Optimal control of path-dependent McKean-Vlasov SDEs in infinite dimension

Andrea COSSO[*]   Fausto GOZZI[†]   Idris KHARROUBI[‡]
Huyên PHAM[§]   Mauro ROSESTOLATO [¶]

Abstract

We study the optimal control of path-dependent McKean-Vlasov equations valued in Hilbert spaces motivated by non Markovian mean-field models driven by stochastic PDEs. We first establish the well-posedness of the state equation, and then we prove the dynamic programming principle (DPP) in such a general framework. The crucial law invariance property of the value function V is rigorously obtained, which means that V can be viewed as a function on the Wasserstein space of probability measures on the set of continuous functions valued in Hilbert space. We then define a notion of pathwise measure derivative, which extends the Wasserstein derivative due to Lions [62], and prove a related functional Itô formula in the spirit of Dupire [33] and Wu and Zhang [73]. The Master Bellman equation is derived from the DPP by means of a suitable notion of viscosity solution. We provide different formulations and simplifications of such a Bellman equation notably in the special case when there is no dependence on the law of the control.

Mathematics Subject Classification (2010): 93E20, 60K35, 49L25.

Keywords: Path-dependent McKean-Vlasov SDEs in Hilbert space, dynamic programming principle, pathwise measure derivative, functional Itô calculus, Master Bellman equation, viscosity solutions.

---

[*]University of Bologna, Italy; andrea.cosso@unibo.it

[†]Luiss University, Roma, Italy; fgozzi@luiss.it

[‡]LPSM, UMR CNRS 8001, Sorbonne University and Université de Paris; idris.kharroubi@upmc.fr

[§]LPSM, UMR CNRS 8001, Sorbonne University and Université de Paris; pham@lpsm.paris

[¶]Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, 73100 Lecce, Italy; mauro.rosestolato@unisalento.it.

Contents

1 Introduction

Given two real separable Hilbert spaces H and K, let us consider the nonlinear stochastic differential equation (SDE) on H in the form:

$$dX_t = AX_t dt + b_t(X, \mathbb{P}_{X_{\cdot \wedge t}}, \alpha_t, \mathbb{P}_{\alpha_t}) dt + \sigma_t(X, \mathbb{P}_{X_{\cdot \wedge t}}, \alpha_t, \mathbb{P}_{\alpha_t}) dB_t, \tag{1.1}$$

over a finite interval $[0, T]$. Here, $A \colon \mathcal{D}(A) \subset H \to H$ is the generator of a C_0-semigroup of contractions in H, and $B = (B_t)_{t \geq 0}$ is a K-valued cylindrical Brownian motion on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{F}^B$ its completed natural filtration. The coefficients b and σ, valued respectively in H and $\mathcal{L}_2(K; H)$ (the space of Hilbert-Schmidt operators from K to H), depend on time, on the whole path of the state process X, on an input control process α, that is an $\mathbb{F}^B$-adapted process valued in some Borel space U, and furthermore on the distribution of the state/control process.

Equation (1.1) is referred to as controlled McKean-Vlasov SDE in Hilbert spaces, and we are interested in the optimal control for (1.1) by minimizing, over control processes α, a functional in the form

$$J(X_0, \alpha) = \mathbb{E}\Big[\int_0^T f_t(X, \mathbb{P}_{X_{\cdot \wedge t}}, \alpha_t, \mathbb{P}_{\alpha_t}) dt + g(X, \mathbb{P}_X) \Big],$$

given running cost and terminal cost functions f and g.

When the coefficients b, σ, f, g do not depend on the law of the state process, the control of equation (1.1) is motivated by various kinds of stochastic partial differential equations (SPDEs) like stochastic heat equations (see e.g. [20, 48, 49, 64, 44]) stochastic reaction-diffusion equations (see e.g.[25, 26]), stochastic porous media equations (see e.g. [7]), singular stochastic dissipative equations (see e.g. [68]), stochastic Burgers and Navier-Stokes equations (see e.g. [28, 67, 53]), Zakai equation in filtering (see e.g. [54]), stochastic first-order equations (see e.g. [47]), stochastic delay equations (see e.g. [43][69], [51, 52, 55] [14]). We refer also to the lecture notes and monographs [29], [6], [35], for an account on this topic.

The main novelty of this paper is to consider a mean-field dependence on the coefficients of the infinite-dimensional stochastic differential equation (1.1), and to study the corresponding control problem. Mean-field diffusion processes, also called McKean-Vlasov equations, in finite dimension have a long history with the pioneering works [58], [65], and later on with the seminal paper [70] in the framework of propagation of chaos. The control of such equations has attracted an increasing interest since the emergence of mean-field game theory initiated independently in [61] and [56], aiming at describing control of large systems or population of interacting particles, and has generated over the last few years numerous contributions, see e.g. [66], [27], [32], and the reference monographs [10] and [23].

In addition to the infinite-dimensional feature of the McKean-Vlasov equation (1.1), we emphasize the path-dependency (in a nonanticipative way) of the coefficients b, σ, f, g, on the state process as well as on its distribution. This general setting, which is motivated by various applications, see e.g. Example 2.12, seems to be considered for the first time in the present paper. It is worth noting that, also in the mean-field game theory, some research papers (see [24, 41]) started to look at the cases when the state of the agents follows a delay equation. As far as we know, only cases with explicit solution are studied up to now, and we think it would be worth to build a general theory in such cases, on the line of what we do here for McKean-Vlasov control problems.

Our basic objective here is to extend to our infinite-dimensional path-dependent setting the tools required in the dynamic programming approach for McKean-Vlasov control problems.

In the finite-dimensional case, i.e. $H = \mathbb{R}^d$, and in the Markovian case, i.e. without path-dependency of the coefficients, the Wasserstein derivative in the lifted sense of Lions [62], turns out to be a convenient notion of measure derivative when combined with Itô's formula along the flow of probability measures (see [18]) in order to define the Master equation in mean-field game/control. These concepts have been recently extended to the path-dependent case in [73] with a functional Itô formula in the McKean-Vlasov setting.

Our contributions. Our first main result is to prove the crucial law invariance property of the value function to the control problem (see Theorem 3.6), which implies that the value function can be considered as a function on the Wasserstein space of probability measures on $C([0,T]; H)$. We also state and provide a direct proof of the dynamic programming principle in this context (see Theorem 3.4 and Corollary 3.9). Next, we introduce a notion of pathwise derivative in Wasserstein space and a related functional Itô formula in our infinite-dimensional McKean-Vlasov context that extend the concepts in [73]. Equipped with these tools, we can then derive from the dynamic programming principle the associated Master HJB equation, which is a PDE where the state variable is a probability measure on $C([0,T]; H)$. For such PDE, we provide equivalent formulations and simplifications, notably in the special case when there is no dependence of the coefficients on the law of the control. We define an intrinsic notion of viscosity solution in $\mathcal{P}_2(C([0,T]; H))$, the space of square-integrable probability measures on $C([0,T]; H)$, together with the viscosity property of the value function. Comparison principle for Master Bellman equation is postponed to further investigation, as it is already a challenging issue in the finite-dimensional case where only partial results exist in the literature, see [73] and [19].

We also point out that our results clarify and improve in particular some statements from the finite-dimensional case, like the law invariance property (see Remark 3.8) and the dynamic programming principle.

The outline of the paper is organized as follows. In Section 2, we present the notations and formulate the McKean-Vlasov state equation valued in Hilbert space: due to the generality of the setting basic results on well-posedness and approximation of this equation are not known and are carefully proved. Section 3 is devoted to the formulation of the optimal control problem, the dynamic programming principle and law invariance property of the associated value function. We introduce in Section 4 the notion of pathwise derivative in Wasserstein space and the related functional Itô formula. Section 5 is concerned with the derivation of the Master Bellman equation and the viscosity property of the value function. Finally, Appendices A, B, C, D, E, F collect some technical results used throughout the paper.

2 Controlled path-dependent McKean-Vlasov SDEs in Hilbert spaces

2.1 Notations and assumptions

State space and functional analytic setting. We fix two real separable Hilbert spaces H and K, with inner products $\langle \cdot, \cdot \rangle_H, \langle \cdot, \cdot \rangle_K$ and induced norms $|\cdot|_H, |\cdot|_K$, respectively, omitting the subscripts H or K when clear from the context. We denote by $\mathcal{L}(K; H)$ (resp. $\mathcal{L}(H)$) the space of bounded linear operators from K to H (resp. H to H). We endow $\mathcal{L}(K; H)$ with the operator norm

$\| \cdot \|_{\mathcal{L}(K;H)}$ defined by

$$\|F\|_{\mathcal{L}(K;H)} = \sup_{k \in K, \, k \neq 0} \frac{|Fk|_H}{|k|_K}, \qquad F \in \mathcal{L}(K;H).$$

Similarly, we endow $\mathcal{L}(H)$ with the corresponding norm $\| \cdot \|_{\mathcal{L}(H)}$. We also denote by $\mathcal{L}_2(K;H)$ the space of Hilbert-Schmidt operators from K to H, that is the set of all $F \in \mathcal{L}(K;H)$ such that

$$\sum_{n \in \mathbb{N}} |Fe_n|_H^2 < +\infty$$

for some orthonormal basis $\{e_n\}_{n \in \mathbb{N}}$ of K. We endow $\mathcal{L}_2(K;H)$ with the norm $\| \cdot \|_{\mathcal{L}_2(K;H)}$ defined by

$$\|F\|_{\mathcal{L}_2(K;H)} = \sqrt{\sum_{n \in \mathbb{N}} |Fe_n|_H^2}, \qquad F \in \mathcal{L}_2(K;H).$$

We recall that the definitions of $\mathcal{L}_2(K;H)$ and $\| \cdot \|_{\mathcal{L}_2(K;H)}$ do not depend on the choice of the orthonormal basis $\{e_n\}_{n \in \mathbb{N}}$ of K.

We now fix a finite time horizon $T > 0$ and consider the *state space* of our optimal control problem which is given by the set $C([0,T];H)$ of continuous H-valued functions on $[0,T]$. Given $x \in C([0,T];H)$ and $t \in [0,T]$, we denote by x_t the value of x at time t and we set $x_{\cdot \wedge t} := (x_{s \wedge t})_{s \in [0,T]}$. Notice that $x_t \in H$, while $x_{\cdot \wedge t} \in C([0,T];H)$. We endow $C([0,T];H)$ with the uniform norm $\| \cdot \|_T$ defined as

$$\|x\|_T = \sup_{s \in [0,T]} |x_s|_H, \qquad x \in C([0,T];H).$$

Notice that $(C([0,T];H), \| \cdot \|_T)$ is a Banach space. We denote by $\mathscr{B}$ the Borel σ-algebra of $C([0,T];H)$. Finally, for every $t \in [0,T]$ we introduce the seminorm $\| \cdot \|_t$ defined as

$$\|x\|_t = \|x_{\cdot \wedge t}\|_T, \qquad x \in C([0,T];H).$$

Spaces of probability measures and Wasserstein distance. Given a metric space M, if $\mathscr{M}$ denotes its Borel σ-algebra, we denote by $\mathcal{P}(M)$ the set of all probability measures on $(M, \mathscr{M})$. We endow $\mathcal{P}(M)$ with the topology of weak convergence. When M is a Polish space S, with metric d_S, we also define, for $q \geq 1$,

$$\mathcal{P}_q(S) := \left\{ \mu \in \mathcal{P}(S) \colon \int_S d_S(x_0, x)^q \mu(dx) < +\infty \right\},$$

where $x_0 \in S$ is arbitrary. This set is endowed with the q-Wasserstein distance defined as

$$\mathcal{W}_q(\mu, \mu') := \inf \left\{ \int_{S \times S} d_S(x, y)^q \, \pi(dx, dy) \colon \pi \in \mathcal{P}(S \times S) \right.$$

$$\left. \text{such that } \pi(\cdot \times S) = \mu \text{ and } \pi(S \times \cdot) = \mu' \right\}^{\frac{1}{q}}, \qquad q \geq 1,$$

for every $\mu, \mu' \in \mathcal{P}_q(S)$. The space $\left(\mathcal{P}_q(S), \mathcal{W}_q\right)$ turns out to be a Polish space (see for instance [71, Theorem 6.18]).

5

Probabilistic setting. We fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a K-valued cylindrical Brownian motion $B = (B_t)_{t \geq 0}$ is defined (see e.g. [29, Section 4.1] and [35, Remark 1.89] on the definition of cylindrical Brownian motion). We denote by $\mathbb{F}^B = (\mathcal{F}_t^B)_{t \geq 0}$ the $\mathbb{P}$-completion of the filtration generated by B[1]. Notice that $\mathbb{F}^B$ is also right-continuous (see [35, Lemma 1.94]), so, in particular, it satisfies the usual conditions. We assume that there exists a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$ satisfying the following standing assumptions.

Standing Assumption ($\mathbf{A}_\mathcal{G}$).

i) $\mathcal{G}$ and $\mathcal{F}_\infty^B$ are independent;

ii) $\mathcal{G}$ is "rich enough" in the sense that the following property holds:

$$\mathcal{P}_2\big(C([0,T];H)\big) = \big\{\mathbb{P}_\xi \text{ with } \xi\colon [0,T] \times \Omega \to H \text{ continuous and}$$
$$\mathcal{B}([0,T]) \otimes \mathcal{G}\text{-measurable process satisfying } \mathbb{E}\big[\|\xi\|_T^2\big] < \infty\big\},$$

i.e. for every $\mu \in \mathcal{P}_2(C([0,T];H))$ there exists a continuous and $\mathcal{B}([0,T]) \otimes \mathcal{G}$-measurable process $\xi\colon [0,T] \times \Omega \to H$, satisfying $\mathbb{E}\|\xi\|_T^2 < \infty$, such that ξ has law equal to μ.

As stated in the following lemma (take $\mathcal{H} = \mathcal{G}$ in Lemma 2.1), property $(\mathbf{A}_\mathcal{G})$-ii) holds if and only if there exists a $\mathcal{G}$-measurable random variable $U_\mathcal{G}\colon \Omega \to \mathbb{R}$ having uniform distribution on $[0,1]$ (see also Remark 2.2).

Lemma 2.1. *On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consider a sub-σ-algebra $\mathcal{H} \subset \mathcal{F}$. The following statements are equivalent.*

1) *There exists a $\mathcal{H}$-measurable random variable $U_\mathcal{H}\colon \Omega \to \mathbb{R}$ having uniform distribution on $[0,1]$.*

2) *$\mathcal{H}$ is "rich enough" in the sense that the following property holds:*

$$\mathcal{P}_2\big(C([0,T];H)\big) = \big\{\mathbb{P}_\xi \text{ with } \xi\colon [0,T] \times \Omega \to H \text{ continuous and}$$
$$\mathcal{B}([0,T]) \otimes \mathcal{H}\text{-measurable process satisfying } \mathbb{E}\big[\|\xi\|_T^2\big] < \infty\big\}.$$

Remark 2.2. *Using the same notations as in Lemma 2.1, if the probability space $(\Omega, \mathcal{H}, \mathbb{P})$ is atomless (namely, for any $E \in \mathcal{H}$ such that $\mathbb{P}(E) > 0$ there exists $F \in \mathcal{H}$, $F \subset E$, such that $0 < \mathbb{P}(F) < \mathbb{P}(E)$) then property 1), or equivalently 2), of Lemma 2.1 holds (see for instance [23, vol. I, p. 352]).*

Remark 2.3. *The additional randomness (other than B) coming from the σ-algebra $\mathcal{G}$ is used for the initial condition ξ of the state equation (2.4) (notice that it is necessary to consider random initial conditions in order to state and prove the dynamic programming principle, Theorem 3.4, where for instance the initial condition at time s is given by the random variable $X^{t,\xi,\alpha}$). However, we remark that whenever $t > 0$ the σ-algebra $\mathcal{G}$ can be replaced by $\mathcal{F}_t^B$; in other words, the initial condition ξ can be taken only $\mathcal{F}_t^B$-measurable. As a matter of fact, for every $t > 0$, it holds that:*

---

[1]Notice that it may be not obvious to define the natural filtration for cylindrical Brownian motion as, in principle, it may depend on the choice of the reference system where such process is considered, which, in general, is not unique. However, as noted in [35, Remark 1.89]) this will not affect the class of integrable processes and, consequently, the filtration.

i) $\mathcal{F}_t^B$ and $\sigma(B_s - B_t,\ s \geq t)$ are independent (in item i) of $(\mathbf{A}_\mathcal{G})$ we have imposed the stronger condition that $\mathcal{G}$ and $\mathcal{F}_\infty^B$ have to be independent; however, if we consider only the control problem with initial time t, then the assumption imposed here is enough.

ii) $\mathcal{F}_t^B$ satisfies the property of being "rich enough" or, equivalently, it satisfies property 1) of Lemma 2.1.

Therefore, the σ-algebra $\mathcal{G}$ is really necessary only for the control problem with initial time $t = 0$. In fact, this allows to define the value function v (see (3.10)) for every pair (t, μ) in $[0, T] \times \mathcal{P}_2(C([0,T]; H))$, with μ being the law of ξ. On the other hand, if we do not use $\mathcal{G}$, v is defined on $(0, T] \times \mathcal{P}_2(C([0,T]; H))$ and for $t = 0$ is defined only at the Dirac measures.

Proof of Lemma 2.1. 1) $\implies$ 2). Fix $\mu \in \mathcal{P}_2(C([0,T]; H))$. Our aim is to find a process $\xi \colon [0, T] \times \Omega \to H$ continuous and $\mathcal{B}([0,T]) \otimes \mathcal{H}$-measurable with law equal to μ. To this end, consider the probability space $([0, 1], \mathcal{B}([0, 1]), \lambda)$, where λ denotes the Lebesgue measure on the unit interval. Given such a μ, it follows from Theorem 3.19 in [59] that there exists a measurable function $\Xi \colon [0, 1] \to C([0, T]; H)$ such that the image (or push forward) measure of λ by Ξ is equal to μ. Now, denote

$$\xi_t(\omega) := \Xi(U_\mathcal{H}(\omega))_t, \qquad \forall\ (t, \omega) \in [0, T] \times \Omega,$$

where the subscript in $\Xi(U_\mathcal{H}(\omega))_t$ denotes the valuation at time t of the continuous function $\Xi(U_\mathcal{H}(\omega))$.

Notice that ξ is a continuous process with law equal to μ. Moreover, for every fixed $t \in [0, T]$, ξ_t is $\mathcal{H}$-measurable. Since ξ has continuous paths, it follows that ξ is also $\mathcal{B}([0, T]) \otimes \mathcal{H}$-measurable (see for instance [30], Chapter IV, Theorem 15). This concludes the proof of the implication 1) $\implies$ 2).

2) $\implies$ 1). The claim follows from Lemma F.1. $\qquad \square$

We denote by $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ the filtration defined as

$$\mathcal{F}_t = \mathcal{G} \vee \mathcal{F}_t^B, \qquad t \geq 0.$$

Notice that $\mathbb{F}$ satisfies the usual conditions of completeness and right-continuity. We then denote by $\mathbf{S}_2(\mathbb{F})$ (resp. $\mathbf{S}_2(\mathcal{G})$) the set of H-valued continuous $\mathbb{F}$-progressively measurable (resp. $\mathcal{B}([0,T]) \otimes \mathcal{G}$-measurable) processes ξ such that

$$\|\xi\|_{\mathbf{S}_2} := \mathbb{E}\big[\|\xi\|_T^2\big]^{\frac{1}{2}} < \infty.$$

Control processes. The *space of control actions*, denoted by U, satisfies the following standing assumption.

Standing Assumption $(\mathbf{A}_\mathrm{U})$. U *is a Borel space (see for instance Definition 7.7 in [11]), namely a Borel subset of some Polish space E. $\mathscr{U}$ denotes its Borel σ-algebra.*

Remark 2.4. *Our Assumption $(\mathbf{A}_\mathrm{U})$ is quite general. Indeed in most applications it is enough that U is a Polish space or even a Hilbert space, as in the examples of Example 2.12.*

Finally, we denote by $\mathcal{U}$ the *space of control processes*, namely the family of all $\mathbb{F}$-progressively measurable processes $\alpha \colon [0, T] \times \Omega \to \mathrm{U}$.

Assumptions on the coefficients of the state equation. We consider a linear, possibly unbounded, operator $A : \mathcal{D}(A) \subset H \to H$ and two functions

$$b, \, \sigma : \, [0,T] \times C([0,T];H) \times \mathcal{P}_2\big(C([0,T];H)\big) \times \mathrm{U} \times \mathcal{P}(\mathrm{U}) \longrightarrow \, H, \, \mathcal{L}_2(K;H),$$

where we recall that $\mathcal{P}(\mathrm{U})$ is endowed with the topology of weak convergence. We impose the following assumptions on A, b, σ.

Assumption $(\mathbf{A}_{A,b,\sigma})$.

(i) *A generates a C_0-semigroup of pseudo-contractions $\{e^{tA}, \, t \geq 0\}$ in H. Hence, there exists and $\eta \in \mathbb{R}$ such that*

$$\|e^{tA}\|_{\mathcal{L}(H)} \leq e^{\eta t}. \tag{2.1}$$

(ii) *The functions b and σ are measurable.*

(iii) *There exists a constant L such that*

$$|b_t(x,\mu,u,\nu) - b_t(x',\mu',u,\nu)|_H \leq L\big(\|x - x'\|_t + \mathcal{W}_2(\mu,\mu')\big),$$
$$\|\sigma_t(x,\mu,u,\nu) - \sigma_t(x',\mu',u,\nu)\|_{\mathcal{L}_2(K;H)} \leq L\big(\|x - x'\|_t + \mathcal{W}_2(\mu,\mu')\big),$$
$$|b_t(0,\delta_0,u,\nu)|_H + \|\sigma_t(0,\delta_0,u,\nu)\|_{\mathcal{L}_2(K;H)} \leq L,$$

for all $(t,u,\nu) \in [0,T] \times \mathrm{U} \times \mathcal{P}(\mathrm{U})$, $(x,\mu),(x',\mu') \in C([0,T];H) \times \mathcal{P}_2(C([0,T];H))$, with δ_0 being the Dirac measure at 0, namely the probability measure on $C([0,T];H)$ putting mass equal to 1 to the constant path 0.

Remark 2.5. *Notice that, from the Lipschitz property of b and σ with respect to the variable $x \in C([0,T];H)$, it follows that b and σ satisfies the following non-anticipativity property:*

$$b_t(x,\mu,u,\nu) = b_t(x_{\cdot \wedge t},\mu,u,\nu), \qquad \sigma_t(x,\mu,u,\nu) = \sigma_t(x_{\cdot \wedge t},\mu,u,\nu),$$

for every $(t,x,\mu,u,\nu) \in [0,T] \times C([0,T];H) \times \mathcal{P}_2(C([0,T];H)) \times \mathrm{U} \times \mathcal{P}(\mathrm{U})$.

Remark 2.6. *The optimal control problem of McKean-Vlasov SDEs is sometimes called extended or generalized (see [1, 27]) when the coefficients also depend on the law of the control process $\mathbb{P}_{\alpha_s}$ (as in the present framework, see equation (2.4) below) or, more generally, on the joint law $\mathbb{P}_{(X_{\cdot \wedge s},\alpha_s)}$. Notice however that, under the assumptions below, the latter case is only apparently more general than our framework. As a matter of fact, consider for simplicity only the drift coefficient b and suppose that it is replaced by a function $\bar{b}$ from $[0,T] \times C([0,T];H) \times \mathrm{U} \times \mathcal{P}(C([0,T];H) \times \mathrm{U})$ to H. Recall from $(\mathbf{A}_{A,b,\sigma})$ that the Lipschitz continuity of b with respect to the law of the path reads as*

$$|b_t(x,\mu,u,\nu) - b_t(x,\mu',u,\nu)|_H \leq L\mathcal{W}_2(\mu,\mu'), \tag{2.2}$$

for all $(t,x,u,\nu) \in [0,T] \times C([0,T];H) \times \mathrm{U} \times \mathcal{P}(\mathrm{U})$, $\mu,\mu' \in \mathcal{P}_2(C([0,T];H))$. In other words, the Lipschitz continuity is imposed on the law of the state variable, not on the law of the control variable. If we do the same for the coefficient $\bar{b}$, we get the following ($\pi(\cdot \times \mathrm{U})$ is the law of the state variable, while $\pi(C([0,T];H) \times \cdot)$ is the law of the control variable):

$$|\bar{b}_t(x,u,\pi) - \bar{b}_t(x,u,\pi')|_H \leq L\mathcal{W}_2\big(\pi(\cdot \times \mathrm{U}), \pi'(\cdot \times \mathrm{U})\big), \tag{2.3}$$

for every $(t, x, u) \in [0, T] \times C([0, T]; H) \times \mathrm{U}$, $\pi, \pi' \in \mathcal{P}(C([0, T]; H) \times \mathrm{U})$ *with* $\pi(\cdot \times \mathrm{U}), \pi'(\cdot \times \mathrm{U}) \in \mathcal{P}_2(C([0, T]; H))$ *and* $\pi(C([0, T]; H) \times \cdot) = \pi'(C([0, T]; H) \times \cdot)$. *We require that* $\pi(C([0, T]; H) \times \cdot) = \pi'(C([0, T]; H) \times \cdot)$ *since, as in* (2.2), *the law of the control variable is fixed. More precisely, using the notation of* (2.2),

$$\mu = \pi(\cdot \times \mathrm{U}), \qquad \mu' = \pi'(\cdot \times \mathrm{U}), \qquad \nu = \pi(C([0, T]; H) \times \cdot) = \pi'(C([0, T]; H) \times \cdot).$$

It follows directly from assumption (2.3) *that* $\bar{b}_t(x, u, \pi) = \bar{b}_t(x, u, \pi')$ *whenever the marginals of* π *and* π' *coincide:* $\pi(\cdot \times \mathrm{U}) = \pi'(\cdot \times \mathrm{U})$ *and* $\pi(C([0, T]; H) \times \cdot) = \pi'(C([0, T]; H) \times \cdot)$. *This shows that* $\bar{b}$ *depends on* π *only through its marginals.*

2.2 State equation

Given an initial time $t \in [0, T]$, an initial path $\xi \in \mathbf{S}_2(\mathbb{F})$, a control process $\alpha \in \mathcal{U}$, the state process evolves according to the following controlled path-dependent McKean-Vlasov stochastic differential equation:

$$\begin{cases} dX_s = AX_s + b_s\big(X, \mathbb{P}_X, \alpha_s, \mathbb{P}_{\alpha_s}\big)ds + \sigma_s\big(X, \mathbb{P}_X, \alpha_s, \mathbb{P}_{\alpha_s}\big)dB_s & s > t \\ X_s = \xi_s & s \leq t. \end{cases} \tag{2.4}$$

Definition 2.7. *Fix* $t \in [0, T]$, $\xi \in \mathbf{S}_2(\mathbb{F})$, $\alpha \in \mathcal{U}$. *A **mild solution** of* (2.4) *is a process* $X = (X_s)_{s \in [0, T]}$ *in* $\mathbf{S}_2(\mathbb{F})$ *satisfying*

$$X_s = e^{((s-t)\vee 0)A} \xi_{s \wedge t} + \int_t^{s \vee t} e^{(s-r)A} b_r\big(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big) dr$$

$$+ \int_t^{s \vee t} e^{(s-r)A} \sigma_r\big(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big) dB_r \qquad \forall s \in [0, T], \ \mathbb{P}\text{-}a.s.$$

Proposition 2.8. *Fix* $t \in [0, T]$, $\xi \in \mathbf{S}_2(\mathbb{F})$, $\alpha \in \mathcal{U}$. *Under* $(\mathbf{A}_{A,b,\sigma})$, *equation* (2.4) *admits a unique mild solution* $X^{t,\xi,\alpha} \in \mathbf{S}_2(\mathbb{F})$. *The map*

$$[0, T] \times \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F}), (t, \xi) \mapsto X^{t,\xi,\alpha}$$

is jointly continuous in (t, ξ), *uniformly with respect to* $\alpha \in \mathcal{U}$, *and Lipschitz continuous in* ξ, *uniformly in* t *and* α. *Moreover,* $X^{t,\xi,\alpha} = X^{t,\xi_{\cdot \wedge t},\alpha}$ *and there exists a constant* C, *independent of* t, ξ, α, *such that*

$$\big\|X^{t,\xi,\alpha}\big\|_{\mathbf{S}_2} \leq C\big(1 + \|\xi_{\cdot \wedge t}\|_{\mathbf{S}_2}\big). \tag{2.5}$$

Proof. See Appendix A. □

Remark 2.9. *From Proposition* 2.8 *we have, in particular, for* $r, t \in [0, T], \xi \in \mathbf{S}_2(\mathbb{F})$,

$$\begin{aligned} \limsup_{r \to t} \sup_{\alpha \in \mathcal{U}} \|X_{r \wedge \cdot}^{\alpha, t, \xi} - \xi_{t \wedge \cdot}\|_{\mathbf{S}_2} &\leq \lim_{r \to t} \Big(\sup_{\alpha \in \mathcal{U}} \|X_{r \wedge \cdot}^{\alpha, t, \xi} - X_{r \wedge \cdot}^{\alpha, r, \xi}\|_{\mathbf{S}_2} + \|\xi_{r \wedge \cdot} - \xi_{t \wedge \cdot}\|_{\mathbf{S}_2}\Big) \\ &\leq \lim_{r \to t} \Big(\sup_{\alpha \in \mathcal{U}} \|X^{\alpha, t, \xi} - X^{\alpha, r, \xi}\|_{\mathbf{S}_2} + \|\xi_{r \wedge \cdot} - \xi_{t \wedge \cdot}\|_{\mathbf{S}_2}\Big) = 0, \end{aligned} \tag{2.6}$$

where we have used the fact that $X_{r \wedge \cdot}^{\alpha, r, \xi} = \xi_{r \wedge \cdot}$.

