SIAM J. CONTROL OPTIM.
Vol. 48, No. 7, pp. 4624–4651

STOCHASTIC EQUATIONS WITH DELAY: OPTIMAL CONTROL VIA BSDEs AND REGULAR SOLUTIONS OF HAMILTON–JACOBI–BELLMAN EQUATIONS*

MARCO FUHRMAN†, FEDERICA MASIERO‡, AND GIANMARIO TESSITORE‡

Abstract. We consider an Itô stochastic differential equation with delay, driven by Brownian motion, whose solution, by an appropriate reformulation, defines a Markov process X with values in a space of continuous functions $\mathbf{C}$, with generator $\mathcal{L}$. We then consider a backward stochastic differential equation depending on X, with unknown processes (Y, Z), and we study properties of the resulting system, in particular we identify the process Z as a deterministic functional of X. We next prove that the forward-backward system provides a suitable solution to a class of parabolic partial differential equations on the space $\mathbf{C}$ driven by $\mathcal{L}$, and we apply this result to prove a characterization of the fair price and the hedging strategy for a financial market with memory effects. We also include applications to optimal stochastic control of differential equation with delay: in particular we characterize optimal controls as feedback laws in terms of the process X.

Key words. stochastic delay differential equations, backward stochastic differential equations, quadratic variation, optimal stochastic control

AMS subject classifications. 60H30, 34K50, 60H07, 93E20

DOI. 10.1137/080730354

1. Introduction.

We will consider stochastic differential equations with delay (SDDEs) on a finite interval of the form

$$\begin{cases} dy_t = b(t, y_{t+\cdot})\, dt + \sigma(t, y_{t+\cdot})\, dW_t, \qquad t \in [0, T], \\ y_\theta = x(\theta),\ \theta \in [-r, 0] \end{cases} \tag{1.1}$$

for an unknown process $(y_t)_{t\in[-r,T]}$ in $\mathbb{R}^n$. Here $r > 0$ is the maximum delay taken into account, and we use the notation $y_{t+\cdot} = (y_{t+\theta})_{\theta\in[-r,0]}$. It is customary (see, for instance, [14]) and convenient to introduce the space $\mathbf{C} = C([-r, 0]; \mathbb{R}^n)$ and the $\mathbf{C}$-valued process $X = (X_t)_{t\in[0,T]}$ defined by

$$X_t(\theta) = y_{t+\theta}, \qquad \theta \in [-r, 0].$$

With this notation, $b(t, \cdot)$ and $\sigma(t, \cdot)$ are functions defined on $\mathbf{C}$ and the equation can be written

$$\begin{cases} dy_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \quad t \in [0, T], \\ X_0 = x \in \mathbf{C}. \end{cases}$$

SDDEs are a classical subject: in the standard reference book [18] (see also [19]) basic results are established: existence and uniqueness of solutions, regular dependence on parameters, Markov property of X as a $\mathbf{C}$-valued process, characterization of its generator. In [7] long time asymptotics are studied in detail.

*Received by the editors July 16, 2008; accepted for publication (in revised form) April 28, 2010; published electronically August 17, 2010.
http://www.siam.org/journals/sicon/48-7/73035.html

†Politecnico di Milano, Dipartimento di Matematica, piazza Leonardo da Vinci 32, 20133 Milano, Italy (marco.fuhrman@polimi.it).
‡Dipartimento di Matematica e Applicazioni, Università di Milano Bicocca, via Cozzi 53, 20125 Milano, Italy (federica.masiero@unimib.it, gianmario.tessitore@unimib.it).

In this paper we will present new results on optimal control problems for SDDEs. Moreover, since the Markov character of solutions allows for the application of dynamic programming arguments, we will also prove new results on the corresponding Hamilton–Jacobi–Bellman equation. More generally, we will consider a class of semilinear versions of the parabolic Kolmogorov equation associated to the process X. This class includes, as a very special case, some infinite-dimensional variants of the Black–Scholes equation for the fair price of an option, of great interest in mathematical finance and already considered in [5].

The main tool will be the use of techniques from the theory of backward stochastic differential equations (BSDEs) in the sense of Pardoux–Peng, first considered in the nonlinear case in the paper [24]. We refer to the monographs [8], [23] for an exposition of the basic theory. The BSDE approach that we follow consists of addressing (1.1), but with generic initial values $t \in [0,T]$ and $x \in \mathbf{C} = C([-r,0];\mathbb{R}^n)$, and then coupling with another equation of backward type, with unknown processes (Y,Z). More precisely, one considers the forward-backward system

$$\text{(1.2)} \qquad \begin{cases} dy_\tau^{t,x} = b(\tau, X_\tau^{t,x})\, d\tau + \sigma(\tau, X_\tau^{t,x})\, dW_\tau, \quad \tau \in [t,T] \subset [0,T], \\ X_t^{t,x} = x, \\ dY_\tau^{t,x} = \psi(\tau, X_\tau^{t,x}, Y_\tau^{t,x}, Z_\tau^{t,x})\, d\tau + Z_\tau^{t,x}\, dW_\tau, \\ Y_T^{t,x} = \phi(X_T^{t,x}), \end{cases}$$

where $\psi : [0,T] \times \mathbf{C} \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ and $\phi : \mathbf{C} \to \mathbb{R}$ are given functions. One can then define a (deterministic) function $v : [0,T] \times \mathbf{C} \to \mathbb{R}$ setting $v(t,x) = Y_t^{t,x}$. Markovianity of system (1.2) immediately yields that $Y_\tau^{t,x} = v(\tau, X_\tau^{t,x})$. In addition we prove that

$$\text{(1.3)} \qquad Z_\tau^{t,x} = \nabla_0 v\left(\tau, X_\tau^{t,x}\right) \sigma(\tau, X_\tau^{t,x}),$$

where ∇_0 is a differential operator defined by

$$\text{(1.4)} \qquad \nabla_0 v(t,x) = \nabla_x v(t,x)(\{0\}).$$

To explain the above expression we recall that the gradient $\nabla_x v(t,x)$ at point $(t,x) \in [0,T] \times \mathbf{C}$ is an element of the dual space $\mathbf{C}^*$, hence an n-tuple of finite Borel measures on $[-r,0]$. Thus, $\nabla_0 v(t,x)$ is a vector in $\mathbb{R}^n$ whose components are the masses at point 0 of the components of $\nabla_x v(t,x)$. We stress the fact that, as it is customary when relating BSDEs and PDEs (see, for instance, [25]), the above "identification of Z" is one of the main technical points of this paper. The proof presented in section 3 is performed by computing the joint quadratic variation of $(v(\tau, X_\tau^{t,x}))_{\tau\in[t,T]}$ and $(W_\tau)_{\tau\in[t,T]}$, using Malliavin calculus techniques.

We are then able to prove that v is the unique solution (in a suitable mild sense) of a semilinear parabolic equation of the form

$$\text{(1.5)} \qquad \begin{cases} \dfrac{\partial v(t,x)}{\partial t} + \mathcal{L}_t v(t,x) = \psi(t,x,v(t,x), \nabla_0 v(t,x)\, \sigma(t,x)), \\ v(T,x) = \phi(x), \qquad t \in [0,T],\ x \in \mathbf{C}, \end{cases}$$

where $\mathcal{L}_t$ is the generator of the Markov process $(X_\tau^{t,x})$ (see [18], [19] or our Remark 2.4).

If one considers the controlled SDDE

$$\text{(1.6)} \qquad \begin{cases} dy_s^u = b(s, X_s^u)\, ds + \sigma(s, X_s^u)\, [h(s, X_s^u, u_s)\, ds + \, dW_s], \quad s \in [t,T], \\ X_t = x, \end{cases}$$

where the solution depends on a control process $u(\cdot)$ taking values in a space U, and $h : [0,T] \times \mathbf{C} \times U \to \mathbb{R}^d$ is given, and one tries to minimize a cost functional

$$J(t,x,u(\cdot)) = \mathbb{E}\int_t^T g(u_s)\, ds + \mathbb{E}\,\phi(X_T^u), \tag{1.7}$$

where $g : U \to [0,\infty)$, then (1.5) is the associated Hamilton–Jacobi–Bellman equation, provided the Hamiltonian function $\psi : [0,T] \times \mathbf{C} \times \mathbb{R}^d \to \mathbb{R}$ is defined by the formula

$$\psi\,(t,x,z) = \inf\left\{g(u) + zh\,(t,x,u) : u \in U\right\}, \qquad t \in [0,T], x \in \mathbf{C}, z \in \mathbb{R}^d.$$

This way we eventually prove that v coincides with the value function of the control problem and that $\nabla_0 v$ occurs in the construction of the optimal feedback.

Although BSDEs were known to be useful tools in the study of control problems and nonlinear partial differential equations, applications to infinite-dimensional state spaces are more recent and difficult; see, e.g., [11], [12] for the case of a Hilbert space, and [17] for some related results on Banach spaces. In these papers, as well as in the present one, the solution of (1.5) is understood in the so-called mild sense. Special difficulties arise if the space $\mathbf{C}$ is used as the state space of the basic stochastic process X. In particular, even in the deterministic case, it is not clear how to formulate the state equation as an evolution equation in $\mathbf{C}$; see [14]. The reason for choosing to work in $\mathbf{C}$ is to allow for great generality on the coefficients b, σ of the SDDEs as well as on the cost functional of the control problem. For instance, the functional ϕ occurring in (1.7) could have the form

$$\phi(z) = \int_{[-r,0]} g(z(\theta))\, \mu(d\theta), \qquad z \in \mathbf{C}, \tag{1.8}$$

for some $g \in C^1(\mathbb{R})$ and some (finite signed) measure μ on $[-r,0]$. The special case when μ is supported on a finite number of points is of particular interest and could be studied by direct methods, but it is included in our results. More generally, if $\phi_1, \ldots, \phi_n$ are functionals with the form (1.8) corresponding to functions $g_1, \ldots, g_n \in C^1(\mathbb{R})$, and if $h \in C^1(\mathbb{R}^n)$, then the functional $\phi(z) = h(\phi_1(z), \ldots, \phi_n(z))$, can also be treated by our methods. One could avoid the use of the space $\mathbf{C}$ by looking at X as a process with values in the space $L^2([0,T];\mathbb{R}^n)$ instead. This was the approach taken in [13]. However, this leads to restrictions on the applicability of the corresponding results.

Optimal control problems for SDDEs have been thoroughly investigated in recent years. The book [4] is a systematic exposition of the state of the existing theory in all aspects and contains an extensive bibliography. Typically, one of the main achievements of optimal control theory is the characterization of the value function as the unique viscosity solution of a Hamilton–Jacobi–Bellman equation. Unfortunately, the proof of uniqueness reported in [4] seems to contain a gap (see the inequality at the bottom of page 175 as well as the subsequent arguments) and, therefore, we prefer not to rely on this result. In fact, we do not prove any result in the framework of viscosity solutions. Indeed, in our paper (see section 6) we assume stronger conditions, namely a special form for the control system (1.6) and differentiability assumptions on the data b, σ, ϕ, ψ with respect to the space variable $x \in \mathbf{C}$. Under these assumptions we consider a different notion of solution and are able to prove that a unique solution exists and has further properties, in particular differentiability. We note that the existence of the gradient of v is of special interest in optimal control theory, since not only does it occur in the (mild) formulation of the corresponding Hamilton–Jacobi–Bellman equation, but it also allows us to characterize the optimal controls via feedback laws and to prove existence of optimal controls after appropriate formulation.

Parabolic equations on the space $\mathbf{C}$ of the form (1.5) have also been considered for other purposes, in particular as an infinite dimensional generalization of the Black–Scholes equation for the fair price of an option, in case the market models or the claim exhibit memory effects. In [4] the existence and uniqueness result for a viscosity solution is stated but the proof has the same problems as in the case of the Hamilton–Jacobi–Bellman equation. In [5], by a verification theorem, it is shown that if the price of the claim is once differentiable in time and twice in space and, in addition, for all times it belongs to the domain of the generator of the shift operator $\mathcal{S}$ (see Remark 2.4), then it solves the generalized Black–Scholes equation in a classical way. Here, see section 7, we prove that if the coefficients in the market and the claim are differentiable, then the price is the unique mild solution of the generalized Black–Scholes equation. Moreover, the special operator ∇_0 defined in (1.4) occurs in the construction of the hedging strategy. Although more natural than in [5], admittedly our assumptions (in particular differentiability of the claim) are not totally satisfactory for applications. We finally mention that in a similar spirit some formulae of Black–Scholes type are proved in [1] for markets with delay effects.

The plan of the paper is as follows: in section 2 we introduce notation and review some results on SDDEs, adding some precision on regularity properties of the solution, concerning in particular their Malliavin derivative. Section 3 is devoted to proving Theorem 3.1, which is the key to many subsequent results; here the operator ∇_0 is introduced. In section 4 we present the forward-backward system (1.2) and prove, in particular, formula (1.3). Section 5 is devoted to the study of (1.5): it is proved that a unique mild solution exists and is connected to the solution of the forward-backward system (1.2) by formula (1.3). In section 6 we study the optimal control problem; we prove in particular that the value function of the control problem is a solution (in the mild sense) of the Hamilton–Jacobi–Bellman equation; moreover, we show that the so-called fundamental relation holds, we give criteria for optimality of feedback controls, and we prove existence of optimal controls in the weak sense. Finally in section 7 it is shown how (1.5) may arise as the Black–Scholes equation in a financial market with memory effects and we give explicit conditions for its solvability.

2. Preliminary results on stochastic delay differential equations.

2.1. Notations. In this paper we consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a standard Wiener process $W = (W_t)_{t\geq 0}$ with values in $\mathbb{R}^d$. We denote by $(\mathcal{F}_t)_{t\geq 0}$ the natural filtration of W augmented in the usual way by the sets of $\mathbb{P}$-measure 0.

