HAL
open science

On some systems controlled by the structure of their memory

Giuseppe Buttazzo, Guillaume Carlier, Rabah Tahraoui

To cite this version:

Giuseppe Buttazzo, Guillaume Carlier, Rabah Tahraoui. On some systems controlled by the structure of their memory. SIAM Journal on Control and Optimization, 2010, 48 (8), pp.5241-5253. 10.1137/100781821 . hal-00637401

HAL Id: hal-00637401

https://hal.science/hal-00637401v1

Submitted on 1 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On some systems controlled by the structure of their memory

G. Buttazzo [*], G. Carlier, R. Tahraoui [†]

July 31, 2010

Abstract

We consider an optimal control problem governed by an ODE with memory playing the role of a control. We show the existence of an optimal solution and derive some necessary optimality conditions. Some examples are then discussed.

Keywords: Optimal control problems, memory terms, Wasserstein distance.

2000 Mathematics Subject Classification: 49K25, 34K35, 49K22, 93C30.

1 Introduction

1.1 Motivation

Dynamics with lags or with more general memory structure (deviated arguments, integro-differential equations...) arise in many different settings in engineering, economics, ecology, biology, modelling of financial time series... It is typically the case when studying the optimal performances of a system in which the response to a given input occurs not instantaneously but only after a certain elapse of time. We refer for instance to the classical book of Bellman and Cooke [2] for a general overview of such functional equations.

[*]Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo, 5, 56127 Pisa, ITALY `buttazzo@dm.unipi.it`

[†]CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE `carlier@ceremade.dauphine.fr`, `tahraoui@ceremade.dauphine.fr`

To illustrate what follows, let us start with a simple differential equation with a deviated argument:

$$\dot{x}(t) = f(x(\theta(t)), t \in [0,1],\ x(0) = x_0 \tag{1.1}$$

where θ, the *deviation*, is a function that satisfies $0 \leq \theta(t) \leq t$ (nonanticipativity), for every time t. Dynamics of the form (1.1) are nonlocal and may arise in various applied settings such as population dynamics or biology (incubation of a disease, maturation phenomena...). We refer the reader to Kuang's book [13] and the references therein for an overview of dynamics with memory in the biosciences, another motivation for related problems in economics can be found for instance in [11]. In the literature on equations with deviated arguments, the deviation θ is usually taken as given. However, it is natural to consider concrete situations where the deviation is not perfectly known. For instance, if some-possibly noisy-observation x^* is available for the state variable, then the simplest way to *estimate* θ is by *least* squares, that is by minimizing

$$\int_0^1 |x(t) - x^*(t)|^2 dt$$

with respect to θ, x being linked to θ by (1.1). One can further add constraints on θ or add a penalization term of the form

$$\int_0^1 g(t, \theta(t)) dt$$

which represents a memory cost in the energy to be minimized. This leads to optimization problems of the form

$$\inf \int_0 j(t, x(t)) dt + \int_0^1 g(t, \theta(t)) dt$$

with respect to x and θ related by the state equation (1.1). We shall see in the sequel, that the previous problem is in fact ill-posed in general and therefore needs to be suitably relaxed. The relaxation of the problem leads to replace (1.1) by

$$\dot{x}(t) = \int_0^t f(x(s)) d\nu_t(s),\ x(0) = x_0 \tag{1.2}$$

where ν_t is a probability measure supported on $[0, t]$ (so that (1.1) corresponds to the case where $\nu_t = \delta_{\theta(t)}$ is a Dirac mass). Note that this formulation is very much in the spirit of relaxation by Young measures in the calculus variations and optimal control theory (see for instance [3], [15]).

Considering directly (i.e. without invoking relaxation) the state equation (1.2) controlled by ν_t is natural as well in population dynamics. Indeed, quoting Kuang's own words: *...more realistic models for single species growths should take into account both the changing of the environment and the effects of time delay*, which leads in [12] to dynamics of the form (1.2). Another motivation comes from *autoregressive models* where one seeks to estimate the dynamics of some state variable by regressing it on its past. In discrete time, an autoregressive process is a stochastic process that satisfies a relation of the form:

$$X_t = \sum_{k=0}^{K} \alpha_k f(t-k, X_{t-k}) + \varepsilon_t$$

where ε_t are independent and identically distributed random variables. Such processes thus have some memory (the range of the memory being the integer K, called the order of the process), and the fact that the previous dynamics is given by a convolution captures some stationarity of the memory structure. This is a particular case of

$$X_t = \sum_{s \leq k} \nu_t(s) f(s, X_s) + \varepsilon_t.$$

Passing to continuous time in the previous equation strongly suggests that estimating the memory structure of the process by least squares naturally leads to solve a control problem with a state equation of the form (1.2).

1.2 Control formulation

The present paper deals with the optimal control of equations of the form

$$\dot{x}(t) = \langle f(., x(.)), \nu_t \rangle = \int_0^t f(s, x(s))\, d\nu_t(s),\ t \in [0,1],\ x(0) = x_0, \tag{1.3}$$

where the control is a family of probability measures $t \mapsto \nu_t$ such that ν_t is supported on $[0,t]$ (nonanticipativity). The special case where $\nu_t = \delta_{\theta(t)}$ is a Dirac mass corresponds to the deviated equation

$$\dot{x}(t) = f\big(\theta(t), x(\theta(t))\big),\ x(0) = x_0, \tag{1.4}$$

where the control is the deviation θ satisfying $\theta(t) \leq t$ for all $t \in [0,1]$.