9

Let $\{A_n\}_{n\in\mathbb{N}}$ be the Yosida approximation of A, i.e. $A_n = nA(n-A)^{-1}$, for $n \in \mathbb{N}, n > \eta$, with η as in (2.1). Denote by S^n the uniformly continuous semigroup generated by A_n. Notice that S^n is a pseudo-contraction semigroup for all $n \in \mathbb{N}$, $n > \eta$, and that, for some $\tilde{\eta} > 0$, $\|S^n_t\|_{\mathcal{L}(H)} \leq e^{\tilde{\eta}t}$ uniformly for $n \in \mathbb{N}, t \geq 0$. In particular, we can apply Proposition 2.8 to obtain existence of a unique mild solution $X^{n,t,\xi,\alpha}$ to the following equation:

$$\begin{cases} dX^n_s = A_n X^n_s + b_s\big(X^n, \mathbb{P}_{X^n}, \alpha_s, \mathbb{P}_{\alpha_s}\big)ds + \sigma_s\big(X^n, \mathbb{P}_{X^n}, \alpha_s, \mathbb{P}_{\alpha_s}\big)dB_s, & s > t, \\ X^n_s = \xi_s, & s \leq t. \end{cases} \quad (2.7)$$

Proposition 2.10. *There exists a constant $C > 0$ such that*

$$\sup_{\substack{\alpha\in\mathcal{U} \\ n\in\mathbb{N} \\ t\in[0,T]}} \|X^{n,t,\xi,\alpha} - X^{n,t,\xi',\alpha}\|_{\mathbf{S}_2} \leq C\|\xi - \xi'\|_{\mathbf{S}_2}, \qquad \forall \xi, \xi' \in \mathbf{S}_2(\mathbb{F}). \quad (2.8)$$

Moreover,

$$\lim_{\substack{t\to t' \\ n\to\infty}} \|X^{n,t,\xi,\alpha} - X^{t',\xi,\alpha}\|_{\mathbf{S}_2} = 0, \qquad \forall \alpha \in \mathcal{U}, \xi \in \mathbf{S}_2(\mathbb{F}), t' \in [0,T]. \quad (2.9)$$

Proof. See Appendix A. $\square$

Remark 2.11. *In the third inequality of Assumption $(\mathbf{A}_{A,b,\sigma})$-(iv) we assume, for simplicity, the boundedness of b and σ with respect to the controls. In many applications (e.g. in linear quadratic control cases see Example 2.12 below) the state equation contains unbounded control terms. It is therefore interesting to understand what happens in such cases. Assume that U is a closed subset of a Hilbert space and that the right-hand side of the third inequality of Assumption $(\mathbf{A}_{A,b,\sigma})$-(iv) is replaced by $L(1+|u|_{\mathrm{U}})$. In this case Proposition 2.8 still holds, assuming that $\alpha. \in L^2([0,T]\times\Omega;\mathrm{U})$, with the estimate (2.5) replaced by*

$$\|X^{t,\xi,\alpha}\|_{\mathbf{S}_2} \leq C\left[1 + \|\xi_{\cdot\wedge t}\|_{\mathbf{S}_2} + \left(\mathbb{E}\int_t^T |\alpha_s|^2_{\mathrm{U}}ds\right)\right]. \quad (2.10)$$

The only difference is that the joint continuity in (t,ξ) is uniform only with respect to α belonging to the bounded sets of $L^2([0,T]\times\Omega;\mathrm{U})$. Consequently also the limit in (2.6) is uniform only in the same sense. Similarly, also Proposition 2.10 still holds but with the supremum in α belonging to the bounded sets of $L^2([0,T]\times\Omega;\mathrm{U})$.

Example 2.12. *Examples of problems where the state equation has the above structure.*

(i) *Lifecycle optimal portfolio problems. This family of problems introduces, together with the standard equation for the wealth $x(\cdot)$ used in Merton model, another state variable y which is the labor income of the agent. The equation for the labor income (which is one of the state equations[2] of the optimal portfolio problem) is, in the simplest case, the one of a geometric Brownian motion (see e.g. [34]). It is however natural, for a more realistic description of such dynamics, to introduce two extensions in the equation for y:*

  – *first, as proposed in the concluding remarks of [34] and done (in different cases) in [14, 13, 12], to add a path-dependent term in the drift and/or in the diffusion;*

---

[2]The equation for y is not controlled, however y is part of the state as it appears in the wealth dynamics, which is controlled.

– second, as proposed in the introduction of [31]), to add, in the drift, a mean-field term depending on the distribution $\mathbb{P}_{y(t)}$ of y itself at time t.[3]

Consequently, it makes sense to model the dynamics of the labor income "y" using a one-dimensional stochastic delay ODE of McKean-Vlasov type as follows (here $\phi \in L^2(-d, 0)$ is a given datum providing the weight of the past income into the the current trend, and Z is a one-dimensional Brownian motion).

$$dy(t) = \left[b_0\big(\mathbb{P}_{y(t)}\big) + \int_{-d}^{0} y(t + \xi)\, \phi(\xi)\, d\xi \right] dt + \sigma\, y(t)\, dZ(t).$$

Such equations can be rephrased as SDEs in the Hilbert space $\mathbb{R} \times L^2(-d, 0)$ and the resulting dynamics falls into the class treated in the present section. In [31] the authors study only the case where b_0 is a linear function[4] of the expectation $\mathbb{E}[y(t)]$ but, as explained there, other choices are possible, such as "the median wage or the truncated average above a certain level, within the company or even within the profession".

The setting of the present paper allows to cover not only the case when b_0 above is nonlinear but also more general (and still interesting, like e.g. the average income of the last years) cases where it depends on the law of the past of y.

Clearly, the problem becomes much more difficult than the one treated in [31] and we should consider the results of this paper as a first step to study such types of models.

(ii) *Optimal investment with vintage capital.* These are typical partial equilibrium models arising in Economics (see e.g. [8, 9, 36, 40]) where the state variable "x" is the capital stock and the control variable is the investment "u", both depending on time $t \geq 0$ and vintage $s \in [0, \bar{s}]$ (here $\bar{s} > 0$ is the maximum possible vintage): capital goods indexed with small s embody newer technologies. In the above papers x, in the simplest cases, is required to satisfy a first-order PDE of the following type:

$$\frac{\partial x(t, s)}{\partial t} + \frac{\partial x(t, s)}{\partial s} = -\delta x(t, s) + u(t, s),$$

where δ is a depreciation rate of the capital goods, which is kept constant for simplicity. This PDE can be easily rewritten as an ODE in the space $H := L^2(0, \bar{s})$. If one wants to take into account stochastic disturbances (similarly to what is done in [42] in a case without vintage), the state equation, written in $L^2(0, \bar{s})$, becomes the following infinite dimensional SDE (here B is a cylindrical Wiener process):

$$dx(t) = [Ax(t) - \delta x(t) + Cu(t)]\, dt + \sigma(x(t))\, dB(t),$$

for suitable linear operators A, C, δ and $\sigma \colon H \to \mathcal{L}(H)$. Here $x(t), u(t)$ stand for $x(t, \cdot)$ and $u(t, \cdot)$, respectively.

Given such a controlled infinite dimensional SDE, the objective to be maximized depends,

---

[3] As written at pages 2-3 of [31], "the labor income y_i of an agent i is benchmarked against the labor incomes of a population $y^N := (y_1, y_2, \ldots, y_N)$ of N agents with comparable tasks or ranks among the profession such as the level of full professor, associate professor, actuary, trader, risk manager etc., where one usually uses some wage level $b(y^N)$ as a reference to declare whether that agent has a superior, fair or inferior labor income compared with her peers".

[4] This is needed there to get a simplified HJB equation and to find explicit solutions of it.

11

beyond the control u, also on the production $Q(t)$ provided by the capital stock. The expression of $Q(t)$, as given for example in [40], is

$$Q(t) := \int_0^{\bar{s}} f(t-s)v(s)x(t,s)ds,$$

where f takes into account the technological progress and v embodies learning and spillover effects. As observed e.g in [46, 57], effects of this type can be modelled taking f and v depending on the distribution of $x(t)$ and $u(t)$. Moreover, effects like the so-called time to build (see e.g. [60, 4]) call for path-dependency for such coefficients (see, for a deterministic infinite dimensional modelling of time-to-build, [4, 5, 63, 38, 39]).

(iii) *Optimal consumption in spatial growth models in Economics.* A way to model capital accumulation in spatial growth models (see e.g. [17]) is to assume that the capital $x(t,\xi)$ at time $t \geq 0$ in the position ξ (here we take, for simplicity, $\xi \in S^1$, the one dimensional sphere) satisfies a second order PDE like

$$\frac{\partial x(t,\xi)}{\partial t} = \frac{\partial^2 x(t,\xi)}{\partial \xi^2} + a(t,\xi)x(t,\xi) - c(t,\xi), \qquad \xi \in S^1, \, t \geq 0,$$

where a is a productivity coefficient and c the consumption. If we modify such PDE to take into account the time delay d due to time-to-build, it takes the following form:

$$\frac{\partial x(t,\xi)}{\partial t} = \frac{\partial^2 x(t,\xi)}{\partial \xi^2} + a(t,\xi)x(t-d,\xi) - c(t,\xi), \qquad \xi \in S^1, \, t \geq 0.$$

By considering stochastic disturbances (as done, e.g., in [16] for the case without space variable and in [50] in the spatial growth framework) and of the mean-field dependence on the productivity (as argued in the previous example) we get the following path-dependent SPDE of McKean-Vlasov type in the space $H := L^2(S^1)$

$$dx(t) = \left[Ax(t) + h\left(x(t-d), \mathbb{P}_{x(t)}\right) - c(t)\right] dt + \sigma(x(t))dB(t),$$

where B is a cylindrical Wiener process, A is a suitable linear second order differential operator, $h \colon H \times \mathcal{P}(H) \to H$ and $\sigma \colon H \to \mathcal{L}(H)$. Here $x(t), c(t)$ stand for $x(t,\cdot)$ and $c(t,\cdot)$, respectively. Again this problem falls into the class we treat in the present paper.

3 The optimal control problem

3.1 Reward functional and lifted value function

We are given two functions

$$f \colon [0,T] \times C([0,T];H) \times \mathcal{P}_2\big(C([0,T];H)\big) \times \mathrm{U} \times \mathcal{P}(\mathrm{U}) \longrightarrow \mathbb{R}$$
$$g \colon C([0,T];H) \times \mathcal{P}_2\big(C([0,T];H)\big) \longrightarrow \mathbb{R}$$

on which we impose the following assumptions.

Assumption ($\mathbf{A}_{f,g}$).

(i) *The functions f and g are measurable.*

(ii) *The function f satisfies the non-anticipativity property:*

$$f_t(x, \mu, u, \nu) = f_t(x_{\cdot \wedge t}, \mu, u, \nu),$$

for all $(t, x, \mu, u, \nu) \in [0, T] \times C([0, T]; H) \times \mathcal{P}_2(C([0, T]; H)) \times U \times \mathcal{P}(U)$.

(iii) *There exists a locally bounded function $h \colon [0, \infty) \to [0, \infty)$ such that*

$$|f_t(x, \mu, u, \nu)| \le h(\mathcal{W}_2(\mu, \delta_0))(1 + \|x\|_t^2), \qquad |g(x, \mu)| \le h(\mathcal{W}_2(\mu, \delta_0))(1 + \|x\|_T^2),$$

for all $(t, x, \mu, u, \nu) \in [0, T] \times C([0, T]; H) \times \mathcal{P}_2(C([0, T]; H)) \times U \times \mathcal{P}(U)$.

We will also need the following continuity assumption on f and g.

Assumption $(\mathbf{A}_{f,g})_{\mathrm{cont}}$. *The function f is locally uniformly continuous in (x, μ) uniformly with respect to (t, u, ν). Similarly, g is locally uniformly continuous. More precisely, it holds that: for every $\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $\delta = \delta(\varepsilon, n) > 0$ such that, for every $(t, u, \nu) \in [0, T] \times U \times \mathcal{P}(U)$, $(x, \mu), (x', \mu') \in C([0, T]; H) \times \mathcal{P}_2(C([0, T]; H))$, with $\|x\|_T + \mathcal{W}_2(\mu, \delta_0) \le n$ and $\|x'\|_T + \mathcal{W}_2(\mu', \delta_0) \le n$,*

$$\|x - x'\|_t + \mathcal{W}_2(\mu, \mu') \le \delta$$
$$\implies \quad |f(t, x, \mu, u, \nu) - f(t, x', \mu', u, \nu)| \le \varepsilon \quad \text{and} \quad |g(x, \mu) - g(x', \mu')| \le \varepsilon.$$

Under Assumptions $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$, from Proposition 2.8 we get that the *reward functional* J, given by

$$J(t, \xi, \alpha) = \mathbb{E}\left[\int_t^T f_s\big(X^{t,\xi,\alpha}, \mathbb{P}_{X_{\cdot \wedge s}^{t,\xi,\alpha}}, \alpha_s, \mathbb{P}_{\alpha_s}\big) ds + g\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}}\big) \right],$$

is well-defined for any $(t, \xi, \alpha) \in [0, T] \times \mathbf{S}_2(\mathbb{F}) \times \mathcal{U}$. We then consider the function $V \colon [0, T] \times \mathbf{S}_2(\mathbb{F}) \longrightarrow \mathbb{R}$, to which we refer as the *lifted value function*, defined as

$$V(t, \xi) = \sup_{\alpha \in \mathcal{U}} J(t, \xi, \alpha), \qquad \forall (t, \xi) \in [0, T] \times \mathbf{S}_2(\mathbb{F}). \tag{3.1}$$

Remark 3.1. *Recall from Proposition 2.8 that $X^{t,\xi,\alpha}$ only involves the values of ξ up to time t, namely it holds that $X^{t,\xi,\alpha} = X^{t,\xi_{\cdot \wedge t},\alpha}$. As a consequence, both J and V satisfy the* non-anticipativity *property:*

$$J(t, \xi, \alpha) = J(t, \xi_{\cdot \wedge t}, \alpha), \qquad V(t, \xi) = V(t, \xi_{\cdot \wedge t}),$$

for every $(t, \xi) \in [0, T] \times \mathbf{S}_2(\mathbb{F})$, $\alpha \in \mathcal{U}$.

Remark 3.2. *In Remark 2.11 we looked at the case when U is a Hilbert space and the coefficients of the state equation have linear growth in the controls. In such a case, in order to have a well-defined reward functional J, we need some compensating term in the current reward f. A typical assumption which guarantees that J is well-defined is that the first inequality of Assumption $(\mathbf{A}_{f,g})$-(iii) is replaced by*

$$f_t(x, \mu, u, \nu) \le h(\mathcal{W}_2(\mu, \delta_0))(1 + \|x\|_t^2) - C|u|_U^\theta,$$

13

for some $\theta > 1$ and $C > 0$. This would include some typical linear-quadratic control cases. For example in the case of optimal investment problems mentioned in Example 2.12-(ii) a typical form of f would be (recall that here $H = L^2(0, \bar{s})$)

$$f_t(x, \mu, u, \nu) = f_t(x, u) = e^{-rt}[R(Q(t)) - \langle a_2, u(t) \rangle_H - \langle Mu(t), u(t) \rangle_H], \tag{3.2}$$

where r is the interest rate, R is suitable one variable function (possibly linear or quadratic) $a_2 \in L^2(0, \bar{s})$, and M is a suitable multiplication operator in $L^2(0, \bar{s})$ (see e.g. [36, 40]).

Proposition 3.3. *Suppose that $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$ hold. The function V satisfies a quadratic growth condition: there exists a constant C such that*

$$|V(t, \xi)| \leq C \left(1 + \|\xi_{\cdot \wedge t}\|_{\mathbf{S}_2}^2 \right), \tag{3.3}$$

for every $(t, \xi) \in [0, T] \times \mathbf{S}_2(\mathbb{F})$. Moreover, if in addition $(\mathbf{A}_{f,g})_{\mathrm{cont}}$ holds, the map $V : [0, T] \times \mathbf{S}_2(\mathbb{F}) \to \mathbb{R}$ is jointly continuous.

Proof. We split the proof into two steps.

Step 1. *Proof of estimate (3.3).* From the definition of J, we have

$$|J(t, \xi, \alpha)| \leq \mathbb{E}\left[\int_t^T \left| f_s\big(X^{t,\xi,\alpha}, \mathbb{P}_{X_{\cdot \wedge s}^{t,\xi,\alpha}}, \alpha_s, \mathbb{P}_{\alpha_s}\big) \right| ds \right] + \mathbb{E}\left[\left| g\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}}\big) \right| \right].$$

By the quadratic growth of f and g, together with estimate (2.5), we see that there exists a constant C such that

$$|J(t, \xi, \alpha)| \leq C \left(1 + \|\xi_{\cdot \wedge t}\|_{\mathbf{S}_2}^2 \right), \tag{3.4}$$

for every $(t, \xi, \alpha) \in [0, T] \times \mathbf{S}_2(\mathbb{F}) \times \mathcal{U}$. Then, estimate (3.3) follows directly from the definition of V and the fact that (3.4) holds uniformly with respect to $\alpha \in \mathcal{U}$.

Step 2. *Continuity of V.* We begin noticing that, for every $(t, \xi), (s, \eta) \in [0, T] \times \mathbf{S}_2(\mathbb{F})$,

$$|V(t, \xi) - V(s, \eta)| \leq \sup_{\alpha \in \mathcal{U}} |J(t, \xi, \alpha) - J(s, \eta, \alpha)|.$$

Then, the continuity of V follows once we prove that J is continuous in (t, ξ) uniformly with respect to α, namely that the following property holds: for every $\varepsilon > 0$ and every $(t, \xi) \in [0, T] \times \mathbf{S}_2(\mathbb{F})$, there exists $\delta = \delta(\varepsilon, t, \xi) > 0$ such that, for every $(s, \eta, \alpha) \in [0, T] \times \mathbf{S}_2(\mathbb{F}) \times \mathcal{U}$,

$$|t - s| \leq \delta \quad \text{and} \quad \|\xi - \eta\|_{\mathbf{S}_2} \leq \delta \quad \Longrightarrow \quad |J(t, \xi, \alpha) - J(s, \eta, \alpha)| \leq \varepsilon.$$

Such a property is a straightforward consequence of the last statement of Proposition 2.8 and of assumption $(\mathbf{A}_{f,g})_{\mathrm{cont}}$. $\qquad \square$

3.2 Dynamic programming principle for V

In this section we prove the dynamic programming principle for the lifted value function V defined in (3.1).

Theorem 3.4. *Suppose that $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$ hold. The lifted value function V satisfies the **dynamic programming principle**: for every $t, s \in [0, T]$, with $t \leq s$, and every $\xi \in \mathbf{S}_2(\mathbb{F})$ it holds that*

$$V(t, \xi) = \sup_{\alpha \in \mathcal{U}} \left\{ \mathbb{E}\left[\int_t^s f_r\big(X^{t,\xi,\alpha}, \mathbb{P}_{X_{\cdot \wedge r}^{t,\xi,\alpha}}, \alpha_r, \mathbb{P}_{\alpha_r}\big) dr \right] + V\big(s, X^{t,\xi,\alpha}\big) \right\}.$$

Proof. Set

$$\Lambda(t,\xi) := \sup_{\alpha \in \mathcal{U}} \left\{ \mathbb{E} \left[\int_t^s f_r\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\, dr \right] + V\big(s, X^{t,\xi,\alpha}\big) \right\}.$$

Step 1. *Proof of the inequality* $\Lambda(t,\xi) \geq V(t,\xi)$. For every fixed $\alpha \in \mathcal{U}$, the lifted value function at $(s, X^{t,\xi,\alpha})$ is given by

$$V(s, X^{t,\xi,\alpha}) = \sup_{\beta \in \mathcal{U}} \mathbb{E} \left[\int_s^T f_r\Big(X^{s,X^{t,\xi,\alpha},\beta}, \mathbb{P}_{X^{s,X^{t,\xi,\alpha},\beta}_{\cdot \wedge r}}, \beta_r, \mathbb{P}_{\beta_r}\Big)\, dr + g\Big(X^{s,X^{t,\xi,\alpha},\beta}, \mathbb{P}_{X^{s,X^{t,\xi,\alpha},\beta}}\Big) \right].$$

Choosing $\beta = \alpha$, we find

$$V(s, X^{t,\xi,\alpha}) \geq \mathbb{E} \left[\int_s^T f_r\Big(X^{s,X^{t,\xi,\alpha},\alpha}, \mathbb{P}_{X^{s,X^{t,\xi,\alpha},\alpha}_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\Big)\, dr + g\Big(X^{s,X^{t,\xi,\alpha},\alpha}, \mathbb{P}_{X^{s,X^{t,\xi,\alpha},\alpha}}\Big) \right].$$

By the uniqueness property for equation (2.4) stated in Proposition 2.8, we obtain the flow property

$$X^{t,\xi,\alpha} = X^{s,X^{t,\xi,\alpha},\alpha}.$$

Hence

$$V(s, X^{t,\xi,\alpha}) \geq \mathbb{E} \left[\int_s^T f_r\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\, dr + g\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}}\big) \right].$$

Adding to both sides the quantity $\mathbb{E} \int_t^s f_r(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r})\, dr$, we get

$$\Lambda(t,\xi) \geq \mathbb{E} \left[\int_t^T f_r\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\, dr + g\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}}\big) \right].$$

As the latter inequality holds true for every $\alpha \in \mathcal{U}$, we conclude that $\Lambda(t,\xi) \geq V(t,\xi)$.

Step 2. *Proof of the inequality* $\Lambda(t,\xi) \leq V(t,\xi)$. For every $\varepsilon > 0$, let $\alpha^\varepsilon \in \mathcal{U}$ be such that

$$\Lambda(t,\xi) \leq \mathbb{E} \left[\int_t^s f_r\big(X^{t,\xi,\alpha^\varepsilon}, \mathbb{P}_{X^{t,\xi,\alpha^\varepsilon}_{\cdot \wedge r}}, \alpha_r^\varepsilon, \mathbb{P}_{\alpha_r^\varepsilon}\big)\, dr \right] + V(s, X^{t,\xi,\alpha^\varepsilon}) + \varepsilon. \tag{3.5}$$

From the definition of $V(s, X^{t,\xi,\alpha^\varepsilon})$, it follows that there exists $\beta^\varepsilon \in \mathcal{U}$ such that

$$V(s, X^{t,\xi,\alpha^\varepsilon}) \leq \mathbb{E} \left[\int_s^T f_r\Big(X^{s,X^{t,\xi,\alpha^\varepsilon},\beta^\varepsilon}, \mathbb{P}_{X^{s,X^{t,\xi,\alpha^\varepsilon},\beta^\varepsilon}_{\cdot \wedge r}}, \beta_r^\varepsilon, \mathbb{P}_{\beta_r^\varepsilon}\Big)\, dr \right. \tag{3.6}$$
$$\left. + g\Big(X^{s,X^{t,\xi,\alpha^\varepsilon},\beta^\varepsilon}, \mathbb{P}_{X^{s,X^{t,\xi,\alpha^\varepsilon},\beta^\varepsilon}}\Big) \right] + \varepsilon.$$

Set

$$\gamma^\varepsilon = \alpha^\varepsilon \mathbb{1}_{[0,s]} + \beta^\varepsilon \mathbb{1}_{(s,T]}.$$

Notice that $\gamma^\varepsilon \in \mathcal{U}$. Using again the uniqueness property for equation (2.4), we get

$$X^{s,X^{t,\xi,\alpha^\varepsilon},\beta^\varepsilon} = X^{t,\xi,\gamma^\varepsilon}.$$

Hence, (3.6) becomes

$$V(s, X^{t,\xi,\alpha^\varepsilon}) \leq \mathbb{E}\left[\int_s^T f_r\big(X^{t,\xi,\gamma^\varepsilon}, \mathbb{P}_{X^{t,\xi,\gamma^\varepsilon}_{\cdot \wedge r}}, \gamma^\varepsilon_r, \mathbb{P}_{\gamma^\varepsilon_r}\big)\, dr + g\big(X^{t,\xi,\gamma^\varepsilon}, \mathbb{P}_{X^{t,\xi,\gamma^\varepsilon}}\big)\right] + \varepsilon.$$

Then, by (3.5) it follows that

$$\Lambda(t,\xi) \leq \mathbb{E}\left[\int_t^s f_r\big(X^{t,\xi,\alpha^\varepsilon}, \mathbb{P}_{X^{t,\xi,\alpha^\varepsilon}_{\cdot \wedge r}}, \alpha^\varepsilon_r, \mathbb{P}_{\alpha^\varepsilon_r}\big)\, dr + \int_s^T f_r\big(X^{t,\xi,\gamma^\varepsilon}, \mathbb{P}_{X^{t,\xi,\gamma^\varepsilon}_{\cdot \wedge r}}, \gamma^\varepsilon_r, \mathbb{P}_{\gamma^\varepsilon_r}\big)\, dr \right.$$

$$\left. + g\big(X^{t,\xi,\gamma^\varepsilon}, \mathbb{P}_{X^{t,\xi,\gamma^\varepsilon}}\big)\right] + 2\varepsilon. \tag{3.7}$$

From the definition of γ^ε, we see that

$$X^{t,\xi,\alpha^\varepsilon}_{\cdot \wedge s} = X^{t,\xi,\gamma^\varepsilon}_{\cdot \wedge s}.$$

As a consequence, we can rewrite (3.7) in terms of the only process $X^{t,\xi,\gamma^\varepsilon}$ as

$$\Lambda(t,\xi) \leq \mathbb{E}\left[\int_t^T f_r\big(X^{t,\xi,\gamma^\varepsilon}, \mathbb{P}_{X^{t,\xi,\gamma^\varepsilon}_{\cdot \wedge r}}, \gamma^\varepsilon_r, \mathbb{P}_{\gamma^\varepsilon_r}\big) dr + g\big(X^{t,\xi,\gamma^\varepsilon}, \mathbb{P}_{X^{t,\xi,\gamma^\varepsilon}}\big)\right] + 2\varepsilon \leq V(t,\xi) + 2\varepsilon.$$

The claim follows from the arbitrariness of ε. $\qquad\square$

Remark 3.5. *It is worth noticing that, despite the stochastic setting, there is no issue of measurability in the proof of the dynamic programming principle. This is a consequence of the fact that the function V depends on the whole random variable ξ, so that the proof of the dynamic programming principle can be done proceeding along the same lines as in the case of deterministic optimal control.*

3.3 Law invariance property of the lifted value function V

In the present section we introduce the *value function* of the optimal control problem, which is a real-valued map defined on $[0,T] \times \mathcal{P}_2(C([0,T]; H))$ (see (3.10)). In order to define such a value function, it is necessary to prove that the map V satisfies the following *law invariance property:* for every $t \in [0,T]$ and every $\xi, \eta \in \mathbf{S}_2(\mathbb{F})$ it holds that

$$V(t,\xi) = V(t,\eta).$$

This is the subject of the next theorem.

Theorem 3.6. *Suppose that $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$ hold. Fix $t \in [0,T]$ and $\xi, \eta \in \mathbf{S}_2(\mathbb{F})$, with $\mathbb{P}_\xi = \mathbb{P}_\eta$. Suppose that there exist two random variables U_ξ and U_η having uniform distribution on $[0,1]$, being $\mathcal{F}_t$-measurable and such that ξ and U_ξ (resp. η and U_η) are independent. Then, it holds that*

$$V(t,\xi) = V(t,\eta).$$

*If in addition $(\mathbf{A}_{f,g})_{\mathrm{cont}}$ holds, the map V satisfies the **law invariance property:** for every $t \in [0,T]$ and every $\xi, \eta \in \mathbf{S}_2(\mathbb{F})$, with $\mathbb{P}_\xi = \mathbb{P}_\eta$, it holds that*

$$V(t,\xi) = V(t,\eta).$$

Proof. We split the proof into two steps.

Step 1. *Only* $(\mathbf{A}_{A,b,\sigma})$ *and* $(\mathbf{A}_{f,g})$ *hold.* Fix $t \in [0, T]$, $\xi, \eta \in \mathbf{S}_2(\mathbb{F})$, with $\mathbb{P}_\xi = \mathbb{P}_\eta$, and let U_ξ, U_η be $\mathcal{F}_t$-measurable random variables with uniform distribution on $[0, 1]$, such that ξ and U_ξ (resp. η and U_η) are independent.

By Remark 3.1 we can assume that $\xi = \xi_{\cdot \wedge t}$ and $\eta = \eta_{\cdot \wedge t}$, so, in particular, both ξ and η are $\mathcal{B}([0, T]) \otimes \mathcal{F}_t$-measurable (this is needed in order to apply Lemma B.2 of Appendix B). Now, given $\alpha \in \mathcal{U}$ consider the function a: $[0, T] \times \Omega \times C([0, T]; H) \times [0, 1] \to \mathrm{U}$ introduced in Lemma B.2. By (B.1) we have

$$\Big((\xi_s)_{s \in [0,T]}, (\mathrm{a}_s(\xi, U_\xi))_{s \in [t,T]}, (B_s - B_t)_{s \in [t,T]}\Big) \overset{\mathscr{L}}{=} \Big((\xi_s)_{s \in [0,T]}, (\alpha_s)_{s \in [t,T]}, (B_s - B_t)_{s \in [t,T]}\Big),$$

where $\overset{\mathscr{L}}{=}$ stands for equality in law (between random objects defined on $(\Omega, \mathcal{F}, \mathbb{P})$). Then, notice that (here we use again that ξ and η are $\mathcal{B}([0, T]) \otimes \mathcal{F}_t$-measurable, so, in particular, they are independent of $(B_s - B_t)_{s \in [t,T]}$)

$$\Big((\xi_s)_{s \in [0,T]}, (\alpha_s)_{s \in [t,T]}, (B_s - B_t)_{s \in [t,T]}\Big) \overset{\mathscr{L}}{=} \Big((\eta_s)_{s \in [0,T]}, (\beta_s)_{s \in [t,T]}, (B_s - B_t)_{s \in [t,T]}\Big), \qquad (3.8)$$

where

$$\beta := \big(\mathrm{a}_s(\eta, U_\eta)\big)_{s \in [0,T]}.$$

Observe that $\beta \in \mathcal{U}$ and, by (3.8),

$$\big((X_s^{t,\xi,\alpha})_{s \in [t,T]}, (\alpha_s)_{s \in [t,T]}\big) \overset{\mathscr{L}}{=} \big((X_s^{t,\eta,\beta})_{s \in [t,T]}, (\beta_s)_{s \in [t,T]}\big),$$

where the above equality in law can be deduced from (3.8) proceeding along the same lines as in the proof of Proposition 1.137 in [35]. As a consequence, it holds that

$$J(t, \xi, \alpha) = J(t, \eta, \beta).$$

Hence $J(t, \xi, \alpha) \leq V(t, \eta)$. From the arbitrariness of α, we deduce that $V(t, \xi) \leq V(t, \eta)$. Changing the roles of ξ and η we get the opposite inequality, from which we deduce that $V(t, \xi) = V(t, \eta)$.

Step 2. *Assumptions* $(\mathbf{A}_{A,b,\sigma})$, $(\mathbf{A}_{f,g})$, $(\mathbf{A}_{f,g})_{\mathrm{cont}}$ *hold.* Fix $t \in [0, T]$ and $\xi, \eta \in \mathbf{S}_2(\mathbb{F})$, with $\mathbb{P}_\xi = \mathbb{P}_\eta$. As in the previous step, we exploit Remark 3.1 and take $\xi = \xi_{\cdot \wedge t}$, $\eta = \eta_{\cdot \wedge t}$ (so, in particular, both ξ and η are $\mathcal{B}([0, T]) \otimes \mathcal{F}_t$-measurable; this is needed in order to apply Lemma B.3).