For fixed $r > 0$, we introduce the space

$$\mathbf{C} = C([-r, 0]; \mathbb{R}^n)$$

of continuous functions from $[-r, 0]$ to $\mathbb{R}^n$, endowed with the usual norm $|f|_{\mathbf{C}} = \sup_{\theta\in[-r,0]} |f(\theta)|$. We will consider $\mathbf{C}$-valued stochastic processes: for $T > 0$ we say that a $\mathbf{C}$-valued process $(X_t)_{t\in[0,T]}$ belongs to the space $\mathcal{S}^p([0, T]; \mathbf{C})$ $(1 \leq p < \infty)$ if its path are $\mathbf{C}$-continuous $\mathbb{P}$-a.s. and the norm

$$\|X\|^p_{\mathcal{S}^p([0,T];\mathbf{C})} = \mathbb{E} \sup_{t\in[0,T]} |X_t|^p_{\mathbf{C}} = \mathbb{E} \sup_{t\in[0,T]} \sup_{\theta\in[-r,0]} |X_t(\theta)|^p$$

is finite. Here and in the following, if no confusion is possible, we denote the norm of $\mathbb{R}^n$, $\mathbb{R}^d$ and $\mathbb{R}^{nd}$ by $|\cdot|$.

We next define several classes of differentiable functions between Banach spaces, first introduced in [11] in connection with stochastic processes, which allow us to formulate several regularity results in a compact way.

In the following, if E and K are Banach spaces, we denote by $\mathcal{G}^1(E,K)$ the space of functions $u : E \to K$ such that 1) u is continuous; 2) u is Gâteaux differentiable on E, with Gâteaux differential at point $x \in E$ denoted by $\nabla u(x) \in L(E,K)$ (the latter being the space of bounded linear operators from E to K, endowed with its usual norm); and 3) for every $h \in E$, the map $x \to \nabla u(x)\, h$ is continuous from E to K. We note that the map $x \to \nabla u(x)$ is not required to be continuous from E to $L(E,K)$: if this happens, then u is also Fréchet differentiable.

We say that a function $v : [0,T] \times E \to K$ belongs to $\mathcal{G}^{0,1}([0,T] \times E, K)$ if 1) v is continuous; 2) for every $t \in [0,T]$, $v(t,\cdot)$ is Gâteaux differentiable on E, with Gâteaux differential at point $x \in E$ denoted by $\nabla_x v(t,x) \in L(E,K)$; and 3) for every $h \in E$, the map $(t,x) \to \nabla_x v(t,x)\, h$ is continuous from $[0,T] \times E$ to K.

Now suppose $E = C([a,b];\mathbb{R}^n)$, where $a,b \in \mathbb{R}$, $a < b$. We recall that the dual space of $C([a,b])$ is the space of finite Borel measures on $[a,b]$, endowed with the variation norm. Identifying E with the product space $C([a,b])^n$ in the obvious way, we conclude that the dual space E^* of E can be identified with the space of n-tuples $\mu = (\mu_k)_{k=1}^n$, where each μ_k is a finite Borel measure on $[a,b]$, and the value of μ at an element $g = (g_k)_{k=1}^n \in C([a,b])^n$, where $g_k \in C([a,b])$, is denoted

$$\int_{[a,b]} g(\theta) \cdot \mu(d\theta) = \sum_{k=1}^n \int_{[a,b]} g_k(\theta)\, \mu_k(d\theta).$$

Let $v : [0,T] \times \mathbf{C} \to \mathbb{R}$ be a function such that $v(t,\cdot)$ is Gâteaux differentiable on $\mathbf{C}$ for every $t \in [0,T]$. Then the gradient $\nabla_x v(t,x)$ at point $(t,x) \in [0,T] \times \mathbf{C}$ is an n-tuple of finite Borel measures on $[-r,0]$. We denote by $|\nabla_x v(t,x)|$ its total variation norm and define

$$\nabla_0 v(t,x) = \nabla_x v(t,x)(\{0\}); \tag{2.1}$$

i.e., $\nabla_0 v(t,x)$ is a vector in $\mathbb{R}^n$ whose components $\nabla_0^k v(t,x)$ $(k = 1,\dots,n)$ are the masses at point 0 of the components of $\nabla_x v(t,x)$.

Remark 2.1. In the following, a basic role will be played by the space $\mathcal{G}^{0,1}([0,T] \times \mathbf{C}, \mathbb{R})$: according to the previous definitions, it consists of real continuous functions v on $[0,T] \times \mathbf{C}$ such that, for every $t \in [0,T]$, $v(t,\cdot)$ is Gâteaux differentiable on $\mathbf{C}$, with Gâteaux differential at point $x \in \mathbf{C}$ denoted by $\nabla_x v(t,x)$ (an n-tuple of finite Borel measures on $[-r,0]$), such that the map

$$(t,x) \to \langle \nabla_x v(t,x), h \rangle_{\mathbf{C}^*,\mathbf{C}} = \int_{[-r,0]} h(\theta) \cdot \nabla_x v(t,x)(d\theta)$$

is continuous on $[0,T] \times \mathbf{C}$, for every $h \in \mathbf{C}$.

2.2. Stochastic delay differential equations. We fix $T > 0$ and consider the following stochastic delay differential equation for an unknown process $(y_t)_{t\in[0,T]}$ taking values in $\mathbb{R}^n$:

$$\begin{cases} dy_t = b(t, y_{t+\cdot})\, dt + \sigma(t, y_{t+\cdot})\, dW_t, \qquad t \in [0,T], \\ y_\theta = x(\theta),\ \theta \in [-r,0], \end{cases} \tag{2.2}$$

where $y_{t+\cdot}$ denotes the past trajectory from time $t-r$ up to time t, namely, $y_{t+\cdot} = (y_{t+\theta})_{\theta\in[-r,0]}$, and $r > 0$ is the delay. $b(t,\cdot)$ and $\sigma(t,\cdot)$ are functions of the past trajectory of y and they are defined on the space of continuous functions, namely, $b : [0,T] \times \mathbf{C} \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbf{C} \to \mathbb{R}^{nd}$, where $\mathbb{R}^{nd}$ is identified with $L(\mathbb{R}^d, \mathbb{R}^n)$ the space of linear operators from $\mathbb{R}^d$ to $\mathbb{R}^n$. The function $x \in \mathbf{C}$ is the initial condition. We will refer to (2.2) as the *delay equation.*

We make the following assumptions on the coefficients of (2.2).

HYPOTHESIS 2.2.

1. *The functions $b : [0,T] \times \mathbf{C} \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbf{C} \to \mathbb{R}^{nd}$ are continuous and there exists a constant $K > 0$ such that for all $t \in [0,T]$ and $y(\cdot) \in \mathbf{C}$*
$$|b(t, y(\cdot))| + |\sigma(t, y(\cdot))| \leq K\,(1 + |y(\cdot)|_{\mathbf{C}});$$
2. *there exists a constant $L > 0$ such that for all $t \in [0,T]$ and $y(\cdot), z(\cdot) \in \mathbf{C}$*
$$|b(t, y(\cdot)) - b(t, z(\cdot))| + |\sigma(t, y(\cdot)) - \sigma(t, z(\cdot))| \leq L\,|y(\cdot) - z(\cdot)|_{\mathbf{C}};$$
3. *for all $t \in [0,T]$, we have $b(t,\cdot) \in \mathcal{G}^1(\mathbf{C}, \mathbb{R}^n)$ and $\sigma(t,\cdot) \in \mathcal{G}^1(\mathbf{C}, \mathbb{R}^{nd})$.*

In the following, we collect some results on the existence and uniqueness of a solution to (2.2) and on its regular dependence on the initial condition. It turns out that there exists a continuous solution, so we can define a $\mathbf{C}$-valued process $X = (X_t)_{t\in[0,T]}$ by

$$X_t(\theta) = y_{t+\theta}, \qquad \theta \in [-r, 0]. \tag{2.3}$$

We notice that if $t + \theta < 0$, then $y_{t+\theta} = x(t+\theta)$. We will use the notations y^x, y^x_t, X^x, or X^x_t to indicate dependence on the starting point $x \in \mathbf{C}$.

THEOREM 2.3. *If points 1 and 2 of Hypothesis 2.2 hold true, then there exists a unique continuous adapted solution of the delay equation (2.2), and, moreover, the process $(X_t)_{t\in[0,T]}$ belongs to $\mathcal{S}^p([0,T]; \mathbf{C})$ for every $p \geq 2$ and*

$$\|X\|^p_{\mathcal{S}^p([0,T];\mathbf{C})} = \mathbb{E} \sup_{t\in[-r,T]} |y_t|^p \leq C$$

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

In addition, the map $x \to X^x$ is Lipschitz continuous from $\mathbf{C}$ to $\mathcal{S}^p([0,T]; \mathbf{C})$; more precisely,

$$\|X^{x_1} - X^{x_2}\|_{\mathcal{S}^p([0,T];\mathbf{C})} = \left(\mathbb{E} \sup_{t\in[-r,T]} |y^{x_1}_t - y^{x_2}_t|^p\right)^{1/p} \leq L \sup_{\theta\in[-r,0]} |x_1(\theta) - x_2(\theta)|$$

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

If we further assume that point 3 of Hypothesis 2.2 holds true, then the map $x \to X^x$ belongs to the space $\mathcal{G}^1(\mathbf{C}, \mathcal{S}^p([0,T]; \mathbf{C}))$.

Proof. For the proof (in the case of $p = 2$), we refer to [18, Chapter II]: we refer to Theorem 2.1 for the existence and uniqueness of the solution of (2.2), to Theorem 3.1 for the Lipschitz dependence of this solution on the initial datum, and to Theorem 3.2 for the differentiability of the solution with respect to the initial datum. See also [19, Theorems I.1 and I.2]. The proof in the case of $p > 2$ can be performed in a similar way. □

Let us introduce a delay equation similar to (2.2) but with the initial condition given at time $t \in [0,T]$:

$$\begin{cases} dy^{t,x}_\tau = b(\tau, y^{t,x}_{\tau+\cdot})\, d\tau + \sigma(\tau, y^{t,x}_{\tau+\cdot})\, dW_\tau, & \tau \in [t,T], \\ y^{t,x}_{t+\theta} = x(\theta), \quad \theta \in [-r, 0]. \end{cases} \tag{2.4}$$

We introduce the $\mathbf{C}$-valued process given by

$$X_\tau^{t,x}(\theta) = y_{\tau+\theta}^{t,x}, \qquad \theta \in [-r, 0]. \tag{2.5}$$

By [18, Chapter III, Theorem 2.1], the $\mathbf{C}$-valued process $(X_\tau^{t,x})_{\tau \in [t,T]}$ is a Markov process with transition semigroup, acting on bounded and Borel measurable $\phi : \mathbf{C} \to \mathbb{R}$, given by

$$P_{t,\tau}[\phi](x) = \mathbb{E}\,\phi(X_\tau^{t,x}), \qquad 0 \le t \le \tau \le T,\ x \in \mathbf{C}. \tag{2.6}$$

Remark 2.4. The transition semigroup $(P_{t,\tau})$ has been extensively studied in the literature; see, e.g., [18] and [19]. Although our techniques essentially bypass the difficulties related to the characterization of the generator $(P_{t,\tau})$ and of its domain, we briefly recall a result which will appear in section 5 in the formulation of the Kolmogorov equation. For simplicity, let us consider the autonomous case in (2.4): b and σ do not depend on time and $s = 0$, so we consider the one parameter semigroup $(P_t)_{t\in[0,T]}$. The transition semigroup (P_t) is never strongly continuous on the space $\mathbf{C}$, nevertheless it admits a weakly continuous generator $\mathcal{L}$; see [18, chapter IV] and [19, chapter II]. Let $S_t : \mathbf{C} \to \mathbf{C}$ denote the shift operator, and let $\mathcal{S}$ denote the weak generator of the corresponding semigroup. To derive a formula for the generator $\mathcal{L}$ we need to augment $\mathbf{C}$ by adding an n-dimensional direction. $\mathcal{L}$ will be equal to the sum of the generator of the shift semigroup $\mathcal{S}$ and a second order linear partial differential operator along this new direction. Let $F_n := \{v1_0 : v \in \mathbb{R}^n\}$ and $\mathbf{C} \oplus F_n := \{f + v1_0 : f \in \mathbf{C}, v \in \mathbb{R}^n\}$ with the norm $\|f + v1_0\|_{\mathbf{C}\oplus F_n} := |f|_{\mathbf{C}} + |v|$. Suppose that $\phi : \mathbf{C} \to \mathbb{R}$ is twice continuously Fréchet differentiable and let $f \in \mathbf{C}$. Then the Fréchet derivatives $\nabla\phi(f)$ and $\nabla^2\phi(f)$ have unique weakly continuous linear and bilinear extensions

$$\overline{\nabla\phi(f)} : \mathbf{C} \oplus F_n \to \mathbb{R}, \qquad \overline{\nabla^2\phi(f)} : (\mathbf{C} \oplus F_n) \times (\mathbf{C} \oplus F_n) \to \mathbb{R}.$$

Comparing with (1.4) we notice that $\overline{\nabla\phi(f)}(1_0) = \nabla_0\phi(f)$. We are ready to introduce $\mathcal{L}$. Suppose that $\phi : \mathbf{C} \to \mathbb{R}$, $\phi \in D(\mathcal{S})$, and ϕ is sufficiently smooth (e.g., ϕ is twice continuously differentiable and its derivatives are globally bounded and Lipschitz continuous). Then $\phi \in D(\mathcal{L})$ and for all $f \in \mathbf{C}$

$$\mathcal{L}(\phi)(f) = \mathcal{S}(\phi)(f) + \overline{\nabla\phi(f)}(b(f)1_0) + \frac{1}{2}\sum_{i=1}^{n} \overline{\nabla^2\phi(f)}(\sigma(f)(e_i)1_0, \sigma(f)(e_i)1_0), \tag{2.7}$$

where $\{e_i\}_{i=1}^n$ is any basis of $\mathbb{R}^n$.

We conclude this remark observing that if $\mathbf{C}$ is replaced by $L^2([-r,0];\mathbb{R}^n)$, then $\mathcal{L}$ takes a much simpler form; see, for instance, [13, page 314].

2.3. Differentiability in the Malliavin sense. Our aim now is to compute the Malliavin derivative of the solution of the delay equation. We start by recalling some basic definitions from the Malliavin calculus. We refer the reader to the book [21] for a detailed exposition.