We will consider the minimization of the functional

$$\int_0^1 j(t, x(t))\, dt + h(x(1)) + \int_0^1 \Big(\int_0^1 g(t,s)\, d\nu_t(s) \Big)\, dt \tag{1.5}$$

where x is related to ν by the state equation (1.3). An example of relevant function g is $g(t,s) = \lambda|t-s|^p$ with $p \geq 1$ and $\lambda \geq 0$. In this case the last term in the previous functional is $\lambda \int_0^1 W_p^p(\delta_t, \nu_t)\, dt$ where $W_p(\delta_t, \nu_t)$ is the p-Wasserstein distance between ν_t and the Dirac measure at t

$$W_p^p(\delta_t, \nu_t) = \int_{[0,t]} |t-s|^p \, d\nu_t(s).$$

The interpretation of the Wasserstein term in the functional is therefore a penalization of long-term memory.

Let us insist here on the fact that the optimal control problem (1.3)-(1.5) in which the memory structure is unknown and acts as a control is, as far as we know, somehow unusual. For variational or optimal control problems where a deviation or a memory structure is given and one looks for a classical optimal control, we refer to [16], [4], [5] (necessary optimality conditions) or [6] (dynamic programming approach leading to a Hamilton-Jacobi equation in infinite dimensions).

The paper is organized as follows. Section 2 is devoted to some preliminary results. In section 3, we prove existence of solutions for (1.5) and show that (1.5) is the natural relaxation of the corresponding optimization problem posed over deviation functions. In section 4, we establish optimality conditions. Finally section 5 is devoted to some examples.

2 Preliminaries

Let $(\nu_t)_{t\in[0,1]}$ be a Borel family of probability measures (meaning that $t \mapsto \int g\, d\nu_t$ is Borel for every continuous g) such that $\nu_t([0,t]) = 1$ for every t; the short notation (ν_t) will be used in the following to denote the family $(\nu_t)_{t\in[0,1]}$. Let $f \in C^0([0,1] \times \mathbb{R}^d, \mathbb{R}^d)$ satisfy the Lipschitz condition that there is a $k \geq 0$ such that:

$$|f(t,x) - f(t,y)| \leq k|x-y|, \ \forall t \in [0,1], \ \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{2.1}$$

The next result gives the existence and uniqueness of a solution to the state equation (1.3):

Proposition 2.1. *Let $f \in C^0([0,1]\times\mathbb{R}^d, \mathbb{R}^d)$ satisfy the Lipschitz assumption (2.1), (ν_t) be as above and $x_0 \in \mathbb{R}^d$. Then the Cauchy problem*

$$\dot{x}(t) = \langle f(., x(.)), \nu_t \rangle \, , \ \forall t \in [0,1], \ x(0) = x_0 \tag{2.2}$$

admits a unique continuous solution x. Moreover x is Lipschitz continuous and satisfies $\|x\|_{W^{1,\infty}} \leq M$ for a constant M that only depends on k, $|x_0|$ and $\sup_{t\in[0,1]} |f(t,0)|$.

Proof. We proceed as in [5]. Let $\lambda > k$ and equip $C^0([0,1],\mathbb{R}^d)$ with the norm

$$\|x\|_\lambda := \sup_{t\in[0,1]} e^{-\lambda t}|x(t)|.$$

For every $x \in C^0([0,1],\mathbb{R}^d)$, define the continuous function Hx by

$$Hx(t) := x_0 + \int_0^t \langle f(.,x(.)),\nu_s\rangle\, ds \qquad \forall t\in[0,1].$$

For x and y continuous and $t\in[0,1]$, one deduces from (2.1) and the nonanticipativity of (ν_t):

$$|Hx(t)-Hy(t)| \le k\|x-y\|_\lambda \int_0^t\int_0^s e^{\lambda\tau}\, d\nu_s(\tau)\, ds \le \frac{k}{\lambda}\|x-y\|_\lambda e^{\lambda t}.$$

One then deduces that H is a contraction and therefore has a unique fixed point in C^0. The second claim easily follows. □

In what follows, $\mathcal{L}^1$ denotes the Lebesgue measure on $[0,1]$. Let (ν_t) be admissible for our problem which means that it is a Borel family of probability measures such that $\nu_t([0,t]) = 1$ for every t. It will be convenient in the sequel to define $\gamma := (\nu_t)\otimes\mathcal{L}^1$ that is the probability measure on $[0,1]^2$ defined by

$$\int_{[0,1]^2} \varphi(t,s)\, d\gamma(t,s) = \int_0^1 \Big(\int_0^1 \varphi(t,s)\, d\nu_t(s)\Big)\, dt \tag{2.3}$$

for every $\varphi\in C^0([0,1]^2,\mathbb{R})$. The admissibility of $(\nu_t)_t$ is equivalent to require that $\gamma := (\nu_t)\otimes\mathcal{L}^1$ belongs to the set

$$\Gamma := \{\gamma \text{ probability on } [0,1]^2 \ :\ \gamma(T)=1,\ \pi_{1\#}\gamma = \mathcal{L}^1\}. \tag{2.4}$$

where T is the triangle $T := \{(t,s)\in[0,1]^2 \ :\ s\le t\}$ and $\pi_{1\#}\gamma$ denotes the first marginal of γ. Given $\gamma\in\Gamma$, the disintegration Theorem (see [8]) enables one to disintegrate γ as $\gamma = (\nu_t)\otimes\mathcal{L}^1$ with (ν_t) admissible. In the sequel, under the assumptions of proposition 2.1, the solution of the Cauchy problem (2.2) will be denoted x_γ. Let us also remark that x_γ can be characterized by the weak form of (2.2) that can be conveniently written in terms of γ as:

$$\int_0^1 \dot\varphi(t)\cdot x_\gamma(t)\, dt = -\int_{[0,1]^2} \varphi(t)\cdot f(s,x_\gamma(s))\, d\gamma(t,s), \quad x_\gamma(0)=x_0 \tag{2.5}$$

for every $\varphi\in C^1_c((0,1),\mathbb{R}^d)$. The other obvious advantage of formulating the problem in terms of $\gamma\in\Gamma$ is that Γ is weakly* compact.