Substep 2.1. *The discrete case.* Suppose that

$$\mathbb{P}_\xi = \sum_{i=1}^m p_i \, \delta_{x_i},$$

for some $\{x_1, \ldots, x_m\} \subset C([0, T]; H)$, with $x_i \neq x_j$ if $i \neq j$, where δ_{x_i} is the Dirac measure at x_i and $p_i > 0$, with $\sum_{i=1}^m p_i = 1$. Then, by Lemma B.3 there exist two $\mathcal{F}_t$-measurable random variables U_ξ and U_η, with uniform distribution on $[0, 1]$, such that ξ and U_ξ (resp. η and U_η) are independent. The claim then follows from **Step 1.**

Substep 2.2. *The general case.* In the general case, we rely on the continuity of the map $\xi \mapsto V(t, \xi)$, defined from $\mathbf{S}_2(\mathbb{F})$ into $\mathbb{R}$, which follows from Proposition 3.3. More precisely, we proceed

by approximating ξ and η. For $n \in \mathbb{N}$, let $\{C_i^n\}_{i \in \mathbb{N}}$ be a partition of $C([0,T];H)$ of Borel sets such that $\operatorname{diam}(C_i^n) < 2^{-n}$. For each $i \in \mathbb{N}$, choose $x_i^n \in C_i^n$. Then, define

$$\tilde{\xi}_n := \sum_{i=1}^{\infty} x_i^n \, \mathbb{1}_{C_i^n}(\xi), \qquad \tilde{\eta}_n := \sum_{i=0}^{\infty} x_i^n \, \mathbb{1}_{C_i^n}(\eta).$$

Notice that $\tilde{\xi}_n, \tilde{\eta}_n \in \mathbf{S}_2(\mathbb{F})$ and $\tilde{\xi}_n \to \xi$, $\tilde{\eta}_n \to \eta$ uniformly with respect to $\omega \in \Omega$. Moreover, $\tilde{\xi}_n$ and $\tilde{\eta}_n$ have the same law. By a diagonal argument, we can choose $N_n \in \mathbb{N}$ such that the sequences $\{\xi_n\}_{n \in \mathbb{N}}$ and $\{\eta_n\}_{n \in \mathbb{N}}$, defined by

$$\xi_n := \sum_{i=1}^{N_n} x_i^n \, \mathbb{1}_{C_i^n}(\xi), \qquad \eta_n := \sum_{i=1}^{N_n} x_i^n \, \mathbb{1}_{C_i^n}(\eta), \tag{3.9}$$

converge respectively to ξ and η, both $\mathbb{P}$-a.s. and in $L^2(\Omega; C([0,T];H))$. From **Substep 2.1** we have

$$V(t, \xi_n) = V(t, \eta_n), \qquad \forall\, n \in \mathbb{N}.$$

Then, using the continuity of V, we can pass to the limit as $n \to \infty$ and conclude that $V(t, \xi) = V(t, \eta)$. $\qquad\qquad\square$

Remark 3.7. *Suppose that* $(\mathbf{A}_{A,b,\sigma})$, $(\mathbf{A}_{f,g})$, $(\mathbf{A}_{f,g})_{\mathrm{cont}}$ *hold. Thanks to the law invariance property stated in Theorem 3.6, in the definition of V we can consider only $\xi \in \mathbf{S}_2(\mathcal{G})$ rather than $\xi \in \mathbf{S}_2(\mathbb{F})$ (recall that it was necessary to take $\xi \in \mathbf{S}_2(\mathbb{F})$ in order to state and prove the dynamic programming principle, Theorem 3.4, where for instance the initial condition at time s is $X^{t,\xi,\alpha}$ and $X^{t,\xi,\alpha} \in \mathbf{S}_2(\mathbb{F})$ but, in general, $X^{t,\xi,\alpha} \notin \mathbf{S}_2(\mathcal{G})$).*

Remark 3.8. *In the finite-dimensional and non-path-dependent case, the law invariance property was already addressed in [27], Proposition 3.1, under only $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$. Notice however that the proof of such a proposition is based on the measurable selection theorem stated in [2], Corollary 18.23, which is unfortunately not true. For this reason, Theorem 3.6 is also relevant in the finite-dimensional and non-path-dependent setting. Moreover, we emphasize that assuming only $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$ is not enough for the validity of the law invariance property. To this regard, we give the following example.*

Example. *Let $T = 1$, $H = \mathbb{R}^3$, $K = \mathbb{R}$, $\mathrm{U} = [0,1]$. We consider a non-path-dependent setting. The coefficients*

$$b, \, \sigma, \, f \colon [0,T] \times \mathbb{R}^3 \times \mathcal{P}_2(\mathbb{R}^3) \times [0,1] \times \mathcal{P}([0,1]) \longrightarrow \mathbb{R}^3, \, \mathbb{R}^3, \, \mathbb{R}$$

and $g \colon \mathbb{R}^3 \times \mathcal{P}_2(\mathbb{R}^3) \to \mathbb{R}$ are given by

$$b_t(x, \mu, u, \nu) := \begin{pmatrix} 0 \\ u \\ 0 \end{pmatrix}, \quad \sigma_t(x, \mu, u, \nu) := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad f_t(x, \mu, u, \nu) := 0, \quad g(x, \mu) := \mathbb{1}_{\{\mu = \mu_0\}}$$

for every $(t, x, \mu, u, \nu) \in [0,T] \times \mathbb{R}^3 \times \mathcal{P}_2(\mathbb{R}^3) \times [0,1] \times \mathcal{P}([0,1])$, where $\mu_0 \in \mathcal{P}_2(\mathbb{R}^3)$ is the probability measure defined as

$$\mu_0 := \mathrm{Unif}\,(0,1) \otimes \mathrm{Bern}(1/2) \otimes \mathcal{N}(0,1),$$

with $\mathrm{Unif}\,(0,1)$ being the uniform distribution on $[0,1]$, $\mathrm{Bern}(1/2)$ the Bernoulli distribution with parameter $1/2$, $\mathcal{N}(0,1)$ the standard Gaussian distribution.

We now fix the probabilistic setting. Consider the probability spaces $(\Omega^\circ, \mathcal{F}^\circ, \mathbb{P}^\circ)$ and $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ where $\Omega^\circ = [0,1]$, $\mathcal{F}^\circ$ its Borel σ-algebra and $\mathbb{P}^\circ$ is the Lebesgue measure on the unit interval, while $\Omega^1 = \{\omega^1 \in C([0,1]; \mathbb{R}) \colon \omega^1_0 = 0\}$, $\mathcal{F}^1$ its Borel σ-algebra and $\mathbb{P}^1$ is the Wiener measure on $(\Omega^1, \mathcal{F}^1)$. We then define $\Omega := \Omega^\circ \times \Omega^1$, $\mathcal{F}$ the completion of $\mathcal{F}^\circ \otimes \mathcal{F}^1$ with respect to $\mathbb{P}^\circ \otimes \mathbb{P}^1$ and by $\mathbb{P}$ the extension of $\mathbb{P}^\circ \otimes \mathbb{P}^1$ to $\mathcal{F}$. We also denote by $\mathcal{G} := \mathcal{F}^\circ \otimes \{\emptyset, \Omega^1\}$ the canonical extension of $\mathcal{F}^\circ$ to the product space Ω. Finally, we denote by $B = (B_t)_{t \in [0,1]}$ the canonical process $B_t(\omega^\circ, \omega^1) := \omega^1_t$, $\forall t \in [0,1]$. Notice that, under the probability measure $\mathbb{P}$, the process B is a real-valued Brownian motion. Then, in the present context the lifted value function is given by

$$V(t, \xi) = \sup_{\alpha \in \mathcal{U}} \mathbb{E}\left[\mathbb{1}_{\left\{\mathbb{P}_{X_1^{t,\xi,\alpha}} = \mu_0\right\}}\right], \qquad \forall (t, \xi) \in [0,1] \times L^2(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^3),$$

where

$$\xi = \begin{pmatrix} \xi^1 \\ \xi^2 \\ \xi^3 \end{pmatrix} \qquad \text{and} \qquad X_1^{t,\xi,\alpha} = \begin{pmatrix} \xi^1 \\ \xi^2 + \int_t^1 \alpha_s \, ds \\ \xi^3 + B_1 - B_t \end{pmatrix}.$$

Now, let $\xi \colon \Omega \to \mathbb{R}^3$ be given by

$$\xi(\omega^\circ, \omega^1) := \begin{pmatrix} \omega^\circ \\ 0 \\ 0 \end{pmatrix}, \qquad \forall (\omega^\circ, \omega^1) \in \Omega.$$

Notice that ξ^1 has distribution $\mathrm{Unif}\,(0,1)$. Moreover, ξ is $\mathcal{G}$-measurable and generates the σ-algebra $\mathcal{G}$ itself, namely $\mathcal{G} = \sigma(\xi)$. Define $\eta \colon \Omega \to \mathbb{R}^3$ and $Z \colon \Omega \to [0,1]$ by

$$\eta := \begin{pmatrix} 2\xi \, \mathbb{1}_{\{\xi \leq 1/2\}} + (2\xi - 1) \, \mathbb{1}_{\{\xi > 1/2\}} \\ 0 \\ 0 \end{pmatrix}, \qquad Z := \mathbb{1}_{\{\xi \leq 1/2\}}.$$

Notice that η and Z are $\mathcal{G}$-measurable and independent. Moreover, the first component of η, namely η^1, has distribution $\mathrm{Unif}\,(0,1)$, while Z has distribution $\mathrm{Bern}(1/2)$.

Let us prove that $V(0, \xi) \neq V(0, \eta)$ and, in particular, $V(0, \xi) = 0$ while $V(0, \eta) = 1$. If the initial condition at time $t = 0$ is η, then taking the control process $\alpha_s^* = Z$, $\forall s \in [0, T]$, we get

$$X_1^{0,\eta,\alpha^*} = \begin{pmatrix} \eta^1 \\ Z \\ B_1 \end{pmatrix}.$$

Notice that $\mathbb{P}_{X_1^{0,\eta,\alpha^*}} = \mu_0$, which proves that $V(0, \eta) = 1$. On the other hand, when the initial condition is ξ, for every $\alpha \in \mathcal{U}$ we have

$$X_1^{0,\xi,\alpha}(\omega^\circ, \omega^1) = \begin{pmatrix} \omega^\circ \\ \int_0^1 \alpha_s(\omega^\circ, \omega^1) \, ds \\ \omega^1_1 \end{pmatrix}, \qquad \forall (\omega^\circ, \omega^1) \in \Omega.$$

19

Let us prove that $V(0,\xi) = 0$. Suppose on the contrary that $V(0,\xi) = 1$. This implies that $X_1^{0,\xi,\alpha}$ has distribution μ_0, so in particular $\int_0^1 \alpha_s\,ds$ has Bernoulli distribution with parameter $1/2$. Recalling that the control process takes values in $[0,1]$, there exists some set $E \in \mathcal{F}$ such that $\alpha_s(\omega) = 1_E(\omega)$, $\forall\,(s,\omega) \in [0,1] \times \Omega$. Since α is progressively measurable, $E \in \mathcal{G}$; in other words, α depends only on ω°. It follows that the random variable $\int_0^1 \alpha_s\,ds$ cannot be independent of ξ (and, a fortiori, of (ξ, B_1)), unless it is a constant. This contradicts the fact that $\int_0^1 \alpha_s\,ds$ has Bernoulli distribution of parameter $1/2$ and proves that $V(0,\xi) = 0$.

In conclusion, under assumptions $(\mathbf{A}_{A,b,\sigma})$, $(\mathbf{A}_{f,g})$, $(\mathbf{A}_{f,g})_{\mathrm{cont}}$, we can define the *value function* $v\colon [0,T] \times \mathcal{P}_2(C([0,T];H)) \to \mathbb{R}$ as

$$v(t,\mu) = V(t,\xi), \qquad \forall\,(t,\mu) \in [0,T] \times \mathcal{P}_2(C([0,T];H)), \tag{3.10}$$

for any $\xi \in \mathbf{S}_2(\mathbb{F})$ with $\mathbb{P}_\xi = \mu$. By Theorem 3.4 we immediately deduce the dynamic programming principle for v.

Corollary 3.9. *Suppose that* $(\mathbf{A}_{A,b,\sigma})$, $(\mathbf{A}_{f,g})$, $(\mathbf{A}_{f,g})_{\mathrm{cont}}$ *hold. The value function v satisfies the **dynamic programming principle**: for every $t,s \in [0,T]$, with $t \leq s$, and every $\mu \in \mathcal{P}_2(C([0,T];H))$ it holds that*

$$v(t,\mu) = \sup_{\alpha \in \mathcal{U}} \left\{ \mathbb{E}\left[\int_t^s f_r\big(X^{t,\xi,\alpha}, \mathbb{P}_{X^{t,\xi,\alpha}_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\,dr \right] + v\big(s, \mathbb{P}_{X^{t,\xi,\alpha}}\big) \right\},$$

for any $\xi \in \mathbf{S}_2(\mathbb{F})$ with $\mathbb{P}_\xi = \mu$.

Remark 3.10. *Recall from Remark 3.1 that V is non-anticipative, namely $V(t,\xi) = V(t,\xi_{\cdot \wedge t})$, for every $(t,\xi) \in [0,T] \times \mathbf{S}_2(\mathbb{F})$. As a consequence, the value function v satisfies the following non-anticipativity property:*

$$v(t,\mu) = v(t,\mu_{[0,t]}),$$

for every $(t,\mu) \in [0,T] \times C([0,T];H)$, where we denote by $\mu_{[0,t]}$ the measure $\mu \circ \big((x_s)_{s\in[0,T]} \mapsto (x_{s\wedge t})_{s\in[0,T]}\big)^{-1}$.

4 Pathwise derivatives in the Wasserstein space and Itô's formula

This section is devoted to the proof of Itô's formula for a real-valued function φ defined on $[0,T] \times \mathcal{P}_2(C([0,T];H))$. Such a formula involves the so-called pathwise derivatives in the Wasserstein space that we now define. In the present section we substantially follow [73, Section 2] (see also [72]), extending their framework to our more general setting with H being a real separable Hilbert space (not necessarily a Euclidean space as in [73]).

4.1 Notations

In order to define the pathwise derivatives, we need to extend the canonical space $C([0,T];H)$ to the space of càdlàg paths $D([0,T];H)$, which we endow with the Skorokhod topology (in what follows, we denote paths in $D([0,T];H)$ using $\hat{\cdot}$ in order to distinguish them from paths in $C([0,T];H)$; we do the same with other mathematical objects). In order to introduce the Skorohod topology, we

follow [15] and introduce the following metric on $D([0,T];H)$ (which corresponds to metric d° in [15, formula (12.16)]):

$$d_{\mathrm{Sk}}(\hat{x},\hat{y}) := \inf_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \left\{ \max\left(\sup_{\substack{s<t\\s,t\in[0,T]}} \left|\log\frac{\boldsymbol{\lambda}(t)-\boldsymbol{\lambda}(s)}{t-s}\right|, \sup_{t\in[0,T]}\left|\hat{x}_t - \hat{y}_{\boldsymbol{\lambda}(t)}\right|_H\right)\right\},$$

where $\boldsymbol{\Lambda}$ denotes the set of strictly increasing and continuous maps $\boldsymbol{\lambda}\colon [0,T]\to[0,T]$, satisfying $\boldsymbol{\lambda}(0)=0$ and $\boldsymbol{\lambda}(T)=T$. We recall from [15, Theorem 12.2] (in [15] the case $H=\mathbb{R}$ is considered, however the same proof applies to the general case of a real separable Hilbert space H) that $(D([0,T];H), d_{\mathrm{Sk}})$ is a Polish space. Moreover, d_{Sk} induces on $C([0,T];H)$ the topology of the uniform convergence (see [15], p. 124), and $C([0,T];H)$ is a Borel subset of $\big(D([0,T];H), d_{\mathrm{Sk}}\big)$ (apply e.g. result number (2) at p. 67 in [37], with $C=X=\big(C([0,T];H),\|\cdot\|_T\big)$, $Y=\big(D([0,T],H),d_{\mathrm{Sk}}\big)$, and as f take the canononical embedding). We define the spaces

$$\mathscr{H} := [0,T]\times\mathcal{P}_2\big(C([0,T];H)\big), \qquad \hat{\mathscr{H}} := [0,T]\times\mathcal{P}_2\big(D([0,T];H)\big).$$

For every $(t,\mu)\in\mathscr{H}$, we denote by $\mu_{[0,t]}$ the measure $\mu\circ\big((x_s)_{s\in[0,T]}\mapsto(x_{s\wedge t})_{s\in[0,T]}\big)^{-1}$. We define similarly $\hat{\mu}_{[0,t]}$, for every $(t,\hat{\mu})\in\hat{\mathscr{H}}$. We then equip $\mathscr{H}$ and $\hat{\mathscr{H}}$ with the following pseudo-distances, respectively:

$$d_{\mathscr{H}}\big((t,\mu),(t',\mu')\big) := \left(|t-t'|^2 + \mathcal{W}_2\big(\mu_{[0,t]},\mu'_{[0,t']}\big)^2\right)^{\frac{1}{2}}, \qquad (t,\mu),(t',\mu')\in\mathscr{H},$$

$$d_{\hat{\mathscr{H}}}\big((t,\hat{\mu}),(t',\hat{\mu}')\big) := \left(|t-t'|^2 + \widehat{\mathcal{W}}_2\big(\hat{\mu}_{[0,t]},\hat{\mu}'_{[0,t']}\big)^2\right)^{\frac{1}{2}}, \qquad (t,\hat{\mu}),(t',\hat{\mu}')\in\hat{\mathscr{H}},$$

where $\widehat{\mathcal{W}}_2$ is defined as (denoting $\mathrm{D}:=D([0,T];H)$)

$$\widehat{\mathcal{W}}_2(\hat{\mu},\hat{\mu}') := \inf\left\{ \int_{\mathrm{D}\times\mathrm{D}} d_{\mathrm{Sk}}(\hat{x},\hat{y})^2\,\hat{\pi}(d\hat{x},d\hat{y})\colon \hat{\pi}\in\mathcal{P}(\mathrm{D}\times\mathrm{D})\right.$$
$$\left. \text{such that } \hat{\pi}(\cdot\times\mathrm{D})=\hat{\mu} \text{ and } \hat{\pi}(\mathrm{D}\times\cdot)=\hat{\mu}'\right\}^{\frac{1}{2}},$$

for every $\hat{\mu},\hat{\mu}'\in\mathcal{P}_2(D([0,T];H))$.

Remark 4.1. *Notice that a function $\varphi\colon\mathscr{H}\to\mathbb{R}$ (resp. $\hat{\varphi}\colon\hat{\mathscr{H}}\to\mathbb{R}$) that is measurable with respect to $d_{\mathscr{H}}$ (resp. $d_{\hat{\mathscr{H}}}$) satisfies the* non-anticipativity *property:*

$$\varphi(t,\mu)=\varphi(t,\mu_{[0,t]}), \quad \forall\,(t,\mu)\in\mathscr{H} \qquad \left(\text{resp. } \hat{\varphi}(t,\hat{\mu})=\hat{\varphi}(t,\hat{\mu}_{[0,t]}), \quad \forall\,(t,\hat{\mu})\in\hat{\mathscr{H}}\right).$$

Remark 4.2. *Notice that there is a natural injection*

$$i_0 : \mathcal{P}_2\big(C([0,T];H)\big)\to\mathcal{P}_2\big(D([0,T];H)\big), \qquad \mu\mapsto i_0(\mu)$$

where $i_0(\mu)(E)=\mu(E\cap C([0,T];H))$ for all E Borel subset of $D([0,T];H)$. This induces an injection $i=\mathscr{H}\to\hat{\mathscr{H}}$ given by $i(t,\mu)=(t,i_0(\mu))$. We claim that the restriction of $d_{\hat{\mathscr{H}}}$ to $i(\mathscr{H})\times i(\mathscr{H})$ gives rise to the same topology on $i(\mathscr{H})$ induced by $d_{\mathscr{H}}$ through the injection i. Indeed, this is a

consequence of the following property. Let $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_2(C([0,T];H))$ and $\mu \in \mathcal{P}_2(C([0,T];H))$.
Then, denoting $\mathrm{C} := C([0,T];H)$, it holds that

$$\mathcal{W}_2(\mu_n, \mu) := \inf\left\{ \int_{\mathrm{C}\times\mathrm{C}} \|x - y\|_T^2\, \pi(dx, dy)\colon \pi \in \mathcal{P}(\mathrm{C}\times\mathrm{C}),\ \pi(\cdot \times \mathrm{C}) = \mu_n,\ \pi(\mathrm{C}\times\cdot) = \mu \right\}^{\frac{1}{2}} \overset{n\to\infty}{\longrightarrow} 0$$

if and only if (denoting $\mathrm{D} := D([0,T];H)$)

$$\widehat{\mathcal{W}}_2(i_0(\mu_n), i_0(\mu))$$

$$:= \inf\left\{ \int_{\mathrm{D}\times\mathrm{D}} d_{\mathrm{Sk}}(\hat{x}, \hat{y})^2\, \hat{\pi}(d\hat{x}, d\hat{y})\colon \hat{\pi} \in \mathcal{P}(\mathrm{D}\times\mathrm{D}),\ \hat{\pi}(\cdot \times \mathrm{D}) = i_0(\mu_n),\ \hat{\pi}(\mathrm{D}\times\cdot) = i_0(\mu) \right\}^{\frac{1}{2}} \overset{n\to\infty}{\longrightarrow} 0.$$

To prove the latter equivalence we first observe that the above probability measure $\hat{\pi}$ with marginals $i_0(\mu_n)$ and $i_0(\mu)$ satisfies $\hat{\pi}(\mathrm{C}\times\mathrm{C}) = 1$. Then, the following equality holds:

$$\widehat{\mathcal{W}}_2(i_0(\mu_n), i_0(\mu)) = \inf\left\{ \int_{\mathrm{C}\times\mathrm{C}} d_{\mathrm{Sk}}(x, y)^2\, \pi(dx, dy)\colon \pi \in \mathcal{P}(\mathrm{C}\times\mathrm{C}),\ \pi(\cdot \times \mathrm{C}) = \mu_n,\ \pi(\mathrm{C}\times\cdot) = \mu \right\}^{\frac{1}{2}}.$$

Now, we recall from [71, Definition 6.8 and Theorem 6.9]) that $\lim_{n\to\infty} \mathcal{W}_2(\mu_n, \mu) = 0$ if and only if for all continuous functions $\varphi\colon (C([0,T];H), \|\cdot\|_T) \to \mathbb{R}$ with $|\varphi(x)| \leq c(1 + \|x\|_T^2)$, $c \in \mathbb{R}$, one has

$$\int_{\mathrm{C}} \varphi(x)\, \mu_n(dx) \overset{n\to\infty}{\longrightarrow} \int_{\mathrm{C}} \varphi(x)\, \mu(dx). \tag{4.1}$$

Similarly, since $(C([0,T];H), d_{\mathrm{Sk}})$ is a Radon separable metric space[5] (see [3, Definition 5.1.4]), from the proof of Proposition 7.1.5. in [3] we deduce that $\lim_{n\to\infty} \widehat{\mathcal{W}}_2(i_0(\mu_n), i_0(\mu)) = 0$ if and only if for all continuous functions $\varphi\colon (C([0,T];H), d_{\mathrm{Sk}}) \to \mathbb{R}$ with $|\varphi(x)| \leq c(1 + d_{\mathrm{Sk}}(x,0)^2)$, $c \in \mathbb{R}$, one has

$$\int_{\mathrm{C}} \varphi(x)\, \mu_n(dx) \overset{n\to\infty}{\longrightarrow} \int_{\mathrm{C}} \varphi(x)\, \mu(dx). \tag{4.2}$$

Hence, recalling that the Skorohod topology relativized to $C([0,T];H)$ coincides with the uniform topology, it follows that the set of real-valued continuous functions on $(C([0,T];H), \|\cdot\|_T)$ coincides with the set of real-valued continuous on $(C([0,T];H), d_{\mathrm{Sk}})$. Finally, concerning the subquadratic growth, it holds that $d_{\mathrm{Sk}}(x,0) = \|x\|_T$, which shows that the class of functions involved in (4.1) and (4.2) are the same.

We also introduce the lifted spaces

$$\mathfrak{H} := [0,T] \times L^2\big(\Omega; C([0,T];H)\big), \qquad \hat{\mathfrak{H}} := [0,T] \times L^2\big(\Omega; D([0,T];H)\big).$$

Remark 4.3. *To alleviate notation, in the present Section 4 we work on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ adopted in the rest of the paper. Notice however that, for the definition of pathwise derivatives, $(\Omega, \mathcal{F}, \mathbb{P})$ can be replaced by any other probability space which supports a random variable having uniform distribution on $[0,1]$. See also Remark 2.2.*

Finally, we introduce the following notation.

---

[5]This comes from the fact that such a space is topologically equivalent to the Polish space $(C([0,T];H), \|\cdot\|_T)$.

Notation (Ntn$_{\hat{\mathbb{P}}}$). *For every $\xi \in L^2(\Omega; C([0,T]; H))$, we denote by $\hat{\mathbb{P}}_\xi$ the law of ξ on $D([0,T]; H)$, while we recall that $\mathbb{P}_\xi$ denotes the law of ξ on $C([0,T]; H)$. So, in particular, $\hat{\mathbb{P}}_\xi \in \mathcal{P}_2(D([0,T]; H))$, while $\mathbb{P}_\xi \in \mathcal{P}_2(C([0,T]; H))$. Clearly, it holds that $\hat{\mathbb{P}}_\xi(\mathfrak{B}) = \mathbb{P}_\xi(\mathfrak{B})$, for every Borel subset $\mathfrak{B}$ of $C([0,T]; H)$. Notice that in [73] the probability $\hat{\mathbb{P}}_\xi$ is denoted simply by $\mathbb{P}_\xi$ (see the beginning of Section 2.4 in [73]).*

4.2 Pathwise derivatives for a map $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ and Itô's formula

We start with the definition of pathwise time derivative for a map $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$.

Definition 4.4. *Let $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ be a non-anticipative function. We say that $\hat{\varphi}$ is pathwise differentiable in time at $(t, \hat{\mu}) \in \hat{\mathscr{H}}$, with $t < T$, if the following limit exists and is finite:*

$$\partial_t \hat{\varphi}(t, \hat{\mu}) := \lim_{\delta \to 0^+} \frac{\hat{\varphi}(t + \delta, \hat{\mu}_{[0,t]}) - \hat{\varphi}(t, \hat{\mu})}{\delta}.$$

At time $t = T$, we define

$$\partial_t \hat{\varphi}(t, \hat{\mu}) := \lim_{t \to T^-} \partial_t \hat{\varphi}(t, \hat{\mu}),$$

*when the limit exists and is finite. We refer to $\partial_t \hat{\varphi}$ as the **pathwise time derivative (or horizontal derivative) of $\hat{\varphi}$ at** $(t, \hat{\mu})$. If $\partial_t \hat{\varphi}$ exists everywhere as a function $\hat{\mathscr{H}} \to \mathbb{R}$, we refer to it as the **pathwise time derivative of** $\hat{\varphi}$.*

Remark 4.5. *Notice that $\partial_t \hat{\varphi}$ is a non-anticipative function, namely $\partial_t \hat{\varphi}(t, \hat{\mu}) = \partial_t \hat{\varphi}(t, \hat{\mu}_{[0,t]})$, for every $(t, \hat{\mu}) \in \hat{\mathscr{H}}$.*

In order to define the pathwise measure derivative, we need to consider the lifting of $\hat{\varphi}$.

Definition 4.6. *Given $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$, we say that $\hat{\Phi} \colon \hat{\mathfrak{H}} \to \mathbb{R}$ is a **lifting** of $\hat{\varphi}$ if*

$$\hat{\Phi}(t, \hat{\xi}) = \hat{\varphi}(t, \mathbb{P}_{\hat{\xi}}), \qquad \forall\, (t, \hat{\xi}) \in \hat{\mathfrak{H}},$$

where we recall that $\mathbb{P}_{\hat{\xi}}$ stands for the law of the random variable $\hat{\xi} \in L^2(\Omega; D([0,T]; H))$.

Definition 4.7. *Let $\hat{\Phi} \colon \hat{\mathfrak{H}} \to \mathbb{R}$ be a non-anticipative function, namely $\hat{\Phi}(t, \hat{\xi}) = \hat{\Phi}(t, \hat{\xi}_{\cdot \wedge t})$, for every $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$. We say that $\hat{\Phi}$ is **pathwise differentiable in space** at $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$ if there exists $D\hat{\Phi}(t, \hat{\xi}) \in L^2(\Omega; H)$ such that*

$$\lim_{Y \to 0} \frac{\left| \hat{\Phi}(t, \hat{\xi} + Y \mathbb{1}_{[t,T]}) - \hat{\Phi}(t, \hat{\xi}) - \mathbb{E}\left[\langle D\hat{\Phi}(t, \hat{\xi}), Y \rangle_H \right] \right|}{|Y|_{L^2(\Omega; H)}} = 0.$$

*We refer to $D\hat{\Phi}(t, \hat{\xi})$ as the **pathwise space derivative (or vertical derivative) of $\hat{\Phi}$ at** $(t, \hat{\xi})$. If $D\hat{\Phi}$ exists everywhere as a function $\hat{\mathfrak{H}} \to L^2(\Omega; H)$, we refer to it as the **pathwise space derivative of** $\hat{\varphi}$.*

Remark 4.8. *Notice that, if $\hat{\Phi}$ is pathwise differentiable in space at $(t, \hat{\xi})$, then it is pathwise differentiable in space at $(t, \hat{\xi}')$ for every $\hat{\xi}' \in L^2(\Omega; D([0,T]; H))$ such that $\hat{\xi}_{t \wedge} = \hat{\xi}'_{t \wedge}$ $\mathbb{P}$-a.s., and, in such a case, $D\hat{\Phi}(t, \hat{\xi}) = D\hat{\Phi}(t, \hat{\xi}')$ $\mathbb{P}$-a.s.*

23

We can then give, similarly to [22, Definition 5.22], the following definition.

Definition 4.9. *Let* $\hat{\varphi}\colon \hat{\mathscr{H}} \to \mathbb{R}$ *be a non-anticipative function and* $(t, \hat{\mu}) \in \hat{\mathscr{H}}$. *We say that* $\hat{\varphi}$ *is* **pathwise differentiable in measure** *at* $(t, \hat{\mu})$ *if its lifting* $\hat{\Phi}$ *is pathwise differentiable in space at some* $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$ *such that* $\mathbb{P}_{\hat{\xi}} = \hat{\mu}$.

Moreover, we say that $\hat{\varphi}$ *admits* **pathwise measure derivative** *at* $(t, \hat{\mu})$ *if its lifting* $\hat{\Phi}$ *is pathwise differentiable in space at every* $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$ *such that* $\mathbb{P}_{\hat{\xi}} = \hat{\mu}$ *and if there exists a masurable function* $\hat{g}\colon D([0,T];H) \to H$ *such that, for all* $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$ *with* $\mathbb{P}_{\hat{\xi}} = \hat{\mu}$,

$$D\hat{\Phi}(t, \hat{\xi}) = \hat{g}(\hat{\xi}) \qquad \mathbb{P}\text{-}a.s. \tag{4.3}$$

The map $\hat{g}$ *(which is* $\hat{\mu}$-*a.s. uniquely determined) is called* **pathwise measure derivative of** $\hat{\varphi}$ **at** $(t, \hat{\mu})$[6].

