ON SOME OPTIMAL CONTROL PROBLEMS GOVERNED BY A STATE EQUATION WITH MEMORY

Guillaume Carlier[1] and Rabah Tahraoui[1]

Abstract. The aim of this paper is to study problems of the form:

$$\inf_{u \in V} J(u) \text{ with } J(u) := \int_0^1 L(s, y_u(s), u(s)) \mathrm{d}s + g(y_u(1))$$

where V is a set of admissible controls and y_u is the solution of the Cauchy problem:

$$\begin{cases} \dot{x}(t) = \langle f(., x(.)), \nu_t \rangle + u(t), \ t \in (0, 1), \\ x(0) = x_0, \end{cases}$$

and each ν_t is a nonnegative measure with support in $[0, t]$. After studying the Cauchy problem, we establish existence of minimizers, optimality conditions (in particular in the form of a nonlocal version of the Pontryagin principle) and prove some regularity results. We also consider the more general case where the control also enters the dynamics in a nonlocal way.

1. Introduction and examples

The aim of this paper is to study deterministic optimal control problems of the form

$$\inf_{u \in V} J(u) \text{ with } J(u) := \int_0^1 L(s, y_u(s), u(s)) \mathrm{d}s + g(y_u(1)). \tag{1}$$

Here, u is the control variable and the state equation governing the dynamics of the state variable $y = y_u$ is an integrodifferential equation modelling memory or delay effects. More precisely, y_u is the solution of the Cauchy problem

$$\begin{cases} \dot{x}(t) = \displaystyle\int_{[0,t]} f(s, x(s)) \mathrm{d}\nu_t(s) + u(t) = \langle f(., x(.)), \nu_t \rangle + u(t), \ t \in (0, 1), \\ x(0) = x_0. \end{cases}$$

---

[1] Université Paris Dauphine, CEREMADE, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, France;
`carlier@ceremade.dauphine.fr; tahraoui@ceremade.dauphine.fr`

In the state equation above, $(\nu_t)_t$ denotes a given measurable family of nonnegative measures such that ν_t has its support in $[0, t]$ (*i.e.* only the past of the trajectory is taken into account). Let us remark that in the previous equation, the control enters the dynamics only through its current value $u(t)$ and not through its past values. The case of a more general state equation of the form:

$$\dot{x}(t) = \langle f(., x(.)), \nu_t \rangle + \langle u(.), \mu_t \rangle \tag{2}$$

$((\mu_t)_t$ being a measurable family of nonnegative measures) may also be relevant in applications and will also be treated in the paper. There are of course many variants of controlled dynamics of the form (2) like:

$$\dot{x}(t) = F(t, x(t), u(t), \langle f(t, ., x(.)), \nu_t \rangle, \langle g(t, ., u(.)), \mu_t \rangle). \tag{3}$$

For the sake of simplicity, we will restrict ourselves to (2) and claim that the main ideas of the paper can be adapted to the more general case of (3).

Problems of the form above arise in different applied settings both in engineering and decision sciences. 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. Such problems have in general been modelled by delayed or deviating arguments differential equations (see examples below). Let us remark that such equations are particular cases of the equations dealt with in the present paper. Before going further, let us consider some examples.

Example 1. *Problems with lags or deviating arguments*

The simplest case is the case of a delayed equation of the form

$$\dot{y}(t) = \begin{cases} u(t) & \text{if } t < \tau \\ f(y(t - \tau)) + u(t) & \text{if } t \geq \tau \end{cases}$$

where $\tau > 0$ is a given delay. This corresponds of course to

$$\nu_t = \begin{cases} 0 & \text{if } t < \tau \\ \delta_{t-\tau} & \text{if } t \geq \tau. \end{cases}$$

The study of delayed differential equations and their control may be traced back to the early 60's (in particular the book of Bellman and Cooke [2], see also Oguztöreli [13]) and there is nowadays a huge literature on the subject.

A natural generalization of time delayed equations is the case of deviating argument equations of the form

$$\dot{y}(t) = f(y(\theta(t))) + u(t)$$

where θ is a given *deviation* function (with $0 \leq \theta(t) \leq t$ say). In this case $\nu_t = \delta_{\theta(t)}$. More generally, one can consider the superposition of several (or even a continuum of) deviations, this corresponds to ν_t of the form:

$$\nu_t := \sum_{i=1}^{k} \alpha_i(t) \delta_{\theta_i(t)} \text{ or more generally } \nu_t := \int_A \alpha(t, a) \delta_{\theta(t,a)} \mathrm{d}\mu(a).$$

The study of general deviating arguments problems has started in the 60's (see the book of El'sgol'ts [9]). More recently, in a calculus of variations framework, Drakhlin and Stepanov [6] obtained lower semi-continuity results for integral functionals with deviating arguments and Samassi and Tahraoui [16,17] obtained optimality conditions for such functionals (see also Samassi's thesis [15]). Let us also mention that problems with deviating arguments where the deviation is unknown *a priori* and determined by optimizing some functional give rise to additional difficulties (see Jouini, Koehl and Touzi [11,12] for a problem of this type arising in economics).