We consider again a standard Wiener process $W = (W_t)_{t\ge 0}$ in $\mathbb{R}^d$ and the Hilbert space $L^2([0,T];\mathbb{R}^d)$ of Borel measurable, square summable functions on $[0,T]$ with values in $\mathbb{R}^d$, with its natural inner product. This can be identified with the product space $(L^2([0,T]))^d$ or with the space $L^2(\mathcal{T})$, where the measure space $\mathcal{T} := [0,T] \times \{1,\dots,d\}$ is endowed with the product of the Lebesgue measure on $[0,T]$ and the counting measure on $\{1,\dots,d\}$. Elements $h \in L^2([0,T];\mathbb{R}^d)$ may be denoted $\{h^j(s),\ s \in [0,T],\ j = 1,\dots,d\}$ or $\{h^j\}$, where $h^j \in L^2([0,T])$.

For every $h \in L^2([0,T];\mathbb{R}^d)$ we denote

$$W(h) = \int_0^T h(s) \cdot dW_s = \sum_{j=1}^d \int_0^T h^j(s)\, dW_s^j.$$

W is an isometry of $L^2([0,T];\mathbb{R}^d)$ onto a Gaussian subspace of $L^2(\Omega)$, called the first Wiener chaos. Given a Hilbert space K, let S_K be the set of K-valued random variables F of the form

$$F = \sum_{r=1}^m f_r(W(h_1), \ldots, W(h_n))\, k_r,$$

where $h_1, \ldots, h_n \in L^2([0,T];\mathbb{R}^d)$, $\{k_r\}$ is a basis of K, and $f_1, \ldots, f_m$ are infinitely differentiable functions $\mathbb{R}^n \to \mathbb{R}$ bounded together with all their derivatives. The Malliavin derivative DF of $F \in S_K$ is defined as the process $\{D_s^j F;\ s \in [0,T],\ j \in \{1, \ldots, d\}\}$ given by

$$D_s^j F = \sum_{r=1}^m \sum_{k=1}^n \partial_k f_r(W(h_1), \ldots, W(h_n))\, h_k^j(s)\, k_r,$$

with values in K; by ∂_k we denote the partial derivatives with respect to the kth variable. It is known that the operator $D : S_K \subset L^2(\Omega;K) \to L^2(\Omega \times [0,T] \times \{1, \ldots, d\}; K) = L^2(\Omega \times \mathcal{T}; K)$ is closable. We denote by $\mathbb{D}^{1,2}(K)$ the domain of its closure, endowed with the graph norm, and we use the same letter to denote D and its closure:

$$D : \mathbb{D}^{1,2}(K) \subset L^2(\Omega;K) \to L^2(\Omega \times \mathcal{T}; K).$$

The adjoint operator of D,

$$\delta :\ \mathrm{dom}\,(\delta) \subset L^2(\Omega \times \mathcal{T}; K) \to L^2(\Omega; K),$$

is called Skorohod integral. For a process $u = \{u_s^j;\ s \in [0,T],\ j \in \{1, \ldots, d\}\} \in \mathrm{dom}(\delta)$ we will also use the notations

$$\delta(u) = \int_0^T u_s\, \hat{d}W_s = \sum_{j=1}^d \int_0^T u_s^j\, \hat{d}W_s^j.$$

It is known that $\mathrm{dom}(\delta)$ contains every $(\mathcal{F}_t)$-predictable process in $L^2(\Omega \times \mathcal{T}; K)$ and for such processes the Skorohod integral coincides with the Itô integral; $\mathrm{dom}(\delta)$ also contains the class $\mathbb{L}^{1,2}(K)$, the latter being defined as the space of processes $u \in L^2(\Omega \times \mathcal{T}; K)$ such that $u_t^j \in \mathbb{D}^{1,2}(K)$ for a.e. $t \in [0,T]$ and every j, and there exists a measurable version of $D_s^i u_t^j$ satisfying

$$\|u\|^2_{\mathbb{L}^{1,2}(K)} = \|u\|^2_{L^2(\Omega \times \mathcal{T}; K)} + \mathbb{E} \sum_{i,j=1}^d \int_0^T \int_0^T \|D_s^i u_t^j\|_K^2\, dt\, ds < \infty.$$

Moreover, $\|\delta(u)\|^2_{L^2(\Omega;K)} \le \|u\|^2_{\mathbb{L}^{1,2}(K)}$. We note that the space $\mathbb{L}^{1,2}(K)$ is isometrically isomorphic to $L^2(\mathcal{T}; \mathbb{D}^{1,2}(K))$.

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

If $K = \mathbb{R}$ or $K = \mathbb{R}^n$, we write $\mathbb{D}^{1,2}$ and $\mathbb{L}^{1,2}$ instead of $\mathbb{D}^{1,2}(K)$ and $\mathbb{L}^{1,2}(K)$, respectively.

We now introduce the Malliavin derivative for a functional of a stochastic process. In the remainder of this section we set $E = C([-r,T];\mathbb{R}^n)$. If $f \in \mathcal{G}^1(E,\mathbb{R}^n)$, then, according to the notation introduced above,

$$\langle \nabla f(x), g\rangle_{E^*,E} = \int_{[-r,T]} g(\theta)\cdot \nabla f(x)(d\theta), \qquad x, g \in E.$$

If y is a continuous stochastic process with time parameter $[-r,T]$, then $f(y_\cdot)$ is a random variable. We wish to state a chain rule for the Malliavin derivative of $f(y_\cdot)$. We will restrict ourselves to the case when y is adapted; more precisely, its restriction to $[0,T]$ is adapted to $(\mathcal{F}_t)_{t\in[0,T]}$ and its restriction to $[-r,0]$ is *deterministic*. Clearly, $Dy_t = 0$ for $t \in [-r,0]$. Following [15, Lemma 2.6], we have the following basic result (we note that in [15] derivatives are understood in the sense of Fréchet, but the same arguments apply to the present situation).

LEMMA 2.5. *For $E = C([-r,T];\mathbb{R}^n)$, let $f \in \mathcal{G}^1(E,\mathbb{R})$ be a Lipschitz continuous function. Assume that $y = (y_t)_{t\in[-r,T]}$ is a process in $\mathbb{R}^n$ satisfying the following conditions:*

1. *y is a continuous adapted process and $\mathbb{E}\sup_{t\in[-r,T]}|y_t|^2 < \infty$;*
2. *$y \in L^2([-r,T],\mathbb{D}^{1,2})$ and the process $\{D_s y_t,\ 0 \le s \le t \le T\}$ admits a version such that, for every $s \in [0,T]$, $\{D_s y_t,\ t\in[s,T]\}$ is a continuous process and*

$$\mathbb{E}\int_0^T \sup_{t\in[s,T]} |D_s y_t|^2 ds < \infty.$$

Then $f(y_\cdot) \in \mathbb{D}^{1,2}$ and its Malliavin derivative is given by the formula: for $j = 1,\dots,d$ and a.e. $s \in [0,T]$ we have, $\mathbb{P}$-a.s.,

$$D_s^j(f(y_\cdot)) = \langle \nabla f(y_\cdot), D_s^j y_\cdot\rangle_{E^*,E} = \int_{[-r,T]} D_s^j y_\theta \cdot \nabla f(y_\cdot)(d\theta). \tag{2.8}$$

Next, we establish when the solution of the delay equation is Malliavin differentiable; moreover, we write a stochastic (functional) differential equation satisfied by the Malliavin derivative. We substantially follow [15, Theorem 4.1].

THEOREM 2.6. *Let Hypothesis 2.2 be satisfied. Then the solution $(y_t)_{t\in[-r,T]}$ satisfies conditions 1 and 2 in Lemma 2.5. Moreover, $y_t \in \mathbb{D}^{1,2}$ for every $t \in [0,T]$ and the following equation holds: for $j = 1,\dots,d$ and every $s\in[0,T]$ we have, $\mathbb{P}$-a.s.,*

$$\begin{cases} D_s^j y_t = \sigma(s, y_{s+\cdot}) + \displaystyle\int_s^t \int_{[-r,0]} D_s^j y_{t+\theta}\cdot \nabla_x b(t, y_{t+\cdot})(d\theta)\, dt \\ \qquad\qquad + \displaystyle\int_s^t \int_{[-r,0]} D_s^j y_{t+\theta}\cdot \nabla_x \sigma(t, y_{t+\cdot})(d\theta)\, dW_t, \qquad t\in[s,T], \\ D_s^j y_t = 0, \qquad t \in [-r,s). \end{cases} \tag{2.9}$$

Finally, for every $p \in [2,\infty)$ and $s \in [0,T]$ we have

$$\mathbb{E}\int_0^T \sup_{t\in[s,T]} |D_s y_t|^p ds < \infty. \tag{2.10}$$

Proof. Except for the final statement, the proof can be achieved with techniques similar to the ones indicated in the proof of Theorem 4.1 in [15]. The only minor difference is that we consider a general delay differential equation, while in [15] the coefficients depend on the past behavior of the solution only after time 0. However, the same arguments apply.

The proof of the final statement follows by standard estimates on (2.10), taking into account that $\nabla_x b$ and $\nabla_x \sigma$ are bounded in the total variation norm. ∎

COROLLARY 2.7. *Suppose that the assumptions of Theorem* 2.6 *hold true, let* $\mathbf{C} = C([-r,0];\mathbb{R}^n)$ *and* $(X_t)_{t\in[0,T]}$ *be the* $\mathbf{C}$*-valued process defined by* (2.3). *Suppose that* $f \in \mathcal{G}^1(\mathbf{C};\mathbb{R})$ *satisfies*

$$|\nabla f(x)| \le C(1 + |x|_{\mathbf{C}})^m, \qquad x \in \mathbf{C},$$

for some $C > 0$ *and* $m \ge 0$.

Then for every $t \in [0,T]$, $f(X_t) = f(y_{t+\cdot})$ *belongs to* $\mathbb{D}^{1,2}$, *and for* $j = 1,\dots,d$ *we have, for a.e.* $s \in [0,T]$, $\mathbb{P}$*-a.s.,*

$$D_s^j(f(X_t)) = \langle \nabla f(X_t), D_s^j y_{t+\cdot}\rangle_{\mathbf{C}^*,\mathbf{C}} = \int_{[-r,0]} D_s^j y_{t+\theta} \cdot \nabla f(X_t)(d\theta). \tag{2.11}$$

Proof. The conclusion follows immediately from Lemma 2.5 and Theorem 2.6 if f is a Lipschitz function. The general case can be proved by approximating f by a sequence of Lipschitz functions obtained by a standard truncation procedure. ∎

Remark 2.8. The first result on Malliavin differentiability of the solution of a functional stochastic differential equations was proved in [16]. In that paper the aim was to prove that y_t belongs to the domain of the generator of the Ornstein–Uhlenbeck semigroup of the Malliavin calculus; therefore, more restrictive assumptions were assumed on the coefficients of (2.2). In particular, they were required to be twice differentiable.

3. A result on joint quadratic variations. The aim of this section is to state and prove a technical result, Theorem 3.1, which will be used in the rest of this paper. To state this theorem we need to recall some definitions concerning joint quadratic variations of stochastic processes and to introduce a differential operator, denoted ∇_0, which will also play a basic role in what follows.

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

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

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

This definition is taken from [27], except that we do not require that the convergence in probability holds uniformly with respect to time. In [27] the process $\langle X,Y\rangle$ is called generalized covariation process; several properties are investigated in [28], [29], often in connection with the stochastic calculus introduced in [26]. With respect to the classical definition, the present one has some technical advantages that are useful when dealing with convergence issues (compare, for instance, the proof of Theorem 3.1).

In the following, we will consider joint quadratic variations over different intervals, which is defined by obvious modifications.

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

If X and Y are stochastic integrals with respect to the Wiener process, then the joint quadratic variation as defined above coincides with the classical one. A similar conclusion holds for general semimartingales; see [27, Proposition 1.1].

We set $\mathbf{C} = C([-r,0];\mathbb{R}^n)$ and, for every $t \in [0,T]$ and $x \in \mathbf{C}$, we let $\{X_s^{t,x}, s \in [t,T]\}$ denote the process defined by the equality (2.5), obtained as a solution to (2.4). In particular, it is a $\mathbf{C}$-valued process with continuous paths and adapted to the filtration $\{\mathcal{F}_{[t,s]}, s \in [t,T]\}$. $X_s^{t,x}(\omega)$ is measurable in (ω, s, t, x).

Let $u : [0,T] \times \mathbf{C} \to \mathbb{R}$ be a function such that $u(t,\cdot)$ is Gâteaux differentiable on $\mathbf{C}$ for every $t \in [0,T]$. Then the gradient $\nabla_x u(t,x)$ at point $(t,x) \in [0,T] \times \mathbf{C}$ is an n-tuple of finite Borel measures on $[-r,0]$. We denote by $|\nabla_x u(t,x)|$ its total variation norm and we denote $\nabla_0 u(t,x) = \nabla_x u(t,x)(\{0\})$; compare (2.1). Thus, $\nabla_0 u(t,x)$ is a vector in $\mathbb{R}^n$ whose components $\nabla_0^k u(t,x)$ $(k = 1,\dots,n)$ are the masses at point 0 of the components of $\nabla_x u(t,x)$.

We denote by W^i $(i = 1,\dots,d)$, the ith component of the Wiener process W, by σ^i the ith column of the $n \times d$ matrix σ, and by σ_k^i $(k = 1,\dots,n)$, its components.

THEOREM 3.1. *Assume that $u : [0,T] \times \mathbf{C} \to \mathbb{R}$ is a Borel measurable function such that $u(t,\cdot) \in \mathcal{G}^1(\mathbf{C},\mathbb{R})$ for every $t \in [0,T]$ and*

$$|u(t,x)| + |\nabla_x u(t,x)| \le C(1 + |x|)^m \tag{3.1}$$

for some $C > 0$, $m \ge 0$, and for every $t \in [0,T]$, $x \in \mathbf{C}$.

Then for every $x \in \mathbf{C}$, $i = 1,\dots,d$, and $0 \le t \le T' < T$, the processes $\{u(s, X_s^{t,x}), s \in [t,T]\}$ and W^i admit a joint quadratic variation on the interval $[t,T']$, given by the formula

$$\begin{aligned}\langle u(\cdot, X_\cdot^{t,x}), W^i\rangle_{[t,T']} &= \int_t^{T'} \sigma^i(s, X_s^{t,x}) \cdot \nabla_0 u(s, X_s^{t,x})\, ds \\ &= \sum_{k=1}^n \int_t^{T'} \sigma_k^i(s, X_s^{t,x}) \cdot \nabla_0^k u(s, X_s^{t,x})\, ds.\end{aligned}$$

Proof. For the sake of simplicity we write the proof in the case $t = 0$, the general case being deduced by the same arguments.