Finally, if $\hat{\varphi}$ *admits pathwise measure derivative at every* $(t, \hat{\mu}) \in \hat{\mathfrak{H}}$, *then the function*

$$\partial_\mu \hat{\varphi}\colon \hat{\mathscr{H}} \times D([0,T];H) \longrightarrow H, \qquad (t, \hat{\mu}, \hat{x}) \longmapsto \partial_\mu \hat{\varphi}(t, \hat{\mu}, \hat{x})$$

such that, for every $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$, $\partial_\mu \hat{\varphi}(t, \mathbb{P}_{\hat{\xi}}, \cdot)$ *is measurable and* $D\hat{\Phi}(t, \hat{\xi}) = \partial_\mu \hat{\varphi}(t, \mathbb{P}_{\hat{\xi}}, \hat{\xi})$ $\mathbb{P}$-*a.s., is called the* **pathwise measure derivative of** $\hat{\varphi}$[7].

In the following lemma we state, under general conditions, the existence of the pathwise measure derivative. In order to do it, we proceed similarly to what is done in [73, Section 2.3] (see also [22, Section 5.3]) in the finite-dimensional case. The proof of the lemma is postponed in Appendix C.

Lemma 4.10. *Fix* $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$. *Let* $\hat{\varphi}\colon \hat{\mathscr{H}} \to \mathbb{R}$ *be such that its lifting* $\hat{\Phi}$ *admits a continuous pathwise space derivative* $D\hat{\Phi}$ *on the set* $\{t\} \times \mathcal{O}_{\hat{\xi}}$, *where* $\mathcal{O}_{\hat{\xi}}$ *is a neighborhood of* $\hat{\xi}$ *in* $L^2(\Omega; D([0,T];H))$. *Then, there exists a measurable function* $\hat{g}\colon D([0,T];H) \to H$ *and*

$$D\hat{\Phi}(t, \hat{\xi}) = \hat{g}(\hat{\xi}), \qquad \mathbb{P}\text{-}a.s. \tag{4.4}$$

Let $\hat{\xi}'$ *be such that* $\mathbb{P}_{\hat{\xi}'} = \mathbb{P}_{\hat{\xi}}$. *If in addition* $\hat{\Phi}$ *admits a continuous pathwise space derivative on the set* $\{t\} \times \mathcal{O}_{\hat{\xi}'}$, *with* $\mathcal{O}_{\hat{\xi}'}$ *being a neighborhood of* $\hat{\xi}'$, *then* (4.4) *holds true with* $\hat{\xi}$ *replaced by* $\hat{\xi}'$.

Hence, if the pathwise space derivative $D\hat{\Phi}(t, \hat{\xi})$ *at* $(t, \hat{\xi})$ *exists for every* $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$ *and if* $D\hat{\Phi}$ *is continuous, then there exists the pathwise measure derivative* $\partial_\mu \hat{\varphi}$ *of* $\hat{\varphi}$.

If in addition the map $\partial_\mu \hat{\varphi}(t, \cdot, \cdot)\colon \mathcal{P}_2(D([0,T];H)) \times D([0,T];H) \to \mathbb{R}$ *is continuous for every* $t \in [0,T]$, *then* $\partial_\mu \hat{\varphi}$ *is uniquely defined.*

Finally, assume that the pathwise space derivative $D\hat{\Phi}$ *exists everywhere and is uniformly continuous. Then* $\partial_\mu \hat{\varphi}$ *is measurable.*

Remark 4.11 (**Non-anticipativity property of** $\partial_\mu \hat{\varphi}$.). *Let* $\hat{\varphi}\colon \hat{\mathscr{H}} \to \mathbb{R}$ *be a non-anticipative function. Suppose that there exists the pathwise measure derivative of* $\hat{\varphi}$. *Then, thanks to Remark 4.8 and equality* (4.3), $\partial_\mu \hat{\varphi}$ *is a non-anticipative function in the sense that, for every* $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$,

$$\partial_\mu \hat{\varphi}(t, \mathbb{P}_{\hat{\xi}})(\hat{\xi}) = \partial_\mu \hat{\varphi}(t, (\mathbb{P}_{\hat{\xi}})_{[0,t]})(\hat{\xi}_{\cdot \wedge t}) \qquad \mathbb{P}\text{-}a.s.$$

---

[6]Notice that, if a function is pathwise differentiable in measure at some point and the related pathwise space derivative is continuous (at least in a neighborhood), then it admits the pathwise measure derivative at that point, see e.g. [21, Theorem 6.5] or [23, Proposition 5.25] in finite dimension and our Lemma 4.10. Without the continuity assumption of the pathwise space derivative such a result is not obvious. See also [45].

[7]We stress the fact that $\partial_\mu \hat{\varphi}(t, \mathbb{P}_{\hat{\xi}}, \cdot)$ is *a-priori* uniquely determined only $\mathbb{P}_{\hat{\xi}}$-a.s.

or, equivalently,

$$\partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x}) = \partial_\mu \hat{\varphi}(t, \hat{\mu}_{[0,t]})(\hat{x}_{\cdot \wedge t}), \qquad \hat{\mu}\text{-}a.e.$$

for every $(t, \hat{\mu}, \hat{x}) \in \hat{\mathscr{H}} \times D([0,T]; H)$.

Finally, we define the pathwise derivative of second-order $\partial_x \partial_\mu \hat{\varphi}$.

Definition 4.12. *Let $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ be a non-anticipative function and $(t, \hat{\mu}) \in \hat{\mathscr{H}}$. Suppose that:*

1) there exists the pathwise measure derivative $\partial_\mu \hat{\varphi}$;

2) for every $t \in [0,T]$, the map $\partial_\mu \hat{\varphi}(t, \cdot)(\cdot) \colon \mathcal{P}_2(D([0,T]; H)) \times D([0,T]; H) \to H$ is continuous (hence, by Lemma 4.10, $\partial_\mu \hat{\varphi}$ is uniquely determined).

*Given $\hat{x} \in D([0,T]; H)$, we say that $\hat{\varphi}$ is **pathwise differentiable in measure and space at** $(t, \hat{\mu}, \hat{x})$ if there exists an operator $\partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x}) \in \mathcal{L}(H)$ such that*

$$\lim_{h \to 0} \frac{\left| \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x} + h \, \mathbb{1}_{[t,T]}) - \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x}) - \partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x}) h \right|_H}{|h|_H} = 0.$$

*We refer to $\partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x})$ as the **second-order pathwise derivative in measure and space of** $\hat{\varphi}$ **at** $(t, \hat{\mu}, \hat{x})$.*

*If $\partial_x \partial_\mu \hat{\varphi}$ exists everywhere as function $\hat{\mathscr{H}} \times D([0,T]; H) \to \mathcal{L}(H)$, we refer to it as the **pathwise derivative in measure and space of** $\hat{\varphi}$.*

Remark 4.13. *Recalling Remark 4.11, we see that $\partial_x \partial_\mu \hat{\varphi}$ is a non-anticipative function, namely it holds that $\partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{\xi}) = \partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu}_{[0,t]})(\hat{\xi}_{\cdot \wedge t})$, $\mathbb{P}$-a.s., for every $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$, with $\mathbb{P}_{\hat{\xi}} = \hat{\mu}$, or, equivalently,*

$$\partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x}) = \partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu}_{[0,t]})(\hat{x}_{\cdot \wedge t}), \qquad \hat{\mu}\text{-}a.e.$$

for every $(t, \hat{\mu}, \hat{x}) \in \hat{\mathscr{H}} \times D([0,T]; H)$.

Definition 4.14. *We denote by $\boldsymbol{C}^{1,2}(\hat{\mathscr{H}})$ the set of non-anticipative functions $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ such that:*

1) *the lifting $\hat{\Phi}$ of $\hat{\varphi}$ admits a continuous pathwise space derivative $D\hat{\Phi}$ on $\hat{\mathfrak{H}}$ (hence, by Lemma 4.10, there exists the pathwise measure derivative $\partial_\mu \hat{\varphi}$);*

2) *$\hat{\varphi}, \partial_\mu \hat{\varphi}$ are continuous;*

3) *there exist the pathwise time derivative $\partial_t \hat{\varphi}$, the second-order pathwise derivative in measure and space $\partial_x \partial_\mu \hat{\varphi}$, and $\partial_t \hat{\varphi}, \partial_x \partial_\mu \hat{\varphi}$ are continuous.*

Definition 4.15. *We denote by $\boldsymbol{C}_b^{1,2}(\hat{\mathscr{H}})$ the set of $\hat{\varphi} \in \boldsymbol{C}^{1,2}(\hat{\mathscr{H}})$ such that $\hat{\varphi}, \partial_t \hat{\varphi}, \partial_\mu \hat{\varphi}, \partial_x \partial_\mu \hat{\varphi}$ are bounded.*

We end this section with the Itô formula. The proof is postponed in Appendix D.

Theorem 4.16. *Fix $t \in [0, T]$ and let $\xi \in \mathbf{S}_2(\mathbb{F})$. Let also $F \colon [0, T] \times \Omega \to H$, $G \colon [0, T] \times \Omega \to \mathcal{L}_2(K; H)$ be square-integrable and $\mathbb{F}$-progressively measurable processes, so, in particular,*

$$\int_0^T \mathbb{E}[|F_s|_H^2]\, ds < \infty, \qquad \int_0^T \mathbb{E}\big[\mathrm{Tr}(G_s G_s^*)\big]\, ds < \infty.$$

Consider the process $X = (X_s)_{s \in [0,T]}$ given by[8]

$$X_s = \xi_{s \wedge t} + \int_t^{s \vee t} F_r\, dr + \int_t^{s \vee t} G_r\, dB_r, \qquad \forall\, s \in [0, T]. \tag{4.5}$$

*If $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ is in $\boldsymbol{C}_b^{1,2}(\hat{\mathscr{H}})$, then the following **Itô formula** holds:*

$$\hat{\varphi}(s, \hat{\mathbb{P}}_{X_{\cdot \wedge s}}) = \hat{\varphi}(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}}) + \int_t^s \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})\, dr + \int_t^s \mathbb{E}\left[\langle F_r, \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr$$

$$+ \frac{1}{2} \int_t^s \mathbb{E}\left[\mathrm{Tr}\left(G_r G_r^* \partial_x \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\right)\right] dr, \tag{4.6}$$

for every $s \in [t, T]$ (for the definition of $\hat{\mathbb{P}}_{X_{\cdot \wedge r}}$ see $(\mathbf{Ntn}_{\hat{\mathbb{P}}})$).

Remark 4.17. *Proceeding along the same lines as in the proof of Theorem 4.16, it is possible to prove Itô's formula for a larger class of functions than $\boldsymbol{C}_b^{1,2}(\hat{\mathscr{H}})$, for example weakening the boundedness assumption of $\partial_\mu \hat{\varphi}$ by assuming linear growth with respect to $\hat{\xi}$.*

Remark 4.18. *Following [73], we notice that in the last term of Itô's formula (4.6) we can replace $\partial_x \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})$ by its symmetrization $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})$, where $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi} \colon \hat{\mathscr{H}} \times D([0, T]; H) \to \mathcal{L}(H)$ is defined as*

$$\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x}) \coloneqq \frac{1}{2}\Big(\partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x}) + \big(\partial_x \partial_\mu \hat{\varphi}(t, \hat{\mu})(\hat{x})\big)^*\Big), \tag{4.7}$$

for every $(t, \hat{\mu}, \hat{x}) \in \hat{\mathscr{H}} \times D([0, T]; H)$.

In the finite-dimensional case, it is proved in [23, Remark 5.98] that $\partial_x \partial_\mu \hat{\varphi}$ is already symmetric. Notice however that such a proof is quite involved and it is still an open problem to show that it remains valid in the present infinite-dimensional framework.

4.3 Pathwise derivatives for a map $\varphi \colon \mathscr{H} \to \mathbb{R}$ and Itô's formula

In the present section we use several times the notation $(\mathbf{Ntn}_{\hat{\mathbb{P}}})$, namely we use the superscript $\hat{\cdot}$ to denote the natural extension to $D([0, T]; H)$ of a probability measure on $C([0, T]; H)$.

Definition 4.19. *Given $\varphi \colon \mathscr{H} \to \mathbb{R}$ and a non-anticipative map $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$, we say that $\hat{\varphi}$ is **consistent** with φ if (for the definition of $\hat{\mathbb{P}}_\xi$ see $(\mathbf{Ntn}_{\hat{\mathbb{P}}})$)*

$$\varphi(t, \mathbb{P}_\xi) = \hat{\varphi}(t, \hat{\mathbb{P}}_\xi), \tag{4.8}$$

for every $(t, \xi) \in \mathfrak{H}$, with $\xi \in \mathbf{S}_2(\mathbb{F})$ (namely, $\xi \in L^2(\Omega; C([0, T]; H))$ and it is $\mathbb{F}$-progressively measurable).

---

[8]In what follows, we will always implicitly refer to Itô processes only by continuous versions.

Remark 4.20. *Notice that we can replace $\mathbf{S}_2(\mathbb{F})$ with $\mathbf{S}_2(\mathcal{G})$ in Definition 4.19, as a matter of fact the sets $\{\mathbb{P}_\xi \colon \xi \in \mathbf{S}_2(\mathbb{F})\}$ and $\{\mathbb{P}_\xi \colon \xi \in \mathbf{S}_2(\mathcal{G})\}$ are equal and coincide with $\mathcal{P}_2(C([0,T];H))$, as it follows from property $(\mathbf{A}_\mathcal{G})$-ii). This also shows that equality (4.8) characterizes φ in terms of $\hat{\varphi}$ for every pair $(t,\mu) \in \mathscr{H}$.*

Next result is crucial in order to define pathwise derivatives for a map $\varphi \colon \mathscr{H} \to \mathbb{R}$, as it states a consistency property for the pathwise derivatives themselves.

Lemma 4.21. *Let $\hat{\varphi}_1, \hat{\varphi}_2 \in \boldsymbol{C}_b^{1,2}(\hat{\mathscr{H}})$. If (for the definition of $\hat{\mathbb{P}}_\xi$ see $(\mathbf{Ntn}_{\hat{\mathbb{P}}})$)*

$$\hat{\varphi}_1(t, \hat{\mathbb{P}}_\xi) = \hat{\varphi}_2(t, \hat{\mathbb{P}}_\xi), \qquad \forall\, (t,\xi) \in \mathfrak{H}, \ with \ \xi \in \mathbf{S}_2(\mathbb{F}),$$

then

$$\partial_t \hat{\varphi}_1(t, \hat{\mathbb{P}}_\xi) = \partial_t \hat{\varphi}_2(t, \hat{\mathbb{P}}_\xi), \tag{4.9}$$

$$\partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_\xi)(\xi) = \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_\xi)(\xi), \qquad \mathbb{P}\text{-a.s.} \tag{4.10}$$

$$\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_\xi)(\xi) = \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_\xi)(\xi), \qquad \mathbb{P}\text{-a.s.} \tag{4.11}$$

for every $(t,\xi) \in \mathfrak{H}$, with $\xi \in \mathbf{S}_2(\mathbb{F})$, where $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}$ is defined by (4.7).

Proof. See Appendix E. $\qquad\qquad\square$

Remark 4.22. *We do not address here the consistency property of $\partial_x \partial_\mu \hat{\varphi}$ (for hints on its proof we refer to [73], see the paragraph just after Theorem 2.9), as Itô's formula (and hence the Hamilton-Jacobi-Bellman equation) only depends on $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}$, see Remark 4.18.*

Remark 4.23. *By the non-anticipativity property of the pathwise derivatives (see Remarks 4.5, 4.11, 4.13), it follows that equalities (4.9)-(4.10)-(4.11) hold if and only if*

$$\partial_t \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}}) = \partial_t \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}}),$$

$$\partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) = \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}), \qquad \mathbb{P}\text{-a.s.}$$

$$\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) = \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}), \qquad \mathbb{P}\text{-a.s.}$$

for every $(t,\xi) \in \mathfrak{H}$, with $\xi \in \mathbf{S}_2(\mathbb{F})$.

Using Lemma 4.21, we can now define the class $\boldsymbol{C}_b^{1,2}(\mathscr{H})$.

Definition 4.24. *We denote by $\boldsymbol{C}^{1,2}(\mathscr{H})$ (respectively $\boldsymbol{C}_b^{1,2}(\mathscr{H})$) the set of maps $\varphi \colon \mathscr{H} \to \mathbb{R}$ for which there exists $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ such that $\hat{\varphi}$ is consistent with φ and $\hat{\varphi} \in \boldsymbol{C}^{1,2}(\hat{\mathscr{H}})$ (respectively $\boldsymbol{C}_b^{1,2}(\hat{\mathscr{H}})$). Then, we define (for the definition of $\hat{\mathbb{P}}_\xi$ see $(\mathbf{Ntn}_{\hat{\mathbb{P}}})$)*

$$\partial_t \varphi(t, \mathbb{P}_\xi) := \partial_t \hat{\varphi}(t, \hat{\mathbb{P}}_\xi),$$

$$\partial_\mu \varphi(t, \mathbb{P}_\xi)(\cdot) := \partial_\mu \hat{\varphi}(t, \hat{\mathbb{P}}_\xi)(\cdot),$$

$$\partial_x^{\mathrm{sym}} \partial_\mu \varphi(t, \mathbb{P}_\xi)(\cdot) := \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}(t, \hat{\mathbb{P}}_\xi)(\cdot),$$

for every $(t,\xi) \in \mathfrak{H}$, with $\xi \in \mathbf{S}_2(\mathbb{F})$.

We can finally state the Itô formula.

Theorem 4.25. *Fix $t \in [0, T]$ and let $\xi \in \mathbf{S}_2(\mathbb{F})$. Let also $F \colon [0, T] \times \Omega \to H$, $G \colon [0, T] \times \Omega \to$*
$\mathcal{L}_2(K; H)$ be square integrable and $\mathbb{F}$-progressively measurable process, so, in particular,

$$\int_0^T \mathbb{E}[|F_s|_H^2] \, ds < \infty, \qquad \int_0^T \mathbb{E}\big[\mathrm{Tr}(G_s G_s^*)\big] \, ds < \infty.$$

Consider the process $X = (X_s)_{s \in [0, T]}$ given by

$$X_s = \xi_{s \wedge t} + \int_t^{s \vee t} F_r \, dr + \int_t^{s \vee t} G_r \, dB_r, \qquad \forall \, s \in [0, T].$$

*If $\varphi \colon \mathscr{H} \to \mathbb{R}$ is in $\boldsymbol{C}_b^{1,2}(\mathscr{H})$, then the following **Itô formula** holds:*

$$\varphi(s, \mathbb{P}_{X_{\cdot \wedge s}}) = \varphi(t, \mathbb{P}_{\xi_{\cdot \wedge t}}) + \int_t^s \partial_t \varphi(r, \mathbb{P}_{X_{\cdot \wedge r}}) \, dr + \int_t^s \mathbb{E}\left[\langle F_r, \partial_\mu \varphi(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr$$

$$+ \frac{1}{2} \int_t^s \mathbb{E}\left[\mathrm{Tr}\left(G_r G_r^* \partial_x \partial_\mu \varphi(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\right)\right] dr, \tag{4.12}$$

for every $s \in [t, T]$.

Proof. By Definition 4.24 there exists $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ such that $\hat{\varphi}$ is consistent with φ and $\hat{\varphi} \in$
$\boldsymbol{C}_b^{1,2}(\hat{\mathscr{H}})$. As a consequence, by Theorem 4.16 we have the Itô formula

$$\hat{\varphi}(s, \hat{\mathbb{P}}_{X_{\cdot \wedge s}}) = \hat{\varphi}(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}}) + \int_t^s \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}}) \, dr + \int_t^s \mathbb{E}\left[\langle F_r, \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr$$

$$+ \frac{1}{2} \int_t^s \mathbb{E}\left[\mathrm{Tr}\left(G_r G_r^* \partial_x \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\right)\right] dr,$$

for every $s \in [t, T]$. Using the fact that $\hat{\varphi}$ is consistent with φ and recalling the definition of pathwise
derivatives of φ (see Definition 4.24), we obtain the claimed Itô formula for φ. $\qquad \square$

In order to apply Itô's formula to our case we need the following variant of Theorem 4.25 (for a
similar result, see Proposition 1.165 in [35]).

Theorem 4.26. *Fix $t \in [0, T]$ and let $\xi \in \mathbf{S}_2(\mathbb{F})$ (namely, $\xi \in L^2(\Omega; C([0, T]; H))$ and it is $\mathbb{F}$-*
progressively measurable). Let A, b, σ be as in Assumption $(\mathbf{A}_{A, b, \sigma})$. Let $X = X^{t, \xi, \alpha}$ be the unique
mild solution of (2.4). Let $\varphi \colon \mathscr{H} \to \mathbb{R}$ belong to $\boldsymbol{C}_b^{1,2}(\mathscr{H})$. Assume also that, for all $(t, \mu, x) \in$
$\mathscr{H} \times C([0, T]; H)$, $\partial_\mu \varphi(t, \mu)(x) \in D(A^*)$ and that the map

$$\mathscr{H} \times C([0, T]; H) \longrightarrow H \qquad (t, \mu, x) \longmapsto A^* \partial \varphi_\mu(t, \mu)(x)$$

*is continuous and bounded [9]. Then the following variant of **Itô formula** holds:*

$$\varphi(s, \mathbb{P}_{X_{\cdot \wedge s}}) = \varphi(t, \mathbb{P}_{\xi_{\cdot \wedge t}}) + \int_t^s \partial_t \varphi(r, \mathbb{P}_{X_{\cdot \wedge r}}) \, dr + \int_t^s \mathbb{E}\left[\langle X_r, A^* \partial_\mu \varphi(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr$$

$$+ \int_t^s \mathbb{E}\left[\langle b_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right), \partial_\mu \varphi(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr \tag{4.13}$$

$$+ \frac{1}{2} \int_t^s \mathbb{E}\left[\mathrm{Tr}\left(\sigma_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right) \sigma_r^*\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right) \partial_x \partial_\mu \varphi(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\right)\right] dr,$$

for every $s \in [t, T]$.

---

[9]Indeed, in view of (2.5) here we could ask only linear growth of $\partial_\mu \varphi$ in x

Proof. The proof can be done proceeding along the same lines as in the proof of Proposition 1.165 in [35]. We provide a sketch of proof.

First of all if A is a bounded operator from Theorem 4.25 we immediately get (4.13). Now take A possibly unbounded and consider its Yosida approximations A_n, for $n \in \mathbb{N}$. Call, as in (2.7), X^n the solution of the state equation when A is replaced by A_n. Then, from (4.13) we get

$$
\varphi(s, \mathbb{P}_{X^n_{\cdot \wedge s}}) = \varphi(t, \mathbb{P}_{\xi_{\cdot \wedge t}}) + \int_t^s \partial_t \varphi(r, \mathbb{P}_{X^n_{\cdot \wedge r}}) \, dr + \int_t^s \mathbb{E}\left[\langle X^n_r, A_n^* \partial_\mu \varphi(r, \mathbb{P}_{X^n_{\cdot \wedge r}})(X^n_{\cdot \wedge r})\rangle_H\right] dr
$$

$$
+ \int_t^s \mathbb{E}\left[\langle b_r\left(X^n, \mathbb{P}_{X^n_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right), \partial_\mu \varphi(r, \mathbb{P}_{X^n_{\cdot \wedge r}})(X^n_{\cdot \wedge r})\rangle_H\right] dr
$$

$$
+ \frac{1}{2} \int_t^s \mathbb{E}\left[\mathrm{Tr}\left(\sigma_r\left(X^n, \mathbb{P}_{X^n_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right) \sigma_r^*\left(X^n, \mathbb{P}_{X^n_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right) \partial_x \partial_\mu \varphi(r, \mathbb{P}_{X^n_{\cdot \wedge r}})(X^n_{\cdot \wedge r})\right)\right] dr,
$$

for every $s \in [t, T]$. Now the convergence of all terms above follows applying, in a straightforward way, the result of Proposition 2.10. $\qquad\square$

5 Hamilton-Jacobi-Bellman equation

5.1 Viscosity properties of the value function

For every $t \in [0, T]$, we introduce the set

$$
\mathcal{M}_t := \left\{\mathfrak{a} \colon \Omega \to \mathrm{U} \colon \mathfrak{a} \text{ is } \mathcal{F}_t\text{-measurable}\right\}.
$$

Note that, since the filtration $\mathcal{F}_t$ is right-continuous, we have $\mathcal{M}_t = \cap_{\varepsilon > 0} \mathcal{M}_{t+\varepsilon}$. We now consider the following Hamilton-Jacobi-Bellman (HJB) equation:

$$
\begin{aligned}
0 = \; & \partial_t w(t, \mu) + \mathbb{E}\langle \xi_t, A^* \partial_\mu w(t, \mu)(\xi)\rangle_H \\
& + \sup_{\mathfrak{a} \in \mathcal{M}_t} \left\{ \mathbb{E}\left[f_t\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right) + \left\langle b_t\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right), \partial_\mu w(t, \mu)(\xi)\right\rangle_H\right] \right. \\
& \left. + \frac{1}{2} \mathbb{E}\left[\mathrm{Tr}\left(\sigma_t\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right) \sigma_t^*\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right) \partial_x \partial_\mu w(t, \mu)(\xi)\right)\right] \right\},
\end{aligned}
\tag{5.1}
$$

for $(t, \mu) \in \mathscr{H}$, $t < T$, $\xi \in \mathbf{S}_2(\mathcal{G})$ such that $\mathbb{P}_\xi = \mu$, with terminal condition

$$
w(T, \mu) = \mathbb{E}[g(\xi, \mu)], \qquad \text{for } \mu \in \mathcal{P}_2(C([0, T]; H)), \; \xi \in \mathbf{S}_2(\mathcal{G}) \text{ such that } \mathbb{P}_\xi = \mu. \tag{5.2}
$$

Definition 5.1. *We say that a function $w \colon \mathscr{H} \to \mathbb{R}$ belongs to the space $\boldsymbol{C}^{1,2}_{b, A^*}(\mathscr{H})$ if it satisfies the following regularity assumptions:*

(i) $w \colon \mathscr{H} \to \mathbb{R}$ belongs to $\boldsymbol{C}^{1,2}_b(\mathscr{H})$;

(ii) for all $(t, \mu, \xi) \in \mathscr{H} \times \mathbf{S}_2(\mathbb{F})$, $\partial_\mu \varphi(t, \mu)(\xi) \in L^2(\Omega; D(A^*))$ and the map

$$
\mathscr{H} \times \mathbf{S}_2(\mathbb{F}) \longrightarrow L^2(\Omega; H), \qquad (t, \mu, \xi) \longmapsto A^* \varphi(t, \mu)(\xi)
$$

is continuous and bounded.

Definition 5.2. *We say that a function $w \colon \mathscr{H} \to \mathbb{R}$ is a classical solution to the HJB equation* (5.1) *with terminal condition* (5.2)*, if it belongs to the space $\boldsymbol{C}^{1,2}_{b,A^*}(\mathscr{H})$ and satisfies* (5.1)-(5.2)*.*

Using Theorem 4.26 and Corollary 3.9 we are able to prove the following result.