In the sequel, we shall always assume that $f\in C^0([0,1]\times\mathbb{R}^d,\mathbb{R}^d)$ satisfy the Lipschitz assumption (2.1) so that x_γ is well-defined for every $\gamma\in\Gamma$.

Lemma 2.2. *If $(\gamma_n) \in \Gamma^{\mathbb{N}}$ weakly* converges to γ, then x_{γ_n} converges to x_γ in $C^0([0,1],\mathbb{R}^d)$.*

Proof. Setting $x_n := x_{\gamma_n}$, it follows from proposition 2.1 and Ascoli-Arzelà's theorem that (x_n) is precompact in C^0. Let y be the uniform limit of some (not relabeled) subsequence. Let $\varphi \in C^1_c((0,1),\mathbb{R}^d)$, one has for every n

$$\int_0^1 \dot{\varphi}(t) \cdot x_n(t)\, dt = -\int_{[0,1]^2} \varphi(t) \cdot f(s, x_n(s))\, d\gamma_n(t,s),\ \ x_n(0) = x_0. \tag{2.6}$$

Since x_n converges uniformly to y, and γ_n converges weakly* to γ, passing to the limit in (2.6) one deduces from the continuity of f that $y = x_\gamma$, and by a standard compactness argument we deduce that the whole sequence converges to x_γ. □

3 The optimization problem

3.1 Existence of optimal controls and relaxation

We are now interested in the optimization problem

$$\inf_{\gamma \in \Gamma} J(\gamma) \tag{3.1}$$

where

$$J(\gamma) := \int_0^1 j(t, x_\gamma(t))\, dt + h(x_\gamma(1)) + \int_T g(t,s)\ d\gamma(t,s).$$

Theorem 3.1. *We make the following assumptions:*

- *the function $j : [0,1] \times \mathbb{R}^d \to [0,+\infty]$ is Borel measurable, nonnegative and $j(t,\cdot)$ is lower semicontinuous for a.e. $t \in [0,1]$;*
- *the function $h : \mathbb{R}^d \to [0,+\infty]$ is nonnegative and lower semicontinuous;*
- *the function $g : T \to [0,+\infty]$ is nonnegative and lower semicontinuous on T;*
- *there exists $\gamma_0 \in \Gamma$ such that $J(\gamma_0) < +\infty$.*

Then the optimal control problem (3.1) *has a finite value and admits a solution.*

Proof. Thanks to lemma 2.2, one easily deduces from the assumptions above that J is weakly* lower-semicontinuous (for the lower semicontinuity of the last term express g as the supremum of continuous functions) and the existence claim then follows from the weak* compactness of Γ. $\square$

Let us define the set of admissible deviations:

$$\Theta := \{\theta : [0,1] \to [0,1] \text{ Borel}, \ \theta(t) \leq t, \ \text{for a.e. } t \in [0,1]\}$$

and the optimal deviation problem

$$\inf_{\theta \in \Theta} F(\theta), \qquad \text{with } F(\theta) := J\big((\delta_{\theta(t)}) \otimes \mathcal{L}^1\big) \tag{3.2}$$

that is the restriction of (3.1) to deviation (or delay) functions. Then (3.1) is the relaxation of (3.2) in the following sense (see for instance [3, 15] for a general presentation of relaxation theory for variational problems).

Theorem 3.2. *If $j \in C^0([0,1] \times \mathbb{R}^d)$, $h \in C^0(\mathbb{R}^d)$ and $g \in C^0(T)$, then for every $\gamma \in \Gamma$, there exists a sequence θ_n in Θ such that $F(\theta_n)$ converges to $J(\gamma)$. In particular*

$$\min_{\gamma \in \Gamma} J(\gamma) = \inf_{\theta \in \Theta} F(\theta).$$

Proof. It follows from lemma 2.2 that J is weakly* continuous. Fix now $\gamma \in \Gamma$; it is a well-known result in the theory of Young measures (see for instance Theorem 9.3 in [1]) that there exists a sequence of Borel maps $\sigma_n : [0,1] \mapsto [0,1]$ such that $(\delta_{\sigma_n(t)}) \otimes \mathcal{L}^1$ converges weakly* to γ. Defining $\theta_n(t) := \min\{\sigma_n(t), t\}$ so that $\theta_n \in \Theta$, it is enough to prove that $(\delta_{\theta_n(t)}) \otimes \mathcal{L}^1$ converges weakly* to γ to conclude. Let $\varphi \in C^0([0,1]^2)$ and let us denote by ω_φ the modulus of continuity of φ. For every $\delta > 0$, we have

$$\Big| \int_0^1 \varphi(t, \theta_n(t)) - \varphi(t, \sigma_n(t))\, dt \Big| \leq 2\|\varphi\|_\infty \mathcal{L}^1(\{t \ : \ \sigma_n(t) \geq t + \delta\}) + \omega_\varphi(\delta).$$

From the weak* convergence of $(\delta_{\sigma_n(t)}) \otimes \mathcal{L}^1$ to γ and from the fact that the support of γ is included in the triangle T, it is easy to deduce that for every $\delta > 0$, $\mathcal{L}^1(\{t \ : \ \sigma_n(t) \geq t + \delta\})$ tends to 0. The weak* convergence of $(\delta_{\theta_n(t)}) \otimes \mathcal{L}^1$ to γ then follows directly. $\square$

It is easy to build from the previous result examples where there is no optimal delay function for (3.2). Indeed, let us take $d = 1$, $f(x) = x$, $x_0 = 1$, $h = 0$, $j(t, x) = |x - x^*(t)|^2$ with $x^*(t) = -1 + 2e^{t/2}$ and $g(t, s) = s(t - s)$. By construction $x^* = x_{\gamma^*}$ for $\gamma^* = \frac{1}{2}(\delta_0 + \delta_t) \otimes \mathcal{L}^1$ so that $J(\gamma^*) = 0$ and γ^* is optimal. If θ was an optimal delay for (3.2) then one should have $F(\theta) = 0$ so that $\theta(t) \in \{0, t\}$ a.e. and $x_\theta = x^*$ (where slightly abusing notations x_θ denotes $x_{(\delta_{\theta(t)})\otimes\mathcal{L}^1}$). From the state equation one should then also have

$$e^{t/2} = x^*(\theta(t)) = -1 + 2e^{\theta(t)/2}, \text{ a.e.}$$

which contradicts $\theta(t) \in \{0, t\}$ a.e..

