STOCHASTIC MAXIMUM PRINCIPLE FOR SPDES WITH DELAY

GIUSEPPINA GUATTERI, FEDERICA MASIERO, AND CARLO ORRIERI

ABSTRACT. In this paper we develop necessary conditions for optimality, in the form of the stochastic Pontryagin maximum principle, for the optimal control problem of a class of infinite dimensional stochastic evolution equations with delay in the state. In the cost functional we allow the final cost to depend on the history of the state. To treat such kind of cost functionals we introduce a new form of anticipated backward stochastic differential equations which plays the role of dual equation associated to the control problem.

1. INTRODUCTION

In this paper we study the stochastic maximum principle (SMP) for controlled stochastic partial differential equations with delay in the state with a cost functional which depend on the past trajectory also in the final cost.

Maximum principle for evolution equations have been proved in [13], see also the recent papers [6], [7], [4], and [10] for the Banach space case. For what concerns maximum principle for delay equations, after the introduction of the anticipated backward stochastic differential equations (AB-SDEs) in the paper [20], a wide literature have developed. We mention, among others, [1] where a problem with pointwise delay in the state and in the control is studied, [18] where a controlled state equation driven by a Brownian motion and by a Poisson random measure is taken into account, [2] where the linear quadratic case is considered.
In [16] the authors study the stochastic maximum principle for delay evolution equations: the state equation is for some aspects more general than ours, on the contrary, as in the whole recent literature cited above, the final cost does not depend on the past: in view of applications this is a strong restriction.
This is also the main novelty of the present paper towards [19], where the authors consider a stochastic maximum principle for delay evolution equations, and for noise with also a jump component. In general, final costs depending on the past trajectory can not be reduced to current costs depending on the past, as we discuss after Example 1. In order to be able to treat them we have to introduced a new form of Anticipated BSDEs which are studied in Section 4.
Moreover in [16] and in [19] the authors deal with special types of dependence on the past: the evaluation at time $t - \delta$, and this is the only type of dependence in [16], and the mean with respect to the Lebesgue measure, denoted by $A(t, \cdot)$. We are able to consider more general dependence on the past, since the dependence in the drift of the equations as well in the current cost is of integral type, but with respect to general regular measures μ_f and μ_l.

In the present paper we formulate the maximum principle by means of an adjoint equation which turns out to be an ABSDE of a new form. Such equations admit a unique solution only in mild sense, and this, for our general operator A, cannot be deduced by means of the Itô formula as in [19] but exploiting the Martingale Representation Theorem, along the lines of [14].

As we just said the recent literature based on ABSDEs does not take into account the case when the final cost depends on the past trajectory of the state. As far as we know, such a general cost is studied only in [15] for finite dimensional systems, where the adjoint equation is solved directly by the authors, without using the theory of ABSDEs. Towards [15], the novelty of the present paper is that we are able to treat infinite dimensional stochastic control problems.

We also mention that stochastic control problems with delay in the state can also be treated by the dynamic programming principle and the solution of the related Hamilton-Jacobi-Bellmann equation, as it can be done for problems without delay. For stochastic control problems with delay this approach is carried out in [9] in the finite dimensional case and in [23] in the infinite dimensional case. Both in these two cases the problem is not reformulated in a product space given by the present times the past trajectory, but it is directly addressed, as we propose here formulating the stochastic maximum principle by means of an ABSDE.

Comparing the two approaches, we notice that by the stochastic maximum principle we are able to treat state equations (see (3) below) more general than the state equations treated in [9] and [23]: in the aforementioned papers the state equation has to satisfy the structure condition, meaning that the control affects the system only through the noise. Nevertheless, by the Pontryagin stochastic maximum principle we can only give necessary conditions for optimality for a state equation like (3). On the other side in [9] and in [23], the authors arrive at more general results, being able to characterize the value function and the feedback law of the optimal controls.

Before going into details of the matter of the paper, let us present a concrete case we can deal with. Consider a controlled stochastic heat equation in one dimension, with Neumann boundary conditions, and with delay in the state:

$$
\begin{cases}
\dfrac{\partial y}{\partial t}(t,\xi) = \Delta y(t,\xi) + \displaystyle\int_{-d}^{0} \tilde{f}(t, y(t+\theta,\xi), u(t,\xi))\mu_f(d\theta)dt + \tilde{g}(t,\xi)\dot{W}(t,\xi), \\
\hspace{6cm} t \in [0,T],\ \xi \in [0,1], \\
y(\theta,\xi) = x(\theta,\xi), \quad \theta \in [-d,0] \\
\dfrac{\partial}{\partial \xi}y(t,\xi) = 0, \quad \xi = 0,1,\ t \in [0,T].
\end{cases}
\tag{1}
$$

Here $\dot{W}(t,\xi)$ is a space time white noise and $u : \Omega \times [0,T] \times [0,1] \to \mathbb{R}$ is a control process such that $u \in L^2(\Omega \times [0,T], L^2([0,1]))$. The maximum delay is given by $d > 0$ and μ_f is a regular measure on the interval $[-d,0]$. The dependence on the past is given in the drift by

$$
\int_{-d}^{0} \tilde{f}(t, y(t+\theta,\xi), u(t,\xi))\mu_f(d\theta),
$$

so it is of integral type, but with respect to a general regular measure μ_f. The objective is to minimize a cost functional of the form

$$
J(u(\cdot)) = \mathbb{E}\int_0^T \int_{-d}^0 \int_0^1 \tilde{l}(t,\xi,y(t+\theta),u(t,\xi))d\xi\mu_l(d\theta)dt + \mathbb{E}\int_{T-d}^T \int_0^1 \tilde{h}(y(\theta,\xi))d\xi\mu_h(d\theta), \tag{2}
$$

where μ_l and μ_h are regular measures on $[-d,0]$ and $[T-d,T]$, respectively.

Equation (1) can be reformulated as an evolution equation in the Hilbert space $H = L^2([0,1])$, see Section 2.2 for more details, and it will be of the form of the following abstract evolution equation we study in the paper.

Namely in a real and separable Hilbert space H we consider the following controlled state equation:

$$
\begin{cases}
dX(t) = \left[AX(t) + \displaystyle\int_{-d}^0 f(t, X(t+\theta), u(t))\mu_f(d\theta)\right]dt + g(t)dW(t), \\
X_0(\theta) = x(\theta), \qquad \theta \in [-d,0].
\end{cases}
\tag{3}
$$

We assume that A is the generator of a strongly continuous semigroup in H, W is a cylindrical Wiener process in another real and separable Hilbert space K and $f : [0,T] \times H \times U \to H$ is a Lipschitz continuous map with respect to the state X, uniformly with respect to the control u. We refer to Section 2.2 for concrete examples of stochastic partial differential equations we can treat.

The control problem associated to (3) is to minimize the cost functional

$$
J(u(\cdot)) = \mathbb{E}\int_0^T \int_{-d}^0 l(t, X(t+\theta), u(t))\mu_l(d\theta)dt + \mathbb{E}\int_{T-d}^T h(X(\theta))\mu_h(d\theta). \tag{4}
$$

Both in the state equation and in the cost functional the dependence on the past trajectory is forced to be of integral type with respect to a general regular measure. The novelty of introducing the dependence on the past trajectory in the final cost leads to an additional term in the adjoint

equation; namely the adjoint equation, in its mild formulation, is given by

$$
\begin{aligned}
p(t) = & \int_{t \vee (T-d)}^{T} S\left(s-t\right)' D_x h(X(s)) \mu_h(ds) \\
& + \int_{t}^{T} S(s-t)' \mathbb{E}^{\mathcal{F}_s} \int_{-d}^{0} D_x f(s, X(s), u(s-\theta))' p(s-\theta) \mu_f(d\theta)\, ds \\
& + \int_{t}^{T} S(s-t)' \mathbb{E}^{\mathcal{F}_s} \int_{-d}^{0} D_x l(s, X(s+\theta), u(s)) \mu_l(d\theta) ds - \int_{t}^{T} S(t-s)' q(s)\, dW(s),
\end{aligned}
\tag{5}
$$

$$
p(T-\theta) = 0,\ q(T-\theta) = 0, \quad \text{a.e. } \theta \in [-d, 0),\ \mathbb{P} - \text{a.s..}
$$

If we consider a general regular measure μ_h, not necessarily absolute continuous with respect to the Lebesgue measure, the differential form of equation (5) does not make sense, since the term

$$
\frac{d}{dt}\left[\int_{t \vee (T-d)}^{T} S\left(s-t\right)' D_x h(X(s)) \mu_h(ds)\right]
$$

is not well defined, while the integral form (5) makes sense: this allows to give the notion of mild solution also in this case. In order to be able to work with differentials, we will consider an ABSDE where μ_h is approximated by a sequence of regular measures $(\mu_h^n)_{n \geq 1}$ absolute continuous with respect to the Lebesgue measure on $[T-d, T]$, so that once we have also applied the resolvent operator of A, the differential form of this approximating ABSDE makes sense. Having at hand these tools, we are able to state necessary conditions for the optimality: if $(\bar{X}, \bar{u})$ is an optimal pair for the control problem, given a pair of processes

$$
(p, q) \in L_{\mathcal{F}}^2(\Omega \times [0, T+d]; H) \times L_{\mathcal{F}}^2(\Omega \times [0, T+d]; \mathcal{L}_2(K; H))
$$

which are solution to the ABSDE (5), the following variational inequality holds

$$
\langle \frac{\partial}{\partial u} \mathcal{H}(t, \bar{X}_t, \bar{u}(t), p(t)), w - \bar{u}(t) \rangle \geq 0.
$$

Here $\mathcal{H} : [0, T] \times C([-d, 0], H) \times U \times H \to \mathbb{R}$ is the so called Hamiltonian function and it is given by

$$
\mathcal{H}(t, x, u, p) = \langle F(t, x, u), p \rangle_H + L(t, x, u),
\tag{6}
$$

where $F : C([-d, 0], H) \times U \to H$ and $L : [0, T] \times C([-d, 0], H) \times U \to H$ are defined by

$$
F(x, u) := \int_{-d}^{0} f(x(\theta), u) \mu_f(d\theta), \quad L(t, x, u) := \int_{-d}^{0} l(t, x(\theta), u) \mu_l(d\theta).
$$

The paper is organized as follows. In Section 2 we present the problem and state our assumptions. Moreover we give some examples of models we can treat with our techniques. Section 3 is devoted to the study of the state equation and of the related first variation process. In Section 4 we introduce the new form of ABSDE which is the essential tool we need to formulate the stochastic maximum principle, and finally in Section 5 we state and prove the stochastic maximum principle.

2. Assumptions and preliminaries

Throughout the paper, we denote by H and K two real and separable Hilbert spaces with inner product $\langle \cdot, \cdot \rangle_H$ and $\langle \cdot, \cdot \rangle_K$ respectively (if no confusion is possible we will use $\langle \cdot, \cdot \rangle$). We denote by $\mathcal{L}(K; H)$ and $\mathcal{L}_2(K; H)$ the space of linear bounded operators and the Hilbert space of Hilbert-Schmidt operators from K to H, respectively.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let $W(t)$ be a cylindrical Wiener process with values in K. We endow $(\Omega, \mathcal{F}, \mathbb{P})$ with the natural filtration $(\mathcal{F}_t)_{t \geq 0}$ generated by W and augmented in the usual way with the family of $\mathbb{P}$-null sets of $\mathcal{F}$. For any $p \geq 1$ and $T > 0$ we define

- $L_{\mathcal{F}}^p(\Omega \times [0, T]; H)$, the set of all $(\mathcal{F}_t)_{t \geq 0}$-progressively measurable processes with values in H such that

$$
\|X\|_{L_{\mathcal{F}}^p(\Omega \times [0, T]; H)} = \left(\mathbb{E} \int_0^T |X(t)|_H^p dt\right)^{1/p} < \infty;
$$

- $L^p_{\mathcal{F}}(\Omega; C([0,T]; H))$, the set of all $(\mathcal{F}_t)_{t \geq 0}$-progressively measurable processes with values in H such that

$$\|X\|_{L^p_{\mathcal{F}}(\Omega; C([0,T];H))} = \left(\mathbb{E} \sup_{t \in [0,T]} |X(t)|^p_H \right)^{1/p} < \infty.$$

2.1. **Formulation of the control problem.** Let U be another separable Hilbert space and we denote by U_c a convex non empty subspace of U. By admissible control we mean an $(\mathcal{F}_t)_{t \geq 0}$-progressively measurable process with values in U_c such that

$$\mathbb{E} \int_0^T |u(t)|^2_U dt < \infty. \tag{7}$$

We denote by $\mathcal{U}_{ad}$ the space of admissible controls.

Remark 1. *Since we are not dealing with control dependent noise, we could remove the convexity assumption on U_c without introducing the second order adjoint equation. Without the convexity assumption the results we achieve can be proved with the spike variation method, in this case the formulation of the maximum principle in Theorem 6 cannot be given in local form, but in global form, see the book [22].*

The present paper is concerned with the study of the following controlled delay evolution equation in H

$$\begin{cases} dX(t) = \left[AX(t) + \int_{-d}^0 f(t, X_t(\theta), u(t)) \mu_f(d\theta) \right] dt + g(t) dW(t), \\ X_0 = x(\theta), \qquad \theta \in [-d, 0], \end{cases} \tag{8}$$

where μ_f is a regular measure on $[-d, 0]$ with finite total variation, see e.g. [5], paragraph 9.9. Moreover $f : [0,T] \times H \times U \to H$, $g : [0,T] \to \mathcal{L}(K; H)$ and $x(\cdot)$ is a given path. By X_t we mean the past trajectory up to time $t - d$, namely

$$X_t(\theta) = X(t + \theta), \qquad \theta \in [-d, 0]. \tag{9}$$

Let us denote by $E := C([-d, 0], H)$ the space of continuous functions from $[-d, 0]$ to H, endowed with the supremum norm. It turns out that there exists a continuous solution to (8), so we can define an E-valued process $X = (X_t)_{t \in [0,T]}$ by formula (9).