We fix $x \in \mathbf{C}$, $T' \in (0,T)$, and we denote $X^{0,x}$ by X for simplicity. Thus, $X_t = y(t + \cdot)$, $t \in [0,T]$, satisfies

$$dy(t) = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \qquad X_0 = x.$$

We will use the results on the Malliavin derivatives stated in Theorem 2.6, and, in particular, formula (2.9) that, in view of (2.11), can be written in the form

$$D_s y(t) = \sigma(s, X_s) + \int_s^t D_s[b(r, X_r)]\, dr + \int_s^t D_s[\sigma(r, X_r)]\, dW_r \tag{3.2}$$

for $0 \le t \le s \le T$. Noting that $\nabla_x b(t,x)$ and $\nabla_x \sigma(t,x)$ are bounded by the Lipschitz constant L of $b(t,\cdot)$ and $\sigma(t,\cdot)$, it follows from (2.11) that for every $r \in [0,T]$

$$\begin{aligned}\|D_\cdot[b(r, X_r)]\|^2 &\le L^2 \int_0^T \sup_{t \in [s,T]} |D_s y(t)|^2\, ds, \\ \|D_\cdot[\sigma(r, X_r)]\|^2 &\le L^2 \int_0^T \sup_{t \in [s,T]} |D_s y(t)|^2\, ds,\end{aligned} \tag{3.3}$$

where $\|\cdot\|$ denotes the norm in $L^2([0,T];\mathbb{R}^d)$.

We have to prove that

$$C^\epsilon := C^\epsilon_{[0,T']}(u(\cdot, X_\cdot), W^i) = \frac{1}{\epsilon}\int_0^{T'} (u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))(W^i_{t+\epsilon} - W^i_t)\, dt$$
$$\to \int_0^{T'} \sigma^i(t, X_t)\cdot \nabla_0 u(t, X_t)\, dt$$

in probability, as $\epsilon \to 0$.

We need to rewrite C_ϵ in an appropriate way, fixing $\epsilon > 0$ so small that $T'+\epsilon \le T$. We first explain our argument by writing down some informal passages: by the rules of Malliavin calculus we have, for a.a. $t \in [0, T']$,

(3.4)

$$(u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))(W^i_{t+\epsilon} - W^i_t) = (u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))e^*_i \int_t^{t+\epsilon} dW_s$$
$$= \int_t^{t+\epsilon} D^i_s(u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))\, ds + \int_t^{t+\epsilon} (u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))e^*_i \hat{d}W_s,$$

where the symbol $\hat{d}W$ denotes the Skorohod integral, and by e_i we denote the ith component of the canonical basis of $\mathbb{R}^d$ and by e^*_i its transpose (row) vector. Integrating over $[0, T']$ with respect to t and interchanging integrals gives

$$\epsilon\, C^\epsilon = \int_0^{T'} \int_t^{t+\epsilon} D^i_s(u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))\, ds\, dt$$
$$+ \int_0^{T'+\epsilon} \int_{(s-\epsilon)^+}^{s\wedge T'} (u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))\, dt\; e^*_i\; \hat{d}W_s. \tag{3.5}$$

To justify (3.4) and (3.5) rigorously we proceed as follows. To shorten notation we define

$$v_t = (u(t+\epsilon, X_{t+\epsilon}) - u(t, X_t))\, 1_{[0,T']}(t), \qquad t \in [0, T],$$
$$A^\epsilon = \{(t, s) \in [0, T] \times [0, T]\ :\ 0 \le t \le T', t \le s \le t+\epsilon\}.$$

Using Corollary 2.7 and formula (2.10), it is easy to show that, for all t, v_t belongs to $\mathbb{D}^{1,2}$ and the process $v_t\, 1_{A^\epsilon}(t, \cdot)$ belongs to $L^2(\Omega \times [0, T])$. By [22, Theorem 3.2] (see also [21, section 1.3.1, equation (1.49)]) we conclude that $v_t\, 1_{A^\epsilon}(t, \cdot)\, e^*_i$ is Skorohod integrable and the formula

$$\int_0^T v_t\, 1_{A^\epsilon}(t, s)\, e^*_i\; \hat{d}W_s = v_t \int_0^T 1_{A^\epsilon}(t, s)\; e^*_i\; \hat{d}W_s - \int_0^T D^i_s v_t\, 1_{A^\epsilon}(t, s)\, ds =: z_t \tag{3.6}$$

holds provided z_t belongs to $L^2(\Omega)$. Since $\int_0^T 1_{A^\epsilon}(t, s)\; \hat{d}W_s$ coincides with the Itô integral $\int_0^T 1_{A^\epsilon}(t, s)\; dW_s = (W_{t+\epsilon} - W_t)1_{[t,T']}(t)$, it is in fact easy to verify that we even have $z \in L^2(\Omega \times [0, T])$; thus (3.6) holds for a.a. t, and (3.6) yields (3.4) for a.a. $t \in [0, T']$.

Next we wish to show that the process $\int_0^T v_t 1_{A^\epsilon}(t, \cdot)\, dt\, e_i$ is Skorohod integrable and to compute its integral, which occurs in the right-hand side of (3.5). For arbitrary

$G \in \mathbb{D}^{1,2}$, by the definition of the Skorohod integral and by (3.6),

$$\begin{aligned}\mathbb{E}\int_0^T \left\langle \int_0^T v_t 1_{A^\epsilon}(t,s)\,dt\,e_i, D_s G\right\rangle_{\mathbb{R}^d} ds &= \int_0^T \mathbb{E}\int_0^T \langle v_t 1_{A^\epsilon}(t,s)\,e_i, D_s G\rangle_{\mathbb{R}^d}\ ds\,dt \\ &= \int_0^T \mathbb{E}\left[G\int_0^T v_t 1_{A^\epsilon}(t,s)\ e_i^*\ \hat{d}W_s\right]\ dt \\ &= \mathbb{E}\left[G\int_0^T z_t\ dt\right].\end{aligned}$$

This shows, by definition, that $\int_0^T v_t 1_{A^\epsilon}(t,\cdot)\,dt\,e_i$ is Skorohod integrable and

$$\int_0^T\int_0^T v_t 1_{A^\epsilon}(t,s)\ dt\ e_i^*\ \hat{d}W_s = \int_0^T z_t\ dt = \int_0^T\int_0^T v_t 1_{A^\epsilon}(t,s)\ e_i^*\ \hat{d}W_s\ dt.$$

Recalling (3.6) we obtain

$$\begin{aligned}&\int_0^T\int_0^T v_t 1_{A^\epsilon}(t,s)\ dt\ e_i^*\ \hat{d}W_s \\ &\quad= \int_0^T v_t(W^i_{t+\epsilon} - W^i_t)\ 1_{[t,T']}(t)\ dt - \int_0^T\int_0^T D^i_s v_t\,e_i\,1_{A^\epsilon}(t,s)\ ds\ dt,\end{aligned}$$

and (3.5) is proved.

Recalling that $D_s(u(t,X_t)) = 0$ for $s > t$ by adaptedness, and using the chain rule (2.11) for the Malliavin derivative, we have, for a.a. s,t with $s \in [t, t+\epsilon]$,

$$\begin{aligned}D_s(u(t+\epsilon, X_{t+\epsilon}) - u(t,X_t)) &= D_s(u(t+\epsilon, X_{t+\epsilon})) \\ &= \int_{[-r,0]} D_s y(t+\epsilon+\theta)\cdot\nabla_x u(t+\epsilon, X_{t+\epsilon})(d\theta),\end{aligned}$$

and from (3.5) we deduce

$$\begin{aligned}C^\epsilon =&\ \frac{1}{\epsilon}\int_0^{T'}\int_t^{t+\epsilon}\int_{[-r,0]} D^i_s y(t+\epsilon+\theta)\cdot\nabla_x u(t+\epsilon, X_{t+\epsilon})(d\theta)\ ds\ dt \\ &+\ \frac{1}{\epsilon}\int_0^{T'+\epsilon}\int_{(s-\epsilon)^+}^{s\wedge T'} (u(t+\epsilon, X_{t+\epsilon}) - u(t,X_t))\ dt\ e_i^*\ \hat{d}W_s \\ &=: I_1^\epsilon + I_2^\epsilon.\end{aligned}$$

Now we let $\epsilon \to 0$, and we first claim that $I_2^\epsilon \to 0$ in probability. To prove this, it is enough to show that the process $\frac{1}{\epsilon}\int_0^T (u(t+\epsilon, X_{t+\epsilon}) - u(t,X_t))\ 1_{A^\epsilon}(t,\cdot)\ dt$ converges to 0 in $\mathbb{L}^{1,2}$. Indeed, since the Skorohod integral is a bounded linear operator from $\mathbb{L}^{1,2}$ to $L^2(\Omega)$, this implies that

$$I_2^\epsilon = \int_0^T \frac{1}{\epsilon}\int_0^T (u(t+\epsilon, X_{t+\epsilon}) - u(t,X_t))1_{A^\epsilon}(t,s)\ dt\ e_i^*\ \hat{d}W_s \to 0$$

in $L^2(\Omega)$. We prove, more generally, that for an arbitrary element $y \in \mathbb{L}^{1,2}(\mathbb{R})$, if we set

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

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

$$\begin{aligned}
|T^\epsilon(y)_s|^2_{\mathbb{D}^{1,2}(\mathbb{R})} &\le \frac{1}{\epsilon^2}\int_0^T 1_{A^\epsilon}(t,s)\,dt \int_0^T |y_{t+\epsilon}-y_t|^2_{\mathbb{D}^{1,2}(\mathbb{R})}\,1_{A^\epsilon}(t,s)\,dt\\
&\le \frac{1}{\epsilon}\int_0^T |y_{t+\epsilon}-y_t|^2_{\mathbb{D}^{1,2}(\mathbb{R})}\,1_{A^\epsilon}(t,s)\,dt,
\end{aligned}$$

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

This shows the required bound and completes the proof that $I_2^\epsilon \to 0$ as $\epsilon \to 0$.

Now we proceed to compute the limit of I_1^ϵ. We note that, by adaptedness, $D_s y(t+\epsilon+\theta) = 0$ for $s > t+\epsilon+\theta$, so that

$$I_1^\epsilon = \frac{1}{\epsilon}\int_0^{T'}\int_t^{t+\epsilon}\int_{[s-t-\epsilon,0]} D_s^i y(t+\epsilon+\theta)\cdot\nabla_x u(t+\epsilon,X_{t+\epsilon})(d\theta)\,ds\,dt.$$

For fixed t, let us exchange integrals with respect to ds and $\nabla_x u(t+\epsilon,X_{t+\epsilon})(d\theta)$ obtaining

$$I_1^\epsilon = \frac{1}{\epsilon}\int_0^{T'}\int_{[-\epsilon,0]}\int_t^{t+\epsilon+\theta} D_s^i y(t+\epsilon+\theta)\,ds\cdot\nabla_x u(t+\epsilon,X_{t+\epsilon})(d\theta)\,dt.$$

Next, we replace $D_s y(t+\epsilon+\theta)$ by the expression given by (3.2) and obtain

$$\begin{aligned}
I_1^\epsilon = {}&\frac{1}{\epsilon}\int_0^{T'}\int_{[-\epsilon,0]}\int_t^{t+\epsilon+\theta}\sigma^i(s,X_s)\,ds\cdot\nabla_x u(t+\epsilon,X_{t+\epsilon})(d\theta)\,dt\\
&+\frac{1}{\epsilon}\int_0^{T'}\int_{[-\epsilon,0]}\int_t^{t+\epsilon+\theta}\int_s^{t+\epsilon+\theta} D_s^i[b(r,X_r)]\,dr\,ds\cdot\nabla_x u(t+\epsilon,X_{t+\epsilon})(d\theta)\,dt\\
&+\frac{1}{\epsilon}\int_0^{T'}\int_{[-\epsilon,0]}\int_t^{t+\epsilon+\theta}\int_s^{t+\epsilon+\theta} D_s^i[\sigma(r,X_r)]\,dW_r\,ds\cdot\nabla_x u(t+\epsilon,X_{t+\epsilon})(d\theta)\,dt\\
&=: J_1^\epsilon + J_2^\epsilon + J_3^\epsilon.
\end{aligned}$$

We first show that $J_3^\epsilon \to 0$ in $L^1(\Omega)$. Since, by (3.1),

$$|\nabla_x u(t+\epsilon,X_{t+\epsilon})| \le C\Big(1+\sup_{t\in[0,T]}|X_t|_{\mathbf{C}}\Big)^m,$$

then, using the notation $|\nabla_x u(t+\epsilon, X_{t+\epsilon})|(d\theta)$ to indicate the total variation measure, we have

$$
\begin{aligned}
|J_3^\epsilon| &\le \frac{1}{\epsilon}\int_0^{T'}\int_{[-\epsilon,0]}\int_t^{t+\epsilon+\theta}\left|\int_s^{t+\epsilon+\theta} D_s^i[\sigma(r,X_r)]\,dW_r\right|\,ds\,|\nabla_x u(t+\epsilon,X_{t+\epsilon})|(d\theta)\,dt\\
&\le C\Big(1+\sup_{t\in[0,T]}|X_t|_{\mathbf{C}}\Big)^m\frac{1}{\epsilon}\int_0^{T'}\sup_{\theta\in[-\epsilon,0]}\int_t^{t+\epsilon+\theta}\left|\int_s^{t+\epsilon+\theta} D_s^i[\sigma(r,X_r)]\,dW_r\right|\,ds\ \,dt\\
&\le C\Big(1+\sup_{t\in[0,T]}|X_t|_{\mathbf{C}}\Big)^m\frac{1}{\epsilon}\int_0^{T'}\int_t^{t+\epsilon}\sup_{\theta\in[-\epsilon,0]}\left|\int_s^{t+\epsilon+\theta} D_s^i[\sigma(r,X_r)]\,dW_r\right|\,ds\ \,dt.
\end{aligned}
$$