3.2 Solutions with finitely many Dirac masses

A simple application of a refinement of the well-known Carathéodory's theorem (see [14]), attributed to Fenchel and Bunt in Theorem 1.3.7. of [10] (see [9] pages 40–41 for a complete proof), gives the existence of optimal controls involving only finitely many Dirac masses.

Proposition 3.3. *Assume that f satisfies (2.1) and f and g are continuous, then for every $\gamma = (\nu_t) \otimes \mathcal{L}^1 \in \Gamma$ there exists $\eta = (\mu_t) \otimes \mathcal{L}^1$ such that μ_t has finite support with at most cardinality $d + 1$ such that $x_\gamma = x_\eta$ and*

$$\int_0^1 g(t, s)\, d\nu_t(s) = \int_0^1 g(t, s)\, d\mu_t(s) \qquad a.e.\ t.$$

In particular, in the optimization problem (3.1) it is enough to optimize over measures in Γ of the form $(\nu_t) \otimes \mathcal{L}^1$ with ν_t a convex combination of at most $d + 1$ (d if $g = 0$) Dirac masses.

Proof. Let $\gamma \in \Gamma$, $x := x_\gamma$ and define for every $t \in [0, 1]$ the compact set

$$S_{x,t} := \big\{(f(s, x(s)), g(t, s)) \ : \ s \in [0, t]\big\} \subset \mathbb{R}^{d+1}.$$

By construction, for almost every t one has

$$\Big(\dot{x}(t), \int_0^t g(t, s)\, d\nu_t(s)\Big) \in \mathrm{co}(S_{x,t}).$$

Since $S_{x,t}$ is a connected set, it follows from the Fenchel-Bunt theorem (see [10] and [9]) that $\big(\dot{x}(t), \int_0^t g(t, s)\, d\nu_t(s)\big)$ may be expressed as a convex combination of at most $d+1$ points in $S_{x,t}$. Hence there exists a discrete probability μ_t on $[0, t]$ with at most $d + 1$ points in its support such that

$$\dot{x}(t) = \langle f(., x(.)), \mu_t \rangle\,, \qquad \int_0^1 g(t, s)\, d\nu_t(s) = \int_0^1 g(t, s)\, d\mu_t(s).$$

The fact that the discrete measure μ_t can be chosen measurable in t follows from standard measurable selection arguments (see for instance [7]). $\square$

4 Necessary conditions of optimality

In this section, we look for optimality conditions for (3.1). In what follows, we further assume that j, h and g are continuous and that

- h is of class C^1,
- $j(t,.)$ and $f(t,.)$ are differentiable for every $t \in [0,1]$ and $\nabla_x j$ and $\nabla_x f$ are continuous on $[0,1] \times \mathbb{R}^d$.

Let $\gamma = (\nu_t) \otimes \mathcal{L}^1$ be a solution to (3.1), $\eta = (\mu_t) \otimes \mathcal{L}^1 \in \Gamma$ and $\varepsilon \in (0,1)$; then

$$\frac{1}{\varepsilon}[J(\gamma + \varepsilon(\eta - \gamma)) - J(\gamma)] \geq 0. \tag{4.1}$$

As usual, our aim is to pass to the limit as $\varepsilon \to 0^+$ and to express the optimality conditions obtained this way in the form of some maximum principle, which will be achieved by introducing some suitable adjoint variable. To shorten notation, we set $\gamma_\varepsilon = \gamma + \varepsilon(\eta - \gamma)$, $x = x_\gamma$ and $x_\varepsilon = x_{\gamma_\varepsilon}$.

Lemma 4.1. *As $\varepsilon \to 0^+$, $z_\varepsilon := \varepsilon^{-1}(x_\varepsilon - x)$ converges uniformly on $[0,1]$ to z, the solution of the linearized system:*

$$\dot{z}(t) = \langle Az, \nu_t \rangle + \langle a, \mu_t - \nu_t \rangle, \qquad t \in [0,1],\ z(0) = 0 \tag{4.2}$$

where A and a are the continuous functions:

$$A(s) := \nabla_x f(s, x(s)), \quad a(s) := f(s, x(s)), \quad \forall s \in [0,1]. \tag{4.3}$$

Proof. First, let us remark that by the same arguments used in proposition 2.1, equation (4.2) posseses a unique (Lipschitz) solution z. By construction we have $z_\varepsilon(0) = 0$ and

$$\dot{z}_\varepsilon(t) = \left\langle \frac{1}{\varepsilon}[f(., x + \varepsilon z_\varepsilon) - f(., x)], \nu_t \right\rangle + \langle f(., x_\varepsilon), \mu_t - \nu_t \rangle \tag{4.4}$$

from which, by the Lipschitz assumption on f, we easily deduce that z_ε is bounded in $W^{1,\infty}$ (in particular $\|x_\varepsilon - x\|_{W^{1,\infty}} = O(\varepsilon)$). By Ascoli-Arzela's theorem, z_ε then possesses a cluster point y in C^0. Passing to the limit in (4.4) along a convergent subsequence then easily yields $y = z$ and thus by compactness the whole family $(z_\varepsilon)_\varepsilon$ converges to z as $\varepsilon \to 0^+$. $\square$