Associated to the controlled state equation (8) we consider a cost functional of the following form:

$$J(u(\cdot)) = \mathbb{E} \int_0^T \int_{-d}^0 l(t, X_t(\theta), u(t)) \mu_l(d\theta) dt + \mathbb{E} \int_{T-d}^T h(X(\theta)) \mu_h(d\theta), \tag{10}$$

where we stress the fact that the final cost depends on the history of the process. Here $l : [0,T] \times H \times U \to \mathbb{R}$, $h : H \to \mathbb{R}$ and μ_l, μ_h are real valued regular measures.

The goal is to minimize the cost functional over all admissible controls. We say that a control $\bar{u}$ is optimal if

$$J(\bar{u}) = \inf_{u(\cdot) \in \mathcal{U}} J(u(\cdot)). \tag{11}$$

Example 1. *Here we give some examples of measures μ_f and μ_l we can treat.*

- *Linear combinations of Dirac functions, that is there exist $k_1, ..., k_n \in [-d, 0]$ such that for every $A \in \mathcal{B}([-d, 0])$*

$$\mu(A) := \sum_{i=1}^n \alpha_i \delta_{k_i}(A), \; k_i \in [-d, 0], \; i = 1, ..., n.$$

- *Discrete measures with infinite support, as an example we mention the probability measure on $[-d, 0]$ given by*

$$\mu(A) := \sum_{i=1}^\infty 2^{-i} \delta_{k_i}(A), \quad k_i = \frac{-d}{i}, \; i \geq 1.$$

- *Measures which are absolute continuous with respect to the Lebesgue measure on $[-d, 0]$, that is there exists $w \in L^1([-d, 0])$ such that*

$$\mu(A) = \int_{-d}^0 w(\theta) d\theta$$

If in the previous points we replace $[-d, 0]$ with $[T - d, T]$ we obtain examples of measures μ_h we can treat.

We notice that if the measure μ_h associated to the final cost is the sum of a measure which is absolutely continuous with respect to the Lebesgue measure in $[T - d, T]$ and of the dirac measure at T, that is

$$\mu_h(A) = \int_A w(\theta)d\theta + \delta_T(A),$$

then the cost (10) can be rewritten as

$$J(u(\cdot)) = \mathbb{E} \int_0^T \int_{-d}^0 l(t, X_t(\theta), u(t))\mu_l(d\theta)dt + \mathbb{E} \int_{T-d}^T h(X(\theta))w(\theta)d\theta + \mathbb{E}h(X(T))$$

$$= \mathbb{E} \int_0^T \left[\int_{-d}^0 l(t, X_t(\theta), u(t))\mu_l(d\theta) + 1_{[T-d,T]}(t)h(X(t))w(t) \right] dt + \mathbb{E}h(X(T)).$$

So in this case the cost can be transformed into a new running cost and a "standard" final cost not depending on the past. The adjoint equation turns out to be a "standard" ABSDE

$$p(t) = \int_{t \vee (T-d)}^T D_x h(X(s))w(s)ds$$

$$+ \int_t^T S(s - t)' \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 D_x f(s - \theta, X(s), u(s - \theta))'p(s - \theta)\mu_f(d\theta)\, ds$$

$$+ \int_t^T S(s - t)' \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 D_x l(s, X(s + \theta), u(s))\mu_l(d\theta)ds - \int_t^T S(t - s)'q(s)\, dW(s),$$

$$p(T) = D_x h(X(T)),\ p(T - \theta) = 0,\ q(T - \theta) = 0, \quad \text{a.e. } \theta \in [-d, 0),\ \mathbb{P} - \text{a.s.}$$

This equation can be also written in differential form as

$$- dp(t) = A'p(t)dt + 1_{(T-d,T)}(t)D_x h(X(t))w(t)dt + \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 D_x l(X(t + \theta), u(t))\mu_l(d\theta)dt$$

$$+ \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 D_x f(X(t), u(t - \theta))'p(t - \theta)\mu_f(d\theta)\, dt - q(t)\, dW(t),$$

$$p(T) = D_x h(X(T)),\ p(T - \theta) = 0,\ q(T - \theta) = 0, \quad \text{a.e. } \theta \in [-d, 0),\ \mathbb{P} \text{ a.s.}$$

This case is a special and simpler case towards the very general final cost we can treat, and for which we are not able to write directly the adjoint equation in differential form.

Remark 2. *We notice that with the approach we present in this paper we can treat costs which are well defined in the space of continuous functions $E = C([-d, 0], H)$ but not in the space of square integrable functions $L^2([-d, 0], H)$, this is the case if the final cost is defined by means of a measure μ_h given e.g. by*

$$\mu_h(A) := \sum_{i=1}^\infty 2^{-i}\delta_{k_i}(A), \quad k_i = T - \frac{d}{i},\ i \geq 1,\ A \in \mathcal{B}([T - d, T]).$$

Therefore, the direct approach we propose allows to treat more general functions than the one based on the reformulation of the problem in the product space $H \times L^2([-d, 0], H)$, which follows the lines of what is done for finite dimensional control problems, see e.g. [3] and references therein.

We now state our assumptions on the coefficients of equation (8). In the following, given three real and separable Hilbert spaces H_1, H_2 and H_3, we denote by $\mathcal{G}^1(H_1, H_2)$ the class of mappings $F : H_1 \to H_2$ such that F is continuous, Gâteaux differentiable on H_1 and $DF : H_1 \to \mathcal{L}(H_1, H_2)$ is strongly continuous. If $H_2 = \mathbb{R}$, we write $\mathcal{G}^1(H_1)$ for $\mathcal{G}^1(H_1, \mathbb{R})$. We denote by $\mathcal{G}^{0,1}([0, T] \times H_1, H_2)$ the class of continuous functions, which are Gâteaux differentiable with respect to $x \in H_1$, and such that $D_x F : [0, T] \times H_1 \to H_2$ is strongly continuous. We also use the immediate generalization to the class $\mathcal{G}^{0,1,1}([0, T] \times H_1 \times H_3, H_2)$ We refer to [11], Section 2.2, for more details on the classes $\mathcal{G}^1$ and $\mathcal{G}^{0,1}$.

Hypothesis 1. *On the coefficients of the controlled state equation we assume that*

(H0) *the measures μ_f and μ_l are finite regular measures on $[-d, 0]$ and μ_h is a finite regular measure on $[T - d, T]$;*

(H1) *the operator $A : D(A) \subset H \to H$ is the infinitesimal generator of a C_0-semigroup $S(t)$ in H and there exist $M > 0$, $\omega > 0$ such that*

$$|S(t)|_{\mathcal{L}(H)} \le Me^{\omega t};$$

(H2) *the map $f : [0, T] \times H \times U \to H$ is measurable, continuous with respect to x and u, and it is Lipschitz continuous with respect to x, uniformly with respect to $t \in [0, T]$ and to $u \in U_c$: $\forall x, y \in H$ there exists $C_1 > 0$ such that*

$$|f(t, x, u) - f(t, y, u)| \le C_1 |x - y|, \quad t \in [0, T], \ u \in U_c;$$

and moreover for all $u \in U_c$, $\forall x \in H$ we assume that

$$|f(t, x, u)| \le C_1 (1 + |x| + |u|), \qquad t \in [0, T], \ u \in U_c;$$

(H3) *the map f is Gâteaux differentiable in x and u, moreover $f \in \mathcal{G}^{0,1,1}([0, T] \times H \times U_c, H)$, and due to the Lipschitz property of f with respect to x we get that $D_x f$ is bounded,*

$$\|D_x f(t, x, u)\|_{\mathcal{L}(H)} \le C_1, \quad \forall t \in [0, T], \ x \in H, \ u \in U_c,$$

moreover we assume that there exists a constant $C_2 > 0$ such that

$$\|D_u f(t, x, u)\|_{\mathcal{L}(U;H)} \le C_2, \quad \forall t \in [0, T], \ x \in H, \ u \in U_c;$$

(H4) *the map $g : [0, T] \to \mathcal{L}(K, H)$ is such that $\forall k \in K$ the map $gk : [0, T] \to H$ is measurable; for every $s > 0$, $t \in [0, T]$ $e^{sA}g(t) \in \mathcal{L}_2(K, H)$ and the following estimate holds true*

$$|e^{sA}g(t)|_{\mathcal{L}_2(K,H)} \le Cs^{-\gamma},$$

for some constant $C > 0$ and $0 \le \gamma < \dfrac{1}{2}$.

(H5) *the map $l : [0, T] \times H \times U \to \mathbb{R}$ belongs to $\mathcal{G}^{0,1,1}([0, T] \times H \times U_c)$ and there exists $j > 0$ and a constant $C_3 > 0$ such that*

$$|D_x l(t, x, u)|_H + |D_u l(t, x, u)|_U \le C_3 (1 + |x|_H^j + |u|_U);$$

(H6) *the map $h : H \to \mathbb{R}$ belongs to $\mathcal{G}^1(H)$ and there exists $k > 0$ and a constant $C_4 > 0$ such that*

$$|D_x h(x)|_H \le C_4 (1 + |x|_H^k)$$

Remark 3. *We notice that even if U_c is convex, the introduction of a multiplicative noise term of the form $g(t, x)$ would add extra difficulties if the noise is not of trace class, since in the first adjoint equation a new term of the form $D_x g(t, x)q$ would appear, and this is not in general well defined, see [8]. We thus decided to avoid this additional difficulty.*

In the next Section we collect some results on existence and uniqueness of a solution to equation (8) and on its regular dependence on the initial condition.

Now we notice that the drift of the state equation (8) can be also rewritten as a linear functional in E, and this linear functional enjoys some regularity inherited by Hypothesis 1 on the coefficients.

Remark 4. *Associated to the coefficient f of the state equation (8), we can define a map $F : [0, T] \times E \times U \to H$ defined by, $\forall t \in [0, T]$, $x \in E$, $u \in U$,*

$$F(t, x, u) := \int_{-d}^0 f(t, x(\theta), u)\mu_f(d\theta). \tag{12}$$

The state equation can be rewritten as

$$\begin{cases} dX(t) = [AX(t) + F(t, X_t, u(t))] \, dt + g(t)dW(t), \\ X(0) = x(\theta), \qquad \theta \in [-d, 0] \end{cases} \tag{13}$$

The same can be done for the cost functional: we can define two maps $L : [0, T] \times E \times U \to \mathbb{R}$ and $H : E \to \mathbb{R}$ defined by, $\forall t \in [0, T]$, $x \in E$, $u \in U$,

$$L(t, x, u) := \int_{-d}^0 l(t, x(\theta), u)\mu_l(d\theta), \qquad H(x) := \int_{T-d}^T h(x(\theta))\mu_h(d\theta) = \int_{-d}^0 h(x(\theta))\mu_h^T(d\theta), \tag{14}$$

where in the last passage we have defined μ_h^T as the measure on $[-d, 0]$ given by

$$\mu_h^T(A) := \mu_h(A + T), \ \forall A \in \mathcal{B}([-d, 0]).$$

The cost functional can be rewritten as

$$J(u(\cdot)) = \mathbb{E} \int_0^T L(t, X_t, u(t)) dt + \mathbb{E} H(X_T). \tag{15}$$

In the following Lemma we show that, under Hypothesis 1 on f, l, h, the maps F, L, H defined in (12) and in (14) satisfy continuity and differentiability property with respect to the trajectory, which belongs to the space E.

Lemma 1. *Let Hypothesis 1 holds true and let F be defined by (12) and L, H be defined by (14). Then F, L, H turn out to be continuous mappings, moreover $F \in \mathcal{G}^{0,1,1}([0,T] \times E \times U_c, H)$, $L \in \mathcal{G}^{0,1,1}([0,T] \times E \times U_c)$ and $H \in \mathcal{G}^1(E)$*

Proof. Let us briefly show, for the reader's convenience, one of the previous implications. Suppose that f satisfies Hypothesis 1, then the map $F(t, x, u) := \int_{-d}^0 f(t, x(\theta), u) \mu_f(d\theta)$ is Gateaux differentiable as a function of x, that is as a function from $E := C([-d, 0], H)$ to H. Indeed, if $h \in E$ we have

$$\begin{aligned}
D_x F(t, x, u)[h] &:= \lim_{\varepsilon \to 0} \frac{F(t, x + \varepsilon h, u) - F(t, x, u)}{\varepsilon} \\
&= \frac{1}{\varepsilon} \int_{-d}^0 [f(t, x(\theta) + \varepsilon h(\theta), u) - f(t, x(\theta), u)] \mu_f(d\theta) \\
&= \int_{-d}^0 D_x f(t, x(\theta), u) h(\theta) \mu_f(d\theta) = \langle D_x f(t, x, u) \mu_f, h \rangle_{E', E}
\end{aligned} \tag{16}$$

where we used the Gateaux differentiability of the map f. In particular, $\|D_F(t, x, u)\|_{\mathcal{L}(E, H)} \leq C$. The other claims follow in the same way. $\qquad\square$

2.2. **Some controlled equations we can treat.** We present here some equations with delay that we can treat with our techniques.

A first class of models include reaction diffusion equations, such as the stochastic heat equation with Neumann boundary conditions (1), already mentioned in the Introduction.