Taking the $L^1(\Omega)$ norm of both sides and using the Hölder and the Doob maximal inequality, we have

$$
\begin{aligned}
&\|J_3^\epsilon\|_{L^1(\Omega)}\\
&\le C\Big\|\Big(1+\sup_{t\in[0,T]}|X_t|_{\mathbf{C}}\Big)^m\Big\|_{L^2(\Omega)}\frac{1}{\epsilon}\int_0^{T'}\int_t^{t+\epsilon}\left\|\sup_{\theta\in[-\epsilon,0]}\left|\int_s^{t+\epsilon+\theta} D_s^i[\sigma(r,X_r)]\,dW_r\right|\right\|_{L^2(\Omega)}ds\,dt\\
&\le \frac{C}{\epsilon}\int_0^{T'}\int_t^{t+\epsilon}\left\|\int_s^{t+\epsilon} D_s^i[\sigma(r,X_r)]\,dW_r\right\|_{L^2(\Omega)}ds\ \,dt\\
&= \frac{C}{\epsilon}\int_0^{T'}\int_t^{t+\epsilon}\left(\int_s^{t+\epsilon}\mathbb{E}\,\left|D_s^i[\sigma(r,X_r)]\right|^2\,dr\right)^{1/2}ds\ \,dt.
\end{aligned}
$$

Denoting for simplicity $h(s,r)=\mathbb{E}\left|D_s^i[\sigma(r,X_r)]\right|^2$, we obtain

$$
\begin{aligned}
\|J_3^\epsilon\|_{L^1(\Omega)} &\le \frac{C}{\epsilon}\int_0^{T'}\int_t^{t+\epsilon}\left(\int_t^{t+\epsilon} h(s,r)\,dr\right)^{1/2} ds\,dt\\
&= \frac{C}{\sqrt{\epsilon}}\int_0^{T'}\left(\int_t^{t+\epsilon}\int_t^{t+\epsilon} h(s,r)\,dr\,ds\right)^{1/2} dt\\
&\le C\sqrt{T'}\left(\int_0^{T'}\left[\frac{1}{\epsilon}\int_t^{t+\epsilon}\int_t^{t+\epsilon} h(s,r)\,dr\,ds\right]\,dt\right)^{1/2}.
\end{aligned}
$$

Let us note that $h\in L^1([0,T]^2)$, since by (3.3) we have

$$
\int_0^T\int_0^T h(s,r)\,dr\,ds \le \mathbb{E}\int_0^T\|D_\cdot[\sigma(r,X_r)]\|^2dr \le L^2T\int_0^T\mathbb{E}\sup_{t\in[s,T]}|D_s y(t)|^2ds<\infty.
$$

Let us define the operator $\mathcal{A}_\epsilon : L^1([0,T]^2)\to L^1([0,T])$ by

$$
(\mathcal{A}_\epsilon k)(t)=\frac{1}{\epsilon}\int_t^{(t+\epsilon)\wedge T}\int_t^{(t+\epsilon)\wedge T} k(s,r)\,dr\,ds,\qquad k\in L^1([0,T]^2).
$$

Then we have $\|J_3^\epsilon\|_{L^1(\Omega)}\le C\sqrt{T'}\|\mathcal{A}_\epsilon h\|^{1/2}_{L^1([0,T])}$, so to prove that $J_3^\epsilon\to 0$ in $L^1(\Omega)$ it is enough to show that $\mathcal{A}_\epsilon k\to 0$ in $L^1([0,T])$ for every $k\in L^1([0,T]^2)$. This is obvious if k is in the space of bounded functions on $[0,T]^2$, a dense subspace of $L^1([0,T]^2)$. So it is enough to show that $\|\mathcal{A}_\epsilon k\|_{L^1([0,T])}\le C\|k\|_{L^1([0,T]^2)}$ for some constant C and for every $k \in L^1([0,T]^2)$. This follows from the inequalities

$$\int_0^T |\mathcal{A}_\epsilon k(t)|\, dt \le \frac{1}{\epsilon}\int_0^T \int_0^T \int_0^T |k(s,r)| 1_{t<s<(t+\epsilon)\wedge T} 1_{t<r<(t+\epsilon)\wedge T} dr\, ds\, dt$$
$$= \frac{1}{\epsilon}\int_0^T \int_0^T |k(s,r)| \left[\int_0^T 1_{(s-\epsilon)^+<t<s} 1_{(r-\epsilon)^+<t<s} dt\right] ds\, dr \le \int_0^T \int_0^T |k(s,r)|\, ds\, dr,$$

since the term in square brackets is less than or equal to ϵ.

This finishes the proof that $J_3^\epsilon \to 0$ in $L^1(\Omega)$, hence in probability. In a similar and simpler way one proves that $J_2^\epsilon \to 0$ in probability.

To finish the proof of the proposition it remains to compute the limit of J_1^ϵ. Exchanging integrals with respect to ds and $\nabla_x u(t+\epsilon, X_{t+\epsilon})(d\theta)$, and then using another change of variable, we have

$$\begin{aligned}
J_1^\epsilon &= \frac{1}{\epsilon}\int_0^{T'} \int_{[-\epsilon,0]} \int_t^{t+\epsilon+\theta} \sigma^i(s,X_s)\, ds \cdot \nabla_x u(t+\epsilon, X_{t+\epsilon})(d\theta)\, dt \\
&= \frac{1}{\epsilon}\int_0^{T'} \int_t^{t+\epsilon} \int_{[s-t-\epsilon,0]} \sigma^i(s,X_s) \cdot \nabla_x u(t+\epsilon, X_{t+\epsilon})(d\theta)\, ds\, dt \\
&= \frac{1}{\epsilon}\int_0^{T'} \int_t^{t+\epsilon} \sigma^i(s,X_s) \cdot \nabla_x u(t+\epsilon, X_{t+\epsilon})([s-t-\epsilon,0])\, ds\, dt \\
&= \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \int_{t-\epsilon}^{t} \sigma^i(s,X_s) \cdot \nabla_x u(t, X_t)([s-t,0])\, ds\, dt \\
&= \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \int_{t-\epsilon}^{t} \sigma^i(t,X_t) \cdot \nabla_x u(t, X_t)([s-t,0])\, ds\, dt \\
&\quad + \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \int_{t-\epsilon}^{t} \{\sigma^i(s,X_s) - \sigma^i(t,X_t)\} \cdot \nabla_x u(t, X_t)([s-t,0])\, ds\, dt \\
&=: H_1^\epsilon + H_2^\epsilon.
\end{aligned}$$

Next, we show that $H_2^\epsilon \to 0$, $\mathbb{P}$-a.s. Since

$$|\nabla_x u(t+\epsilon, X_{t+\epsilon})| \le C\Big(1 + \sup_{t\in[0,T]} |X_t|_{\mathbf{C}}\Big)^m,$$

we have

$$\begin{aligned}
|H_2^\epsilon| &\le C\Big(1 + \sup_{t\in[0,T]} |X_t|_{\mathbf{C}}\Big)^m \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \int_{t-\epsilon}^{t} |\sigma^i(s,X_s) - \sigma^i(t,X_t)|\, ds\, dt \\
&\le C\Big(1 + \sup_{t\in[0,T]} |X_t|_{\mathbf{C}}\Big)^m \int_0^T \frac{1}{\epsilon}\int_{(t-\epsilon)^+}^{t} |\sigma^i(s,X_s) - \sigma^i(t,X_t)|\, ds\, dt.
\end{aligned}$$

Let us fix $\omega \in \Omega$ and note that $\mathbb{P}$-a.s., $\sigma^i(\cdot, X_\cdot) \in L^1([0,T])$. Let us define the operator $\mathcal{B}_\epsilon : L^1([0,T]) \to L^1([0,T])$ as

$$(\mathcal{B}_\epsilon k)(t) = \frac{1}{\epsilon}\int_{(t-\epsilon)^+}^{t} |k(s) - k(t)|\, ds, \qquad k \in L^1([0,T]).$$

Then we have $|H_2^\epsilon| \le C(1+\sup_{t\in[0,T]}|X_t|_{\mathbf{C}})^m \|\mathcal{B}_\epsilon\sigma^i(\cdot,X.)\|_{L^1([0,T])}$, $\mathbb{P}$-a.s., so to prove that $H_2^\epsilon \to 0$ in probability it is enough to show that $\mathcal{B}_\epsilon k \to 0$ in $L^1([0,T])$ for every $k\in L^1([0,T])$. This is obvious if k is in the space of continuous functions on $[0,T]$, a dense subspace of $L^1([0,T])$. So it is enough to show that $\|\mathcal{B}_\epsilon k\|_{L^1([0,T])} \le C\|k\|_{L^1([0,T])}$ for some constant C and for every $k\in L^1([0,T])$. This follows from the inequality

$$|(\mathcal{B}_\epsilon k)(t)| \le \frac{1}{\epsilon}\int_{(t-\epsilon)^+}^t |k(s)|\,ds + |k(t)|,$$

which implies

$$\begin{aligned}
\int_0^T |(\mathcal{B}_\epsilon k)(t)|\,dt &\le \|k\|_{L^1([0,T])} + \frac{1}{\epsilon}\int_0^T\int_{(t-\epsilon)^+}^t |k(s)|\,ds\,dt \\
&= \|k\|_{L^1([0,T])} + \frac{1}{\epsilon}\int_0^T\int_s^{(s+\epsilon)\wedge T} |k(s)|\,dt\,ds \\
&\le \|k\|_{L^1([0,T])} + \int_0^T |k(s)|\,ds = 2\|k\|_{L^1([0,T])}.
\end{aligned}$$

This finishes the proof that $H_2^\epsilon \to 0$ $\mathbb{P}$-a.s., hence in probability.

It remains to consider the term

$$\begin{aligned}
H_1^\epsilon &= \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \sigma^i(t,X_t)\cdot\int_{t-\epsilon}^t \nabla_x u(t,X_t)([s-t,0])\,ds\,dt \\
&= \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \sigma^i(t,X_t)\cdot\int_{t-\epsilon}^t\int_{[s-t,0]} \nabla_x u(t,X_t)(d\theta)\,ds\,dt \\
&= \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \sigma^i(t,X_t)\cdot\int_{[-\epsilon,0]}\int_{t-\epsilon}^{t+\theta} ds\nabla_x u(t,X_t)(d\theta)\,dt \\
&= \frac{1}{\epsilon}\int_\epsilon^{T'+\epsilon} \sigma^i(t,X_t)\cdot\int_{[-\epsilon,0]} (\theta+\epsilon)\,\nabla_x u(t,X_t)(d\theta)\,dt \\
&= \int_\epsilon^{T'+\epsilon} \sigma^i(t,X_t)\cdot\int_{[-r,0]} \left(1+\frac{\theta}{\epsilon}\right)^+ \nabla_x u(t,X_t)(d\theta)\,dt.
\end{aligned}$$

We clearly have, $\mathbb{P}$-a.s.,

$$\begin{aligned}
\int_{[-r,0]} \left(1+\frac{\theta}{\epsilon}\right)^+ \nabla_x u(t,X_t)(d\theta) &\to \int_{[-r,0]} 1_{\{0\}}(\theta)\,\nabla_x u(t,X_t)(d\theta) \\
&= \nabla_x u(t,X_t)(\{0\}) = \nabla_0 u(t,X_t),
\end{aligned}$$

and by dominated convergence, $\mathbb{P}$-a.s.,

$$H_1^\epsilon \to \int_0^{T'} \sigma^i(t,X_t)\cdot\nabla_0 u(t,X_t)\,dt.$$

This shows that C^ϵ converges in probability and its limit is

$$\langle u(\cdot,X.),W^i\rangle_{[0,T']} = \int_0^{T'} \sigma^i(t,X_t)\cdot\nabla_0 u(t,X_t)\,dt. \qquad \square$$

4. The forward-backward system with delay. In this section we will discuss existence, uniqueness, and regular dependence on the initial data of the following forward-backward system: for given $t \in [0,T]$ and $x \in \mathbf{C} = C([-r,0];\mathbb{R}^n)$,

$$
\left\{
\begin{array}{l}
dy_\tau = b(\tau, X_\tau)\, d\tau + \sigma(\tau, X_\tau)\, dW_\tau, \quad \tau \in [t,T] \subset [0,T], \\
X_t = x, \\
dY_\tau = \psi(\tau, X_\tau, Y_\tau, Z_\tau)\, d\tau + Z_\tau\, dW_\tau, \\
Y_T = \phi(X_T).
\end{array}
\right. \tag{4.1}
$$

Here we use the notation $X_\tau(\theta) = y_{\tau+\theta}$, $\theta \in [-r,0]$, as before, so the first equation in (4.1) is the same as (2.4). We extend the definition of X setting $X_s = x$ for $0 \leq s \leq t$. The second equation in (4.1), namely,

$$
\left\{
\begin{array}{ll}
dY_\tau = \psi(\tau, X_\tau, Y_\tau, Z_\tau)\, d\tau + Z_\tau\, dW_\tau, & \tau \in [0,T], \\
Y_T = \phi(X_T),
\end{array}
\right. \tag{4.2}
$$

is of backward type. Under suitable assumptions on the coefficients $\psi : [0,T] \times \mathbf{C} \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ and $\phi : \mathbf{C} \to \mathbb{R}$, we will look for a solution consisting of a pair of predictable processes, taking values in $\mathbb{R} \times \mathbb{R}^d$, such that Y has continuous paths and

$$
\|\, (Y,Z)\, \|^2_{\mathbb{K}_{cont}} := \mathbb{E} \sup_{\tau \in [0,T]} |Y_\tau|^2 + \mathbb{E} \int_0^T |Z_\tau|^2\, d\tau < \infty;
$$

see, e.g., [24]. In the following, we denote by $\mathbb{K}_{cont}\,([0,T])$ the space of such processes.

The solution of (4.1) will be denoted by $(X_\tau, Y_\tau, Z_\tau)_{\tau \in [0,T]}$, or, to stress the dependence on the initial time t and on the initial datum x, by $(X^{t,x}_\tau, Y^{t,x}_\tau, Z^{t,x}_\tau)_{\tau \in [0,T]}$.

