

STOCHASTIC MAXIMUM PRINCIPLE FOR PROBLEMS WITH DELAY WITH DEPENDENCE ON THE PAST THROUGH GENERAL MEASURES

GIUSEPPINA GUATTERI

Dipartimento di Matematica, Politecnico di Milano
via Bonardi 9, 20133 Milano, Italia

FEDERICA MASIERO*

Dipartimento di Matematica e Applicazioni, Università di Milano-Bicocca
via Cozzi 55, 20125 Milano, Italia

(Communicated by Qi Lü)

ABSTRACT. We prove a stochastic maximum principle for a control problem where the state equation is delayed both in the state and in the control, and both the running and the final cost functionals may depend on the past trajectories. The adjoint equation turns out to be a new form of linear anticipated backward stochastic differential equations (ABSDEs in the following), and we prove a direct formula to solve these equations.

1. Introduction. In this paper we consider a controlled state equation for the process x with delay both in the state and in the control u , namely x is the solution to the following stochastic delay controlled equation in \mathbb{R}^n driven by an m -dimensional Brownian motion W :

$$\begin{cases} dx(t) = b(t, x_t, u_t)dt + \sigma(t, x_t, u_t)dW(t), \\ x(\theta) = x_0(\theta), \quad u(\theta) = \eta(\theta), \quad \theta \in [-d, 0]. \end{cases} \quad (1)$$

Here and throughout the paper we use the notation $x_t(\theta) = x(t+\theta)$, $u_t(\theta) = u(t+\theta)$, with $\theta \in [-d, 0]$ to denote the past trajectory of x and u from $t-d$ up to time t . We consider admissible controls u , that are progressively measurable and square integrable processes taking values in a convex set $U \subset \mathbb{R}^k$: in this case the stochastic maximum principle can be formulated in terms of the first order adjoint equation. We are able to allow a quite general dependence on the past trajectories x_t and u_t of the state and of the control, namely the drift and diffusion can be written as

$$\begin{aligned} f(t, x, u) &= \bar{f}(t, \int_{-d}^0 x(\theta)\mu_1(d\theta), \int_{-d}^0 u(\theta)\mu_3(d\theta)), \\ g(t, x, u) &= \bar{g}(t, \int_{-d}^0 x(\theta)\mu_2(d\theta), \int_{-d}^0 u(\theta)\mu_4(d\theta)), \end{aligned} \quad (2)$$

2020 *Mathematics Subject Classification.* 60H10, 93E20.

Key words and phrases. Stochastic maximum principle, delay, anticipated backward stochastic differential equations.

* Corresponding author: Federica Masiero.

where $\mu_1, \mu_2, \mu_3, \mu_4$ are finite regular measures on $[-d, 0]$, and $\bar{f}(t, \cdot, \cdot)$ and $\bar{g}(t, \cdot, \cdot)$ are Lipschitz continuous and differentiable. Associated to equation (1) we consider the cost functional

$$J(u(\cdot)) = \mathbb{E} \int_0^T l(t, x_t, u_t) dt + \mathbb{E} h(x_T) \tag{3}$$

that we have to minimize over all admissible controls. Also in the running cost l and in the final cost h we can allow a general dependence on the past trajectories x_t and x_T : there exist finite regular measures μ_5, μ_6, μ such that the current cost and the final cost can be written as follows

$$\begin{aligned} l(t, x, u) &= \bar{l}(t, \int_{-d}^0 x(\theta) \mu_5(d\theta), \int_{-d}^0 u(\theta) \mu_6(d\theta)), \\ h(x) &= \bar{h}(\int_{-d}^0 x(\theta) \mu(d\theta)), \quad \forall x \in C([-d, 0], \mathbb{R}^n), \end{aligned} \tag{4}$$

Such kind of dependence is rather general even though some relevant cases cannot be covered, like the dependence on the supremum of the history of the path, for example $h(x) = h(\sup_{\theta \in [-d, 0]} x(\theta))$.

We choose to attach our problem by means of the stochastic maximum principle since the dynamic programming approach stochastic optimal control problems governed by delay equations with delay in the control are usually harder to study than the ones when the delay appears only in the state. The main difficulty is that the associated Hamilton Jacobi Bellman equation is an infinite dimensional second order semilinear PDE, which is not trivial to solve, see e.g. [10, 11, 12]. Indeed the delay in the control cannot be directly treated by means of the dynamic programming principle, and in order to remove such delay in the control the problem must be turned into an infinite dimensional stochastic control problem that, unlike the infinite dimensional formulation of problem with delay in the state, does not satisfy the so called structure condition according to which the control affect the system as a perturbation of the noise, and moreover in many situations, including the case of pointwise delay, the control operator is unbounded. More general cases can be treated by applying the so called randomization method, see e.g. [1]: with this approach it is possible to characterize the value function but no conditions on the optimal control can be given.

On the contrary, studying a stochastic optimal control problem by means of the stochastic maximum principle allows to get conditions on the optimal control.

When studying our control problem by means of the stochastic maximum principle the adjoint equation turns out to be the following ABSDE for the pair of processes $(p, q) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T], \mathbb{R}^{n \times m})$,

$$\left\{ \begin{aligned} p(t) &= \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 \bar{l}_x(s - \theta, x(s - \theta), u(s - \theta)) \mu_5(d\theta) ds \\ &+ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 p(s - \theta) \bar{f}_x(s - \theta, x(s - \theta), u(s - \theta)) \mu_1(d\theta) ds \\ &+ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q(s - \theta) \bar{g}_x(s - \theta, x(s - \theta), u(s - \theta)) \mu_2(d\theta) ds \\ &+ \int_t^T q(s) dW_s + \int_{t \vee (T-d)}^T \bar{h}_x(x_T) \mu(d\theta), \\ p(T - \theta) &= 0, \quad q(T - \theta) = 0, \quad \forall \theta \in [-d, 0). \end{aligned} \right. \tag{5}$$

Notice that equation (5) does not make sense in differential form: the term

$$\frac{d}{dt} \int_{t \vee (T-d)}^T \bar{h}_x(x_T) p(T-\theta) \mu(d\theta)$$

is well defined when μ is absolutely continuous with respect to the Lebesgue measure. In order to be able to work with differentials, we will consider an ABSDE where μ is approximated by a sequence of finite regular measures $(\mu^n)_{n \geq 1}$ absolutely continuous with respect to the Lebesgue measure on $[T-d, T]$, so that the differential

$$\frac{d}{dt} \int_{t \vee (T-d)}^T \bar{h}_x(x_T) p(T-\theta) \mu^n(d\theta)$$

makes sense. In this way, for the approximating ABSDE the differential form makes sense.

The ABSDE (5) is a new type of linear ABSDEs, already considered in [14]: in the present paper for the solution of (5) we are able to give a representation which is the analogous of the one for linear BSDEs.

With these tools in hands, we are able to state necessary conditions for the optimality in terms of the pair of processes (p, q) : let (\bar{x}, \bar{u}) be an optimal pair and let $u^\rho = \bar{u} + \rho \bar{v}$, where \bar{v} is another admissible control, then, \mathbb{P} - a.s. and for a.e. $t \in [0, T]$:

$$\begin{aligned} & \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 (\bar{f}(t-\theta, \bar{x}_{t-\theta}, u_{t-\theta}^\rho) - \bar{f}(t-\theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta})) p(t-\theta) \mu_3(d\theta) \\ & \quad + \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 (l(t, \bar{x}_{t-\theta}, u_{t-\theta}^\rho) - l(t, \bar{x}_{t-\theta}, \bar{u}_{t-\theta})) \mu_6(d\theta) \\ & \quad + \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 (\bar{g}(t-\theta, \bar{x}_{t-\theta}, u_{t-\theta}^\rho) - \bar{g}(t-\theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta})) q(t-\theta) \mu_4(d\theta) \geq 0. \end{aligned} \quad (6)$$

Notice that this formula can be rewritten in a differential way under stronger assumptions on the coefficients, see Section 3, formula (57).

The results achieved by means of the stochastic maximum principle can be applied to a stochastic optimal control problem arising in advertisement models with delay and to an optimal portfolio problem with execution delay, we refer to Sections 4 and 5 for details.

After the introduction of anticipated backward stochastic differential equations (ABSDEs) in the paper [22], the stochastic maximum principle for delay equations has been widely studied in the literature. We mention, among others, [4], where a problem with pointwise delay in the state and in the control is studied, [19], where a controlled state equation driven by a Brownian motion and by a Poisson random measure is taken into account, and the delay affects the system by means of terms with a more restrictive structure than the one considered in (2), indeed, the measures $\mu_j, j = 1, \dots, 4$ all reduce to the same measure, absolutely continuous with respect to the Lebesgue measure, and with exponential density.

In the present paper we study the stochastic maximum principle for stochastic control problems where the state equation may present delay in the state and in the control and where in the associated cost functional we allow dependence on the past trajectory also in the final cost. Following the standard steps in the variational approach for control problems, we formulate the maximum principle by means of an adjoint equation. The novelty is that the adjoint equation turns out to be an

ABSDE of a more general form than the ones introduced in [22] and generalized in [23]. This is due to the fact that we allow dependence on the past trajectory also in the final cost. It turns out that the adjoint equation is not regular enough to be an Itô process and so to prove the stochastic maximum principle we must introduce some suitable regularized approximating problems.

The dependence on the past trajectory in the final cost has been studied also in [14], for an infinite dimensional evolution equation with delay only in the state and no control dependent noise. The adjoint ABSDE considered in [14] is similar to the one we handle here; in the present paper the ABSDE is solved directly by an extension of the formulas for linear BSDEs.

Concerning the recent literature based on ABSDEs, we are able to consider more general dependence on the past trajectory and moreover we can study the case when the final cost depends on the past trajectory of the state. As far as we know, such a general case is studied only in [17], with a direct functional analytic method, and the authors do not take into account the delay in the control.

The paper is organized as follows: in Section 2 we study the new form of linear ABSDEs, in Section 3 we present the control problem and after studying the variation of the state with respect to the variation of the control we formulate and prove the stochastic maximum principle, in Sections 4 and 5 the results are applied respectively to a stochastic dynamic model in marketing for problems of optimal advertising and to an optimal portfolio problem with execution delay.

1.1. Notations. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, $W(t)$ an m -dimensional Brownian motion and let $(\mathcal{F}_t)_{t \geq 0}$ be the natural filtration associated to W , augmented in the usual way with the family of \mathbb{P} -null sets of \mathcal{F} .

For any $p \in [1, \infty]$ and $T > 0$ we define

- $\mathcal{L}_{\mathcal{F}}^p(\Omega \times [0, T]; \mathbb{R}^k)$, the set of all (\mathcal{F}_t) -progressive processes with values in \mathbb{R}^k such that the norm

$$\begin{aligned} \|Y\|_{\mathcal{L}_{\mathcal{F}}^p(\Omega \times [0, T]; \mathbb{R}^k)}^p &= \left(\mathbb{E} \int_0^T |Y(t)|_{\mathbb{R}^k}^p dt \right)^{1/p} && \text{if } p < \infty, \\ \|Y\|_{\mathcal{L}_{\mathcal{F}}^\infty(\Omega \times [0, T]; \mathbb{R}^k)} &= \text{ess sup}_{\omega \in \Omega, t \in [0, T]} |Y_t(\omega)| && \text{if } p = \infty \end{aligned}$$

is finite. Here we take the *ess sup* with respect to $dt \otimes d\mathbb{P}$.

- $\mathcal{L}_{\mathcal{F}}^p(\Omega; C([0, T]; \mathbb{R}^k))$, the set of all (\mathcal{F}_t) -progressive and continuous processes with values in \mathbb{R}^k such that the norm

$$\begin{aligned} \|Y\|_{\mathcal{L}_{\mathcal{F}}^p(\Omega; C([0, T]; \mathbb{R}^k))}^p &= \mathbb{E} \sup_{t \in [0, T]} |Y_t|^p && \text{if } p < \infty, \\ \|Y\|_{\mathcal{L}_{\mathcal{F}}^\infty(\Omega; C([0, T]; \mathbb{R}^k))} &= \text{ess sup}_{\omega \in \Omega} \sup_{t \in [0, T]} |Y_t(\omega)| && \text{if } p = \infty \end{aligned}$$

is finite. Elements of this space are identified up to indistinguishability. We will denote the space as $\mathcal{S}_{\mathcal{F}}^p([0, T])$.