Here we present a controlled stochastic heat equation in one dimension, with Dirichlet boundary conditions, and with delay in the state:

$$\begin{cases}
\dfrac{\partial y}{\partial t}(t, \xi) = \Delta y(t, \xi) + \displaystyle\int_{-d}^0 \tilde{f}(\xi, y(t + \theta, \xi), u(t, \xi)) \mu_f(d\theta) dt + \tilde{g}(t, \xi) \dot{W}(t, \xi), \\
\hphantom{\dfrac{\partial y}{\partial t}(t, \xi) = \Delta y(t, \xi)} t \in [0, T], \ \xi \in [0, 1], \\
y(\theta, \xi) = x(\theta, \xi), \quad \theta \in [-d, 0], \\
y(t, \xi) = 0, \quad \xi = 0, 1, \ t \in [0, T].
\end{cases} \tag{17}$$

The process $\dot{W}(t, \xi)$ is a space time white noise and $u : \Omega \times [0, T] \times [0, 1] \to \mathbb{R}$ is the control process such that $u \in L^2(\Omega \times [0, T], L^2([0, 1]))$. The maximum delay is given by $d > 0$ and μ_f is a regular measure on the interval $[-d, 0]$. The cost functional we can study is given by

$$J(u(\cdot)) = \mathbb{E} \int_0^T \int_{-d}^0 \int_0^1 \tilde{l}(t, \xi, y(t + \theta, \xi), u(t, \xi)) d\xi \mu_l(d\theta) dt + \mathbb{E} \int_{T-d}^T \int_0^1 \tilde{h}(y(\theta, \xi), \xi) d\xi \mu_h(d\theta). \tag{18}$$

The measure μ_l is a regular measure on $[-d, 0]$, and μ_h is a regular measure on $[T - d, T]$.

In abstract reformulation equation (17) is an evolution equation in the Hilbert space $L^2([0, 1])$ and the space of controls U is itself given by $H = L^2([0, 1])$: equation (17) can be rewritten as

$$\begin{cases}
dX(t) = \left[AX(t) + \displaystyle\int_{-d}^0 f(t, (X(t + \theta), u(t)) \mu_f(d\theta) \right] dt + g(t) dW(t), \\
X_0(\theta) = x(\theta), \quad \theta \in [-d, 0],
\end{cases} \tag{19}$$

where $X(t)(\xi) := y(t, \xi)$, A is the Laplace operator with Dirichlet boundary conditions, f is the Nemytskii operator associated to $\tilde{f}$, and g is the multiplicative operator associated to $\tilde{g}$ by

$$(g(t)z)(\xi) := \tilde{g}(t, \xi) z(\xi), \text{ for a.e. } \xi \in [0, 1], \ \forall z \in L^2([0, 1]).$$

The process W is a cylindrical Wiener process in $L^2([0,1])$. In abstract reformulation the cost functional can be rewritten as

$$J(u(\cdot)) = \mathbb{E}\int_0^T \int_{-d}^0 l(t, X(t+\theta), u(t))\mu_l(d\theta)dt + \mathbb{E}\int_{T-d}^T h(X(\theta))\mu_h(d\theta) \qquad (20)$$

where

$$l(t, X(t+\theta), u(t)) = \int_0^1 \tilde{l}(t, \xi, X(t+\theta)(\xi), u(t,\xi))d\xi, \text{ and } h(X(\theta)) = \int_0^1 \tilde{h}(y(\theta,\xi),\xi)d\xi.$$

We now make some assumptions on the coefficients and on the cost functional so that Hypothesis 1 is satisfied.

Hypothesis 2. *The functions $\tilde{f}$, $\tilde{g}$, $\tilde{l}$ and $\tilde{h}$ are all measurable and real valued. Moreover*
  (1) *$\tilde{f}$ is defined on $[0,1] \times \mathbb{R} \times \mathbb{R}$ and there exists a constant $L > 0$ such that*

$$\left|\tilde{f}(\xi, x, u) - \tilde{f}(\xi, y, u)\right| \le L|x - y|,$$

  for almost all $\xi \in [0,1]$, for all $x, y \in \mathbb{R}$ and $u \in U$. Moreover for almost all $\xi \in [0,1]$, $\forall u \in \mathbb{R}$, $\tilde{f}(\xi, \cdot, u) \in C_b^1(\mathbb{R})$ and $\forall x \in \mathbb{R}$, $\tilde{f}(\xi, x, \cdot) \in C_b^1(\mathbb{R})$;
  (2) *$\tilde{g}$ is defined on $[0,1] \times [0,T]$ and there exist a constants $K > 0$ such that for almost all $\xi \in [0,1]$ and for $t \in [0,T]$*

$$|\tilde{g}(\xi, t)| \le K;$$

  (3) *$\tilde{l}: [0,T] \times [0,1] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is measurable and for a.a. $t \in [0,T]$, $\xi \in [0,1]$, the map $\tilde{l}(\tau, \xi, \cdot, \cdot): \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is continuous. There exists c_1 square integrable on $[0,1]$ such that*

$$\left|\tilde{l}(t, \xi, x, u) - \tilde{l}(t, \xi, y, u)\right| \le c_1(\xi)|x - y|,$$

  for $\xi \in [0,1]$, $x, y \in \mathbb{R}$, $t \in [0,T]$, and $u \in \mathbb{R}$. Moreover $\left|\tilde{l}(t, \xi, x, u)\right| \le c_2(\xi)$, with c_2 integrable on $[0,1]$. In addition, for a.a. $\xi \in [0,1]$ and $t \in [0,T]$, and for $x, u \in \mathbb{R}$ $\tilde{l}(t, \xi, \cdot, u) \in C_b^1(\mathbb{R})$ and $\tilde{l}(t, \xi, x, \cdot) \in C_b^1(\mathbb{R})$.
  (4) *$\tilde{h}$ is defined on $\mathbb{R} \times [0,1] \times \mathbb{R}$ and $\tilde{h}(\cdot, \xi)$ is uniformly continuous, uniformly with respect to $\xi \in [0,1]$; moreover $\left|\tilde{h}(x, \xi)\right| \le c_3(\xi)$, with c_3 integrable on $[0,1]$;*
  (5) *$x_0 \in L^2([0,1])$.*

It is immediate to see that if Hypothesis 2 holds true for the coefficients $\tilde{f}$, $\tilde{g}$, $\tilde{l}$ and $\tilde{h}$, then f, g, l and h satisfy Hypothesis 1.

Another class of systems we can treat is given by controlled stochastic wave equations: we consider, for $0 \le t \le T$ and $\xi \in [0,1]$, the following wave equation:

$$\begin{cases} \frac{\partial^2}{\partial t^2}y(t,\xi) = \frac{\partial^2}{\partial \xi^2}y(t,\xi) + \int_{-d}^0 \tilde{f}(\xi, y(t+\theta,\xi), u(t,\xi))\mu_f(d\theta) + \dot{W}(t,\xi) \\ y(t,0) = y(t,1) = 0, \\ y(0,\xi) = x_0(\xi), \\ \frac{\partial y}{\partial t}(t,\xi)|_{t=0} = x_1(\xi), \end{cases} \qquad (21)$$

where $\dot{W}_t(\xi)$ is a space-time white noise on $[0,T] \times [0,1]$ and u is the control process in $L^2(\Omega \times [0,T], L^2(0,1))$. The cost functional we are interested in is the following

$$J(x_0, x_1, u) = \mathbb{E}\int_0^T \int_{-d}^0 \int_0^1 \tilde{l}(t, \xi, y(t+\theta,\xi), u(t,\xi))d\xi\mu_l(d\theta)dt$$
$$+ \mathbb{E}\int_{T-d}^T \int_0^1 \tilde{h}(\xi, y(\theta,\xi))\mu_h(d\theta)d\xi. \qquad (22)$$

The measures μ_f and μ_l are regular measures on $[-d, 0]$, and μ_h is a regular measure on $[T-d, T]$. On the drift $\tilde{f}$ and on the costs $\tilde{l}$ and $\tilde{h}$ we make the following assumptions:

Hypothesis 3. *The functions $\tilde{f}$, $\tilde{l}$ and $\tilde{h}$ are all measurable and real valued. Moreover*

(1) $\tilde{f}$ is defined on $[0,1] \times \mathbb{R} \times \mathbb{R}$ and there exists a constant $C > 0$ such that, for a.a. $\xi \in [0,1]$, $\forall x, y, u \in \mathbb{R}$

$$\left| \tilde{f}(\xi, x, u) - \tilde{f}(\xi, y, u) \right| \leq C \left| x - y \right|.$$

Moreover for almost all $\xi \in [0,1]$, for all $x, y \in \mathbb{R}$, $f(\xi, \cdot, u) \in C_b^1(\mathbb{R})$ and $f(\xi, x, \cdot) \in C_b^1(\mathbb{R})$.

(2) $\tilde{l} : [0,T] \times [0,1] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is such that for a.a. $t \in [0,T]$, $\xi \in [0,1]$, the map $\tilde{l}(\tau, \xi, \cdot, \cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is continuous. There exists c_1 square integrable on $[0,1]$ such that

$$\left| \tilde{l}(t, \xi, x, u) - \tilde{l}(t, \xi, y, u) \right| \leq c_1(\xi) \left| x - y \right|,$$

for $\xi \in [0,1]$, $x, y \in \mathbb{R}$, $t \in [0,T]$ and $u \in \mathbb{R}$. There exists c_2 integrable on $[0,1]$ such that $\left| \tilde{l}(t, \xi, x, u) \right| \leq c_2(\xi)$. Moreover for a.a. $\xi \in [0,1]$ and $t \in [0,T]$, and for $x, u \in \mathbb{R}$ $\tilde{l}(t, \xi, \cdot, u) \in C_b^1(\mathbb{R})$ and $\tilde{l}(t, \xi, x, \cdot) \in C_b^1(\mathbb{R})$.

(3) $\tilde{h}$ is defined on $\mathbb{R} \times [0,1] \times \mathbb{R}$ and $\tilde{h}(\cdot, \xi)$ is uniformly continuous, uniformly with respect to $\xi \in [0,1]$; moreover $\left| \tilde{h}(x, \xi) \right| \leq c_3(\xi)$, with c_3 integrable on $[0,1]$ and for almost all $\xi \in [0,1]$ $\tilde{h}(\cdot, \xi) \in C_b^1(\mathbb{R})$;

(4) $x_0, x_1 \in L^2([0,1])$.

We want to write equation (21) in an abstract form. We denote by Λ the Laplace operator with Dirichlet boundary conditions and we introduce the Hilbert space

$$H = L^2([0,1]) \oplus \mathcal{D}\left(\Lambda^{-\frac{1}{2}}\right) = L^2([0,1]) \oplus H^{-1}([0,1]).$$

On H we define the operator A by

$$\mathcal{D}(A) = H_0^1([0,1]) \oplus L^2([0,1]), \qquad A \begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} 0 & I \\ -\Lambda & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}, \quad \text{for every } \begin{pmatrix} y \\ z \end{pmatrix} \in \mathcal{D}(A).$$

We also set $f : H \times L^2([0,1]) \to L^2([0,1])$ given by

$$f\left(\begin{pmatrix} y \\ z \end{pmatrix}, u \right)(\xi) := \tilde{f}(\xi, y(\xi), u(\xi)), \quad \text{for all } \begin{pmatrix} y \\ z \end{pmatrix} \in H, \xi \in [0,1],$$

and $g : L^2([0,1]) \longrightarrow H$ with

$$gu = \begin{pmatrix} 0 \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ I \end{pmatrix} u,$$

for all $u \in L^2([0,1])$.

Equation (21) can be rewritten in an abstract way as an equation in H of the following form:

$$\begin{cases} dX(t) = AX(t)dt + g \int_{-d}^0 f((X(t+\theta), u(t))\, dt + g dW_t, & t \in [0,T] \\ X_0 = x. \end{cases} \tag{23}$$

The cost functional (22)

$$J(u(\cdot)) = \mathbb{E} \int_0^T \int_{-d}^0 l(t, X(t+\theta), u(t)) \mu_l(d\theta) dt + \mathbb{E} \int_{T-d}^T h(X(\theta)) \mu_h(d\theta) \tag{24}$$

where

$$l(t, X(t+\theta), u(t)) = \int_0^1 \tilde{l}(t, \xi, X_1(t+\theta, \xi), u(t, \xi)) d\xi, \text{ and } h(X(\theta)) = \int_0^1 \tilde{h}(X_1(\theta, \xi), \xi) d\xi.$$

where X_1, X_2 are the first and the second component of an element $X \in H$, namely $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

It is immediate to see that if Hypothesis 3 holds true for the coefficients $\tilde{f}$, $\tilde{l}$ and $\tilde{h}$, then f, l and h satisfy Hypothesis 1.

We notice that, since in its abstract formulation equation (21) reads as (23), it satisfies the structure condition, and so the control problem could be treated by solving the associated Hamilton Jacobi Bellmann equation, as in [23]. With our techniques we can treat equations more general than equation (21), for example with a diffusion term $\tilde{g} : [0,1] \to \mathbb{R}$ not invertible.

3. Analysis of the state equation

In this Section we study the state equation (8) and its behaviour with respect to a convex perturbation of the optimal control. Here and in the following, for the sake of brevity, we consider a drift f and a current cost l which do not depend on time. A mild solution to equation (8) is an $\mathcal{F}_t$-progressively measurable process, satisfying $\mathbb{P}$-a.s., for $t \in [0,T]$ the integral equation

$$X(t) = S(t)x(0) + \int_0^t S(t-s) \int_{-d}^0 f(X_s(\theta), u(s))\mu_f(d\theta)ds + \int_0^t S(t-s)g(s)dW(s). \tag{25}$$

We refer e.g. to [3] for the basic properties of mild solution of evolution equations. In the following Proposition we prove existence of a mild solution for equation (8), for every admissible control u, and we show that this mild solution has smooth dependence with respect to the initial condition.

Proposition 1. *Let Hypothesis 1, points H0, H1, H2, H4, hold true. Then there exists a unique mild solution X to equation (8) and for every $p \geq 1$ there exists a constant $c > 0$ such that*

$$\mathbb{E} \sup_{t \in [-d,T]} |X(t)|_H^p \leq c(1 + |x|)_{C([-d,0];H)}^p. \tag{26}$$

Proof. This result is collected in [23], Theorem 3.2. In [23] no proof is given, referring to [11], where the case of H-valued stochastic evolution equations without delay is considered.

