^1(H, \mathbb{R})$;
(ii) for all $x \in H$, $h \in H$, we have

$$|u(x)| \leq C\ (1 + |x|)^C, \qquad |\nabla_x u(x)h| \leq C\ |h|\ (1 + |x|)^C,$$

for some constant $C > 0$;
(iii) the following equality holds, for every $x \in H$ and $T \geq 0$:

$$u(x) = e^{-\lambda T}\ P_T[u](x) - \int_0^T e^{-\lambda\tau}\ P_\tau\Big[F\Big(\cdot, u(\cdot), \nabla u(\cdot)\, \sigma(\cdot)\Big)\Big](x)\ d\tau. \tag{6.144}$$

Together with Eq. (6.133) we again consider the backward equation for $0 \leq \tau \leq T < \infty$

$$\begin{aligned} Y(\tau; x) - Y(T; x) + \int_\tau^T Z(s; x))\, dW(s) + \lambda \int_\tau^T Y(s; x)\, ds \\ = - \int_\tau^T F(X(s; x), Y(s; x), Z(s; x))\, ds, \end{aligned} \tag{6.145}$$

where $F : H \times \mathbb{R} \times \Xi^* \to \mathbb{R}$ and λ are the same occurring in the nonlinear stationary Kolmogorov equation. Under the stated assumptions, Proposition 6.60 gives a unique solution $\{(X(\tau; x), Y(\tau; x), Z(\tau; x))\}_{\tau \geq 0}$ of the forward–backward system (6.138).

We can now state one of our main results.

Theorem 6.64 *Assume that Hypothesis 6.57 holds and that F satisfies the conditions in Hypothesis 6.59.*

Then there exists a $\widehat{\lambda} \in \mathbb{R}$ such that, for every $\lambda > \widehat{\lambda}$, the nonlinear stationary Kolmogorov equation (6.143) has a unique mild solution. The solution u is given by the formula

$$u(x) = Y(0; x), \tag{6.146}$$

where $\{(X(\tau; x), Y(\tau; x), Z(\tau; x))\}_{\tau \geq 0}$ is the solution of the backward-forward system 6.138), and it satisfies

$$|u(x)| \leq C\ (1 + |x|)^{m+1}, \qquad |\nabla u(x)h| \leq C\ |h|(1 + |x|)^{[(m+1)^2]},$$

for some constant C and every $x, h \in H$.

Remark 6.65 The constant $\widehat{\lambda}$ in the statement of the theorem can be chosen equal to (6.140). ■

Remark 6.66 From Remark 6.62 it follows immediately that if, in addition to Hypothesis 6.57 and 6.59, we assume that $F(\cdot, 0, 0)$ is bounded and F satisfies Hypothesis 6.59 with $m = 0$, then $\widehat{\lambda}$ can be chosen equal to (6.141) instead of (6.140). Moreover, in this case, we have $|u(x)| \le C$, $|\nabla u(x)h| \le C\ |h|$ for some constant C and every $x, h \in H$. ∎

Finally, we again apply the above results to a control problem. We mainly wish to show here what frameworks can be covered.

Let again H and Ξ denote real separable Hilbert spaces (the state space and the noise space, respectively) and let Λ be a Polish space (the control space). For $t \in [0, +\infty)$ a *generalized reference probability space* is given by $\mu = (\Omega, \mathscr{F}, \mathscr{F}_s, \mathbb{P}, W)$, where

- $(\Omega, \mathscr{F}, \mathbb{P})$ is a complete probability space;
- $(\mathscr{F}_s)_{s\ge 0}$ is a filtration in it, satisfying the usual conditions;
- $(W(s))_{s\ge 0}$ is a cylindrical $\mathbb{P}$-Wiener process in Ξ, with respect to the filtration $\mathscr{F}_s$.

