SIAM J. CONTROL OPTIM.
Vol. 62, No. 3, pp. 1490–1520

OPTIMAL CONTROL OF STOCHASTIC DELAY DIFFERENTIAL EQUATIONS AND APPLICATIONS TO PATH-DEPENDENT FINANCIAL AND ECONOMIC MODELS*

FILIPPO DE FEO†, SALVATORE FEDERICO‡, AND ANDRZEJ ŚWIĘCH§

Abstract. In this manuscript we consider a class of optimal control problems of stochastic differential delay equations. First, we rewrite the problem in a suitable infinite-dimensional Hilbert space. Then, using the dynamic programming approach, we characterize the value function of the problem as the unique viscosity solution of the associated infinite-dimensional Hamilton-Jacobi-Bellman equation. Finally, we prove a $C^{1,\alpha}$-partial regularity of the value function. We apply these results to path dependent financial and economic problems (Merton-like portfolio problem and optimal advertising).

Key words. stochastic optimal control, viscosity solutions in Hilbert spaces, partial regularity, stochastic delay equations, path-dependent equations, Merton problem

MSC codes. 49L25, 93E20, 49K45, 60H15, 49L20, 35R15, 49L12

DOI. 10.1137/23M1553960

1. Introduction.

In this paper, we consider a class of stochastic optimal control problems with infinite horizon with delays in the state equation. Precisely, the state equation is a stochastic delay differential equation (SDDE) in $\mathbb{R}^n$ of the form

$$dy(t)=b_0\left(y(t),\int_{-d}^{0}a_1(\xi)y(t+\xi)\,d\xi,u(t)\right)dt+\sigma_0\left(y(t),\int_{-d}^{0}a_2(\xi)y(t+\xi)\,d\xi,\,u(t)\right)dW(t),$$

with initial data $y(0)=x_0$ and $y(\xi)=x_1(\xi)$ for $\xi\in[-d,0]$. Here, u is a control process ranging in a suitable set of admissible processes $\mathcal{U}$, and the goal is to minimize, for $u(\cdot)\in\mathcal{U}$, a functional of the form

$$J(x;u(\cdot))=\mathbb{E}\left[\int_0^\infty e^{-\rho t}l(y(t),u(t))dt\right],\quad x=(x_0,x_1).$$

Our goal is to employ the dynamic programming approach and characterize the value function V for the problem as the unique solution of the Hamilton-Jacobi-Bellman (HJB) equation in an appropriate sense, and prove its suitable regularity

*Received by the editors February 17, 2023; accepted for publication (in revised form) December 29, 2023; published electronically May 20, 2024.
https://doi.org/10.1137/23M1553960
Funding: The work of the first author was partially funded by the Italian Ministry of University and Research (MIUR), within PRIN project 2017FKHBA8 001 (The Time-Space Evolution of Economic Activities: Mathematical Models and Empirical Applications) and by the INdAM (Istituto Nazionale di Alta Matematica) - GNAMPA (Gruppo Nazionale per l'Analisi Matematica, la Probabilitá e loro Applicazioni) Project "Riduzione del modello in sistemi dinamici stocastici in dimensione infinita a due scale temporali." Salvatore Federico was partially funded by the Italian Ministry of University and Research (MIUR), within PRIN project 2017FKHBA8 001 (The Time-Space Evolution of Economic Activities: Mathematical Models and Empirical Applications) and is affiliated with INdAM-GNAMPA.
†Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133, Milano, Italy (filippo.defeo@polimi.it).
‡Department of Mathematics, University of Bologna, Piazza di Porta S. Donato, 5, 40126 Bologna, Italy (s.federico@unibo.it).
§School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 USA (swiech@math.gatech.edu).

properties, having in mind construction of optimal feedback controls. As it is well known, the main difficulty for delay problems is in the lack of Markovianity, which prevents a direct application of the dynamic programming method. In fact, even though the dynamic programming principle can be proved (see [61]), it is not immediately clear how to derive an HJB equation, which is, in general, intrinsically infinite-dimensional, as the initial datum x_1 is a function. If the delay kernels a_1, a_2 have a special structure, the HJB equation can be reduced to a finite-dimensional one (see, e.g., [62]). However, this is not the case in general and other approaches are needed to tackle the problem. A possible method consists of developing and using an Itô's formula based on differential calculus for equations with delay, leading to a theory of the so-called path-dependent PDEs (see, e.g., [6, 7, 21, 31, 32, 33, 76, 77] and the references therein). Another approach, which is the one we follow here, is to lift the state equation to an infinite-dimensional Banach or Hilbert space (depending on the regularity of the data), in order to regain Markovianity. This is done at a cost of moving to infinite-dimension.[1] To be more precise, the state equation and the cost functional are then rewritten in a suitable infinite-dimensional space as

$$dY(t) = [AY(t) + b(Y(t), u(t))]dt + \sigma(Y(t), u(t))\, dW(t), \qquad Y(0) = x = (x_0, x_1),$$

and

$$J(x; u(\cdot)) = \mathbb{E}\left[\int_0^\infty e^{-\rho t} L(Y(t), u(t))dt\right],$$

with suitable A, b, σ, L. This is explained in section 3. Once this is done, one may try to employ techniques of stochastic optimal control in infinite-dimensional spaces and study the associated infinite-dimensional HJB equation. We approach this infinite-dimensional HJB equation by means of viscosity solutions, whose theory is developed better in Hilbert spaces (see, e.g., [23, 63] for first order equations and deterministic problems; [36] for second order equations and stochastic problems). Thus we take the data allowing one to rewrite the state equation in the Hilbert space $X := \mathbb{R}^n \times L^2([-d, 0]; \mathbb{R}^n)$. The HJB equation on X has the form

$$\rho v(x) - \langle Ax, Dv(x)\rangle + \tilde{H}(x, D_{x_0}v(x), D^2_{x_0}v(x)) = 0, \quad x = (x_0, x_1(\cdot)) \in X,$$

where the Hamiltonian $\tilde{H}$ only involves the derivatives with respect to the finite-dimensional component x_0 (see section 5).

PDEs in Hilbert spaces have been studied following at least four different approaches, based on various notions of solutions. We recall them here, together with their variants, and then describe how these approaches were applied to stochastic optimal control problems coming from delay problems.

(i) Classical solutions (e.g., see [36, Chapter 2]). Classical solutions are rare as the regularity required for this notion of solution is typically hard to obtain.

(ii) Viscosity solutions (see, e.g., [63, Chapter 6] and [36, Chapter 3] for a general overview, respectively, in the deterministic and the stochastic case). This approach is particularly suitable to treat first and second order fully nonlinear degenerate HJB equations.

(iii) Mild solutions in spaces of continuous functions via fixed point methods. This method was initiated in the deterministic case in [4] and then developed by many other authors (see [36, Chapter 4] for an overview).

[1]For the procedure of rewriting deterministic delay differential equations, we refer the reader to [8, Part II, Chapter 4]. For the stochastic case, one may consult [20, 24, 53, 54, 46] for the Hilbert space case and [69, 70, 47] for the Banach space case. A "mixed" approach is employed in [40].

(iv) Mild solutions in (L^2, μ) spaces, where μ is an invariant measure of the uncontrolled system (see, e.g., [36, Chapter 5] and [52]).

(v) Mild solutions in spaces of continuous functions via backward stochastic differential equations (BSDEs) methods (see, e.g. [36, Chapter 6] for a complete picture and also the original works [49] and [50]). This method relies on an extension to infinite dimension of the celebrated BSDEs approach to semilinear HJB equation initiated by [73] and then developed by many other authors.

(vi) Regular solutions via convex regularization procedures. This approach was introduced mostly to study the parabolic case and requires a strong regularity of the data (e.g., the initial condition of the parabolic HJB must be convex and C^2); see [4]. It was then developed further only for first order equations (e.g., see [36, Chapter 4, Bibliographic Notes]).

(vii) Explicit (classical) solutions. This method is applied only in special cases, typically for linear-quadratic and linear-power problems; still it may provide interesting applications to economic theory. For an overview, see [36, Chapter 4, Section 10].

Regarding applications of these various approaches to stochastic optimal control problems arising from delay equations, the general "state of the art" is the following.

- The approach based on mild solutions was successfully employed using the three methods (iii)–(v) to treat such problems. However, the main drawback of these methods is that they work only for semilinear HJB equations (i.e., when there is no control in the diffusion coefficient) and they also require many technical assumptions on the data. Smoothing properties of the transition semigroup associated to the linear part of the equation are needed for the fixed point approach.[2] The linearity of the state equation, the so-called *structure condition,* that is the requirement that the range of the control operator is contained in the range of the noise, as well as constraints on the data guaranteeing the existence of an invariant measure, are needed for the (L^2, μ)[3] approach. A special structure condition is also needed for the BSDEs approach.[4] In all cases there are also some limits on the generality of the coefficients of the delay state equation, leaving out some interesting cases arising in applications; for instance, portfolio problems, where the control naturally acts in the diffusion, leading to a fully nonlinear HJB equation.
- Regarding the approach based on (vi), such methods were employed in [37, 38, 39] to study a deterministic optimal investment problem with vintage capital.[5]
- The approach based on (vii) was employed to study deterministic and stochastic problems with differential delay equations, for example in [34, 2, 3, 11, 12, 13].
- The theory of viscosity solutions (ii) was first applied to deterministic control problems. The notion of the so-called B-continuous viscosity solutions in infinite dimension from [23] is employed in [42, 43, 45], where a class optimal control problem with delays and state constraints was considered: the value

[2]This property intrinsically does not hold for infinite-dimensional systems coming from delay equations; the problem can be circumvented by looking at *partial* smoothing properties, see [55, 56, 57, 66].

[3]See [53].

[4]See, e.g., [48] and also [36, Chapter 6, section 6.6].

[5]Even if, strictly speaking, these are not optimal control problems with delay, their infinite-dimensional formulation shares the same features with the latter, as the unbounded operator is the first derivative.

function was proved to be a viscosity solution of the HJB equation and a partial C^1-regularity of the value function was obtained. This regularity result allowed one to construct optimal feedback controls. Another concept of viscosity solution was used in [17], where the value function was characterized as the unique solution in that sense. There have also been some results in the stochastic case. We refer to [85, 86, 87], where approaches using appropriately defined viscosity solutions in spaces of right-continuous functions and continuous functions were studied. In [79] the authors prove existence, uniqueness, and partial regularity of viscosity solutions to Kolmogorov equations[6] related to stochastic delay equations. Our paper can be considered as an extension of this paper to the case of fully nonlinear HJB equations.

We now describe in detail the results of this paper and compare them with the related literature. First, after rewriting the problem as an infinite-dimensional control problem (see Proposition 3.1), in order to apply the theory of viscosity solutions in Hilbert spaces [36], we rewrite it further by introducing a maximal dissipative operator $\tilde{A}$ in the state equation (see Proposition 3.3) and introduce an operator B satisfying the so-called weak B-condition (see Proposition 3.5). Then, we prove that the data of the problem satisfy some regularity conditions with respect to the norm induced by the operator $B^{1/2}$ (see Lemma 4.1). This enables us to characterize the value function of the problem as the unique viscosity solution of the infinite-dimensional HJB equation (our first main result, Theorem 5.4). To the best of our knowledge, this is the first existence and uniqueness result for fully nonlinear HJB equations in Hilbert spaces related to a general class of stochastic optimal control problems with delays involving controls in the diffusion coefficient.

Unfortunately, due to the lack of good regularity results, in general the notion of viscosity solution does not provide tools to construct optimal controls. In particular, verification theorems in the context of viscosity solutions are difficult to implement, especially in the stochastic case.[7] For these reasons, obtaining regularity results (even only partial) is very important. In this respect, we have to mention that some of the aforementioned papers go exactly in this direction. In [42], in a purely deterministic framework, using the convexity of the value function, the authors are able to prove a (finite-dimensional) C^1-regularity of the value function with respect to the x_0-component. This result is the basis for construction of optimal feedback controls via an ad hoc verification theorem using viscosity property of the value function, a goal obtained in [43]. In our paper, we have to deal with the stochastic framework and we do not impose conditions ensuring the convexity of the value function, so the techniques of [42] to prove the desired partial regularity are not applicable. Instead, we rely on smoothing properties of the noise to obtain a similar result. This approach is inspired by the arguments of [79] for Kolmogorov equations and a finite-dimensional reduction procedure which first appeared in [64]. We prove the $C^{1,\alpha}$ partial regularity of $V(x_0, x_1)$ with respect to the x_0-component (see Theorem 6.5) under rather general assumptions (we only require some standard Lipschitz conditions on the data, uniform in the control variable and uniform ellipticity condition of the diffusion on

[6]Control problems are not considered there.

[7]To have an idea about how much the problem is tricky and delicate, the reader may look at, in finite dimension, [84, Chapter 5], [58], and [59]; for deterministic optimal control problems in finite dimension, see also [5]. For the infinite-dimensional case, the situation is clearly even more technical: in the deterministic setting some formulations can be found in [63, Chapter 6], [16], and [35]; in the stochastic case we mention a recent paper [80]. Optimal feedback controls are constructed for a class of problems with bounded evolution in [67].

bounded sets). Fixing the infinite-dimensional component x_1, we reduce the infinite-dimensional HJB equation (5.5) to a uniformly elliptic second order finite-dimensional PDE. Then standard elliptic regularity results give the required $C^{1,\alpha}$-regularity. However, our method is different and more efficient with respect to the one used in [79]. In [79], an approximating procedure is employed: infinite-dimensional SDEs with smoothed coefficients and Yosida approximations of the unbounded operator are considered, with the corresponding value functions for a smoothed out payoff function. Then, it is proved there that the finite-dimensional sections of the approximating value functions are viscosity solutions of certain linear finite-dimensional parabolic equations for which $C^{1,\alpha}$-estimates hold and the result follows by passing to the limit. In our paper we simplify considerably the argument, avoiding this complex approximating procedure by using deeper results from the theory of L^p-viscosity solutions [15, 81, 82]. And our method works for fully nonlinear equations and hence for control problems with controls in the drift and the diffusion coefficients.

The $C^{1,\alpha}$ partial regularity result is interesting on its own and seems to be the first one for fully nonlinear second order HJB equations with unbounded operators in Hilbert spaces. From the point of view of the control problem, it is particularly relevant as, under some additional natural assumptions, it allows one to define a possible optimal feedback control. However, to prove that this control is actually optimal is not an easy task. We will address this in a future publication. Concerning regularity results for HJB equations related to delay problems in the existing literature, we observe that in [65, 48] the full Gateaux differentiability of the solution is obtained by means of an approach via BSDEs, assuming the differentiability of the data, and applied to problems with delays (see also [36, Chapter 6] in Hilbert spaces). For mild solutions in (L^2, μ) spaces, in [36, Chapter 5] a first-order regularity of the solution in a Sobolev sense is proved (applications to delay problems are provided in [36, subsection 5.6]). In [66], the Gateaux differentiability of the solution is obtained by means of a partial smoothing of the semigroup. We also mention that in the case of bounded HJB equations (unrelated to delay problems), some $C^{1,1}$ regularity results for viscosity solutions were obtained in [9, 10, 51, 64, 67] for first and second order HJB equations in Hilbert spaces and spaces of probability measures.