Example 2. *A deterministic advertising model*

Arguing that there is a time lag between advertising expenditure and the corresponding effect on the goodwill level, Gozzi and Marinelli proposed in [10] a stochastic optimal control advertising model with delay. The state variable y is the stock of advertising goodwill and the control variable u models the intensity of advertising spending. The dynamics of y is assumed to be governed by a stochastic delay differential equation of the form

$$\begin{cases} dy_t = \left[a_0 y(t) + b_0 u(t) + \int_0^r (a_1(s)y(t-s) + b_1(s)u(t-s))\mathrm{d}s \right] \mathrm{d}t + \sigma \mathrm{d}W_t \\ (y, u) \text{ prescribed on } [-r, 0] \end{cases}$$

where W_t stands for the standard one-dimensional Brownian motion, $a_0 \leq 0$, $b_0 \geq 0$ are given constants, a_1 and b_1 are given L^2 functions on $[-r, 0]$ with $b_1 \geq 0$. In the deterministic case, $\sigma = 0$, and we are left with a special case of the dynamics (2).

Example 3. *Variants of Ramsey's economic growth model*

In [14], Ramsey proposed a celebrated model of economic growth. In finite horizon, it reads as:

$$\sup \int_0^T \mathrm{e}^{-\delta t} U(c(t))\mathrm{d}t + g(k(T)) \ : \ \dot{k}(t) = f(t, k(t)) - c(t), k(0) = k_0.$$

The state variable k is the capital, the control c is the consumption rate, and f is the production function. In the original model $i(s) := f(s, k(s)) - c(s)$ represents the investment at time t and writing $\dot{k}(t) = f(t, k(t)) - c(t)$ amounts to assuming that saving instantaneously converts into capital. This is a strong assumption and it is more realistic to assume that the capital growth depends on past investments according to a relation of the form

$$\dot{k}(t) = \int_0^t i(s)\mathrm{d}\nu_t(s) = \int_0^t (f(s, k(s)) - c(s))\mathrm{d}\nu_t(s).$$

We refer to Boucekkine *et al.* [3] for a variant of Ramsey's model with delays. There are actually many related problems in economics, management of natural resources and finance (optimal harvesting, models with age-structured populations...). For extensions to the stochastic control framework and more applications, we refer for instance to Elsanosi, Øksendal and Sulem [8].

In Section 2, the Cauchy problem is studied in details. Section 3 deals with the existence of optimal controls, the material of these sections is rather standard but is included for the sake of completeness. The main novelty consists of the derivation of optimality conditions obtained in Section 4. In the case of an unconstrained control, we first establish the Euler-Lagrange equations of the problem, then apply the optimality conditions to obtain some regularity results. We then consider the case of (convex) constraints on the control and prove that in this case, the optimality conditions can be written in the form of a Pontryagin principle with a nonlocal Hamiltonian. Finally, Section 5 extends the previous result to the more general state equation (2).

2. On the Cauchy problem

Our first aim is to solve the Cauchy problem:

$$\begin{cases} \dot{x}(t) &= \langle f(., x(.)), \nu_t \rangle + v, \ t \in (0, 1), \\ x(0) &= x_0. \end{cases} \tag{4}$$

Let us consider the following assumptions:

- **(H1)** $f \in C^0([0, 1] \times \mathbb{R}^d, \mathbb{R}^d)$, and there exists $(a, b) \in \mathbb{R}_+^2$ such that $|f(t, x)| \leq a|x| + b$, for all $(t, x) \in [0, 1] \times \mathbb{R}^d$;

- **(H'1)** for every $r > 0$, there exists k such that for all $(t, x, y) \in [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d$ such that $|x| \leq r$ and $|y| \leq r$, $|f(t, x) - f(t, y)| \leq k|x - y|$;
- **(H2)** for each $t \in [0, 1]$, ν_t is a nonnegative finite measure such that $\nu_t((t, 1]) = 0$ and $t \mapsto \nu_t$ is measurable in the sense that $t \mapsto \langle g, \nu_t \rangle$ is measurable for every $g \in C^0([0, 1], \mathbb{R})$;
- **(H3)** defining $\alpha(t) := \nu_t([0, 1]) = \nu_t([0, t])$, we assume $\alpha \in L^1(0, 1)$.