Given such μ, for every starting point $x \in H$ we will consider the following controlled state equation

$$\begin{cases} dX(s;x) = (AX(s;x) + b(X(s;x)) + \sigma(X(s;x))R(X(s;x), a(s)))\ ds \\ \qquad\qquad\qquad\qquad\qquad\qquad + \sigma(X(s;x))\, dW(s),\ \ s \in [0,\infty), \\ X(0) = x \in H. \end{cases} \tag{6.147}$$

In (6.147) and below the equation is understood in the mild sense. $a(\cdot) : \Omega \times [0, +\infty) \to \Lambda$ is the control process, which is always assumed to be progressively measurable with respect to $\{\mathscr{F}_s\}_{s\ge 0}$. On the coefficients A, b, σ, R precise assumptions will be formulated in Hypothesis 6.67 below. As in Sect. 6.5 we will impose on R only measurability and boundedness assumptions. As mentioned, this requires some care in the formulation of the control problem. We again call $(a(\cdot), X(\cdot))$ an *admissible control pair* if $a(\cdot)$ is an $\mathscr{F}_s$-progressively measurable process with values in Λ and $X(\cdot)$ is a mild solution to (6.147) corresponding to $a(\cdot)$. To every admissible control pair we associate the cost:

$$J^{\mu}(x; a(\cdot), X(\cdot)) = \mathbb{E} \int_0^{\infty} e^{-\lambda s} l(X(s;x), a(s))\, ds,$$

where l is a suitable real function. As in the parabolic case, see Sect. 6.5, the optimal control problem in the extended strong formulation consists in minimizing the functional $J^{\mu}(x; a(\cdot), X(\cdot))$ over all admissible control pairs (a, X), and characterizing the value function

$$V^{\mu}(x) = \inf_{(a(\cdot), X(\cdot))} J^{\mu}(x; a(\cdot), X(\cdot; x)).$$

We will also address the optimal control problem in the extended weak formulation, which consists in further minimizing with respect to all generalized reference probability spaces, i.e., in characterizing the value function

$$\overline{V}(x) = \inf_{\mu} V^{\mu}(x).$$

The corresponding Hamiltonian function is defined for all $x \in H$, $z \in \Xi^*$ setting

$$F_0(x, z) = \inf_{a \in \Lambda} \left(l(x, a) + z\, R(x, a) \right). \tag{6.148}$$

We also define as usual

$$\Gamma(x, z) = \{a \in \Lambda : F_0(x, z) = l(x, a) + z\, R(x, a)\}.$$

We make the following assumption.

Hypothesis 6.67 The following holds:

(1) A, b and σ satisfy Hypothesis 6.57.
(2) $R : H \times \Lambda \to \Xi$ is Borel measurable and $|R(x, a)|_{\Xi} \leq L_R$ for a suitable constant $L_R > 0$ and all $x \in H$, $a \in \Lambda$.
(3) $l : H \times \Lambda \to \mathbb{R}$ is continuous and satisfies $|l(x, u)| \leq K_l(1 + |x|^{m_l})$ for suitable constants $K_l > 0$, $m_l \geq 0$ and all $x \in H$, $u \in \Lambda$.
(4) F_0 belongs to $\mathcal{G}^1(H \times \Lambda^*, \mathbb{R})$ and satisfies Hypothesis 6.59 (to avoid confusion we denote by m_F the constant m introduced in Hypothesis 6.59) We also notice that by its definition F_0 is Lipschitz with respect to z with Lipschitz constant L_R.
(5) Finally, we fix here $p > 2$, q and β satisfying (6.139) with $m = m_F$, and such that $q > m_F$.

In the following $\eta(q)$ is the constant introduced in Proposition 6.58.