Theorem 5.3. *Let Assumptions $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})_{\mathrm{cont}}$ hold. Assume also that b, σ, f are uniformly continuous in t, uniformly with respect to the other variables. Assume that the value function v (see* (3.10)*) belongs to the space $\boldsymbol{C}^{1,2}_{b,A^*}(\mathscr{H})$. Then v is a classical solution of* (5.1)-(5.2)*.*

Proof. From Corollary 3.9 we know that, for every $(t,\mu) \in \mathscr{H}$, for every $\xi \in \mathbf{S}_2(\mathbb{F})$ such $\mathbb{P}_\xi = \mu$, for every $\alpha \in \mathcal{U}$, and for every $h > 0$ sufficiently small,

$$0 = \sup_{\alpha \in \mathcal{U}} \left\{ \mathbb{E}\left[\frac{1}{h}\int_t^{t+h} f_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right) dr\right] + \frac{1}{h}\left[v\big(t+h, \mathbb{P}_X\big) - v(t,\mu)\right] \right\}, \tag{5.3}$$

where, for simplicity, we wrote simply X in place of $X^{t,\xi,\alpha}$. Now we use (4.13) getting

$$\frac{1}{h}\left[v\big(t+h, \mathbb{P}_{X^{t,\xi,\alpha}}\big) - v(t,\mu)\right] = \frac{1}{h}\int_t^{t+h} \partial_t v(r, \mathbb{P}_{X_{\cdot \wedge r}})\, dr + \frac{1}{h}\int_t^{t+h} \mathbb{E}\left[\langle X_r, A^*\partial_\mu v(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr$$

$$+ \frac{1}{h}\int_t^{t+h} \mathbb{E}\left[\langle b_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right), \partial_\mu v(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr \tag{5.4}$$

$$+ \frac{1}{2h}\int_t^{t+h} \mathbb{E}\left[\mathrm{Tr}\left(\sigma_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right)\sigma_r^*\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\right) \partial_x \partial_\mu v(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\right)\right] dr,$$

We now show, as in the typical proof of this result, the two inequalities. First take any $\mathfrak{a} \in \mathcal{M}_t$ and consider the control $\alpha \in \mathcal{U}$ defined as

$$\alpha_s = 0, \quad s \in [0,t), \qquad \alpha_s = \mathfrak{a}, \quad s \in [t,T]$$

Then, from (5.3) and (5.4) we get

$$0 \geq \frac{1}{h}\int_t^{t+h} \mathbb{E}f_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right) dr + \frac{1}{h}\int_t^{t+h} \partial_t v(r, \mathbb{P}_{X_{\cdot \wedge r}})\, dr$$

$$+ \frac{1}{h}\int_t^{t+h} \mathbb{E}\left[\langle X_r, A^*\partial_\mu v(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr$$

$$+ \frac{1}{h}\int_t^{t+h} \mathbb{E}\left[\langle b_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right), \partial_\mu v(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\rangle_H\right] dr \tag{5.5}$$

$$+ \frac{1}{2h}\int_t^{t+h} \mathbb{E}\left[\mathrm{Tr}\left(\sigma_r\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right)\sigma_r^*\left(X, \mathbb{P}_{X_{\cdot \wedge r}}, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right) \partial_x \partial_\mu v(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})\right)\right] dr,$$

Using the regularity of v and the continuity properties of b, σ, f we get, passing to the limit for $h \to 0^+$

$$0 \geq \partial_t v(t,\mu) + \mathbb{E}\langle \xi_t, A^*\partial_\mu v(t,\mu)(\xi)\rangle_H \mathbb{E}\left[f_t\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right) + \langle b_t\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right), \partial_\mu v(t,\mu)(\xi)\rangle_H\right]$$

$$+ \frac{1}{2}\mathbb{E}\left[\mathrm{Tr}\left(\sigma_t\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right)\sigma_t^*\left(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\right)\partial_x \partial_\mu v(t,\mu)(\xi)\right)\right] \tag{5.6}$$

Then the inequality $\geq$ follows by the arbitrariness of $\mathfrak{a}$. We now prove the opposite inequality. For every $\varepsilon > 0$ we take $\alpha^\varepsilon \in \mathcal{U}$ such that, denoting by X^ε the corresponding state trajectory,

$$-\varepsilon \leq \mathbb{E}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} f_r\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha^\varepsilon_r, \mathbb{P}_{\alpha^\varepsilon_r}\big)\,dr\right] + \frac{1}{\varepsilon}\Big[v\big(t+\varepsilon, \mathbb{P}_{X^\varepsilon}\big) - v(t,\mu)\Big]\bigg\}, \tag{5.7}$$

Now we apply Ito's formula (5.4) above, for $h = \varepsilon$, getting

$$-\varepsilon \leq \mathbb{E}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} f_r\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha^\varepsilon_r, \mathbb{P}_{\alpha^\varepsilon_r}\big)\,dr\right] + \frac{1}{\varepsilon}\int_t^{t+\varepsilon} \partial_t v(r, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}})\,dr$$

$$+ \frac{1}{\varepsilon}\int_t^{t+\varepsilon} \mathbb{E}\left[\langle X^\varepsilon_r, A^*\partial_\mu v(r, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}})(X^\varepsilon_{\cdot\wedge r})\rangle_H\right]dr$$

$$+ \frac{1}{\varepsilon}\int_t^{t+\varepsilon} \mathbb{E}\left[\langle b_r\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big), \partial_\mu v(r, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}})(X^\varepsilon_{\cdot\wedge r})\rangle_H\right]dr \tag{5.8}$$

$$+ \frac{1}{2\varepsilon}\int_t^{t+\varepsilon} \mathbb{E}\left[\mathrm{Tr}\big(\sigma_r\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\sigma_r^*\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\partial_x\partial_\mu v(r, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}})(X^\varepsilon_{\cdot\wedge r})\big)\right]dr,$$

By Remark 2.9 we obtain that, as $\varepsilon \to 0$, $X^\varepsilon_{\cdot\wedge r} \to \xi_{\cdot\wedge t}$ and $\mathbb{P}_{X^\varepsilon_{\cdot\wedge r}} \to \mathbb{P}_{\xi_{\cdot\wedge t}}$, hence the second and third integrals of the above right-hand side converge to

$$\partial_t v(t,\mu) + \mathbb{E}\left[\langle \xi_t, A^*\partial_\mu v(t,\mu)(\xi)\rangle_H\right]$$

The remaining integrals of the right-hand side of (5.8) can be rewritten as

$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon}\bigg(\mathbb{E}\left[f_r\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha^\varepsilon_r, \mathbb{P}_{\alpha^\varepsilon_r}\big) + \langle b_r\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big), \partial_\mu v(r, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}})(X^\varepsilon_{\cdot\wedge r})\rangle_H\right] \tag{5.9}$$

$$\mathbb{E}\left[\mathrm{Tr}\big(\sigma_r\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\sigma_r^*\big(X^\varepsilon, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}}, \alpha_r, \mathbb{P}_{\alpha_r}\big)\partial_x\partial_\mu v(r, \mathbb{P}_{X^\varepsilon_{\cdot\wedge r}})(X^\varepsilon_{\cdot\wedge r})\big)\right]\bigg)dr$$

Recall now that, by our assumptions, b and σ are uniformly continuous in (t, x, μ) uniformly with respect to the other variables and that f is locally uniformly continuous in (x, μ) uniformly with respect to the other variables. Hence, using again Remark 2.9 we obtain that (5.9) can be rewritten as

$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon}\bigg(\mathbb{E}\left[f_t\big(\xi, \mu, \alpha^\varepsilon_r, \mathbb{P}_{\alpha^\varepsilon_r}\big) + \langle b_t\big(\xi, \mu, \alpha_r, \mathbb{P}_{\alpha_r}\big), \partial_\mu v(t,\mu)(\xi)\rangle_H\right] \tag{5.10}$$

$$\mathbb{E}\left[\mathrm{Tr}\big(\sigma_t\big(\xi, \mu, \alpha_r, \mathbb{P}_{\alpha_r}\big)\sigma_r^*\big(\xi, \mu, \alpha_r, \mathbb{P}_{\alpha_r}\big)\partial_x\partial_\mu v(t,\mu)(\xi)\big)\right]\bigg)dr + \rho(\varepsilon)$$

where $\rho(\varepsilon) \to 0$ as $\varepsilon \to 0$. It follows

$$0 \leq \varepsilon + \rho(\varepsilon) + \partial_t v(t,\mu) + \mathbb{E}\langle \xi_t, A^*\partial_\mu v(t,\mu)(\xi)\rangle_H$$

$$+ \sup_{\mathfrak{a}\in\mathcal{M}_{t+\varepsilon}}\bigg\{\mathbb{E}\left[f_t\big(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\big) + \langle b_t\big(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\big), \partial_\mu v(t,\mu)(\xi)\rangle_H\right] \tag{5.11}$$

$$+ \frac{1}{2}\mathbb{E}\left[\mathrm{Tr}\big(\sigma_t\big(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\big)\sigma_t^*\big(\xi, \mu, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\big)\partial_x\partial_\mu v(t,\mu)(\xi)\big)\right]\bigg\}$$

The conclusion then follows invoking the second part of Lemma F.2. $\qquad\square$

Now, we provide the definition of viscosity solution that we shall use.

Definition 5.4. *We say that a function $w\colon \mathscr{H} \to \mathbb{R}$ is a viscosity subsolution (respectively super-solution) to the HJB equation (5.1) with terminal condition (5.2), if:*

- *$w(T, \mu) \leq (\text{respectively} \geq) \mathbb{E}[g(\xi, \mu)]$, for $\mu \in \mathcal{P}_2(C([0,T]; H))$, $\xi \in \mathbf{S}_2(\mathcal{G})$ such that $\mathbb{P}_\xi = \mu$;*

- *for $(t, \mu) \in \mathscr{H}$ and for every test function $\varphi \in \boldsymbol{C}^{1,2}_{b,A^*}(\mathscr{H})$ such that $w - \varphi$ has a maximum at (t, μ) (with value 0), one has that (5.1)-(5.2) is satisfied with the inequality $\leq$ (respectively $\geq$) in place of the equality and with φ in place of w.*

Moreover, w is called a viscosity solution of (5.1)-(5.2) if it is both a viscosity subsolution and a viscosity supersolution.

Theorem 5.5. *Let Assumptions $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})_{\mathrm{cont}}$ hold. Assume also that b, σ, f are uniformly continuous in t, uniformly with respect to the other variables. Then, the value function v is a viscosity solution of (5.1)-(5.2).*

Proof. The proof follows exactly the same lines as in the proof of Theorem 5.3, simply replacing v with φ. $\square$

5.2 Alternative forms of the HJB equation

We derive alternative forms of the Hamilton-Jacobi-Bellman equation (5.1) relying on technical results reported in Appendix F. We first need to introduce the following sets:

- $\mathcal{M}_\mathcal{G}$ is the set
$$\mathcal{M}_\mathcal{G} := \big\{ \mathfrak{a}\colon \Omega \to \mathrm{U}\colon \mathfrak{a} \text{ is } \mathcal{G}\text{-measurable} \big\};$$

- $\check{\mathcal{M}}$ is the set of Borel-measurable maps $\check{\mathrm{a}}\colon C([0,T]; H) \times [0,1] \to \mathrm{U}$;

- $\mathcal{M}$ is the set of Borel-measurable maps $\mathrm{a}\colon C([0,T]; H) \to \mathrm{U}$.

The technical results reported in Appendix F provides the following key proposition.

Proposition 5.6. *Suppose that $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$ hold. Let $(t, \mu) \in \mathscr{H}$, $w \in \boldsymbol{C}^{1,2}_b(\mathscr{H})$, and define $F\colon C([0,T]; H) \times \mathrm{U} \times \mathcal{P}(\mathrm{U}) \to \mathbb{R}$ by*

$$F(x, u, \nu) := f_t(x, \mu, u, \nu) + \big\langle b_t(x, \mu, u, \nu), \partial_\mu w(t, \mu)(x) \big\rangle_H$$
$$+ \frac{1}{2} \mathrm{Tr}\Big(\sigma_t(x, \mu, u, \nu) \sigma_t^*(x, \mu, u, \nu) \partial_x \partial_\mu w(t, \mu)(x) \Big),$$

for every $(x, u, \nu) \in C([0,T]; H) \times \mathrm{U} \times \mathcal{P}(\mathrm{U})$. Let also $\xi \in \mathbf{S}_2(\mathcal{G})$ with $\mathbb{P}_\xi = \mu$.

1) *Suppose that ξ is such that there exists a $\mathcal{G}$-measurable random variable U_ξ having uniform distribution on $[0,1]$ and being independent of ξ (by Lemma F.1 we know that for each μ there exist at least one ξ, with $\mathbb{P}_\xi = \mu$, satisfying this property). Then, it holds that*

$$\sup_{\mathfrak{a} \in \mathcal{M}_t} \mathbb{E}\big[F\big(\xi, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\big) \big] = \sup_{\mathfrak{a} \in \mathcal{M}_\mathcal{G}} \mathbb{E}\big[F\big(\xi, \mathfrak{a}, \mathbb{P}_\mathfrak{a}\big) \big] = \sup_{\check{\mathrm{a}} \in \check{\mathcal{M}}} \mathbb{E}\big[F\big(\xi, \check{\mathrm{a}}(\xi, U_\xi), \mathbb{P}_{\check{\mathrm{a}}(\xi, U_\xi)}\big) \big]. \tag{5.12}$$

2) *Suppose that F does not depend on its last argument, namely $F = F(x, u)$. Then, it holds that*

$$\sup_{\mathfrak{a} \in \mathcal{M}_t} \mathbb{E}\big[F(\xi, \mathfrak{a})\big] = \sup_{\mathbf{a} \in \mathcal{M}} \mathbb{E}\big[F(\xi, \mathbf{a}(\xi))\big] = \mathbb{E}\Big[\operatorname*{ess\,sup}_{u \in \mathrm{U}} F(\xi, u)\Big]. \tag{5.13}$$

Proof. Equalities (5.12) are a direct consequence of Lemma F.2, while (5.13) follows directly from equalities (F.4) and (F.5) of Lemma F.3. □

Remark 5.7. *Notice that the requirement that F does not depend on ν in item 2) of Proposition 5.6 is necessary for the validity of the first equality in* (5.13) *(we do not consider the second equality in* (5.13) *in this case, as it is not clear how to write the last quantity in* (5.13) *when F depends also on ν). As a matter of fact, consider the following example.*

Example. *Take $\mathrm{U} = [0, 1]$, endowed with its Borel σ-algebra, and let F be given by*

$$F(x, u, \nu) = -\mathcal{W}_2(\nu, \lambda), \qquad \forall (x, u, \nu) \in C([0, T]; H) \times [0, 1] \times \mathcal{P}([0, 1]),$$

with λ being the Lebesgue measure on the unit interval. Let also ξ be constant and identically equal to some fixed path $\bar{x} \in C([0, T]; H)$. Moreover, denote by $U_{\mathcal{G}}$ a $\mathcal{G}$-measurable random variable having distribution λ, whose existence follows from Lemma 2.1. Then, for every $t \in [0, T]$,

$$\sup_{\mathfrak{a} \in \mathcal{M}_t} \big(-\mathcal{W}_2(\mathbb{P}_{\mathfrak{a}}, \lambda) \big) = 0$$

and the supremum is attained at $\mathfrak{a}^*$, where $\mathfrak{a}^* := U_{\mathcal{G}}$. On the other hand, if $\mathbf{a} \in \mathcal{M}$ then $\mathbf{a}(\xi)$ is equal to the constant $\mathbf{a}(\bar{x})$, so, in particular, $\mathbb{P}_{\mathbf{a}(\xi)} = \delta_{\mathbf{a}(\bar{x})}$. This implies that

$$\sup_{\mathbf{a} \in \mathcal{M}} \big(-\mathcal{W}_2(\mathbb{P}_{\mathbf{a}(\xi)}, \lambda) \big) = \sup_{c \in [0,1]} \big(-\mathcal{W}_2(\delta_c, \lambda) \big) = -\inf_{c \in [0,1]} \left(\int_0^1 |c - r|^2 \, dr \right)^{\frac{1}{2}}$$

$$= -\inf_{c \in [0,1]} \sqrt{c^2 - c + \frac{1}{3}} = -\frac{1}{12}.$$

Remark 5.8. *Suppose that $F = F(x, u)$ and define $F^*: C([0, T]; H) \to \mathbb{R} \cup \{+\infty\}$ as*

$$F^*(x) := \sup_{u \in \mathrm{U}} F(x, u), \qquad \forall x \in C([0, T]; H).$$

If F^* is measurable, then $\operatorname{ess\,sup}_{u \in \mathrm{U}} F(\xi, u) = F^*(\xi)$, $\mathbb{P}$-a.s., and the essential supremum appearing in (5.13) *can be replaced with the supremum, so that we obtain*

$$\sup_{\mathfrak{a} \in \mathcal{M}_t} \mathbb{E}\big[F(\xi, \mathfrak{a})\big] = \mathbb{E}\Big[\sup_{u \in \mathrm{U}} F(\xi, u)\Big].$$

Notice that, under assumptions $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$, it follows from Proposition 7.47 in [11] that F^* is lower semi-analytic (see Definition 7.21 in [11] for the definition of lower semi-analytic). However, we cannot in general say that F^* is measurable (see for instance the discussion at the end of Section B.5 in [11]). Sufficient conditions ensuring the measurability of F^* are given for instance in Proposition 7.32 of [11] and read as follows:

(a) *If $F: C([0, T]; H) \times \mathrm{U} \to \mathbb{R}$ is lower semi-continuous and U is compact, then F^* is lower semi-continuous.*

(b) *If $F\colon C([0,T];H) \times \mathrm{U} \to \mathbb{R}$ is upper semi-continuous, then F^* is upper semi-continuous.*

It follows from Proposition 5.6 that, under $(\mathbf{A}_{A,b,\sigma})$ and $(\mathbf{A}_{f,g})$, the Hamilton-Jacobi-Bellman equation (5.1) can also be written in the following two alternative forms:

$$
\begin{cases}
\partial_t w(t,\mu) + \mathbb{E}\langle \xi_t, A^*\partial_\mu w(t,\mu)(\xi)\rangle_H \\
+ \sup_{\mathfrak{a}\in\mathcal{M}_\mathcal{G}} \Big\{ \mathbb{E}\big[f_t\big(\xi,\mu,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big) + \big\langle b_t\big(\xi,\mu,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big), \partial_\mu w(t,\mu)(\xi)\big\rangle_H\big] \\
+ \dfrac{1}{2}\mathbb{E}\Big[\mathrm{Tr}\Big(\sigma_t\big(\xi,\mu,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big)\sigma_t^*\big(\xi,\mu,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big)\partial_x\partial_\mu w(t,\mu)(\xi)\Big)\Big]\Big\} = 0, & (t,\mu)\in\mathscr{H},\, t<T, \\
w(T,\mu) = \mathbb{E}[g(\xi,\mu)], & \mu\in\mathscr{P}_2(C([0,T];H)),
\end{cases}
\tag{5.14}
$$

or, alternatively,

$$
\begin{cases}
\partial_t w(t,\mu) + \mathbb{E}\langle \xi_t, A^*\partial_\mu w(t,\mu)(\xi)\rangle_H + \sup_{\breve{\mathrm{a}}\in\breve{\mathcal{M}}} \Big\{ \mathbb{E}\big[f_t\big(\xi,\mu,\breve{\mathrm{a}}(\xi,U_\xi),\mathbb{P}_{\breve{\mathrm{a}}(\xi,U_\xi)}\big) \\
+\big\langle b_t\big(\xi,\mu,\breve{\mathrm{a}}(\xi,U_\xi),\mathbb{P}_{\breve{\mathrm{a}}(\xi,U_\xi)}\big), \partial_\mu w(t,\mu)(\xi)\big\rangle_H\big] \\
+\dfrac{1}{2}\mathbb{E}\Big[\mathrm{Tr}\Big(\sigma_t\sigma_t^*\big(\xi,\mu,\breve{\mathrm{a}}(\xi,U_\xi),\mathbb{P}_{\breve{\mathrm{a}}(\xi,U_\xi)}\big)\partial_x\partial_\mu w(t,\mu)(\xi)\Big)\Big]\Big\} = 0, & (t,\mu)\in\mathscr{H},\, t<T, \\
w(T,\mu) = \mathbb{E}[g(\xi,\mu)], & \mu\in\mathscr{P}_2(C([0,T];H)),
\end{cases}
\tag{5.15}
$$

where, in both (5.14) and (5.15), $\xi \in \mathbf{S}_2(\mathcal{G})$, with $\mathbb{P}_\xi = \mu$, is such that there exists a $\mathcal{G}$-measurable random variable U_ξ having uniform distribution on $[0,1]$ and being independent of ξ (we recall that, by Lemma F.1, for each μ there exists at least one ξ, with $\mathbb{P}_\xi = \mu$, satisfying this latter property).

Now, suppose that $(\mathbf{A}_{A,b,\sigma})$, $(\mathbf{A}_{f,g})$ hold and also that the coefficients b, σ, f do not depend on their last argument (namely, $b = b_t(x,\mu,u)$, $\sigma = \sigma_t(x,\mu,u)$, $f = f_t(x,\mu,u)$). Then, by Proposition 5.6 we deduce that the Hamilton-Jacobi-Bellman equation (5.1) can also be written in the following two alternative forms:

$$
\begin{cases}
\partial_t w(t,\mu) + \mathbb{E}\langle \xi_t, A^*\partial_\mu w(t,\mu)(\xi)\rangle_H \\
+ \sup_{\mathrm{a}\in\mathcal{M}} \Big\{ \mathbb{E}\big[f_t\big(\xi,\mu,\mathrm{a}(\xi)\big) + \big\langle b_t\big(\xi,\mu,\mathrm{a}(\xi)\big), \partial_\mu w(t,\mu)(\xi)\big\rangle_H\big] \\
+\dfrac{1}{2}\mathbb{E}\Big[\mathrm{Tr}\Big(\sigma_t\big(\xi,\mu,\mathrm{a}(\xi)\big)\sigma_t^*\big(\xi,\mu,\mathrm{a}(\xi)\big)\partial_x\partial_\mu w(t,\mu)(\xi)\Big)\Big]\Big\} = 0, & (t,\mu)\in\mathscr{H},\, t<T, \\
w(T,\mu) = \mathbb{E}[g(\xi,\mu)], & \mu\in\mathscr{P}_2(C([0,T];H)),
\end{cases}
\tag{5.16}
$$

for every $\xi \in \mathbf{S}_2(\mathcal{G})$ with $\mathbb{P}_\xi = \mu$, or, alternatively:

$$
\begin{cases}
\partial_t w(t,\mu) + \mathbb{E}\langle \xi_t, A^*\partial_\mu w(t,\mu)(\xi)\rangle_H \\
+\mathbb{E}\Big[\operatorname*{ess\,sup}_{u\in\mathrm{U}} \Big[f_t(\xi,\mu,u) + \big\langle b_t(\xi,\mu,u), \partial_\mu w(t,\mu)(\xi)\big\rangle_H \\
+\dfrac{1}{2}\mathrm{Tr}\Big(\sigma_t(\xi,\mu,u)\sigma_t^*(\xi,\mu,u)\partial_x\partial_\mu w(t,\mu)(\xi)\Big)\Big]\Big] = 0, & (t,\mu)\in\mathscr{H},\, t<T, \\
w(T,\mu) = \mathbb{E}[g(\xi,\mu)], & \mu\in\mathscr{P}_2(C([0,T];H)),
\end{cases}
\tag{5.17}
$$

for every $\xi \in \mathbf{S}_2(\mathcal{G})$ with $\mathbb{P}_\xi = \mu$.

Remark 5.9. *In the case of the optimal investment problem outlined in Example 2.12-(ii) and in Remark 3.2, equation (3.2), the HJB equation can be written as follows (here we explicitly write that the production Q depends on the state process ξ_t and on its distribution μ):*

$$
\begin{cases}
\partial_t w(t,\mu) + \mathbb{E}\langle \xi_t, A^* \partial_\mu w(t,\mu)(\xi)\rangle_H \\
+\mathbb{E}\Big[e^{-rt} R(Q(t,\xi_t,\mu) + \delta(\xi,\mu)\, \langle \xi_t, \partial_\mu w(t,\mu)(\xi)\rangle_H \\
+ \operatorname{ess\,sup}_{u\in U}\Big[\langle Cu, \partial_\mu w(t,\mu)(\xi)\rangle_H - e^{-rt}(\langle a_1, u\rangle_H + \langle Mu, u\rangle_H)\Big] \\
+\dfrac{1}{2}\sigma^2 \operatorname{Tr}\Big(\partial_x \partial_\mu w(t,\mu)(\xi)\Big)\Big] = 0, & (t,\mu)\in\mathcal{H},\, t<T, \\
w(T,\mu) = \mathbb{E}[g(\xi,\mu)], & \mu\in\mathscr{P}_2(C([0,T];H)),
\end{cases}
\tag{5.18}
$$

for every $\xi\in \mathbf{S}_2(\mathcal{G})$ with $\mathbb{P}_\xi = \mu$. Note that in the above case the $\operatorname{ess\,sup}$ appearing in the Hamiltonian can be explicitly computed.

A State equation: proofs

We collect in the following lemma some continuity results for contractions in Banach spaces that we will use to obtain the needed continuity properties of the mild solution to the state equation.

Lemma A.1. *Let $\mathcal{R}$ be a non-empty set, $\mathcal{T}$ be a topological space, (M,d) be a metric space, Y be a Banach space, $\gamma\in[0,1)$. Let w be a modulus of continuity. Let $h\colon \mathcal{R}\times\mathcal{T}\times M\times Y \to Y$ be such that:*

$$
|h(r,x,m,y) - h(r,x,m',y')| \le w(d(m,m')) + \gamma|y-y'|, \qquad \forall\, r\in\mathcal{R},\, x\in\mathcal{T},\, m,m'\in M,\, y,y'\in Y.
$$

Let $\mathcal{E}\subset 2^{\mathcal{R}}$ be a set of subsets of $\mathcal{R}$. Assume that, if $\{x_\iota\}_{\iota\in\mathcal{I}}\subset\mathcal{T}$ is a net converging to $x\in\mathcal{T}$, then

$$
\limsup_{\iota}\ \sup_{r\in E} |h(r,x_\iota,m,y) - \varphi(r,x,m,y)| = 0, \qquad \forall\, m\in M,\, y\in Y,\, E\in\mathcal{E}.
$$

Denote by $\varphi\colon \mathcal{R}\times\mathcal{T}\times M \to Y$ the fixed-point map associated with h, i.e. φ is the unique map satisfying

$$
h(r,x,m,\varphi(r,x,m)) = \varphi(r,x,m), \qquad \forall\, r\in\mathcal{R},\, x\in\mathcal{T},\, m\in M.
$$

Then, given a net $\{x_\iota\}_{\iota\in\mathcal{I}}\subset\mathcal{T}$ converging to x, it holds that

$$
\limsup_{\iota}\ \sup_{r\in E} |\varphi(r,x_\iota,m) - \varphi(r,x,m')| \le \frac{1}{1-\gamma} w(d(m,m')), \qquad \forall\, m,m'\in M,\, E\in\mathcal{E}.
\tag{A.1}
$$

Proof. Write

$$
\begin{aligned}
|\varphi(r,x_\iota,m) - \varphi(r,x,m')| &= |h(r,x_\iota,m,\varphi(r,x_\iota,m)) - h(r,x,m',\varphi(r,x,m'))| \\
&\le |h(r,x_\iota,m,\varphi(r,x_\iota,m)) - h(r,x_\iota,m,\varphi(r,x,m'))| \\
&\quad + |h(r,x_\iota,m,\varphi(r,x,m')) - h(r,x,m,\varphi(r,x,m'))| \\
&\quad + |h(r,x,m,\varphi(r,x,m')) - h(r,x,m',\varphi(r,x,m'))| \\
&\le \gamma|\varphi(r,x_\iota,m) - \varphi(r,x,m')| \\
&\quad + |h(r,x_\iota,m,\varphi(r,x,m')) - h(r,x,m,\varphi(r,x,m'))| + w(d(m,m')).
\end{aligned}
$$

Then

$$\sup_{r \in E} |\varphi(r, x_\iota, m) - \varphi(r, x, m')| \le \frac{1}{1 - \gamma} \sup_{r \in E} |h(r, x_\iota, m, \varphi(r, x, m')) - h(r, x, m, \varphi(r, x, m'))|$$
$$+ \frac{1}{1 - \gamma} w(d(m, m'))$$

and the claim follows taking the limit with respect to ι. $\qquad\square$

Proof of Proposition 2.8. The proof is based, as usual, on a contraction argument. We introduce the space $L^2_\mathbb{F}(H)$ (resp. $L^2_\mathbb{F}(\mathcal{L}_2(K; H))$) of all square-integrable $\mathbb{F}$-progressively measurable processes on $[0, T]$ taking values in H (resp. $\mathcal{L}_2(K; H)$), normed respectively by $\|X\|_{L^2_\mathbb{F}(H)} := \left(\mathbb{E}\left[\int_0^T |X_s|^2_H ds\right]\right)^{1/2}$ and $\|\Phi\|_{L^2_\mathbb{F}(H)} := \left(\mathbb{E}\left[\int_0^T |\Phi_s|^2_{\mathcal{L}_2(K;H)} ds\right]\right)^{1/2}$. For the sake of brevity, in what follows we will denote $S_t := e^{At}$, $b^\alpha_t(X, \mathbb{P}_X) := b_t(X, \mathbb{P}_X, \alpha_t, \mathbb{P}_{\alpha_t})$, $\sigma^\alpha_t(X, \mathbb{P}_X) := \sigma_t(X, \mathbb{P}_X, \alpha_t, \mathbb{P}_{\alpha_t})$. For $t \in [0, T]$ and $\alpha \in \mathcal{U}$, let us define

$$
\begin{aligned}
&\mathrm{id}^S_t : \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F}), && \xi \mapsto \xi.\mathbf{1}_{[0,t]}(\cdot) + S_{.-t}\xi_t \mathbf{1}_{(t,T]}(\cdot) \\
&F_{b^\alpha} : \mathbf{S}_2(\mathbb{F}) \to L^2_\mathbb{F}(H), && X \mapsto b^\alpha_.(X, \mathbb{P}_X) \\
&F_{\sigma^\alpha} : \mathbf{S}_2(\mathbb{F}) \to L^2_\mathbb{F}(\mathcal{L}_2(K; H)), && X \mapsto \sigma^\alpha_.(X, \mathbb{P}_X) \\
&S \star_t \# : L^2_\mathbb{F}(H) \to \mathbf{S}_2(\mathbb{F}), && X \mapsto \mathbf{1}_{[t,T]}(\cdot) \int_t^. S_{.-s} X_s ds \\
&S \overset{dB}{\star}_t \# : L^2_\mathbb{F}(\mathcal{L}_2(K; H)) \to \mathbf{S}_2(\mathbb{F}), && \Phi \mapsto \mathbf{1}_{[t,T]}(\cdot) \int_t^. S_{.-s} \Phi_s dB_s.
\end{aligned}
\tag{A.2}
$$

We briefly explain why the functions above are well-defined. Regarding id^S_t, it holds that

$$\mathrm{id}^S_t(\xi) = \xi_{.\wedge t} + \mathbf{1}_{(t,T]}(\cdot)(S_{.-t} - I)\xi_t, \tag{A.3}$$

which clearly shows that $\mathrm{id}^S_t(\xi) \in \mathbf{S}_2(\mathbb{F})$. Regarding F_{b^α}, due to the measurability assumptions on b, we have that $F_{b^\alpha}(X)$ is progressively measurable. Moreover, recalling Assumption $(\mathbf{A}_{A,b,\sigma})$-(iii), we have

$$\|F_{b^\alpha}(X)\|^2_{L^2_\mathbb{F}(H)} \le 3L^2 \mathbb{E}\left[\int_0^T \left(1 + 2\|X\|^2_t\right) dt\right] < \infty. \tag{A.4}$$

In the same way, we obtain the measurability of $F_{\sigma^\alpha}(X)$ and

$$\|F_{\sigma^\alpha}(X)\|^2_{L^2_\mathbb{F}(H)} \le 3L^2 \mathbb{E}\left[\int_0^T \left(1 + 2\|X\|^2_t\right) dt\right] < \infty. \tag{A.5}$$

Regarding $S \star_t X$, for $X \in L^2_\mathbb{F}(H)$, it is not difficult to see that it is continuous, $\mathbb{F}$-adapted, and that

$$\|S \star_t X\|^2_{\mathbf{S}_2} \le e^{2\eta(T-t)}(T - t)\|X\|^2_{L^2_\mathbb{F}(H)}. \tag{A.6}$$

Finally, regarding $S \overset{dB}{\star}_t \Phi$, for $\Phi \in L^2_\mathbb{F}(\mathcal{L}_2(K; H))$, by [35, Theorem 1.111] we know that the $\mathbb{F}$-adapted process $\left\{\mathbf{1}_{[t,T]}(t') \int_t^{t'} S_{t'-s} \Phi_s ds\right\}_{t' \in [0,T]}$ admits a continuous version, that we name $S \overset{dB}{\star}_t \Phi$, and that

$$\|S \overset{dB}{\star}_t \Phi\|_{\mathbf{S}_2} \le C_{\eta,T}\|\Phi\|_{L^2_\mathbb{F}(\mathcal{L}_2(K;H))}, \tag{A.7}$$

where $C_{\eta,T}$ is a constant depending only on η, T.
We now define the map

$$\psi \colon \mathcal{U} \times [0,T] \times \mathbf{S}_2(\mathbb{F}) \times \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F})$$

by

$$\psi(\alpha, t, \xi, X) \;=\; \mathrm{id}_t^S(\xi) + S \star_t F_{b^\alpha}(X) + S \overset{dB}{\star_t} F_{\sigma^\alpha}(X), \quad \forall (\alpha, t, \xi, X) \in \mathcal{U} \times [0,T] \times \mathbf{S}_2(\mathbb{F}) \times \mathbf{S}_2(\mathbb{F}).$$

Claim I. *For fixed $\xi, X \in \mathbf{S}_2(\mathbb{F})$, $\psi(\alpha, t, \xi, X)$ is continuous in t, uniformly in $\alpha \in \mathcal{U}$.*
The continuity of $t \mapsto \mathrm{id}_t^S(\xi)$ follows from Lebesgue's dominated convergence theorem and the fact that, for every fixed $\omega \in \Omega$, $\mathrm{id}_{t'}^S(\xi)(\omega)$ converges uniformly to $\mathrm{id}_t^S(\xi)(\omega)$ as $t' \to t$.
Moreover, for $t', t \in [0, T]$, $t' < t$, we have

$$S \star_{t'} F_{b^\alpha}(X) - S \star_t F_{b^\alpha}(X) \;=\; \mathbf{1}_{[t',t]}(\cdot) \int_{t'}^{\cdot} S_{\cdot - s} F_{b^\alpha}(X)_s ds + \mathbf{1}_{[t,T]}(\cdot) \int_{t'}^{t} S_{\cdot - s} F_{b^\alpha}(X)_s ds.$$

Then

$$\|S \star_{t'} F_{b^\alpha}(X) - S \star_t F_{b^\alpha}(X)\|_T \;\le\; 2 e^{\eta T} \int_{t'}^{t} |F_{b^\alpha}(X)_s|_H ds.$$