In the sequel, we shall simply write C^0, L^p, $W^{1,p}$, BV instead respectively of $C^0([0, 1], \mathbb{R}^d)$, $L^p((0, 1), \mathbb{R}^d)$, $W^{1,p}((0, 1), \mathbb{R}^d)$, BV$((0, 1), \mathbb{R}^d)$. For x and y in $\mathbb{R}^d$, the usual inner product of x and y will be denoted $x \cdot y$.

2.1. **Existence, uniqueness**

Before we go further, under assumption **(H3)**, we have the following lemma:

Lemma 1. *Let $\lambda > 0$ and define for every $t \in [0, 1]$,*

$$\varphi_\lambda(t) := \int_0^t e^{\lambda(s-t)} \alpha(s) ds$$

then φ_λ converges uniformly to 0 on $[0, 1]$ as $\lambda \to +\infty$.

Proof. Let $\varepsilon > 0$, and let $\delta > 0$ be such that $\int_A \alpha \leq \varepsilon$ for every Borel set A with Lebesgue measure less than δ. Then for every $t \in [0, \delta]$, $\varphi_\lambda(t) \leq \varepsilon$ and for $t \in [\delta, 1]$, one has:

$$0 \leq \varphi_\lambda(t) = \int_0^{t-\delta} e^{\lambda(s-t)} \alpha(s) ds + \int_{t-\delta}^t e^{\lambda(s-t)} \alpha(s) ds \leq e^{-\lambda \delta} \|\alpha\|_{L^1} + \varepsilon.$$

The desired conclusion follows. $\square$

Proposition 1. *Let us assume that $v \in L^1$, that **(H1)**, **(H2)** and **(H3)** hold and let us denote by S_v the set of $W^{1,1}$ solutions of (4). We then have:*

1. *$S_v \neq \emptyset$, and for all $x \in S_v$, $\|x\|_\infty \leq M$ for some constant $M = M(|x_0|, a, b, \|\alpha\|_{L^1}, \|v\|_{L^1})$;*
2. *S_v is compact for both C^0 and $W^{1,1}$ topologies;*
3. *if, in addition **(H'1)** is satisfied then S_v consists of a single point y_v and the map $v \in L^1 \mapsto y_v \in W^{1,1}$ is locally Lipschitz.*

Proof. 1. For $v \in L^1$ and $x \in C^0$, define for all $t \in [0, 1]$:

$$T_v x(t) := x_0 + \int_0^t (\langle f(., x(.)), \nu_s \rangle + v(s)) ds.$$

It is obvious under our assumptions that $T_v(C^0) \subset W^{1,1} \subset C^0$. Let x and y be in C^0, we have:

$$\|T_v x - T_v y\|_\infty \leq \|\alpha\|_{L^1} \|f(., x(.)) - f(., y(.))\|_\infty$$

which proves that T_v is continuous (for the uniform topology of C^0).

Let $\lambda > 0$ and define for every $x \in C^0$:

$$\|x\|_\lambda := \sup_{t \in [0,1]} e^{-\lambda t} |x(t)|. \tag{5}$$

Of course, $(C^0, \|.\|_\lambda)$ is a Banach space. Let $x \in C^0$, we have:

$$\langle |f(., x(.))|, \nu_s \rangle \leq \langle a|x| + b, \nu_s \rangle \leq (a|x_0| + a\|x - x_0\|_\lambda e^{\lambda s} + b)\alpha(s)$$

defining $C := \|v\|_{L^1} + (a|x_0| + b)\|\alpha\|_{L^1}$, we then get:

$$|T_v x(t) - x_0| \le C + a\|x - x_0\|_\lambda \int_0^t e^{\lambda s}\alpha(s)\mathrm{d}s$$

this yields:

$$\|T_v x - x_0\|_\lambda \le C + a\|x - x_0\|_\lambda \max_{t\in[0,1]} \varphi_\lambda(t). \qquad (6)$$

Let us denote by $\overline{B}_\lambda(x_0, r)$ the closed ball of C^0 with center x_0 and radius r for the norm $\|.\|_\lambda$. It then follows from (6) and Lemma 1 that for any $r > C$, $T_v(\overline{B}_\lambda(x_0, r)) \subset \overline{B}_\lambda(x_0, r)$ when λ is chosen sufficiently large. Now, let $x \in \overline{B}_\lambda(x_0, r)$ and $t_2 > t_1$ be in $[0, 1]$, we have for some nonnegative constant C':

$$|T_v x(t_2) - T_v x(t_1)| \le \int_{t_1}^{t_2} (C'\alpha(s) + |v(s)|)\mathrm{d}s.$$