Lemma 6.68 *Assume that* $\lambda > 0$ *satisfies*

$$\lambda > \frac{L_R m_l}{2(q - m_l)} - \eta(q)\, m_l. \tag{6.149}$$

Then the cost functional is well defined and $J(x_0; a(\cdot), X(\cdot)) < \infty$ *for all* $x_0 \in H$ *and all generalized reference probability spaces.*

By Theorem 6.64, for all $\lambda > \widehat{\lambda}$ (the constant $\widehat{\lambda}$ can be chosen equal to (6.140) with $L_z = L_R$) the stationary Hamilton–Jacobi–Bellman equation relative to the above stated problem, written formally as

$$\mathcal{A}v(x) = \lambda v(x) + F_0(x, \nabla v(x)\sigma(x)), \qquad x \in H, \tag{6.150}$$

admits a unique mild solution, in the sense of Definition 6.63, which we will denote by v.

We are in a position to solve the control problem:

Theorem 6.69 *Assume Hypothesis 6.67 and suppose that λ satisfies:*

$$\lambda > \left(-\beta + \frac{L_R^2}{2}\right) \vee \left(-\beta + \frac{L_R}{2(p-1)}\right) \vee \left(\frac{L_R m_l}{2(q-m_l)} - \eta(q)\, m_l\right). \quad (6.151)$$

Then the following holds

(1) For all generalized reference probability space μ and all admissible control pairs $(a(\cdot), X(\cdot))$ we have $J^\mu(x; a(\cdot), X(\cdot)) \geq v(x)$.
It follows that $V^\mu(x) \geq v(x)$ for every μ, and so $\overline{V}(x) \geq v(x)$.

(2) For all μ and all admissible control pairs (a, X), the equality $J^\mu(x; a(\cdot), X(\cdot)) = v(x)$ holds if and only if the following feedback law is satisfied:

$$a(s) \in \Gamma(X(s), \nabla_x v(X(s))\, \sigma(X(s))), \quad \mathbb{P}\text{-a.s. for a.a. } s \in [t, T]. \quad (6.152)$$

We again have existence of the optimal control in the extended weak formulation.

Theorem 6.70 *If in addition to the assumptions of the above theorem we suppose that $\Gamma(x, z)$ is non-empty for all $x \in H$ and $z \in \Xi^*$. Let $\gamma : H \times \Xi^* \to \Lambda$ be a measurable selection of Γ (which exists, see Theorem 8.2.10, in [20]). Then there exists at least one generalized reference probability space $\overline{\mu}$ and an admissible control pair $(\overline{a}(\cdot), \overline{X}(\cdot))$ for which (6.152) holds. In particular, it follows that $V_t^{\overline{\mu}}(x) = v(t, x)$ and so $\overline{V}(t, x) = v(t, x)$. In the space $\overline{\mu}$ the process $\overline{X}$ is a mild solution to the closed loop equation:*

$$\begin{cases} d\overline{X}(s) = A\overline{X}(s)\, ds + \sigma(\overline{X}(s))\, R\Big(\overline{X}(s),\ \gamma(s, \overline{X}(s), \nabla_x v(\overline{X}(s))\, \sigma(\overline{X}(s)))\Big)\, ds \\ \qquad\qquad +b(\overline{X}(s))\, ds + \sigma(\overline{X}(s))\, d\overline{W}(s), \qquad s \in [t, T], \\ \overline{X}(0) \ = x \in H, \end{cases} \quad (6.153)$$

the feedback law takes the form

$$\overline{a}(s) = \gamma(\overline{X}(s), \nabla_x v(\overline{X}(s))\, \sigma(\overline{X}(s))), \qquad \mathbb{P}\text{-a.s. for a.e. } s \in [0, T],$$

and the pair $(\overline{a}(\cdot), \overline{X}(\cdot))$ is optimal for the control problem in the extended weak formulation.

Remark 6.71 If, in addition to points *1–4* of Hypothesis 6.67, we also assume that l is bounded and Lipschitz in x uniformly in $u \in \mathcal{U}$, then it is easily verified that $F_0(\cdot, 0)$ is bounded and F_0 satisfies Hypothesis 6.59 with $m = 0$. Thus by Remark 6.62 the results of Theorem 6.69 can be improved in the following way.