- $\mathcal{L}_{\mathcal{F}}^p(\Omega; B([0, T]; \mathbb{R}^k))$, the set of (\mathcal{F}_t) -progressive measurable and a.s bounded trajectories processes with values in \mathbb{R}^k such that the norm

$$\begin{aligned} \|Y\|_{\mathcal{L}_{\mathcal{F}}^p(\Omega; B([0, T]; \mathbb{R}^k))}^p &= \mathbb{E} \sup_{t \in [0, T]} |Y_t|^p && \text{if } p < \infty, \\ \|Y\|_{\mathcal{L}_{\mathcal{F}}^\infty(\Omega; B([0, T]; \mathbb{R}^k))} &= \text{ess sup}_{\omega \in \Omega} \sup_{t \in [0, T]} |Y_t(\omega)| && \text{if } p = \infty \end{aligned}$$

is finite. Elements of this space are identified up to indistinguishability. We will denote the space as $\mathcal{B}_{\mathcal{F}}^p([0, T])$

Throughout the paper given a progressive measurable process $y \in \mathcal{L}_{\mathcal{F}}^1(\Omega \times [0, T], \mathbb{R}^k)$, for $0 \leq s \leq t$ by $\mathbb{E}^{\mathcal{F}_s} y(t)$ we denote the optional projection of y into \mathcal{F}_s .

2. A new form of anticipated backward stochastic differential equations.

In this section we study ABSDEs which have the suitable form to be the adjoint equations in problems with delay we treat in the present paper. We will consider a stochastic differential equation of backward type, and on its coefficients we make the following assumptions.

Hypothesis 2.1. Let $f \in \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}^n)$, $g \in \mathcal{L}^\infty_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}^{n \times n})$, $h \in \mathcal{L}^\infty_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}^m)$ and $\xi \in \mathcal{B}^2_{\mathcal{F}}([T - d, T])$ where $0 < d \leq T$ is fixed. Let μ be a finite regular measure on $[T - d, T]$ and denote by $|\mu|$ its total variation, see [6].

Here we refer to the definition of regular measure, given in [6], page 156, according to which given a closed interval I a non negative measure μ defined on the σ -algebra $\mathcal{B}(I)$ of all the Borel sets of I is called *regular* if $\forall A \in \mathcal{B}(I)$

$$\begin{aligned} \mu(A) &= \sup\{\mu(F) : F \subseteq A, F \text{ is closed}\} \\ \mu(A) &= \inf\{\mu(G) : G \supseteq A, G \text{ is open}\} \end{aligned}$$

To be more precise in [6] such a measure is called strongly regular, but we follow most of the literature where such a measure is called *regular measure*.

If $h_1, h_2 \in \mathbb{R}^m$, we denote for brevity with $h_1 h_2$ the scalar product $\langle h_1, h_2 \rangle_{\mathbb{R}^m}$. Let us consider the following linear BSDE

$$\begin{aligned} p(t) &= \int_t^T f(s)ds + \int_t^T g(s)p(s)ds + \int_t^T q(s)h(s)ds + \int_t^T q(s)dW(s) \quad (7) \\ &+ \int_{t \vee (T-d)}^T \xi(s)\mu(ds). \end{aligned}$$

Let \mathbb{Q} be the probability measure, equivalent to the original one \mathbb{P} , such that

$$\tilde{W}(t) := \int_0^t h(s) ds + W(t)$$

is a \mathbb{Q} -Wiener process. Notice that $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(\int_0^T h(s) dB_s)$ thus, thanks to hypothesis 2.1 the two measures are equivalent, and in particular for any $\Theta \in L^1(\Omega, \mathbb{P})$

$$\mathbb{E}_{\mathbb{Q}}|\Theta| \leq C\mathbb{E}|\Theta| \quad \text{and} \quad \mathbb{E}|\Theta| \leq \tilde{C}\mathbb{E}_{\mathbb{Q}}|\Theta| \quad (8)$$

for some C, \tilde{C} depending only on T and the process h .

We have the following formula for the unique solution of the linear BSDE (7), which is the counterpart for the classical formula for the solution of a linear BSDE. We also notice that equation (7) is a linear BSDE with a final datum ξ acting not only at the final time T , but on the whole interval $[T - d, T]$.

Lemma 2.2. *Assume Hypothesis 2.1 holds true, then the BSDE (7) admits a unique adapted solution, that is a pair of processes $(p, q) \in \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}^n) \times \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}^{n \times m})$, satisfying the integral equation (7). The process p is given by the formula*

$$p(t) = \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_{t \vee (T-d)}^T e^{\int_t^s g(u)du} \xi(s)\mu(ds) + \int_t^T e^{\int_t^s g(u)du} f(s)ds \right]. \quad (9)$$

Proof. Under the probability measure \mathbb{Q} equation (7) can be rewritten as

$$p(t) = \int_t^T f(s)ds + \int_t^T g(s)p(s)ds + \int_t^T q(s)d\tilde{W}(s) + \int_{t \vee (T-d)}^T \xi(s)\mu(ds). \quad (10)$$

Let us first prove that the process p given by formula (9) verifies equation (7) for $t \in [T-d, T]$. For $t \in [T-d, T]$ equation (10) implies

$$p(t) = \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_t^T f(s) ds + \int_t^T g(s)p(s) ds + \int_t^T \xi(s)\mu(ds) \right]. \quad (11)$$

Using formula (9) we define

$$\bar{p}(t) := \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_t^T e^{\int_t^s g(u) du} \xi(s)\mu(ds) + \int_t^T e^{\int_t^s g(u) du} f(s) ds \right], \quad t \in [T-d, T]. \quad (12)$$

It is immediate to see that \bar{p} defined in formula (12) satisfies (11), indeed putting formula (12) for \bar{p} in equation (11) we get

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_t^T e^{\int_t^s g(u) du} \xi(s)\mu(ds) + \int_t^T e^{\int_t^s g(u) du} f(s) ds \right] \\ &= \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_t^T f(s) ds + \int_t^T \xi(s)\mu(ds) \right] \\ &+ \int_t^T g(s) \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_s} \left[\int_s^T e^{\int_s^r g(u) du} \xi(r)\mu(dr) + \int_s^T e^{\int_s^r g(u) du} f(r) dr ds \right] \\ &= \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_t^T f(s) ds + \int_t^T \xi(s)\mu(ds) \right] \\ &+ \int_t^T g(s) \left[\int_s^T e^{\int_s^r g(u) du} \xi(r)\mu(dr) + \int_s^T e^{\int_s^r g(u) du} f(r) dr \right] ds. \end{aligned} \quad (13)$$

Changing the order of integration it is immediate to see that

$$\left(\mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_t^T e^{\int_t^s g(u) du} \xi(s)\mu(ds) + \int_t^T e^{\int_t^s g(u) du} f(s) ds \right] \right)_{t \in [0, T]}$$

satisfies the integral equation (11) that corresponds to equation (10), by the usual application of the Martingale Representation Theorem.

Now let us consider the following equation, under \mathbb{Q} again, for $t \in [0, T-d]$:

$$p(t) = \int_t^{T-d} f(s) ds + \int_t^{T-d} g(s)p(s) ds + \int_t^{T-d} q(s) d\tilde{W}(s) + \bar{p}(T-d), \quad (14)$$

that is a standard BSDE with final datum $\bar{p}(T-d)$ defined in (12). The unique solution of (14) is given by:

$$\begin{aligned} \tilde{p}(t) &= \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[e^{\int_t^{T-d} g(u) du} \bar{p}(T-d) + \int_t^{T-d} e^{\int_t^s g(u) du} f(s) ds \right] \\ &= \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[e^{\int_t^{T-d} g(u) du} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_{T-d}} \left[\int_{T-d}^T e^{\int_{T-d}^s g(u) du} \xi(s)\mu(ds) \right. \right. \\ &\quad \left. \left. + \int_{T-d}^T e^{\int_{T-d}^s g(u) du} f(s) ds \right] + \int_t^{T-d} e^{\int_t^s g(u) du} f(s) ds \right], \end{aligned} \quad (15)$$

Hence $p(t)$ defined by (9), can also be written as

$$p(t) := \begin{cases} \bar{p}(t) & t \in [T - d, T], \\ \tilde{p}(t) & t \in [0, T - d] \end{cases} \tag{16}$$

So by construction p together with its corresponding martingale term q , that can be uniquely determined through the variation of the process p with the noise W , is a solution to (10), and hence a weak solution to (7) in the whole time interval $[0, T]$.

In particular, directly from (9) we get that, see also [21, Chapter 5], for any $\beta > 0$, taking also into account (8) and the Burkholder-Davis-Gundy inequalities:

$$\mathbb{E} \sup_{t \in [0, T]} e^{\beta t} |p(t)|^2 \leq C \mathbb{E}_{\mathbb{Q}} \sup_{t \in [0, T]} e^{\beta t} |p(t)|^2 \tag{17}$$

$$\leq C e^{\beta T} \mathbb{E}_{\mathbb{Q}} \left[\sup_{t \in [0, T]} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left| \int_t^T e^{\int_t^s g(u) du} \xi(s) \mu(ds) \right|^2 \right] \tag{18}$$

$$+ \sup_{t \in [0, T]} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left| \int_t^T e^{\int_t^s g(u) du} f(s) ds \right|^2 \Big]$$

$$\leq C e^{\beta T} \mathbb{E}_{\mathbb{Q}} \left[\sup_{t \in [0, T]} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left[\int_{T-d}^T |\xi(s)| |\mu|(ds) \right]^2 + \frac{1}{\beta} \sup_{t \in [0, T]} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \int_0^T e^{\beta s} |f(s)|^2 ds \right]$$

$$\leq C \mathbb{E}_{\mathbb{Q}} \left[e^{\beta T} \sup_{t \in [0, T]} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left(\int_{T-d}^T |\xi(s)| |\mu|(ds) \right)^2 + \frac{1}{\beta} \sup_{t \in [0, T]} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \int_0^T e^{\beta s} |f(s)|^2 ds \right]$$

$$\leq C \left[e^{\beta T} \left(\mathbb{E}_{\mathbb{Q}} \int_{T-d}^T |\xi(s)| |\mu|(ds) \right)^2 + \frac{1}{\beta} \mathbb{E}_{\mathbb{Q}} \int_0^T e^{\beta s} |f(s)|^2 ds \right]$$

$$\leq C \left[e^{\beta T} \left(\mathbb{E} \int_{T-d}^T |\xi(s)| |\mu|(ds) \right)^2 + \frac{1}{\beta} \mathbb{E} \int_0^T e^{\beta s} |f(s)|^2 ds \right]$$

where the constant C may change from line to line but only depends on T and $\|g\|_{\mathcal{L}^{\infty}(\Omega \times [0, T])}$, $\|h\|_{\mathcal{L}^{\infty}(\Omega \times [0, T])}$ and thus

$$\begin{aligned} & \mathbb{E} \sup_{t \in [0, T]} e^{\beta t} |p(t)|^2 \tag{19} \\ & \leq C \left[e^{\beta T} (|\mu|([T - d, T]))^2 \mathbb{E} \sup_{t \in [T-d, T]} |\xi(t)|^2 + \frac{1}{\beta} \mathbb{E} \int_0^T e^{\beta s} |f(s)|^2 ds \right]. \end{aligned}$$

By the latter calculations together with standard considerations see for instance [9], using the Martingale Representation Theorem, we get for all $\beta > 0$,

$$\begin{aligned} & \mathbb{E} \sup_{t \in [0, T]} e^{\beta t} |p(t)|^2 + \mathbb{E} \int_0^T e^{\beta s} |q(s)|^2 ds \tag{20} \\ & \leq c \left[\mathbb{E}_{\mathbb{Q}} \sup_{t \in [0, T]} e^{\beta t} |p(t)|^2 + \mathbb{E}_{\mathbb{Q}} \int_0^T e^{\beta s} |q(s)|^2 ds \right] \end{aligned}$$

$$\begin{aligned} &\leq c \left[e^{\beta T} \left(\mathbb{E}_{\mathbb{Q}} \int_{T-d}^T |\xi(s)| |\mu|(ds) \right)^2 + \frac{1}{\beta} \mathbb{E}_{\mathbb{Q}} \int_0^T e^{\beta s} |f(s)|^2 ds \right], \\ &\leq c \left[e^{\beta T} \left(\mathbb{E} \int_{T-d}^T |\xi(s)| |\mu|(ds) \right)^2 + \frac{1}{\beta} \mathbb{E} \int_0^T e^{\beta s} |f(s)|^2 ds \right], \end{aligned}$$

and these estimates imply

$$\begin{aligned} &\mathbb{E} \int_0^T \frac{\beta}{2} e^{\beta s} |p(s)|^2 ds + \mathbb{E} \int_0^T e^{\beta s} |q(s)|^2 ds \tag{21} \\ &\leq c \left[e^{\beta T} (|\mu|([T-d, T]))^2 \mathbb{E} \sup_{t \in [0, T]} |\xi(t)|^2 + \frac{1}{\beta} \mathbb{E} \int_0^T e^{\beta s} |f(s)|^2 ds \right], \end{aligned}$$

where the constant c depends on $\|g\|_{\mathcal{L}_{\mathcal{F}}^{\infty}(\Omega \times [0, T])}$, $\|h\|_{\mathcal{L}_{\mathcal{F}}^{\infty}(\Omega \times [0, T])}$, and T . Therefore $(p, q) \in \mathcal{S}_{\mathcal{F}}^2([0, T]) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T]; \mathbb{R}^{n \times m})$. Pathwise uniqueness follows by standard arguments since the non-classical terms disappears when one calculates the difference between solutions. By the Yamada-Watanabe type result for weak solutions for BSDEs, see [3], also for BSDEs pathwise uniqueness implies the uniqueness in law; and the pathwise uniqueness together with the existence of the weak solution imply the existence of the strong solution. This concludes the proof. \square

We apply the results collected in Lemma 2.2 to prove existence and uniqueness of a solution of the following anticipated ABSDE (22).