HYPOTHESIS 4.1. *The maps* $\psi : [0,T] \times \mathbf{C} \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ *and* $\phi : \mathbf{C} \to \mathbb{R}$ *are Borel measurable and satisfy the following assumptions:*

1. *there exists* $L > 0$ *such that*

$$
\begin{aligned}
|\psi(t,x,y,z_1) - \psi(t,x,y,z_2)| &\leq L\, |z_1 - z_2|\,, \\
|\psi(t,x,y_1,z) - \psi(t,x,y_2,z)| &\leq L\, |y_1 - y_2|
\end{aligned}
$$

   for every $t \in [0,T]$, $x \in \mathbf{C}$, $y, y_1, y_2 \in \mathbb{R}$, *and* $z, z_1, z_2 \in \mathbb{R}^d$*;*
2. $\psi(t,\cdot,\cdot,\cdot) \in \mathcal{G}^1\left(\mathbf{C} \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R}\right)$ *for every* $t \in [0,T]$*;*
3. *there exist* $K > 0$ *and* $m \geq 0$ *such that*

$$
|\nabla_x \psi(t,x,y,z)| \leq K\,(1 + |x|_{\mathbf{C}} + |y|)^m\,(1 + |z|)
$$

   for every $t \in [0,T]$, $x \in \mathbf{C}$, $y \in \mathbb{R}$, *and* $z \in \mathbb{R}^d$*;*
4. $\phi \in \mathcal{G}^1(\mathbf{C}, \mathbb{R})$ *and there exist* $K > 0$ *and* $m \geq 0$ *such that*

$$
|\nabla \phi(x)| \leq K\,(1 + |x|_{\mathbf{C}})^m\,, \qquad x \in \mathbf{C}.
$$

Under these assumptions we can state a result on existence and uniqueness of a solution of the forward-backward system (4.1) and on its regular dependence on x.

PROPOSITION 4.2. *Assume that Hypotheses* 2.2 *and* 4.1 *hold true. Then the forward-backward system* (4.1) *admits a unique solution* $(X^{t,x}, Y^{t,x}, Z^{t,x}) \in \mathcal{S}^p([0,T];\mathbf{C}) \times \mathbb{K}_{cont}\,([0,T])$ *for every* $(t,x) \in [0,T] \times \mathbf{C}$. *Moreover, the map* $(t,x) \mapsto (X^{t,x}, Y^{t,x}, Z^{t,x})$ *belongs to the space* $\mathcal{G}^1\,([0,T] \times \mathbf{C}, \mathcal{S}^p([0,T];\mathbf{C}) \times \mathbb{K}_{cont}\,([0,T]))$. *Finally, the following estimate holds true: for every* $p \geq 2$ *there exists* $C > 0$ *such that*

$$
\left[\mathbb{E} \sup_{\tau \in [0,T]} \left|\nabla_x Y^{t,x}_\tau\right|^p\right]^{1/p} \leq C\left(1 + |x|^{(m+1)^2}_{\mathbf{C}}\right), \qquad t \in [0,T], x \in \mathbf{C}.
$$

Proof. We give only a sketch of the proof. The forward equation has a unique solution by Theorem 2.3. Existence and uniqueness of the solution of the backward equation follows from the classical result [24].

In Theorem 2.3 we have shown that the map $x \mapsto X^{t,x}$ belongs to $\mathcal{G}^1\left(\mathbf{C}, \mathcal{S}^p\left([0,T]; \mathbf{C}\right)\right)$ for every $2 \leq p < \infty$. Then the proof of continuity and differentiability of $(t,x) \mapsto (X^{t,x}, Y^{t,x}, Z^{t,x})$ in the appropriate norms, as well as the final estimate on $\nabla_x Y_\tau^{t,x}$, can be achieved as in Proposition 5.2 in [11]. The only difference is that in [11] the process $X^{t,x}$ takes values in a Hilbert space, while in our context it takes values in the Banach space $\mathbf{C}$; nevertheless, the same arguments apply (see also [17] for a similar result in Banach spaces). ∎

COROLLARY 4.3. *Assume that Hypotheses* 2.2 *and* 4.1 *hold true. Then the function* $v : [0,T] \times \mathbf{C} \to \mathbb{R}$ *defined by*

$$v(t,x) = Y_t^{t,x}, \qquad t \in [0,T], x \in \mathbf{C}, \tag{4.3}$$

belongs to $\mathcal{G}^{0,1}\left([0,T] \times \mathbf{C}; \mathbb{R}\right)$*. Moreover, there exists* $C > 0$ *such that*

$$|\nabla_x v\,(t,x)| \leq C\left(1 + |x|_{\mathbf{C}}^{(m+1)^2}\right), \qquad t \in [0,T]\,, x \in \mathbf{C}.$$

Finally, for every $t \in [0,T]$ *and* $x \in \mathbf{C}$*, we have,* $\mathbb{P}$*-a.s,*

$$Y_s^{t,x} = v\left(s, X_s^{t,x}\right) \qquad \text{for every } s \in [t,T], \tag{4.4}$$

$$Z_s^{t,x} = \nabla_0 v\left(s, X_s^{t,x}\right)\sigma(s, X_s^{t,x}) \qquad \text{for a.e. } s \in [t,T]. \tag{4.5}$$

Proof. It is well known that $v(t,x)$ is deterministic, and its properties are, therefore, a direct consequence of Proposition 4.2. Equality (4.4) is also a standard consequence of uniqueness of the solution of the backward equation.

To prove (4.5) we consider the joint quadratic variation of $Y^{t,x}$ and the Wiener process W^i on an interval $[t,T']$, with $T' < T$. Taking into account the backward equation we obtain

$$\langle Y^{t,x}, W^i\rangle_{[t,T']} = \int_t^{T'} Z_s^i\, ds.$$

By Theorem 3.1 we have

$$\langle v\left(\cdot, X_\cdot^{t,x}\right), W^i\rangle_{[t,T']} = \int_t^{T'} \sigma^i(s, X_s^{t,x}) \cdot \nabla_0 v(s, X_s^{t,x})\, ds,$$

so that (4.4) implies (4.5). ∎

Remark 4.4. If we strengthen slightly the regularity assumptions and require that, for all $t \in [0,T]$, the functions $b(t,\cdot), \sigma(t,\cdot), \phi$ are continuously Fréchet differentiable on $\mathbf{C}$ and $\psi(t,\cdot,\cdot,\cdot)$ is continuously Fréchet differentiable on $\mathbf{C} \times \mathbb{R} \times \mathbb{R}^d$, then we can prove, with only minor changes in the proofs, that the function v defined in (4.3) is Fréchet differentiable with respect to x and the Fréchet derivative is a continuous function from $[0,T] \times \mathbf{C}$ to the dual space $\mathbf{C}^*$ with respect to the usual norm (i.e., the variation norm).

Remark 4.5. In the context of Proposition 4.2, the law of the solution $(X^{t,x}, Y^{t,x}, Z^{t,x})$ is uniquely determined by x and the coefficients b, σ, ψ, ϕ. Since $v(t,x)$ is deterministic, hence determined by its law, we conclude that the function v is a functional of the coefficients b, σ, ψ, ϕ and does not depend on the particular choice of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ nor on the Wiener process W.

5. Nonlinear parabolic equations. Let us consider again the Markov process $\{X_\tau^{t,x},\ 0 \leq t \leq \tau \leq T,\ x \in \mathbf{C}\}$, defined by the formula (2.5), starting from the family of solutions to (2.4). Let us denote by $(\mathcal{L}_t)_{t\in[0,T]}$ the corresponding generator. Thus, each $\mathcal{L}_t$ is a second order differential operator acting on a suitable domain consisting of real functions defined on $\mathbf{C}$. In the autonomous case, a description of the generator, denoted by $\mathcal{L}$, was given in section 2, Remark 2.4. In this section we treat semilinear parabolic equations driven by $(\mathcal{L}_t)$, which are generalizations of the Kolmogorov equations. We will introduce a concept of solution, called the mild solution, that does not require a description of the generators. In what follows the notation $\mathcal{L}_t$ will be used only in a formal way.

The parabolic equations that we study have the following form:

$$\begin{cases} \dfrac{\partial v(t,x)}{\partial t} + \mathcal{L}_t v(t,x) = \psi(t,x,v(t,x),\nabla_0 v(t,x)\,\sigma(t,x)), \\ v(T,x) = \phi(x), \qquad t\in[0,T],\ x\in\mathbf{C}, \end{cases} \tag{5.1}$$

with unknown function $v:[0,T]\times\mathbf{C}\to\mathbb{R}$ and given coefficients $\psi:[0,T]\times\mathbf{C}\times\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}$ and $\phi:\mathbf{C}\to\mathbb{R}$. We recall the notation $\nabla_0 v(t,x)$ introduced in (2.1). We note, in particular, that $\nabla_0 v(t,x)$ is a vector in $\mathbb{R}^n$ whose components are denoted $\nabla_0^k v(t,x)$ $(k=1,\ldots,n)$. If we denote by $\sigma_k^i(t,x)$ $(k=1,\ldots,n,\ i=1,\ldots,d)$ the components of the matrix $\sigma(t,x)$, then $\nabla_0 v(t,x)\,\sigma(t,x)$ denotes the vector in $\mathbb{R}^d$ whose components are $\sum_{k=1}^n \nabla_0^k v(t,x)\sigma_k^i(t,x)$, $(i=1,\ldots,d)$.

Recalling the definition of the transition semigroup $P_{t,\tau}$ given in (2.6), and writing the variation of constants formula for a solution to (5.1), we formally obtain

$$v(t,x) = P_{t,T}[\phi](x) - \int_t^T P_{t,\tau}\Big[\psi(\cdot,v(\tau,\cdot),\nabla_0 v(\tau,\cdot)\,\sigma(\tau,\cdot))\Big](x)\,d\tau, \qquad t\in[0,T],\ x\in\mathbf{C}. \tag{5.2}$$

We notice that this formula is meaningful if $\nabla_0 v$ is well defined and provided ϕ and ψ satisfy some growth and measurability conditions. This way we arrive at the following definition of a mild solution of the semilinear Kolmogorov equation (5.1).

DEFINITION 5.1. *A function $v:[0,T]\times\mathbf{C}\to\mathbb{R}$ is a mild solution of the semilinear Kolmogorov equation* (5.1) *if $v\in\mathcal{G}^1([0,T]\times\mathbf{C};\mathbb{R})$, there exist $C>0$, $q\geq 0$ such that*

$$|v(t,x)| + |\nabla_x v(t,x)| \leq C(1+|x|)^q, \qquad t\in[0,T],\ x\in\mathbf{C}, \tag{5.3}$$

and the equality (5.2) *holds.*

The space $\mathcal{G}^1([0,T]\times\mathbf{C};\mathbb{R})$ was described in Remark 2.1. $|\nabla_x v(t,x)|$ denotes the total variation norm of the $\mathbb{R}^n$-valued finite Borel measure $\nabla_x v(t,x)$ on $[-r,0]$.

THEOREM 5.2. *Assume that Hypotheses* 2.2 *and* 4.1 *hold true. Then there exists a unique mild solution v of* (5.1)*. The function v coincides with the one introduced in Corollary* 4.3.

Proof. The proof is similar to the proof of Theorem 6.2 in [11] and Theorem 6.1 in [12]; nevertheless, we give a sketch below.

At first we prove existence. For fixed $t\in[0,T]$ and $x\in\mathbf{C}$, let $(X_\tau^{t,x},Y_\tau^{t,x},Z_\tau^{t,x})_{\tau\in[0,T]}$ denote the solution of the forward-backward system (4.1), and let $v(t,x)$ be defined by equality (4.3). The required regularity and growth conditions of the function v were proved in Corollary 4.3, so it remains to prove that v satisfies equality (5.2). Since the pair $(Y^{t,x},Z^{t,x})$ is a solution to the backward equation (4.2), we have

$$Y_t^{t,x} + \int_t^T Z_\tau^{t,x}dW_\tau = \phi\left(X_T^{t,x}\right) + \int_t^T \psi\left(\tau,X_\tau^{t,x},Y_\tau^{t,x},Z_\tau^{t,x}\right)d\tau.$$

Taking expectation and applying formulae (4.4) and (4.5), we get the equality (5.2).

It remains to prove uniqueness. Let v be a mild solution to (5.1), so that for every $s \in [t,T] \subset [0,T]$,

$$v(s,x) = \mathbb{E}\phi(X_T^{s,x}) + \mathbb{E}\int_s^T \psi(\tau, X_\tau^{s,x}, v(\tau, X_\tau^{s,x}), \nabla_0 v(\tau, X_\tau^{s,x})\,\sigma(\tau, X_\tau^{s,x}))\,d\tau.$$

By the Markov property of $X_\tau^{s,x}$ it follows that

$$v\left(s, X_s^{t,x}\right) = \mathbb{E}^{\mathcal{F}_s}\eta - \mathbb{E}^{\mathcal{F}_s}\int_t^s \psi\left(\tau, X_\tau^{t,x}, v(\tau, X_\tau^{t,x}), \nabla_0 v\left(\tau, X_\tau^{t,x}\right)\sigma\left(\tau, X_\tau^{t,x}\right)\right)d\tau,$$

where we have defined $\eta = \phi\left(X_T^{t,x}\right) + \int_t^T \psi(\tau, X_\tau^{t,x}, v(\tau, X_\tau^{t,x}), \nabla_0 v(\tau, X_\tau^{t,x})\,\sigma(\tau, X_\tau^{t,x}))\,d\tau$. By the representation theorem of martingales (see, e.g., [6, Theorem 8.2]), there exists a predictable process $\widetilde{Z} \in L^2\left(\Omega \times [0,T], \mathbb{R}^d\right)$, such that $\mathbb{E}^{\mathcal{F}_s}\eta = \int_t^s \widetilde{Z}_\tau dW_\tau + v(t,x)$, $s \in [t,T]$. So

$$v\left(s, X_s^{t,x}\right) = v(t,x) + \int_t^s \widetilde{Z}_\tau dW_\tau - \int_t^s \psi\left(\tau, X_\tau^{t,x}, v(\tau, X_\tau^{t,x}), \nabla_0 v\left(\tau, X_\tau^{t,x}\right)\sigma(\tau, X_\tau^{t,x})\right)d\tau. \tag{5.4}$$

Now we compute the joint quadratic variation with W^i of the processes occurring at both sides of this equality, on an interval $[t,T'] \subset [t,T)$. Considering the right-hand side we obtain $\int_t^{T'} \widetilde{Z}_\tau^i d\tau$ by the rules of stochastic calculus. By Theorem 3.1 we have $\langle v(\cdot, X_\cdot^{t,x}), W^i\rangle_{[t,T']} = \int_t^{T'} \sigma^i(\tau, X_\tau^{t,x})\nabla_0 v(\tau, X_\tau^{t,x})\ d\tau$. Therefore, we have $\widetilde{Z}_\tau = \sigma(\tau, X_\tau^{t,x})\nabla_0 v(\tau, X_\tau^{t,x})$ and substituting in (5.4) we get

$$\begin{aligned} v(s, X_s^{t,x}) = \phi\left(X_T^{t,x}\right) &- \int_s^T \nabla_0 v\left(\tau, X_\tau^{t,x}\right)\sigma(\tau, X_\tau^{t,x})dW_\tau \\ &+ \int_s^T \psi\left(\tau, X_\tau^{t,x}, v(\tau, X_\tau^{t,x}), \nabla_0 v\left(\tau, X_\tau^{t,x}\right)\sigma(\tau, X_\tau^{t,x})\right)d\tau. \end{aligned}$$

By comparing with the backward equation in (4.1) we see that the pairs of processes $(Y_s^{t,x}, Z_s^{t,x})$ and $(v(s, X_s^{t,x}), \nabla_0 v(s, X_s^{t,x})\sigma(s, X_s^{t,x}))$, $s \in [t,T]$, solve the same equation. By uniqueness of the solution we have $Y_s^{t,x} = v(s, X_s^{t,x})$, $s \in [t,T]$, and for $s = t$ we get $Y_t^{t,x} = v(t,x)$. □

Remark 5.3. The proof of uniqueness is based on an application of Theorem 3.1. Inspection of the proof shows that uniqueness holds in a larger class of functions. Namely, if a Borel measurable function $v : [0,T] \times \mathbf{C} \to \mathbb{R}$ satisfies $v(t,\cdot) \in \mathcal{G}^1(\mathbf{C}, \mathbb{R})$ for every $t \in [0,T]$, and the inequality

$$|v(t,x)| + |\nabla_x v(t,x)| \leq C(1+|x|)^q, \qquad t \in [0,T], x \in \mathbf{C},$$

holds for some $C > 0$, $q \geq 0$, and (5.2) holds, then v coincides with the solution constructed in Theorem 3.1.