Instead of Hypothesis 6.67 point *5* it is enough to take $q > p > 2$ and $\beta < \eta(q) \wedge 0$. Moreover, instead of (6.151) it is enough to assume

$$\lambda > -\beta + \left(\frac{L_R^2}{2} \vee \frac{L_R}{2(p-1)} \right).$$

■

6.12.1 The Heat Equation with Multiplicative Noise

Finally, we show how the assumptions on the controlled heat equation in Sect. 6.11 have to be adapted to fit this last framework. We again consider a stochastic heat equation with additive white noise in dimension 1 (for $t \geq 0$, $\xi \in [0, 1]$):

$$\begin{cases} \dfrac{\partial}{\partial t} x(t,\xi) = \dfrac{\partial^2}{\partial \xi^2} x(t,\xi) + f_0(\xi, x(t,\xi)) \\ \qquad\qquad\qquad\qquad\qquad +\sigma_0(\xi, x(t,\xi)) r(\xi)\, a(t,\xi) + \sigma_0(\xi)\, \frac{\partial}{\partial t} \mathcal{W}(t,\xi), \\ x(t,0) = x(t,1) = 0, \\ x(0,\xi) = x_0(\xi), \end{cases} \tag{6.154}$$

and the cost functional:

$$J(x_0; a(\cdot), x(\cdot)) = \mathbb{E} \int_0^\infty \int_0^1 e^{-\lambda t} \left[\ell_0(\xi, x(t,\xi)) + |a(t,\xi)|^2 \right] d\xi\, dt. \tag{6.155}$$

The assumptions and notations are the same as in Sect. 6.11 except that:

- σ_0 depends on x as well. We assume that it is bounded, differentiable with respect to x and Lipschitz with respect to x, uniformly in ξ.
- We relax the assumptions on ℓ_0. Namely, we assume that ℓ_0 is defined on $[0, 1] \times \mathbb{R}$. Moreover, for a.a. $\xi \in [0, 1]$, the map $\ell_0(\xi, \cdot)$ is in $C^1(\mathbb{R}, \mathbb{R})$ and

$$|\ell_0(\xi, 0)| \leq c_0(\xi), \quad \left| \frac{\partial}{\partial \eta} \ell_0(\xi, \eta) \right| \leq c_1(\xi), \quad \text{with} \int_0^1 \left(c_0(\xi) + c_1^2(\xi) \right) d\xi < +\infty. \tag{6.156}$$

- We restrict our analysis to controls taking values in a ball of $L^2([0, 1])$. Namely, we assume:

$$\int_0^1 a^2(t,\xi) d\xi \leq 1, \quad \mathbb{P} \otimes dt\text{-a.s.}$$

The problem can be rewritten in the abstract way exactly as in Sect. 6.11 with the difference that now:

$$\inf_{a \in H: |a| \leq 1} (|a|_H^2 + z(Ra)) = \inf_{a \in H: |a| \leq 1} (|a|_H^2 + (R^* z) a) = \Psi(R^* z),$$

where (with the standard identifications):

$$\Psi(p) = \begin{cases} -(1/4)|p|^2_{L^2([0,1])} & \text{if } |p|_{L^2([0,1])} \le 2 \\ -|p| + 1 & \text{if } |p|_{L^2([0,1])} > 2 \end{cases}, \qquad R^* z = rz.$$

In addition, $\inf_{a \in H}(|a|^2_H + z(Ra))$ is a minimum achieved for $a = \psi(R^* z)$ where

$$\psi(p) = \begin{cases} -(1/2)p & \text{if } |p|_{L^2([0,1])} \le 2, \\ -p/|p| & \text{if } |p|_{L^2([0,1])} > 2. \end{cases}$$

So $F_0(x, z) = l(x) + \Psi(R^* z)$ belongs to $\mathcal{G}^1(H \times H^*, \mathbb{R})$. As a consequence of Theorems 6.69 and 6.70 we have the following result.