We also provide two applications of our results. First, we consider a Merton-type portfolio optimization problem with path-dependency features in the dynamics of the risky asset.[8] Merton's problem with path-dependency features in the stock price was studied in [72] (see also the references therein), where an exponential structure of the delay kernel is assumed; the problem is approached by employing the methods of [62] to reduce the infinite-dimensional HJB equation to a finite-dimensional one. Other results in this direction are in [11, 12, 13], where the authors consider the life-cycle optimal portfolio choice problem faced by an agent receiving labor income whose dynamic has delays, while the dynamics of the risky assets are Markovian. As a second application, we illustrate the stochastic optimal advertising problem with delays, studied in the literature in [53].

This paper is organized as follows. In section 2 we introduce the problem and state the main assumptions. In section 3 we rewrite the problem in an infinite-dimensional setting. In section 4 we prove some preliminary estimates for solutions of the state equation and the value function. In section 5 we introduce the notion of viscosity solution of the HJB equation and state a theorem about the existence and uniqueness

[8]See [68] for the original problem formulation and [60] for a complete exposition with a quantitative analysis on why path-dependent models are important in financial modeling.

of viscosity solutions, and characterize the value function as the unique viscosity solution. In section 6 we prove, under additional assumptions, a partial $C^{1,\alpha}$ regularity result for the value function. Section 7 is devoted to the two applications.

2. The optimal control problem: Setup and assumptions. We denote by $M^{m\times n}$ the space of real valued $m\times n$-matrices and we denote by $|\cdot|$ the Euclidean norm in $\mathbb{R}^n$ as well as the norm of elements of $M^{m\times n}$ regarded as linear operators from $\mathbb{R}^m$ to $\mathbb{R}^n$. We will write $x\cdot y$ for the inner product in $\mathbb{R}^n$. Given $d>0$, we consider the standard Lebesgue space $L^2 := L^2([-d,0];\mathbb{R}^n)$ of square integrable functions from $[-d,0]$ to $\mathbb{R}^n$. We denote by $\langle\cdot,\cdot\rangle_{L^2}$ the inner product in L^2 and by $|\cdot|_{L^2}$ the norm. We also consider the standard Sobolev space $W^{1,2} := W^{1,2}([-d,0];\mathbb{R}^n)$ of functions in L^2 admitting weak derivative in L^2, endowed with the inner product $\langle f,g\rangle_{W^{1,2}} := \langle f,g\rangle_{L^2} + \langle f',g'\rangle_{L^2}$ and norm $|f|_{W^{1,2}} := (|f|^2_{L^2} + |f'|^2_{L^2})^{\frac{1}{2}}$, which render it a Hilbert space. It is well known that the space $W^{1,2}$ can be identified with the space of absolutely continuous functions from $[-d,0]$ to $\mathbb{R}^n$.

Let $\tau = (\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P},W)$ be a reference probability space, that is $(\Omega,\mathcal{F},\mathbb{P})$ is a complete probability space, $W=(W(t))_{t\geq 0}$ is a standard $\mathbb{R}^q$-valued Wiener process, $W(0)=0$, and $(\mathcal{F}_t)_{t\geq 0}$ is the augmented filtration generated by W. We consider the following controlled SDDE:

(2.1)
$$\begin{cases} dy(t) = b_0\left(y(t), \int_{-d}^0 a_1(\xi)y(t+\xi)\,d\xi, u(t)\right)dt + \sigma_0\left(y(t), \int_{-d}^0 a_2(\xi)y(t+\xi)\,d\xi,\, u(t)\right)dW(t),\\ y(0)=x_0, \quad y(\xi)=x_1(\xi) \quad \forall \xi\in[-d,0), \end{cases}$$

where $d>0$ is the maximum delay and
(i) $x_0\in\mathbb{R}^n$, $x_1\in L^2([-d,0];\mathbb{R}^n)$ are the initial conditions;
(ii) $b_0\colon \mathbb{R}^n\times\mathbb{R}^h\times U\to\mathbb{R}^n$, $\sigma_0\colon\mathbb{R}^n\times\mathbb{R}^h\times U\to M^{n\times q}$;
(iii) $a_i:[-d,0]\to M^{h\times n}$ for $i=1,2$ and if a_i^j is the jth row of $a_i(\cdot)$ for $j=1,\dots,h$, then $a_i^j\in W^{1,2}$ and $a_i^j(-d)=0$.

The precise assumptions on b_0,σ_0 will be given below.

Remark 2.1. In (2.1), setting the initial value, we have separated the role of the present $y(0)$ from the role of past $y(\xi)$ for $\xi\in[-d,0)$. This is due to the fact that we want to embed the SDDE into an infinite-dimensional *product space* framework which in our case is the Hilbert space $\mathbb{R}^n\times L^2$.[9] The choice of working within a product space framework allows more flexibility in the study of the HJB equation, as it naturally separates, in the infinite-dimensional formulation, the "present" component $y(0)$ from the "past" component $y(\xi)_{\xi\in[-d,0)}$ and allows one to study the regularity of its solution only with respect to the former one, which is finite dimensional. Other possible kinds of infinite-dimensional formulations in the spirit of the product space approach have been introduced in the literature. We refer to [48, 47]. Their advantage, compared to our approach, is in treating functionals with pointwise delay in the SDDE (cf. Remark 2.2 below). The variation of the "present component" in these papers is connected to the notion of the vertical derivative in the so-called *functional Itô's calculus* (see, e.g., [18, 19]).

[9] Alternatively, one could, for instance, embed the SDDE in the infinite-dimensional space Banach space $C([-d,0];\mathbb{R}^n)$ (then it is natural to take the initial datum $y(\xi)$, $\xi\in[-d,0]$ in $C([-d,0];\mathbb{R}^n)$; for a comparison between these two approaches, we refer to [8, Part II, Chapter 4] in the deterministic setting. However, note that $C([-d,0];\mathbb{R}^n)$ is embedded in $\mathbb{R}^n\times L^2$; so, if $x_1\in C([-d,0];\mathbb{R}^n)$, nothing prevents us to take $x_0 = x_1(0)$, so to set $y(\xi)=x_1(\xi)$ for every $\xi\in[-d,0]$.

Remark 2.2. As in [42, 79], we assume that the delay kernels a_i^j are regular functions: $a_i^j \in W^{1,2}$. Hence, we cannot treat the general case of a_i^j being a measure (e.g., the pointwise delay given by Dirac delta δ_{-d} are not covered by our approach). However, we point out that, even though a more general case of a measure can be handled by the approaches via mild solutions or BSDE's (see, e.g., [36]), here we are able to cover the case of control in the diffusion term, leading to a fully nonlinear HJB equation, and we also avoid assuming differentiability of the coefficients.

Remark 2.3. As in [42, 79], the condition $a_i(-d)=0$, for $i=1,2$, is technical (see Remark 4.2). Yet, it is not too restrictive in applications: indeed, one can always take a slightly larger $d_\varepsilon := d+\varepsilon$, and extend a_i to an absolutely continuous $M^{h\times n}$-valued function a_i^ε over $[-d_\varepsilon, 0]$ in such a way that $a_i(-d_\varepsilon)=0$. Clearly, using the above procedure, the structure of (2.1) changes and so does its solution. However, some convergences can be guaranteed. We refer to Remark 3.2 for more details.

We consider the following infinite horizon optimal control problem. Given $x=(x_0,x_1)\in\mathbb{R}^n\times L^2$, we define a cost functional of the form

$$J(x;u(\cdot)) = \mathbb{E}\left[\int_0^\infty e^{-\rho t} l(y^{x,u}(t),u(t))dt\right], \tag{2.2}$$

where $\rho>0$ is the discount factor, $l\colon \mathbb{R}^n\times U\to\mathbb{R}$ is the running cost and $U\subset\mathbb{R}^p$. For every reference probability space τ we consider the set of control processes

$$\mathcal{U}_\tau = \{u(\cdot):\Omega\times[0,+\infty)\to U : u(\cdot)\,\text{is}\,(\mathcal{F}_t)\text{-progressively measurable}\}.$$

We define

$$\mathcal{U} = \bigcup_\tau \mathcal{U}_\tau,$$

where the union is taken over all reference probability spaces τ. The goal is to minimize $J(x,u(\cdot))$ over all $u(\cdot)\in\mathcal{U}$. This is a standard setup of a stochastic optimal control problem (see [84, 36]) used to apply the dynamic programming approach. We remark (see, e.g., [36, section 2.3.2]) that

$$\inf_{u(\cdot)\in\mathcal{U}} J(x,u(\cdot)) = \inf_{u(\cdot)\in\mathcal{U}_\tau} J(x,u(\cdot))$$

for every reference probability space τ so the optimal control problem is, in fact, independent of the choice of a reference probability space. We will assume the following conditions.

Assumption 2.4. The functions b_0,σ_0 are continuous and such that there exist constants $L,C>0$ such that, for every $x,x_1,x_2\in\mathbb{R}^n$, $z,z_1,z_2\in\mathbb{R}^h$ and every $u\in U$,

$$\begin{aligned}
&|b_0(x,z,u)| \le C(1+|x|+|z|),\\
&|\sigma_0(x,z,u)| \le C(1+|x|+|z|),\\
&|b_0(x_2,z_2,u)-b_0(x_1,z_1,u)| \le L(|x_2-x_1|+|z_2-z_1|),\\
&|\sigma_0(x_2,z_2,u)-\sigma_0(x_1,z_1,u)| \le L(|x_2-x_1|+|z_2-z_1|).
\end{aligned}$$

Under Assumption 2.4, by [78, Theorem IX.2.1], for each initial datum $x:=(x_0,x_1)\in\mathbb{R}^n\times L^2([-d,0];\mathbb{R}^n)$ and each control $u(\cdot)\in\mathcal{U}$, there exists a unique (up to indistinguishability) strong solution to (2.1) and this solution admits a version with continuous paths that we denote by $y^{x;u}$. The proof that the assumptions of [78, Theorem IX.2.1] are satisfied can be found in [46, Proposition 2.5].

Assumption 2.5. $l\colon \mathbb{R}^n \times U \to \mathbb{R}$ is continuous and is such that the following hold.

(i) There exist constants $K, m > 0$, such that

$$|l(z,u)| \le K(1+|z|^m) \quad \forall y \in \mathbb{R}^n,\ \forall u \in U. \tag{2.3}$$

(ii) There exists a local modulus of continuity for l, uniform in $u \in U$, i.e., for each $R > 0$, there exists a nondecreasing function $\omega_R : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{r\to 0^+} \omega_R(r) = 0$ and

$$|l(z,u) - l(z',u)| \le \omega_R(|z-z'|) \tag{2.4}$$

for every $z, z' \in \mathbb{R}^n$ such that $|z|, |z'| \le R$ and every $u \in U$.

We will later show that, suitably reformulating the state equation in an infinite-dimensional framework, the cost functional is well defined and finite for a sufficiently large discount factor $\rho > 0$ (see Assumption 4.4).

Throughout this paper we will write $C > 0, \omega, \omega_R$ to indicate, respectively, a constant, a modulus continuity, and a local modulus of continuity, which may change from place to place if the precise dependence on other data is not important.

3. The equivalent infinite dimensional Markovian representation. The optimal control problem at hand is not Markovian due to the delay. In order to regain Markovianity and approach the problem by dynamic programming, following a well-known procedure (see [8, Part II, Chapter 4] for deterministic delay equations and [20, 26, 42] for the stochastic case), we reformulate the state equation by lifting it to an infinite-dimensional space.

We define $X := \mathbb{R}^n \times L^2$. An element $x \in X$ is a couple $x = (x_0, x_1)$, where $x_0 \in \mathbb{R}^n$, $x_1 \in L^2$; sometimes, we will write $x = \left[\begin{smallmatrix} x_0 \\ x_1 \end{smallmatrix}\right]$. The space X is a Hilbert space when endowed with the inner product

$$\langle x, z\rangle_X := x_0 \cdot z_0 + \langle x_1, z_1\rangle_{L^2} = x_0 z_0 + \int_{-d}^0 x_1(\xi)\cdot z_1(\xi)\, d\xi, \quad x = (x_0,x_1),\ z = (z_0,z_1) \in X.$$

The induced norm, denoted by $|\cdot|_X$, is then

$$|x|_X = \left(|x_0|^2 + \int_{-d}^0 |x_1(\xi)|_{L^2}^2\, d\xi\right)^{1/2}, \quad x = (x_0,x_1) \in X.$$

For $R > 0$, we denote

$$B_R := \{x \in X : |x|_X < R\}, \quad B_R^0 := \{x_0 \in \mathbb{R}^n : |x_0| < R\},$$
$$B_R^1 := \{x_1 \in L^2[-d,0] : |x_1|_{L^2} < R\}$$

to be the open balls of radius R in X, $\mathbb{R}^n$, and L^2, respectively. We denote by $\mathcal{L}(X)$ the space of bounded linear operators from X to X, endowed with the operator norm $|L|_{\mathcal{L}(X)} = \sup_{|x|_X=1} |Lx|_X$. An operator $L \in \mathcal{L}(X)$ can be seen as

$$Lx = \begin{bmatrix} L_{00} & L_{01} \\ L_{10} & L_{11} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}, \quad x = (x_0,x_1) \in X,$$

where $L_{00}\colon \mathbb{R}^n \to \mathbb{R}^n$, $L_{01}\colon L^2 \to \mathbb{R}^n$, $L_{10}\colon \mathbb{R}^n \to L^2$, $L_{11}\colon L^2 \to L^2$ are bounded linear operators. Moreover, given two separable Hilbert spaces $(H, \langle\cdot,\cdot\rangle_H), (K, \langle\cdot,\cdot\rangle_K)$, we denote by $\mathcal{L}_1(H,K)$ the space of trace-class operators endowed with the norm

$$|L|_{\mathcal{L}_1(H,K)} = \inf\left\{\sum_{i\in\mathbb{N}} |a_i|_H |b_i|_K : Lx = \sum_{i\in\mathbb{N}} b_i \langle a_i, x\rangle_H, a_i \in H, b_i \in K\ \forall i \in \mathbb{N}\right\}.$$

We also denote by $\mathcal{L}_2(H,K)$ the space of Hilbert–Schmidt operators from H to K endowed with the norm $|L|_{\mathcal{L}_2(H,K)} = (\mathrm{Tr}(L^*L))^{1/2}$. When $H = K$ we simply write $\mathcal{L}_1(H)$, $\mathcal{L}_2(H)$. We denote by $S(H)$ the space of self-adjoint operators in $\mathcal{L}(H)$. If $Y, Z \in S(H)$, we write $Y \le Z$ if $\langle Yx, x\rangle \le \langle Zx, x\rangle$ for every $x \in H$.