Under our differentiability assumption, using lemma 4.1 and passing to the limit in (4.1) we then get:

$$\int_0^1 B \cdot z \, dt + b \cdot z(1) + \int_0^1 \langle g(t,.), \mu_t - \nu_t \rangle \, dt \geq 0. \tag{4.5}$$

where z is related to γ and η by the linearized equation (4.2) and

$$B(s) := \nabla_x j(s, x(s)) \quad \forall s \in [0,1], \qquad b := \nabla h(x(1)). \tag{4.6}$$

To make condition (4.5) tractable, we shall introduce, as usual in control theory, an adjoint state; to do so we shall need a few notations and preliminaries. Recall that $\gamma \in \Gamma$ is given by the disintegration $\gamma = (\nu_t) \otimes \mathcal{L}^1$ and let $\nu := \pi_{2\#}\gamma$ be the second marginal of γ defined by

$$\int_0^1 \varphi(s) \, d\nu(s) = \int_{[0,1]^2} \varphi(s) \, d\gamma(t,s) = \int_0^1 \langle \varphi, \nu_t \rangle \, dt, \qquad \forall \varphi \in C^0([0,1]).$$

Invoking again the disintegration theorem, γ admits the disintegration $\gamma = \nu \otimes (\nu_s^*)$ that is (ν_s^*) is a Borel family of probability measures and for every test-function $\varphi \in C^0([0,1]^2)$ one has

$$\int_{[0,1]^2} \varphi \, d\gamma = \int_0^1 \langle \varphi(.,s), \nu_s^* \rangle \, d\nu(s) = \int_0^1 \langle \varphi(t,.), \nu_t \rangle \, dt. \tag{4.7}$$

Note that the requirement that γ is supported by T implies that $\nu_s^*([s,1]) = 1$. Since γ has $\mathcal{L}^1$ as first marginal, (4.7) also holds for test functions of the form $(t,s) \mapsto q(t)\varphi(s)$ with φ continuous but q only L^1. This enables us, for $q \in L^1$, to define $\langle q, \nu_s^* \rangle \, \nu$ as the finite measure defined by

$$\langle \varphi, \langle q, \nu_s^* \rangle \, \nu \rangle := \int_0^1 q(t) \, \langle \varphi, \nu_t \rangle \, dt = \int_{[0,1]^2} q(t)\varphi(s) \, d\gamma(t,s) \quad \forall \varphi \in C^0([0,1]).$$

The adjoint of (4.2) will then be expressed using the measures ν_s^* and ν, as usual, it is an equation of backward type which is formally written as:

$$\dot{q} = B - A^T \langle q, \nu_s^* \rangle \, \nu, \qquad q(1) = -b \tag{4.8}$$

where the continuous functions B and A are defined by (4.6) and (4.3) respectively, A^T denotes the transpose of A and the measure $\langle q, \nu_s^* \rangle \, \nu$ is defined as above. Equation (4.8) has to be understood in an integral sense, a solution of (4.8) is then by definition an $L^1((0,1), \mathbb{R}^d)$ function q such that

$$q(t) = -b - \int_t^1 B \, ds + \int_{[t,1]} A^T(s) \, \langle q, \nu_s^* \rangle \, d\nu(s) \qquad \text{for a.e. } t \in [0,1]. \tag{4.9}$$

Such a solution belongs to $\mathrm{BV}((0,1),\mathbb{R}^d)$, the space of curves in $\mathbb{R}^d$ of bounded variation, and the weak formulation of (4.8) reads as

$$\int_0^1 q\cdot\dot\varphi\,dt = -\int_0^1 B\cdot\varphi\,dt - b\cdot\varphi(1) + \int_{[0,1]^2}\Big(A(s)\varphi(s)\cdot q(t)\Big)\,d\gamma(t,s) \quad (4.10)$$

for every $\varphi\in C^1([0,1]),\mathbb{R}^d)$ such that, $\varphi(0)=0$. Note that it is easy to see in this case that (4.10) also holds for every test-function, $\varphi\in W^{1,\infty}((0,1),\mathbb{R}^d)$ such that $\varphi(0)=0$. Since $q\in\mathrm{BV}((0,1),\mathbb{R}^d)$ the left-limit $q(1^-)$ is well defined, and the condition $q(1)=-b$ in (4.8) has to be intended as $q(1^-)=-b$; this convention for BV functions will be used in the rest of the paper. The well-posedness of (4.8) follows from the following result.

Lemma 4.2. *Let $B\in C^0([0,1],\mathbb{R}^d)$, $A\in C^0([0,1],\mathbb{R}^{d\times d})$ and $b\in\mathbb{R}^d$. Then (4.8) admits a unique solution $q\in\mathrm{BV}((0,1),\mathbb{R}^{\mathrm{d}})$. Moreover if z is the solution of (4.2) then*

$$\int_0^1 B\cdot z\,dt + b\cdot z(1) = -\int_0^1 q(t)\,\langle a,\mu_t-\nu_t\rangle\,dt. \quad (4.11)$$

Proof. Take $\lambda>\|A\|_\infty$ and equip $L^1((0,1),\mathbb{R}^d)$ with the norm

$$\|q\| := \int_0^1 e^{\lambda t}|q(t)|\,dt.$$

In view of (4.9), it is natural, for every $q\in L^1$ to define Kq by

$$Kq(t) := -b - \int_t^1 B\,ds + \int_{[t,1]} A^T(s)\,\langle q,\nu_s^*\rangle\,d\nu(s) \quad \text{for a.e. } t\in[0,1],$$

so that $Kq\in\mathrm{BV}((0,1),\mathbb{R}^d)$. Let q_1 and q_2 be in L^1 and $t\in(0,1)$, let $\varepsilon\in(0,t)$ and φ_ε be a continuous function such that $0\le\varphi_\varepsilon\le 1$, $\varphi_\varepsilon=1$ on $[t,1]$ and $\varphi_\varepsilon=0$ on $[0,t-\varepsilon]$. We then have:

$$\begin{aligned}
|Kq_1(t)-Kq_2(t)| &\le \|A\|_\infty\int_0^1 \varphi_\varepsilon\,\langle|q_1-q_2|,\nu_s^*\rangle\,d\nu\\
&= \|A\|_\infty\int_0^1 \langle\varphi_\varepsilon,\nu_\tau\rangle\,|q_1(\tau)-q_2(\tau)|\,d\tau\\
&\le \|A\|_\infty\int_0^1 \nu_\tau([t-\varepsilon,\tau])|q_1(\tau)-q_2(\tau)|\,d\tau\\
&\le \|A\|_\infty\int_{t-\varepsilon}^1 |q_1-q_2|\,d\tau.
\end{aligned}$$