This proves that $T_v(\overline{B}_\lambda(x_0, r))$ is uniformly equicontinuous, we thus deduce from Ascoli's theorem that $T_v(\overline{B}_\lambda(x_0, r))$ is relatively compact in C^0. From Schauder's fixed point theorem we deduce that T_v admits at least one fixed point in $\overline{B}_\lambda(x_0, r)$ which proves that S_v is nonempty. S_v is obviously closed in C^0 since T_v is continuous. Choosing λ such that $\max_{t\in[0,1]} \varphi_\lambda(t) \le 1/(2a)$, we deduce from (6) that $S_v \subset \overline{B}_\lambda(x_0, 2C)$ which establishes the first claim of the proposition.

2. Since $S_v = T_v(S_v)$ and $T_v(\overline{B}_\lambda(x_0, 2C))$ is relatively compact in C^0, we can conclude that S_v is compact in C^0. We claim that S_v is also compact in $W^{1,1}$, indeed let $(x_n)_n$ be a sequence of elements of S_v, up to some subsequence we may assume that x_n converges uniformly to some $x \in S_v$. Since, one has:

$$|\dot{x} - \dot{x}_n| \le \alpha \|f(., x(.)) - f(., x_n(.))\|_\infty$$

it thus follows that x_n converges uniformly to x in $W^{1,1}$.

3. Finally, let us assume that f satisfies the local Lipschitz condition and let x_1 and x_2 belong to S_v. We already know that $S_v \subset C^0([0,1], \overline{B}_M)$ for some constant M; let us denote by k the Lipschitz constant of f with respect to its second argument on $[0,1] \times \overline{B}_M$. For each $t \in [0,1]$, we then have:

$$|T_v x_1(t) - T_v x_2(t)| \le \int_0^t k\langle |x_1 - x_2|, \nu_s\rangle \mathrm{d}s \le k\|x_1 - x_2\|_\lambda \int_0^t e^{\lambda s}\alpha(s)\mathrm{d}s$$

hence:

$$\|T_v x_1 - T_v x_2\|_\lambda = \|x_1 - x_2\|_\lambda \le \|x_1 - x_2\|_\lambda k \max \varphi_\lambda$$

again choosing λ large enough and using Lemma 1, we get $x_1 = x_2$.

Let $\rho > 0$ and v_1, v_2 be in the centered ball of L^1 of radius ρ and set $y_i := y_{u_i}$ for $i = 1, 2$. Since

$$y_2(t) - y_1(t) = \int_0^t (\langle f(., y_2(.)) - f(., y_1(.)), \nu_s\rangle)\mathrm{d}s + \int_0^t (u_2 - u_1),$$

we deduce from 1. and **(H'1)**, that there exists $k = k(\rho)$ such that for every $\lambda > 0$ and every $t \in [0,1]$, one has:

$$|y_1(t) - y_2(t)|e^{-\lambda t} \le k\|y_1 - y_2\|_\lambda \varphi_\lambda(t) + \|u_1 - u_2\|_{L^1}.$$

Hence for λ large enough ($\lambda \ge \lambda(\rho)$ say), one has:

$$\|y_1 - y_2\|_\lambda \le 2\|u_1 - u_2\|_{L^1}.$$

Thus, there exists $C = C(\rho)$ such that $\|y_1 - y_2\|_\infty \le C(\rho)\|u_1 - u_2\|_{L^1}$. We deduce the desired result by remarking that

$$|\dot{y}_1(t) - \dot{y}_2(t)| \le |u_1(t) - u_2(t)| + k\alpha(t)\|y_1 - y_2\|_\infty. \qquad \square$$

Remark. When $d = 1$ and f is sublinear and nondecreasing (but not locally Lipschitz) with respect to its second argument, it is possible (by suitably regularizing f from above of from below) to show that the Cauchy problem admits a unique largest and a unique smallest solution.

2.2. Continuous dependence and linearization

In this paragraph, we always assume **(H1)**, **(H'1)**, **(H2)** and **(H3)**.