We define the operator $A : D(A) \subset X \to X$:

$$Ax = \begin{bmatrix} 0 \\ {x_1}' \end{bmatrix}, \quad D(A) = \left\{x = (x_0, x_1) \in X :\ x_1 \in W^{1,2},\ x_1(0) = x_0\right\}.$$

By [8, Theorem 4.2], the operator A is the generator of a strongly continuous semigroup e^{tA} on X, whose explicit expression is

$$e^{At}x = \begin{bmatrix} x_0 \\ I_{[-d,0]}(t+\cdot)x_1(t+\cdot) + I_{[0,\infty)}(t+\cdot)x_0 \end{bmatrix}, \quad x = (x_0, x_1) \in X.$$

Notice that

$$|e^{tA}|_{\mathcal{L}(X)} \le (2(1+d))^{1/2} \quad \forall t \ge 0. \tag{3.1}$$

The adjoint of the operator A is the operator (see [42, Proposition 3.4])

$$A^*x = \begin{bmatrix} x_1(0) \\ -x_1' \end{bmatrix}, \quad D(A^*) = \left\{x = (x_0, x_1) \in X : x_1 \in W^{1,2},\ x_1(-d) = 0\right\}. \tag{3.2}$$

We now define $b\colon X \times U \to X$ (with a small abuse of notation for $b_0(x,u)$) by

$$b(x,u) = \begin{bmatrix} b_0(x,u) \\ 0 \end{bmatrix} = \begin{bmatrix} b_0\left(x_0, \int_{-d}^0 a_1(\xi)x_1(\xi)\,d\xi, u\right) \\ 0 \end{bmatrix}, \quad x = (x_0, x_1) \in X,\ u \in U,$$

and $\sigma\colon X \times U \to \mathcal{L}(\mathbb{R}^q, X)$ (again with a small abuse of notation for $\sigma_0(x,u)$) by

$$\sigma(x,u)w = \begin{bmatrix} \sigma_0(x,u)w \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_0\left(x_0, \int_{-d}^0 a_2(\xi)x_1(\xi)\,d\xi, u\right) w \\ 0 \end{bmatrix},$$
$$x = (x_0, x_1) \in X,\ u \in U,\ w \in \mathbb{R}^q.$$

Given $x \in X$ and a control process $u(\cdot) \in \mathcal{U}$, we consider the following infinite-dimensional stochastic differential equation:

$$\begin{cases} dY(t) = [AY(t) + b(Y(t), u(t))]dt + \sigma(Y(t), u(t))\,dW(t), \\ Y(0) = x. \end{cases} \tag{3.3}$$

It is well known (see, e.g., [26]) that there exists a unique mild solution to (3.3), that is an X-valued progressively measurable stochastic process Y satisfying

$$Y(t) = e^{At}x + \int_0^t e^{A(t-s)}b(Y(s), u(s))ds + \int_0^t e^{A(t-s)}\sigma(Y(s), u(s))dW(s) \quad \forall t \ge 0.$$

The infinite-dimensional stochastic differential equation (3.3) is linked to (2.1) by the following result; see [46, Theorem 3.4] (cf. also the original result in the linear case [20]).

PROPOSITION 3.1. *Let Assumption* 2.4 *hold. Given* $x \in X$ *and* $u(\cdot) \in \mathcal{U}$, *let* $y^{x,u}$ *be the unique strong solution to* (2.1), *and let* $Y^{x,u}$ *be the unique mild solution to* (3.3). *Then*

$$Y^{x,u}(t) = (y^{x,u}(t), y^{x,u}(t+\cdot)|_{[-d,0]}) \quad \forall t \geq 0.$$

Proposition 3.1 provides a Markovian reformulation of the optimal control problem in the Hilbert space X. Indeed, the functional (2.2) can be rewritten in X as

$$J(x;u(\cdot)) = \mathbb{E}\left[\int_0^\infty e^{-\rho t}[L(Y^{x,u}(t),u(t))]dt\right], \tag{3.4}$$

where $L: X \times U \to \mathbb{R}$ is defined by

$$L(x,u) := l(x_0,u), \quad x = (x_0,x_1) \in X, \ u \in U.$$

The value function V for this problem is defined by

$$V(x) := \inf_{u(\cdot)\in\mathcal{U}} J(x;u(\cdot)).$$

Remark 3.2. If we use the procedure described in Remark 2.3, we may embed all the "ε-problems" in the space $X_1 = \mathbb{R}^n \times L^2([-d-1,0];\mathbb{R}^n)$. Then, calling $Y^\varepsilon(t)$ the state of the infinite-dimensional representation for the "ε-problems", it is possible to prove that there exists $\lambda > 0$ such that

$$\mathbb{E}\left[|Y^\varepsilon(t) - Y(t)|^2_{X_1}\right] \leq C\varepsilon^{1/2}e^{\lambda t}|x|^2_{X_1} \xrightarrow{\epsilon\to 0} 0. \tag{3.5}$$

Then, if $J^\varepsilon, V^\varepsilon$ are the cost functional and the value function associated to the ε-problems, using (3.5) and taking ρ large enough, one can show that

$$\lim_{\epsilon\to 0} V^\varepsilon(x) = V(x).$$

On the other hand, the convergence of the derivatives of V^ε to the derivatives of V cannot be proved a priori. Hence, the convergence of a (candidate) optimal feedback control $u^*_\varepsilon(\cdot)$ constructed by optimal synthesis to a (candidate) optimal feedback control $u^*(\cdot)$ for the original problem is not clear.

3.1. Rewriting of the state equation using a maximal monotone operator. In this subsection we rewrite the state equation (3.3) using a maximal dissipative operator $\tilde{A}$. This is needed to be in the framework of the theory of viscosity solutions to the associated HJB equation (see [36, Chapter 3]). To do this we introduce the operator $\tilde{A}\colon D(\tilde{A}) \subset X \to X$ defined by $\tilde{A} := A - (x_0, 0)$, i.e.,

$$\tilde{A}x = \begin{bmatrix} -x_0 \\ x_1{}' \end{bmatrix}, \quad D(\tilde{A}) = D(A) = \left\{x = (x_0,x_1) \in X : x_1 \in W^{1,2}, \ x_1(0) = x_0\right\}. \tag{3.6}$$

PROPOSITION 3.3. *The operator* $\tilde{A}$ *defined in* (3.6) *is maximal dissipative.*

Proof. Let $x \in D(\tilde{A})$. Taking into account that $x_1(0) = x_0$, we have

$$\begin{aligned}\langle \tilde{A}x, x\rangle_X &= -|x_0|^2 + \int_{-d}^0 \langle x'_1(\xi), x_1(\xi)\rangle d\xi = -|x_0|^2 + \left[\frac{|x_1(\xi)|^2}{2}\right]_{-d}^0 \\ &= -\frac{|x_0|^2}{2} - |x_1(-d)|^2 \leq 0,\end{aligned}$$

so $\tilde{A}$ is dissipative.

In order to prove that $\tilde{A}$ is maximal dissipative we need to show that $\mathcal{R}(\lambda I - \tilde{A}) = X$ for some $\lambda > 0$. This means that we have to show that there exists some $\lambda > 0$ such that, for each $y = (y_0, y_1) \in X$, we can find $x = (x_0, x_1) \in D(\tilde{A})$ such that $\lambda x - \tilde{A}x = y$, i.e.,

$$\begin{cases} \lambda x_0 + x_0 = y_0, \\ \lambda x_1 - {x_1}' = y_1. \end{cases} \tag{3.7}$$

Indeed this is true for each $\lambda > 0$. Let $\lambda > 0$ and take an arbitrary $y = (y_0, y_1) \in X$. By the first equation in (3.7), we have

$$x_0 = \frac{1}{1+\lambda} y_0.$$

Now recall that if $x = (x_0, x_1) \in D(\tilde{A})$, we must have $x_1(0) = x_0 = y_0/(1+\lambda)$. The second equation of (3.7) is then the ODE

$$x_1'(\xi) = \lambda x_1(\xi) - y_1(\xi) \quad \forall \xi \in [-d, 0], \quad x_1(0) = y_0/(1+\lambda).$$

Its unique solution $x_1 \in W^{1,2}$ is given by

$$x_1(\xi) = \frac{1}{1+\lambda} e^{\lambda \xi} y_0 + \int_\xi^0 e^{-\lambda r} y_1(r) dr \quad \forall \xi \in [-d, 0].$$

Therefore, we found the unique solution $x = (x_0, x_1) \in D(\tilde{A})$ to the abstract equation $(\lambda I - \tilde{A})x = y$, showing the maximality of $\tilde{A}$. ∎

By Proposition 3.3, we can now rewrite (3.3) using the maximal dissipative operator $\tilde{A}$ as

$$\begin{cases} dY(t) = \left[\tilde{A}Y(t) + \tilde{b}(Y(t), u(t))\right] dt + \sigma(Y(t), u(t))\, dW(t), \\ Y(0) = x \in X, \end{cases} \tag{3.8}$$

where $\tilde{b}\colon X \times U \to X$ is defined by

$$\tilde{b}(x,u) = b(x,u) + \begin{bmatrix} x_0 \\ 0 \end{bmatrix} = \begin{bmatrix} b_0\left(x_0, \int_{-d}^0 a_1(\xi) x_1(\xi)\, d\xi, u\right) + x_0 \\ 0 \end{bmatrix}, \quad x = (x_0, x_1) \in X, \ u \in U.$$

Since the operator $\tilde{A}$ is the sum of A and a bounded operator, by [30, Corollary 1.7] we conclude that (3.8) is equivalent to (3.3) and they have the same (unique) mild solution given, in terms of $\tilde{A}$, by

$$Y(t) = e^{\tilde{A}t}x + \int_0^t e^{\tilde{A}(t-s)}\, \tilde{b}(Y(s), u(s)) ds + \int_0^t e^{\tilde{A}(t-s)} \sigma(Y(s), u(s)) dW(s). \tag{3.9}$$

3.2. Weak *B*-condition. In this subsection we recall the concept of weak B-condition (for $\tilde{A}$) and introduce an operator B satisfying it. This concept is fundamental in the theory of viscosity solutions in Hilbert spaces (see [36, Chapter 3]), which will be used in this paper. We first notice that by (3.2) and the definition of $\tilde{A}$, the adjoint operator $\tilde{A}^* : D(\tilde{A}^*) \subset X \to X$ is given by

$$\tilde{A}^* x = \begin{bmatrix} x_1(0) - x_0 \\ -{x_1}' \end{bmatrix}, \quad D(\tilde{A}^*) = D(A^*) = \left\{ x = (x_0, x_1) \in X : x_1 \in W^{1,2}([-d, 0]; \mathbb{R}^n), \right.$$
$$\left. x_1(-d) = 0 \right\}.$$

DEFINITION 3.4 (See [36, Definition 3.9]). *We say that $B \in \mathcal{L}(X)$ satisfies the weak B-condition (for $\tilde{A}$) if the following hold:*

(i) *B is strictly positive, i.e., $\langle Bx, x\rangle_X > 0$ for every $x \neq 0$;*

(ii) *B is self-adjoint;*

(iii) *$\tilde{A}^* B \in \mathcal{L}(X)$;*

(iv) *there exists $C_0 \geq 0$ such that*

$$\langle \tilde{A}^* Bx, x\rangle_X \leq C_0 \langle Bx, x\rangle_X \quad \forall x \in X.$$

We construct an operator B satisfying the weak B-condition. Let $\tilde{A}^{-1}$ be the inverse of the operator $\tilde{A}$. Its explicit expression can be derived from the proof of the second part of Proposition 3.3 with $y = -x$ and $\lambda = 0$ (which does not invalidate the calculations in the proof) and is given by

$$\tilde{A}^{-1}x = \left(-x_0, -x_0 - \int_{\cdot}^{0} x_1(\xi)d\xi\right) \quad \forall x = (x_0, x_1) \in X. \tag{3.10}$$

Notice that $\tilde{A}^{-1} \in \mathcal{L}(X)$. Moreover, since $\tilde{A}^{-1}$ is continuous as an operator from X to $W^{1,2}$, and the embedding $W^{1,2} \hookrightarrow L^2$ is compact, $\tilde{A}^{-1} : L^2 \to L^2$ is compact. Define now

$$B := (\tilde{A}^{-1})^* \tilde{A}^{-1} = (\tilde{A}^*)^{-1} \tilde{A}^{-1} \in \mathcal{L}(X). \tag{3.11}$$

B is compact by the compactness of $\tilde{A}^{-1}$.

PROPOSITION 3.5. *The operator B defined in* (3.11) *satisfies the weak B-condition for $\tilde{A}$ with $C_0 = 0$.*

Proof. It is immediate to see that $B \in \mathcal{L}(X)$, $\tilde{A}^* B = \tilde{A}^{-1} \in \mathcal{L}(X)$, and B is self-adjoint. Moreover,

$$\langle Bx, x\rangle_X = \langle \tilde{A}^{-1}x, \tilde{A}^{-1}x\rangle_X = |\tilde{A}^{-1}x|_X \geq 0 \quad \forall x \in X.$$

We now show that B is strictly positive. Let $x \neq 0$. If $x_0 \neq 0$, we have $|\tilde{A}^{-1}x|_X > 0$. On the other hand, if $x_0 = 0$, then we must have $x_1 \neq 0$ and then the function $\int_{\cdot}^{0} x_1(\xi)d\xi \not\equiv 0$. In both cases, by (3.10), we deduce the strict positivity of B.

Finally, by the dissipativity of $\tilde{A}$, we have

$$\langle A^* Bx, x\rangle_X = \langle \tilde{A}^{-1}x, x\rangle_X = \langle y, \tilde{A}y\rangle_X \leq 0$$

by taking $y = \tilde{A}^{-1}x$. The claim is proved. ∎

Observe that

$$Bx = \begin{bmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}, \qquad x = (x_0, x_1) \in X. \tag{3.12}$$

By the strict positivity of B, $B_{00} \in M^{n\times n}$ is strictly positive and B_{11} is strictly positive as an operator $L^2 \to L^2$. Moreover, since B is strictly positive and self-adjoint, the operator $B^{1/2} \in \mathcal{L}(X)$ is well defined, self-adjoint, and strictly positive. We introduce the $|\cdot|_{-1}$-norm on X by

$$\begin{aligned} |x|_{-1}^2 &= \langle B^{1/2}x, B^{1/2}x\rangle_X = \langle Bx, x\rangle_X \\ &= \langle (\tilde{A}^{-1})^* \tilde{A}^{-1}x, x\rangle_X = \langle \tilde{A}^{-1}x, \tilde{A}^{-1}x\rangle_X = |\tilde{A}^{-1}x|_X^2 \quad \forall x \in X. \end{aligned} \tag{3.13}$$

We define

$$X_{-1} := \quad \text{the completion of } X \text{ under } |\cdot|_{-1},$$

which is a Hilbert space endowed with the inner product

$$\langle x, y\rangle_{-1} := \langle B^{1/2}x, B^{1/2}y\rangle_X = \langle Bx, y\rangle_X = \langle \tilde{A}^{-1}x, \tilde{A}^{-1}y\rangle_X.$$

Notice that $|x|_{-1} \le |\tilde{A}^{-1}|_{\mathcal{L}(X)}|x|_X$; in particular, we have $(X, |\cdot|) \hookrightarrow (X_{-1}, |\cdot|_{-1})$. Moreover, strict positivity of B ensures that the operator $B^{1/2}$ can be extended to an isometry

$$B^{1/2} \colon (X_{-1}, |\cdot|_{-1}) \to (X, |\cdot|_X).$$

By (3.13) and an application of [26, Proposition B.1], we have $\text{Range}(B^{1/2}) = \text{Range}((\tilde{A}^{-1})^*)$. Since $\text{Range}((\tilde{A}^{-1})^*) = D(\tilde{A}^*)$, we have

$$\text{Range}\big(B^{1/2}\big) = D(\tilde{A}^*). \tag{3.14}$$

By (3.14), the operator $\tilde{A}^* B^{1/2}$ is well defined on the whole space X. Moreover, since $\tilde{A}^*$ is closed and $B^{1/2} \in \mathcal{L}(X)$, $\tilde{A}^* B^{1/2}$ is a closed operator. Thus, by the closed graph theorem, we have

$$\tilde{A}^* B^{1/2} \in \mathcal{L}(X). \tag{3.15}$$

Remark 3.6. In the infinite-dimensional theory of viscosity solutions it is only required that $\tilde{A}^* B \in \mathcal{L}(X)$ (condition (iii) of Definition 3.4). Such an operator can be constructed for any maximal dissipative operator $\tilde{A}$ (see, e.g., [36, Theorem 3.11]). In the case of the present paper, in addition, we also have $\tilde{A}^* B^{1/2} \in \mathcal{L}(X)$. Hence, here B is better than a generic B that one usually uses in the standard theory.