This implies, recalling Assumption $(\mathbf{A}_{A,b,\sigma})$(iii),

$$\sup_{\alpha \in \mathcal{U}} \lim_{|t'-t| \to 0} \|S \star_{t'} F_{b^\alpha}(X) - S \star_t F_{b^\alpha}(X)\|_{\mathbf{S}_2} \;=\; 0.$$

Finally, take again $t', t \in [0, T]$, $t' < t$. Then

$$S \overset{dB}{\star_{t'}} F_{\sigma^\alpha}(X) - S \overset{dB}{\star_t} F_{\sigma^\alpha}(X) \;=\; \mathbf{1}_{[t',t]}(\cdot) \int_{t'}^{\cdot} S_{\cdot - s} F_{\sigma^\alpha}(X)_s dB_s + \mathbf{1}_{[t,T]}(\cdot) \int_{t'}^{t} S_{\cdot - s} F_{\sigma^\alpha}(X)_s dB_s.$$

By [35, Theorem 1.111], we have

$$\|S \overset{dB}{\star_{t'}} F_{\sigma^\alpha}(X) - S \overset{dB}{\star_t} F_{\sigma^\alpha}(X)\|_{\mathbf{S}_2}^2 \;\le\; C'_{\eta,T} \|\mathbf{1}_{[t',t]} F_{\sigma^\alpha}(X)\|_{L_{\mathbb{F}}^2(H)}^2$$

and then, after recalling Assumption $(\mathbf{A}_{A,b,\sigma})$(iii),

$$\sup_{\alpha \in \mathcal{U}} \lim_{|t'-t| \to 0} \|S \overset{dB}{\star_{t'}} F_{\sigma^\alpha}(X) - S \overset{dB}{\star_t} F_{\sigma^\alpha}(X)\|_{\mathbf{S}_2} \;=\; 0.$$

Claim II. *$\psi(\alpha, t, \xi, X)$ is Lipschitz continuous in ξ, uniformly in t, X, α.*
For $\alpha \in \mathcal{U}$, $t \in [0, T]$, $\xi, \xi' \in \mathbf{S}_2(\mathbb{F})$, $X \in \mathbf{S}_2(\mathbb{F})$, we have

$$\|\psi(\alpha, t, \xi, X) - \psi(\alpha, t, \xi', X)\|_{\mathbf{S}_2}^2 \;=\; \|\mathrm{id}_t^S(\xi - \xi')\|_{\mathbf{S}_2}^2 \;\le\; 2(1 + e^{\eta T})\|\xi - \xi'\|_{\mathbf{S}_2}^2.$$

Now, for $a, b \in [0, T]$, $a < b$, let us consider the restriction of ψ to the time interval $[a, b]$ and stopped at time b, namely

$$\psi_{a,b} \colon \mathcal{U} \times [a,b] \times \mathbf{S}_2(\mathbb{F}) \times \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F}), \qquad (\alpha, t, \xi, X) \mapsto \psi(\alpha, t, \xi, X)_{b \wedge \cdot}.$$

Claim III. *There exists $\varepsilon > 0$ such that, if $b - a \le \varepsilon$, then $\psi_{a,b}(\alpha, t, \xi, X)$ is a contraction in X, uniformly in α, t, ξ, a, b, namely: for some $\gamma \in [0, 1)$,*

$$\|\psi_{a,b}(\alpha, t, \xi, X)_{b \wedge \cdot} - \psi_{a,b}(\alpha, t, \xi, X')_{b \wedge \cdot}\|_{\mathbf{S}_2} \;\le\; \gamma \|X - X'\|_{\mathbf{S}_2},$$

for all $\alpha \in \mathcal{U}, t \in [a,b], (\xi, X, X') \in \mathbf{S}_2(\mathbb{F})^3$ and for all $a,b \in [0,T], a < b, b - a \leq \varepsilon$.
Let $a, b \in [0,T]$, $a < b$, $t \in [a,b]$, $\alpha \in \mathcal{U}$, $\xi \in \mathbf{S}_2(\mathbb{F})$, $X, X' \in \mathbf{S}_2(\mathbb{F})$. Notice that

$$\psi(\alpha, t, \xi, X) - \psi(\alpha, t, \xi, X') = S \star_t \left(F_{b^\alpha}(X) - F_{b^\alpha}(X') \right) + S \overset{dB}{\star_t} \left(F_{\sigma^\alpha}(X) - F_{\sigma^\alpha}(X') \right) \quad \text{(A.8)}$$

and

$$\|\psi_{a,b}(\alpha, t, \xi, X) - \psi_{a,b}(\alpha, t, \xi, X')\|_T = \|\psi(\alpha, t, \xi, X) - \psi(\alpha, t, \xi, X')\|_b. \quad \text{(A.9)}$$

Moreover, recalling Assumption $(\mathbf{A}_{A,b,\sigma})$-(iii), we have

$$\|S \star_t \left(F_{b^\alpha}(X) - F_{b^\alpha}(X') \right)\|_b^2 \leq e^{2\eta T} \left(\int_a^b \left| F_{b^\alpha}(X)_s - F_{b^\alpha}(X')_s \right| ds \right)^2$$

$$\leq 2L^2 e^{2\eta T}(b-a)^2 \left(\|X - X'\|_T^2 + \mathcal{W}_2^2(\mathbb{P}_X, \mathbb{P}_{X'}) \right).$$

Then

$$\mathbb{E}\left[\|S \star_t \left(F_{b^\alpha}(X) - F_{b^\alpha}(X') \right)\|_b^2 \right] \leq 4L^2 e^{2\eta T}(b-a)^2 \|X - X'\|_{\mathbf{S}_2}. \quad \text{(A.10)}$$

By [35, Theorem 1.111] there exists a constant $C_{\eta,T}$, depending only on η, T, such that

$$\mathbb{E}\left[\|S \overset{dB}{\star_t} \left(F_{\sigma^\alpha}(X) - F_{\sigma^\alpha}(X') \right)\|_b^2 \right] \leq C_{\eta,T}^2 \mathbb{E}\left[\int_a^b \|F_{\sigma^\alpha}(X)_s - F_{\sigma^\alpha}(X')_s\|_{\mathcal{L}_2(K;H)}^2 ds \right]$$

$$\leq 4L^2 C_{\eta,T}^2 (b-a) \|X - X'\|_{\mathbf{S}_2}^2, \quad \text{(A.11)}$$

where for the last inequality we have used Assumption $(\mathbf{A}_{A,b,\sigma})$-(iii) again. By (A.8), (A.9), (A.10), and (A.11), we have, if $b - a < 1$,

$$\|\psi_{a,b}(\alpha, t, \xi, X) - \psi_{a,b}(\alpha, t, \xi, X')\|_{\mathbf{S}_2} \leq C_{L,\eta,T}(b-a)^{1/2} \|X - X'\|_{\mathbf{S}_2},$$

where $C_{L,\eta,T}$ is a constant depending only on L, η, T. We can then choose $\varepsilon \in (0,1)$ such that $C_{L,\eta,T}\varepsilon^{1/2} < 1/2$. Then the map $\psi_{a,b}(\alpha, t, \xi, \cdot)$ is a 1/2-contraction, uniformly in α, t, ξ and in $a, b \in [0,T]$, whenever $a < b, b - a \leq \varepsilon$.

Claim IV. *For $\varepsilon > 0$ as in Claim III, and whenever $a, b \in [0,T]$, $a < b$, $b - a \leq \varepsilon$, there exists a unique mild solution[10] $X^{t,\xi,\alpha}$ to equation (2.4) on the interval $[a,b]$, for any $\alpha \in \mathcal{U}, t \in [a,b], \xi \in \mathbf{S}_2(\mathbb{F})$.*

Let ε be as in *Claim III*. Then, for any α, t, ξ, by the Banach contraction principle, there exists a unique fixed point $X^{t,\xi,\alpha}$ to $\psi_{a,b}(\alpha, t, \xi, \cdot)$. Clearly $X^{t,\xi,\alpha}$ is a mild solution to (2.4) on the interval $[a,b]$.

Claim V. *For ε as in Claim III, and uniformly for $a, b \in [0,T]$, $a < b$, $b - a \leq \varepsilon$, the map*

$$\varphi_{a,b} \colon \mathcal{U} \times [a,b] \times \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F}), \qquad (\alpha, t, \xi) \mapsto X^{t,\xi,\alpha} \quad \text{(A.12)}$$

is continuous in (t,ξ), uniformly in α, and Lipschitz-continuous in ξ, uniformly in α, t.

Let ε be as in *Claim III*. We apply Lemma A.1 with $\mathcal{R} = \mathcal{U}$, $\mathcal{T} = [a,b]$, $M = \mathbf{S}_2(\mathbb{F})$, $Y = \mathbf{S}_2(\mathbb{F})$, $\mathcal{E} = 2^\mathcal{R}$. Then, by *Claims I,II,III*, we get

$$\lim_{\substack{t \to t' \\ t \in [a,b]}} \sup_{\alpha \in \mathcal{U}} \|\varphi_{a,b}(\alpha, t, \xi) - \varphi_{a,b}(\alpha, t', \xi')\|_{\mathbf{S}_2} \leq 4(1 + e^{\eta T})^{1/2} \|\xi - \xi'\|_{\mathbf{S}_2}, \qquad \forall t' \in [a,b], \xi, \xi' \in \mathbf{S}_2.$$

---

[10]We say that $X^{t,\xi,\alpha}$ is a mild solution *on the interval* $[a,b]$ if equation (2.4) is solved in the mild sense for $s \in [a,b]$.

Claim VI. *For any* $\alpha \in \mathcal{U}, t \in [0,T], \xi \in \mathbf{S}_2(\mathbb{F})$, *there exists a unique mild solution* $X^{t,\xi,\alpha}$ *to* (2.4), *and the map*

$$\varphi \colon \mathcal{U} \times [0,T] \times \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F}), \ (\alpha,t,\xi) \mapsto X^{t,\xi,\alpha} \tag{A.13}$$

is continuous in t,ξ, *uniformly in* α, *and Lipschitz continuous in* ξ, *uniformly in* t,α.
Pick $\varepsilon > 0$ as in *Claim III*. Choose $a_0 = 0 < a_1 < \ldots < a_n = T$ with $a_{i+1} - a_i \leq \varepsilon$. Define

$$\hat{\varphi}_{a_i} \colon \mathcal{U} \times [a_i, T] \times \mathbf{S}_2(\mathbb{F}) \to \mathcal{U} \times [a_i, T] \times \mathbf{S}_2(\mathbb{F})$$

by $\hat{\varphi}_{a_i}(\alpha, t, \xi) = (\alpha, a_{i+1}, \varphi_{a_i, a_{i+1}}(\alpha, t, \xi))$ if $t \in [a_i, a_{i+1}]$ and $\hat{\varphi}_{a_i}(\alpha, t, \xi) = (\alpha, t, \xi_{a_{i+1}\wedge \cdot})$ if $t > a_{i+1}$.
If we now define $\varphi(\alpha, t, \xi)$ to be the second component of

$$\hat{\varphi}_{a_n} \left(\hat{\varphi}_{a_{n-1}} \left(\ldots \hat{\varphi}_{a_1} \left(\varphi_{a_0}(\alpha, t, \xi) \right) \ldots \right) \right),$$

we can easily check, thanks to *Claim V*, that $\varphi(\alpha, t, \xi)$ is the unique mild solution $X^{t,\xi,\alpha}$ to (2.4), and the map (A.13) has the desired regularity properties.

Claim VII. $X^{t,\xi,\alpha} = X^{t,\xi_{t\wedge \cdot},\alpha}$.
This is due to the fact that, for any $\alpha \in \mathcal{U}$, $a, b \in [0,T]$, $a < b$, $t \in [a,b]$, $\xi \in \mathbf{S}_2(\mathbb{F})$, $X \in \mathbf{S}_2(\mathbb{F})$,

$$\psi_{a,b}(\alpha, t, \xi, X) \ = \ \psi_{a,b}(\alpha, t, \xi_{t\wedge \cdot}, X).$$

Hence, the unique fixed point of $\psi_{a,b}(\alpha, t, \xi, \cdot)$ has be the same of $\psi_{a,b}(\alpha, t, \xi_{t\wedge \cdot}, \cdot)$.

Claim VIII. *There exists a constant* C *such that*

$$\|X^{t,\xi,\alpha}\|_{\mathbf{S}_2} \ \leq \ C(1 + \|\xi_{t\wedge \cdot}\|_{\mathbf{S}_2}), \qquad \forall\, t \in [0,T], \ \xi \in \mathbf{S}_2, \ \alpha \in \mathcal{U}.$$

Due to *Claim VI*, we only need to show that

$$\sup_{\alpha \in \mathcal{U}, t \in [0,T]} \|X^{t,0,\alpha}\|_{\mathbf{S}_2} < \infty. \tag{A.14}$$

We have, by using (A.4), (A.5), (A.6), (A.7), with T replaced with $t' \in [0,T]$,

$$\mathbb{E} \left[\|X^{t,0,\alpha}\|_{t'}^2 \right] \ \leq \ 2 \left(\|S \star_t \left(\mathbf{1}_{[0,t']} F_{b^\alpha}(X^{t,0,\alpha}) \right) \|_{\mathbf{S}_2}^2 + \|S \overset{dB}{\star_t} \left(\mathbf{1}_{[0,t']} F_{\sigma^\alpha}(X^{t,0,\alpha}) \right) \|_{\mathbf{S}_2}^2 \right)$$

$$\leq \ 2 \left(3L^2 e^{2\eta T} T + 3L^2 C_{\eta,T}^2 \right) \int_0^{t'} (1 + 2\mathbb{E} \left[\|X^{t,0,\alpha}\|_s^2 \right]) ds.$$

An application of Gronwall's inequality yields

$$\mathbb{E} \left[\|X^{t,0,\alpha}\|_{t'}^2 \right] \ \leq \ C, \qquad \forall\, t' \in [0,T],$$

for some C independent of α, t, which proves (A.14) and then (2.5), after recalling *Claim VII*. $\qquad \square$

Proof of Proposition 2.10. For $\alpha \in \mathcal{U}$, let $F_{b^\alpha}, F_{\sigma^\alpha}$ be as in (A.2). Then, let $\mathrm{id}_t^{S^n}, S^n \star_t \#, S^n \overset{dB}{\star_t} \#$ be defined as in (A.2) by replacing S with S^n. Denote $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Let us now define

$$\widetilde{\psi} \colon \mathcal{U} \times \overline{\mathbb{N}} \times [0,T] \times \mathbf{S}_2(\mathbb{F}) \times \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F})$$

by

$$\widetilde{\psi}(\alpha, n, t, \xi, X) \ = \ \mathrm{id}_t^{S^n}(\xi) + S^n \star_t F_{b^\alpha}(X) + S^n \overset{dB}{\star_t} F_{\sigma^\alpha}(X),$$

39

for all $\alpha \in \mathcal{U}, n \in \overline{\mathbb{N}}, t \in [0,T], \xi \in \mathbf{S}_2(\mathbb{F}), X \in \mathbf{S}_2(\mathbb{F})$, where we set $S^\infty := S$. Let

$$\widetilde{\psi}_{a,b}(\alpha, n, t, \xi, X) = \widetilde{\psi}(\alpha, n, t, \xi, X)_{b \wedge \cdot}.$$

whenever $a, b \in [0,T], a < b, t \in [a,b]$. Due to the uniform boundedness of the Yosida approximation, and by arguing as in the proof of *Claims II,III* of Proposition 2.8, one can show that

$\widetilde{\psi}(\alpha, n, t, \xi, X)$ *is Lipschitz continuous in* ξ, *uniformly in* α, n, t, X. *Moreover, there exists* $\varepsilon > 0$ *such that, if* $b - a < \varepsilon$, *then* $\widetilde{\psi}_{a,b}(\alpha, n, t, \xi, X)$ *is a contraction in* X, (A.15) *uniformly in* α, n, t, ξ, a, b.

Now we show that, for $\alpha \in \mathcal{U}, t' \in [0,T], \xi, X \in \mathbf{S}_2(\mathbb{F})$,

$$\lim_{\substack{t \to t' \\ n \to \infty}} \|\widetilde{\psi}(\alpha, n, t, \xi, X) - \widetilde{\psi}(\alpha, \infty, t', \xi, X)\|_{\mathbf{S}_2} = 0. \tag{A.16}$$

First, notice that, for $\mathbb{P}$-a.e. $\omega \in \Omega$, the range of $\xi(\omega)$ is compact. Since $S_t^n x \to S_t x$ uniformly for $t \in [0,T]$ and $x \in K$, whenever $K \subset H$ is compact, an application of Lebesgue's dominated convergence theorem provides

$$\lim_{n \to \infty} \|\mathrm{id}_{t'}^{S^n}(\xi) - \mathrm{id}_{t'}^S(\xi)\|_{\mathbf{S}_2(\mathbb{F})} = 0, \qquad \forall t' \in [0,T], \xi \in \mathbf{S}_2(\mathbb{F}).$$

Secondly, for $\alpha \in \mathcal{U}, X \in \mathbf{S}_2(\mathbb{F})$, after defining

$$f_n(r) = \sup_{t \in [0,T]} |(S_t^n - S_t) b_r(X, \mathbb{P}_X, \alpha_r, \mathbb{P}_{\alpha_r})|^2,$$

we have, by Lebesgue's dominated convergence theorem,

$$\lim_{n \to \infty} \|S^n \star_{t'} F_{b^\alpha}(X) - S \star_{t'} F_{b^\alpha}(X)\|_{\mathbf{S}_2}^2 = \lim_{n \to \infty} \mathbb{E}\left[\sup_{t \in [t',T]} \left|\int_{t'}^t \left(S_{t-r}^n - S_{t-r}\right) b_r(X, \mathbb{P}_X, \alpha_r, \mathbb{P}_{\alpha_r}) dr\right|^2\right]$$

$$\leq \lim_{n \to \infty} \mathbb{E}\left[\int_{t'}^T f_n(r) dr\right] = 0.$$

Thirdly, by [35, Proposition 1.112], we have

$$\lim_{n \to \infty} \|S^n \overset{dB}{\star}_{t'} F_{\sigma^\alpha}(X) - S \overset{dB}{\star}_{t'} F_{\sigma^\alpha}(X)\|_{\mathbf{S}_2(\mathbb{F})} = 0, \qquad \forall \alpha \in \mathcal{U}, t \in [0,T], X \in \mathbf{S}_2(\mathbb{F}).$$

Putting together the above partial results, we get

$$\lim_{n \to \infty} \|\widetilde{\psi}(\alpha, n, t', \xi, X) - \widetilde{\psi}(\alpha, \infty, t', \xi, X)\|_{\mathbf{S}_2} = 0.$$

Then, to prove (A.16), it is enough to show that

$$\lim_{t \to t'} \sup_{n \in \mathbb{N}} \|\widetilde{\psi}(\alpha, n, t, \xi, X) - \widetilde{\psi}(\alpha, n, t', \xi, X)\|_{\mathbf{S}_2} = 0.$$

But this can be obtained by arguing as in the proof of *Claim I* of Proposition 2.8, due to the uniform boundedness $\|S_t^n\|_{\mathcal{L}(H)} \leq e^{\widetilde{\eta} t}$, for $n \in \mathbb{N}, t \geq 0$.
Now, for any small ε as in (A.15), and when $a, b \in [0,T], a < b, b - a < \varepsilon$, denote by

$$\widetilde{\varphi}_{a,b} : \mathcal{U} \times \overline{\mathbb{N}} \times [a,b] \times \mathbf{S}_2(\mathbb{F}) \to \mathbf{S}_2(\mathbb{F}), \qquad (\alpha, n, t, \xi) \mapsto X^{n,t,\xi,\alpha}$$

the fixed-point map associated with $\widetilde{\psi}_{a,b}$ (similarly as done for $\psi_{a,b}, \varphi_{a,b}$ in the proof of Proposition 2.8). Notice that $X^{n,t,\xi,\alpha}$ is the unique mild solution of (2.7) (resp. (2.4)), when $n \in \mathbb{N}$ (resp. $n = \infty$), on the interval $[a,b]$. Thanks to (A.15) and (A.16), we can apply Lemma A.1 with $\mathcal{R} = \mathcal{U}$, $\mathcal{T} = \overline{\mathbb{N}} \times [a,b]$, $M = \mathbf{S}_2(\mathbb{F})$, $Y = \mathbf{S}_2(\mathbb{F})$, $\mathcal{E} = \{\{\alpha\}\}_{\alpha \in \mathcal{U}}$, and obtain

$$\sup_{\substack{n \in \mathbb{N} \\ t \in [a,b]}} \|\widetilde{\varphi}_{a,b}(\alpha, n, t, \xi) - \widetilde{\varphi}_{a,b}(\alpha, n, t, \xi')\|_{\mathbf{S}_2} \ \leq \ C\|\xi - \xi'\|_{\mathbf{S}_2}, \qquad \forall\, \alpha \in \mathcal{U},\, \xi, \xi' \in \mathbf{S}_2(\mathbb{F}), \quad (A.17)$$

and

$$\lim_{\substack{t \to t' \\ t \in [a,b] \\ n \to \infty}} \|\widetilde{\varphi}_{a,b}(\alpha, n, t, \xi) - \widetilde{\varphi}_{a,b}(\alpha, \infty, t', \xi)\|_{\mathbf{S}_2} \ = \ 0, \qquad \forall\, \alpha \in \mathcal{U}, \quad (A.18)$$

for some constant C, uniformly for $a, b \in [0,T]$, $a < b$, $b - a < \varepsilon$, $t' \in [a,b]$. To conclude the proof it is enough to use (A.17) and (A.18) iteratively, recalling the relation between the mild solution on the subinterval $[a,b] \subset [0,T]$ and the global mild solution on $[0,T]$ (arguing as in the proof of *Claim VI* of Proposition 2.8, simply replacing $[a_i, T]$ with $\overline{\mathbb{N}} \times [a_i, T]$). $\qquad\square$

B   Law invariance property of V: technical results

In the present appendix we prove three technical results that are needed in the proof of the law invariance property (Theorem 3.6). The first technical result corresponds to Theorem 6.10 in [59]. For the convenience of the reader, we restate it here using the notation adopted in the paper and in the form needed for the proof of Theorem 3.6.

Lemma B.1. *Consider a probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$, a measurable space $(E, \mathscr{E})$, a Borel space $(\mathrm{U}, \mathscr{U})$. Consider also two random variables $\Gamma \colon \hat{\Omega} \to E$ and $\alpha \colon \hat{\Omega} \to \mathrm{U}$. Suppose that there exists a random variable $\hat{U} \colon \hat{\Omega} \to \mathbb{R}$, having uniform distribution on $[0,1]$, such that Γ and $\hat{U}$ are independent. Then, there exists a measurable function $\mathrm{a} \colon E \times [0,1] \to \mathrm{U}$ satisfying*

$$\big(\Gamma, \mathrm{a}(\Gamma, \hat{U})\big) \overset{\mathscr{L}_{\hat{\Omega}}}{=} (\Gamma, \alpha),$$

where $\overset{\mathscr{L}_{\hat{\Omega}}}{=}$ stands for equality in law (between random objects defined on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$).

Proof. See Theorem 6.10 in [59]. $\qquad\square$

Before stating next result, we introduce the following notation. For every $t \in [0,T]$, let $\mathbb{F}^{B,t} = (\mathcal{F}_s^{B,t})_{s \geq 0}$ be the $\mathbb{P}$-completion of the filtration generated by $(B_{s \vee t} - B_t)_{s \geq 0}$. Let also $Prog(\mathbb{F}^{B,t})$ be the σ-algebra of $[0,T] \times \Omega$ of all $\mathbb{F}^{B,t}$-progressive sets (recall that a set $C \subset [0,T] \times \Omega$ is called $\mathbb{F}^{B,t}$-*progressive* if the corresponding indicator function $\mathbb{1}_C$ is an $\mathbb{F}^{B,t}$-progressively measurable process; notice that the family of all $\mathbb{F}^{B,t}$-progressive sets is a σ-algebra).

Lemma B.2. *Let $t \in [0,T]$, $\alpha \in \mathcal{U}$, $\xi \in \mathbf{S}_2(\mathbb{F})$, with ξ being $\mathcal{B}([0,T]) \otimes \mathcal{F}_t$-measurable. Suppose that there exists an $\mathcal{F}_t$-measurable random variable U_ξ, having uniform distribution on $[0,1]$ and being independent of ξ. Then, there exists a measurable function*

$$\mathrm{a} \colon \big([0,T] \times \Omega \times C([0,T];H) \times [0,1], Prog(\mathbb{F}^{B,t}) \otimes \mathscr{B} \otimes \mathcal{B}([0,1])\big) \longrightarrow (\mathrm{U}, \mathscr{U})$$

such that $(\mathrm{a}_s(\xi, U_\xi))_{s \in [0,T]} \in \mathcal{U}$, $\mathrm{a}_s(\cdot, \cdot)$ *is constant for every* $s < t$, *moreover*

$$\left((\xi_s)_{s \in [0,T]}, (\mathrm{a}_s(\xi, U_\xi))_{s \in [t,T]}, (B_s - B_t)_{s \in [t,T]} \right) \overset{\mathscr{L}}{=} \left((\xi_s)_{s \in [0,T]}, (\alpha_s)_{s \in [t,T]}, (B_s - B_t)_{s \in [t,T]} \right), \quad \text{(B.1)}$$

where $\overset{\mathscr{L}}{=}$ *stands for equality in law (between random objects defined on* $(\Omega, \mathcal{F}, \mathbb{P})$*).*

Proof. Denote

$$\hat{\Omega} = [0,T] \times \Omega, \qquad \hat{\mathcal{F}} = \mathcal{B}([0,T]) \otimes \mathcal{F}, \qquad \hat{\mathbb{P}} = \lambda_T \otimes \mathbb{P},$$

with λ_T being the uniform distribution on $([0,T], \mathcal{B}([0,T]))$. Then, consider the canonical extension of U_ξ to $\hat{\Omega}$, which will be denoted by $\hat{U}_\xi$ (notice that $\hat{U}_\xi$ has uniform distribution on $[0,1]$ and is independent of ξ). Let also $(\bar{E}, \bar{\mathscr{E}})$ be the measurable space defined as: $\bar{E} = [0,T] \times \Omega$ and $\bar{\mathscr{E}} = Prog(\mathbb{F}^{B,t})$. Then, let $\mathcal{I}^{B,t} \colon \hat{\Omega} \to \bar{E}$ be the identity map. Finally, define $\Gamma = (\mathcal{I}^{B,t}, \xi)$. Notice that Γ is a random variable taking values in the measurable space $(E, \mathscr{E})$, with $E = \bar{E} \times C([0,T]; H) = [0,T] \times \Omega \times C([0,T]; H)$ and $\mathscr{E} = \bar{\mathscr{E}} \otimes \mathscr{B} = Prog(\mathbb{F}^{B,t}) \otimes \mathscr{B}$. We also observe that Γ and $\hat{U}_\xi$ are independent. We can then apply Lemma B.1, from which it follows the existence of a map $\bar{\mathrm{a}} \colon [0,T] \times \Omega \times C([0,T]; H) \times [0,1] \to \mathrm{U}$, measurable with respect to the σ-algebra $Prog(\mathbb{F}^{B,t}) \otimes \mathscr{B} \otimes \mathcal{B}([0,1])$, such that

$$\left(\Gamma, \bar{\mathrm{a}}(\Gamma, \hat{U}_\xi) \right) \overset{\mathscr{L}_{\hat{\Omega}}}{=} \left(\Gamma, \alpha \right),$$

where $\mathscr{L}_{\hat{\Omega}}$ stands for equality in law between random objects defined on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$. Then, we deduce

$$\left((\xi_s)_{s \in [0,T]}, (\bar{\mathrm{a}}_s(\xi, U_\xi))_{s \in [0,T]}, (B_s - B_t)_{s \in [t,T]} \right) \overset{\mathscr{L}}{=} \left((\xi_s)_{s \in [0,T]}, (\alpha_s)_{s \in [0,T]}, (B_s - B_t)_{s \in [t,T]} \right),$$

where we recall that $\mathscr{L}$ stands for equality in law between random objects defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Finally, let $\mathrm{a} \colon [0,T] \times \Omega \times C([0,T]; H) \times [0,1] \to \mathrm{U}$ be the map given by

$$\mathrm{a}_s(\omega, x, u) = u_0 \, \mathbb{1}_{[0,t)}(s) + \bar{\mathrm{a}}_s(\omega, x, u) \, \mathbb{1}_{[t,T]}(s), \quad \forall (s, \omega, x, u) \in [0,T] \times \Omega \times C([0,T]; H) \times \mathrm{U},$$

where $u_0 \in \mathrm{U}$ is arbitrarily chosen. We have that $(\mathrm{a}_s(\xi, U_\xi))_{s \in [0,T]}$ is $\mathbb{F}$-progressively measurable (here we use that $\mathrm{a}_s(\cdot, \cdot)$ is constant for every $s < t$, ξ is $\mathcal{B}([0,T]) \otimes \mathcal{F}_t$-measurable and U_ξ is $\mathcal{F}_t$-measurable). So, in particular, $(\mathrm{a}_s(\xi, U_\xi))_{s \in [0,T]} \in \mathcal{U}$ and equality (B.1) holds. $\qquad \square$

Lemma B.3. *Let* $t \in [0,T]$ *and* $\xi \in \mathbf{S}_2(\mathbb{F})$, *with* ξ *being* $\mathcal{B}([0,T]) \otimes \mathcal{F}_t$-*measurable. Suppose that there exists* $\{x_1, \dots, x_m\} \subset C([0,T]; H)$, *with* $x_i \neq x_j$ *if* $i \neq j$, *such that*

$$\mathbb{P}_\xi = \sum_{i=1}^m p_i \, \delta_{x_i},$$

where δ_{x_i} *is the Dirac measure at* x_i *and* $p_i > 0$, *with* $\sum_{i=1}^m p_i = 1$. *Then, there exists an* $\mathcal{F}_t$-*measurable random variable* U_ξ *having uniform distribution on* $[0,1]$ *and being independent of the* $\mathcal{F}_t$-*measurable map*

$$\tilde{\xi} \colon \Omega \to C([0,T]; H), \quad \omega \mapsto \xi(\omega).$$

Proof. We recall from Lemma 2.1 that there exists a $\mathcal{G}$-measurable (so, in particular, $\mathcal{F}_t$-measurable) random variable $U_{\mathcal{G}}$ with uniform distribution on $[0,1]$. If $U_{\mathcal{G}}$ is independent of $\tilde{\xi}$, then we take $U_\xi = U_{\mathcal{G}}$, otherwise we proceed as follows. Denote

$$E_i := \{\omega \in \Omega \colon \xi(\omega) = x_i\}, \qquad \forall i = 1, \dots, m.$$

Since ξ is $\mathcal{B}([0,T]) \otimes \mathcal{F}_t$-measurable, it follows that each E_i belongs to $\mathcal{F}_t$. Now, for each $i = 1, \ldots, m$, define the function $F_i \colon [0,1] \to [0,1]$ as follows

$$F_i(r) := \mathbb{P}\big(\{U_{\mathcal{G}} \leq r\} \cap E_i\big), \qquad \forall\, r \in [0,1].$$

Notice that F_i satisfies the following properties:

- F_i is a non-decreasing function;

- F_i is continuous;

- $F_i(0) = 0$ and $F_i(1) = \mathbb{P}(E_i) = p_i > 0$.