Lemma 2. *If a sequence v_n converges weakly in L^1 to some v then y_{v_n} converges uniformly to y_v on $[0, 1]$.*

Proof. Let us set $y_n := y_{v_n}$ and let us prove that y_n is uniformly equicontinuous (we already know that it is bounded in L^∞ by Prop. 1). On the one hand, we know from Proposition 1 that there exists a constant C such that for all n and all t, h such that $[t, t+h]$, one has:

$$|y_n(t+h) - y_n(t)| \le \int_t^{t+h} (C\alpha + |v_n|).$$

On the other hand, v_n satisfies Dunford-Pettis criterion hence is uniformly integrable. We therefore deduce from Ascoli's theorem that (y_n) is precompact in C^0. Let z be some limit point of (y_n) for the C^0 norm, if we establish that $z = y_v$, the proof will be complete. To prove that $z = y_v$ it is enough to pass to the limit in:

$$y_n(t) = x_0 + \int_0^t (\langle f(., y_n(.)), \nu_s \rangle + v_n(s)) ds$$

and to invoke the uniqueness result of Proposition 1. $\qquad \square$

For the following result, we further assume that for every $t \in [0, 1]$, $f(t, .)$ is of class C^1 on $\mathbb{R}^d$, we denote by $D_x f(t, y)$ the Jacobian matrix of $f(t, .)$ at the point y and assume that $D_x f$ is a continuous function of both variables t and x. We then have:

Proposition 2. *Let $(u, v) \in L^1 \times L^1$ and $\varepsilon > 0$. Define $y := y_u$ and $y_\varepsilon := y_{u+\varepsilon v}$, then:*

$$\frac{y_\varepsilon - y}{\varepsilon} \to h \quad in \ W^{1,1} \ as \ \varepsilon \to 0^+$$

where h is the solution of the linearized state equation:

$$\begin{cases} \dot{h}(t) = \langle D_x f(., y(.))h(.), \nu_t \rangle + v(t), \ t \in (0, 1), \\ h(0) = 0. \end{cases} \qquad (7)$$

Proof. For $\varepsilon > 0$, let us set $h_\varepsilon := (y_\varepsilon - y)/\varepsilon$. We know from Proposition 1 that h_ε is bounded in $W^{1,1}$ hence in C^0, hence there is a constant k such that for every $\varepsilon > 0$ and $t \in (0, 1)$:

$$|\dot{h}_\varepsilon(t)| \le k\alpha(t) + |v(t)|.$$

Since the rightmost member of this inequality is L^1, the family (h_ε) is precompact in C^0. Let η be some limit point for the C^0 norm of (h_ε), $\eta = \lim_j h_{\varepsilon_j}$ and let us prove that $\eta = h$. For all t and j, we have:

$$h_{\varepsilon_j}(t) = \int_0^t \left(\left\langle \frac{f(., y(.) + \varepsilon_j h_{\varepsilon_j}(.)) - f(., y(.))}{\varepsilon_j}, \nu_s \right\rangle + v(s) \right) ds$$

letting j tend to ∞ we get:

$$\eta(t) = \int_0^t (\langle D_x f(., y(.))\eta(.), \nu_s\rangle + v(s))\mathrm{d}s.$$

Since h is the unique solution of the linearized equation (7), this proves that $h = \eta$ and that the whole h_ε converges to h in C^0 as ε tends to 0. To prove that there is also convergence in $W^{1,1}$ we remark that:

$$\|\dot{h}_\varepsilon - \dot{h}\|_{L^1} \leq \|\alpha\|_{L^1}\delta_\varepsilon$$

where

$$\delta_\varepsilon = \left\| \frac{f(., y(.) + \varepsilon h_\varepsilon(.)) - f(., y(.))}{\varepsilon} - D_x f(., y(.))h(.) \right\|_\infty$$

which tends to 0 as $\varepsilon \to 0^+$. □

3. Existence of optimal controls

From now on, we will always assume **(H1)**, **(H'1)**, **(H2)** and **(H3)**. In addition we assume the following:

- **(H4)** g is l.s.c., L is a normal integrand from $[0, 1] \times \mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R} \cup +\infty$ (*i.e.* L is Borel and for *a.e.* $t \in [0, 1]$, $L(t, ., .)$ is l.s.c.) and $L(t, x, .)$ is convex for every $(t, x) \in [0, 1] \times \mathbb{R}^d$;
- **(H5)** there exists a convex l.s.c. function $\Psi \colon \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{+\infty} \Psi(\xi)/\xi = +\infty$ and $G \in L^1((0, 1), \mathbb{R})$ such that:

$$L(t, x, v) \geq \Psi(|v|) + G(t), \ \forall (t, x, v) \in [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d;$$

- **(H6)** V is a closed convex subset of L^1 and there exists $u_0 \in V$ such that $J(u_0) < +\infty$.