By (3.10), we immediately notice that

$$|x_0| \le |x|_{-1}, \quad \forall x = (x_0, x_1) \in X. \tag{3.16}$$

Since B is a compact, self-adjoint, and strictly positive operator on X, by the spectral theorem B admits a set of eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}} \subset (0, +\infty)$ such that $\lambda_i \to 0^+$ and a corresponding set $\{f_i\}_{i\in\mathbb{N}} \subset X$ of eigenvectors forming an orthonormal basis of X. By taking $\{e_i\}_{i\in\mathbb{N}}$ defined by $e_i := \frac{1}{\sqrt{\lambda_i}} f_i$, we then get an orthonormal basis of X_{-1}. We set $X^N := \text{Span}\{f_1, \dots f_N\} = \text{Span}\{e_1, \dots e_N\}$ for $N \ge 1$, and let $P_N \colon X \to X$ be the orthogonal projection onto X_N and $Q_N := I - P_N$. Since $\{e_i\}_{i\in\mathbb{N}}$ is an orthogonal basis of X_{-1}, the projections P_N, Q_N extend to orthogonal projections in X_{-1} and we will use the same symbols to denote them. We notice that

$$BP_N = P_N B, \quad BQ_N = Q_N B. \tag{3.17}$$

Therefore, since $|BQ_N|_{\mathcal{L}(X)} = |Q_N B|_{\mathcal{L}(X)}$ and B is compact, we get

$$\lim_{N\to\infty} |BQ_N|_{\mathcal{L}(X)} = 0. \tag{3.18}$$

4. Estimates for the state equation and the value function.

In this section we prove estimates for solutions of the state equation, the cost functional, and the value function. We first derive regularity properties for $\tilde{b}, \sigma, L$.

LEMMA 4.1. *Let Assumptions 2.4 and 2.5 hold. There exists $C>0$ and a local modulus of continuity ω such that the following hold true for every $x,y\in X, u\in U$:*

$$|\tilde b(x,u)-\tilde b(y,u)|\le C|x-y|_{-1}, \tag{4.1}$$

$$\langle \tilde b(x,u)-\tilde b(y,u), B(x-y)\rangle_X \le C|x-y|_{-1}^2, \tag{4.2}$$

$$|\tilde b(x,u)|\le C(1+|x|_X), \tag{4.3}$$

$$|\sigma(y,u)-\sigma(x,u)|_{\mathcal{L}_2(\mathbb{R}^q,X)}\le C|x-y|_{-1}, \tag{4.4}$$

$$|\sigma(x,u)|_{\mathcal{L}_2(\mathbb{R}^q,X)}\le C(1+|x|_X), \tag{4.5}$$

$$|L(x,u)-L(y,u)|\le \omega\left(|x-y|_{-1},R\right), \tag{4.6}$$

$$|L(x,u)|\le C\left(1+|x|_X^m\right). \tag{4.7}$$

Moreover,

$$\lim_{N\to\infty}\sup_{u\in U}\operatorname{Tr}\left[\sigma(x,u)\sigma(x,u)^*BQ_N\right]=0 \quad \forall x\in X. \tag{4.8}$$

Proof of (4.1) *and* (4.2). By our assumptions, we have $(0,a_1^i)\in D(\tilde A^*)$ for every $i\le h$, where a_1^i are the rows of the matrix a_1. Then,

$$\begin{aligned}\left|\int_{-d}^0 a_1(\xi)x_1(\xi)d\xi\right|^2 &= \sum_{i=1}^h |\langle(0,a_1^i),x\rangle_X|^2=\sum_{i=1}^h |\langle(0,a_1^i),\tilde A\tilde A^{-1}x\rangle_X|^2\\ &\le \sum_{i=1}^h |\langle\tilde A^*(0,a_1^i),\tilde A^{-1}x\rangle_X|^2\le\sum_{i=1}^h|\tilde A^*(0,a_1^i)|_X^2|x|_{-1}^2=C|x|_{-1}^2.\end{aligned} \tag{4.9}$$

Thus, by (4.9) and (3.16), we get

$$\begin{aligned}|\tilde b(x,u)-\tilde b(y,u)|_{-1}&\le C|\tilde b(x,u)-\tilde b(y,u)|_X\\ &\le\left|b_0\left(x_0,\int_{-d}^0 a_1(\theta)x_1(\theta)d\theta,u\right)-b_0\left(y_0,\int_{-d}^0 a_1(\theta)y_1(\theta)d\theta,u\right)\right|\\ &\quad+|x_0-y_0|\le C\left(|x_0-y_0|+\left|\int_{-d}^0 a_1(\theta)(x_1(\theta)-y_1(\theta))d\theta\right|\right)\\ &\le C|x-y|_{-1}.\end{aligned} \tag{4.10}$$

Inequality (4.2) follows trivially from (4.1). ◻

Proof of (4.3). The estimate follows easily from Assumption 2.4, the definition of $\tilde b$ and the fact that $a_1^j\in W^{1,2}$ for $j=1,\dots,h$. ◻

Proof of (4.4). We have

$$\begin{aligned}&|\sigma(y,u)-\sigma(x,u)|_{\mathcal{L}_2(\mathbb{R}^q,X)}\\ &\quad\le C\left|\sigma_0\left(x_0,\int_{-d}^0 a_2(\theta)x_1(\theta)d\theta,u\right)-\sigma_0\left(y_0,\int_{-d}^0 a_2(\theta)y_1(\theta)d\theta,u\right)\right|.\end{aligned}$$

Since $(0,a_2)\in D(\tilde A^*)$, by (4.9) with a_2 in place of a_1, we have

$$\begin{aligned}&\left|\sigma_0\left(x_0,\int_{-d}^0 a_2(\theta)x_1(\theta)d\theta,u\right)-\sigma_0\left(y_0,\int_{-d}^0 a_2(\theta)y_1(\theta)d\theta,u\right)\right|\\ &\quad\le C\left(|x_0-y_0|+\left|\int_{-d}^0 a_2(\theta)(x_1(\theta)-y_1(\theta))d\theta\right|\right)\le C|x-y|_{-1}.\end{aligned}$$ ◻

Proof of (4.5). Inequality (4.5) follows from the definition of $\sigma\colon X\times U\to\mathcal{L}(\mathbb{R}^q,X)$, since

$$|\sigma(x,u)|^2_{\mathcal{L}_2(\mathbb{R}^q,X)}=\sum_{i=1}^n|\sigma_0(x,u)v_i|^2\le C|\sigma_0(x,u)|^2\le C(1+|x|^2_X),$$

where $\{v_i\}_{i=1,\dots,q}$ is the canonical basis of $\mathbb{R}^q$. □

Proof of (4.6). The proof follows from the definition of L, (2.4), and (3.16), as we have

$$|L(x,u)-L(y,u)|\le\omega_R(|x_0-y_0|)\le\omega_R(|x-y|_{-1}).$$ □

Proof of (4.7). The proof follows from the definition of L and (2.3). □

Proof of (4.8). We notice that by [36, Appendix B] and (4.5), we have

$$\begin{aligned}|\operatorname{Tr}[\sigma(x,u)\sigma(x,u)^*BQ_N]|&\le|\sigma(x,u)\sigma(x,u)^*BQ_N|_{\mathcal{L}_1(X)}\\&\le|\sigma(x,u)\sigma(x,u)^*|_{\mathcal{L}_1(X)}|BQ_N|_{\mathcal{L}(X)}\\&\le|\sigma(x,u)|^2_{\mathcal{L}_2(\mathbb{R}^q,X)}|BQ_N|_{\mathcal{L}(X)}\le C(1+|x|^2)|BQ_N|_{\mathcal{L}(X)}.\end{aligned}$$

Thus we obtain (4.8) by taking the supremum over u, letting $N\to\infty$ and using (3.18). □

Remark 4.2. The requirements $a_1(-d)=a_2(-d)=0$ are, in general, necessary to get (4.9). Indeed, consider, for example, the case $a_1(\cdot)\equiv1$. In such a case the sequence

$$x^N=(x_0^N,x_1^N),\quad x_0^N=0,\ x_1^N=NI_{[-d,-d+1/N]},\quad N\ge1,$$

is such that

$$\left|\int_{-d}^0a_1(\xi)x_1^N(\xi)d\xi\right|=1\quad\forall N\ge1,\qquad|x^N|_{-1}\to0\text{ when }N\to\infty,$$

and then (4.9) cannot be satisfied.

We recall [36, Proposition 3.24]. Set

$$\rho_0:=\begin{cases}0&\text{if }m=0,\\Cm+\frac12C^2m&\text{if }0<m<2,\\Cm+\frac12C^2m(m-1)&\text{if }m\ge2,\end{cases}$$

where C is the constant appearing in (4.3) and (4.5), and m is the constant from Assumption 2.5 and (4.7).

PROPOSITION 4.3 (see [36, Proposition 3.24]). *Let Assumption* 2.4 *hold, and let* $\lambda>\rho_0$. *Let* $Y(t)$ *be the mild solution of* (3.8) *with initial datum* $x\in X$ *and control* $u(\cdot)\in\mathcal{U}$. *Then, there exists* $C_\lambda>0$ *such that*

$$\mathbb{E}[|Y(t)|^m_X]\le C_\lambda(1+|x|^m_X)e^{\lambda t}\quad\forall t\ge0.$$

We need the following assumption.

Assumption 4.4. $\rho>\rho_0$.

PROPOSITION 4.5. *Let Assumptions* 2.4, 2.5, *and* 4.4 *hold. There exists* $C>0$ *such that*

$$|J(x;u(\cdot))| \leq C(1+|x|_X^m) \quad \forall x \in X,\ \forall u(\cdot) \in \mathcal{U}.$$

Hence,

$$|V(x)| \leq C(1+|x|_X^m) \quad \forall x \in X.$$

Proof. By (4.7) and Proposition 4.3 applied with $\lambda = (\rho+\rho_0)/2$, we have

$$|J(x,u(\cdot))| \leq C \int_0^\infty e^{-\rho t}\mathbb{E}[(1+|Y(t)|_X)^m]dt \leq C(1+|x|_X^m) \quad \forall x \in X,\ \forall u(\cdot) \in \mathcal{U}.$$

The estimate on V follows from this. ∎

Next, we show continuity properties of V. We recall first the notion of B-continuity (see [36, Definition 3.4]).

DEFINITION 4.6. *Let* $B \in \mathcal{L}(X)$ *be a strictly positive self-adjoint operator. A function* $u : X \to \mathbb{R}$ *is said to be* B*-upper semicontinuous (respectively,* B*-lower semicontinuous) if, for any sequence* $\{x_n\}_{n\in\mathbb{N}} \subset X$ *such that* $x_n \rightharpoonup x \in X$ *and* $Bx_n \to Bx$ *as* $n \to \infty$*, we have*

$$\limsup_{n\to\infty} u(x_n) \leq u(x) \qquad \textit{(respectively,} \quad \liminf_{n\to\infty} u(x_n) \geq u(x)).$$

A function $u : X \to \mathbb{R}$ *is said to be* B*-continuous if it is both* B*-upper semicontinuous and* B*-lower semicontinuous.*

We remark that, since the operator B defined in (3.11) is compact, in our case B-upper/lower semicontinuity is equivalent to the weak sequential upper/lower semicontinuity, respectively.

PROPOSITION 4.7. *Let Assumptions* 2.4, 2.5, *and* 4.4 *hold. For every* $R>0$, *there exists a modulus of continuity* ω_R *such that*

$$|V(x)-V(y)| \leq \omega_R(|x-y|_{-1}) \quad \forall x,y \in X,\ \textit{s.t. } |x|,|y| \leq R. \tag{4.11}$$

Hence V *is* B*-continuous and thus weakly sequentially continuous.*

Proof. We prove the estimate

$$|J(x,u)-J(y,u)| \leq \omega_R(|x-y|_{-1}) \quad \forall x,y \in X:\ |x|,|y| \leq R, \forall u(\cdot) \in U,$$

as in [36, Proposition 3.73], since the assumptions of the latter are satisfied due to Lemma 4.1. Then, (4.11) follows. As for the last claim, we observe that by (4.11) and by [36, Lemma 3.6(iii)], V is B-continuous in X. ∎

We point out that V may not be continuous with respect to the $|\cdot|_{-1}$ norm in the whole X.

5. HJB equation: Viscosity solutions. In this section we characterize V as the unique B-continuous viscosity solution to the associated HJB equation. Given $v \in C^1(X)$, we denote by $Dv(x)$ its Fréchet derivative at $x \in X$ and we write

$$Dv(x) = \begin{bmatrix} D_{x_0}v(x) \\ D_{x_1}v(x) \end{bmatrix},$$

where $D_{x_0}v(x), D_{x_1}v(x)$ are the partial Fréchet derivatives. For $v \in C^2(X)$, we denote by $D^2v(x)$ its second order Fréchet derivative at $x \in X$ which we will often write as

$$D^2v(x) = \begin{bmatrix} D^2_{x_0^2}v(x) & D^2_{x_0x_1}v(x) \\ D^2_{x_1x_0}v(x) & D^2_{x_1^2}v(x) \end{bmatrix}.$$

We define the Hamiltonian function $H : X \times X \times S(X) \to \mathbb{R}$ by

$$\begin{aligned}
&H(x,p,Z)\\
&\quad := \sup_{u\in U}\left\{\langle -\tilde{b}(x,u),p\rangle - \frac{1}{2}\mathrm{Tr}(\sigma(x,u)\sigma(x,u)^*Z) - L(x,u)\right\}\\
&\quad = -x_0\cdot p_0 + \sup_{u\in U}\Bigg\{ - b_0\left(x_0, \int_{-d}^0 a_1(\xi)x_1(\xi)\,d\xi, u\right)\cdot p_0\\
&\qquad - \frac{1}{2}\mathrm{Tr}\left[\sigma_0\left(x_0, \int_{-d}^0 a_2(\xi)x_1(\xi)\,d\xi, u\right)\sigma_0\left(x_0, \int_{-d}^0 a_2(\xi)x_1(\xi)\,d\xi, u\right)^T Z_{00}\right] - l(x_0,u)\Bigg\}\\
&\quad = -x_0\cdot p_0 + \sup_{u\in U}\left\{ - b_0(x,u)\cdot p_0 - \frac{1}{2}\mathrm{Tr}\left[\sigma_0(x,u)\,\sigma_0(x,u)^T Z_{00}\right] - l(x_0,u)\right\}\\
&\quad =: \tilde{H}(x,p_0,Z_{00}).
\end{aligned}$$

By [36, Theorem 3.75] the Hamiltonian H satisfies the following properties.