Letting $\varepsilon \to 0^+$ we get

$$|Kq_1(t) - Kq_2(t)| \leq \|A\|_\infty \int_t^1 |q_1 - q_2| \, d\tau. \tag{4.12}$$

Multiplying (4.12) by $e^{\lambda t}$, integrating and using Fubini's theorem then yields

$$\begin{aligned} \|Kq_1 - Kq_2\| &\leq \|A\|_\infty \int_0^1 e^{\lambda t} \Big(\int_t^1 |q_1 - q_2| \, d\tau \Big) \, dt \\ &= \|A\|_\infty \int_0^1 |q_1(\tau) - q_2(\tau)| \Big(\int_0^\tau e^{\lambda t} \, dt \Big) \, d\tau \\ &\leq \frac{\|A\|_\infty}{\lambda} \|q_1 - q_2\|. \end{aligned}$$

Since we have chosen $\lambda > \|A\|_\infty$, the map K is a contraction and the existence and uniqueness of a solution to (4.8) then follows from Banach-Picard's fixed point theorem.

Let us now establish (4.11). Using z (which is Lipschitz) as test-function in (4.10), we first get

$$\int_0^1 q \cdot \dot{z} \, dt = - \int_0^1 B \cdot z \, dt - b \cdot z(1) + \int_{[0,1]^2} (A(s)z(s)) \cdot q(t) \, d\gamma(t,s)$$

then using (4.2) yields

$$\int_0^1 q \cdot \dot{z} \, dt = \int_{[0,1]^2} (A(s)z(s)) \cdot q(t) \, d\gamma(t,s) + \int_0^1 q(t) \cdot \langle a, \mu_t - \nu_t \rangle \, dt$$

which proves (4.11). □

Theorem 4.3. *Under the assumptions of this section, if $\gamma \in \Gamma$, $\gamma = (\nu_t) \otimes \mathcal{L}^1 = \nu \otimes (\nu_s^*)$ solves (3.1) and $x = x_\gamma$ then one has for $\mathcal{L}^1$-a.e. t*

$$\mathrm{spt}(\nu_t) \subset \mathrm{argmax}_{s \in [0,t]} \big\{ q(t) \cdot f(s, x(s)) - g(t,s) \big\}, \tag{4.13}$$

where q is the adjoint variable whose dynamics is given by

$$q(1) = -\nabla h(x(1)), \;\; \dot{q}(s) = \nabla_x j(s, x(s)) - \nabla_x f(s, x(s))^T \langle q, \nu_s^* \rangle \, \nu(s). \tag{4.14}$$

Proof. Let $\gamma = (\nu_t)\otimes\mathcal{L}^1 \in \Gamma$ be optimal for (3.1); then for every $\eta = (\mu_t)\otimes\mathcal{L}^1$, it follows from (4.5), (4.6) and (4.11) that

$$\int_0^1 \langle q(t)\cdot f(.,x(.)) - g(t,.), \nu_t - \mu_t\rangle \, dt \geq 0$$

with q defined by (4.14). Since in the previous inequality, μ_t is an arbitrary probability measure supported on $[0,t]$, (4.13) directly follows. □

Remark 4.4. In the case where f is linear in x (that is $f(s,x) = A(s)x$) and $j(t,.)$ and h are convex (so that $J(\gamma)$ is convex in x_γ, but in general not in γ), then the following condition is sufficient for optimality:

$$\mathrm{spt}(\nu_t) \subset \mathrm{argmax}_{s\in[0,t]} \left\{q(t)\cdot(A(s)x_\eta(s)) - g(t,s)\right\}, \tag{4.15}$$

for every $\eta \in \Gamma$ and a.e. $t \in [0,1]$. Indeed, if $\gamma = (\nu_t)\otimes\mathcal{L}^1$, $x = x_\gamma$, q is the adjoint variable defined as in theorem 4.3 and $\eta = (\mu_t)\otimes\mathcal{L}^1 \in \Gamma$, by the above convexity assumptions and the same manipulations as before (integration by parts and disintegration), we get:

$$\begin{aligned} J(\eta) - J(\gamma) &\geq \int_0^1 \nabla_x j(s,x(s))\cdot(x_\eta(s) - x(s))ds + \int_{[0,1]^2} g(t,s)d(\eta-\gamma)(t,s) \\ &= \int_0^1 \langle q(t)\cdot Ax_\eta(.) - g(t,.), \nu_t - \mu_t\rangle \, dt \end{aligned}$$

which proves that (4.15) is a sufficient optimality condition. Of course, (4.15) is not necessary in general and is difficult to exploit since it involves every admissible η. However, we shall give a simple example in section 5 where the sufficient condition (4.15) together with the necessary condition of theorem 4.3 actually enables one to compute an optimal control.