6. Application to stochastic optimal control. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space, satisfying the usual conditions, and let W be an $\mathbb{R}^d$-valued standard Wiener process with respect to $(\mathcal{F}_t)$ and $\mathbb{P}$. We consider the following controlled functional stochastic equation on an interval $[t,T] \subset [0,T]$:

$$\begin{cases} dy_s^u = b(s, y_{s+\cdot}^u)\,ds + \sigma(s, y_{s+\cdot}^u)\,[h(s, y_{s+\cdot}^u, u_s)\,ds + \ dW_s], \\ y_{t+\theta}^u = x(\theta), \quad \theta \in [-r, 0]. \end{cases} \tag{6.1}$$

The coefficients b and σ satisfy the previous assumptions. $u(\cdot)$ denotes the control and y^u the corresponding solution. We assume that the controls are $(\mathcal{F}_t)$-predictable processes with values in a given measurable space $(U,\mathcal{U})$. The function $h : [0,T] \times \mathbf{C} \times U \to \mathbb{R}^d$ is measurable and bounded. We introduce again the process

$$X_s^u = y_{s+\cdot}^u = \{y_{s+\theta}^u,\ \theta \in [-r,0]\}, \qquad s \in [t,T], \tag{6.2}$$

which now depends on the control and takes values in $\mathbf{C} = C([-r,0];\mathbb{R}^n)$, so that (6.1) can be rewritten as

$$\begin{cases} dy_s^u = b(s,X_s^u)\, ds + \sigma(s,X_s^u)\,[h(s,X_s^u,u_s)\, ds +\ dW_s], & s \in [t,T], \\ X_t = x. \end{cases} \tag{6.3}$$

We introduce the cost functional to minimize

$$J(t,x,u(\cdot)) = \mathbb{E}\int_t^T g(u_s)\, ds + \mathbb{E}\,\phi(y_{T+\cdot}^u) = \mathbb{E}\int_t^T g(u_s)\, ds + \mathbb{E}\,\phi(X_T^u), \tag{6.4}$$

where $g : U \to [0,\infty)$ and $\phi : \mathbf{C} \to \mathbb{R}$ are given functions.

Remark 6.1. Without any substantial change, we could consider more general cost functionals of the form

$$J(t,x,u(\cdot)) = \mathbb{E}\int_t^T [\ell(y_s^u) + g(u_s)]\ ds + \mathbb{E}\phi(y_{T+\cdot}^u), \tag{6.5}$$

where $\ell : \mathbb{R}^n \to \mathbb{R}$. In fact, this kind of cost can be put in the form (6.4) as follows: first note that in (6.1) we can assume $r \geq T$, possibly extending the functions b and σ in the obvious way; next, we define, for $x \in \mathbf{C}$,

$$\phi_0(x) = \int_{t-T}^0 \ell(x(s))\, ds,$$

so that $\phi_0(X_T^u) = \int_t^T \ell(y_s^u)\, ds$, and we conclude that

$$J(t,x,u(\cdot)) = \mathbb{E}\int_t^T g(u_s)\, ds + \mathbb{E}[(\phi_0 + \phi)(X_T^u)],$$

which has the required form. In a similar way, under suitable assumptions, one could consider even more general costs of the form

$$J(t,x,u(\cdot)) = \mathbb{E}\int_t^T \ell(s,y_s^u,u_s)\, ds + \mathbb{E}\,\phi(y_{T+\cdot}^u).$$

However, we limit ourselves to cost functionals with the structure of (6.4).

To proceed further we need to introduce the Hamiltonian function $\psi : [0,T] \times \mathbf{C} \times \mathbb{R}^d \to \mathbb{R}$ defined, for $t \in [0,T]$, $x \in \mathbf{C}$, $z \in \mathbb{R}^d$, by the formula

$$\psi\,(t,x,z) = \inf\,\{g(u) + zh\,(t,x,u) : u \in U\} \tag{6.6}$$

and the corresponding, possibly empty, set of minimizers

$$\Gamma\,(t,x,z) = \{u \in U,\ \ g(u) + zh\,(t,x,u) = \psi\,(t,x,z)\}\,. \tag{6.7}$$

Remark 6.2. By the Filippov Theorem (see, e.g., [2, Theorem 8.2.10, p. 316]), if U is a complete metric space equipped with its Borel σ-algebra, g is continuous, h is measurable bounded, with $u \mapsto h(t,x,u)$ continuous on U, and if Γ takes nonempty values (as is always the case if U is compact), then Γ admits a measurable selection; i.e., there exists a Borel measurable map $\Gamma_0 : [0,T] \times \mathbf{C} \times \mathbb{R}^d \to U$ such that $\Gamma_0(t,x,z) \in \Gamma(t,x,z)$ for $t \in [0,T]$, $x \in E$, $z \in \mathbb{R}^d$.

We are now ready to formulate the assumptions we need.

HYPOTHESIS 6.3.

1. *$(U,\mathcal{U})$ is a measurable space, $g : U \to [0,\infty)$ is measurable, $h : [0,T]\times\mathbf{C}\times U \to \mathbb{R}^d$ is measurable and bounded.*
2. *The Hamiltonian ψ defined in* (6.6) *satisfies the requirements of points* 2 *and* 3 *of Hypothesis* 4.1.
3. *The function $\phi : \mathbf{C} \to \mathbb{R}$ satisfies the requirements of point* 4 *in Hypothesis* 4.1, *namely, it belongs to $\mathcal{G}^1(\mathbf{C},\mathbb{R})$ and there exist $K > 0$ and $m \geq 0$ such that*
$$|\nabla\phi(x)| \leq K\left(1 + |x|_{\mathbf{C}}\right)^m, \qquad x \in \mathbf{C}.$$

Remark 6.4.

1. Hypothesis 6.3 is stronger than Hypothesis 4.1. Indeed, point 1 of Hypothesis 4.1 is a straightforward consequence of the fact that h is assumed to be bounded.
2. In the case $U \subset \mathbb{R}^k$, $h(t,x,u) = u$, the previous assumptions require, in particular, that the set U where control processes take values should be bounded.
3. The assumptions on the Hamiltonian function ψ can be easily verified in specific cases. For instance, if $h(t,x,u) = u$ as before, U is a closed ball of $\mathbb{R}^k$ centered at the origin, and $g(u) = g_0(|u|^p)$ for some $p > 1$ and some convex function $g_0 : [0,\infty) \to [0,\infty)$ such that $g \in C^1([0,\infty))$ and $g'(0) > 0$, then the Hamiltonian is differentiable with respect to z and ψ satisfies points 2 and 3 of Hypothesis 4.1.

Now let us consider a probability space $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$, a standard Wiener process $\widetilde{W}$ in $\mathbb{R}^d$, and the following forward-backward system:

$$\begin{cases} dy_\tau = b(\tau,X_\tau)\,d\tau + \sigma(\tau,X_\tau)\,d\widetilde{W}_\tau, \quad \tau \in [t,T] \subset [0,T], \\ X_t = x, \\ dY_\tau = \psi(X_\tau,Z_\tau)\,d\tau + Z_\tau\,d\widetilde{W}_\tau, \\ Y_T = \phi(X_T). \end{cases} \tag{6.8}$$

By Remark 4.5, the function $v : [0,T] \times \mathbf{C} \to \mathbb{R}$ defined by the equality

$$v(t,x) = Y_t^{t,x} \tag{6.9}$$

is a functional of the coefficients b, σ, ψ, ϕ and does not depend on the particular choice of $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$ nor on the Wiener process $\widetilde{W}$.

In the following proposition we show that the function v, defined in this way by means of an appropriate forward-backward stochastic differential system, plays a basic role in the control problem.

To begin we notice that if ψ is the Hamiltonian defined in (6.6) and ϕ is the final cost in functional (6.4), then (5.1) is the Hamilton–Jacobi–Bellman equation related to the the present stochastic optimal control problem. In particular, Theorem (5.2) implies that v defined in (6.9) is the unique mild solution of Hamilton–Jacobi–Bellman equation (5.1).

Then we obtain, by a customary Girsanov transform argument (see [9]), a version of the so-called *fundamental relation*.

PROPOSITION 6.5. *Assume that Hypotheses* 2.2 *and* 6.3 *hold true, and that the cost functional is given in* (6.4). *Let v be defined in* (6.9). *Then for every $t \in [0,T]$ and $x \in \mathbf{C}$ and for every admissible control $u(\cdot)$, we have*

$$v(t,x) = J(t,x,u(\cdot)) + \mathbb{E}\int_t^T [\psi(s, X_s^u, Z_s^u) - Z_s^u h(s, X_s^u, u_s) - g(u_s)]\, ds. \tag{6.10}$$

In particular, $v(t,x) \le J(t,x,u(\cdot))$.

Proof. The proof follows from the same arguments used in the proof of Theorem 7.2 in [11] and is, therefore, omitted. ∎

The equality (6.10) immediately gives the following consequences.

PROPOSITION 6.6. *Let $t \in [0,T]$ and $x \in \mathbf{C}$ be fixed. Assume that the set-valued map Γ has nonempty values and assume that $\Gamma_0 : [0,T] \times \mathbf{C} \times \mathbb{R}^d \to U$ is a measurable selection. Moreover, suppose that a control $u(\cdot)$ satisfies*

$$u_\tau = \Gamma_0(\tau, X_\tau^u, Z_\tau^u), \qquad \mathbb{P}\text{-a.s. for a.e. } \tau \in [t,T]. \tag{6.11}$$

Then $J(t,x,u(\cdot)) = v(t,x)$ (thus $u(\cdot)$ is optimal), and the optimal pair $(u(\cdot), X^u)$ satisfies the feedback law

$$u_\tau = \Gamma_0(\tau, X_\tau^u, \nabla_0 v(\tau, X_\tau^u)\sigma(\tau, X_\tau^u)), \qquad \mathbb{P}\text{-a.s. for a.e. } \tau \in [t,T]. \tag{6.12}$$

We note that (6.12) follows from (6.11) and (4.5).

However, we cannot prove the existence of an optimal control satisfying (6.11) (and hence (6.12)). Such a control can be shown to exist if there exists a solution to the so-called closed-loop equation

(6.13)

$$\begin{cases} dy_\tau = b(\tau, X_\tau)\, d\tau + \sigma(\tau, X_\tau)[h(\tau, X_\tau, \Gamma_0(\tau, X_\tau, \nabla_0 v(\tau, X_\tau)\, \sigma(\tau, X_\tau)))d\tau + dW_\tau], \\ \hfill \tau \in [t,T], \\ X_t(\theta) = x(\theta), \qquad \theta \in [-r, 0], \end{cases}$$

since in this case one can define an optimal control setting

$$u_\tau = \Gamma_0(\tau, X_\tau, \nabla_0 v(\tau, X_\tau)\sigma(\tau, X_\tau)).$$

However, under the present assumptions, we cannot guarantee that the closed-loop equation has a solution in the usual strong sense. To circumvent this difficulty we will revert to a weak formulation of the optimal control problem.

6.1. Weak formulation of the optimal control problem. We formulate the optimal control problem in the weak sense following the approach of [10, see, e.g., Chapter III]. The main advantage is that we will be able to solve the closed loop equation in a weak sense, and hence to find an optimal control, even if the feedback law is nonsmooth.

Initially, we are given the set U and the functions b, σ, h, g, ϕ. By an *admissible control system* we mean

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P}, W, u(\cdot), X^u),$$

where $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is a filtered probability space satisfying the usual conditions, W is an $\mathbb{R}^d$-valued standard Wiener process with respect to $(\mathcal{F}_t)$ and $\mathbb{P}$, u is an $(\mathcal{F}_t)$-predictable process with values in U, X^u satisfies (6.2)–(6.3). An admissible control system will be briefly denoted by (W, u, X^u) in the following. Our aim is now to minimize the cost functional

$$J(t, x, (W, u, X^u)) = \mathbb{E} \int_t^T g(u_s)\, ds + \mathbb{E}\, \phi(X_T^u) \tag{6.14}$$

over all the admissible control systems (W, u, X^u). We can prove the following results.