Under assumptions **(H1)**, **(H'1)**, **(H2)**, **(H3)**, **(H4)**, **(H5)** and **(H6)**, one easily obtains existence of a solution to (1):

Proposition 3. *There exists $\overline{u} \in V$ such that $J(\overline{u}) \leq J(u)$ for all $u \in V$.*

Proof. Let $u_n \in V^{\mathbb{N}}$ be a minimizing sequence of J over V. It follows from **(H5)** and Dunford-Pettis theorem that u_n admits a (not relabeled) subsequence that converges weakly in L^1 to some $\overline{u}$, by convexity $\overline{u} \in V$. It then follows from Lemma 2 that $y_n := y_{u_n}$ converges uniformly to $\overline{y} := y_{\overline{u}}$. It then follows from our assumptions and Theorem 2.1, p. 243 of Ekeland and Temam [7], that:

$$J(\overline{u}) \leq \varliminf J(u_n)$$

which proves the result. □

4. Optimality conditions

4.1. Euler-Lagrange equations

In this paragraph, we assume the following:

- **(H7)** for every $t \in [0, 1]$, $f(t, .)$ is of class C^1 and $D_x f$ is continuous;
- **(H8)** g is of class C^1, L is continuous and for every $t \in [0, 1]$, $(x, v) \mapsto L(t, x, v)$ is of class C^1 and $\nabla_x L$ and $\nabla_v L$ are continuous;
- **(H9)** there exists $p > 1$ and $c > 0$ such that L satisfies **(H5)** with $\Psi(\xi) = c\xi^p$, $\alpha \in L^p$, there exists $G_1 \in L^{p'}((0, 1), \mathbb{R})$ (p' being the conjugate exponent of p), $G_2 \in L^1((0, 1), \mathbb{R})$, and for each $r > 0$ there exists $\beta_r \geq 0$ such that for all $(t, x, v) \in [0, 1] \times \overline{B}_r \times \mathbb{R}^d$:

$$|\nabla_v L(t, x, v)| \leq G_1(t) + \beta_r(1 + |u|^{p-1}), \ |\nabla_x L(t, x, v)| \leq G_2(t) + \beta_r(1 + |u|^p).$$

We then consider the following problem

$$\inf_{u \in L^p} J(u) \text{ with } J(u) := \int_0^1 L(s, y_u(s), u(s))\mathrm{d}s + g(y_u(1)). \tag{8}$$

Under our assumptions, let us remark that for every $v \in L^p$, $y_v \in W^{1,p}$ and that, if L is convex in v, then (8) admits at least one solution in L^p. Under the previous assumptions, we also have, as a direct consequence of Proposition 2:

Lemma 3. *J is Gâteaux-differentiable on L^p, and for all $(u,v) \in L^p \times L^p$ one has:*

$$\langle J'(u), v \rangle = \int_0^1 \left(\nabla_x L(t, y_u(t), u(t))h_v(t) + \nabla_v L(t, y_u(t), u(t)) \cdot v(t) \right) \mathrm{d}t + \nabla g(y_u(1)) \cdot h_v(1)$$

where $h_v \in W^{1,p}$ is the solution of the linearized state equation:

$$\begin{cases} \dot{h}(t) = \langle D_x f(., y_u(.))h(.), \nu_t \rangle + v(t), \ t \in (0,1), \\ h(0) = 0. \end{cases} \tag{9}$$

Let us denote by $\mathcal{L}^1$ the Lebesgue measure on $[0,1]$ and let us define the nonnegative measure $\gamma := \nu_t \otimes \mathcal{L}^1$ on $[0,1]^2$, and define ν as the second marginal of γ. Using test-functions, γ and ν are then defined by:

$$\int \phi(t,s)\mathrm{d}\gamma(t,s) = \int_0^1 \left(\int_0^1 \phi(t,s)\mathrm{d}\nu_t(s) \right) \mathrm{d}t, \ \forall \phi \in C^0([0,1]^2, \mathbb{R}),$$

$$\int \phi(s)\mathrm{d}\nu(s) = \int_0^1 \left(\int_0^1 \phi(s)\mathrm{d}\nu_t(s) \right) \mathrm{d}t, \ \forall \phi \in C^0([0,1], \mathbb{R}).$$

Using the disintegration theorem (see for instance the book of Dellacherie and Meyer [5] or the appendix in the lecture notes of Ambrosio [1]) we may also write $\gamma = \nu \otimes \nu_s^*$ where ν_s^* is a measurable family of probability measures on $[0,1]$. We recall that $\phi \in L^1(\gamma)$ if and only if:

- for $\mathcal{L}^1$-a.e. $t \in [0,1]$, $\phi(t,.) \in L^1(\nu_t)$, and
- $t \mapsto \langle \phi(t,.), \nu_t \rangle \in L^1(\mathcal{L}^1)$

which is also equivalent to

- for ν-a.e. $s \in [0,1]$, $\phi(.,s) \in L^1(\nu_s^*)$, and
- $s \mapsto \langle \phi(.,s), \nu_s^* \rangle \in L^1(\nu)$.