Remark 4.5. Let us indicate, without giving details, that one may obtain in a similar way as in theorem 4.3 optimality conditions for the optimal control of the slightly more general state equation containing a local term:

$$\dot{x}(t) = \eta(t,x(t)) + \langle f(.,x(.)), \nu_t\rangle\,,\ x(0) = x_0.$$

Indeed, considering the minimization of the cost J (which has the same form as before), one gets, by similar arguments as previously, that any minimizer $\gamma = (\nu_t)\otimes\mathcal{L}^1$ has to satisfy the statement of theorem 4.3, the only modification being that the adjoint equation now reads as

$$\dot{q}(s) = \nabla_x j(s,x(s)) - \nabla_x\eta(s,x(s))^T q(s) - \nabla_x f(s,x(s))^T \langle q, \nu_s^*\rangle\, \nu(s). \tag{4.16}$$

5 Examples

5.1 Scalar case

We start by considering the scalar ODE with memory

$$\dot{x}(t) = \alpha \langle x, \nu_t \rangle, \qquad x(0) = 1$$

where $\alpha > 0$ is a parameter and (ν_t) is an admissible control. It is clear that if we want to minimize a cost like

$$J(\gamma) = \int_0^1 x_\gamma(t)\, dt$$

the best choice for the control is $\gamma = (\nu_t) \otimes \mathcal{L}^1$ with $\nu_t = \delta_0$ for all $t \in [0,1]$, which gives

$$x(t) = 1 + \alpha t, \qquad J_{min} = 1 + \frac{\alpha}{2}.$$

Analogously, for a cost like

$$J(\gamma) = -x_\gamma(1)$$

the best choice is $\gamma = (\nu_t) \otimes \mathcal{L}^1$ with $\nu_t = \delta_t$ for all $t \in [0,1]$, which gives

$$x(t) = e^{\alpha t}, \qquad J_{min} = -e^{\alpha}.$$

Take now the cost

$$J(\gamma) = a \int_0^1 x_\gamma(t)\, dt - b x_\gamma(1)$$

with $a, b > 0$ and with no penalization in the use of memory (i.e. $g = 0$). The Pontryagin principle of theorem 4.3 gives the adjoint equation

$$\dot{q}(t) = a - \alpha \langle q, \nu_t^* \rangle \nu(t), \qquad q(1) = b \tag{5.1}$$

and the necessary condition of optimality

$$\mathrm{spt}(\nu_t) \subset \mathrm{argmax}_{s \in [0,t]} \left\{ \alpha q(t) x_\gamma(s) \right\} \tag{5.2}$$

since x_γ is increasing, the latter is equivalent to

$$\nu_t = \begin{cases} \delta_t & \text{if } q(t) > 0 \\ \delta_0 & \text{if } q(t) < 0. \end{cases}$$

Thanks to remark 4.4 and the fact that x_η is nondecreasing for every admissible η then the conditions (5.1)-(5.2) are in fact *sufficient*. Since $q(1) = b > 0$,

we necessarily have $\nu_t = \delta_t$ for t close to 1. Now, it is natural, in view of the previous considerations to look for an optimal control in the form

$$\nu_t = \begin{cases} \delta_0 & \text{if } t \in (0, t_0) \\ \delta_t & \text{if } t \in (t_0, 1) \end{cases}$$

where $t_0 \in [0, 1]$ will be determined in such a way that the previous necessary and sufficient optimality conditions hold. For such a family of Dirac masses, a direct computation yields

$$\nu = t_0 \delta_0 + \mathcal{L}^1_{[t_0,1]}$$

and

$$\nu_s^* = \begin{cases} t_0^{-1} \mathcal{L}^1_{[0,t_0]} & \text{if } s = 0 \text{ and } t_0 \neq 0, \\ \delta_0 & \text{if } s = 0 \text{ and } t_0 = 0, \\ \delta_t & \text{if } s \in (t_0, 1) \end{cases}$$

(note that ν_s^* is only defined for ν-a.e. s). Integrating (5.1) on $(t_0, 1]$ then yields

$$q(t) = \frac{a}{\alpha} + \left(b - \frac{a}{\alpha}\right) e^{\alpha(1-t)}, \ t \in (t_0, 1).$$

There are now two cases: either

$$\frac{b}{a} \geq \frac{e^\alpha - 1}{\alpha e^\alpha} \tag{5.3}$$

in which case, q remains positive on $(0, 1)$ and thus $t_0 = 0$ and $\nu_t = \delta_t$ for every $t \in [0, 1]$ is optimal. Or (5.3) does not hold; in this case, set

$$t_0 = 1 - \frac{1}{\alpha} \ln\left(\frac{a}{a - \alpha b}\right) \in (0, 1).$$

Now let $t \in (0, t_0)$, integrating (5.1) between t and t_0, we simply get $q(t) = a(t - t_0)$ so that q is negative on $(0, t_0)$ and then $\nu_t = \delta_0$ for $t \in (0, t_0)$ and $\nu_t = \delta_t$ for $t \in (t_0, 1)$ is an optimal control.

5.2 A two-dimensional example

Let us now consider the two-dimensional linear-quadratic-like problem that consists in minimizing

$$\int_0^1 \frac{1}{2} \left(a(t)x^2(t) + b(t)y^2(t)\right) dt$$

with respect to the admissible control (ν_t) when the state equations for x and y are

$$\dot{x}(t) = \alpha \langle x, \nu_t \rangle, \ \dot{y}(t) = \beta \langle y, \nu_t \rangle$$

together with the initial conditions $(x(0), y(0)) = (x_0, y_0)$. The weight functions a and b are continuous (but not necessarily positive). We also assume that $\alpha > 0$, $\beta > 0$ and $x_0 > 0$, $y_0 < 0$ which guarantees that x is increasing and positive and y is decreasing and negative. The adjoint equations read as

$$\dot{q}_1 = ax - \alpha \langle q_1, \nu_s^* \rangle \nu, \quad \dot{q}_2 = by - \alpha \langle q_2, \nu_s^* \rangle \nu, \quad q_1(1) = q_2(1) = 0. \tag{5.4}$$

Furthermore, the maximum principle of theorem 4.3 gives

$$\operatorname{spt} \nu_t \subset \operatorname{argmax}_{s \in [0,t]} \big\{ q_1(t)x(s) + q_2(t)y(s) \big\}.$$

Since x is increasing and y is decreasing, we thus deduce that:

$$\begin{aligned} q_1(t) > 0,\ q_2(t) < 0 &\Rightarrow \nu_t = \delta_t, \\ q_1(t) < 0,\ q_2(t) > 0 &\Rightarrow \nu_t = \delta_0. \end{aligned}$$

Now, if $a(1) > 0$ and $b(1) > 0$, it is easy to deduce from (5.4) and the fact that $\operatorname{spt} \nu_s^* \subset [s, 1]$ (and that $\nu(\{1\}) = 0$) that $q_1(t) < 0$ and $q_2(t) > 0$ for $t < 1$ sufficiently close to 1 and therefore the optimal ν's have to satisfy $\nu_t = \delta_0$ for t close to 1.