As before, let $(W_t)_{t \geq 0}$ be a standard m -dimensional Brownian motion, and on the coefficients we make the following assumptions.

Hypothesis 2.3. $f \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T], \mathbb{R}^n)$, $g \in \mathcal{L}_{\mathcal{F}}^{\infty}(\Omega \times [0, T+d], \mathbb{R}^{n \times n})$ $h \in \mathcal{L}_{\mathcal{F}}^{\infty}(\Omega \times [0, T+d], \mathbb{R}^m)$ and $\xi \in \mathcal{B}_{\mathcal{F}}^2([T-d, T])$ where $0 < d \leq T$ is fixed. Let μ be a finite regular measure on $[T-d, T]$, and μ_1 and μ_2 be finite regular measures on $[-d, 0]$.

Remark 1. The results in this Section can be extended from measures μ, μ_1, μ_2 in Hypothesis 2.3 to vector valued finite regular measures, that allow to consider more general dependence on the past trajectory.

We will prove existence and uniqueness of the following anticipated BSDEs of backward type:

$$\begin{cases} p(t) = \int_t^T f(s) ds + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 g(s-\theta) p(s-\theta) \mu_1(d\theta) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q(s-\theta) h(s-\theta) \mu_2(d\theta) ds + \int_t^T q_s dW_s \\ \quad + \int_{t \vee (T-d)}^T \xi(s) \mu(ds) \\ p(T-\theta) = 0, q(T-\theta) = 0 \quad \forall \theta \in [-d, 0). \end{cases} \tag{22}$$

The ABSDE (22) is of the form of equation introduced in [22] and generalized in [23], with the difference that it is given a final datum acting not only in $[T, T+d)$, but also in $[T-d, T]$, see also [14]. Notice that as soon as the process q belongs to $\mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T+d], \mathbb{R}^{n \times m})$ the term $\mathbb{E}^{\mathcal{F}} \cdot \int_{-d}^0 q(\cdot-\theta) h(\cdot-\theta) \mu_2(d\theta)$ has meaning since:

$$\mathbb{E} \int_0^T \int_{-d}^0 |q(t-\theta)|^2 |h(t-\theta)|^2 |\mu_2|(d\theta) dt$$

$$\leq |\mu|^2([-d, 0]) \|h\|_{\mathcal{L}^\infty}^2 \mathbb{E} \int_0^{T+d} |q(\rho)|^2 d\rho < +\infty$$

Theorem 2.4. *Let Hypothesis 2.3 holds true. Then the ABSDE (22) admits a unique adapted solution, that is a pair of processes $(p, q) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$, satisfying the integral equation (22).*

Proof. We notice that by the data of our problem, if it exists, the pair of processes (p, q) solution to the ABSDE (22), is such that $p(t) = q(t) = 0$ for $t \in (T, T + d)$.

Let us consider the more general equation, for $(\xi, \eta) \in \mathcal{L}_{\mathcal{F}}^2(\Omega; B([T - d, T + d]; \mathbb{R}^n)) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [T, T + d], \mathbb{R}^{n \times m})$

$$\left\{ \begin{aligned} p(t) &= \int_t^T f(s) ds + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 g(s - \theta) p(s - \theta) \mu_1(d\theta) ds \\ &+ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q(s - \theta) h(s - \theta) \mu_2(d\theta) ds + \int_t^T q(s) dW(s) \\ &+ \int_{t \vee (T-d)}^T \xi(s) \mu(ds) \\ p(T - \theta) &= \xi(\theta), \quad q(T - \theta) = \eta(\theta) \quad \forall \theta \in [-d, 0). \end{aligned} \right. \quad (23)$$

We prove existence of a solution by a fixed point argument on the space $\mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$ endowed with the equivalent norm

$$\|(p, q)\|_\beta = \mathbb{E} \int_0^{T+d} |p(s)|^2 e^{\beta s} ds + \int_0^{T+d} |q(s)|^2 e^{\beta s} ds, \quad (24)$$

with $\beta > 0$ to be chosen in the following.

Given $(y, z) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$ we define the map $\Gamma : \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m}) \rightarrow \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$.

The pair $(p, q) := \Gamma(y, z)$ is given by the pair of processes solution of the following BSDE given in integral form:

$$\left\{ \begin{aligned} p(t) &= \int_t^T f(s) ds + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 g(s - \theta) y(s - \theta) \mu_1(d\theta) ds \\ &+ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 z(s - \theta) h(s - \theta) \mu_2(d\theta) ds \\ &+ \int_t^T q(s) dW_s + \int_{t \vee (T-d)}^T \xi(s) \mu(ds) \\ p(T - \theta) &= \xi(\theta), \quad q(T - \theta) = \eta(\theta) \quad \forall \theta \in [-d, 0). \end{aligned} \right. \quad (25)$$

Thanks to Lemma 2.2 it turns out that $(p, q) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T], \mathbb{R}^{n \times m})$, and together with the condition given in (25) it turns out that $(p, q) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$. So Γ is well defined. Next we prove that Γ is a contraction. Let $y, \bar{y} \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n)$ and $z, \bar{z} \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$, and set $\hat{y} = y - \bar{y}$, $\hat{z} = z - \bar{z}$. We denote $(\hat{p}, \hat{q}) = \Gamma(y, z) - \Gamma(\bar{y}, \bar{z})$. So

$$\begin{aligned} \hat{p}(t) &= \int_t^T f(s) ds + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 g(s - \theta) \hat{y}(s - \theta) \mu_1(d\theta) ds \\ &+ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 \hat{z}(s - \theta) h(s - \theta) \mu_2(d\theta) ds + \int_t^T \hat{q}(s) dW_s. \end{aligned} \quad (26)$$

Equation (26) is a special case of the BSDE (7), whose existence and uniqueness have been studied in Lemma 2.2. By estimate (19) we get (here and in the following c is a constant whose value can change from line to line)

$$\begin{aligned}
 & \mathbb{E} \int_0^T \left(\frac{\beta}{2} |\hat{p}(s)|^2 + |\hat{q}(s)|^2 \right) e^{\beta s} ds & (27) \\
 & \leq \frac{2c}{\beta} \mathbb{E} \int_0^T \left| \int_{-d}^0 g(s-\theta) \hat{y}(s-\theta) \mu_1(d\theta) \right|^2 ds + \int_{-d}^0 |\hat{z}(s-\theta) h(s-\theta) \mu_2(d\theta)|^2 ds \\
 & \leq \frac{2c}{\beta} \left\{ \mathbb{E} \int_0^T \int_{-d}^0 |\hat{y}(s-\theta)|^2 |\mu_1|(d\theta) ds + \mathbb{E} \int_0^T \int_{-d}^0 |\hat{z}(s-\theta)|^2 |\mu_2|(d\theta) ds \right\} \\
 & \leq \frac{2c}{\beta} \left\{ \int_{-d}^0 \left[\mathbb{E} \int_0^T |\hat{y}(s-\theta)|^2 ds \right] |\mu_1|(d\theta) \right. \\
 & \quad \left. + \int_{-d}^0 \left[\mathbb{E} \int_t^T |\hat{z}(s-\theta)|^2 ds \right] |\mu_2|(d\theta) \right\} \\
 & = \frac{2c}{\beta} \left\{ \int_{-d}^0 \left[\mathbb{E} \int_0^T |\hat{y}(s)|^2 ds \right] |\mu_1|(d\theta) + \int_{-d}^0 \left[\mathbb{E} \int_t^T |\hat{z}(s)|^2 ds \right] |\mu_2|(d\theta) \right\} \\
 & \leq c \frac{2}{\beta} \|(\hat{y}, \hat{z})\|_{\beta}.
 \end{aligned}$$

By choosing $\beta > 0$ such that $c \frac{2}{\beta} < 1$ we have proved that Γ is a contraction, and its unique fixed point is the unique solution of the ABSDE (22). \square

Equation (22) can be written in differential form if we make some additional assumptions on the measure μ . If we assume that

$$\mu = c\delta_T + \tilde{\mu}, \quad c \in \mathbb{R},$$

where $\tilde{\mu}$ is a measure on $(T-d, T)$ absolutely continuous with respect to the Lebesgue measure, equation (22) can be written in differential form as

$$\begin{cases} -dp(t) = f(t)dt + \mathbb{E}^{\mathcal{F}^t} \int_{-d}^0 g(t-\theta)p(t-\theta)\mu_1(d\theta)dt \\ + \mathbb{E}^{\mathcal{F}^t} \int_{-d}^0 q(t-\theta)h(t-\theta)\mu_2(d\theta)dt + q(t)dW(t) + \xi(t)\eta_{\tilde{\mu}}(t)dt \\ p(T) = c\xi(T), \quad p(T-\theta) = 0, \quad q(T-\theta) = 0 \quad \forall \theta \in [-d, 0), \end{cases} \quad (28)$$

where for $t \in [T-d, T]$, $\eta_{\tilde{\mu}}$ is the Radon-Nikodym derivative of $\tilde{\mu}$ with respect to the Lebesgue measure, that is $\eta_{\tilde{\mu}}$ is defined by the relation

$$\tilde{\mu}(dt) = \eta_{\tilde{\mu}}(t)dt.$$

If we do not make additional assumptions on the measure μ , the differential form of equation (22) does not make sense, since the term

$$d \int_{t \vee (T-d)}^T \xi(s)\mu(ds)$$

is not well defined for $t \in [T-d, T]$. In the following, we build an approximating ABSDE whose differential form makes sense. This approximating ABSDE is obtained by a suitable approximation of μ : the construction of this sequence of approximating measures $(\mu_n)_{n \geq 1}$ is given in the following Lemma, which is the analogous of

Lemma 5 in [14]. In the following, given $I \subset \mathbb{R}$ we denote with $C_b(I, \mathbb{R})$ the space of bounded and continuous functions from I to \mathbb{R} ; with the notation $\lambda_{[T-d, T]}$, we denote the Lebesgue measure on $[T - d, T]$.

Lemma 2.5. *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 measures on $[T - d, T]$, absolutely continuous with respect to $\lambda_{[T-d, T]}$ and such that*

$$\bar{\mu} = \lim_{n \rightarrow \infty} \bar{\mu}_n, \quad \text{in the sense of the narrow convergence,} \tag{29}$$

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

$$\int_{T-d}^T f d\bar{\mu} = \lim_{n \rightarrow \infty} \int_{T-d}^T f d\bar{\mu}_n \tag{30}$$

Notice that we can apply the previous Lemma to the approximation of the measure μ by defining $\bar{\mu}$ such that for any $A \in \mathcal{B}([T - d, T])$

$$\bar{\mu}(A) = \mu(A \setminus \{T\}) : \tag{31}$$

the measure $\bar{\mu}$ is obtained from the original measure μ , by subtracting to μ its mass in $\{T\}$. Lemma 2.5 ensures that there exists a sequence of measures $(\bar{\mu}_n)_{n \geq 1}$, on $[T - d, T]$, which are absolutely continuous with respect to the Lebesgue measure on $[T - d, T]$ and converge to $\bar{\mu}$.

The next step is to build an approximation, in a sense that we are going to precise, of the equation (22), by approximating $\bar{\mu}$ obtained by μ in (31).