LEMMA 5.1. *Let Assumptions* 2.4 *and* 2.5 *hold.*

(i) *H is uniformly continuous on bounded subsets of $X \times X \times S(X)$.*

(ii) *For every $x, p \in X$ and every $Y, Z \in S(X)$ such that $Z \le Y$, we have*

$$H(x,p,Y) \le H(x,p,Z). \tag{5.1}$$

(iii) *For every $x, p \in X$ and every $R > 0$, we have*

(5.2)
$$\lim_{N\to\infty} \sup\left\{|H(x,p,Z+\lambda BQ_N) - H(x,p,Z)| : \ |Z_{00}| \le R, \ |\lambda| \le R\right\} = 0.$$

(iv) *For every $R > 0$ there exists a modulus of continuity ω_R such that*

$$\begin{aligned}
&H\left(z, \frac{B(z-y)}{\varepsilon}, Z\right) - H\left(y, \frac{B(z-y)}{\varepsilon}, Y\right)\\
&\quad \ge -\omega_R\left(|z-y|_{-1}\left(1 + \frac{|z-y|_{-1}}{\varepsilon}\right)\right)
\end{aligned} \tag{5.3}$$

for every $\varepsilon > 0$, $y, z \in X$ such that $|y|_X, |z|_X \le R$, $Y, Z \in \mathcal{S}(X)$ satisfying

$$Y = P_NYP_N, \quad Z = P_NZP_N,$$

and

$$\frac{3}{\varepsilon}\begin{pmatrix} BP_N & 0 \\ 0 & BP_N \end{pmatrix} \le \begin{pmatrix} Y & 0 \\ 0 & -Z \end{pmatrix} \le \frac{3}{\varepsilon}\begin{pmatrix} BP_N & -BP_N \\ -BP_N & BP_N \end{pmatrix}.$$

(v) *If $C > 0$ is the constant in* (4.3) *and* (4.5), *then, for every $x \in X, p, q \in X, Y, Z \in \mathcal{S}(X)$,*

(5.4)
$$|H(x,p+q,Y+Z) - H(x,p,Y)| \le C(1+|x|_X)|q_0| + \frac{1}{2}C^2(1+|x|_X)^2|Z_{00}|.$$

The HJB equation associated with the optimal control problem is the infinite-dimensional PDE

$$\rho v(x) - \langle \tilde{A}x, Dv(x)\rangle + H(x, Dv(x), D^2 v(x)) = 0, \quad x \in X. \tag{5.5}$$

We recall the definition of B-continuous viscosity solution from [36].

Definition 5.2.

(i) $\phi\colon X \to \mathbb{R}$ *is a regular test function if*

$$\phi \in \Phi := \{\phi \in C^2(X) : \phi \text{ is } B\text{-lower semicontinuous and } \phi, D\phi, D^2\phi, A^*D\phi \text{ are uniformly continuous on } X\};$$

(ii) $g\colon X \to \mathbb{R}$ *is a radial test function if*

$$g \in \mathcal{G} := \{g \in C^2(X) : g(x) = g_0(|x|_X) \text{ for some } g_0 \in C^2([0,\infty)) \text{ nondecreasing}, \; {g_0}'(0) = 0\}.$$

Note that, if $g \in \mathcal{G}$, we have

$$Dg(x) = \begin{cases} g_0'(|x|_X)\frac{x}{|x|_X} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases} \tag{5.6}$$

We say that a function is locally bounded if it is bounded on bounded subsets of X.

Definition 5.3.

(i) *A locally bounded B-upper semicontinuous function $v : X \to \mathbb{R}$ is a viscosity subsolution of* (5.5) *if, whenever $v - \phi - g$ has a local maximum at $x \in X$ for $\phi \in \Phi, g \in \mathcal{G}$, then*

$$\rho v(x) - \langle x, \tilde{A}^* D\phi(x)\rangle_X + H(x, D\phi(x) + Dg(x), D^2\phi(x) + D^2 g(x)) \leq 0.$$

(ii) *A locally bounded B-lower semicontinuous function $v : X \to \mathbb{R}$ is a viscosity supersolution of* (5.5) *if, whenever $v + \phi + g$ has a local minimum at $x \in X$ for $\phi \in \Phi$, $g \in \mathcal{G}$, then*

$$\rho v(x) + \langle x, \tilde{A}^* D\phi(x)\rangle_X + H(x, -D\phi(x) - Dg(x), -D^2\phi(x) - D^2 g(x)) \geq 0.$$

(iii) *A viscosity solution of* (5.5) *is a function $v : X \to \mathbb{R}$ which is both a viscosity subsolution and a viscosity supersolution of* (5.5).

Define

$$\mathcal{S} := \{u\colon X \to \mathbb{R} : \exists k \geq 0 \text{ satisfying (5.7) and } \tilde{C} \geq 0 \text{ such that } |u(x)| \leq \tilde{C}(1 + |x|_X^k)\},$$

where

$$\begin{cases} k < \frac{\rho}{C + \frac{1}{2}C^2} & \text{if } \frac{\rho}{C+\frac{1}{2}C^2} \leq 2, \\ Ck + \frac{1}{2}C^2 k(k-1) < \rho & \text{if } \frac{\rho}{C+\frac{1}{2}C^2} > 2, \end{cases} \tag{5.7}$$

and C is the constant appearing in (4.3) and (4.5).

We can now state the theorem characterizing V as the unique viscosity solution of (5.5) in $\mathcal{S}$.

THEOREM 5.4. *Let Assumptions* 2.4, 2.5, *and* 4.4 *hold. The value function* V *is the unique viscosity solution of* (5.5) *in the set* $\mathcal{S}$.

Proof. Notice that $V \in \mathcal{S}$ by Proposition 4.5.

The proof of the fact that V is the unique viscosity solution of the HJB equation can be found in [36, Theorem 3.75] as all assumptions of this theorem are satisfied due to Lemma 5.1. The reader can also check the comparison theorem [36, Theorem 3.56]. ∎

Remark 5.5. We remark that Theorem 5.4 also holds in the deterministic case, i.e., when $\sigma(x,u)=0$. (in which case we may take $\rho_0 = Cm$ and $k < \rho/C$ in (5.7)). The theory of viscosity solutions handles well degenerate HJB equations, i.e., when the Hamiltonian satisfies

$$H(x,p,Y) \le H(x,p,Z)$$

for every $Y,Z \in S(X)$ such that $Z \le Y$. Hence viscosity solutions can be used in connection with the dynamic programming method for optimal control of stochastic differential equations in the case of degenerate noise in the state equation, in particular, when it completely vanishes (deterministic case). This is not possible using the mild solutions approach due to its various limitations described in the introduction (for more on this, see [36]).

6. Partial regularity. In this section we prove partial regularity of V with respect to the x_0-variable. To do this we assume the following.

Assumption 6.1. For every $R > 0$ there exists $K_R > 0$ such that

$$|V(y) - V(x)| \le K_R |y-x|_{-1} \quad \forall x,y \in X, |x|,|y| \le R.$$

Assumption 6.1 is satisfied in many natural cases when the cost function $l(\cdot,u)$ is Lipschitz continuous as we illustrate in the following example.

Example 6.2. Suppose that $l(\cdot,u)$ is Lipschitz continuous, uniformly in u, and $\rho > C + \frac{C^2|B|_{\mathcal{L}(X)}}{2}$, where C is the constant from (4.2) and (4.4). Indeed, fix $x,y \in X$, $u(\cdot) \in \mathcal{U}$ and denote by $X(s)$, $Y(s)$ the mild solutions of the state equation with initial data x,y, respectively, and the same control $u(\cdot)$. Then, by [36, Lemma 3.20], we have

$$\mathbb{E}\left[|X(r) - Y(r)|_{-1}^2\right] \le C(r)|x-y|_{-1}^2, \tag{6.1}$$

where, recalling that the constant of the weak B-condition in our case is $C_0 = 0$,

$$C(r) = e^{(2C + C^2|B|_{\mathcal{L}(X)})r}.$$

Since $L(x,u) = l(x_0,u)$, the Lipschitz continuity of $l(\cdot,u)$ (uniform in u), (3.16), (6.1), and $\rho > C + \frac{C^2|B|_{\mathcal{L}(X)}}{2}$ yield

$$\begin{aligned}
&\mathbb{E}\int_0^\infty e^{-\rho r}|L(X(r),u(r)) - L(Y(r),u(r))|dr \le C_1 \int_0^\infty e^{-\rho r}\mathbb{E}|X_0(r) - Y_0(r)|dr \\
&\le C_1 \int_0^\infty e^{-\rho r}\mathbb{E}|X(r)-Y(r)|_{-1}dr \le C_1 \int_0^\infty e^{-\rho r} e^{\left(C + \frac{C^2|B|_{\mathcal{L}(X)}}{2}\right)r} dr |x-y|_{-1} \\
&\le C_2|x-y|_{-1}
\end{aligned}$$

and the claim easily follows.

Remark 6.3. Notice that, if we only assume that $l(\cdot,u)$ is locally Lipschitz continuous, uniformly in u, the above proof would not work and, by using an argument outlined in the proof of Proposition 4.7, we would only get that V is uniformly continuous with respect to the $|\cdot|_{-1}$ norm on bounded sets of X.

Finally, we assume local uniform nondegeneracy of $\sigma_0\colon \mathbb{R}^n \times L^2 \times U \to \mathbb{R}^n \times \mathbb{R}^q$ (recall the definition of $\sigma_0(x,u)$ in section 3).

Assumption 6.4. For every $R>0$ there exists $\lambda_R>0$ such that

$$\sigma_0(x,u)\sigma_0(x,u)^T \geq \lambda_R I \quad \forall x \text{ such that } |x|_X \leq R \;\; \forall u \in U.$$

For every $\bar{x}_1 \in L^2$ we define

$$V^{\bar{x}_1}(x_0) := V(x_0,\bar{x}_1) \quad \forall x_0 \in \mathbb{R}^n.$$

Theorem 6.5 is the main result of this section.

THEOREM 6.5. *Let Assumptions* 2.4, 2.5, 4.4, 6.1, *and* 6.4 *hold. For every* $p>n$ *and every fixed* $\bar{x}_1 \in L^2$, *we have* $V^{\bar{x}_1} \in W^{2,p}_{loc}(\mathbb{R}^n)$; *thus, by Sobolev embedding,* $V^{\bar{x}_1} \in C^{1,\alpha}_{loc}(\mathbb{R}^n)$ *for all* $0<\alpha<1$. *Moreover, for every* $R>0$, *there exists* $C_R>0$ *such that*

$$|V^{\bar{x}_1}|_{W^{2,p}(B_R)} \leq C_R \qquad \forall \bar{x}_1 \textit{ such that } |\bar{x}_1|_{L^2} \leq R.$$

Finally, $D_{x_0}V$ *is continuous with respect to the* $|\cdot|_{-1}$ *norm on bounded sets of* X. *In particular,* $D_{x_0}V$ *is continuous in* X.

Remark 6.6. We point out that Assumption 6.4 is a nondegeneracy assumption for the finite-dimensional valued matrix $\sigma_0\colon \mathbb{R}^n \times L^2 \times U \to \mathbb{R}^n \times \mathbb{R}^q$, not for the operator $\sigma: X \times U \to \mathcal{L}(\mathbb{R}^q, X)$ (recall its definition in section 3), which is (highly) degenerate. Hence, classical regularization methods via Bismut-Elworthy-Li formula (see, e.g., [26]), which assume nondegeneracy of σ, cannot be applied. On the other hand, the Bismut-Elworthy-Li formula has been extended to degenerate cases when the pseudo-inverse of σ is not bounded (see, e.g., [1] and the references therein). These results have been applied to optimal control problems in the case of uncontrolled diffusion term and semilinear HJB equation. However, these methods seem to rely strongly on the *semilinear structure* of the HJB equation and it is not clear how to extend them to our *fully nonlinear* HJB equation, where the viscosity approach seems to be the only one employable.

With respect to the aforementioned literature, we also point out that our Theorem 6.5 provides local $C^{1,\alpha}$-partial regularity of the value function V, without assumptions of differentiability of the coefficients; whereas, in the approaches via mild solutions/BSDE's, one usually gets Gateaux differentiability in the whole space under assumptions of regularity of the coefficients. Hence, when restricted to $\mathbb{R}^n$, our result is stronger.

Before presenting the proof, we explain its basic idea. Since the range of $\sigma(x,u)$ is finite dimensional (the x_1-component of $\sigma(x,u)w$ for $w \in \mathbb{R}^q$ is 0), we first show that the function $V^{\bar{x}_1}$, defined on a finite-dimensional space, is a viscosity subsolution (respectively, supersolution) of a finite-dimensional equation (6.8) (respectively, (6.9)) in $\mathbb{R}^n$. This part contains the most technical difficulties. Once this is done, we can then apply the theory of L^p-viscosity solutions (e.g., [82]) to obtain that $V^{\bar{x}_1}$ is an L^p-viscosity solution of a fully nonlinear, uniformly elliptic PDE equation in $\mathbb{R}^n$ with a locally bounded right-hand side function, i.e., (6.10). Then, the regularity theory for uniformly elliptic equations yields $V^{\bar{x}_1} \in W^{2,p}_{\text{loc}}(\mathbb{R}^n)$ for every $p > n$. As said in the introduction, this approach is inspired by the arguments of [79] for Kolmogorov equations, where approximating value functions $V_n^{\bar{x}_1}$ are proved to be L^p-viscosity solutions of some finite-dimensional parabolic PDE for which $C^{1,\alpha}$-estimates hold; the result follows by passing to the limit. The method we present here is a refinement of that one: we avoid the complex approximating procedure employed in [79] by using new and deeper results from the theory of L^p-viscosity solutions and are able to prove directly that $V^{\bar{x}_1}$ is an L^p-viscosity solution of a finite-dimensional elliptic PDE (6.10), which then gives the desired $C^{1,\alpha}$-regularity.

Proof. We organize the proof in several steps.