Moreover, if $\phi \in L^1(\gamma)$, then:

$$\int_{[0,1]^2} \phi(t,s)\mathrm{d}\gamma(t,s) := \int_0^1 \langle \phi(t,.), \nu_t \rangle \mathrm{d}t = \int_0^1 \langle \phi(.,s), \nu_s^* \rangle \mathrm{d}\nu(s).$$

Let us finally remark that since ν_t is supported by $[0,t]$, ν_s^* is supported by $[s,1]$. For every matrix A, we shall denote by A^T the transpose of A. The Euler-Lagrange equation of (8) then reads as:

Theorem 1. *Let $\overline{u}$ be a solution of (8) and $\overline{y} := y_{\overline{u}}$ be the corresponding trajectory. Slightly abusing notations, let us simply denote*

$$D_x f(., \overline{y}(.))^T \langle \nabla_v L(\overline{y}, \overline{u}), \nu_.^* \rangle \ : \ t \in [0,1] \mapsto D_x f(t, \overline{y}(t))^T \langle \nabla_v L(., \overline{y}(.), \overline{u}(.)), \nu_t^* \rangle.$$

Then the following Euler-Lagrange equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla_v L(., \overline{y}(.), \overline{u}(.)) \right) = \nabla_x L(., \overline{y}(.), \overline{u}(.)) - (D_x f(., \overline{y}(.)))^T \langle \nabla_v L(\overline{y}, \overline{u}), \nu_.^* \rangle)\nu \tag{10}$$

and the transversality condition:

$$\nabla_v L(1, \overline{y}(1), \overline{u}(1)) = -\nabla g(\overline{y}(1)) \tag{11}$$

are satisfied in the $(W^{1,p})'$ *sense, which means that for every* $x \in W^{1,p}$ *such that* $x(0) = 0$, *one has:*

$$0 = \nabla g(\overline{y}(1)) \cdot x(1) + \int_0^1 (\nabla_x L(t, \overline{y}(t), \overline{u}(t)) \cdot x(t) + \nabla_v L(t, \overline{y}(t), \overline{u}(t)) \cdot \dot{x}(t)) \, \mathrm{d}t$$
$$- \int_0^1 ((D_x f(s, \overline{y}(s)))x(s) \cdot \langle \nabla_v L(., \overline{y}(.), \overline{u}(.)), \nu_s^* \rangle) \, \mathrm{d}\nu(s).$$

Proof. To shorten notations, let us set:

$$\Pi(t) := \nabla_v L(t, \overline{y}(t), \overline{u}(t)), \ A(t) := D_x f(t, \overline{y}(t)),$$
$$\Theta(t) := \nabla_x L(t, \overline{y}(t), \overline{u}(t)). \tag{12}$$

Since $\overline{u} \in L^p$, our assumptions imply that $\Pi \in L^{p'}$, $\Theta \in L^1$ and $A \in C^0$.

Let $x \in W^{1,p}$ be such that $x(0) = 0$ and define:

$$v(t) := \dot{x}(t) - \langle A(.)x(.), \nu_t \rangle$$

so that $v \in L^p$ and $x = h_v$. By Lemma 3 and since $\langle J'(\overline{u}), v \rangle = 0$, we get:

$$\int_0^1 (\Theta \cdot x + \Pi \cdot (\dot{x} - \langle A(.)x(.), \nu_t \rangle)) \mathrm{d}t + \nabla g(\overline{y}(1)) \cdot x(1) = 0. \tag{13}$$