Theorem 6.72 *Under the previous assumption we can find $\hat{\lambda}$ such that, for all $\lambda > \hat{\lambda}$, there exists at least one generalized reference probability space $\overline{\mu} = (\overline{\Omega}, \overline{\mathscr{F}}, \overline{\mathscr{F}}_s, \overline{\mathbb{P}}, \overline{W})$ and an admissible control pair $(\overline{a}(\cdot), \overline{x}(\cdot))$ for which*

$$\overline{V}(x_0) = J^{\overline{\mu}}(x_0; \overline{a}(\cdot), \overline{x}(\cdot)), \qquad x_0 \in L^2([0, 1]).$$

In particular, the triple $(\overline{\mu}, \overline{a}(\cdot), \overline{x}(\cdot))$ is optimal.

The value function $\overline{V}(x_0)$ coincides with the function $v(x_0)$, which is the unique mild solution to the Hamilton–Jacobi–Bellman equation (6.150) in the sense specified by Definition 6.63 (see Theorem 6.64) where (with the standard identifications)

$$F_0(x, \nabla v \sigma) = l(x) + \Psi(R^* \nabla v(x) \sigma(x)) = \ell_0(\cdot, x(\cdot)) + \Psi(r(\cdot)\sigma_0(\cdot, x(\cdot))\nabla v(x)(\cdot)).$$

In the space $\overline{\mu}$ the process $(\overline{x}(s, \cdot))_{s \ge 0}$ is a mild solution to the closed loop equation

$$\begin{cases} \dfrac{\partial}{\partial t}\bar{x}(t, \xi) = \dfrac{\partial^2}{\partial \xi^2}\bar{x}(t, \xi) + f_0(\xi, \bar{x}(t, \xi)) + \sigma_0(\xi, \bar{x}(\xi))\, \dfrac{\partial}{\partial t}\mathcal{W}(t, \xi) \\ \qquad\qquad + \sigma_0(\xi, \bar{x}(\xi)) r(\xi) \psi\left(r(\cdot)\sigma_0(\cdot, \bar{x}(t, \cdot))\nabla v(\bar{x}(t, \cdot))(\cdot)\right)(\xi) dt, \\ \bar{x}(t, 0) = \bar{x}(t, 1) = 0, \\ \bar{x}(0, \xi) = x_0(\xi), \end{cases}$$

and the optimal pair $(\overline{a}(t, \cdot), \overline{x}(t, \cdot))$ satisfies the feedback law equality

$$\overline{a}(t, \cdot) = \psi\left(r(\cdot)\sigma_0(\cdot, \bar{x}(t, \cdot))\nabla v(\bar{x}(t, \cdot))(\cdot)\right).$$

6.13 Bibliographical Notes

The paper [475] by É. Pardoux and S. Peng is generally recognized as the starting point of the theory of Backward Stochastic Differential Equations (BSDEs): there the authors solved a general nonlinear BSDE under Lipschitz assumptions on the coefficients. Earlier results on the linear case were proved by several authors, in particular by J-.M. Bismut and A. Bensoussan, in connection with the so-called Stochastic Maximum Principle (in the sense of Pontryagin). Since the appearance of [475], the theory began to develop quickly, motivated by applications to stochastic optimal control, partial differential equations and mathematical finance. Some standard references are [211, 420, 477, 575].

Here we limit ourselves to a bibliographical account of the main achievements related to BSDEs driven by a Brownian motion in an infinite-dimensional context, i.e., when at least one of the unknown processes (Y, Z) takes values in an infinite-dimensional space or when the BSDE is coupled with another (forward) stochastic differential equation with infinite-dimensional solution process.