Step 1. Fix $\bar{x}_1 \in L^2$, and let $R \geq 1$ be such that $|\bar{x}_1|_{L^2} \leq R$. Let $\bar{x}_0 \in \mathbb{R}^n$ be such that $|\bar{x}_0| \leq R$ and $\varphi \in C^2(\mathbb{R}^n)$ be such that $V^{\bar{x}_1} - \varphi$ has a strict global maximum at $\bar{x}_0$. We assume without loss of generality that the maximum is equal to 0 and that $\varphi > 0$ if $|x_0| > 4R$. We extend φ to X by defining $\tilde{\varphi}(x_0, x_1) := \varphi(x_0)$. With an abuse of notation we will still write $\varphi(x_0)$ for $\tilde{\varphi}(x)$ and $D\varphi(x_0) = (D_{x_0}\varphi(x_0), 0)$. Note that $D\varphi(x_0) \in D(\tilde{A}^*)$ for all $x \in X$. Set $\bar{x} := (\bar{x}_0, \bar{x}_1)$. We consider, for $\varepsilon > 0$, the functions

$$\Phi_\varepsilon(x) = V(x) - \varphi(x_0) - \frac{1}{\varepsilon}|x - (x_0, \bar{x}_1)|^2_{-1} - M(|x|_X - 2R)^{m+4}_+, \quad x = (x_0, x_1) \in X,$$

where M is chosen so that $V(x) < \frac{M}{2}(|x|_X - 2R)^{m+4}_+$ if $|x|_X > 4R$. Then

$$\Phi_\varepsilon(x) < -\frac{M}{2}(|x|_X - 2R)^{m+4}_+ \quad \forall x \in X, |x|_X > 4R. \tag{6.2}$$

Observe that Φ_ε is weakly sequentially upper semicontinuous as V is weakly sequentially continuous by Proposition 4.7, the functions $x \mapsto \varphi(x_0)$ and $x \mapsto |x|^2_{-1}$ are weakly sequentially continuous, and the function $x \mapsto |x|_X$ is weakly sequentially lowersemicontinuous.

We distinguish two cases: (i) $\sup_X \Phi_\varepsilon > 0$ for every $\varepsilon > 0$; (ii) $\Phi_\varepsilon \leq 0$ for some (small) ε.

Case (i). Recall that $V^{\bar{x}_1}(\bar{x}_0) - \varphi(\bar{x}_0) = 0$ and $|\bar{x}|_X \leq \sqrt{2}R$. Then Φ_ε has global maximum at some $\hat{x}^\varepsilon \in X$, with $|\hat{x}^\varepsilon|_X \leq 4R$ and $\Phi_\varepsilon(\hat{x}^\varepsilon) > 0$. Recalling Assumption 6.1 and since $V^{\bar{x}_1} - \varphi$ has a strict global maximum at $\bar{x}_0$ equal to 0, we have

$$\begin{aligned}
0 < \Phi_\varepsilon(\hat{x}^\varepsilon) &\leq V(\hat{x}^\varepsilon) - \varphi(\hat{x}^\varepsilon_0) - \frac{1}{\varepsilon}|\hat{x} - (\hat{x}^\varepsilon_0, \bar{x}_1)|^2_{-1} \\
&= V(\hat{x}^\varepsilon) - V(\hat{x}^\varepsilon_0, \bar{x}_1) + V^{\bar{x}_1}(\hat{x}^\varepsilon_0) - \varphi(\hat{x}^\varepsilon_0) - \frac{1}{\varepsilon}|\hat{x}^\varepsilon - (\hat{x}^\varepsilon_0, \bar{x}_1)|^2_{-1} \\
&\leq K_{4R}|\hat{x}^\varepsilon - (\hat{x}^\varepsilon_0, \bar{x}_1)|_{-1} - \frac{1}{\varepsilon}|(0, \hat{x}^\varepsilon_1 - \bar{x}_1)|^2_{-1} \\
&= K_{4R}|(0, \hat{x}^\varepsilon_1 - \bar{x}_1)|_{-1} - \frac{1}{\varepsilon}|(0, \hat{x}^\varepsilon_1 - \bar{x}_1)|^2_{-1}.
\end{aligned}$$

It thus follows that

$$|(0, \hat{x}^\varepsilon_1 - \bar{x}_1)|_{-1} \leq K_{4R}\varepsilon. \tag{6.3}$$

On the other hand, by the fact that $\sup_X \Phi_\varepsilon > 0$, we have

$$\begin{aligned}
V^{\bar{x}_1}(\bar{x}_0) - \varphi(\bar{x}_0) &= \sup_{\mathbb{R}^n}(V^{\bar{x}_1} - \varphi) = 0 < \sup_X \Phi_\varepsilon = \Phi_\varepsilon(\hat{x}^\varepsilon) \\
&\leq V(\hat{x}^\varepsilon) - \varphi(\hat{x}^\varepsilon_0) = V(\hat{x}^\varepsilon) - V^{\bar{x}_1}(\hat{x}^\varepsilon_0) + V^{\bar{x}_1}(\hat{x}^\varepsilon_0) - \varphi(\hat{x}^\varepsilon_0).
\end{aligned}$$

Now, taking the $\liminf_{\varepsilon\to 0}$ above, by (6.3) and Assumption 6.1, we obtain

$$V^{\bar{x}_1}(\bar{x}_0) - \varphi(\bar{x}_0) \le \liminf_{\varepsilon\to 0}(V^{\bar{x}_1}(\hat{x}_0^\varepsilon) - \varphi(\hat{x}_0^\varepsilon)).$$

Since $V^{\bar{x}_1} - \varphi$ has a strict global maximum at $\bar{x}_0$ we thus must have

$$\lim_{\varepsilon\to 0} |\hat{x}_0^\varepsilon - \bar{x}_0| = 0. \tag{6.4}$$

In particular, from (6.3) and (6.4) it now follows that

$$\lim_{\varepsilon\to 0} |\hat{x}^\varepsilon - \bar{x}|_{-1} = 0. \tag{6.5}$$

Case (ii). In this case Φ_ε has a maximum at $\bar{x}$ since $\Phi_\varepsilon(\bar{x}) = 0$, so we easily get (6.5) with $\hat{x}^\varepsilon = \bar{x}$.

Step 2. Define $\psi = \phi + g$, where

$$\phi(x) := \varphi(x_0) + \frac{1}{\varepsilon}|x - (x_0, \bar{x}_1)|_{-1}^2, \qquad g(x) := M(|x| - 2R)_+^{m+4}.$$

With these definitions we have

$$V - \phi - g = \Phi_\varepsilon,$$

so that $V - \phi - g$ has a global maximum at $\hat{x}^\varepsilon$. Moreover,

$$\begin{aligned} &D\phi(x) = D\varphi(x_0) + \frac{2}{\varepsilon}B(x - (x_0, \bar{x}_1)),\\ &D_{x_0}\phi(x) = D_{x_0}\varphi(x_0), \end{aligned}$$

and

$$|D^2 g(x)| + |Dg(x)| \le \bar{C}_R \quad \forall |x| \le 4R. \tag{6.6}$$

Notice that, since $D\varphi(x_0) \in D(\tilde{A}^*)$, we have $\phi \in \Phi$, i.e., it is a regular test function according to Definition 5.2. Moreover, clearly $g \in \mathcal{G}$, i.e., it is a radial test function according to Definition 5.2. We will write $\bar{C}_R$ to denote a generic constant, possibly changing from line to line, depending on R and the data of the problem, which is independent of ε, φ, and on $x \in B_R$. Then, since V is a viscosity subsolution to (5.5), we have

$$\rho V(\hat{x}^\varepsilon) - \left\langle \hat{x}^\varepsilon, \tilde{A}^*(D\varphi(\hat{x}_0^\varepsilon) + \frac{1}{\varepsilon}B(0, \hat{x}_1^\varepsilon - \bar{x}_1)) \right\rangle + H(\hat{x}^\varepsilon, D\psi(\hat{x}^\varepsilon), D^2\psi(\hat{x}^\varepsilon)) \le 0.$$

By (3.15) and (6.3), we have

$$\begin{aligned} \frac{1}{\varepsilon}|\tilde{A}^* B(0, \hat{x}_1^\varepsilon - \bar{x}_1)| &= \frac{1}{\varepsilon}|\tilde{A}^* B^{1/2} B^{1/2}(0, \hat{x}_1^\varepsilon - \bar{x}_1)|\\ &\le |\tilde{A}^* B^{1/2}|_{\mathcal{L}(X)} \frac{1}{\varepsilon}|(0, \hat{x}_1^\varepsilon - \bar{x}_1)|_{-1} \le \bar{C}_R. \end{aligned}$$

The latter two inequalities, Proposition 4.5, (5.4), (6.6), the fact that $|\hat{x}^\varepsilon|_X \le 4R$, and the definition of $\tilde{H}$ imply

$$-\langle \hat{x}^\varepsilon, \tilde{A}^* D\varphi(\hat{x}_0^\varepsilon)\rangle + \tilde{H}(\hat{x}^\varepsilon, D_{x_0}\varphi(\hat{x}_0^\varepsilon), D_{x_0^2}^2\varphi(\hat{x}_0^\varepsilon)) \le \bar{C}_R.$$

Recalling (6.4), we have

$$\lim_{\varepsilon\to 0}\tilde{A}^* D\varphi(\hat{x}_0^\varepsilon) = \tilde{A}^* D\varphi(\bar{x}_0).$$

Hence, letting $\varepsilon \to 0$ in the previous inequality and using (6.5) and the continuity of $\tilde{H}$, we obtain

$$-\langle \bar{x}, \tilde{A}^* D\varphi(\bar{x}_0)\rangle + \tilde{H}(\bar{x}, D_{x_0}\varphi(\bar{x}), D^2_{x_0^2}\varphi(\bar{x})) \le \bar{C}_R.$$

Now, since $V^{\bar{x}_1} - \varphi$ has a strict global maximum at $\bar{x}_0$ and by Assumption 6.1 the function $V^{\bar{x}_1}$ is Lipschitz continuous on $\mathbb{R}^n$, with Lipschitz constant $\bar{C}_R$ independent of $\bar{x}_1$, we have $|D_{x_0}\varphi(\bar{x}_0)| \le \bar{C}_R$. Then, as $D\varphi(\bar{x}_0) \in D(\tilde{A}^*)$ we have $|\tilde{A}^* D\varphi(\bar{x}_0)|_X = |-(D_{x_0}\varphi(\bar{x}_0), 0)|_X \le \bar{C}_R$.

Therefore, by (5.4), $|\bar{x}| \le \sqrt{2}R$ and the definition of $\tilde{H}$, we have

$$\tilde{H}(\bar{x}, 0, D^2_{x_0^2}\varphi(\bar{x}_0)) - \bar{C}_R \le \tilde{H}(\bar{x}, D_{x_0}\varphi(\bar{x}_0), D^2_{x_0^2}\varphi(\bar{x}_0)),$$

so that we obtain

$$\tilde{H}(\bar{x}, 0, D^2_{x_0^2}\varphi(\bar{x}_0)) \le \bar{C}_R,$$

i.e.,

$$\sup_{u\in U}\left[-\frac{1}{2}\mathrm{Tr}(\sigma_0(\bar{x}, u)\sigma_0(\bar{x}, u)^T D^2_{x_0^2}\varphi(\bar{x}_0))\right] \le \bar{C}_R \tag{6.7}$$

for some constant $\bar{C}_R > 0$ independent of φ and $\bar{x}$ if $|\bar{x}_0| \le R, |\bar{x}_1| \le R$. Thus, for every $\bar{x}_1$ with $|\bar{x}_1| \le R$, the function $V^{\bar{x}_1}$ is a viscosity subsolution of the finite-dimensional equation

$$\sup_{u\in U}\left[-\frac{1}{2}\mathrm{Tr}(\sigma_0((x_0, \bar{x}_1), u)\sigma_0((x_0, \bar{x}_1), u)^T D^2 v(x_0))\right] = \bar{C}_R \quad \text{in } \{|x_0| < R\}. \tag{6.8}$$

Similarly, we prove that for every $|\bar{x}_1| \le R$, the function $V^{\bar{x}_1}$ is a viscosity supersolution of

$$\sup_{u\in U}\left[-\frac{1}{2}\mathrm{Tr}(\sigma_0((x_0, \bar{x}_1), u)\sigma_0((x_0, \bar{x}_1), u)^T D^2 v(x_0))\right] = -\bar{C}_R \quad \text{in } \{|x_0| < R\}. \tag{6.9}$$

Step 3. We now employ the theory of L^p-viscosity solutions. Since the readers may not be familiar with it, we will proceed slowly. Hypothesis 6.4 guarantees that the Bellman operator appearing in (6.8) and (6.9) is uniformly elliptic in $\{|x_0| < R\}$. Thus, by [15, Proposition 2.9], for every $p > n$ the function $V^{\bar{x}_1}$ is an L^p-viscosity subsolution of (6.8) and an L^p-viscosity supersolution of (6.9). Now, by [15, Proposition 3.5], the function $V^{\bar{x}_1}$ is twice pointwise differentiable a.e. in $\{|x_0| < R\}$ and by [15, Proposition 3.4] we have that (6.8) and (6.9) are satisfied pointwise a.e. Then, defining the function

$$f(x_0) := \sup_{u\in U}\left[-\frac{1}{2}\mathrm{Tr}(\sigma_0((x_0, \bar{x}_1), u)\sigma_0((x_0, \bar{x}_1), u)^T D^2 V^{\bar{x}_1}(x_0))\right],$$

we have $|f|_{L^\infty(B_R)} \le \bar{C}_R$ (measurability of f is explained, for instance, in [82]). We can then apply [82, Corollary 3] (first such result was stated for L^p-viscosity solutions in [81]) to get that for every $p > n$, the function $V^{\bar{x}_1}$ is an L^p-viscosity solution of

$$\sup_{u\in U}\left[-\frac{1}{2}\mathrm{Tr}(\sigma_0((x_0, \bar{x}_1), u)\sigma_0((x_0, \bar{x}_1), u)^T D^2 v(x_0))\right] = f(x_0) \quad \text{in } \{|x_0| < R\}. \tag{6.10}$$

We now conclude by standard elliptic regularity (see, e.g., [14, Theorem 7.1], together with Remark 1 there, or [81, Theorem 3.1]) that $V^{\bar{x}_1} \in W^{2,p}_{\rm loc}(B_R)$ and $|V^{\bar{x}_1}|_{W^{2,p}(B_{R/2})} \leq C_R$ for some constant C_R. Thus, in particular, by Sobolev embeddings, $V^{\bar{x}_1} \in C^{1,\alpha}_{\rm loc}(B_R)$ for all $0<\alpha<1$.

Step 4. We now prove that $D_{x_0}V$ is continuous in $|\cdot|_{-1}$ norm on bounded sets of X.