Proposition 1. *Let Hypothesis 2.3 holds true and assume $\xi \in \mathcal{S}_{\mathcal{F}}^2([T - d, T])$, let $\bar{\mu}$ be defined by (31), and let us consider $(\bar{\mu}_n)_n$ the approximations of $\bar{\mu}$, absolutely continuous with respect to the Lebesgue measure on $(T - d, T)$. Let us consider the approximating ABSDEs (of “standard” type):*

$$\left\{ \begin{array}{l} p^n(t) = \int_t^T f(s)ds + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 g(s - \theta)p^n(s - \theta)\mu_1(d\theta)ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q^n(s - \theta)h(s - \theta)\mu_2(d\theta)ds \\ \quad + \int_t^T q^n(s)dW_s + \int_{t \vee (T-d)}^T \xi(s)\bar{\mu}_n(ds) + \mu(T)\xi(T), \\ p^n(T - \theta) = 0, \quad q^n(T - \theta) = 0 \quad \forall \theta \in [-d, 0). \end{array} \right. \tag{32}$$

Then the pair (p^n, q^n) , solution to (32) converges in $\mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$ to the pair (p, q) solution to (22).

Proof. Let us first prove that the sequence $(p^n, q^n)_n$ is a Cauchy sequence in $\mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T + d], \mathbb{R}^{n \times m})$. The equation satisfied by $(p^n(t) -$

$p^k(t), q^n(t) - q^k(t)$, $n, k \geq 1$, turns out to be the following ABSDE:

$$\left\{ \begin{array}{l} p^n(t) - p^k(t) = \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 g(s-\theta) (p^n(s-\theta) - p^k(s-\theta)) \mu_1(d\theta) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 (q^n(s-\theta) - q^k(s-\theta)) h(s-\theta) \mu_2(d\theta) ds \\ \quad + \int_t^T (q^n(s) - q^k(s)) dW(s) + \int_{t \vee (T-d)}^T \xi(s) \bar{\mu}_n(ds) \\ \quad - \int_{t \vee (T-d)}^T \xi(s) \bar{\mu}_k(ds), \\ p^n(T-\theta) - p^k(T-\theta) = 0, \quad q^n(T-\theta) - q^k(T-\theta) = 0 \quad \forall \theta \in [-d, 0]. \end{array} \right. \quad (33)$$

Notice that the terms $\int_{t \vee (T-d)}^T \xi(s) \bar{\mu}_n(ds)$, $\int_{t \vee (T-d)}^T \xi(s) \bar{\mu}_k(ds)$, are Ito terms, so equation (33) is a standard ABSDE. By standard estimates, see e.g. [23], Lemma 2.3, formula (3), as $n, k \rightarrow \infty$

$$\begin{aligned} \mathbb{E}|p^n(t) - p^k(t)|^2 + \mathbb{E} \int_t^T |p^n(s) - p^k(s)|^2 ds + \mathbb{E} \int_t^T |q^n(s) - q^k(s)|^2 ds \\ \leq \mathbb{E} \left| \int_{t \vee (T-d)}^T \xi(s) (\bar{\mu}_n(ds) - \bar{\mu}_k(ds)) \right|^2 \rightarrow 0 \end{aligned}$$

by the narrow convergence of the sequence of measures $\bar{\mu}_n$. So the sequence (p^n, q^n) is a Cauchy sequence in $\mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T+d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T+d], \mathbb{R}^{n \times m})$. It remains to show that it converges to (p, q) solution of equation (22). Let us denote

$$(\bar{p}, \bar{q}) = \lim_{n \rightarrow \infty} (p^n, q^n) \text{ in } \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T+d], \mathbb{R}^n) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T+d], \mathbb{R}^{n \times m}).$$

and for every $t \in [0, T]$:

$$\lim_{n \rightarrow +\infty} \mathbb{E}|p^n(t) - \bar{p}(t)|^2 = 0$$

Notice also that $p^n(s-\theta), \bar{p}(s-\theta) = 0$, for $s-\theta > T$: thus we have that, for all $t \in [0, T]$,

$$\begin{aligned} & \mathbb{E} \left| \int_t^T \int_{-d}^0 g(s-\theta) p^n(s-\theta) \mu_1(d\theta) ds - \int_t^T \int_{-d}^0 g(s-\theta) \bar{p}(s-\theta) \mu_1(d\theta) ds \right|^2 \\ &= \mathbb{E} \left| \int_{-d}^0 \int_t^T g(s-\theta) (p^n(s-\theta) - \bar{p}(s-\theta)) ds \mu_1(d\theta) \right|^2 \\ &\leq |g|_{\mathcal{L}^\infty[0, T]}^2 \mathbb{E} \left(\int_{-d}^0 \int_t^T |p^n(s-\theta) - \bar{p}(s-\theta)| ds |\mu_1|(d\theta) \right)^2 \\ &\leq |g|_{\mathcal{L}^\infty[0, T]}^2 \mathbb{E} \left(\int_{-d}^0 \int_{(t+\theta) \vee 0}^{T+\theta} |p^n(\sigma) - \bar{p}(\sigma)| d\sigma |\mu_1|(d\theta) \right)^2 \\ &\leq |g|_{\mathcal{L}^\infty[0, T]}^2 \mathbb{E} \left(\int_{-d}^0 \int_0^T |p^n(\sigma) - \bar{p}(\sigma)| d\sigma |\mu_1|(d\theta) \right)^2 \\ &\leq |g|_{\mathcal{L}^\infty[0, T]}^2 \mathbb{E} \left(\int_{-d}^0 |\mu_1|(d\theta) \int_0^T |p^n(\sigma) - \bar{p}(\sigma)| d\sigma \right)^2 \end{aligned}$$

$$\begin{aligned} &\leq |g|_{\mathcal{L}^\infty[0,T]}^2 \left(\int_{-d}^0 |\mu_1|(d\theta) \right)^2 \mathbb{E} \left(\int_0^T |p^n(\sigma) - \bar{p}(\sigma)| d\sigma \right)^2 \\ &\leq T |g|_{\mathcal{L}^\infty[0,T]}^2 |\mu_1|([-d, 0])^2 \mathbb{E} \int_0^T |p^n(r) - \bar{p}(r)|^2 dr. \end{aligned}$$

So

$$\begin{aligned} &\mathbb{E} \left| \int_t^T \int_{-d}^0 q^n(s - \theta) h(s - \theta) \mu_2(d\theta) ds - \int_t^T \int_{-d}^0 \bar{q}(s - \theta) h(s - \theta) \mu_2(d\theta) ds \right|^2 \\ &\leq T |h|_{\mathcal{L}^\infty[0,T]}^2 |\mu_2|([-d, 0])^2 \mathbb{E} \int_0^T |q^n(r) - \bar{q}(r)|^2 dr \end{aligned}$$

By passing to the limit as $n \rightarrow \infty$ in equation (32) we get

$$\left\{ \begin{aligned} \mathbb{E}^{\mathcal{F}_t} \bar{p}(t) &= \mathbb{E}^{\mathcal{F}_t} \int_t^T f(s) ds + \mathbb{E}^{\mathcal{F}_t} \int_t^T \int_{-d}^0 g(s - \theta) \bar{p}(s - \theta) \mu_1(d\theta) ds \\ &+ \mathbb{E}^{\mathcal{F}_t} \int_t^T \int_{-d}^0 \bar{q}(s - \theta) h(s - \theta) \mu_2(d\theta) ds \\ &+ \mathbb{E}^{\mathcal{F}_t} \left[\int_{t \vee (T-d)}^T \xi(s) \bar{\mu}(ds) + \mu(T) \xi_T \right], \\ \bar{p}(T - \theta), \bar{q}(T - \theta) &= 0 \quad \forall \theta \in [-d, 0]. \end{aligned} \right.$$

It follows immediately that $(\bar{p}, \bar{q}) = (p, q)$ and this concludes the proof. □

3. The controlled problem and the stochastic maximum principle. Let us consider the following controlled state equation in \mathbb{R}^n

$$\begin{cases} dx(t) = f(t, x_t, u_t) dt + g(t, x_t, u_t) dW(t), & t \in [0, T], \\ x(\theta) = \bar{x}(\theta), \quad u(\theta) = \eta(\theta), & \theta \in [-d, 0], \end{cases} \tag{34}$$

where W in this section, for simplicity of notation, will be supposed to be a real standard Brownian motion, and x_t and u_t denote the past trajectories from time $t-d$ up to time t . Moreover \bar{x} and η are the initial paths of the state and of the control respectively, and we assume η to be deterministic and such that $\int_{-d}^0 \eta^2(t) dt < +\infty$. By admissible control we mean an \mathcal{F}_t -progressively measurable process with values in a convex set $U \subset \mathbb{R}^k$.

$$\mathbb{E} \int_{-d}^T |u(t)|^2 dt < \infty, \tag{35}$$

such that $u(\theta) = \eta(\theta)$, \mathbb{P} -a.s. for a.e. $\theta \in [-d, 0]$. We will denote this space of admissible controls by \mathcal{U} .

We want to minimize the following cost functional

$$J(u(\cdot)) = \mathbb{E} \int_0^T l(t, x_t, u_t) dt + \mathbb{E} h(x_T) \tag{36}$$

over all admissible controls. We make the following assumptions on f, g, l, h and on the initial condition \bar{x} . Here and in the following we denote by $E = C_b([-d, 0], \mathbb{R}^n)$ and $K = C_b([-d, 0], \mathbb{R}^k)$.

Hypothesis 3.1. Let μ_i , $i = 1, \dots, 6$ and μ be finite regular measures.

We assume that f , g , l and h are defined for any $x \in E$ and any $u \in K$ in terms of $\bar{f} : \Omega \times [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^n$, $\bar{g} : \Omega \times [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^n$, $\bar{l} : [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}$ and $\bar{h} : \mathbb{R}^n \rightarrow \mathbb{R}$ as follows

$$\begin{aligned} f(t, x, u) &= \bar{f}(t, \int_{-d}^0 x(\theta)\mu_1(d\theta), \int_{-d}^0 u(\theta)\mu_3(d\theta)), \\ g(t, x, u) &= \bar{g}(t, \int_{-d}^0 x(\theta)\mu_2(d\theta), \int_{-d}^0 u(\theta)\mu_4(d\theta)), \\ l(t, x, u) &= \bar{l}(t, \int_{-d}^0 x(\theta)\mu_5(d\theta), \int_{-d}^0 u(\theta)\mu_6(d\theta)), \\ h(x) &= \bar{h} \left(\int_{-d}^0 x(\theta)\mu(d\theta) \right). \end{aligned}$$

Here and in the following in the drift f , and correspondingly in \bar{f} , and in the diffusion term g , and correspondingly in \bar{g} , we omit the dependence on ω . We assume that for each \mathcal{F}_t , $t \in [0, T]$ -adapted processes, $x, u \in \mathcal{L}_{\mathcal{F}}^2(\Omega, B([0, T; \mathbb{R}^k]))$, the processes $\bar{f}(\cdot, x, u)$, $\bar{g}(\cdot, x, u)$ are also \mathcal{F}_t , $t \in [0, T]$ -adapted.

We will assume that \bar{f} , \bar{g} and \bar{l} are Borel measurable and differentiable with respect to the second and to the third variable, that with an abuse of notation we still refer to as x and u . Moreover $\bar{f}_x, \bar{f}_u, \bar{g}_x$ and \bar{g}_u are uniformly bounded, while \bar{l}_x, \bar{l}_u have linear growth with respect to x and u , uniformly in t , finally \bar{h} is differentiable and \bar{h}_x has linear growth too. Moreover we will use the following notations

$$\begin{aligned} \bar{f}_x(t, \int_{-d}^0 x(\theta)\mu_1(d\theta), \int_{-d}^0 u(\theta)\mu_3(d\theta)) &=: \bar{f}_x(t, x, u), \\ \bar{f}_u(t, \int_{-d}^0 x(\theta)\mu_1(d\theta), \int_{-d}^0 u(\theta)\mu_3(d\theta)) &=: \bar{f}_u(t, x, u) \\ \bar{g}_x(t, \int_{-d}^0 x(\theta)\mu_2(d\theta), \int_{-d}^0 u(\theta)\mu_4(d\theta)) &=: \bar{g}_x(t, x, u), \\ \bar{g}_u(t, \int_{-d}^0 x(\theta)\mu_2(d\theta), \int_{-d}^0 u(\theta)\mu_4(d\theta)) &=: \bar{g}_u(t, x, u) \\ \bar{l}_x(t, \int_{-d}^0 x(\theta)\mu_5(d\theta), \int_{-d}^0 u(\theta)\mu_6(d\theta)) &=: \bar{l}_x(t, x, u), \\ \bar{l}_u(t, \int_{-d}^0 x(\theta)\mu_5(d\theta), \int_{-d}^0 u(\theta)\mu_6(d\theta)) &=: \bar{l}_u(t, x, u) \\ \bar{h}_x \left(\int_{-d}^0 x(\theta)\mu(d\theta) \right) &=: \bar{h}_x(x). \end{aligned}$$

Remark 2. In analogy to Remark 1, all the results in this Section and throughout the paper can be extended to measures μ_i , $i = 1, \dots, 6$, μ considered in Hypothesis 3.1 possibly vector valued finite regular measures with values respectively in \mathbb{R}^{j_i} , $i = 1, \dots, 6$, \mathbb{R}^j , with $j_i, j \geq 1$, $i = 1, \dots, 6$.