To our knowledge, the first result on BSDEs when the process Y evolves in an infinite-dimensional space is that of Bensoussan [45] concerning the linear case. A highly non-trivial extension of the nonlinear case originally addressed by Pardoux and Peng in the infinite-dimensional context is in [350], followed by [558] and by some results in [284, 285]. The case of dissipative coefficients is considered in [129, 130]. A special class of backward equations, called of Volterra type, are studied in the Hilbert space case in [11, 12].

The Stochastic Maximum Principle, which is not treated in this chapter, remains one of the main sources of interest for studying BSDEs with infinite-dimensional process Y. Although the equation is linear in this case, the occurrence of unbounded coefficients often makes the study technically challenging. After the reference [45] already mentioned, the papers [196, 349] treat the maximum principle for a general controlled evolution equation in a Hilbert space. Applications to concrete controlled stochastic PDEs can be found in [598] for equations linear in the state, and in [280]. The case of a controlled stochastic PDE with additive noise and dissipative drift is treated in [282]. The treatise [414] is entirely devoted to the Stochastic Maximum Principle in infinite dimension.

A special mention is deserved for the study of the stochastic backward Hamilton–Jacobi–Bellman equation, introduced in [481] and further studied in [85]. Representation formulae for equations of similar type are proved in [549].

Many other cases of concrete stochastic PDEs of backward type have been studied, as objects of intrinsic interest and not necessarily related to stochastic optimal control problems, see for instance [197–199, 348, 419, 421, 504, 505, 552], and the subject is developing quickly.

Very often a scalar BSDE (i.e., where the process Y is real-valued) is introduced, coupled with a forward equation representing the dynamics of a controlled process evolving in an infinite-dimensional case, driven by a finite- or infinite-dimensional Brownian motion. This is the situation addressed in this chapter. As seen above, the process Y is then related to the value function of the optimal control problem and, in the Markovian case, it is used to represent or to construct a solution (in an appropriate sense) to the corresponding Hamilton–Jacobi–Bellman (HJB) equation. The first systematic study of this type for controlled stochastic equations in Hilbert space is in [284–286]. More general coefficients (for instance, of dissipative type), or more general growth conditions, were studied in [75–77, 351], see also [593–595].

Often, better results are obtained by a combination of probabilistic arguments on the BSDE and an analytic study of the HJB equation, as in [432, 433]. In [435, 438, 442] very general Hamiltonians are addressed. Smoothing effects of the HJB equation, due to a nondegenerate diffusion coefficient of the controlled equation, were studied in [283, 440].

The case of linear controlled evolution equations and quadratic cost also lead to stochastic backward equations of Riccati type, when the coefficients are perturbed by noise. In the infinite-dimensional framework we cite [333–335, 414, 415].

Applications to models with delay or memory effects can be found in several of the previous references. Memory effects are explicitly studied by BSDE techniques for the heat equation in [131] and for controlled stochastic Volterra equations in [63, 132]. Related results can be found in [600].

A special branch of the literature is devoted to the case when the controlled equation is a stochastic PDE with Brownian noise acting on the boundary conditions, often in combination with a control process on the boundary as well. We mention [181, 332, 437, 591, 592]. We also cite [331] for a version of the Stochastic Maximum Principle in this framework and [62] for the related case of dynamical boundary conditions.

Although in the large majority of the mentioned papers the state space is a Hilbert space, there are a few papers related to extensions to Banach space-valued processes: see [281, 436, 596].

BSDEs can be used to address other stochastic optimization problems, even when the controlled systems evolves in an infinite-dimensional space. In [182, 278] ergodic optimal control problems are studied, whereas applications of BSDEs to the theory of stochastic differential games are given in [274, 275], where games with an infinite number of players are considered.

More specific topics are treated in [273] (connections with conditioned processes in Hilbert spaces) and [330] (strongly coupled infinite-dimensional forward–backward systems, i.e., when the forward equations depends on the unknown pair (Y, Z) solution to the backward equation).