The proof follows by techniques in [11] and in [17], we just give a sketch. We introduce the map

$$\Gamma : L_{\mathcal{F}}^p(\Omega, C([0,T], E)) \mapsto L_{\mathcal{F}}^p(\Omega, C([0,T], E))$$

defined by, $\forall \theta \in [-d,0]$,

$$\left(\Gamma(Y)_t\right)(\theta) := \tilde{\Gamma}(Y)_{t+\theta}, \tag{27}$$

where for all $s \in [-d,T]$, we define $\tilde{\Gamma}_s : L_{\mathcal{F}}^p(\Omega, C([0,T], H)) \mapsto L_{\mathcal{F}}^p(\Omega, C([0,T], H))$ as

$$\tilde{\Gamma}(Y)_s = x(s), \qquad s \in [-d,0], \tag{28}$$

$$\tilde{\Gamma}(Y)_s = S(s)x(0) + \int_0^s S(s-r) \int_{-d}^0 f(Y(r+\theta), u(r))\mu_f(d\theta)dr$$

$$+ \int_0^t S(s-r)g(r)dW(r), \quad s \in [0,T].$$

It turns out that Γ is well defined from $L_{\mathcal{F}}^p(\Omega, C([0,T], E))$ to $L_{\mathcal{F}}^p(\Omega, C([0,T], E))$ and it is a contraction: to this aim it suffices to notice that

$$\mathbb{E} \sup_{t \in [-d,T]} |X(t)|^p = \mathbb{E} \sup_{t \in [0,T]} \|X_t\|_E^p$$

and to apply arguments in [11], Propositions 3.2 and 3.3. The crucial point is the fact that f is Lipschitz continuous and the measure μ_f has finite total variation. $\square$

Let $(\bar{X}(\cdot), \bar{u}(\cdot))$ be an optimal pair. For any $w(\cdot) \in \mathcal{U}_{ad}$ we can define the perturbed control

$$u^\rho(\cdot) := \bar{u}(\cdot) + \rho v(\cdot) \tag{29}$$

where $v(\cdot) = w(\cdot) - \bar{u}(\cdot)$ and $0 \leq \rho \leq 1$. The perturbed control $u^\rho(\cdot)$ is admissible due to the convexity of the set U_c and the corresponding state is denoted by X^ρ.

Lemma 2. *Under Hypothesis 1, points H0, H1, H2 and H4, the following holds*

$$\mathbb{E} \sup_{t \in [0,T]} |X^\rho(t) - \bar{X}(t)|_H^2 \to 0, \quad as \ \rho \to 0.$$

Proof. Let us write the equation satisfied by $X^\rho - \bar{X}$ in mild form

$$X^\rho(t) - \bar{X}(t) = \int_0^t S(t-s) \int_{-d}^0 \left[f(X^\rho(s+\theta), u^\rho(s)) - f(\bar{X}(s+\theta), \bar{u}(s)) \right] \mu_f(d\theta)ds. \tag{30}$$

Then we have

$$|X^\rho(t) - \bar{X}(t)|_H^2 \le C \int_0^t \|S(t-s)\|_{\mathcal{L}(H)}^2 \int_{-d}^0 |f(X^\rho(s+\theta), u^\rho(s)) - f(\bar{X}(s+\theta), \bar{u}(s))|_H^2 |\mu_f|(d\theta) ds$$

$$\le C \int_0^t \|S(t-s)\|_{\mathcal{L}(H)}^2 \int_{-d}^0 |f(X^\rho(s+\theta), u^\rho(s)) - f(\bar{X}(s+\theta), u^\rho(s))|_H^2 |\mu_f|(d\theta) ds$$

$$+ C \int_0^t \|S(t-s)\|_{\mathcal{L}(H)}^2 \int_{-d}^0 |f(\bar{X}(s+\theta), u^\rho(s)) - f(\bar{X}(s+\theta), \bar{u}(s))|_H^2 |\mu_f|(d\theta) ds.$$

Thanks to Hypotheses H1, H2 and the finiteness of the total variation measure $|\mu_f|$, taking the supremum in $t \in [0,T]$ we obtain

$$|X^\rho(t) - \bar{X}(t)|_H^2 \le K \int_0^t \sup_{r \in [0,s]} |X^\rho(r) - \bar{X}(r)|_H^2 ds + \nu^\rho(t)$$

$$\le K \int_0^T \sup_{r \in [0,s]} |X^\rho(r) - \bar{X}(r)|_H^2 ds + \nu^\rho(T) \tag{31}$$

where K depends only on T and we denoted by $\nu^\rho(t)$ the quantity

$$\nu^\rho(T) := \int_0^T \int_{-d}^0 |f(\bar{X}(s+\theta), u^\rho(s)) - f(\bar{X}(s+\theta), \bar{u}(s))|_H^2 |\mu_f|(d\theta) ds.$$

Let us notice that, by H2, by estimates (26), and the definition of u^ρ in (29), we can apply the Dominated Convergence Theorem. By the continuity of the map f with respect to u, we get $\nu^\rho(t) \to 0$ if $\rho \to 0$. Now we can take the supremum over the entire interval $[0,T]$ to obtain

$$\sup_{t \in [0,T]} |X^\rho(t) - \bar{X}(t)|_H^2 \le K \int_0^T \sup_{t \in [0,s]} |X^\rho(t) - \bar{X}(t)|_H^2 ds + \nu^\rho(T).$$

Using the Gronwall Lemma we get

$$\mathbb{E} \sup_{t \in [0,T]} |X^\rho(t) - \bar{X}(t)|_H^2 \le \tilde{K} \mathbb{E} \left[\nu^\rho(T) \right], \tag{32}$$

and letting $\rho \to 0$ we get the required result. $\qquad \square$

Now we can introduce the first variation process $Y(t)$, which satisfies the following equation

$$\begin{cases} \dfrac{d}{dt} Y(t) = AY(t) + \displaystyle\int_{-d}^0 D_x f(X(t+\theta), u(t)) Y(t+\theta) \mu_f(d\theta) \\ \qquad\qquad + \displaystyle\int_{-d}^0 D_u f(X(t+\theta), u(t)) \mu_f(d\theta) v(t) \\ Y_0 = 0. \end{cases} \tag{33}$$

This equation is well-posed in a mild sense, as we show in the following

Proposition 2. *Let Hypothesis 1 be in force. Then equation* (33) *admits a unique mild solution solution, i.e. a progressive H-valued process $Y \in L_{\mathcal{F}}^2(\Omega \times [0,T]; H)$ such that for all $t \in [0,T]$, $\mathbb{P}$-a.s.*

$$Y(t) = \int_0^t S(t-s) \int_{-d}^0 [D_x f(X(s+\theta), u(s)) Y(s+\theta) + D_u f(X(s+\theta), u(s)) v(s)] \, \mu_f(d\theta) ds \tag{34}$$

Proof. The proof follows like the one of Proposition 1, the equation here is linear and there is no diffusion term. $\qquad \square$

Once we have defined the the first variation process Y, we can write an expansion up to the first order of the perturbed trajectory of the state

$$X^\rho(t) = \bar{X}(t) + \rho Y(t) + R^\rho(t), \quad t \in [0,T]. \tag{35}$$

The aim is to show that the rest goes to zero in the right topology, i.e.

$$\lim_{\rho \to 0} \frac{1}{\rho^2} \mathbb{E} \sup_{t \in [0,T]} |R^\rho(t)|^2 = 0,$$

This is the content of the following

Lemma 3. *Let Hypothesis 1, points H0, H1, H2, H3, H4, holds. Then the process $\tilde{X}^\rho$ defined as*

$$\tilde{X}^\rho(t) = \frac{X^\rho(t) - \bar{X}(t)}{\rho} - Y(t), \tag{36}$$

satisfies the following

$$\lim_{\rho \to 0} \mathbb{E} \sup_{t \in [0,T]} |\tilde{X}^\rho(t)|^2 = 0. \tag{37}$$

Proof. To lighten the notation, here we adopt the convention introduced in (9). The process $\tilde{X}^\rho$ defined in (36) is the solution to the following integral equation

$$\tilde{X}^\rho(t) = \int_0^t S(t-s)\frac{1}{\rho}\int_{-d}^0 \big[f(\bar{X}_s(\theta) + \rho Y_s(\theta) + \rho \tilde{X}_s^\rho(\theta), \bar{u}(s) + \rho v(s)) - f(\bar{X}_s(\theta), \bar{u}(s))\big]\mu_f(d\theta)ds$$

$$- \int_0^t S(t-s)\int_{-d}^0 \big[D_x f(\bar{X}_s(\theta), \bar{u}(s))Y_s(\theta) + D_u f(\bar{X}_s(\theta), \bar{u}(s))v(s)\big]\mu_f(d\theta)ds$$

with $\tilde{X}^\rho(0) = 0$ as initial datum. Via standard computations we obtain

$$\tilde{X}^\rho(t) = \int_0^t S(t-s)\int_0^1\int_{-d}^0 D_x f\big(\bar{X}_s(\theta) + \lambda\rho(Y_s(\theta) + \tilde{X}_s^\rho(\theta)), \bar{u}(s)\big)\tilde{X}_s^\rho(\theta)\,\mu_f(d\theta)d\lambda ds$$

$$+ \int_0^t S(t-s)\int_0^1\int_{-d}^0 \big[D_x f\big(\bar{X}_s(\theta) + \lambda\rho(Y_s(\theta) + \tilde{X}_s^\rho(\theta)), \bar{u}(s)\big)$$

$$- D_x f(\bar{X}_s(\theta), \bar{u}(s))\big]Y_s(\theta)\mu_f(d\theta)d\lambda ds$$

$$+ \int_0^t S(t-s)\int_0^1\int_{-d}^0 \big[D_u f\big(\bar{X}_s(\theta) + \lambda\rho(Y_s(\theta) + \tilde{X}_s^\rho(\theta)), \bar{u}(s) + \lambda\rho v(s)\big)$$

$$- D_u f(\bar{X}_s(\theta), \bar{u}(s))\big]v(s)\,\mu_f(d\theta)d\lambda ds,$$

so that

$$|\tilde{X}^\rho(t)|_H^2 \leq K\int_0^t \|S(t-s)\|_{\mathcal{L}(H)}^2 \Big|\int_{-d}^0 \tilde{X}_s^\rho(\theta)\mu_f(d\theta)\Big|_H^2 ds$$

$$+ \int_0^T \|S(T-t)\|_{\mathcal{L}(H)}^2\Big|\int_0^1\int_{-d}^0\big[D_x f\big(\bar{X}_t(\theta) + \lambda\rho(Y_t(\theta) + \tilde{X}_t^\rho(\theta)), \bar{u}(t)\big)$$

$$- D_x f(\bar{X}_t(\theta), \bar{u}(t))\big]Y_t(\theta)\mu_f(d\theta)d\lambda\Big|_H^2 dt$$

$$+ \int_0^T \|S(T-t)\|_{\mathcal{L}(H)}^2\Big|\int_0^1\int_{-d}^0\big[D_u f\big(\bar{X}_t(\theta) + \lambda\rho(Y_t(\theta) + \tilde{X}_t^\rho(\theta)), \bar{u}(t) + \lambda\rho v(t)\big)$$

$$- D_u f(\bar{X}_t(\theta), \bar{u}(t))\big]v(t)\,\mu_f(d\theta)d\lambda\Big|_H^2 dt$$

$$\leq K\int_0^T \Big|\int_{-d}^0 \tilde{X}_s^\rho(\theta)\mu_f(d\theta)\Big|_H^2 ds + \nu_\rho(T),$$

where the constant K only depends on T and $\nu_\rho(T)$ is defined by

$$\nu_\rho(T) = \int_0^T \|S(T-t)\|_{\mathcal{L}(H)}^2\Big|\int_0^1\int_{-d}^0\big[D_u f\big(\bar{X}_t(\theta) + \lambda\rho(Y_t(\theta) + \tilde{X}_t^\rho(\theta)), \bar{u}(t) + \lambda\rho v(t)\big)$$

$$- D_u f(\bar{X}_t(\theta), \bar{u}(t))\big]v(t)\,\mu_f(d\theta)d\lambda\Big|_H^2 dt. \tag{38}$$

By the boundedness of $D_x f(\cdot)$ and $D_u f(\cdot)$, we have that $\nu_\rho(t) \to 0$ as $\rho \to 0$.

$$|\tilde{X}^\rho(t)|_H^2 \leq K\int_0^T \sup_{r\in[0,s]} |\tilde{X}^\rho(r)|_H^2 ds + \nu_\rho(T)$$

$$\leq K\int_0^T \sup_{r\in[0,s]} |\tilde{X}^\rho(r)|_H^2 ds + \nu_\rho(T), \tag{39}$$

Now we can take the supremum on the left hand side

$$\mathbb{E}\sup_{t\in[0,T]}|\tilde{X}^\rho(t)|_H^2 \le K\mathbb{E}\int_0^T \sup_{r\in[0,s]}|\tilde{X}^\rho(r)|_H^2 ds + \mathbb{E}\left[\nu_\rho(T)\right]. \tag{40}$$

Using the Gronwall inequality and the convergence $\nu_\rho(t) \to 0$, as $\rho \to 0$, we get the result. $\qquad\square$

4. A new form of Anticipated Backward SPDE

In this section we study ABSDEs with suitable form to be the adjoint equation in problems with delay. The solvability of this class of equations allows to formulate a stochastic maximum principle for infinite dimensional controlled state equations with delay, with final cost functional depending on the history of the process.

Many recent papers, we cite among others [1], [2], [16], [18] and [19], deal with similar problems, but only in the case of final cost not depending on the past of the process. Such a general case is treated in the paper [15] for finite dimensional systems, and it is here proved by means of ABSDEs.