We notice that when the coefficients are stochastic, under Hypothesis 3.1 existence and uniqueness of a solution of equation (34) holds true, see e.g. [18], Theorem I.1

We notice that the terms

$$\int_{-d}^0 u(\theta)\mu_3(d\theta), \quad \int_{-d}^0 u(\theta)\mu_4(d\theta), \quad \int_{-d}^0 u(\theta)\mu_6(d\theta)$$

appearing respectively in the drift f , in the diffusion g and in the current cost l do not make sense in a standard way and for every $t \in [0, T]$ as soon as the control u is not assumed to be integrable with respect to the measures μ_3, μ_4, μ_6 , but only square integrable with respect to the Lebesgue measure in $[-d, 0]$.

So it is necessary to give a precise meaning to the state equation and to the current cost. First of all we want to clarify that for any $u \in \mathcal{U}$ equation (34) is well defined, indeed for any $u \in \mathcal{U}$ and any finite regular measure $\tilde{\mu}$ we have that:

$$\mathbb{E} \int_0^T \int_{-d}^0 |u(t + \theta)|^2 |\tilde{\mu}|(d\theta) dt \leq |\tilde{\mu}|([-d, 0]) \mathbb{E} \int_{-d}^T |u(\rho)|^2 d\rho < +\infty$$

thus

$$\int_{-d}^0 |u(t + \theta)|^2 |\tilde{\mu}|(d\theta) < +\infty, \quad \text{a.s. for a.e. } t \in [0, T].$$

Then we can deduce that, thanks to Hypothesis 3.1, for all $x \in \mathcal{S}_{\mathcal{F}}^p([-d, T])$ and $u \in \mathcal{U}$ the processes are square integrable: $f(t, x_t, u_t) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T]; \mathbb{R}^n)$, $g(t, x_t, u_t) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T]; \mathbb{R}^n)$. In a similar way it follows that the current cost is well defined.

Moreover for any $u \in \mathcal{U}$, there exists a solution $x = x^u \in \mathcal{S}_{\mathcal{F}}^p([-d, T])$: the result follows in the same way as for controlled stochastic delay equations without delay in the control, and it is substantially contained e.g. in [18], where stochastic delay equations with random drift and diffusion are solved.

Next we want to show that the adjoint equation of a delay equation is of the form of ABSDE (22), and it allows to formulate a stochastic maximum principle for finite dimensional controlled state equations with delay, and in the case of final cost functional depending on the history of the process.

Many recent papers, see e.g. [4], [5], deal with similar problems, but only in the simpler case of final cost functionals not depending on the past of the process. Moreover only the case of pointwise delay is considered, or in some cases the past affects the system at time t by terms of the form

$$\int_{-d}^0 e^{-\lambda\theta} \xi(t + \theta) d\theta$$

where ξ may coincide with the state x of the system, and/or with the control u . These two choices coincide respectively with taking the measures $\mu_i, i = 1, \dots, 6$ delta Dirac measures and measures absolutely continuous with respect to the Lebesgue measure and with exponential density.

In the present paper we are able to handle $\mu_i, i = 1, \dots, 6$ finite regular measures on $[-d, 0]$: such a general case is treated in the paper [16], only in the case without delay in the control and it is here proved by means of anticipated BSDEs.

In order to write the adjoint equation, at first we study the variation of the state: let us consider the pair (x, u) , where x is solution to equation (34) and u is the control process in this equation, and let $v \in \mathcal{U}$ be another admissible control; set $\bar{v} = v - \bar{u}$ and

$$u^\rho = \bar{u} + \rho \bar{v}. \tag{37}$$

Also u^ρ turns out to be an admissible control. Let x^ρ be the solution of equation (34) corresponding to the admissible control u^ρ and let y be the solution of the following linear equation

$$\begin{cases} dy(t) = \int_{-d}^0 \bar{f}_x(t, \bar{x}_t, \bar{u}_t) y_t(\theta) \mu_1(d\theta) dt + \int_{-d}^0 \bar{f}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}_t(\theta) \mu_3(d\theta) dt + \\ \quad + \int_{-d}^0 \bar{g}_x(t, \bar{x}_t, \bar{u}_t) y_t(\theta) \mu_2(d\theta) dW_t + \int_{-d}^0 \bar{g}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}_t(\theta) \mu_4(d\theta) dW_t \\ y(\theta) = 0, \forall \theta \in [-d, 0]. \end{cases} \tag{38}$$

With an immediate extension of Theorem 3.2 in [17] to the case with delay in the control, we have the following first order expansion

$$x^\rho(t) = \bar{x}(t) + \rho y(t) + R^\rho(t), \quad t \in [0, T], \quad \lim_{\rho \rightarrow 0} \frac{1}{\rho^2} \mathbb{E} \sup_{t \in [0, T]} |R^\rho(t)|^2 = 0. \tag{39}$$

We are going to prove that equation (22) with

$$\begin{aligned} f(t) &= \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{l}_x(t - \theta, \bar{x}(t - \theta), \bar{u}(t - \theta))(d\theta), \\ g(t) &= \bar{f}_x(t, \bar{x}_t, \bar{u}_t), \quad h(t) = \bar{g}_x(t, \bar{x}_t, \bar{u}_t), \quad \xi(t) = h_x(\bar{x}_T). \end{aligned} \tag{40}$$

is the adjoint equation in the control problem with cost functional (36). We notice with the coefficients given by (40) the BSDE (22) is solvable by Theorem 2.4 since Hypothesis 2.3 is satisfied.

To prove that (22) is the adjoint equation, for a.a. $\tau \in [0, T]$, $x \in E$, $u \in L^2([-d, T]; U)$, $p, q \in \mathbb{R}^n$, we define the Hamiltonian function as

$$\begin{aligned} \mathcal{H}(t, x, u, p, q) &= \bar{f} \left(\tau, \int_{-d}^0 x(\theta) \mu_1(d\theta), \int_{-d}^0 u(\theta) \mu_3(d\theta) \right) p \\ &\quad + \bar{g} \left(\tau, \int_{-d}^0 x(\theta) \mu_2(d\theta), \int_{-d}^0 u(\theta) \mu_4(d\theta) \right) q \\ &\quad + \bar{l} \left(\tau, \int_{-d}^0 x(\theta) \mu_5(d\theta), \int_{-d}^0 u(\theta) \mu_6(d\theta) \right) \\ &= \bar{f}(\tau, x, u) p + \bar{g}(\tau, x, u) q + l(\tau, x, u), \end{aligned} \tag{41}$$

where the last expression will be used, with an abuse of notation, to shorten the formulas. Notice that the Hamiltonian function is not defined for every τ , as discussed at the beginning of this Section, due to the fact that \bar{f} , \bar{g} and \bar{l} depend respectively on the terms $\int_{-d}^0 u(t - \theta + \eta) \mu_3(d\eta)$, $\int_{-d}^0 u(t - \theta + \eta) \mu_4(d\eta)$ and $\int_{-d}^0 u(t - \theta + \eta) \mu_6(d\eta)$.

The Hamiltonian function turns out to be a p -integrable function in time for any $p \geq 1$, and so for any function $v \in L^q([0, T])$ the integral

$$\int_0^T \mathcal{H}(t, x, u, p, q) v(t) dt, \quad t \in [0, T], \quad x \in E, u \in L^2([-d, T]; U), p, q \in \mathbb{R}^n.$$

makes sense, and this integral appears in the proof of the stochastic maximum principle, see the next Theorem on the stochastic maximum principle.

In the formulation of the stochastic maximum principle, the adjoint ABSDE turns out to be nothing else than equation (22), with with f, g, h and ξ given in (40).

Theorem 3.2. *Let Hypothesis 3.1 holds true. Let (p, q) be the unique solution of the ABSDE*

$$\left\{ \begin{aligned} p(t) &= \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 \bar{l}_x(s - \theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_5(d\theta) ds \\ &+ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 p(s - \theta) \bar{f}_x(s - \theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_1(d\theta) ds \\ &+ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q(s - \theta) \bar{g}_x(s - \theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_2(d\theta) ds \\ &+ \int_t^T q(s) dW_s + \int_{t \vee (T-d)}^T \bar{h}_x(\bar{x}_T) \mu(d\theta) \\ p(T - \theta) &= 0, q(T - \theta) = 0 \quad \forall \theta \in [-d, 0]. \end{aligned} \right. \tag{42}$$

Let (\bar{x}, \bar{u}) be an optimal pair for the optimal control problem of minimizing the cost functional (36) related to the controlled state equation (34). Then the following condition holds:

$$\begin{aligned} &\langle v(t) - \bar{u}(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{f}_u(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) p(t - \theta) \mu_3(d\theta) \rangle \\ &+ \langle v(t) - \bar{u}(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{g}_u(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) q(t - \theta) \mu_4(d\theta) \rangle \\ &+ \langle v(t) - \bar{u}(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{l}_u(t, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) \mu_6(d\theta) \rangle \geq 0 \quad dt \times \mathbb{P} - a.e.; \end{aligned} \tag{43}$$

for all $v \in \mathcal{U}$.

Remark 3. We notice that in equation (42) and in condition (43) the terms $\mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{g}_x(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) q(t - \theta) \mu_2(d\theta)$ and $\mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{g}_u(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) q(t - \theta) \mu_4(d\theta)$ make sense only when integrated with respect to t as we already pointed out for the control terms.

As it is well known, the stochastic maximum principle can be reformulated without differentiability assumptions on the coefficients as stated in Hypothesis 3.1. In the place of differentiability, we assume that \bar{f} and \bar{g} are Lipschitz continuous with respect to x, u , \bar{l} is locally Lipschitz continuous with respect to x, u , and \bar{l} is locally Lipschitz continuous. In this case condition (44) can be replaced by a condition on the variation of the Hamiltonian function. Namely let v be another admissible control, set $\bar{v} = v - \bar{u}$ and $u^\rho = \bar{u} + \rho \bar{v}$, condition (44) can be substituted by

$$\begin{aligned} &\mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 (\bar{f}(t - \theta, \bar{x}_{t-\theta}, u_{t-\theta}^\rho) - \bar{f}(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta})) p(t - \theta) \mu_3(d\theta) \\ &+ \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 (\bar{g}(t - \theta, \bar{x}_{t-\theta}, u_{t-\theta}^\rho) - \bar{g}(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta})) q(t - \theta) \mu_4(d\theta) \\ &+ \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 (\bar{l}(t - \theta, \bar{x}_{t-\theta}, u_{t-\theta}^\rho) - \bar{l}(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta})) \mu_6(d\theta) \geq 0, \end{aligned} \tag{44}$$

$dt \times d\mathbb{P} - a.e.$. This form of the maximum principle can be obtained in a similar to the differentiable case, without writing the variation of the coefficients in terms of derivatives.

Finally we notice that, unlike the undelayed case, both conditions (43) and (44) cannot be expressed with any derivative or variation of the Hamiltonian.

Proof of Theorem 3.2. As we already pointed out, see the comment after the proof of Theorem 2.4, the adjoint equation (42) is not regular enough to perform directly the usual proof of the maximum principle. Thus during the proof we must introduce some suitable regularized approximating problem to apply the Itô formula and deduce the necessary condition (43).