THEOREM 6.7. *Assume that Hypotheses* 2.2 *and* 6.3 *hold true, and that the cost functional is given in* (6.14). *Let v be defined in* (6.9). *Then for every $t \in [0, T]$ and $x \in \mathbf{C}$ and for all admissible control system (W, u, X^u), we have*

$$J(t, x, (W, u, X^u)) \geq v(t, x)$$

and the equality holds if and only if

$$u_\tau \in \Gamma(\tau, X_\tau^u, \nabla_0 v(\tau, X_\tau^u)\sigma(\tau, X_\tau^u)), \qquad \mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T]\,.$$

Moreover, assume that the set-valued map Γ has nonempty values and it admits a measurable selection $\Gamma_0 : [0, T] \times \mathbf{C} \times \mathbb{R}^d \to U$. Then an admissible control system (W, u, X^u) satisfying the feedback law

$$u_\tau = \Gamma_0(X_\tau^u, \nabla_0 v(\tau, X_\tau^u)\sigma(\tau, X_\tau^u)), \qquad \mathbb{P}\text{-a.s. for a.a. } \tau \in [t, T]$$

is optimal.

Finally, the closed loop equation (6.13) *admits a weak solution $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}, W, X)$ which is unique in law and setting*

$$u_\tau = \Gamma_0\left(\tau, X_\tau, \nabla_0 v(\tau, X_\tau)\sigma(\tau, X_\tau)\right),$$

we obtain an optimal admissible control system (W, u, X).

Proof. The proof follows from the fundamental relation (6.10) and the same arguments leading to Proposition 6.6 and the remarks following it. The only difference here is the solvability of the closed loop equation in a weak sense, which is, however, a standard application of a Girsanov change of measure. ◻

7. Application to pricing. We consider a financial market, of Black and Scholes type, with one risky asset, whose price at time t is denoted by S_t, and one non-risky asset, whose price is denoted by B_t. We assume the following prices evolution:

$$\begin{cases} dS_t = \mu(t, S_{t+\cdot})\, S_t\, dt \;+\; \sigma(t, S_{t+\cdot})\, S_t\, dW_t, & t \in [0, T], \\ S_\theta = s_\theta, & \theta \in [-r, 0], \\ dB_t = \rho B_t\, dt, & t \in [0, T], \\ B_0 = 1, \end{cases} \tag{7.1}$$

where $\rho > 0$, $r > 0$, and $s \in \mathbf{C} = C([-r, 0], \mathbb{R})$. We notice that the coefficients μ and σ depend on the past trajectory: $S_{t+\cdot}$ stands for the past trajectory of length r, i.e., $S_{t+\cdot} = (S_{t+\theta})_{\theta\in[-r,0]}$. Moreover, we consider a contingent claim of the form

$$\phi(S_{T+\cdot}),$$

where $\phi : \mathbf{C} \to \mathbb{R}$. If $r > T$, then the claim depends on the whole evolution in time of the prices of the shares; see [3], [20], or [30] and the references within for a general discussion on such kinds of options, usually referred to as *path-dependent*.

We denote by π_t the value of the investor's portfolio invested in the risky asset at time t. π is called a trading strategy; we will consider only predictable trading strategies which are square-integrable, i.e., $\mathbb{E} \int_0^T |\pi_t|^2 dt < \infty$. We notice that the value V_t of the corresponding self-financing portfolio satisfies

$$dV_t = \rho V_t \, dt + \pi_t \, \sigma(t, S_{t+\cdot}) \, \theta(t, S_{t+\cdot}) \, dt + \pi_t \, \sigma(t, S_{t+\cdot}) \, dW_t, \tag{7.2}$$

where

$$\theta(t, S_{t+\cdot}) = \frac{\mu(t, S_{t+\cdot}) - \rho}{\sigma(t, S_{t+\cdot})} \tag{7.3}$$

is called the risk premium.

At time T the investor has to pay a contingent claim of the form $\phi(S_{T+\cdot})$, where $\phi : \mathbf{C} \to \mathbb{R}$ is some given function. The pricing problem is to find and characterize pairs (π, V_0) consisting of a strategy π and an initial capital $V_0 \in \mathbb{R}$ such that

$$V_T = \phi(S_{T+\cdot}).$$

π is then called a hedging strategy and V_0 is called the fair price of the claim at time $t = 0$.

Throughout this section we assume the following.

HYPOTHESIS 7.1.

1. $(W_t)_{t\geq 0}$ *is a real Wiener process defined in a complete probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $(\mathcal{F}_t)_{t\geq 0}$ *is the filtration generated by* W *augmented with null sets.*
2. $\mu : [0, T] \times \mathbf{C} \to \mathbb{R}$ *is Borel measurable and bounded, and there exists* $L > 0$ *such that*
$$|\mu(t, f^1) f^1(0) - \mu(t, f^2) f^2(0)| \leq L |f_1 - f_2|_{\mathbf{C}} \tag{7.4}$$
for all $t \in [0, T]$, $f^1, f^2 \in \mathbf{C}$*; moreover,* $\mu(t, \cdot) \in \mathcal{G}^1(\mathbf{C}, \mathbb{R})$ *for all* $t \in [0, T]$.
3. $\sigma : [0, T] \times \mathbf{C} \to \mathbb{R}$ *is Borel measurable and there exists* $c > 0$ *such that*
$$|\sigma(t, f)| \geq c, \tag{7.5}$$
for every $f \in \mathbf{C}$*, so that the risk premium in* (7.3) *is well defined and bounded; moreover,*
$$|\sigma(t, f^1) f^1(0) - \sigma(t, f^2) f^2(0)| \leq L |f_1 - f_2|_{\mathbf{C}} \tag{7.6}$$
for a suitable $L > 0$ *and for all* $t \in [0, T]$, $f^1, f^2 \in \mathbf{C}$*; finally,* $\sigma(t, \cdot) \in \mathcal{G}^1(\mathbf{C}, \mathbb{R})$ *for all* $t \in [0, T]$
4. $\phi \in \mathcal{G}^1(\mathbf{C}, \mathbb{R})$ *satisfies* $|\nabla \phi(x)| \leq C(1 + |x|_{\mathbf{C}})^m$ *for all* $x \in \mathbf{C}$ *and some* $C > 0$ *and* $m \geq 0$.

By the Girsanov theorem there exists a probability measure, called risk-neutral probability, for which

$$\overline{W}_t = \int_0^t \theta(\tau, S_{\tau+\cdot}) \, d\tau + W_t, \qquad t \in [0, T],$$

is a Wiener process. Then

$$dS_t = \rho\, S_t\, dt + \sigma(t, S_{t+\cdot})\, S_t\, d\overline{W}_t, \qquad dV_t = \rho V_t\, dt + \pi_t \sigma(t, S_{t+\cdot})\, d\overline{W}_t.$$

The existence of a hedging strategy can be established as follows: using the results of section 4 we first find a solution to the following forward-backward stochastic differential system:

$$\begin{cases} dS_t = \rho\, S_t\, dt + \sigma\,(t, S_{t+\cdot})\, S_t\, d\overline{W}_t, \quad t \in [0,T], \\ S_{0+\cdot} = s, \\ dV_t = \rho V_t\, dt + Z_t\, d\overline{W}_t, \\ V_T = \phi(S_{T+\cdot}). \end{cases} \tag{7.7}$$

Next, recalling (7.5), we note that the required hedging strategy can be recovered from the process Z setting $\pi_t = Z_t/\sigma(t, S_{t+\cdot})$.

However, a better characterization of the hedging strategy and the fair price of the claim can be obtained. We first consider, for arbitrary $t \in [0,T]$ and $s \in \mathbf{C}$, the following forward-backward system, which generalizes (7.7):

$$\begin{cases} dS_\tau^{t,s} = \rho\, S_\tau^{t,s}\, dt + \sigma\,(\tau, S_{\tau+\cdot}^{t,s})\, S_\tau^{t,s}\, d\overline{W}_\tau, \quad \tau \in [t,T], \\ S_{t+\cdot}^{t,s} = s, \\ dV_\tau^{t,s} = \rho V_\tau^{t,s}\, d\tau + Z_\tau^{t,s}\, d\overline{W}_\tau, \\ V_T^{t,s} = \phi(S_{T+\cdot}^{t,s}), \end{cases}$$

with unknown triple $(S_\tau^{t,s}, V_\tau^{t,s}, Z_\tau^{t,s})$. Setting $X_\tau^{t,s} = S_{\tau+\cdot}^{t,s}$, then X is a Markov process in $\mathbf{C}$ with generator $\mathcal{L}$. We finally define

$$v(t,s) = V_t^{t,s}, \qquad t \in [0,T],\ s \in \mathbf{C}.$$

It follows from Corollary 4.5 that $Z_\tau^{t,s} = \nabla_0 v(\tau, X_\tau^{t,x})\ \sigma(\tau, X_\tau^{t,x})$. We conclude that the fair price and the hedging strategy are uniquely determined as

$$V_0 = v(0,s), \qquad \pi_t = \frac{Z_t^{0,s}}{\sigma(t, X_t^{0,s})} = \nabla_0 v(t, X_t^{0,s}) = \nabla_0 v(t, S_{t+\cdot}).$$

Moreover, see Theorem 5.2, $v(t,s)$ is characterized as the unique mild solution of the equation

$$\begin{cases} \dfrac{\partial v(t,x)}{\partial t} + \mathcal{L}v(t,x) = \rho\, v(t,x), \\ u(T,x) = \phi(x), \qquad t \in [0,T],\ x \in \mathbf{C}, \end{cases} \tag{7.8}$$

which can be considered as a generalization of the Black–Scholes equation to the present setting.

REFERENCES

[1] M. Arriojas, Y. Hu, S.-E. A. Mohammed, and G. Pap, *A delayed Black and Scholes formula*, Stoch. Anal. Appl., 25 (2007), pp. 471–492.

[2] J. P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser Boston, Boston, 1990.

[3] T. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, Oxford, 1998.

[4] M.-H. Chang, *Stochastic Control of Hereditary Systems and Applications*, Stoch. Mod. Appl. Probab. 59, Springer, New York, 2008.

[5] M.-H. Chang and R. K. Youree, *Infinite-dimensional Black-Scholes equation with hereditary structure*, Appl. Math. Optim., 56 (2007), pp. 395–424.
[6] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia Math. Appl. 44, Cambridge University Press, Cambridge, UK, 1992.
[7] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite-Dimensional Systems*, London Math. Soc. Lecture Note Ser. 229, Cambridge University Press, Cambridge, UK, 1996.
[8] N. El Karoui and L. Mazliak, eds., *Backward Stochastic Differential Equations*, Pitman Res. Notes Math. Ser. 364, Longman, Harlow, UK, 1997.
[9] N. El Karoui, S. Peng, and M. C. Quenez, *Backward stochastic differential equations in finance*, Math. Finance, 7 (1997), pp. 1–71.
[10] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Appl. Math. (N.Y.) 25, Springer-Verlag, New York, 1993.
[11] M. Fuhrman and G. Tessitore, *Nonlinear Kolmogorov equations in infinite dimensional spaces: The backward stochastic differential equations approach and applications to optimal control*, Ann. Probab., 30 (2002), pp. 1397–1465.
[12] M. Fuhrman and G. Tessitore, *Infinite horizon backward stochastic differential equations and elliptic equations in Hilbert spaces*, Ann. Probab., 32 (2004), pp. 607–660.
[13] M. Fuhrman and G. Tessitore, *Generalized directional gradients, backward stochastic differential equations and mild solutions of semilinear parabolic equations*, Appl. Math. Optim., 51 (2005), pp. 279–332.
[14] J. Hale, *Theory of Functional Differential Equations*, Appl. Math. Sci. 3, Springer-Verlag, New York, 1971.
[15] Y. Hu, *A generalized Haussmann's formula*, Stochastic Anal. Appl., 11 (1993), pp. 49–60.
[16] S. Kusuoka and D. Stroock, *Applications of the Malliavin calculus.* I, in Stochastic Analysis (Katata/Kyoto, 1982), North–Holland Math. Library 32, North–Holland, Amsterdam, 1984, pp. 271–306.
[17] F. Masiero, *Stochastic optimal control problems and parabolic equations in Banach spaces*, SIAM J. Control Optim., 47 (2008), pp. 251–300.
[18] S. E. A. Mohammed, *Stochastic Functional Differential Equations*, Res. Notes Math. 99, Pitman, Boston, 1984.
[19] S. E. A. Mohammed, *Stochastic Differential Systems with Memory: Theory, Examples and Applications*, in Stochastic Analysis and Related Topics. VI. Proceedings of the 6th Oslo-Silivri Workshop (Geilo, 1996), Progr. Probab. 42, Birkhäuser Boston, Boston, 1998.
[20] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling*, Appl. Math. (N.Y.) 36, Springer Verlag, Berlin, 1997.
[21] D. Nualart, *The Malliavin Calculus and Related Topics*, Probab. Appl., Springer-Verlag (N.Y.), New York, 1995.
[22] D. Nualart and É. Pardoux, *Stochastic calculus with anticipative integrands*, Probab. Theory Related Fields, 78 (1988), pp. 535–581.
[23] É. Pardoux, *BSDE's, weak convergence and homogeneization of semilinear PDE's*, in Nonlinear Analysis, Differential Equations and Control, F. H. Clarke and R. J. Stern, eds., Kluwer Academic Publishers, Dordrecht, 1999, pp. 503–549.
[24] É. Pardoux and S. G. Peng, *Adapted solution of a backward stochastic differential equation*, Systems Control Lett., 14 (1990), pp. 55–61.
[25] É. Pardoux and S. Peng, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, in Stochastic Partial Differential Equations and Their Applications, Lecture Notes in Control and Inform. Sci. 176, B. L. Rozowskii and R. B. Sowers, eds., Springer, Berlin, 1992, pp. 200–217.
[26] F. Russo and P. Vallois, *Forward, backward and symmetric stochastic integration*, Probab. Theory Related Fields, 97 (1993), pp. 403–421.
[27] F. Russo and P. Vallois, *The generalized covariation process and Itô formula*, Stochastic Process. Appl., 59 (1995), pp. 81–104.
[28] F. Russo and P. Vallois, *Itô formula for C^1-functions of semimartingales*, Probab. Theory Related Fields, 104 (1996), pp. 27–41.
[29] F. Russo and P. Vallois, *Stochastic calculus with respect to continuous finite quadratic variation processes*, Stochastics Stochastics Rep., 70 (2000), pp. 1–40.
[30] P. Willmott, J. Dewynne, and S. Howison, *Option Pricing*, Oxford Financial Press, Oxford, 1993.