We will consider the following infinite dimensional ABSDE, which in integral form is given by

$$
\begin{aligned}
p(t) = &\int_{t\vee(T-d)}^T S\left(s-t\right)' D_x h(X(s))\mu_h(ds)\\
&+ \int_t^T S(s-t)' \mathbb{E}^{\mathcal{F}_s}\int_{-d}^0 D_x f(X(s),u(s-\theta))'p(s-\theta)\mu_f(d\theta)\,ds\\
&+ \int_t^T S(s-t)' \mathbb{E}^{\mathcal{F}_s}\int_{-d}^0 D_x l(X(s+\theta),u(s))\mu_l(d\theta)ds - \int_t^T S(t-s)'q(s)\,dW(s), \quad t\in[0,T]\\
p(t) = &\,0, \qquad \text{for all } t\in]T,T+d], \mathbb{P}-\text{a.s.}, \qquad q(t)=0 \quad \text{a.e. } t\in]T,T+d], \mathbb{P}-\text{a.s.}.
\end{aligned}
\tag{41}
$$

If we do not make additional assumption on the measure μ_h, the differential form of equation (41) does not make sense, since the term

$$d_t\left[\int_{t\vee(T-d)}^T S\left(s-t\right)' D_x h(X(s))\mu_h(ds)\right]$$

is not well defined for every $t \in [0,T]$, while the notion of mild solution is meaningful also in this case.

Definition 1. *We say that* (41) *admits a mild solution if there is a couple of processes*

$$(p,q)\in L^2_\mathcal{F}(\Omega\times[0,T+d];H)\times L^2_\mathcal{F}(\Omega\times[0,T+d];\mathcal{L}_2(K;H))$$

that solves equation (41) *for all $t\in[0,T+d]$.*

We will prove existence and uniqueness for such equation by an approximation procedure. We start by recalling the following apriori estimates.

Lemma 4. *Let $\gamma \in L^2_\mathcal{F}(\Omega\times[0,T];H)$ be a progressively measurable process and $\xi\in L^2_{\mathcal{F}_T}(\Omega;H)$. Then the mild solution to*

$$-dp(t) = A'p(t)dt + \gamma(t)dt - q(t)dW(t), \qquad p(T)=\xi$$

satisfies the following a priori estimate $\forall t\in[0,T]$, for every $\beta>0$:

$$\mathbb{E}\int_0^T\left(|p(t)|_H^2 + |q(t)|_{\mathcal{L}_2(K;H)}^2\right)e^{\beta t}dt \le C\left[\mathbb{E}[|\xi|_H^2 e^{\beta T}] + \frac{2}{\beta}\mathbb{E}\int_0^T|\gamma(t)|_H^2 e^{\beta t}dt\right] \tag{42}$$

for some constant C depending on T and the constants appearing in Hypothesis 1.

Proof. The solution to this equation was proved in [14], Proposition 2.1, see also [12] for the estimates with exponential weights, Theorem 4.4 in particular estimate (4.13) and (4.14). $\qquad\square$

Suppose first that the final cost is independent of the past, which is to say that in Hypothesis 1 μ_h coincides with δ_T, the Dirac measure in T. If the final cost is independent on the past, in

the adjoint equation (41) the terminal condition is given at the final time T and it is equal to $D_x h(X(T)) \in L^2_{\mathcal{F}_T}(\Omega; H)$. Let us consider the following anticipated backward SPDE (ABSDE):

$$-dp(t) = A'p(t)\, dt + \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 D_x f(X(t), u(t-\theta))' p(t-\theta) \mu_f(d\theta) dt$$

$$+ \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 D_x l(X(t+\theta), u(t)) \mu_l(d\theta) dt - q(t)\, dW(t) \tag{43}$$

$$p(T) = D_x h(X(T))$$

$$p(t) = 0, \text{ for all } t \in]T, T+d], \mathbb{P}-\text{a.s.}, \ q(t) = 0 \text{ a.e. } t \in]T, T+d], \mathbb{P}-\text{a.s.}.$$

We extend the notion of mild solutions to the anticipating case.

Definition 2. *We say that equation (43) admits a mild solution if there is a couple of processes*

$$(p, q) \in L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H) \times L^2_{\mathcal{F}}(\Omega \times [0, T+d]; \mathcal{L}_2(K; H))$$

that solves the following equation, for all $t \in [0, T]$:

$$p(t) = S(T-t)' D_x h(X(T)) + \int_t^T S(s-t)' \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 D_x f(X(s), u(s-\theta))' p(s-\theta) \mu_f(d\theta) ds \tag{44}$$

$$+ \int_t^T S(s-t)' \int_{-d}^0 D_x l(X(s+\theta), u(s)) \mu_l(d\theta) ds - \int_t^T S(t-s)' q(s) dW(s)$$

And $p(t) = 0$ for all $t \in]T, T+d], \mathbb{P}-\text{a.s.}, \ q(t) = 0$ a.e. $t \in]T, T+d], \mathbb{P}-\text{a.s.}.$

Theorem 4. *Suppose that Hypothesis 1 holds, with $\mu_h = \delta_T$. Then the anticipated backward SPDE (44) admits a unique mild solution $(p, q) \in L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H) \times L^2_{\mathcal{F}}(\Omega \times [0, T+d]; \mathcal{L}_2(K; H))$.*
Moreover we have that, for β large enough:

$$\mathbb{E} \int_0^{T+d} \left(\frac{1}{2} |p(t)|_H^2 + |q(t)|_{\mathcal{L}_2(K;H)}^2 \right) e^{\beta t} dt \le C \left[1 + \mathbb{E} \sup_{t \in [-d, T]} |X(t)|_H^{2(j \vee k)} + \mathbb{E} \int_0^T |u(t)|_U^2\, dt \right] \tag{45}$$

for some constant C depending on β, T and the constants appearing in Hypothesis 1.

Proof. In order to prove the existence of a solution to (44) we want to construct a contraction map, as in [20] in the finite dimensional case, but with some differences due to the infinite dimensional setting. Given a pair of processes $(y, z) \in L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H) \times L^2_{\mathcal{F}}(\Omega \times [0, T+d]; \mathcal{L}_2(K; H))$, we define the contraction map $\Gamma : L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H) \times L^2_{\mathcal{F}}(\Omega \times [0, T+d]; \mathcal{L}_2(K; H)) \to L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H) \times L^2_{\mathcal{F}}(\Omega \times [0, T+d]; \mathcal{L}_2(K; H))$ using the mild formulation:

$$\begin{cases} p(t) = S(T-t)' D_x h(X(T)) + \int_t^T S(s-t)' \gamma(s, y_s) ds - \int_t^T S(t-s)' q(s) dW(s) \\ p(t) = 0 \qquad \forall t \in]T, T+d], \mathbb{P}-\text{a.s.} \\ q(t) = 0 \qquad \text{a.e. } t \in]T, T+d], \mathbb{P}-\text{a.s.} \end{cases}$$

where

$$\gamma(s, y_s) = \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 D_x f(X(s), u(s-\theta))' y(s-\theta) \mu_f(d\theta) + \int_{-d}^0 D_x l(X(s+\theta), u(s)) \mu_l(d\theta)$$

that belongs to $L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H)$, thanks to the assumptions on f and l and (26). We define the map Γ from $L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H) \times L^2_{\mathcal{F}}(\Omega \times [0, T+d]; \mathcal{L}_2(K; H))$ to itself, such that $\Gamma(y, z) = (p, q)$ and $\Gamma(y', z') = (p', q')$, where y, y' and z, z' are arbitrary elements of the space just defined. Denote their differences by

$$(\hat{y}, \hat{z}) = (y - y', z - z'), \qquad (\hat{p}, \hat{q}) = (p - p', q - q')$$

Using estimate in Lemma 4, according to the definition of p, p', q, q' we can say that

$$\mathbb{E} \int_0^{T+d} \left(|\hat{p}(t)|_H^2 + |\hat{q}(t)|_{\mathcal{L}_2(K;H)}^2 \right) e^{\beta t} dt \le \frac{2}{\beta} \mathbb{E} \int_0^T |\gamma(t, y_t) - \gamma(t, y_t')|^2 e^{\beta t} dt$$

The right hand side of the estimate can be rewritten using the special form of γ, then

$$\mathbb{E} \int_0^{T+d} \left(|\hat{p}(t)|_H^2 + |\hat{q}(t)|_{\mathcal{L}_2(K;H)}^2 \right) e^{\beta t} dt \leq \frac{2}{\beta} \mathbb{E} \int_0^T \left| \int_{-d}^0 D_x f(X(t), u(t-\theta))' \hat{y}(t-\theta) \mu_f(d\theta) \right|^2 e^{\beta t} dt$$

$$\leq \frac{2C_1}{\beta} \mathbb{E} \int_0^T \int_{-d}^0 |\hat{y}(t-\theta)|_H^2 |\mu_f|(d\theta) e^{\beta t} dt$$

After a change of variable $s = t - \theta$ we obtain

$$\mathbb{E} \int_0^T \int_{-d}^0 |\hat{y}(t-\theta)|_H^2 |\mu_f|(d\theta) e^{\beta t} dt \leq \mathbb{E} \int_0^d \int_{-s}^0 |\hat{y}(s)|_H^2 e^{\beta(s+\theta)} |\mu_f|(d\theta) ds$$

$$+ \mathbb{E} \int_d^T \int_{-d}^0 |\hat{y}(s)|_H^2 e^{\beta(s+\theta)} |\mu_f|(d\theta) ds \qquad (46)$$

$$+ \mathbb{E} \int_T^{T+d} \int_{-d}^{0 \wedge (T-s)} |\hat{y}(s)|_H^2 e^{\beta(s+\theta)} |\mu_f|(d\theta) ds.$$

So there exists a constant C, that depends on T, such that

$$\frac{2}{\beta} \mathbb{E} \int_0^T \int_{-d}^0 |\hat{y}(t-\theta)|_H^2 |\mu_f|(d\theta) e^{\beta t} dt \leq \frac{2C}{\beta} \mathbb{E} \int_0^{T+d} |\hat{y}(t)|_H^2 e^{\beta t} dt. \qquad (47)$$

We can choose β such that the map Γ is a strict contraction. By the Fixed Point Theorem we get existence and uniqueness of a mild solution. Let us come to the estimate. Let (p, q) be the fixed point solution, hence by (42) we have, for a β large enough:

$$\mathbb{E} \int_0^{T+d} \left(\frac{1}{2} |p(t)|_H^2 + |q(t)|_{\mathcal{L}_2(K;H)}^2 \right) e^{\beta t} dt$$

$$\leq C \left[1 + e^{\beta T} \mathbb{E} |D_x h(X(T))|_H^2 + \frac{2e^{\beta T}(T+d)}{\beta} \mathbb{E} \sup_{t \in [-d,T]} |X(t)|^{2j} + \frac{2e^{\beta T}}{\beta} \mathbb{E} \int_0^T |u(t)|^2 dt \right]$$

$$\leq C \left[1 + e^{\beta T} \left[C_4 + \frac{2(T+d)}{\beta} \right] \mathbb{E} \sup_{t \in [-d,T]} |X(t)|^{2(j \vee k)} + \frac{2e^{\beta T}}{\beta} \mathbb{E} \int_0^T |u(t)|^2 dt \right].$$

Notice that the constant C does not depend on β, since it comes from (42). □

In the following, we build an approximating ABSDE whose differential form still makes sense and that can approximate equation (41). Such approximating ABSDE is obtained by a suitable approximation of μ_h: the construction of this sequence of approximating measures $(\mu_h^n)_{n \geq 1}$ is given in the following Lemma.

Lemma 5. *Let $\bar{\mu}$ be a finite regular measure on $[T-d, T]$, such that $\bar{\mu}(\{T\}) = 0$. There exists a sequence $(\bar{\mu}^n)_{n \geq 1}$ of finite regular measure on $[T-d, T]$, absolutely continuous with respect to $\lambda_{[T-d,T]}$, the Lebesgue measure on $[T-d,T]$, such that*

$$\bar{\mu} = w^* - \lim_{n \to \infty} \bar{\mu}^n, \qquad (48)$$

that is for every $\phi \in C_b([T-d, T]; \mathbb{R})$

$$\int_{T-d}^T \phi \, d\bar{\mu} = \lim_{n \to \infty} \int_{T-d}^T \phi \, d\bar{\mu}^n \qquad (49)$$

Proof. We notice that, denoted by $\mathcal{M}$ the set of regular probability measures on a bounded and closed interval I, then $\text{Extr}\mathcal{M} = \{\delta_x : x \in I\}$. Moreover, by the Krein-Milman Theorem it turns out that $\mathcal{M} = \bar{co}\text{Extr}\mathcal{M} = \bar{co}\{\delta_x : x \in I\}$, and so any probability measure in $\mathcal{M}$ can be approximated by a linear convex combination of δ_x, $x \in I$.