For every $i = 1, \ldots, m$ ed $n \in \mathbb{N}$, define

$$r_{i, \frac{k}{2^n}} := \min\left\{ r \in [0,1] \colon F_i(r) = \frac{k}{2^n} p_i \right\}, \qquad k = 1, \ldots, 2^n.$$

Then, for each $n \in \mathbb{N}$ define the random variable $X_n \colon \Omega \to \mathbb{R}$ as

$$X_n(\omega) := \sum_{i=1}^{m} \sum_{k=1}^{2^{n-1}} \mathbb{1}_{\left\{ r_{i, \frac{2k-1}{2^n}} < U_{\mathcal{G}} \leq r_{i, \frac{2k}{2^n}} \right\} \cap E_i}(\omega), \qquad \forall\, \omega \in \Omega.$$

Notice that $\{X_n\}_n$ is a sequence of independent and identically distributed Bernoulli random variables of parameter $1/2$. In addition, $\{X_n\}_n$ and $\tilde{\xi}$ are independent. Then, consider the random variable $U_\xi \colon \Omega \to \mathbb{R}$ given by

$$U_\xi := \sum_{n=1}^{\infty} \frac{1}{2^n} X_n.$$

We have that U_ξ and $\tilde{\xi}$ are independent. Finally, it follows for instance from Lemma 3.20 in [59] that the random variable U_ξ has uniform distribution on $[0,1]$. $\qquad\square$

C   Pathwise measure derivative: law invariance property

We devote this appendix to extend a useful result (firstly proved in [21, Section 6.1] in the case of the set $\mathbb{R}^d$) to the case of the set $C([0,T]; H)$. More precisely, our aim is to prove that $D\hat{\Phi}(t, \hat{\xi})$ only depends on the law of $\hat{\xi}$. Here we substantially follow the idea of [72] and [73, Section 2.3]. However, since our setting is more general, we present the full proof.

C.1   The discrete case

We first consider the case where the random variable $\hat{\xi}$ takes a countable number of values. We assume the following.

Assumption $(\mathbf{A}_\xi)$. *Consider a sequence $\{\hat{x}_i\}_{i \in \mathbb{N}} \subset D([0,T]; H)$, $\hat{x}_i \neq \hat{x}_j$ if $i \neq j$. We assume that the random variable $\hat{\xi}$ has law*

$$\mathbb{P}_{\hat{\xi}} = \sum_{i=1}^{N} p_i \delta_{\hat{x}_i}$$

where $N \in \mathbb{N} \setminus \{0\}$, $\delta_{\hat{x}_i}$ denotes the Dirac measure at $\hat{x}_i$ and $p_i > 0$, with $\sum_{i=1}^{N} p_i = 1$.

Lemma C.1. *Let $(t, \hat{\xi}) \in \hat{\mathfrak{H}}$, with $\hat{\xi}$ as in Assumption* **(A$_\xi$)**. *Let $\hat{\varphi} \colon \hat{\mathscr{H}} \to \mathbb{R}$ be such that its lifting $\hat{\Phi}$ is pathwise differentiable in space at $(t, \hat{\xi})$. Then there exists a measurable function $\hat{g} \colon D([0, T], H) \to H$ such that $\hat{g}(\hat{\xi}) \in L^2(\Omega; H)$ and*

$$D\hat{\Phi}(t, \hat{\xi}) = \hat{g}(\hat{\xi}). \tag{C.1}$$

The function $\hat{g}$ can be defined by $\hat{g}(\hat{x}) := 0$ if $\hat{x} \notin \{\hat{x}_i\}_{i=1,\dots,N}$, and

$$\langle \hat{g}(\hat{x}_i), h \rangle_H := p_i^{-1} \mathbb{E}\left[\left\langle D\hat{\Phi}(t, \hat{\xi}), h \mathbb{1}_{\hat{\xi}^{-1}(\hat{x}_i)} \right\rangle_H \right]$$

$$= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon p_i} \left[\hat{\varphi}\left(t, \sum_{\substack{j=1,\dots,N \\ j \neq i}} p_j \delta_{\hat{x}_j} + p_i \delta_{\hat{x}_i + \varepsilon h \mathbb{1}_{[t,T]}} \right) - \hat{\varphi}\left(t, \sum_{j=1}^N p_j \delta_{\hat{x}_j} \right) \right], \tag{C.2}$$

for all $h \in H$ and $i = 1, \dots, N$.

Proof. Fix $i = 1, \dots, N$ and $h \in H, h \neq 0$. Take any measurable set $A' \subset A_i := \{\hat{\xi} = \hat{x}_i\}$ and set $\hat{\eta} := h \mathbb{1}_{A'} \in L^2(\Omega; H)$. Notice that, for any $\varepsilon > 0$ we have

$$\mathbb{P}_{\hat{\xi} + \varepsilon \hat{\eta}} = \sum_{j \neq i} p_j \delta_{\hat{x}_j} + \mathbb{P}(A') \delta_{\hat{x}_i + \varepsilon h \mathbb{1}_{[t,T]}} + (p_i - \mathbb{P}(A')) \delta_{\hat{x}_i}$$

which depends only on $\mathbb{P}_{\hat{\xi}}$ and on $\mathbb{P}(A')$. We get

$$\mathbb{E}\left[\left\langle D\hat{\Phi}(t, \hat{\xi}), h \mathbb{1}_{A'} \right\rangle_H \right]$$

$$= \lim_{\varepsilon \to 0^+} \frac{\hat{\varphi}\left(t, \sum_{j \neq i} p_j \delta_{\hat{x}_j} + \mathbb{P}(A') \delta_{\hat{x}_i + \varepsilon h \mathbb{1}_{[t,T]}} + (p_i - \mathbb{P}(A')) \delta_{\hat{x}_i} \right) - \hat{\varphi}\left(t, \sum_{j=1}^N p_j \delta_{\hat{x}_j} \right)}{\varepsilon}. \tag{C.3}$$

Notice that the map

$$H \times L^2(\Omega; \mathbb{R}) \longrightarrow L^2(\Omega; H), \qquad (h, \zeta) \longmapsto \zeta h$$

is bilinear and continuous, hence

$$\Lambda \colon H \times L^2(\Omega; \mathbb{R}) \longrightarrow \mathbb{R}, \qquad (h, \zeta) \longmapsto \mathbb{E}[\langle D\hat{\Phi}(t, \hat{\xi}), \zeta h \rangle_H]$$

is a bilinear and continuous form. By the Riesz representation theorem, there exists a (unique) linear and continuous map

$$T \colon H \longrightarrow L^2(\Omega; \mathbb{R})$$

representing Λ, namely

$$\Lambda(h, \zeta) = \mathbb{E}[\langle D\hat{\Phi}(t, \hat{\xi}), \zeta h \rangle_H] = \mathbb{E}[T(h)\zeta] \qquad \forall (h, \zeta) \in H \times L^2(\Omega; \mathbb{R}). \tag{C.4}$$

By (C.4) and (C.3), it follows that $\mathbb{E}[\langle D\hat{\Phi}(t, \hat{\xi}), h \mathbb{1}_{A'} \rangle_H] = \mathbb{E}[T(h) \mathbb{1}_{A'}]$ depends only on the law of $\hat{\xi}$ and on $\mathbb{P}(A')$ for every measurable $A' \subset \{\hat{\xi} = \hat{x}_i\}$. Recall from Lemma 2.1 that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supports a random variable with uniform distribution on $[0, 1]$. We can then apply Lemma 2 of [72], getting that $T(h)$ is $\mathbb{P}$-a.s. constant on $\{\hat{\xi} = \hat{x}_i\}$. For $i = 1, \dots, N$, define the map $\psi_i \colon H \to \mathbb{R}$ by

$$\psi_i(h) := \frac{1}{p_i} \mathbb{E}[T(h) \mathbb{1}_{A_i}], \qquad \forall h \in H.$$

44

Then $\psi_i(h) = T(h)$, $\mathbb{P}$-a.s. on A_i. Notice that, by the very definition of ψ_i and the linearity of Λ, we have

$$\sum_{i=1}^{N} \psi_i(h) \mathbb{E}[\zeta \mathbb{1}_{A_i}] = \sum_{i=1}^{N} \mathbb{E}[T(h)\zeta \mathbb{1}_{A_i}] = \sum_{i=1}^{N} \Lambda(h, \zeta \mathbb{1}_{A_i}) = \Lambda(h, \zeta) = \mathbb{E}[\langle D\hat{\Phi}(t, \hat{\xi}), h\zeta \rangle_H],$$

for all $h \in H$, $\zeta \in L^2(\Omega; \mathbb{R})$. The linearity and continuity of T entails $\psi_i \in H'$. Therefore, it can be identified to some $\phi_i \in H$. Define $\hat{g} \colon D([0,T]; H) \to H$ by

$$\hat{g}(\hat{x}) := \sum_{i=1}^{N} \phi_i \, \delta_{\hat{x}_i}(\hat{x}) \qquad \forall \, \hat{x} \in D([0,T]; H).$$

Notice that, $\mathbb{P}$-a.s. on Ω and for all $h \in H$,

$$\langle \hat{g}(\hat{\xi}), h \rangle_H = \left\langle \sum_{i=1}^{N} \phi_i \, \delta_{\hat{x}_i}(\hat{\xi}), h \right\rangle_H = \left\langle \sum_{i=1}^{N} \phi_i \, \mathbb{1}_{A_i}, h \right\rangle_H$$

$$= \sum_{i=1}^{N} \psi_i(h) \, \mathbb{1}_{A_i} = \sum_{i=1}^{N} T(h) \, \mathbb{1}_{A_i} = T(h). \qquad (\text{C.5})$$

Now let $Y := \sum_{k=1}^{M} a_k \mathbb{1}_{B_k}$ be a simple function, where $M \in \mathbb{N}, a_k \in H, B_k \in \mathcal{F}$. By (C.5) it follows

$$\mathbb{E}[\langle D\hat{\Phi}(t, \hat{\xi}), Y \rangle_H] = \sum_{k=1}^{M} \mathbb{E}[\langle D\hat{\Phi}(t, \hat{\xi}), a_k \mathbb{1}_{B_k} \rangle_H] = \sum_{k=1}^{M} \mathbb{E}\left[T(a_k)\mathbb{1}_{B_k}\right] \qquad (\text{C.6})$$

$$= \sum_{k=1}^{M} \mathbb{E}\left[\langle \hat{g}(\hat{\xi}), a_k \rangle_H \mathbb{1}_{B_k}\right] = \mathbb{E}\left[\langle \hat{g}(\hat{\xi}), Y \rangle_H\right].$$

Since (C.6) holds for any simple function Y, we conclude

$$D\hat{\Phi}(t, \hat{\xi}) = \hat{g}(\hat{\xi}) \qquad \mathbb{P}\text{-a.s.}$$

which is (C.1). Finally, (C.2) comes from (C.1) and (C.3). $\qquad \square$

C.2 The general case

Proof of Lemma 4.10. We proceed by approximation. For $n \in \mathbb{N}$, let $\{D_i^n\}_{i \in \mathbb{N}}$ be a partition of $D([0,T]; H)$ made by Borel sets such that $\text{diam}_{d_{\text{Sk}}}(D_i^n) < 2^{-n}$. Such a partition exists because of the separability of $(D([0,T]; H), d_{\text{Sk}})$. For $i \in \mathbb{N}$, choose $d_i^n \in D_i^n$. Define, for every $n \in \mathbb{N}$,

$$\hat{\zeta}_n := \sum_{i \in \mathbb{N}} d_i^n \, \mathbb{1}_{D_i^n}(\hat{\xi}).$$

Then $d_{\text{Sk}}(\hat{\xi}(\omega), \hat{\zeta}_n(\omega)) \leq 2^{-n}$ for all $\omega \in \Omega$ and all $n \in \mathbb{N}$. Therefore $\hat{\zeta}_n \in L^2(\Omega; D([0,T], H))$ and $\hat{\zeta}_n \to \hat{\xi}$ uniformly. By a diagonal argument, we can choose $N_n \in \mathbb{N}$ such that the sequence $\{\hat{\xi}_n\}_{n \in \mathbb{N}}$, defined by

$$\hat{\xi}_n := \sum_{i=1}^{N_n} d_i^n \mathbb{1}_{D_i^n}(\hat{\xi}) \qquad (\text{C.7})$$

converges to $\hat{\xi}$ both $\mathbb{P}$-a.s. and in $L^2(\Omega; D([0, T]; H))$. By Lemma C.1, there exists for each n a function $\hat{g}_n \colon D([0, T]; H) \to H$ such that

$$\hat{g}_n(\hat{\xi}_n) \in L^2(\Omega; H) \text{ and } D\hat{\Phi}(t, \hat{\xi}_n) = \hat{g}_n(\hat{\xi}_n). \tag{C.8}$$

Define

$$\tilde{g}_n := \sum_{i=1}^{N_n} \hat{g}_n \left(d_i^n \right) \mathbb{1}_{D_i^n}.$$

Notice that $\hat{g}_n(\hat{\xi}_n) = \tilde{g}_n(\hat{\xi})$. Then, by (C.8) and by continuity of $D\hat{\Phi}(t, \cdot)$ in a neighborhood of $\hat{\xi}$, we have $\tilde{g}_n(\hat{\xi}) \to D\hat{\Phi}(t, \hat{\xi})$ in $L^2(\Omega; H)$. Let $\{\tilde{g}_{n_k}(\hat{\xi})\}_k$ be a subsequence converging $\mathbb{P}$-a.s. to $D\hat{\Phi}(t, \hat{\xi})$. Define

$$S := \left\{ \hat{x} \in D([0, T]; H) \colon \{\tilde{g}_{n_k}(\hat{x})\}_k \text{ is convergent} \right\},$$

and $\hat{g} \colon D([0, T]; H) \to H$ by $\hat{g} := \mathbb{1}_S \lim_k \tilde{g}_{n_k}$. Clearly $\hat{g}$ is measurable. Notice that $\mathbb{P}(\hat{\xi} \in S) = 1$, then

$$\lim_{k \to \infty} \tilde{g}_{n_k}(\hat{\xi}) = \hat{g}(\hat{\xi}) \quad \mathbb{P}\text{-a.s. in } D([0, T]; H) \tag{C.9}$$

which provides $\hat{g}(\hat{\xi}) = D\hat{\Phi}(t, \hat{\xi})$.

Finally, if $\hat{\xi}'$ is distributed as $\hat{\xi}$, then we can perform exactly the same steps by replacing $\hat{\xi}, \hat{\xi}_n$ by $\hat{\xi}', \hat{\xi}'_n$ and by choosing the same N_n in (C.7) and then the same $\tilde{g}_n, \hat{g}$. Moreover, $\mathbb{P}(\hat{\xi}' \in S) = 1$ as well. Then (C.9) holds with the same $\hat{g}$ and with $\hat{\xi}$ replaced by $\hat{\xi}'$, which entails $\hat{g}(\hat{\xi}') = D\hat{\Phi}(t, \hat{\xi}')$.

The remaining part of the proof goes exactly as in the proof of Corollary 2.3 of [73]. $\square$

D   Proof of Itô's formula

Proof of Theorem 4.16. Let $\xi, \hat{\varphi}$ be as in the statement. In what follows, we will tacitly make use of the non-anticipative property of $\hat{\varphi}$ and of its derivatives. Let $t \in [0, T), s \in (t, T]$. For $n \in \mathbb{N}$, let

$$t_{-1}^n := t, \qquad t_k^n := t + \frac{k}{n}(s - t) \quad \forall\, k = 1, \dots, n$$

and $X^n = \sum_{k=1}^n X_{t_{k-1}^n} \mathbb{1}_{[t_{k-1}^n, t_k^n)}$. Clearly $\mathbb{E}\left[\|X^n\|_T\right] < \infty$. Observe that

$$X_{t_k^n \wedge \cdot}^n = X_{t_{k-1}^n \wedge \cdot}^n + (X_{t_k^n}^n - X_{t_{k-1}^n}^n)\mathbb{1}_{[t_k^n, T]} \qquad \forall\, k = 1, \dots, n. \tag{D.1}$$

By continuity of X, we clearly have

$$\lim_{n \to \infty} \sup_{r \in [0, T]} \|X_r(\omega) - X_r^n(\omega)\|_H = 0 \qquad \forall \omega \in \Omega. \tag{D.2}$$

Denote

$$X^{n, \theta, k} = X_{t_{k-1}^n \wedge \cdot}^n + \theta(X_{t_k^n}^n - X_{t_{k-1}^n}^n) \qquad \forall \theta \in [0, 1]. \tag{D.3}$$

By continuity of X, if $n \to \infty$ and if $\{k_n\}_{n \in \mathbb{N}} \subset \mathbb{N}$ is a sequence such that $k_n \in \{1, \dots, n\}$ and $t_{k_n}^n \to r$ in $[t, s]$, then

$$\lim_{n \to \infty} \sup_{\theta \in [0, 1]} \|X_{r \wedge \cdot}(\omega) - X_{t_{k_n}^n \wedge \cdot}^{n, \theta, k_n}(\omega)\|_T = 0 \qquad \forall \omega \in \Omega. \tag{D.4}$$

Consider the difference $\hat{\varphi}(s, X^n) - \hat{\varphi}(t, X^n)$, written as

$$\hat{\varphi}(s, \hat{\mathbb{P}}_{X^n}) - \hat{\varphi}(t, \hat{\mathbb{P}}_{X^n}) = \sum_{k=1}^{n} \left(\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^n}) - \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^n_{t_{k-1}^n \wedge \cdot}}) \right)$$

$$+ \sum_{k=1}^{n} \left(\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^n_{t_{k-1}^n \wedge \cdot}}) - \hat{\varphi}(t_{k-1}^n, \hat{\mathbb{P}}_{X^n}) \right). \tag{D.5}$$

Let us firt take in consideration the quantity $\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^n}) - \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^n_{t_{k-1}^n \wedge \cdot}})$. By (D.1) and by the the fact that $\hat{\varphi} \in \boldsymbol{C}_b^{1,2}(\hat{\mathscr{H}})$, we have

$$\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^n}) - \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^n_{t_{k-1}^n \wedge \cdot}}) = \hat{\Phi}(t_k^n, X^n) - \hat{\Phi}(t_k^n, X^n_{t_{k-1}^n \wedge \cdot})$$

$$= \int_0^1 \mathbb{E}\left[\langle D\hat{\Phi}\left(t_k^n, X^{n,\theta,k}, \right), X^n_{t_k^n} - X^n_{t_{k-1}^n} \rangle \right] d\theta$$

$$= \int_0^1 \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}} \right)\left(X^{n,\theta,k} \right), X^n_{t_k^n} - X^n_{t_{k-1}^n} \rangle \right] d\theta$$

$$= \text{(after recalling (4.5))}$$

$$= \int_0^1 \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}} \right)\left(X^{n,\theta,k} \right), \int_{t_{k-1}^n}^{t_k^n} F_r dr \rangle \right] d\theta$$

$$+ \int_0^1 \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}} \right)\left(X^{n,\theta,k} \right), \int_{t_{k-1}^n}^{t_k^n} G_r dB_r \rangle \right] d\theta$$

$$= \mathbf{I}_F^{n,k} + \mathbf{I}_B^{n,k}. \tag{D.6}$$

Observe that, by the boundedness assumption on $\partial_\mu \hat{\varphi}$ and the integrability assumption on F, we can commute the integral $\int_{t_{k-1}^n}^{t_k^n} \cdot dr$ first with $\mathbb{E}$ and then (since the map $[0,T] \mapsto L^1(\Omega; H)$, $r \mapsto F_r$ is measurable and integrable) with $\int_0^1 \cdot d\theta$ [11]. By summing over k the quantity $\mathbf{I}_F^{n,k}$, we then have

$$\sum_{k=1}^{n} \mathbf{I}_F^{n,k} = \sum_{k=1}^{n} \int_{t_{k-1}^n}^{t_k^n} \int_0^1 \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}} \right)\left(X^{n,\theta,k} \right), F_r \rangle \right] d\theta dr$$

$$= \int_t^s \int_0^1 \mathbb{E}\left[\langle \sum_{k=1}^{n} \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}} \right)\left(X^{n,\theta,k} \right) \mathbb{1}_{[t_{k-1}^n, t_k^n)}(r), F_r \rangle \right] d\theta dr. \tag{D.7}$$

By continuity of $\partial_\mu \hat{\varphi}$ and by (D.4), we have

$$\lim_{n \to \infty} \left| \sum_{k=1}^{n} \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}} \right)\left(X^{n,\theta,k}(\omega) \right) \mathbb{1}_{[t_{k-1}^n, t_k^n)}(r) - \partial_\mu \hat{\varphi}\left(r, \hat{\mathbb{P}}_X \right)(X(\omega)) \right|_H = 0 \qquad \forall \omega \in \Omega. \tag{D.8}$$

By Lebesgue's dominated convergence theorem and by (D.8),(D.7), we conclude

$$\lim_{n \to \infty} \sum_{k=1}^{n} \mathbf{I}_F^{n,k} = \int_t^s \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(r, \hat{\mathbb{P}}_X \right)(X), F_r \rangle \right] dr. \tag{D.9}$$

---

[11] But notice that the integral in $\int_0^1 \cdot d\theta$ cannot be commuted with $\mathbb{E}$ before showing the existence of a jointly measurable representant $(\theta, \omega) \mapsto Z(\theta, \omega)$ of $\theta \mapsto \partial_\mu \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\theta,k})(\omega)$.

We now address $\mathbf{I}_B^{n,k}$. In this case, we cannot immediately commute $\int_{t_{k-1}^n}^{t_k^n} \cdot dB_r$ with $\mathbb{E}$, because $X^{n,\theta,k}$ is a-priori only $\mathcal{F}_{t_k^n}$-measurable. We then add a null term and then expand the sum. Indeed, we have

$$\int_0^1 \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}}\right)\left(X^{n,\theta,k-1}\right), \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\rangle\right] d\theta =$$
$$= \int_0^1 \mathbb{E}\left[\int_{t_{k-1}^n}^{t_k^n} \langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}}\right)\left(X^{n,\theta,k-1}\right), G_r dB_r\rangle\right] d\theta = 0.$$

Then, by using the continous second-order pathwise derivative in mesure and space of $\hat{\varphi}$

$$\mathbf{I}_B^{n,k} = \int_0^1 \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}}\right)\left(X^{n,\theta,k}\right), \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\rangle\right] d\theta$$

$$= \int_0^1 \mathbb{E}\left[\langle \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}}\right)\left(X^{n,\theta,k}\right) - \partial_\mu \hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}}\right)\left(X^{n,\theta,k-1}\right), \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\rangle\right] d\theta$$

$$= \int_0^1 \mathbb{E}\left[\langle \int_0^1 \partial_x \partial_\mu \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k})\left(X_{t_k^n}^n - X_{t_{k-1}^n}^n\right)\theta d\varepsilon, \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\rangle\right] d\theta$$

$$= \int_0^1 \mathbb{E}\left[\langle \int_0^1 \partial_x \partial_\mu \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k})\left(\int_{t_{k-1}^n}^{t_k^n} F_r dr\right)\theta d\varepsilon, \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\rangle\right] d\theta$$

$$+ \int_0^1 \mathbb{E}\left[\langle \int_0^1 \partial_x \partial_\mu \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k})\left(\int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right)\theta d\varepsilon, \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\rangle\right] d\theta$$

$$= \mathbf{I}_{BF}^{n,k} + \mathbf{I}_{BB}^{n,k}.$$

(D.10)

Notice that

$$\sum_{k=1}^n \mathbb{E}\left[\left|\langle \partial_x \partial_\mu \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k})\left(\int_{t_{k-1}^n}^{t_k^n} F_r dr\right), \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\rangle\right|\right] \leq$$

$$\leq M \sum_{k=1}^n \left(\mathbb{E}\left[\left|\int_{t_{k-1}^n}^{t_k^n} F_r dr\right|_H^2\right] \mathbb{E}\left[\left|\int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right|_H^2\right]\right)^{1/2}$$

$$\leq M \sum_{k=1}^n \left(\frac{s-t}{n}\mathbb{E}\left[\int_{t_{k-1}^n}^{t_k^n} |F_r|_H^2 dr\right] \mathbb{E}\left[\int_{t_{k-1}^n}^{t_k^n} |G_r|_{\mathcal{L}_2(K;H)}^2 dr\right]\right)^{1/2}$$

$$\leq M \left(\frac{s-t}{n}\right)^{1/2} \sum_{k=1}^n \left(\mathbb{E}\left[\int_{t_{k-1}^n}^{t_k^n} |F_r|_H^2 dr\right] \mathbb{E}\left[\int_{t_{k-1}^n}^{t_k^n} |G_r|_{\mathcal{L}_2(K;H)}^2 dr\right]\right)^{1/2}$$

$$\leq M \left(\frac{s-t}{n}\right)^{1/2} \left(\mathbb{E}\left[\int_t^s |F_r|_H^2 dr\right] \mathbb{E}\left[\int_t^s |G_r|_{\mathcal{L}_2(K;H)}^2 dr\right]\right)^{1/2},$$

which goes to 0 as $n \to \infty$. By Lebesgue's dominated convergence theorem, it then follows

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbf{I}_{BF}^{n,k} = 0.$$

(D.11)

Now, in order to compute $\mathbf{I}_{BB}^{n,k}$, consider first that, for $m \in \mathbb{N}$ and $n \geq m$,

$$\sum_{k=1}^{n} \mathbb{E}\left[\left|\left\langle\left(\partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k}) - \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1})\right)\left(\int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right), \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right\rangle\right|\right]$$

$$\leq \sum_{k=1}^{n} \mathbb{E}\left[\left|\partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k}) - \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1})\right|_{L(H)}\left|\int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right|_H^2\right]$$

$$\leq \mathbb{E}\left[\sup_{\substack{n\in\mathbb{N},\ n\geq m,\\ k=1,\dots,n}}\left|\partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k}) - \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1})\right|_{L(H)}\left|\int_t^s G_r dB_r\right|_H^2\right].$$

By continuity of $\partial_x\partial_\mu\hat{\varphi}$ and by (D.4) it follows

$$\lim_{m\to\infty}\sup_{\substack{n\in\mathbb{N},\ n\geq m,\\ k=1,\dots,n}}\left|\partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k}(\omega)) - \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1}(\omega))\right| = 0,$$

for every $\omega \in \Omega$, and then, by Lebesgue's dominated convergence theorem,

$$\lim_{n\to\infty}\sum_{k=1}^{n} \mathbb{E}\left[\left|\left\langle\left(\partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k}) - \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1})\right)\left(\int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right),\right.\right.\right.$$
$$\left.\left.\left.\int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right\rangle\right|\right] = 0. \quad \text{(D.12)}$$

By (D.12) and by Lebesgue's dominated convergence theorem, we can then write

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbf{I}_{BB}^{n,k} =$$

$$= \lim_{n\to\infty}\sum_{k=1}^{n}\int_0^1 \mathbb{E}\left[\left\langle\int_0^1 \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1})\left(\int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right)\theta d\varepsilon, \int_{t_{k-1}^n}^{t_k^n} G_r dB_r\right\rangle\right]d\theta$$

$$= \lim_{n\to\infty}\sum_{k=1}^{n}\int_0^1\int_0^1 \theta\mathbb{E}\left[\int_{t_{k-1}^n}^{t_k^n}\text{Tr}\left(G_r G_r^* \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1})\right)dr\right]d\varepsilon d\theta$$

$$= \lim_{n\to\infty}\int_0^1\int_0^1 \theta\mathbb{E}\left[\int_t^s \sum_{k=1}^{n}\text{Tr}\left(G_r G_r^* \partial_x\partial_\mu\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}})(X^{n,\varepsilon\theta,k-1})\right)\mathbb{1}_{[t_{k-1}^n,t_k^n)}(r)dr\right]d\varepsilon d\theta.$$
$$\text{(D.13)}$$

Moreover, by continuity of $\partial_x\partial_\mu\hat{\varphi}$ and by (D.4), we have

$$\lim_{n\to\infty}\left|\sum_{k=1}^{n}\partial_x\partial_\mu\hat{\varphi}\left(t_k^n, \hat{\mathbb{P}}_{X^{n,\theta,k}}\right)\left(X^{n,\varepsilon\theta,k}(\omega)\right)\mathbb{1}_{[t_{k-1}^n,t_k^n)}(r) - \partial_x\partial_\mu\hat{\varphi}\left(r, \hat{\mathbb{P}}_X\right)(X(\omega))\right|_H = 0 \qquad \forall\omega\in\Omega.$$

This, together with (D.13), provides

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbf{I}_{BB}^{n,k} = \frac{1}{2}\int_t^s \mathbb{E}\left[\text{Tr}\left(G_r G_r^* \partial_x\partial_\mu\hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot\wedge r}})(X_{\cdot\wedge r})\right)\right]dr. \quad \text{(D.14)}$$

49

We now consider the quantity $\sum_{k=1}^{n} \hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) - \hat{\varphi}(t_{k-1}^n, \hat{\mathbb{P}}_{X^n})$ appearing in (D.5). By using the continuity of the pathwise time derivative $\partial_t \hat{\varphi}$, we can write

$$\sum_{k=1}^{n} \left(\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) - \hat{\varphi}(t_{k-1}^n, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) \right) = \sum_{k=1}^{n} \int_{t_{k-1}^n}^{t_k^n} \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) dr$$

$$= \int_t^s \sum_{k=1}^{n} \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) \mathbb{1}_{[t_{k-1}^n, t_k^n)}(r) dr \tag{D.15}$$

By continuity of $\partial_t \hat{\varphi}$ and by (D.4), we have

$$\lim_{n \to \infty} \left| \sum_{k=1}^{n} \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) \mathbb{1}_{[t_{k-1}^n, t_k^n)}(r) - \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_X) \right| = 0. \tag{D.16}$$

By (D.15),(D.16), and by Lebesgue's dominated convergence theorem, we obtain

$$\lim_{n \to \infty} \sum_{k=1}^{n} \left(\hat{\varphi}(t_k^n, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) - \hat{\varphi}(t_{k-1}^n, \hat{\mathbb{P}}_{X_{t_{k-1}^n \wedge \cdot}}) \right) = \int_t^s \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_X) dr. \tag{D.17}$$

Putting together (D.5), (D.6), (D.9), (D.10), (D.11), (D.14), (D.17), we finally obtain, by recalling also (D.4),

$$\hat{\varphi}(s, \hat{\mathbb{P}}_X) - \hat{\varphi}(t, \hat{\mathbb{P}}_X) = \lim_{n \to \infty} \left(\hat{\varphi}(s, \hat{\mathbb{P}}_{X^n}) - \hat{\varphi}(t, \hat{\mathbb{P}}_{X^n}) \right)$$

$$= \int_t^s \partial_t \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}}) \, dr + \int_t^s \mathbb{E}\left[\langle F_r, \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \rangle_H \right] dr$$

$$+ \frac{1}{2} \int_t^s \mathbb{E}\left[\text{Tr}\left(G_r G_r^* \partial_x \partial_\mu \hat{\varphi}(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \right) \right] dr,$$

which concludes the proof. $\qquad\square$

E   Consistency property of pathwise derivatives

Proof of Lemma 4.21. The consistency of pathwise time derivatives is a direct consequence of their definition (Definition 4.4).