As usual in proving the stochastic maximum principle, we start by writing the variation of the cost functional. Namely, following (37), let (\bar{x}, \bar{u}) be an optimal pair and let v be another admissible control, set $\bar{v} = v - \bar{u}$ and $u^\rho = \bar{u} + \rho\bar{v}$. We can write the variation of the cost functional,

$$\delta J = J(u^\rho(\cdot)) - J(\bar{u}(\cdot)),$$

as

$$\begin{aligned} 0 \leq \delta J &= J(u^\rho(\cdot)) - J(\bar{u}(\cdot)) \\ &= \mathbb{E} \int_0^T l(t, x_t^\rho, u_t^\rho) dt - \mathbb{E} \int_0^T l(t, \bar{x}_t, \bar{u}_t) dt + \mathbb{E} (h(x_T^\rho) - h(\bar{x}_T)) = I_1 + I_2. \end{aligned} \quad (45)$$

Now

$$\begin{aligned} I_1 &= \mathbb{E} \int_0^T l(t, x_t^\rho, u_t^\rho) dt - \mathbb{E} \int_0^T l(t, \bar{x}_t, \bar{u}_t) dt \\ &= \left[\mathbb{E} \int_0^T l(t, x_t^\rho, u_t^\rho) dt - \mathbb{E} \int_0^T l(t, \bar{x}_t, u_t^\rho) dt \right] \\ &\quad + \left[\mathbb{E} \int_0^T l(t, \bar{x}_t, u_t^\rho) dt - \mathbb{E} \int_0^T l(t, \bar{x}_t, \bar{u}_t) dt \right] = J_1 + J_2 \end{aligned}$$

We rewrite (39) as

$$x^\rho(t) = \bar{x}(t) + \rho y(t) + R^\rho(t), \quad t \in [0, T], \quad \lim_{\rho \rightarrow 0} \frac{1}{\rho^2} \mathbb{E} \sup_{t \in [0, T]} |R^\rho(t)|^2 = 0, \quad (46)$$

where y is solution to equation (38). We start by computing J_1 :

$$\begin{aligned} J_1 &= \mathbb{E} \int_0^T \left[l \left(t, \int_{-d}^0 x^\rho(t+\theta) \mu_5(d\theta), \int_{-d}^0 u^\rho(t+\theta) \mu_6(d\theta) \right) \right. \\ &\quad \left. - l \left(t, \int_{-d}^0 \bar{x}(t+\theta) \mu_5(d\theta), \int_{-d}^0 u^\rho(t+\theta) \mu_6(d\theta) \right) \right] dt \\ &= \mathbb{E} \int_0^T \int_0^1 \int_{-d}^0 (x_t^\rho - \bar{x}_t) \bar{l}_x(t, (\bar{x}_t + \lambda(x_t^\rho - \bar{x}_t)), u_t^\rho) \mu_5(d\theta) d\lambda dt \\ &= \mathbb{E} \int_0^T \int_0^1 \int_{-d}^0 (\rho y_t(\theta) + R^\rho(t+\theta)) \bar{l}_x(t, \bar{x}_t + \lambda(x_t^\rho - \bar{x}_t), u_t^\rho) \mu_5(d\theta) d\lambda dt. \end{aligned} \quad (47)$$

By similar computations we obtain the analogous formula for J_2 :

$$\begin{aligned} J_2 &= \mathbb{E} \int_0^T \int_0^1 \int_{-d}^0 (u_t^\rho - \bar{u}_t) \bar{l}_u(t, \bar{x}_t, u_t^\rho + \lambda(u_t^\rho - \bar{u}_t)) \mu_6(d\theta) d\lambda dt \\ &= \mathbb{E} \int_0^T \int_0^1 \int_{-d}^0 \rho \bar{v}(t+\theta) \bar{l}_u(t, \bar{x}_t, u_t^\rho + \lambda(u_t^\rho - \bar{u}_t)) \mu_6(d\theta) d\lambda dt. \end{aligned} \quad (48)$$

Notice that the last term is well defined only when \bar{u}, u^ρ are continuous, i.e. belong to E , but can be extended to the whole $L^2([-d, T]; U)$ by a standard density argument. We now compute I_2 :

$$\begin{aligned} I_2 &= \mathbb{E} (h(x_T^\rho) - h(\bar{x}_T)) \tag{49} \\ &= \mathbb{E} \int_0^1 \int_{-d}^0 (x^\rho(T + \theta) - \bar{x}(T + \theta)) \bar{h}_x(\bar{x}_T + \lambda(x_T^\rho - \bar{x}_T)) \mu(d\theta) \\ &= \mathbb{E} \int_0^1 \int_{-d}^0 (\rho y(T + \theta) + R^\rho(T + \theta)) \bar{h}_x(\bar{x}_T + \lambda(x_T^\rho - \bar{x}_T)) \mu(d\theta) \\ &= \mathbb{E} \int_0^1 \int_{-d}^0 (\rho y(T + \theta) + R^\rho(T + \theta)) \bar{h}_x(\bar{x}_T + \lambda(x_T^\rho - \bar{x}_T)) \mu(d\theta). \end{aligned}$$

Now we follow Lemma 2.5 and we decompose the measure μ into

$$\mu = \bar{\mu} + \mu(\{0\})\delta_0 \tag{50}$$

so that $\bar{\mu}$ turns out to be a finite regular measure on $[-d, 0]$, such that $\bar{\mu}(\{0\}) = 0$. By Lemma 2.5 there exists a sequence $(\bar{\mu}^n)_{n \geq 1}$ of finite regular measures on $[-d, 0]$, absolutely continuous with respect to $\lambda_{[-d, 0]}$, the Lebesgue measure on $[-d, 0]$, such that

$$\bar{\mu} = \lim_{n \rightarrow \infty} \bar{\mu}^n. \tag{51}$$

So following (49), the variation of the final cost can be written as

$$\begin{aligned} I_2 &= \lim_{n \rightarrow \infty} \mathbb{E} \int_0^1 \int_{-d}^0 (\rho y_T(\theta) + R^\rho(T + \theta)) \bar{h}_x(\bar{x}_T + \lambda(x_T^\rho - \bar{x}_T)) d\bar{\mu}_n(\theta) d\lambda \tag{52} \\ &\quad + \mathbb{E} \int_0^1 (\rho y_T(\theta) + R^\rho(T + \theta)) \bar{h}_x(\bar{x}_T + \lambda(x_T^\rho - \bar{x}_T)) \mu(\{0\}) d\lambda \end{aligned}$$

So taking into account the computation for J_1, J_2 and for I_2 that we have performed in (47), (48), (49) and (52), also by dividing both sides of (45) by ρ , and then by letting $\rho \rightarrow 0$ on the right hand side, we get

$$\begin{aligned} 0 &\leq \mathbb{E} \int_0^T \int_{-d}^0 y_t(\theta) \bar{l}_x(t, \bar{x}_t, \bar{u}_t) \mu_5(d\theta) dt + \mathbb{E} \int_0^T \int_{-d}^0 \bar{l}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}_t(\theta) \mu_6(d\theta) dt \tag{53} \\ &\quad + \mu_h(\{0\}) \bar{h}_x(\bar{x}_T) y_T(0) + \lim_{n \rightarrow \infty} \mathbb{E} \int_{-d}^0 y_T(\theta) \bar{h}_x(\bar{x}_T) d\bar{\mu}_n(\theta). \end{aligned}$$

Let J^n the cost obtained from J defined in (36) by replacing $\bar{\mu}$ with its absolute continuous approximation $\bar{\mu}_n$ in the final cost. So the variation δJ^n of J^n is given by

$$\begin{aligned} \delta J^n &:= J^n(u^\rho(\cdot)) - J^n(\bar{u}(\cdot)) = \mathbb{E} \int_0^T \int_{-d}^0 y(t + \theta) \bar{l}_x(t, \bar{x}_t, \bar{u}_t) \mu_5(d\theta) dt \tag{54} \\ &\quad + \mathbb{E} \int_0^T \bar{l}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}_t(\theta) \mu_6(d\theta) dt \\ &\quad + \mu(\{0\}) \bar{h}_x(\bar{x}_T) y_T(0) + \mathbb{E} \int_{-d}^0 y_T(\theta) \bar{h}_x(\bar{x}_T) d\bar{\mu}^n(\theta). \end{aligned}$$

We notice that in the first term, $y_t(\theta) = y(t + \theta) = 0$ if $t + \theta < 0$, and the same holds for \bar{v} .

Since (42) does not make sense in differential form and we cannot apply the Ito formula, we now introduce an approximated version of equation (42), along the lines

we have described at the beginning of the proof.

First we notice that with μ decomposed into $\bar{\mu}$ and $\mu(0)$ as in (50), equation (42) can be rewritten as

$$\left\{ \begin{array}{l} p(t) = \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 \bar{l}_x(s-\theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_5(d\theta) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 p(s-\theta) \bar{f}_x(s-\theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_1(d\theta) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q(s-\theta) \bar{g}_x(s-\theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_2(d\theta) ds \\ \quad + \int_t^T q(s) dW_s + \int_{t \vee (T-d)}^T \bar{h}_x(\bar{x}_T) \bar{\mu}(d\theta) + \mu(\{0\}) \bar{h}_x(\bar{x}_T) \\ p(T-\theta), q(T-\theta) = 0 \quad \forall \theta \in [-d, 0). \end{array} \right. \quad (55)$$

Now we approximate $\bar{\mu}$ by $\bar{\mu}^n$ as in (51) in the ABSDE (55), and so we obtain an approximated version of (55) given by

$$\left\{ \begin{array}{l} p^n(t) = \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 \bar{l}_x(s-\theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_5(d\theta) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 p^n(s-\theta) \bar{f}_x(s-\theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_1(d\theta) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q^n(s-\theta) \bar{g}_x(s-\theta, \bar{x}_{s-\theta}, \bar{u}_{s-\theta}) \mu_2(d\theta) ds \\ \quad + \int_t^T q^n(s) dW_s + \int_{t \vee (T-d)}^T \bar{h}_x(\bar{x}_T) \bar{\mu}^n(d\theta) + \mu(\{0\}) \bar{h}_x(\bar{x}_T) \\ p^n(T-\theta), q^n(T-\theta) = 0 \quad \forall \theta \in [-d, 0). \end{array} \right. \quad (56)$$

Since the differential form of $p^n(t)$ makes sense, we can compute $d\langle y(t), p^n(t) \rangle$:

$$\begin{aligned} & d\langle y(t), p^n(t) \rangle \\ &= \langle dy(t), p^n(t) \rangle + \langle y(t), dp^n(t) \rangle + \left\langle \int_{-d}^0 \bar{g}_x(t, \bar{x}_t, \bar{u}_t) \mu_2(d\theta), q^n(t) \right\rangle dt \\ &+ \left\langle \int_{-d}^0 \bar{g}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t+\theta) \mu_4(d\theta), q^n(t) \right\rangle dt \\ &= \left\langle \int_{-d}^0 y(t+\theta) \bar{f}_x(t, \bar{x}_t, \bar{u}_t) d\mu_1(\theta) dt + \int_{-d}^0 f_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t+\theta) \mu_6(d\theta) dt \right. \\ &+ \int_{-d}^0 y(t+\theta) \bar{g}_x(t, \bar{x}_t, \bar{u}_t) d\mu_2(\theta) dW(t) \\ &+ \int_{-d}^0 \bar{g}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t+\theta) \mu_4(d\theta) dW(t), p^n(t) \rangle \\ &- \langle y(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{l}_x(t-\theta, \bar{x}_{t-\theta}, \bar{u}(t-\theta)) \mu_5(d\theta) \rangle dt \\ &- \langle y(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 p^n(t-\theta) \bar{f}_x(t-\theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) \mu_1(d\theta) \rangle dt \\ &- \langle y(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 q^n(t-\theta) \bar{g}_x(t-\theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) \mu_2(d\theta) \rangle dt \end{aligned}$$

$$\begin{aligned}
 & - \langle y(t), q(t)dW_t \rangle + \langle y(t), \chi_{t>T-d} \bar{h}_x(\bar{x}_T) \frac{d\bar{\mu}^n}{dt} dt \rangle \\
 & - \langle \int_{-d}^0 \bar{g}_x(t, \bar{x}_t, \bar{u}_t) \mu_2(d\theta), q^n(t) \rangle dt.
 \end{aligned}$$