Our aim is to approximate a finite regular measure $\bar{\mu}$ on the interval $I = [T-d, T]$. To this aim, there exist (at least) two positive, finite, regular measures $\bar{\mu}^+$ and $\bar{\mu}^-$ such that $\forall A \in \mathcal{B}([T-d, T])$,

$$\bar{\mu}(A) = \bar{\mu}^+(A) - \bar{\mu}^-(A).$$

Since $\bar{\mu}^+$ and $\bar{\mu}^-$ are positive, finite, regular measures on $[T-d, T]$, it is possible to associate to $\bar{\mu}^+$ and $\bar{\mu}^-$ two regular probability measures $\mathbb{P}^+$ and $\mathbb{P}^-$ such that

$$\mathbb{P}^+ = \frac{\bar{\mu}^+}{\bar{\mu}^+([T-d,T])}, \quad \mathbb{P}^- = \frac{\bar{\mu}^-}{\bar{\mu}^-([T-d,T])}.$$

By what pointed out at the beginning of the proof, each of the two regular probability measures $\mathbb{P}^+$ and $\mathbb{P}^-$ can be approximated by a linear convex combination of δ_x, $x \in I$, namely

$$\bar{\mu}^+ = w^* - \lim_{n\to\infty} \bar{\mu}^+([T-d,T]) \sum_{k=1}^n \alpha_k \delta_{x_k}, \; x_k \in I, \alpha_k \geq 0, \; \sum_{k=1}^n \alpha_k = 1;$$

$$\bar{\mu}^- = w^* - \lim_{n\to\infty} \bar{\mu}^-([T-d,T]) \sum_{k=1}^n \beta_k \delta_{y_k}, \; y_k \in I, \beta_k \geq 0, \; \sum_{k=1}^n \beta_k = 1;$$

and so

$$\bar{\mu} = w^* - \lim_{n\to\infty} \left[\bar{\mu}^+([T-d,T]) \sum_{k=1}^n \alpha_k \delta_{x_k} - \bar{\mu}^-([T-d,T]) \sum_{k=1}^n \beta_k \delta_{y_k} \right],$$

where

$$x_k, \, y_k \in I, \alpha_k \geq 0, \, \sum_{k=1}^n \alpha_k = 1, \, \beta_k \geq 0, \, \sum_{k=1}^n \beta_k = 1.$$

This gives an approximation of $\bar{\mu}$ by means of discrete measures. Now each δ measure can be approximated by means of a measure which is absolutely continuous with respect to the Lebesgue measure. Namely let us set

$$m_n = \min_{k=1,\ldots,n-1} (x_{k+1} - x_k),$$

and let us consider the measures $\lambda_{k,j}$, $j < \frac{m_n}{2}$, absolutely continuous with respect to λ, such that $\forall A \in \mathcal{B}(I)$

$$\lambda_{k,j}(A) = \int_A \frac{1}{j} \chi_{[x_k, x_k+j]}(x) \, dx.$$

It turns out that, for each k,

$$\delta_k = w^* - \lim_{j\to\infty} \lambda_{k,j}.$$

Moreover for every function $\xi \in C_b(I)$, by the mean value Theorem, with $x_{k^*} \in [x_k, x_k+j]$, $j < \frac{m_n}{2}$

$$\left| \sum_{k=1}^n \alpha_k \left[\int_I \xi(x) \, d\delta_{x_k}(x) - \int_I \xi(x) \lambda_{k,j}(x)(dx) \right] \right| = \left| \sum_{k=1}^n \alpha_k \left(\xi(x_k) - \xi(x_{k^*}) \right) \right|$$

$$\leq \sum_{k=1}^n \alpha_k \omega \left(|x_k - x_{k^*}| \right) \leq \omega \left(m_n \right) \sum_{k=1}^n \alpha_k = \omega \left(m_n \right) \to 0 \text{ as } n \to \infty,$$

where $\omega(\cdot)$ is the modulus of continuity of ξ. The desired approximation result holds true. $\qquad \square$

Remark 5. *Notice that the aim is the approximation of the measure μ_h in equation (41), in order to give sense to the differential form of such equation, and for this measure it can be $\mu_h(\{T\}) \neq 0$. To the measure μ_h we can associate another measure $\bar{\mu}_h$ such that for any $A \in \mathcal{B}([T-d,T])$*

$$\bar{\mu}_h(A) = \mu_h(A \setminus \{T\}). \tag{50}$$

Roughly speaking the measure $\bar{\mu}_h$ is obtained by the original measure μ_h, by subtracting to μ_h its mass in $\{T\}$. Lemma 5 ensures that there exists a sequence of measures $(\bar{\mu}_h^n)_{n\geq 1}$, on $[T-d,T]$, which are absolutely continuous with respect to the Lebesgue measure on $[T-d,T]$ and that are w^*-convergent to $\bar{\mu}_h$.

We can define the following regular approximations of equation (41) by approximating $\bar{\mu}_h$ obtained by μ_h in (50). We first introduce a preliminary equation

$$p^n(t) = \int_{t\vee(T-d)}^T S(s-t)' D_x h(X(s)) \bar{\mu}_h^n(ds)$$

$$+ \int_t^T S(s-t)' \Lambda(s) \, ds + \int_t^T S(s-t)' q^n(s) \, dW_s, \qquad t \in [0,T] \tag{51}$$

where $\Lambda \in L^2_{\mathcal{F}}(\Omega \times [0,T]; H)$. We can prove that:

Lemma 6. *Assume Hypothesis 1 and that $\Lambda \in L^2_{\mathcal{F}}(\Omega \times [0,T]; H)$. Then there exists a unique mild solution to (51). Moreover there exists a positive constant C, depending only on T and constants appearing in hypothesis 1, such that for every $\beta > 0$:*

$$\mathbb{E} \int_0^T \left(\frac{1}{2} |p^n(t)|^2_H + |q^n(t)|^2_{\mathcal{L}_2(K;H)} \right) e^{\beta t} dt \tag{52}$$

$$\leq C \Big[\mathbb{E} \int_0^T e^{\beta t} \Big| \int_{t \vee (T-d)}^T S(s-t)' D_x h(X(s)) \bar{\mu}^n_h(ds) \Big|^2_H dt$$

$$+ \frac{2}{\beta} \mathbb{E} \int_0^T |\Lambda(t)|^2_H e^{\beta t} dt + \mathbb{E} \int_0^T e^{\beta t} \Big| \int_{t \vee (T-d)}^T S(s-t)' K(s,t) \bar{\mu}^n_h(ds) \Big|^2_{\mathcal{L}_2(K;H)} dt \Big]$$

where $K \in L^2_{\mathcal{F}}(\Omega \times [0,T] \times [0,T]; \mathcal{L}_2(K;H))$ is such that for every $s \geq t \geq 0$:

$$\mathbb{E}^{\mathcal{F}_t} D_x h(X(s)) = D_x h(X(s)) - \int_t^s K(s,\theta) \, dW_\theta, \quad d\mathbb{P} \times d\theta \text{ a.s.} \tag{53}$$

Proof. Let us notice first of all that $\bar{\mu}^n_h(dt) = \bar{\zeta}^n_h(t) \, dt$, with $\bar{\zeta}^n_h \in L^\infty([T-d,T])$, thus the term $D_x h(X(s)) \bar{\zeta}^n_h(s)$ can be treated as a "normal" forcing term and lemma 4 can be applied to get existence and uniqueness of (51), for every n.

We concentrate on the estimate (52), we consider $t > T - d$, the most critical case:

$$p^n(t) = \mathbb{E}^{\mathcal{F}_t} p^n(t) = \mathbb{E}^{\mathcal{F}_t} \int_t^T S(s-t)' D_x h(X(s)) \bar{\mu}^n_h(ds) + \mathbb{E}^{\mathcal{F}_t} \int_t^T S(s-t)' \Lambda(s) ds \tag{54}$$

by standard estimates we get:

$$\mathbb{E} \int_0^T \frac{1}{2} |p^n(t)|^2_H e^{\beta t} \, dt \leq \int_0^T e^{\beta t} \mathbb{E} \Big| \int_t^T S(s-t)' D_x h(X(s)) \bar{\mu}^n_h(ds) \Big|^2_H dt$$

$$+ \mathbb{E} \int_0^T e^{\beta t} \Big(\int_t^T e^{-\beta s} \, ds \int_t^T e^{\beta s} |S(s-t)' \Lambda(s)|^2_H \, ds \Big) dt$$

$$\leq \int_0^T e^{\beta t} \mathbb{E} \Big| \int_t^T S(s-t)' D_x h(X(s)) \bar{\mu}^n_h(ds) \Big|^2_H dt + \frac{M}{\beta} \int_0^T e^{\beta t} |\Lambda(t)|^2_H \, dt$$

where the constant M depends on T and the constants appearing in Hypothesis 1.

Let us come to the martingale term q^n. Following [14], we get that:

$$q^n(t) = \int_t^T S(s-t)' K(s,t) \bar{\mu}^n_h(ds) + \int_t^T S(s-t)' \tilde{K}(s,t) \, ds, \quad d\mathbb{P} \times dt \text{ a.s.} \tag{55}$$

where K, $\tilde{K}$ are given by the Martingale Representation Theorem respectively in (53) and

$$\mathbb{E}^{\mathcal{F}_t} \Lambda(s) = \Lambda(s) - \int_t^s \tilde{K}(s,\theta) \, dW_\theta, \quad d\mathbb{P} \times d\theta \text{ a.s.} \tag{56}$$

Thus, see also [12], section 4, equation (4.14), we get that

$$\mathbb{E} \int_0^T e^{\beta t} |q^n(t)|^2_{\mathcal{L}_2(K;H)} \, dt \leq 2\mathbb{E} \Big[\int_0^T e^{\beta t} \Big| \int_t^T S(s-t)' K(s,t) \bar{\mu}^n_h(ds) \Big|^2_{\mathcal{L}_2(K;H)} dt + \frac{C}{\beta} \int_0^T e^{\beta t} |\Lambda(t)|^2_H \, dt \Big]$$

where C depends only on T and the constants appearing in Hypothesis 1. The case when $0 \leq t \leq T - d$ can be treated in the same way. $\square$

Theorem 5. *Let Hypothesis 1 holds true, let $\bar{\mu}_h$ be defined by (50), and let us consider $(\bar{\mu}^n_h)_{n \geq 0}$ the w^*-approximations of $\bar{\mu}_h$, absolutely continuous with respect to the Lebesgue measure on $[T-d,T]$.*

Let us consider the approximating ABSDEs (of "standard" type):

$$
\begin{aligned}
p^n(t) = {}& \int_{t\vee(T-d)}^T S(s-t)'D_xh(X(s))\bar{\mu}_h^n(ds) \\
& + \int_t^T S(s-t)'\mathbb{E}^{\mathcal{F}_s}\int_{-d}^0 D_xf(X(s),u(s-\theta))'p^n(s-\theta)\mu_f(d\theta)\,ds \\
& + \int_t^T S(s-t)'\mathbb{E}^{\mathcal{F}_s}\int_{-d}^0 D_xl(X(s+\theta),u(s))\mu_l(d\theta)ds - \int_t^T S(t-s)'q^n(s)\,dW(s) \\
& + S(T-t)'D_xh(X(T))\mu_h(\{T\}), \qquad t\in[0,T] \\
p^n(t) = {}& 0 \qquad \forall t\in]T,T+d], \mathbb{P}-a.s. \quad q^n(t)=0 \qquad a.e.\ t\in]T,T+d], \mathbb{P}-a.s.
\end{aligned}
\tag{57}
$$

Then the pair (p^n,q^n), solution to (57), converges in $L^2_{\mathcal{F}}(\Omega\times[0,T+d];H)\times L^2_{\mathcal{F}}(\Omega\times[0,T+d];\mathcal{L}_2(K;H))$ to the pair (p,q), mild solution to (41). Moreover such solution is unique in $L^2_{\mathcal{F}}(\Omega\times[0,T+d];H)\times L^2_{\mathcal{F}}(\Omega\times[0,T+d];\mathcal{L}_2(K;H))$.

Proof. Let us first prove that the sequence $(p^n,q^n)_n$ is a Cauchy sequence in $L^2_{\mathcal{F}}(\Omega\times[0,T+d];H)\times L^2_{\mathcal{F}}(\Omega\times[0,T+d];\mathcal{L}_2(K,H))$. The equation satisfied by $(p_t^n-p_t^k,q_t^n-q_t^k)$, $n,k\geq 1$, turns out to be an ABSDE:

$$
\begin{aligned}
(p^n-p^k)(t) = {}& \int_{t\vee(T-d)}^T S(s-t)'D_xh(X(s))\bar{\mu}_h^n(ds) - \int_{t\vee(T-d)}^T S(s-t)'D_xh(X(s))\bar{\mu}_h^k(ds) \\
& + \int_t^T S(s-t)'\mathbb{E}^{\mathcal{F}_s}\int_{-d}^0 D_xf(X(s),u(s-\theta))'(p^n-p^k)(s-\theta)\mu_f(d\theta)\,ds \\
& + \int_t^T S(s-t)'(q^n-q^k)(s)\,dW(s)
\end{aligned}
\tag{58}
$$

Now we use estimate (52) with $\Lambda(s)=\mathbb{E}^{\mathcal{F}_s}\int_{-d}^0 D_xf(X(s),u(s-\theta))'(p^n-p^k)(s-\theta)\mu_f(d\theta)$ to get

$$
\mathbb{E}\int_0^{T+d}\left(\frac{1}{2}|(p^n-p^k)(t)|_H^2+|(q^n-q^k)(t)|_{\mathcal{L}_2(K;H)}^2\right)e^{\beta t}dt
\tag{59}
$$

$$
\begin{aligned}
\leq {}& C\Big[\int_0^T e^{\beta t}\mathbb{E}\Big|\int_{t\vee(T-d)}^T S(s-t)'D_xh(X(s))(\bar{\mu}_h^n-\bar{\mu}_h^k)(ds)\Big|_H^2 dt \\
& + \mathbb{E}\int_0^T e^{\beta t}\Big|\int_{t\vee(T-d)}^T S(s-t)'K(s,t)(\bar{\mu}_h^n-\bar{\mu}_h^k)(ds)\Big|_{\mathcal{L}_2(K;H)}^2 dt \\
& + \frac{2}{\beta}\mathbb{E}\int_0^{T+d}|(p^n-p^k)(t)|_H^2 e^{\beta t}dt\Big]
\end{aligned}
$$

Hence, choosing β such that $\frac{2}{\beta}\leq\frac{1}{4}$ we get that

$$
\mathbb{E}\int_0^{T+d}\left(\frac{1}{4}|(p^n-p^k)(t)|_H^2+|(q^n-q^k)(t)|_{\mathcal{L}_2(K;H)}^2\right)e^{\beta t}dt
\tag{60}
$$

$$
\begin{aligned}
\leq {}& C\Big[\mathbb{E}\int_0^T e^{\beta t}\Big|\int_{t\vee(T-d)}^T S(s-t)'D_xh(X(s))(\bar{\mu}_h^n-\bar{\mu}_h^k)(ds)\Big|_H^2 dt \\
& + \mathbb{E}\int_0^T e^{\beta t}\Big|\int_{t\vee(T-d)}^T S(s-t)'K(s,t)(\bar{\mu}_h^n-\bar{\mu}_h^k)(ds)\Big|_{\mathcal{L}_2(K;H)}^2 dt\Big]
\end{aligned}
$$

In order to prove that the sequence (p^n,q^n) is a Cauchy sequence, we need to exploit the w^*-convergence of the sequence of measures $\bar{\mu}_h^n$ as $n\to\infty$.