Let $|x|_X \leq R$ and assume without loss of generality that $x=(0,x_1)$ and $V(0,x_1)=0, D_{x_0}V(0,x_1)=0$. Suppose by contradiction that there is $\varepsilon>0$ and a sequence $x^N=(x_0^N,x_1^N)\to x=(0,x_1)$ in $|\cdot|_{-1}$ norm such that $(x_0^N,x_1^N)\in B_{2R}$ and $|(D_{x_0}V(x_0^N,x_1^N),0)|_{-1}\geq\varepsilon$. We remind that the $|(\cdot,0)|_{-1}$ and the standard norm in $\mathbb{R}^n$ are equivalent. Let K_R be from Assumption 6.1. Observe that, since $V^{x_1}\in C^{1,\alpha}_{\rm loc}(\mathbb{R}^n)$, for every $y_0\in\mathbb{R}^n$ such that

$$|(y_0,0)|_{-1} \leq (K_{3R}|(0,x_1^N-x_1)|_{-1})^{\frac{1}{1+\alpha}} + 2|(x_0^N,0)|_{-1}, \tag{6.11}$$

we have

$$|V(y_0,x_1)| = |V(y_0,x_1)-V(0,x_1)| \leq C|(y_0,0)|_{-1}^{1+\alpha} \leq C(K_{3R}|(0,x_1^N-x_1)|_{-1} + |(x_0^N,0)|_{-1}^{1+\alpha}).$$

Then

$$|V(y_0,x_1^N)| \leq |V(y_0,x_1)| + K_{3R}|(0,x_1^N-x_1)|_{-1} \leq C(K_{3R}|(0,x_1^N-x_1)|_{-1} + |(x_0^N,0)|_{-1}^{1+\alpha}). \tag{6.12}$$

Now observe that by taking

$$y_0 = x_0^N + \frac{D_{x_0}V(x_0^N,x_1^N)}{|(D_{x_0}V(x_0^N,x_1^N),0)|_{-1}}((K_{3R}|(0,x_1^N-x_1)|_{-1})^{\frac{1}{1+\alpha}} + |(x_0^N,0)|_{-1})$$

we have (6.11) so that (6.12) holds for $|V(y_0,x_1^N)|$. Moreover, note that also x_0^N satisfies (6.11) so that we have (6.12) for $|V(x_0^N,x_1^N)|$. Now, since $V^{x_1^N}\in C^{1,\alpha}_{\rm loc}(\mathbb{R}^n)$, we have

$$\begin{aligned} V(y_0,x_1^N) &\geq V(x_0^N,x_1^N) + D_{x_0}V(x_0^N,x_1^N)\cdot(y_0-x_0^N) - C|y_0-x_0^N|^{1+\alpha} \geq V(x_0^N,x_1^N) \\ &\quad + \frac{1}{|(D_{x_0}V(x_0^N,x_1^N),0)|_{-1}}|D_{x_0}V(x_0^N,x_1^N)|^2(K_{3R}|(0,x_1^N-x_1)|_{-1})^{\frac{1}{1+\alpha}} \\ &\quad + |(x_0^N,0)|_{-1}) - C((K_{3R}|(0,x_1^N-x_1)|_{-1})^{\frac{1}{1+\alpha}} + |(x_0^N,0)|_{-1})^{1+\alpha}. \end{aligned}$$

Therefore, using $|(D_{x_0}V(x_0^N,x_1^N),0)|_{-1}\geq\varepsilon$, the fact that (6.12) holds for $|V(x_0^N,x_1^N)|$, and since $|(\cdot,0)|_{-1}$ is an equivalent norm in $\mathbb{R}^n$, we obtain

$$V(y_0,x_1^N) \geq \eta\varepsilon((K_{3R}|x_1^N-x_1|_{-1})^{\frac{1}{1+\alpha}} + |(x_0^N,0)|_{-1}) - C(K_{3R}|(0,x_1^N-x_1)|_{-1} + |(x_0^N,0)|_{-1}^{1+\alpha}),$$

where $\eta>0$ is a constant. Finally, since (6.12) holds for $|V(y_0,x_1^N)|$, we have

$$\eta\varepsilon((K_{3R}|(0,x_1^N-x_1)|_{-1})^{\frac{1}{1+\alpha}} + |(x_0^N,0)|_{-1}) \leq C(K_{3R}|(0,x_1^N-x_1)|_{-1} + |(x_0^N,0)|_{-1}^{1+\alpha}),$$

which is impossible for large N as $\alpha>0$. This concludes the proof of the theorem. ∎

The regularity result is interesting on its own. It possibly can also be used to define an optimal feedback map under some natural assumptions.

Assume that U is compact and that σ_0 does not depend on u. The Hamiltonian then has the form

$$\begin{aligned}
&\mathcal{H}(x, Dv(x), D^2v(x)) \\
&\quad = H(x, Dv(x)) - x_0 \cdot D_{x_0}v(x) - \frac{1}{2}\mathrm{Tr}\left[\sigma_0(x)\sigma_0(x)^T D^2_{x_0^2}v(x)\right] \\
&\quad = \tilde{H}(x, D_{x_0}v(x)) - x_0 \cdot D_{x_0}v(x) \\
&\qquad - \frac{1}{2}\mathrm{Tr}\left[\sigma_0\left(x_0, \int_{-d}^0 a_2(\xi)x_1(\xi)\,d\xi\right)\sigma_0\left(x_0, \int_{-d}^0 a_2(\xi)x_1(\xi)\,d\xi\right)^T D^2_{x_0^2}v(x)\right],
\end{aligned}$$

where

$$\begin{aligned}
H(x, Dv(x)) &= \tilde{H}(x, D_{x_0}v(x)) \\
&= \max_{u\in U}\left\{ -b_0\left(x_0, \int_{-d}^0 a_1(\xi)x_1(\xi)\,d\xi, u\right)\cdot D_{x_0}v(x) - l(x_0,u)\right\}.
\end{aligned}$$

By Theorem 6.5 we can define a candidate optimal feedback map, i.e.,

$$u^*(x) \in \mathrm{argmax}_{u\in U}\left\{ -b_0\left(x_0, \int_{-d}^0 a_1(\xi)x_1(\xi)\,d\xi, u\right)\cdot D_{x_0}V(x) - l(x,u)\right\}.$$

To show that this may lead to the existence of an optimal feedback control is a difficult problem, which passes through a verification theorem with only partial regularity and the study of the closed loop equation (see, e.g., [36, 44] in the context of the approach via mild solutions or [43] in the case of optimal control of deterministic delay equations). This problem is addressed in [27], under the additional assumption that the value function V is $|\cdot|_{-1}$-semiconvex (that is, there exists $C \geq 0$ such that $V(x) + C|x|^2_{-1}$ is convex).

7. Applications.

In this section we provide two examples of possible applications of our approach. The first arises in finance, the second in marketing.

7.1. Merton-like problem with path dependent coefficients.

We consider a financial market composed by a risk-free asset (bond) B and a risky asset (stock) S. The respective dynamics (deterministic for the bond, stochastic for the stock) are given by

$$\begin{aligned}
&db(t) = rb(t)dt, \\
&ds(t) = \mu\left(\int_{-d}^0 a_1(\xi)s(t+\xi)d\xi\right)s(t)dt + v\left(\int_{-d}^0 a_2(\xi)s(t+\xi)d\xi\right)s(t)dW(t),
\end{aligned}$$

with initial data $s(0) = s_0 > 0, s(\xi) = s_1(\xi) > 0$ for every $\xi \in [-d, 0]$, $b(0) = 1$, and W is a real-valued Brownian motion. Moreover,

- (i) $r \geq 0$;
- (ii) a_1, a_2 are given deterministic functions satisfying the assumptions used in the previous sections;
- (iii) $\mu, v : \mathbb{R} \to \mathbb{R}$ are given Lipschitz continuous functions.

The investor chooses a consumption-investment strategy by deciding at time $t \geq 0$ the fraction $u(t) \in [0, 1]$ of the portfolio $z(t)$ to be invested in the risky stock S; the remaining part $1 - u(t)$ is then invested in the bond B (self-financed portfolio with no borrowing and no short selling constraints); The dynamic of the portfolio (wealth) $z(t)$ is then

$$\begin{aligned}dz(t) &= \frac{ds(t)}{s(t)}u(t)z(t) + \frac{db(t)}{b(t)}(1-u(t))z(t)dt\\ &= \left[rz(t) + \left[\mu\left(\int_{-d}^{0} a_1(\xi)s(t+\xi)d\xi\right) - r\right]u(t)z(t)\right]dt\\ &\quad + \nu\left(\int_{-d}^{0} a_2(\xi)s(t+\xi)d\xi\right)u(t)z(t)dW(t),\end{aligned}$$

with $z(0)=z_0$, where $z_0>0$ is the initial value of the portfolio. The stochastic process $u(\cdot)$ is the control process. We use the same setup of the stochastic optimal control problem as the one in section 2. The control set U is now

$$U=[0,1].$$

The state equation of the optimal control problem can be seen as a controlled SDDE in $\mathbb{R}^2$ for the couple $y(t)=(s(t),z(t))$ in the form (2.1), where, for every $x_0=(s_0,z_0)\in\mathbb{R}^2$, $(s_1,z_1)\in L^2$, $u\in U$,

$$b_0\left(x_0, \int_{-d}^{0} a_1(\xi)x_1(\xi)\,d\xi, u\right) = \begin{bmatrix} \mu\left(\int_{-d}^{0} a_1(\xi)s_1(\xi)d\xi\right)s_0 \\ rz_0 + \left[\mu\left(\int_{-d}^{0} a_1(\xi)s_1(\xi)d\xi\right) - r\right]uz_0, \end{bmatrix},$$

$$\sigma_0\left(x_0, \int_{-d}^{0} a_2(\xi)x_1(\xi)\,d\xi,\, u\right)w = \begin{bmatrix} \nu\left(\int_{-d}^{0} a_2(\xi)s_1(\xi)d\xi\right)s_0 w \\ \nu\left(\int_{-d}^{0} a_2(\xi)s_1(\xi)d\xi\right)uz_0 w \end{bmatrix}.$$

The goal of the investor is to solve the following optimization problem:

$$\sup_{u(\cdot)\in\mathcal{U}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}g(z(t))dt\right]$$

for some concave utility function $g\colon\mathbb{R}\to\mathbb{R}$, where $\rho>0$ is a discount factor. The optimization of this kind of functionals arises in mathematical finance, for example in the context of portfolio optimization with random horizon (see, e.g., [41, section 6.1]) or in the context of pension fund management (see, e.g., [28, 40]).

Note that the maximization problem is equivalent to

$$\inf_{u(\cdot)\in\mathcal{U}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}l(z(t))dt\right],$$

where $l(z)=-g(z)$. By considering the infinite-dimensional framework of section 3, if l (or, equivalently, g) satisfies Assumption 2.5, we can use Theorem 5.4 to characterize the value function V as the unique viscosity solution to (5.5).

7.2. Optimal advertising with delays. The following problem is taken from [53]. In the spirit of the model in [53, section 4] we assume that no delay in the control is present. The model for the dynamics of the stock of advertising goodwill $y(s)$ of the product is given by the following controlled SDDE:

$$\begin{cases} dy(t) = \left[a_0y(t) + \int_{-d}^{0} a_1(\xi)y(t+\xi)\,d\xi + c_0u(t)\right]dt + \sigma_0\,dW(t),\\ y(0)=x_0, \quad y(\xi)=x_1(\xi) \quad \forall\xi\in[-d,0),\end{cases}$$

where $d>0$, the control process $u(s)$ models the intensity of advertising spending and W is a real-valued Brownian motion.

(i) $a_0 \leq 0$ is a constant factor of image deterioration in absence of advertising;
(ii) $c_0 \geq 0$ is a constant advertising effectiveness factor;
(iii) $a_1 \leq 0$ is a given deterministic function satisfying the assumptions used in the previous sections which represents the distribution of the forgetting time;
(iv) $\sigma_0 > 0$ represents the uncertainty in the model;
(v) $x_0 \in \mathbb{R}$ is the level of goodwill at the beginning of the advertising campaign;
(vi) $x_1 \in L^2([-d,0];\mathbb{R})$ is the history of the goodwill level.

Again, we use the same setup of the stochastic optimal control problem as the one in section 2 and the control set U is here

$$U = [0, \bar{u}]$$

for some $\bar{u} > 0$. The optimization problem is

$$\inf_{u \in \mathcal{U}} \mathbb{E}\left[\int_0^\infty e^{-\rho s} l(y(s), u(s)) ds\right],$$

where $\rho > 0$ is a discount factor, $l(x,u) = h(u) - g(x)$, with a continuous and convex cost function $h\colon U \to \mathbb{R}$ and a continuous and concave utility function $g\colon \mathbb{R} \to \mathbb{R}$ which satisfies Assumption 2.5.

Setting

$$b_0\left(x_0, \int_{-d}^0 a_1(\xi) x_1(\xi)\, d\xi, u\right) := a_0 x_0 + \int_{-d}^0 a_1(\xi) x_1(\xi)\, d\xi + c_0 u,$$

we are then in the setting of section 2. Therefore, using the infinite-dimensional framework of section 3, we can use Theorem 5.4 to characterize the value function V as the unique viscosity solution to (5.5), and Theorem 6.5 to obtain partial regularity of V.

Remark 7.1. We point out that, since the structure condition (discussed in the introduction) is satisfied here, this model can be completely solved using the approach via BSDEs or mild solutions and optimal feedback laws can be constructed in these cases. However, with the techniques of the present paper, we could also consider a more general model of the form (2.1) and still apply Theorems 5.4 and 6.5.

Acknowledgments. The authors are grateful to the associate editor and the referee for their careful reading of the manuscript and their helpful comments.

REFERENCES

[1] D. Addona, E. Bandini, and F. Masiero, *A nonlinear Bismut-Elworthy formula for HJB equations with quadratic Hamiltonian in Banach spaces*, Nonlinear Differ. Equ. Appl., 27 (2020), 37.
[2] M. Bambi, G. Fabbri, and F. Gozzi, *Optimal policy and consumption smoothing effects in the time-to-build AK model*, Econom. Theory, 50 (2012), pp. 635–669.
[3] M. Bambi, C. Di Girolami, S. Federico, and F. Gozzi, *Generically distributed investments on flexible projects and endogenous growth*, Econom. Theory, 63 (2017), pp. 521–558.
[4] V. Barbu and G. Da Prato, *Hamilton-Jacobi Equations in Hilbert Spaces*, Res. Notes in Math. 86, Pitman, Boston, 1983.
[5] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Systems Control Found. Appl. 12, Birkhäuser, Boston, 1997.
[6] E. Bayraktar and C. Keller, *Path-dependent Hamilton-Jacobi equations in infinite dimensions*, J. Funct. Anal., 275 (2018), pp. 2096–2161.
[7] E. Bayraktar and C. Keller, *Path-dependent Hamilton–Jacobi equations with super-quadratic growth in the gradient and the vanishing viscosity method*, SIAM J. Control Optim., 60 (2022), pp. 1690–1711, https://doi.org/10.1137/21M1395557.

[8] A. Bensoussan, G. Da Prato, M. C. Delfour, and S. K. Mitter, *Representation and Control of Infinite Dimensional Systems*, 2nd ed., Systems Control Found. Appl., Birkhäuser, Boston, 2007.

[9] A. Bensoussan and P. Yam, *Control problem on space of random variables and master equation*, ESAIM Control Optim. Calc. Var., 25 (2019), 10.

[10] A. Bensoussan, P. Graber, and P. Yam, *Control on Hilbert Spaces and Application to Mean Field Type Control Theory*, preprint, https://arxiv.org/abs/2005.10770, 2020.

[11] E. Biffis, F. Gozzi, and C. Prosdocimi, *Optimal portfolio choice with path dependent labor income: The infinite horizon case*, SIAM J. Control Optim., 58 (2020), pp. 1906–1938, https://doi.org/10.1137/19M1259687.