Integrating between 0 and T and taking expectation we obtain

$$\begin{aligned}
 \mathbb{E} \langle y(T), \mu_h(\{0\}) \bar{h}_x(x_T) \rangle &= \mathbb{E} \int_0^T \langle \int_{-d}^0 y(t+\theta) \bar{f}_x(t, x_t, u_t) d\mu_1(\theta) \\
 &+ \int_{-d}^0 f_u(t, x_t, u_t) \bar{v}(t+\theta) \mu_6(d\theta), p^n(t) \rangle dt \\
 &- \mathbb{E} \int_0^T \langle y(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \chi_{t-\theta < T} \bar{l}_x(t-\theta, x_{t-\theta}, u_{t-\theta}) \mu_5(d\theta) \rangle dt \\
 &- \mathbb{E} \int_0^T \langle y(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 p^n(t-\theta) \chi_{t-\theta < T} \bar{f}_x(t-\theta, x_{t-\theta}, u_{t-\theta}) \mu_1(d\theta) \rangle dt \\
 &- \mathbb{E} \int_0^T \langle y(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 q^n(t-\theta) \chi_{t-\theta < T} \bar{g}_x(t-\theta, x_{t-\theta}, u_{t-\theta}) \mu_2(d\theta) \rangle dt \\
 &- \mathbb{E} \int_0^T \langle y(t), \chi_{t>T-d} \bar{h}_x(x_T) \tilde{\mu}^n(t) \rangle dt \\
 &- \mathbb{E} \int_0^T \langle \int_{-d}^0 \bar{g}_x(t, x_t, u_t) d\mu_2(\theta), q^n(t) \rangle dt
 \end{aligned}$$

where

$$\tilde{\mu}^n = \frac{d\bar{\mu}^n}{dt},$$

is the Radon Nikodym derivative of $\bar{\mu}^n$ with respect to the Lebesgue measure. By some change in the time variable and with the optimal pair (\bar{x}, \bar{u}) instead of (x, u) , it turns out that

$$\begin{aligned}
 \delta J^n &= \mathbb{E} \langle y(T), \mu(\{0\}) \bar{h}_x(\bar{x}_T) \rangle + \mathbb{E} \int_0^T \langle y(t), \chi_{t>T-d} \bar{h}_x(\bar{x}_T) \tilde{\mu}^n(t) \rangle dt \\
 &+ \mathbb{E} \int_0^T \langle y(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \chi_{t-\theta < T} \bar{l}_x(t-\theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) \mu_5(d\theta) \rangle dt \\
 &+ \mathbb{E} \int_0^T \langle l_u(t, \bar{x}_t, \bar{u}_t) \int_{-d}^0 \bar{v}(t+\theta) \mu_6(d\theta) \rangle dt \\
 &= \mathbb{E} \int_0^T \langle \int_{-d}^0 \bar{f}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t+\theta) \mu_3(d\theta), p^n(t) \rangle dt \\
 &+ \mathbb{E} \int_0^T \langle \int_{-d}^0 \bar{g}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t+\theta) \mu_4(d\theta), q^n(t) \rangle dt \\
 &+ \mathbb{E} \int_0^T \int_{-d}^0 l_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t+\theta) \mu_6(d\theta) dt.
 \end{aligned}$$

So, taking into account (53)

$$0 \leq \mathbb{E} \int_0^T \langle \int_{-d}^0 \bar{f}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t+\theta) \mu_3(d\theta), p^n(t) \rangle dt$$

$$\begin{aligned}
 &+ \mathbb{E} \int_0^T \left\langle \int_{-d}^0 \bar{g}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t + \theta) \mu_4(d\theta), q^n(t) \right\rangle dt \\
 &+ \mathbb{E} \int_0^T \int_{-d}^0 l_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t + \theta) \mu_6(d\theta) dt
 \end{aligned}$$

and letting $n \rightarrow \infty$ we get

$$\begin{aligned}
 0 &\leq \mathbb{E} \int_0^T \left\langle \int_{-d}^0 \bar{f}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}_t(\theta) \mu_3(d\theta), p(t) \right\rangle dt & (57) \\
 &+ \mathbb{E} \int_0^T \left\langle \int_{-d}^0 \bar{g}_u(t, \bar{x}_t, \bar{u}_t) \bar{v}_t(\theta) \mu_4(d\theta), q(t) \right\rangle dt \\
 &+ \mathbb{E} \int_0^T \int_{-d}^0 l_u(t, \bar{x}_t, \bar{u}_t) \bar{v}(t + \theta) \mu_6(d\theta) dt \\
 &= \mathbb{E} \int_0^T \langle \bar{v}(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{f}_u(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) p(t - \theta) \mu_3(d\theta) \rangle dt \\
 &+ \mathbb{E} \int_0^T \langle \bar{v}(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 \bar{g}_u(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) q(t - \theta) \mu_4(d\theta) \rangle dt \\
 &+ \mathbb{E} \int_0^T \langle \bar{v}(t), \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 l_u(t - \theta, \bar{x}_{t-\theta}, \bar{u}_{t-\theta}) \bar{v}(t) \mu_6(d\theta) \rangle dt
 \end{aligned}$$

which is nothing else than (43) in integral form. The conclusion follows by a standard localization procedure, along the lines given e.g. in [20], end of paragraph 5.4, see also [17], end of the proof of Theorem 5.1, and [24].

4. Delay equations arising in advertising models. We consider a stochastic dynamic model in marketing for problems of optimal advertising. We study, as done in [10] and in [11], stochastic models for optimal advertising starting from the stochastic variant introduced in [13], and also with delay both in the state and in the control, see also [15]. In this model delay in the control corresponds to lags in the effect of advertisement.

So we consider, for $t \in [0, T]$, the following controlled stochastic differential equation in \mathbb{R} with delay in the state and in the control:

$$\begin{cases}
 dy(t) = \left[a_0 y(t) + \int_{-d}^0 y(t + \theta) \mu_a(d\theta) + b_0 u(t) + \int_{-d}^0 u(t + \theta) \mu_b(d\theta) \right] dt \\
 \quad + \sigma_a y(t) dW_t + \sigma_b u(t) dW_t, \\
 y(\theta) = y_0(\theta), \quad \theta \in [-d, 0), \\
 u(\theta) = u_0(\theta), \quad \theta \in [-d, 0).
 \end{cases} \tag{58}$$

In equation (58), y represents the goodwill level, a_0 is a constant factor of image deterioration in absence of advertising, b_0 is a constant representing an advertising effectiveness factor, $\mu_a(\cdot)$ is the distribution of the forgetting time, and $\mu_b(\cdot)$ is the distribution of the time lag between the advertising expenditure u and the corresponding effect on the goodwill level. The diffusion term $\sigma_a y(t)$ accounts for the word of mouth communication, the parameter σ_a is the advertising volatility; the diffusion term, $\sigma_b u(t)$ accounts for the effect of advertising, the parameter σ_b is the communication effectiveness volatility. Moreover, $y_0(0)$ is the level of goodwill at the beginning of the advertising campaign, while $y_0(\cdot)$ is the history of the goodwill level before the initial time, and $u_0(\cdot)$ is the history of the advertising expenditure

before the initial time, too.

We assume the following:

- Hypothesis 4.1.** (i) W is a standard Brownian motion in \mathbb{R} , and $(\mathcal{F}_t)_{t \geq 0}$ is the augmented filtration generated by W ;
 (ii) $a_0, \sigma_a, \sigma_b \in \mathbb{R}$;
 (iii) the control strategy u belongs to \mathcal{U} where

$$\mathcal{U} := \{z \in \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}) : u(t) \in U \text{ a.s.}\}$$

where U is a convex subset of \mathbb{R} ;

- (iv) $d > 0$ is the maximum delay the control takes to affect the system;
 (v) μ_a, μ_b are finite regular measures in $[-d, 0]$ that describe the time that respectively the state and the control take to affect the system.

The objective is to minimize, over all controls in \mathcal{U} , the following finite horizon cost:

$$J(t, x, u) = \mathbb{E} \int_t^T \ell(s, y(s), u(s)) ds + \mathbb{E}\phi(y_T), \tag{59}$$

where ℓ represents the cost of advertisement, and $-\phi$ represents the final utility, that may depend on the trajectory $y_T = y(T + \theta)$, $\theta \in [-d, 0]$. We assume that $\ell : [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is continuous, bounded and differentiable with respect to x and u , moreover the derivatives with respect to x and u satisfy

$$|\ell_x(t, y, u)| + |\ell_u l(t, y, u)| \leq C_3(1 + |y| + |u|),$$

and ϕ is given by

$$\phi(y_T) = \bar{\phi} \left(\int_{-d}^0 y(T + \theta) \mu_{\phi}(d\theta) \right), \tag{60}$$

where $\bar{\phi} : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous and differentiable and μ_{ϕ} is another finite regular measure on $[-d, 0]$.

We consider the adjoint equation for the pair of processes $(p, q) \in \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}) \times \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R})$

$$\left\{ \begin{array}{l} p(t) = \int_t^T \ell_x(s, y(s), u(s)) ds + \int_t^T a_0 p(s) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 p(s - \theta) \mu_a(d\theta) ds + \int_t^T \sigma_a q(s) ds + \int_t^T q_s dW_s \\ \quad + \int_{t \vee (T-d)}^T \bar{\phi}_x(y_T) \mu_{\phi}(d\theta) \\ p(T - \theta) = 0, q(T - \theta) = 0 \quad \forall \theta \in [-d, 0]. \end{array} \right. \tag{61}$$

Theorem 4.2. *Let Hypothesis 4.1 hold true. Let (p, q) be the unique solution of the ABSDE (61). Let (\bar{y}, \bar{u}) be an optimal pair for the optimal control problem of minimizing the cost functional (59) related to the controlled state equation (58). Let v be another admissible control, set $\bar{v} = v - \bar{u}$ and $u^{\rho} = \bar{u} + \rho \bar{v}$, then*

$$\begin{aligned} & b_0 (\bar{u}(t) - u^{\rho}(t)) p(t) + (\bar{u}(t) - u^{\rho}(t)) \mathbb{E}^{\mathcal{F}_t} \int_{-d}^0 p(t - \theta) \mu_b(d\theta) \\ & + \sigma_b (\bar{u}(t) - u^{\rho}(t)) q(t) + \ell_u(t, \bar{y}(t), \bar{u}(t)) (\bar{u}(t) - u^{\rho}(t)) \leq 0 \quad dt \times \mathbb{P} \text{ a.s.} \end{aligned} \tag{62}$$

5. An optimal portfolio problem with execution delay. We consider a generalized Black and Scholes market with one risky asset, whose price at time t is denoted by $S(t)$ and whose past trajectory from time $t - d$ up to time t is denoted by S_t , and one non-risky asset, whose price at time t is denoted by $B(t)$. The result can be extended to the case of a Black and Scholes market with j risky assets, whose prices at time t are denoted by $S^i(t), i = 1, \dots, j$, and one non-risky asset: for the sake of simplicity we limit here to the case of only one risky asset.

The evolution of the prices is given by the following stochastic delay differential equation in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

$$\begin{cases} dS(t) = S(t) [b(t, S_t)dt + \sigma(t, S_t)dW_t], & S(\theta) = \nu_0(\theta), \\ dB(t) = r(t, S_t)B(t)dt, \\ B(0) = B_0 \end{cases} \quad (63)$$

where $W(t)$ is a standard Brownian motion in \mathbb{R} , $(\mathcal{F}_t)_{t \geq 0}$ is the filtration generated by W and augmented with null probability sets and $S_t(\theta) = S(t + \theta)$, $\theta \in [-d, 0]$. The drift b , the diffusion σ and the rate r are given by

$$\begin{aligned} b(t, S_t) &= \bar{b}(t, \int_{-d}^0 S(t + \theta)\mu_{\bar{b}}(d\theta)), & \sigma(t, S_t) &= \bar{\sigma}(t, \int_{-d}^0 S(t + \theta)\mu_{\bar{\sigma}}(d\theta)), \\ r(t, S_t) &= \bar{r}(t, \int_{-d}^0 S(t + \theta)\mu_{\bar{r}}(d\theta)) \end{aligned} \quad (64)$$

where $\mu_{\bar{b}}, \mu_{\bar{\sigma}}, \mu_{\bar{r}}$ are finite regular measures on $[-d, 0]$.

Hypothesis 5.1. On $\bar{b}, \bar{\sigma}$ and \bar{r} we make the following assumptions:

- i) $\mu_{\bar{b}}$ is a regular measure and $\bar{b} : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is measurable. Moreover $\forall s_i \in \mathbb{R}, i = 1, 2$

$$|\bar{b}(t, s_1) - \bar{b}(t, s_2)| \leq c|s_1 - s_2|$$

for some $c > 0$ and for all $t \in [0, T]$ $\bar{b}(t, \cdot)$ is differentiable;

- ii) $\mu_{\bar{\sigma}}$ is a regular measure and $\bar{\sigma} : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is measurable. Moreover $\forall s_i \in \mathbb{R}, i = 1, 2$

$$|\bar{\sigma}(t, s_1) - \bar{\sigma}(t, s_2)| \leq c|s_1 - s_2|$$

for some $c > 0$ and for all $t \in [0, T]$ $\bar{\sigma}(t, \cdot)$ is differentiable;

- iii) $\bar{r} : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is measurable. Moreover $\forall s_i \in \mathbb{R}, i = 1, 2$

$$|r(t, s_1) - r(t, s_2)| \leq c|s_1 - s_2|$$

for some $c > 0$ and for all $t \in [0, T]$ $\bar{r}(t, \cdot)$ is differentiable.