Let us consider the first term, since $s\to S(s-t)'D_xh(X(s))$ is continuous $\mathbb{P}$-a.s. and uniformly bounded we have, by the Dominated Convergence Theorem:

$$
\lim_{n,k\to+\infty}\mathbb{E}\int_0^T e^{\beta t}\Big|\int_{t\vee(T-d)}^T S(s-t)'D_xh(X(s))(\bar{\mu}_h^n-\bar{\mu}_h^k)(ds)\Big|_H^2 dt=0.
\tag{61}
$$

The second term requires some work, since $s \to S(s-t)'K(s,t)$ is not continuous. First of all we notice that by (53) we get, by Hypothesis 1, point (H6),

$$\mathbb{E} \int_0^s |K(s,t)|^2_{\mathcal{L}_2(K;H)} \, dt \leq 4\mathbb{E}|D_x h(X(s))|^2 \leq 4C.$$

Hence, we get

$$\mathbb{E} \int_{T-d}^T \int_0^s |K(s,t)|^2_{\mathcal{L}_2(K;H)} \, dt \, d|\bar{\mu}_h|(ds) \leq 4C|\bar{\mu}_h|([T-d,T]). \tag{62}$$

where $\bar{\mu}_h$ is the weak limit of the measures $\bar{\mu}_h^n$, and $|\bar{\mu}_h|$ is the positive measure given by $\bar{\mu}_h^+ + \bar{\mu}_h^-$.

Now we introduce some processes $K^\varepsilon(s,t)$ such that $s \to K^\varepsilon(s,t)$ has a continuous version for all $t \in [0,T]$, and that can be chosen so that:

$$|K^\varepsilon(s,t)| \leq |K(s,t)|, \quad \text{a.e. } s,t \in [T-d,T] \times [0,T] \text{ and } \mathbb{P}-a.s., \tag{63a}$$

$$\lim_{\varepsilon \to 0} \mathbb{E} \int_0^T \int_{T-d}^T |K^\varepsilon(s,t) - K(s,t)|^2 |\bar{\mu}_h|(ds) \, dt = 0. \tag{63b}$$

Such properties are very well known for $H = \mathbb{R}$, see for instance [21], and can be easily extended to a separable Hilbert space going through a basis expansion.

We can finally conclude noticing that:

$$\mathbb{E} \int_0^T e^{\beta t} \Big| \int_{t \vee (T-d)}^T S(s-t)'K(s,t)(\bar{\mu}_h^n - \bar{\mu}_h^k)(ds) \Big|^2_{\mathcal{L}_2(K;H)} \, dt$$

$$\leq 2\Big[\mathbb{E} \int_0^T e^{\beta t} \Big| \int_{t \vee (T-d)}^T S(s-t)'K^\varepsilon(s,t)(\bar{\mu}_h^n - \bar{\mu}_h^k)(ds) \Big|^2_{\mathcal{L}_2(K;H)} \, dt$$

$$+ \mathbb{E} \int_0^T e^{\beta t} \Big| \int_{t \vee (T-d)}^T S(s-t)'(K(s,t) - K^\varepsilon(s,t))(\bar{\mu}_h^n - \bar{\mu}_h^k)(ds) \Big|^2_{\mathcal{L}_2(K;H)} \, dt\Big]$$

$$\leq C\Big[\mathbb{E} \int_0^T e^{\beta t} \Big| \int_{t \vee (T-d)}^T S(s-t)'K^\varepsilon(s,t)(\bar{\mu}_h^n - \bar{\mu}_h^k)(ds) \Big|^2_{\mathcal{L}_2(K;H)} \, dt$$

$$+ \mathbb{E} \int_0^T \int_{T-d}^T \Big| K(s,t) - K^\varepsilon(s,t) \Big|^2_{\mathcal{L}_2(K;H)} |\bar{\mu}_h|(ds) \, dt\Big].$$

We let first n and k go to infinity to let the first term go to zero. Then, using (63a) and the fact that by the Banach-Steinhaus Theorem

$$|\bar{\mu}_h^n| \leq |\bar{\mu}_h|,$$

we let ε go to zero and we get the convergence to 0 of the whole expression. So (p_n, q_n) is a Cauchy sequence in $L^2_{\mathcal{F}}(\Omega \times [0, T+d]; H) \times L^2_{\mathcal{F}}(\Omega \times [0, T+d]; \mathcal{L}_2(K,H))$, thus there exists a couple $(\bar{p}, \bar{q})$ such that:

$$\mathbb{E} \int_0^{T+d} |(p^n - \bar{p})(s)|^2_H \, ds + \mathbb{E} \int_0^{T+d} |(q^n - \bar{q})(s)|^2_{\mathcal{L}_2(K;H)} \, ds \to 0 \qquad \text{as } n \to \infty. \tag{64}$$

It remains to show the convergence inside the mild formulation of the equation to get that the limit $(\bar{p}, \bar{q})$ fulfils equation (41). Writing again the equation satisfied by the difference $p^n - p^k$, and recalling the measurability property typical of the solutions of backward equations, we get:

$$\mathbb{E}^{\mathcal{F}_t}(p^n - p^k)(t) = \mathbb{E}^{\mathcal{F}_t} \int_{t \vee (T-d)}^T S(s-t)' D_x h(X(s))(\bar{\mu}_h^n - \bar{\mu}_h^k)(ds)$$

$$+ \mathbb{E}^{\mathcal{F}_t} \int_t^T S(s-t)' \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 D_x f(X(s), u(s-\theta))'(p^n - p^k)(s-\theta)\mu_f(d\theta) \, ds \tag{65}$$

Similarly to (61), by the Dominated Convergence Theorem, we have

$$\mathbb{E}\Big| \mathbb{E}^{\mathcal{F}_t} \int_{t \vee (T-d)}^T S(s-t)' D_x h(X(s))(\bar{\mu}_h^n - \bar{\mu}_h^k)(ds) \Big|^2 \to 0 \quad \text{as } n, k \to +\infty. \tag{66}$$

Moreover we get that for some constant M, which depends only on T and the quantities appearing in Hypothesis 1, the following holds

$$\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_t}\int_t^T \mathbb{E}^{\mathcal{F}_s} S(s-t)'\int_{-d}^0 D_x f(X(s), u(s-\theta))'(p^n - p^k)(s-\theta)\mu_f(d\theta)ds\right|^2$$

$$\leq M|\mu_f|^2([-d,0])\mathbb{E}\int_0^{T+d} |(p^n - p^k)(r)|^2 dr \to 0 \quad \text{as } n, k \to +\infty.$$

Thus $\mathbb{E}|p^n(t) - p^k(t)|^2 \to 0$ as $k, n \to \infty$ for every $t \in [0,T]$, and there exists a $p(t)$ such that $\mathbb{E}|p^n(t) - p(t)|^2 \to 0$ as $n \to \infty$ for every $t \in [0,T]$. It can be seen, again by the dominated convergence Theorem, that also $\int_0^{T+d} \mathbb{E}|p^n(t) - p(t)|^2 \, dt \to 0$ as $n \to \infty$ so such limit must coincide with $\bar{p}(t)$ at least $\mathbb{P}$ a.s. for a.e. $t \in [T-d, T]$. We eventually get that $(\bar{p}, \bar{q})$, choosing if necessary p instead of $\bar{p}$, is a mild solution to equation (41) since every term in (57) must converge to its expected limit.

Uniqueness is not a problem since as soon as we calculate the difference of two mild solutions, the data disappear and we can treat the equation as in lemma 4.

$\square$

5. Functional SMP: Necessary form

Now we are able to prove a version of the SMP, in its necessary form, for the control problem with state equation and cost functional given by (8) and (10), respectively.
Let $(\bar{X}, \bar{u})$ be an optimal pair. We start this section by recovering the form of the derivative of the cost functional.

Lemma 7. *The cost functional $J(\cdot)$ is Gateaux differentiable and the derivative has the following form*

$$\frac{d}{d\rho} J(\bar{u}(\cdot) + \rho v(\cdot))|_{\rho=0} = \mathbb{E}\int_0^T \int_{-d}^0 D_x l(\bar{X}(t+\theta), \bar{u}(t)) Y(t+\theta)\mu_l(d\theta)dt$$

$$+ \mathbb{E}\int_0^T \int_{-d}^0 D_u l(\bar{X}(t+\theta), \bar{u}(t))v(t)\mu_l(d\theta)dt + \mathbb{E}\int_{T-d}^T D_x h\left(\bar{X}(\theta)\right) Y(\theta)\mu_h(d\theta)$$

Proof. Let $(\bar{X}, \bar{u})$ be an optimal pair and let w be another admissible control, set $v = w - \bar{u}$ and $u^\rho = \bar{u} + \rho w$. We can write the variation of the cost functional in the form

$$0 \leq \frac{J(u^\rho(\cdot)) - J(\bar{u}(\cdot))}{\rho}$$

$$= \mathbb{E}\frac{1}{\rho}\int_0^T \int_{-d}^0 \left[l(X^\rho(t+\theta), u^\rho(t)) - l(\bar{X}(t+\theta), \bar{u}(t))\right]\mu_l(d\theta)dt$$

$$+ \mathbb{E}\frac{1}{\rho}\int_{T-d}^T \left[h(X^\rho(\theta)) - h(\bar{X}(\theta))\right]\mu_h(d\theta) = I_1 + I_2.$$

Now

$$I_1 = \mathbb{E}\frac{1}{\rho}\int_0^T \int_{-d}^0 \left[l(X^\rho(t+\theta), u^\rho(t)) - l(\bar{X}(t+\theta), u^\rho(t))\right]\mu_l(d\theta)dt$$

$$+ \mathbb{E}\frac{1}{\rho}\int_0^T \int_{-d}^0 \left[l(\bar{X}(t+\theta), u^\rho(t)) - l(\bar{X}(t+\theta), \bar{u}(t))\right]\mu_l(d\theta)dt$$

$$= \mathbb{E}\int_0^T \int_{-d}^0 \int_0^1 \langle D_x l(\bar{X}(t+\theta) + \lambda(X^\rho(t+\theta) - \bar{X}(t+\theta)), u^\rho(t)), (\tilde{X}^\rho(t+\theta) + Y(t+\theta))\rangle d\lambda\mu_l(d\theta)dt$$

$$+ \mathbb{E}\int_0^T \int_{-d}^0 \int_0^1 \langle D_u l(\bar{X}(t+\theta), \bar{u}(t) + \lambda\rho v(t)), v(t)\rangle d\lambda\mu_l(d\theta)dt;$$

and

$$I_2 = \mathbb{E}\frac{1}{\rho}\int_{T-d}^{T}\left[h(X^{\rho}(\theta)) - h(\bar{X}(\theta))\right]\mu_h(d\theta)$$
$$= \mathbb{E}\int_{T-d}^{T}\int_{0}^{1}\langle D_x h\left(\bar{X}(\theta) + \lambda(X^{\rho}(\theta) - \bar{X}(\theta))\right), (\tilde{X}^{\rho}(\theta) + Y(\theta))\rangle d\lambda\mu_h(d\theta).$$

From (37) we know that

$$\lim_{\rho\to 0}\mathbb{E}\sup_{t\in[0,T]}|\tilde{X}^{\rho}(t)|_H^2 = 0$$

so that, sending ρ to 0, we obtain the required equivalence

$$0 \le \frac{d}{d\rho}J(\bar{u}(\cdot) + \rho v(\cdot))|_{\rho=0} = \mathbb{E}\int_{0}^{T}\int_{-d}^{0}\langle D_x l(\bar{X}(t+\theta), \bar{u}(t)), Y(t+\theta)\rangle\mu_l(d\theta)dt$$
$$+ \mathbb{E}\int_{0}^{T}\int_{-d}^{0}\langle D_u l(\bar{X}(t+\theta), \bar{u}(t)), v(t)\rangle\mu_l(d\theta)dt + \mathbb{E}\int_{T-d}^{T}\langle D_x h\left(\bar{X}(\theta)\right), Y(\theta)\rangle\mu_h(d\theta).$$

$\square$

Notice that without the convexity assumption on U_c, the derivative of the cost function does not make sense.

Now define the Hamiltonian associated to the system by setting $\mathcal{H} : [0,T] \times C([-d,0];H) \times U_c \times H \to \mathbb{R}$ as

$$\mathcal{H}(t,x,u,p) = \langle\int_{-d}^{0}f(x(\theta),u)\mu_f(d\theta), p\rangle_H + \int_{-d}^{0}l(x(\theta),u)\mu_l(d\theta) \tag{67}$$
$$= \langle F(x,u), p\rangle_H + L(t,x,u).$$

where in the last equality we adopted the formalism introduced in Remark 4.

Let us state the stochastic maximum principle, where for the Definition of $\bar{X}_t$ we refer to (9).

Theorem 6. *Let Hypothesis 1 be satisfied and suppose that $(\bar{X}, \bar{u})$ is an optimal pair for the control problem. Then there exist a pair of processes $(p,q) \in L_{\mathcal{F}}^2(\Omega \times [0,T+d];H) \times L_{\mathcal{F}}^2(\Omega \times [0,T+d];\mathcal{L}_2(K;H))$ which are the solution to the ABSDE (41) such that the following variational inequality holds*

$$\langle\frac{\partial}{\partial u}\mathcal{H}(t,\bar{X}_t,\bar{u}(t),p(t),q(t)), w - \bar{u}(t)\rangle \ge 0$$

for all $w \in U_c$, $\mathbb{P} \times dt$- a.e., where $\mathcal{H}$ is the Hamiltonian function defined in (67).