[12] S. Biagini, F. Gozzi, and M. Zanella, *Robust portfolio choice with sticky wages*, SIAM J. Financial Math., 13 (2022), pp. 1004–1039, https://doi.org/10.1137/21M1429722.

[13] B. Djehiche, F. Gozzi, G. Zanco, and M. Zanella, *Optimal portfolio choice with path dependent benchmarked labor income: A mean field model*, Stochastic Process. Appl., 145 (2022), pp. 48–85.

[14] L. Caffarelli and X. Cabré, *Fully Nonlinear Elliptic Equations*, Amer. Math. Soc. Colloq. Publ. 43, American Mathematical Society, Providence, RI, 1995.

[15] L. Caffarelli, M. G. Crandall, M. Kocan, and A. Święch, *On viscosity solutions of fully nonlinear equations with measurable ingredients*, Comm. Pure Appl. Math., 49 (1996), pp. 365–397.

[16] P. Cannarsa and H. Frankowska, *Value function and optimality conditions for semilinear control problems*, Appl. Math. Optim., 26 (1992), pp. 139–169.

[17] G. Carlier and R. Tahraoui, *Hamilton-Jacobi-Bellman equations for the optimal control of a state equation with memory*, ESAIM Control Optim. Calc. Var., 16 (2010), pp. 744–763.

[18] R. Cont and D. Fournié, *Change of variable formulas for non-anticipative functionals on path space*, J. Funct. Anal., 259 (2010), pp. 1043–1072.

[19] R. Cont and D. Fournié, *Functional Itô calculus and stochastic integral representation of martingales*, Ann. Probab., 41 (2013), pp. 109–133.

[20] A. Chojnowska-Michalik, *Representation theorem for general stochastic delay equations*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys., 26 (1978), pp. 635–642.

[21] A. Cosso, S. Federico, F. Gozzi, M. Rosestolato, and N. Touzi, *Path-dependent equations and viscosity solutions in infinite dimension*, Ann. Probab., 46 (2018), pp. 126–174.

[22] M. G. Crandall and P. L. Lions, *Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions.* IV. *Hamiltonians with unbounded linear terms*, J. Funct. Anal., 90 (1990), pp. 237–283.

[23] M. G. Crandall and P. L. Lions, *Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions.* V. *Unbounded linear terms and B-continuous solutions*, J. Funct. Anal., 97 (1991), pp. 417–465.

[24] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite-Dimensional Systems*, London Math. Soc. Lecture Note Ser. 229, Cambridge University Press, Cambridge, 1996.

[25] G. Da Prato and J. Zabczyk, *Second Order Partial Differential Equations in Hilbert Spaces*, London Math. Soc. Lecture Note Ser. 293, Cambridge University Press, Cambridge, 2002.

[26] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia Math. Appl. 152, Cambridge University Press, Cambridge, 2014.

[27] F. de Feo and A. Święch, *Optimal Control of Stochastic Delay Differential Equations: Optimal Feedback Controls*, preprint, https://arxiv.org/abs/2309.05029, 2023.

[28] M. Di Giacinto, S. Federico, and F. Gozzi, *Pension funds with a minimum guarantee: A stochastic control approach*, Finance Stoch., 15 (2011), pp. 297–342.

[29] I. Elsanosi, B. Øksendal, and A. Sulem, *Some solvable stochastic control problems with delay*, Stochastics Rep., 71 (2000), pp. 69–89.

[30] K. J. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Grad. Texts in Math. 194, Springer, Berlin, 2000.

[31] I. Ekren, C. Keller, N. Touzi, and J. Zhang, *On viscosity solutions of path dependent PDEs*, Ann. Probab., 42 (2014), pp. 204–236.

[32] I. Ekren, N. Touzi, and J. Zhang, *Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part* I, Ann. Probab., 44 (2016), pp. 1212–1253.

[33] I. Ekren, N. Touzi, and J. Zhang, *Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part* II, Ann. Probab., 44 (2016), pp. 2507–2553.

[34] G. Fabbri and F. Gozzi, *Solving optimal growth models with vintage capital: The dynamic programming approach*, J. Econom. Theory, 143 (2008), pp. 331–373.

[35] G. Fabbri, F. Gozzi, and A. Święch, *Verification theorem and construction of ε-optimal controls for control of abstract evolution equations*, J. Convex Anal., 17 (2010), pp. 611–642.

[36] G. Fabbri, F. Gozzi, and A. Święch, *Stochastic Optimal Control in Infinite Dimension. Dynamic Programming and HJB Equations. With a Contribution by Marco Fuhrman and Gianmario Tessitore*, Probab. Theory Stoch. Model. 82, Springer, Cham, 2017.

[37] S. Faggian, *Regular solutions of first-order Hamilton-Jacobi equations for boundary control problems and applications to economics*, Appl. Math. Optim., 51 (2005), pp. 123–162.

[38] S. Faggian, *Hamilton–Jacobi equations arising from boundary control problems with state constraints*, SIAM J. Control Optim., 47 (2008), pp. 2157–2178, https://doi.org/10.1137/070683738.

[39] S. Faggian and F. Gozzi, *Optimal investment models with vintage capital: Dynamic programming approach*, J. Math. Econom., 46 (2010), pp. 416–437.

[40] S. Federico, *A stochastic control problem with delay arising in a pension fund model*, Finance Stoch., 15 (2011), pp. 421–459.

[41] S. Federico, P. Gassiat, and F. Gozzi, *Utility maximization with current utility on the wealth: Regularity of solutions to the HJB equation*, Finance Stoch., 19 (2015), pp. 415–448.

[42] S. Federico, B. Goldys, and F. Gozzi, *HJB equations for the optimal control of differential equations with delays and state constraints,* I: *Regularity of viscosity solutions*, SIAM J. Control Optim., 48 (2010), pp. 4910–4937, https://doi.org/10.1137/09076742X.

[43] S. Federico, B. Goldys, and F. Gozzi, *HJB equations for the optimal control of differential equations with delays and state constraints,* II: *Verification and optimal feedbacks*, SIAM J. Control Optim., 49 (2011), pp. 2378–2414, https://doi.org/10.1137/100804292.

[44] S. Federico and F. Gozzi, *Verification theorems for stochastic optimal control problems in Hilbert spaces by means of a generalized Dynkin formula*, Ann. Appl. Probab., 28 (2018), pp. 3558–3599.

[45] S. Federico and E. Tacconi, *Dynamic programming for optimal control problems with delays in the control variable*, SIAM J. Control Optim., 52 (2014), pp. 1203–1236, https://doi.org/10.1137/110840649.

[46] S. Federico and P. Tankov, *Finite-dimensional representations for controlled diffusions with delay*, Appl. Math. Optim., 71 (2015), pp. 165–194.

[47] F. Flandoli and G. Zanco, *An infinite-dimensional approach to path-dependent Kolmogorov's equations*, Ann. Probab., 44 (2016), pp. 2643–2693.

[48] M. Fuhrman, F. Masiero, and G. Tessitore, *Stochastic equations with delay: Optimal control via BSDEs and regular solutions of Hamilton–Jacobi–Bellman equations*, SIAM J. Control Optim., 48 (2010), pp. 4624–4651, https://doi.org/10.1137/080730354.

[49] 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.

[50] M. Fuhrman and G. Tessitore, *Infinite horizon backward stochastic differential equations and elliptic equations in Hilbert spaces*, Ann. Probab., 32 (2004), pp. 607–660.

[51] W. Gangbo and A. Mészáros, *Global well-posedness of master equations for deterministic displacement convex potential mean field games*, Comm. Pure Appl. Math., 75 (2022), pp. 2685–2801.

[52] B. Goldys and F. Gozzi, *Second order parabolic Hamilton-Jacobi-Bellman equations in Hilbert spaces and stochastic control: L^2_μ approach*, Stochastic Process. Appl., 116 (2006), pp. 1932–1963.

[53] F. Gozzi and C. Marinelli, *Stochastic optimal control of delay equations arising in advertising models*, in Stochastic Partial Differential Equations and Applications-VII, Lect. Notes Pure Appl. Math. 245, Chapman & Hall/CRC, Boca Raton, FL, 2006, pp. 133–148.

[54] F. Gozzi, C. Marinelli, and S. Savin, *On controlled linear diffusions with delay in a model of optimal advertising under uncertainty with memory effects*, J. Optim. Theory Appl., 142 (2009), pp. 291–321.

[55] F. Gozzi and F. Masiero, *Stochastic optimal control with delay in the control* I: *Solving the HJB equation through partial smoothing*, SIAM J. Control Optim., 55 (2017), pp. 2981–3012, https://doi.org/10.1137/16M1070128.

[56] F. Gozzi and F. Masiero, *Stochastic optimal control with delay in the control* II: *Verification theorem and optimal feedbacks*, SIAM J. Control Optim., 55 (2017), pp. 3013–3038, https://doi.org/10.1137/16M1073637.

[57] F. Gozzi and F. Masiero, *Errata: Stochastic optimal control with delay in the control* I: *Solving the HJB equation through partial smoothing, and stochastic optimal control with delay in the control* II: *Verification theorem and optimal feedbacks*, SIAM J. Control Optim., 59 (2021), pp. 3096–3101, https://doi.org/10.1137/21M1407434.

[58] F. Gozzi, A. Święch, and X. Y. Zhou, *A corrected proof of the stochastic verification theorem within the framework of viscosity solutions*, SIAM J. Control Optim., 43 (2005), pp. 2009–2019, https://doi.org/10.1137/S0363012903428184.

[59] F. Gozzi, A. Święch, and X. Y. Zhou, *Erratum: "A corrected proof of the stochastic verification theorem within the framework of viscosity solutions"*, SIAM J. Control Optim., 48 (2010), pp. 4177–4179, https://doi.org/10.1137/090775567.

[60] J. Guyon and J. Lekeufack, *Volatility is (mostly) path-dependent*, Quant. Finance, 23 (2023), pp. 1221–1258, https://doi.org/10.1080/14697688.2023.2221281.

[61] B. Larssen, *Dynamic programming in stochastic control of systems with delay*, Stoch. Stoch. Rep., 74 (2002), pp. 651–673.

[62] B. Larssen and N. H. Risebro, *When are HJB-equations in stochastic control of delay systems finite dimensional?*, Stochastic Anal. Appl., 21 (2003), pp. 643–671.

[63] X. J. Li and J. M. Yong, *Optimal Control Theory for Infinite-Dimensional Systems*, Systems and Control: Foundations and Applications, Birkhäuser, Boston, 1995.

[64] P. L. Lions, *Viscosity solutions of fully nonlinear second-order equations and optimal stochastic control in infinite dimensions.* I. *The case of bounded stochastic evolutions*, Acta Math., 161 (1988), pp. 243–278.

[65] F. Masiero, *Stochastic optimal control problems and parabolic equations in Banach spaces*, SIAM J. Control Optim., 47 (2008), pp. 251–300, https://doi.org/10.1137/050632725.

[66] F. Masiero and G. Tessitore, *Partial smoothing of delay transition semigroups acting on special functions*, J. Differential Equations, 316 (2022), pp. 599–640.

[67] S. Mayorga and A. Święch, *Finite dimensional approximations of Hamilton–Jacobi–Bellman equations for stochastic particle systems with common noise*, SIAM J. Control Optim., 61 (2023), pp. 820–851, https://doi.org/10.1137/22M1489186.

[68] R. Merton, *Lifetime portfolio selection under uncertainty: The continuous-time case*, Rev. Econom. Statist., 51 (1969), pp. 247–257.

[69] S. E. A. Mohammed, *Stochastic Functional Differential Equations*, Res. Notes in Math. 99, Pitman, Boston, 1984.

[70] S. E. A. Mohammed, *Stochastic differential systems with memory: Theory, examples and applications*, in Stochastic Analysis and Related Topics VI, Progr. Probab. 42, L. Decreusefond, J. Gjerde, B. Øksendal, and A. S. Üstünel, eds., Birkhäuser, Boston, 1998, pp. 1–77.

[71] M. Nerlove and J. K. Arrow, *Optimal advertising policy under dynamic conditions*, Economica, 29 (1962), pp. 129–142.

[72] T. Pang Tao and Y. Yong, *A new stochastic model for stock price with delay effects*, in Proceedings of the Conference on Control and its Applications, Society for Industrial and Applied Mathematics, Philadelphia, 2019, pp. 110–117, https://doi.org/10.1137/1.9781611975758.17.

[73] É. Pardoux and S. G. Peng, *Adapted solution of a backward stochastic differential equation*, Systems Control Lett., 14 (1990), pp. 55–61.

[74] H. Pham, *Continuous-Time Stochastic Control and Optimization with Financial Applications*, Stoch. Model. Appl. Probab. 61, Springer, Berlin, 2009.

[75] P. E. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[76] Z. Ren, N. Touzi, and J. Zhang, *Comparison of viscosity solutions of fully nonlinear degenerate parabolic path-dependent PDEs*, SIAM J. Math. Anal., 49 (2017), pp. 4093–4116, https://doi.org/10.1137/16M1090338.

[77] Z. Ren and M. Rosestolato, *Viscosity solutions of path-dependent PDEs with randomized time*, SIAM J. Math. Anal., 52 (2020), pp. 1943–1979, https://doi.org/10.1137/18M122666X.

[78] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd ed., Grundlehren Math. Wiss. 293, Springer, Berlin, 1999.

[79] M. Rosestolato and A. Święch, *Partial regularity of viscosity solutions for a class of Kolmogorov equations arising from mathematical finance*, J. Differential Equations, 262 (2017), pp. 1897–1930.

[80] W. Stannat and L. Wessels, *Necessary and Sufficient Conditions for Optimal Control of Semilinear Stochastic Partial Differential Equations*, preprint, https://arxiv.org/abs/2112.09639, 2021.

[81] A. Święch, *$W^{1,p}$-interior estimates for solutions of fully nonlinear, uniformly elliptic equations*, Adv. Differential Equations, 2 (1997), pp. 1005–1027.

[82] A. Święch, *Pointwise properties of L^p-viscosity solutions of uniformly elliptic equations with quadratically growing gradient terms*, Discrete Contin. Dyn. Syst., 40 (2020), pp. 2945–2962.
[83] R. B. Vinter and R. H. Kwong, *The infinite time quadratic control problem for linear systems with state and control delays: An evolution equation approach*, SIAM J. Control Optim., 19 (1981), pp. 139–153, https://doi.org/10.1137/0319011.
[84] J. Yong and X. Y. Zhou, *Stochastic Controls, Hamiltonian Systems and HJB Equations*, Appl. Math. 43, Springer, New York, 1999.
[85] J. Zhou, *A class of infinite-horizon stochastic delay optimal control problems and a viscosity solution to the associated HJB equation*, ESAIM Control Optim. Calc. Var., 24 (2018), pp. 639–676.
[86] J. Zhou, *Delay optimal control and viscosity solutions to associated Hamilton-Jacobi-Bellman equations*, Internat. J. Control, 92 (2019), pp. 2263–2273.
[87] J. Zhou, *A notion of viscosity solutions to second-order Hamilton-Jacobi-Bellman equations with delays*, Internat. J. Control, 95 (2022), pp. 2611–2631.