If a and b are everywhere positive, then (5.4) implies $q_1 < 0$ and $q_2 > 0$ on $[0, 1)$ so that (not surprisingly) there is only one optimal ν that is $\nu_t = \delta_0$ for every t. Now let us consider the case where a and b may change sign and let us look for conditions that ensure that $q_1(t) > 0$ and $q_2(t) < 0$ for t close to 0 so that optimal ν's satisfy $\nu_t = \delta_t$ for t close to 0.

Let us assume further that $\alpha < 1$ and $\beta < 1$, then we straightly deduce from the state equation the estimates

$$\|x\|_\infty \leq \frac{x_0}{1-\alpha},\ \|y\|_\infty \leq \frac{|y_0|}{1-\beta} \tag{5.5}$$

using those estimates and the adjoint equation for q_1, we also get

$$\|q_1\|_\infty \leq \frac{\|a\|_\infty x_0}{(1-\alpha)^2}. \tag{5.6}$$

We then get

$$\begin{aligned} -q_1(0) &= \int_0^1 ax\, dt - \alpha \int_0^1 \langle q_1, \nu_s^* \rangle\, d\nu \\ &\leq \int_0^1 a_+ x\, dt - \int_0^1 a_- x\, dt + \frac{\alpha \|a\|_\infty x_0}{(1-\alpha)^2} \\ &\leq x_0 \Big(\frac{1}{1-\alpha} \int_0^1 a_+\, dt - \int_0^1 a_-\, dt + \frac{\alpha \|a\|_\infty}{(1-\alpha)^2} \Big) \end{aligned}$$

so that $q_1(t) < 0$ for t close to 0 as soon as

$$\int_0^1 a_- \, dt > \frac{1}{1-\alpha} \int_0^1 a_+ \, dt + \frac{\alpha \|a\|_\infty}{(1-\alpha)^2}.$$

A similar condition on b ensures that $q_2(t) > 0$ for t close to 0.

Acknowledgments: The work of the authors has been partially supported by their respective institutions through an international cooperation agreement Pise-Paris Dauphine. G.C. gratefully acknowledges the support of the Agence Nationale de la Recherche through the projects OTARIE and EVAMEF. Finally, the authors are grateful to an anonymous referee whose suggestions led to an improved statement in proposition 3.3.

References

[1] L. Ambrosio: *Lecture notes on optimal transport problems.* In "Mathematical aspects of evolving interfaces", Lecture Notes in Mathematics **1812**, Springer-Verlag, Berlin (2003), 1–52.

[2] R. Bellman, K.L. Cooke: *Differential-difference equations.* Mathematics in Science and Engineering, Academic Press, New York-London (1963).

[3] G. Buttazzo: *Semicontinuity, Relaxation and Integral Representation in the Calculus of Variations.* Pitman Res. Notes Math. Ser. **207**, Longman, Harlow (1989).

[4] G. Carlier, A. Houmia, R. Tahraoui: *On Pontryagin's principle for the optimal control of some state equation with memory.* To appear in J. Convex Analysis, **17** no. 3 (2010).

[5] G. Carlier, R. Tahraoui: *On some optimal control problems governed by a state equation with memory.* ESAIM Control Optim. Calc. Var., **14** no. 4 (2008), 725–743.

[6] G. Carlier, R. Tahraoui: *Hamilton-Jacobi-Bellman equations for the optimal control of a state equation with memory.* To appear in ESAIM Control Optim. Calc. Var. (2010).

[7] C. Castaing, M. Valadier: *Convex Analysis and Measurable Multifunctions.* Lecture Notes in Mathematics **580**, Springer-Verlag, Berlin (1977).

[8] C. Dellacherie, P.A. Meyer: *Probabilities and Potential.* Mathematical Studies **29**, North-Holland, Amsterdam (1978).

[9] H.G. Eggleston: *Convexity.* Cambridge Tracts in Mathematics and Mathematical Physics **47**, Cambridge University Press, New York (1958).

[10] J.B. Hiriart-Urruty, C. Lemaréchal: *Fundamentals of Convex Analysis.* Grundlehren text editions, Springer-Verlag, Berlin (2001).

[11] E. Jouini, P.F. Koehl, N. Touzi: *Optimal investment with taxes: an optimal control problem with endogeneous delay.* Nonlinear Anal. TMA, **37** (1999) 31–56.

[12] Y. Kuang: *Global stability for a class of nonlinear nonautonomous delay equations.* Nonlinear Anal. TMA, **17** (1991), 627–634.

[13] Y. Kuang: *Delay differential equations with applications in population dynamics.* Mathematics in Science and Engineering, Academic Press, Boston (1993).

[14] R.T. Rockafellar: *Convex Analysis.* Princeton University Press, Princeton (1972).

[15] T. Roubíček: *Relaxation in Optimization Theory and Variational Calculus.* De Gruyter Series in Nonlinear Analysis and Applications **4**, Walter de Gruyter, Berlin (1997).

[16] L. Samassi, R. Tahraoui: *How to state necessary optimality conditions for control problems with deviating arguments?.* ESAIM Control Optim. Calc. Var., **14** no. 2 (2008), 381–409.