Proof. Firstly we study the duality between the first variation process $Y(\cdot)$ and the adjoint process $p(\cdot)$. Then we rewrite the duality using the variation of the cost functional.

As a first step we are going to prove the following duality formula

$$\mathbb{E}\langle Y(T), D_x h(X(T))\mu_h(\{T\})\rangle + \mathbb{E}\int_{0}^{T}\int_{-d}^{0}\langle D_x l(X(t+\theta), u(t)), Y(t)\rangle\mu_l(d\theta)dt$$
$$= \mathbb{E}\int_{0}^{T}\int_{-d}^{0}\langle D_u f(X(t+\theta), u(t))'p(t), v(t)\rangle\mu_f(d\theta)dt - \mathbb{E}\int_{T-d}^{T}\langle Y(\theta), D_x h(X(\theta))\rangle\bar{\mu}_h(d\theta) \tag{68}$$

which is the crucial relation from which we can formulate the SMP. Let us recall that the dual equation (41) associated to the system can not be written in differential form. Then, in order to obtain the variational inequality, we have to approximate the measure μ_h by means of the sequence $\bar{\mu}_h^n$, once we have subtracted its mass in $\{T\}$. What we obtain is equation (57), which we rewrite

below for the reader's convenience

$$p^n(t) = \int_{t \vee (T-d)}^T S\,(s-t)'\,D_x h(X(s))\bar{\mu}_h^n(ds)$$

$$+ \int_t^T S(s-t)'\,\mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 D_x f(X(s), u(s-\theta))'p^n(s-\theta)\mu_f(d\theta)\,ds$$

$$+ \int_t^T S(s-t)'\,\mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 D_x l(s, X(s+\theta), u(s))\mu_l(d\theta)ds - \int_t^T S(t-s)'q^n(s)\,dW(s) \qquad (69)$$

$$+ S(T-t)'D_x h(X(T))\mu_h(\{T\}), \qquad t \in [0,T]$$

$$p^n(t) = 0, \qquad \text{for all } t \in]T, T+d], \mathbb{P} - \text{a.s.}, \qquad q^n(t) = 0 \quad \text{a.e. } t \in]T, T+d], \mathbb{P} - \text{a.s.}.$$

For all $k \in \mathbb{N}$, let $J_k := k(k-A)^{-1}$, the bounded operators such that $\lim_{k \to +\infty} |J_k x - x|_H = 0$, for all $x \in H$. We set $p^{n,k}(t) := J_k' p^n(t)$ and $q^{n,k}(t) := J_k' q^n(t)$, hence by Dominated Convergence Theorem it can be proved:

$$\mathbb{E} \sup_{t \in [0,T+d]} |p^{n,k}(t) - p^n(t)|_H^2 \to 0, \qquad \text{as } k \to 0; \qquad (70)$$

$$\mathbb{E} \int_0^{T+d} |q^{n,k}(t) - q^n(t)|_{\mathcal{L}(K;H)}^2\,dt \to 0, \qquad \text{as } k \to 0.$$

Notice that the processes $p^{n,k}$ admit an Itô differential:

$$\begin{cases} -dp^{n,k}(t) = \left[A'p^{n,k}(t) + J_k' \int_{-d}^0 D_x f(X(t), u(t-\theta))'p^n(t-\theta)\mu_f(d\theta) \right. \\[2ex] \qquad\qquad \left. + J_k' \int_{-d}^0 D_x l(t, X_t(\theta), u(t))\mu_l(d\theta) \right]dt + J_k' D_x h(X(t))\bar{\zeta}_h^n(t)dt - q^{n,k}(t)dW(t), \\[2ex] p^{n,k}(T) = J_k' D_x h(X(T))\mu_h(\{T\}), \end{cases} \qquad (71)$$

where we denoted by $\bar{\zeta}_h^n(\cdot)$ the Radon-Nykodim derivative of $\bar{\mu}_h^n$, with respect to the Lebesgue measure on $[T-d, T]$.

Let us now consider the k-approximation of the first variation process $Y^k(t) =: J_k Y(t)$, see (33). The equation for Y^k is

$$\begin{cases} \dfrac{dY^k}{dt}(t) = AY^k(t) + J_k \int_{-d}^0 D_x f(Xt+(\theta), u(t))Y(t+\theta)\mu_f(d\theta) \\[2ex] \qquad\qquad + J_k \int_{-d}^0 D_u f(X_t(\theta), u(t))\mu_f(d\theta)v(t), \\[2ex] Y^k(0) = 0. \end{cases} \qquad (72)$$

One can prove, evaluating the difference between Y and Y^k and using standard arguments, that:

$$\mathbb{E} \sup_{t \in [-d,T]} |Y^k(t) - Y(t)|_H^2 \to 0, \qquad \text{as } k \to 0 \qquad (73)$$

Computing the Itô formula for the product $d\langle p^{n,k}(t), Y^k(t)\rangle$, integrating in time and taking expectation we end up with

$$\mathbb{E}\langle Y^k(T), p^{n,k}(T)\rangle_H = \mathbb{E} \int_0^T J_k \langle \int_{-d}^0 D_x f(X(t+\theta), u(t))Y(t+\theta)\mu_f(d\theta), p^{n,k}(t)\rangle_H dt$$

$$+ \mathbb{E} \int_0^T J_k \langle \int_{-d}^0 D_u f(X(t+\theta), u(t))\mu_f(d\theta)v(t), p^{n,k}(t)\rangle_H dt$$

$$- \mathbb{E} \int_0^T \langle Y^k(t), J_k' \int_{-d}^0 D_x l(X(t+\theta), u(t))\mu_l(d\theta)\rangle_H dt \qquad (74)$$

$$- \mathbb{E} \int_0^T \langle Y^k(t), J_k' \int_{-d}^0 D_x f(X(t), u(t-\theta))'p^n(t-\theta)\mu_f(d\theta)\rangle_H dt$$

$$- \mathbb{E} \int_{T-d}^T \langle Y^k(\theta), J_k' D_x h(X(\theta))\bar{\zeta}_h^n(\theta)\rangle d\theta$$

Now we let k tend to ∞ to get the following duality relation

$$\mathbb{E}\langle Y(T), D_x h(X(T))\mu_h(\{T\})\rangle + \mathbb{E}\int_0^T \int_{-d}^0 \langle D_x l(X(+\theta), u(t)), Y(t)\rangle \mu_l(d\theta)dt$$

$$= \mathbb{E}\int_0^T \int_{-d}^0 \langle D_u f(X(t+\theta), u(t))'p^n(t), v(t)\rangle \mu_f(d\theta)dt - \mathbb{E}\int_{T-d}^T \langle Y(\theta), D_x h(X(\theta))\rangle \bar{\mu}_h^n(d\theta) \tag{75}$$

In order to obtain the required duality formula (68) we have to pass to the limit as $n \to \infty$ in (75). This can be done using the result on convergence of measures stated in Lemma 5 and the convergence results proved in Proposition 5.

We thus end up with (68), and recalling that:

$$\mu_h = \bar{\mu}_h + \mu_h(\{T\})\delta_T$$

we can write (68) as:

$$\mathbb{E}\int_{T-d}^T \langle Y(\theta), D_x h(X(\theta))\rangle \mu_h(d\theta) + \mathbb{E}\int_0^T \int_{-d}^0 \langle D_x l(X(t+\theta), u(t)), Y(t)\rangle \mu_l(d\theta)dt$$

$$= \mathbb{E}\int_0^T \int_{-d}^0 \langle D_u f(X(t+\theta), u(t))'p(t), v(t)\rangle \mu_f(d\theta)dt. \tag{76}$$

On the other side, from lemma 7, we know that

$$0 \le \frac{d}{d\rho} J(\bar{u}(\cdot) + \rho v(\cdot))|_{\rho=0} = \mathbb{E}\int_0^T \int_{-d}^0 \langle D_x l(t, \bar{X}(+\theta)), \bar{u}_t), Y_t(\theta)\rangle \mu_l(d\theta)dt$$

$$+ \mathbb{E}\int_0^T \int_{-d}^0 \langle D_u l(t, \bar{X}(t+\theta), \bar{u}_t), v_t\rangle \mu_l(d\theta)dt + \mathbb{E}\int_{T-d}^T \langle D_x h\left(\bar{X}(\theta)\right), Y(\theta)\rangle \mu_h(d\theta) \tag{77}$$

So substituting (76) in (77) we eventually get

$$0 \le \mathbb{E}\int_0^T \int_{-d}^0 \langle D_u f(\bar{X}(t+\theta), \bar{u}(t))'p(t), v(t)\rangle \mu_f(d\theta)dt + \mathbb{E}\int_0^T \int_{-d}^0 \langle D_u l(t, \bar{X}(t+\theta), \bar{u}_t), v_t\rangle \mu_l(d\theta)dt$$

$$= \mathbb{E}\int_0^T \langle \frac{\partial}{\partial u} H(t, \bar{X}_t, \bar{u}(t), p(t), q(t)), v(t)\rangle dt \tag{78}$$

from which the required result holds true. $\square$

REFERENCES

[1] Li Chen and Zhen Wu. Maximum principle for the stochastic optimal control problem with delay and application. *Automatica J. IFAC*, 46(6):1074–1080, 2010.

[2] Li Chen, Zhen Wu, and Zhiyong Yu. Delayed stochastic linear-quadratic control problem and related applications. *Journal of Applied Mathematics*, 46:22 pp. doi:10.1155/2012/835319, 2012.

[3] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[4] Kai Du and Qingxin Meng. A maximum principle for optimal control of stochastic evolution equations. *SIAM J. Control Optim.*, 51(6):4343–4362, 2013.

[5] Yuli Eidelman, Vitali Milman, and Antonis Tsolomitis. *Functional analysis*, volume 66 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2004. An introduction.

[6] Marco Fuhrman, Ying Hu, and Gianmario Tessitore. Stochastic maximum principle for optimal control of SPDEs. *C. R. Math. Acad. Sci. Paris*, 350(13-14):683–688, 2012.

[7] Marco Fuhrman, Ying Hu, and Gianmario Tessitore. Stochastic maximum principle for optimal control of SPDEs. *Appl. Math. Optim.*, 68(2):181–217, 2013.

[8] Marco Fuhrman, Ying Hu, and Gianmario Tessitore. Stochastic maximum principle for optimal control of spdes driven by white noise. *arXiv preprint arXiv:1409.4746*, 2014.

[9] Marco Fuhrman, Federica Masiero, and Gianmario 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.

[10] Marco Fuhrman and Carlo Orrieri. Stochastic Maximum Principle for Optimal Control of a Class of Nonlinear SPDEs with Dissipative Drift. *SIAM J. Control Optim.*, 54(1):341–371, 2016.

[11] Marco Fuhrman and Gianmario 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.

[12] Giuseppina Guatteri and Gianmario Tessitore. On the backward stochastic Riccati equation in infinite dimensions. *SIAM J. Control Optim.*, 44(1):159–194 (electronic), 2005.

[13] Ying Hu and Shi Ge Peng. Maximum principle for semilinear stochastic evolution control systems. *Stochastics Stochastics Rep.*, 33(3-4):159–180, 1990.

[14] Ying Hu and Shi Ge Peng. Adapted solution of a backward semilinear stochastic evolution equation. *Stochastic Anal. Appl.*, 9(4):445–459, 1991.

[15] Ying Hu and Shi Ge Peng. Maximum principle for optimal control of stochastic system of functional type. *Stochastic Anal. Appl.*, 14(3):283–301, 1996.

[16] Qingxin Meng and Yang Shen. Optimal control for stochastic delay evolution equations. *accepted on Appl. Math. Optim.*, (10.1007/s00245-015-9308-2).

[17] Salah-Eldin A. Mohammed. Stochastic differential systems with memory: theory, examples and applications. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, volume 42 of *Progr. Probab.*, pages 1–77. Birkhäuser Boston, Boston, MA, 1998.

[18] Bernt Øksendal, Agnès Sulem, and Tusheng Zhang. Optimal control of stochastic delay equations and time-advanced backward stochastic differential equations. *Adv. in Appl. Probab.*, 43(2):572–596, 2011.

[19] Bernt Øksendal, Agnès Sulem, and Tusheng Zhang. Optimal partial information control of SPDEs with delay and time-advanced backward SPDEs. In *Stochastic analysis and applications to finance*, volume 13 of *Interdiscip. Math. Sci.*, pages 355–383. World Sci. Publ., Hackensack, NJ, 2012.

[20] Shige Peng and Zhe Yang. Anticipated backward stochastic differential equations. *Ann. Probab.*, 37(3):877–902, 2009.

[21] Walter Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, third edition, 1987.

[22] J. Yong and X. Y. Zhou. *Stochastic Controls: Hamiltonian Systems and emphHJB Equations.* New York: Springer-Verlag, 1999.

[23] Jianjun Zhou and Bin Liu. The existence and uniqueness of the solution for nonlinear Kolmogorov equations. *J. Differential Equations*, 253(11):2873–2915, 2012.

(G. Guatteri) Dipartimento di Matematica, Politecnico di Milano. via Bonardi 9, 20133 Milano, Italia
E-mail address: `giuseppina.guatteri@polimi.it`

(F. Masiero) Dipartimento di Matematica e Applicazioni, Università di Milano-Bicocca. via Cozzi 55, 20125 Milano, Italia
E-mail address: `federica.masiero@unimib.it`

(C. Orrieri) Dipartimento di Matematica, Università di Pavia. via Ferrata 1, 27100 Pavia, Italia
E-mail address: `carlo.orrieri01@ateneopv.it`