We now consider the evolution of $V(t)$, the value at time t of the associated self-financing portfolio. We consider an optimal portfolio problem with execution delay, which is inspired by the models studied, in a different context, in [2] in a stochastic impulse control framework, and which is treated also in [8]. Let $d > 0$ be a fixed execution delay time: at time $t > 0$ the investor chooses, on the basis of the information contained in \mathcal{F}_t , to allocate the amount of money $u(t) > 0$ of its portfolio in the risky asset. This is the control process. However, due to the execution delay this order will be executed at time $t + d$ when the price of the risky asset has changed, see [2] for the formulation of this problem in a stochastic impulse control framework. Moreover we allow consumption, and also the investors are allowed to take money from the portfolio V : in the model this is represented by

a further control c .

The state equation for the optimal portfolio with execution delay is similar to the one considered in [7] in the case without delay, see also [8], and it is given by

$$\begin{cases} dV(t) = r(t, S_t)(V(t) - \pi^*(t - d))dt - c(t)dt + \pi^*(t - d) [b(t, S_t)dt \\ \quad + \sigma(t, S_t)dW_t] \\ V(\theta) = \eta(\theta), \pi(\theta) = \pi_0(\theta), \theta \in [-d, 0]. \end{cases} \tag{65}$$

Here we will only consider square-integrable, predictable investment strategies $\pi \in \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R})$.

The aim is to maximize the utility functional over the set of the admissible strategies

$$U(c) = \mathbb{E} \int_0^T [U_1(t, c(t))] dt + \mathbb{E} \left[U_2 \left(\int_{-d}^0 V(T + \theta) \mu_U(d\theta) \right) \right] = \mathbb{E} [U(V_T)], \tag{66}$$

where $U_1 : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ and $U_2 : \mathbb{R} \rightarrow \mathbb{R}$ are given utility functions, U_1 represents the utility from consumption and it is assumed to be continuous, differentiable in the second variable and the derivative with respect to c has linear growth in c , and U_2 represents the utility from the wealth on $[T - d, T]$ and it is assumed to be Lipschitz continuous and differentiable. Here μ_U is another finite regular measure on $[-d, 0]$: the utility is related not only to the final value T , but to the value of the portfolio in the window $[T - d, T]$, and so it depends on $V(T + \theta)$, $\theta \in [-d, 0]$.

At any time $t \in [-d, T]$, the state $X(t) \in \mathbb{R}^2$ is given by the pair

$$X(t) = \begin{pmatrix} S(t) \\ V(t) \end{pmatrix}.$$

So the equation for X is given by

$$\begin{cases} d \begin{pmatrix} S(t) \\ V(t) \end{pmatrix} = \begin{pmatrix} S(t)b(t, S_t) \\ r(t, S_t)(V(t) - \pi(t - d)) - c(t) + \pi(t - d)b(t, S_t) \end{pmatrix} dt \\ \quad + \begin{pmatrix} S(t)\sigma(t, S_t) \\ \pi(t - d)\sigma(t, S_t) \end{pmatrix} dW_t \\ \begin{pmatrix} S(\theta) \\ V(\theta) \end{pmatrix} = \begin{pmatrix} \nu_0(\theta) \\ \eta(\theta) \end{pmatrix} \end{cases} \tag{67}$$

and it turns out to be an equation with delay both in the state and in the control.

Notice that the adjoint processes are given by a pair of processes

$$(p, q) = \left(\begin{pmatrix} p^1 \\ p^2 \end{pmatrix}, \begin{pmatrix} q^1 \\ q^2 \end{pmatrix} \right) \in \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}^2) \times \mathcal{L}^2_{\mathcal{F}}(\Omega \times [0, T], \mathbb{R}^2)$$

solution of the ABSDEs we are going to write, and that it turns out that the pair (p^1, q^1) is identically 0. Indeed

$$\begin{cases} p^1(t) = \int_t^T p^1(s)b(s, S_s) ds + \int_t^T q^1(s)\sigma(s, S_s) ds + \int_t^T q^1(s)dW_s \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_0^0 p^1(s - \theta)S(s - \theta)\bar{b}_x(s - \theta, S_{s-\theta})\mu_{\bar{b}}(d\theta) ds \\ \quad + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q^1(s - \theta)S(s - \theta)\bar{\sigma}_x(s - \theta, S_{s-\theta})\mu_{\bar{\sigma}}(d\theta) ds \\ p(T - \theta) = 0, q(T - \theta) = 0 \quad \forall \theta \in [-d, 0]. \end{cases} \tag{68}$$

The pair of processes $(p^2, q^2) \in \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T], \mathbb{R}) \times \mathcal{L}_{\mathcal{F}}^2(\Omega \times [0, T], \mathbb{R})$ satisfies the following equation:

$$\left\{ \begin{array}{l} p^2(t) = \\ \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 p^2(s-\theta) (V(s-\theta) - \pi(s-\theta-d)) \bar{r}_x(s-\theta, S_{s-\theta}) \mu_{\bar{r}}(d\theta) ds \\ + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 p^2(s-\theta) \pi(s-\theta-d) \bar{b}_x(s-\theta, S_{s-\theta}) \mu_{\bar{b}}(d\theta) ds \\ + \int_t^T \mathbb{E}^{\mathcal{F}_s} \int_{-d}^0 q^2(s-\theta) \pi(s-\theta-d) \bar{\sigma}_x(s-\theta, S_{s-\theta}) \mu_{\bar{\sigma}}(d\theta) ds \\ + \int_t^T q^2(s) dW_s + \int_{t \vee (T-d)}^T U_x(V_T) \mu_U(d\theta) \\ p^2(T-\theta) = 0, q^2(T-\theta) = 0 \quad \forall \theta \in [-d, 0). \end{array} \right. \quad (69)$$

From the maximum principle stated in Theorem 3.2 we deduce the following condition on the optimal strategy for the present problem: notice that the optimality condition can be given only in terms of the pair of processes (p^2, q^2) .

Theorem 5.2. *Let Hypothesis 3.1 holds true. Let (p^2, q^2) be the unique solution of the ABSDE (69) Let $(\bar{X}, \bar{\pi}, \bar{c})$ be an optimal pair for the optimal control problem of minimizing $-U$, where U is defined in (66). For every admissible control (π_1, c_1) set $\bar{\pi}_1 = \pi_1 - \bar{\pi}$, $\bar{c}_1 = c_1 - \bar{c}$ and $\pi^\rho = \bar{\pi} + \rho \bar{\pi}_1$, $c^\rho = \bar{c} + \rho \bar{c}_1$, then*

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_t} [& (r(t+d, \bar{S}_{t+d}) (\pi^\rho(t) - \bar{\pi}(t)) + c^\rho(t+d) - \bar{c}(t+d) \\ & + (\bar{\pi}(t) - \pi^\rho(t)) b(t+d, \bar{S}_{t+d})) p^2(t+d) \\ & + (\bar{\pi}(t) - \pi^\rho(t)) \sigma(t+d, \bar{S}_{t+d}) q^2(t+d)] + (c^\rho(t) - \bar{c}(t)) (U_1)_c(t, \bar{c}(t)) \leq 0 \end{aligned}$$

$dt \times \mathbb{P}$ a.s., where we have set $\bar{X}(t) = \begin{pmatrix} \bar{S}(t) \\ \bar{V}(t) \end{pmatrix}$.

REFERENCES

- [1] E. Bandini, A. Cosso, M. Fuhrman and H. Pham, [Backward SDEs for optimal control of partially observed path-dependent stochastic systems: A control randomization approach](#), *Ann. Appl. Probab.*, **28** (2018), 1634–1678.
- [2] B. Bruder and H. Pham, [Impulse control problem on finite horizon with execution delay](#), *Stochastic Process. Appl.*, **119** (2009), 1436–1469.
- [3] R. Buckdahn, H.-J. Engelbert and A. Răşcanu, [On weak solutions of backward stochastic differential equations](#), *Teor. Veroyatn. Primen.*, **49** (2004), 70–108.
- [4] L. Chen, Z. Wu and Z. Yu, [Maximum principle for the stochastic optimal control problem with delay and application](#), *Automatica J. IFAC*, **46** (2010), 1074–1080.
- [5] L. Chen and Z. Wu, [Delayed stochastic linear-quadratic control problem and related applications](#), *Journal of Applied Mathematics*, **2012** (2012), 835319, 22 pp.
- [6] Y. Eidelman, V. Milman and A. Tsolomitis, *Functional Analysis, An Introduction, Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI, 2004.
- [7] N. El Karoui, S. Peng and M. C. Quenez, [Backward stochastic differential equations in finance](#), *Math. Finance*, **7** (1997), 1–71.
- [8] G. Fabbri and S. Federico, [On the infinite-dimensional representation of stochastic controlled systems with delayed control in the diffusion term](#), *Mathematical Economics Letters*, **2** (2014), 33–44.
- [9] M. Fuhrman and G. Tessitore, [Nonlinear Kolmogorov equations in infinite dimensional spaces: The backward stochastic differential equations approach and applications to optimal control](#), *Ann. Probab.*, **30** (2002), 1397–1465.

- [10] F. Gozzi and C. Marinelli, Stochastic optimal control of delay equations arising in advertising models, *Stochastic partial differential equations and applications—VII, Lect. Notes Pure Appl. Math.*, Chapman & Hall/CRC, Boca Raton, FL, **245** (2006), 133–148.
- [11] F. Gozzi, C. Marinelli and S. Savin, On controlled linear diffusions with delay in a model of optimal advertising under uncertainty with memory effects, *J. Optim. Theory Appl.*, **142** (2009), 291–321.
- [12] F. Gozzi and F. Masiero, Stochastic optimal control with delay in the control II: Verification theorem and optimal feedbacks, *SIAM J. Control Optim.*, **55** (2017), 3013–3038.
- [13] L. Grosset and B. Viscolani, Advertising for a new product introduction: A stochastic approach, *Top*, **12** (2004), 149–167.
- [14] G. Guatteri, F. Masiero and C. Orrieri, Stochastic maximum principle for SPDEs with delay, *Stochastic Process. Appl.*, **127** (2017), 2396–2427.
- [15] R. F. Hartl, Optimal dynamic advertising policies for hereditary processes, *J. Optim. Theory Appl.*, **43** (1984), 51–72.
- [16] Y. Hu and S. Peng, Maximum principle for semilinear stochastic evolution control systems, *Stochastics Stochastics Rep.*, **33** (1990), 159–180.
- [17] Y. Hu and S. Peng, Maximum principle for optimal control of stochastic system of functional type, *Stochastic Anal. Appl.*, **14** (1996), 283–301.
- [18] S.-E. A. Mohammed, Stochastic differential systems with memory: Theory, examples and applications. Stochastic analysis and related topics VI, *Progr. Probab.*, Birkhäuser Boston, Boston, MA, **42** (1998), 1–77.
- [19] B. Øksendal, A. Sulem and T. Zhang, Optimal control of stochastic delay equations and time-advanced backward stochastic differential equations, *Adv. in Appl. Probab.*, **43** (2011), 572–596.
- [20] C. Orrieri, E. Rocca and L. Scarpa, Optimal control of stochastic phase-field models related to tumor growth, *ESAIM Control Optim. Calc. Var.*, forthcoming.
- [21] E. Pardoux and A. Răşcanu, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations, Stochastic Modelling and Applied Probability*, Springer, Cham, 2014.
- [22] S. Peng and Z. Yang, Anticipated backward stochastic differential equations, *Ann. Probab.*, **37** (2009), 877–902.
- [23] Z. Yang and R. J. Elliott, Some properties of generalized anticipated backward stochastic differential equations, *Electron. Commun. Probab.*, **18** (2013), 10.
- [24] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, New York, Springer-Verlag, 1999.

Received February 2020; 1st revision May 2020; final revision September 2020.

E-mail address: giuseppina.guatteri@polimi.it

E-mail address: federica.masiero@unimib.it