We now prove the consistency of the other two derivatives using Itô's formula (Theorem 4.16). Fix $t \in [0, T]$ and let $\xi \in \mathbf{S}_2(\mathbb{F})$. Let also $F \colon [0, T] \times \Omega \to H$ (resp. $G \colon [0, T] \times \Omega \to \mathcal{L}_2(K; H)$) be a integrable (resp. square-integrable) and $\mathbb{F}$-progressively measurable process, so, in particular,

$$\int_0^T \mathbb{E}[|F_s|_H] \, ds < \infty, \qquad \int_0^T \mathbb{E}\left[\text{Tr}(G_s G_s^*)\right] ds < \infty.$$

Consider the process $X = (X_s)_{s \in [0, T]}$ given by

$$X_s = \xi_{s \wedge t} + \int_t^{s \vee t} F_r \, dr + \int_t^{s \vee t} G_r \, dB_r, \qquad \forall \, s \in [0, T].$$

Then, by Theorem 4.16 and Remark 4.18 we have the Itô formulae, for every $s \in [t, T]$,

$$\hat{\varphi}_1(s, \hat{\mathbb{P}}_{X_{\cdot \wedge s}}) = \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}}) + \int_t^s \partial_t \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})\, dr + \int_t^s \mathbb{E} \left[\langle F_r, \partial_\mu \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \rangle_H \right] dr$$

$$+ \frac{1}{2} \int_t^s \mathbb{E} \left[\mathrm{Tr} \left(G_r G_r^* \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \right) \right] dr$$

and

$$\hat{\varphi}_2(s, \hat{\mathbb{P}}_{X_{\cdot \wedge s}}) = \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}}) + \int_t^s \partial_t \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})\, dr + \int_t^s \mathbb{E} \left[\langle F_r, \partial_\mu \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \rangle_H \right] dr$$

$$+ \frac{1}{2} \int_t^s \mathbb{E} \left[\mathrm{Tr} \left(G_r G_r^* \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \right) \right] dr.$$

Since $\hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}}) = \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})$ and $\partial_t \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}}) = \partial_t \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})$, for every $r \in [0, T]$, we find

$$\int_t^s \mathbb{E} \left[\langle F_r, \left(\partial_\mu \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) - \partial_\mu \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \right) \rangle_H \right] dr \tag{E.1}$$

$$+ \frac{1}{2} \int_t^s \mathbb{E} \left[\mathrm{Tr} \left(G_r G_r^* \left(\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) - \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(r, \mathbb{P}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \right) \right) \right] dr = 0,$$

for every $s \in [t, T]$.

Consistency of $\partial_\mu \hat{\varphi}_1$ and $\partial_\mu \hat{\varphi}_2$. Let $Z \colon \Omega \to H$ be an $\mathcal{F}_t$-measurable random variable in $L^2(\Omega; H)$, and define

$$F_s := Z \mathbb{1}_{[t,T]}(s), \qquad G_s := 0, \qquad \forall s \in [0, T].$$

Then, from (E.1) we obtain

$$\int_t^s \mathbb{E} \left[\langle Z, \left(\partial_\mu \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) - \partial_\mu \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \right) \rangle_H \right] dr = 0, \qquad \forall s \in [t, T]. \tag{E.2}$$

From the continuity of the map $r \mapsto \partial_\mu \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) - \partial_\mu \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})$, we get in particular that the integrand of (E.2) is equal to zero at $r = t$, namely

$$\mathbb{E} \left[\langle Z, \left(\partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) - \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) \right) \rangle_H \right] = 0.$$

From the arbitrariness of the $\mathcal{F}_t$-measurable random variable $Z \in L^2(\Omega; H)$, this yields

$$\partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) = \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}), \qquad \mathbb{P}\text{-a.s.}$$

Recalling Remark 4.23, we see that this proves the consistency of $\partial_\mu \hat{\varphi}_1$ and $\partial_\mu \hat{\varphi}_2$.

Consistency of $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1$ and $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2$. Let $\Lambda \colon \Omega \to \mathcal{L}_2(K; H)$ be an $\mathcal{F}_t$-measurable random variable in $L^2(\Omega; \mathcal{L}_2(K; H))$, and define

$$F_s := 0, \qquad G_s := \Lambda \mathbb{1}_{[t,T]}(s), \qquad \forall s \in [0, T].$$

Then, from (E.1) we obtain

$$\int_t^s \mathbb{E} \left[\mathrm{Tr} \left(\Lambda \Lambda^* \left(\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) - \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) \right) \right) \right] dr = 0, \quad \forall s \in [t, T].$$

From the continuity of the map $r \mapsto \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r}) - \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(r, \hat{\mathbb{P}}_{X_{\cdot \wedge r}})(X_{\cdot \wedge r})$, we get

$$\mathbb{E}\left[\mathrm{Tr}\left(\Lambda \Lambda^* \big(\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) - \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) \big) \right) \right] = 0.$$

From the arbitrariness of the $\mathcal{F}_t$-measurable random variable $\Lambda \in L^2(\Omega; \mathcal{L}_2(K; H))$, we conclude that

$$\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}) = \partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2(t, \hat{\mathbb{P}}_{\xi_{\cdot \wedge t}})(\xi_{\cdot \wedge t}), \qquad \mathbb{P}\text{-a.s.}$$

which, together with Remark 4.23, gives the claimed consistency of $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_1$ and $\partial_x^{\mathrm{sym}} \partial_\mu \hat{\varphi}_2$. $\qquad \square$

F   Hamilton-Jacobi-Bellman equation: technical results

In the present appendix we prove three technical results which are used in Section 5 to derive alternative forms of the Hamilton-Jacobi-Bellman equation (5.1).

Lemma F.1. *Given $\mu \in \mathcal{P}_2(C([0,T]; H))$ there exists $\xi \in \mathbf{S}_2(\mathcal{G})$, with $\mathbb{P}_\xi = \mu$, and a $\mathcal{G}$-measurable random variable $U_\xi \colon \Omega \to \mathbb{R}$, with uniform distribution on $[0,1]$, such that the following holds:*

- *ξ and U_ξ are independent.*

Proof. Fix $\mu \in \mathcal{P}_2(C([0,T]; H))$ and consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$, where λ denotes the Lebesgue measure on the unit interval. Denote

$$\bar{\Omega} = [0,1] \times C([0,T]; H), \qquad \bar{\mathcal{F}} = \mathcal{B}([0,1]) \otimes \mathscr{B}, \qquad \bar{\mathbb{P}} = \lambda \otimes \mu.$$

Fix an orthonormal basis $\{e_n\}_{n \in \mathbb{N}}$ of H. Then, let $\mathcal{J} \colon [0,1] \times C([0,T]; H) \to C([0,T]; H)$ be the map defined as

$$\mathcal{J}(r,x) := re_1 + \sum_{n=2}^{\infty} \langle x, e_{n-1} \rangle_H e_n, \qquad \forall (r,x) \in [0,1] \times C([0,T]; H).$$

Let $\tilde{\mu}$ denote the image measure (or push-forward) of $\bar{\mathbb{P}} = \lambda \otimes \mu$ by $\mathcal{J}$. Notice that $\tilde{\mu} \in \mathcal{P}_2(C([0,T]; H))$. Then, from property $(\mathbf{A}_{\mathcal{G}})$-ii) it follows that there exists a continuous and $\mathcal{B}([0,T]) \otimes \mathcal{G}$-measurable process $\tilde{\xi} \colon [0,T] \times \Omega \to H$ with law equal to $\tilde{\mu}$.

Now, define the maps $P \colon C([0,T]; H) \to \mathbb{R}$ and $Q \colon C([0,T]; H) \to C([0,T]; H)$ as

$$P(x) := \langle x_0, e_1 \rangle_H, \qquad Q(x) := \sum_{n=1}^{\infty} \langle x, e_{n+1} \rangle_H e_n, \qquad \forall x \in C([0,T]; H),$$

where we recall that x_0 is the value of the path x at time $t = 0$. Then, denote

$$U_\xi := P(\tilde{\xi}), \qquad \xi := Q(\tilde{\xi}).$$

It is then easy to see that U_ξ and ξ satisfy the claimed properties. $\qquad \square$

Lemma F.2. *Fix $t \in [0,T]$, $\xi \in \mathbf{S}_2(\mathcal{G})$ and let $F \colon C([0,T]; H) \times \mathcal{P}_2(C([0,T]; H)) \times \mathrm{U} \times \mathcal{P}(\mathrm{U}) \to \mathbb{R}$ be a measurable function. Suppose that $\mathbb{E}[|F(\xi, \mathbb{P}_\xi, \mathfrak{a}, \mathbb{P}_\mathfrak{a})|] < +\infty, \forall \mathfrak{a} \in \mathcal{M}_t$. Suppose also that*

$$\text{there exists a } \mathcal{G}\text{-measurable random variable } U_\xi \colon \Omega \to \mathbb{R} \tag{F.1}$$
$$\text{having uniform distribution on } [0,1] \text{ and being independent of } \xi.$$

Then, it holds that

$$\sup_{\mathfrak{a}\in\mathcal{M}_t} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big)\big] = \sup_{\mathfrak{a}\in\mathcal{M}_\mathcal{G}} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big)\big] = \sup_{\breve{a}\in\breve{\mathcal{M}}} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\breve{a}(\xi,U_\xi),\mathbb{P}_{\breve{a}(\xi,U_\xi)}\big)\big]. \qquad \text{(F.2)}$$

Moreover, if F is continuous in the first two variables, uniformly with respect to the last two, then the first equality holds true without assuming (F.1).

Proof. Since $\breve{a}(\xi,U_\xi)\in\mathcal{M}_\mathcal{G}$ and $\mathcal{M}_\mathcal{G}\subset\mathcal{M}_t$, we immediately get

$$\sup_{\mathfrak{a}\in\mathcal{M}_t} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big)\big] \geq \sup_{\mathfrak{a}\in\mathcal{M}_\mathcal{G}} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big)\big] \geq \sup_{\breve{a}\in\breve{\mathcal{M}}} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\breve{a}(\xi,U_\xi),\mathbb{P}_{\breve{a}(\xi,U_\xi)}\big)\big].$$

Then, in order to get (F.2), it remains to prove the inequality

$$\sup_{\breve{a}\in\breve{\mathcal{M}}} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\breve{a}(\xi,U_\xi),\mathbb{P}_{\breve{a}(\xi,U_\xi)}\big)\big] \geq \sup_{\mathfrak{a}\in\mathcal{M}_t} \mathbb{E}\big[F\big(\xi,\mathbb{P}_\xi,\mathfrak{a},\mathbb{P}_\mathfrak{a}\big)\big]. \qquad \text{(F.3)}$$

To this end, denote

$$\hat{\Omega} = [0,T]\times\Omega, \qquad \hat{\mathcal{F}} = \mathcal{B}([0,T])\otimes\mathcal{F}, \qquad \hat{\mathbb{P}} = \lambda_T\otimes\mathbb{P},$$

with λ_T being the uniform distribution on $([0,T],\mathcal{B}([0,T]))$. Given $\mathfrak{a}\in\mathcal{M}_t$, consider the canonical extensions of $\mathfrak{a}$ and U_ξ to $\hat{\Omega}$, which will be denoted respectively by $\hat{\mathfrak{a}}$ and $\hat{U}_\xi$ (notice that $\hat{U}_\xi$ has uniform distribution on $[0,1]$ and is independent of ξ). We can now apply Lemma B.1, with $(E,\mathscr{E})$, Γ, $\hat{U}$ being respectively $(C([0,T];H),\mathscr{B})$, ξ, $\hat{U}_\xi$. Then, it follows the existence of a measurable map $\breve{a}\colon C([0,T];H)\times[0,1]\to\mathrm{U}$ such that

$$\big(\xi,\breve{a}(\xi,\hat{U}_\xi)\big) \overset{\mathscr{L}_{\hat{\Omega}}}{=} \big(\xi,\hat{\mathfrak{a}}\big),$$

where $\mathscr{L}_{\hat{\Omega}}$ stands for equality in law between random objects defined on $(\hat{\Omega},\hat{\mathcal{F}},\hat{\mathbb{P}})$. Then, we deduce

$$\Big((\xi_s)_{s\in[0,T]},\breve{a}(\xi,U_\xi)\Big) \overset{\mathscr{L}}{=} \Big((\xi_s)_{s\in[0,T]},\mathfrak{a}\Big),$$

where $\mathscr{L}$ stands for equality in law between random objects defined on $(\Omega,\mathcal{F},\mathbb{P})$. This implies the validity of inequality (F.3) and proves (F.2).

Finally, suppose that F is continuous in the first two variables, uniformly with respect to the last two. Then, in order to get (F.2), it is enough to proceed as in Step 2 of the proof of Theorem 3.6. More precisely, if ξ is discrete, then by Lemma B.3 there exists an $\mathcal{F}_t$-measurable random variable U_ξ having uniform distribution on $[0,1]$ and being independent of ξ. In the general case, we rely on the continuity of F and we approximate ξ by a sequence of discrete random variables. $\quad\square$

Lemma F.3. *Fix $t\in[0,T]$, $\xi\in\mathbf{S}_2(\mathcal{G})$ and let $F\colon C([0,T];H)\times\mathrm{U}\to\mathbb{R}$ be a measurable function. Suppose that $\mathbb{E}[|F(\xi,\mathfrak{a})|]<+\infty$, $\forall\,\mathfrak{a}\in\mathcal{M}_t$, and also that $\sup_{u\in\mathrm{U}}F(x,u)<+\infty$, $\forall\,x\in C([0,T];H)$. Then, it holds that*

$$\sup_{\mathfrak{a}\in\mathcal{M}_t} \mathbb{E}\big[F(\xi,\mathfrak{a})\big] = \sup_{\mathrm{a}\in\mathcal{M}} \mathbb{E}\big[F(\xi,\mathrm{a}(\xi))\big] \qquad \text{(F.4)}$$

and

$$\sup_{\mathfrak{a}\in\mathcal{M}_t} \mathbb{E}\big[F(\xi,\mathfrak{a})\big] = \mathbb{E}\Big[\operatorname*{ess\,sup}_{u\in\mathrm{U}} F(\xi,u)\Big]. \qquad \text{(F.5)}$$

Remark F.4. *Suppose that ξ and U_ξ are as in Lemma F.1, namely $\xi \in \mathbf{S}_2(\mathcal{G})$, $\mathbb{P}_\xi = \mu$, and ξ is such that there exists a $\mathcal{G}$-measurable random variable U_ξ having uniform distribution on $[0,1]$ and being independent of ξ. Then, formula (F.4) holds without assuming that $\sup_{u \in U} F(x,u) < +\infty$, $\forall\, x \in C([0,T]; H)$. As a matter of fact, in this case, thanks to formula (F.2), it is enough to prove that*

$$\sup_{\check{a} \in \check{\mathcal{M}}} \mathbb{E}\big[F(\xi, \check{a}(\xi, U_\xi))\big] = \sup_{a \in \mathcal{M}} \mathbb{E}\big[F(\xi, a(\xi))\big].$$

Clearly, it holds that $\sup_{\check{a} \in \check{\mathcal{M}}} \mathbb{E}[F(\xi, \check{a}(\xi, U_\xi))] \geq \sup_{a \in \mathcal{M}} \mathbb{E}[F(\xi, a(\xi))]$. On the other hand, for every fixed $\check{a} \in \check{\mathcal{M}}$ we have

$$\mathbb{E}\big[F(\xi, \check{a}(\xi, U_\xi))\big] = \int_0^1 \mathbb{E}\big[F(\xi, \check{a}(\xi, r))\big]\, dr \leq \sup_{a \in \mathcal{M}} \mathbb{E}\big[F(\xi, a(\xi))\big].$$

From the arbitrariness of $\check{a}$ we get the other inequality and we conclude that (F.4) holds.

Proof of Lemma F.3. Since $a(\xi) \in \mathcal{M}_t$, we immediately get

$$\sup_{a \in \mathcal{M}} \mathbb{E}\big[F(\xi, a(\xi))\big] \leq \sup_{\mathfrak{a} \in \mathcal{M}_t} \mathbb{E}\big[F(\xi, \mathfrak{a})\big].$$

In addition, for every fixed $\mathfrak{a} \in \mathcal{M}_t$ we have

$$\mathbb{E}\big[F(\xi, \mathfrak{a})\big] \leq \mathbb{E}\Big[\operatorname*{ess\,sup}_{u \in U} F(\xi, u)\Big].$$

From the arbitrariness of $\mathfrak{a}$, we find $\sup_{\mathfrak{a} \in \mathcal{M}_t} \mathbb{E}[F(\xi, \mathfrak{a})] \leq \mathbb{E}[\operatorname{ess\,sup}_{u \in U} F(\xi, u)]$. Then, in order to prove the validity of both (F.4) and (F.5), it remains to prove that

$$\mathbb{E}\Big[\operatorname*{ess\,sup}_{u \in U} F(\xi, u)\Big] \leq \sup_{a \in \mathcal{M}} \mathbb{E}\big[F(\xi, a(\xi))\big]. \tag{F.6}$$

Let $\mu = \mathbb{P}_\xi$ denote the law of ξ. Suppose for a moment that for every $\varepsilon > 0$ there exists $a_\mu^\varepsilon \in \mathcal{M}$ and a μ-null set $N_\varepsilon \in \mathscr{B}$ such that

$$F(x, u) \leq F(x, a_\mu^\varepsilon(x)) + \varepsilon, \qquad \forall\, x \in C([0,T]; H) \backslash N_\varepsilon,\ \forall\, u \in U. \tag{F.7}$$

Then, in particular, there exists a $\mathbb{P}$-null set $\bar{N}_\varepsilon \in \mathcal{F}$ such that

$$F(\xi(\omega), u) \leq F(\xi(\omega), a_\mu^\varepsilon(\xi(\omega))) + \varepsilon, \qquad \forall\, \omega \in \Omega \backslash \bar{N}_\varepsilon,\ \forall\, u \in U.$$

As a consequence, using the definition of essential supremum for the family of real-valued random variables $\{F(\xi, u)\}_{u \in U}$, we find

$$\operatorname*{ess\,sup}_{u \in U} F(\xi, u) \leq F(\xi, a_\mu^\varepsilon(\xi)) + \varepsilon, \qquad \mathbb{P}\text{-a.s.}$$

This yields

$$\mathbb{E}\Big[\operatorname*{ess\,sup}_{u \in U} F(\xi, u)\Big] \leq \mathbb{E}\big[F(\xi, a_\mu^\varepsilon(\xi))\big] + \varepsilon \leq \sup_{a \in \mathcal{M}} \mathbb{E}\big[F(\xi, a(\xi))\big] + \varepsilon.$$

From the arbitrariness of ε we get inequality (F.6) and we conclude that (F.4) holds.

It remains to prove (F.7). To this end, we use that U is a Borel space and we implement Proposition 7.50 in [11] (in particular, X, Y, D, f in the statement of this latter proposition are given respectively by $C([0,T];H)$, U, $C([0,T];H) \times$ U, F). By Proposition 7.50 in [11] it follows that, for every $\varepsilon > 0$, there exists an analytically measurable function (we refer to Definition 7.20 in [11] for the definition of *analytically measurable* function) $a^\varepsilon \colon C([0,T];H) \to$ U such that

$$F(x, a^\varepsilon(x)) \geq \begin{cases} \sup_{u \in U} F(x,u) - \varepsilon, & \text{if } \sup_{u \in U} F(x,u) < +\infty, \\ 1/\varepsilon, & \text{if } \sup_{u \in U} F(x,u) = +\infty, \end{cases}$$

for all $x \in C([0,T];H)$. Since it holds that $\sup_{u \in U} F(x,u) < +\infty$, $\forall\, x \in C([0,T];H)$, we can rewrite the above inequality simply as

$$\sup_{u \in U} F(x,u) \leq F(x, a^\varepsilon(x)) + \varepsilon, \qquad \forall\, x \in C([0,T];H). \tag{F.8}$$

Using Lemma 7.27 in [11] and the fact that U is Borel-isomorphic to a Borel subset of $[0,1]$, we see that there exists a Borel-measurable function $a^\varepsilon_\mu \colon C([0,T];H) \to$ U such that $a^\varepsilon(x) = a^\varepsilon_\mu(x)$ for μ-a.e. $x \in C([0,T];H)$. Hence, from (F.8) we obtain (F.7). $\qquad\blacksquare$

References

[1] B. Acciaio, J. Backhoff-Veraguas, and R. Carmona. Extended mean field control problems: stochastic maximum principle and transport perspective. *SIAM J. Control Optim.*, 57(6):3666–3693, 2019.

[2] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis*. Springer, Berlin, third edition, 2006. A hitchhiker's guide.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[4] M. Bambi. Endogenous growth and time to build: the ak case. *J. of Ec. Dynamics and Control*, 32(4):1015–1040, 2008.

[5] M. Bambi, G. Fabbri, and F. Gozzi. Optimal policy and consumption smoothing effects in the time-to-build AK model. *Economic Theory*, 50(3):635–669, 2012.

[6] V. Barbu, G. Da Prato, and M. Röckner. Stochastic porous media equations. In *Lecture Notes in Mathematics*. Springer, 2016.

[7] V. Barbu, M. Röckner, and D. Zhang. Optimal control of nonlinear stochastic differential equations on Hilbert spaces. *SIAM J. Control Optim.*, 58(4):2383–2410, 2020.

[8] E. Barucci and F. Gozzi. Optimal investment in a vintage capital model. *Res. Econ.*, 52(2):159–188, 1998.

[9] E. Barucci and F. Gozzi. Technology adoption and accumulation in a vintage-capital model. *J. Econ.*, 74(1):1–38, 2001.

[10] A. Bensoussan, J. Frehse, and P. Yam. *Mean field games and mean field type control theory*. Springer, 2013.

[11] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control. The Discrete-Time Case*. Academic Press, New York, 1978.

[12] S. Biagini, F. Gozzi, and M. Zanella. Robust portfolio choice with sticky wages. *Arxiv:2104.12010*, 2021.

[13] E. Biffis, G. Cappa, F. Gozzi, and M. Zanella. Optimal portfolio choice with path dependent labor income: Finite retirement time. *Arxiv:2101.09732*, 2021.

[14] F. Biffis, E. Gozzi and C. Prosdocimi. Optimal portfolio choice with path dependent labor income: the infinite horizon case. *SIAM J. Control Optim.*, 58(4):1906–1938, 2020.

[15] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[16] Jean-Michel Bismut. Growth and optimal intertemporal allocation of risks. *Journal of Economic Theory*, 10(2):239–257, 1975.

[17] R. Boucekkine, C. Camacho, and G. Fabbri. Spatial dynamics and convergence: The spatial ak model. *Journal of Economic Theory*, 148(6):2719–2736, 2013.

[18] R. Buckdahn, J. Li, S. Peng, and C. Rainer. Mean-field stochastic differential equations and associated PDEs. *Annals of Probability*, 45(2):824–878, 2017.

[19] M. Burzoni, V. Ignazio, M. Reppen, and M. Soner. Viscosity solutions for controlled McKean-Vlasov jump-diffusions. *SIAM Journal on Control and Optimization*, 58(3):1676–1699, 2020.

[20] P. Cannarsa and G. Da Prato. Second-order Hamilton-Jacobi equations in infinite dimensions. *SIAM J. Control Optim.*, 29(2):474–492, 1991.

[21] P. Cardaliaguet. Notes on Mean Field Games (from P.-L. Lions' lectures at Collège de France). https://www.ceremade.dauphine.fr/cardaliaguet/MFG20130420.pdf, 2013.

[22] R. Carmona and F. Delarue. The master equation for large population equilibriums. In *Stochastic Analysis and Applications 2014*, pages 77–128. Springer, 2014.

[23] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications vol I. and II.* Probability Theory and Stochastic Modelling. Springer, 2018.

[24] R. Carmona, J.P. Fouque, S.M. Mousavi, and L. Sun. Systemic risk and stochastic games with delay. *JOTA*, 179:366–399, 2018.

[25] S. Cerrai. Optimal control problems for stochastic reaction-diffusion systems with non-Lipschitz coefficients. *SIAM J. Control Optim.*, 39(6):1779–1816, 2001.

[26] S. Cerrai. Stationary Hamilton-Jacobi equations in Hilbert spaces and applications to a stochastic optimal control problem. *SIAM J. Control Optim.*, 40(3):824–852, 2001.

[27] A. Cosso and H. Pham. Zero-sum stochastic differential games of generalized McKean-Vlasov type. *J. Math. Pures Appl. (9)*, 129:180–212, 2019.

[28] G. Da Prato and A. Debussche. Control of the stochastic Burgers model of turbulence. *SIAM J. Control Optim.*, 37(4):1123–1149, 1999.

[29] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 2014.

[30] C. Dellacherie and P.-A. Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, 1978.

[31] B. Djehiche, F. Gozzi, G. Zanco, and M. Zanella. Optimal portfolio choice with path dependent benchmarked labor income: a mean field model. *Stochastic Processes and Applications (to appear). Arxiv:2009.03922.*, 2020.

[32] M. F. Djete, D. Possamaï, and X. Tan. McKean-Vlasov optimal control: the dynamic programming principle. *Preprint arXiv:1907.08860*, 2019.

[33] B. Dupire. Functional itô calculus. *SSRN: 1435551*, 2009.

[34] P.H. Dybvig and H. Liu. Lifetime consumption and investment: retirement and constrained borrowing. *J. Ec. Theory.*, 145:885–907, 2010.

[35] G. Fabbri, F. Gozzi, and A. Swiech. *Stochastic Control in Infinite Dimension*. Springer, 2017.

[36] S. Faggian and F. Gozzi. Optimal investment models with vintage capital: Dynamic programming approach. *J. Math. Econ.*, 46(4):416–437, 2010.

[37] H. Federer. *Geometric measure theory*. Springer, New York, 1969.

[38] S. Federico, B. Goldys, and F. Gozzi. HJB equations for the optimal control of ddes with state constraints i: Regularity of viscosity solutions. *SIAM Journal on Control and Optimization*, 48(8):416–437, 2010.

[39] S. Federico, B. Goldys, and F. Gozzi. HJB equations for the optimal control of differential equations with delays and state constraints, ii: Optimal feedbacks and approximations. *SIAM Journal on Control and Optimization*, 49:2378–2414, 2011.

[40] G. Feichtinger, R. F. Hartl, P. M. Kort, and V. M. Veliov. Anticipation effects of technological progress on capital accumulation: a vintage capital approach. *J. Econom. Theory*, 126(1):143–164, 2006.

[41] J.P. Fouque and Z. Zhang. Mean field game with delay: A toy model. *Risks*, 90(6):risks6030090, 2018.

[42] Panos Fousekis and James S. Shortle. Anticipation effects of technological progress on capital accumulation: a vintage capital approach. *Am. J. of Agr. Econ.*, 77(4):990–1000, 1995.

[43] 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(7):4624–4651, 2010.

[44] 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(3):1397–1465, 2002.

[45] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)*, 125:119–174, 2019.

[46] Cecilia Garcia-Penalosa and Stephen J. Turnovsky. Growth and income inequality: a canonical model. *Economic Theory*, 28:25–49, 2006.

[47] 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(12):1932–1963, 2006.

[48] F. Gozzi. Regularity of solutions of a second order Hamilton-Jacobi equation and application to a control problem. *Comm. Partial Differential Equations*, 20(5-6):775–826, 1995.

[49] F. Gozzi. Global regular solutions of second order Hamilton-Jacobi equations in Hilbert spaces with locally Lipschitz nonlinearities. *J. Math. Anal. Appl.*, 198(2):399–443, 1996.

[50] F. Gozzi and M. Leocata. A stochastic model of economic growth in time-space. *Arxiv:2104.11128.*, 2021.

[51] 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(5):2981–3012, 2017.

[52] F. Gozzi and F. Masiero. Stochastic optimal control with delay in the control II: Verification theorem and optimal feedbacks. *SIAM J. Control Optim.*, 55(5):3013–3038, 2017.

[53] F. Gozzi, S. S. Sritharan, and A. Święch. Bellman equations associated to the optimal feedback control of stochastic Navier-Stokes equations. *Comm. Pure Appl. Math.*, 58(5):671–700, 2005.

[54] F. Gozzi and A. Święch . Hamilton-Jacobi-Bellman equations for the optimal control of the Duncan-Mortensen-Zakai equation. *J. Funct. Anal.*, 172(2):466–510, 2000.

[55] Fausto Gozzi and Federica 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 [MR3702861]. *SIAM J. Control Optim.*, 59(4):3096–3101, 2021.

[56] M. Huang, P. Caines, and R. Malhamé. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information and Systems*, 6(3):221–252, 2006.

[57] Robert E. Lucas Jr. and Benjamin Moll. Hamilton-Jacobi-Bellman equations for the optimal control of the Duncan-Mortensen-Zakai equation. *J. Pol. Econ.*, 122(1), 2014.

[58] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. University of California Press, Berkeley and Los Angeles, 1956.

[59] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[60] F. Kydland and E.C. Prescott. Time to build and aggregate fluctuations. *Econometrica*, 50:1345–1370, 1982.

[61] J.M. Lasry and P.L. Lions. Mean-field games. *Japanese Journal of Mathematics*, 2:229–260, 2007.

[62] P-L. Lions. Cours au collège de france: Théorie des jeux à champ moyens. Audio Conference, 2006-2012.

[63] S. Federico M. Bambi, C. Di Girolami and F. Gozzi. Generically Distributed Investments on Flexible Projects and Endogenous Growth. *Economic Theory*, 63(2):521–558, 2017.

[64] F. Masiero. Semilinear Kolmogorov equations and applications to stochastic optimal control. *Appl. Math. Optim.*, 51(2):201–250, 2005.

[65] H.P. McKean. Propagation of chaos for a class of nonlinear parabolic equations. *Lecture Series in Di erential Equations*, 7(41-57), 1967.

[66] H. Pham and X. Wei. Dynamic programming for optimal control of stochastic McKean-Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55:1069–1101, 2017.

[67] G. Da Prato and A. Debussche. Dynamic programming for the stochastic Navier-Stokes equations. *M2AN Math. Model. Numer. Anal.*, 34(2):459–475, 2000.

[68] G. Da Prato and M. Röckner. Singular dissipative equations in Hilbert spaces. *Probability Theory and Related Fields*, 124(2):261–303, 2002.

[69] M. Rosestolato and A. Święch. Partial regularity of viscosity solutions for a class of Kolmogorov equations arising from mathematical finance. *J. Differential Equations*, 262(3):1897–1930, 2017.

[70] A-S Sznitman. Topics in propagation of chaos. In *Lecture Notes in Mathematics*, pages 165–251. Springer, 1991.

[71] C. Villani. *Optimal Transport Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer Verlag, 2009.

[72] C. Wu and J. Zhang. An elementary proof for the structure of Wasserstein derivatives. *Preprint arXiv:1705.08046*, 2018.

[73] C. Wu and J. Zhang. Viscosity solutions to parabolic master equations and McKean-Vlasov SDEs with closed-loop controls. *Ann. Appl. Probab.*, 30(2):936–986, 2020.