Let us define for all $(i, j) \in \{1, ..., d\}^2$, and all $(t, s) \in [0, 1]^2$, $\phi_{ij}(t, s) := \Pi_i(t)A_{ij}(s)x_j(s)$. Since both A and x are continuous on $[0, 1]$, $\phi_{ij}(t, .) \in L^1(\nu_t)$ for *a.e.* t and

$$|\langle \phi_{ij}(t, .), \nu_t \rangle| \leq \|A_{ij}\|_\infty \|x_j\|_\infty \alpha(t)|\Pi_i(t)|$$

since $\Pi \in L^{p'}$ and $\alpha \in L^p$, the right-hand side member is L^1, we deduce that $\phi_{ij} \in L^1(\gamma)$. Hence $\phi_{ij}(., s) \in L^1(\nu_s^*)$ for ν-*a.e.* s, $\langle \phi_{ij}(., s), \nu_s^* \rangle \in L^1(\nu)$, and:

$$\int_0^1 (\Pi(t) \cdot \langle A(.)x(.), \nu_t \rangle) \mathrm{d}t = \sum_{1 \leq i,j \leq d} \int_{[0,1]^2} \phi_{ij}(t, s) \mathrm{d}\gamma(t, s)$$
$$= \sum_{1 \leq i,j \leq d} \int_0^1 \langle \phi_{ij}(., s), \nu_s^* \rangle \, \mathrm{d}\nu(s)$$
$$= \sum_{1 \leq i,j \leq d} \int_0^1 (A_{ij}(s)x_j(s) \langle \Pi_i, \nu_s^* \rangle) \mathrm{d}\nu(s)$$
$$= \int_0^1 x(s) \cdot (A(s)^T \langle \Pi, \nu_s^* \rangle) \mathrm{d}\nu(s).$$

With (13), we thus get, for all $x \in W^{1,p}$ such that $x(0) = 0$

$$\nabla g(\overline{y}(1)) \cdot x(1) + \int_0^1 (\Theta \cdot x + \Pi \cdot \dot{x}) \, \mathrm{d}t - \int_0^1 (A(s)x(s) \cdot \langle \Pi, \nu_s^* \rangle) \mathrm{d}\nu(s) = 0.$$

This exactly means that Π satisfies in the $(W^{1,p})'$ sense:

$$\dot{\Pi} = \Theta - (A(.)^T \langle \Pi(.), \nu_.^* \rangle)\nu, \tag{14}$$

$$\Pi(1) = -\nabla g(\overline{y}(1)). \tag{15}$$

$\square$

Example. We aim to illustrate the previous optimality conditions by a simple example of Ramsey type. Let us consider the following problem:

$$\sup \left\{ \int_0^1 e^{-\delta t} U(c(t)) dt + g(k(1)) \; : \; k(0) = k_0, \; \dot{k}(t) = \alpha \int_0^t k(s) d\nu_t(s) - c(s) \right\}.$$

In this problem, c is the consumption (control variable), k is the stock of capital (state variable), U is a utility function, $\delta > 0$ a discount rate and for the sake of simplicity, we have taken a linear production function $f(k) = \alpha k$.

In the classical no-memory case (*i.e.* $\nu_t = \delta_t$), setting

$$v(t) := e^{-\delta t} U'(c(t)), \; \forall t \in [0, 1] \tag{16}$$

the Euler-Lagrange equation simply reads as $\dot{v} = -\alpha v$ hence:

$$U'(c(t)) = e^{(\delta - \alpha)t} U'(c(0)). \tag{17}$$

In the general case, defining v by (16), the Euler-Lagrange equation becomes:

$$\dot{v}(s) = -\alpha \left(\int_0^1 v(t) d\nu_s^*(t) \right) \nu(s).$$

If we specify to the case where $d\nu_t(s) := g(t,s)\chi_{[0,t]}(s)ds$ (with $\chi_{[0,t]}$ the characteristic function of $[0,t]$) then $\nu_s^*(t) = g(t,s)\chi_{[s,1]}(t)dt$ and $d\nu(s) = ds$. Thus, we get the following linear integrodifferential equation for v:

$$\dot{v}(s) = -\alpha \int_s^1 v(t)g(t,s)dt.$$

For instance, taking $g \equiv 1$ leads to $\ddot{v} = \alpha v$ and $\dot{v}(1) = 0$ so that:

$$U'(c(t)) = \frac{U'(c(0))}{1 + e^{2\sqrt{\alpha}}} \left(e^{(\delta + \sqrt{\alpha})t} + e^{2\sqrt{\alpha} + (\delta - \sqrt{\alpha})t} \right).$$

4.2. **Regularity of optimal controls**

Assuming as previously that $\overline{u}$ is a solution of (8), $\overline{y} := y_{\overline{u}}$ and defining Π, Θ and A by (12), let us denote by $\widetilde{\nu}_\Pi$ the vector-valued measure:

$$d\widetilde{\nu}_\Pi := F_\Pi d\nu \text{ where } F_\Pi(s) := A(s)^T \langle \Pi(.), \nu_s^* \rangle. \tag{18}$$

We recall that $\Theta \in L^{p'}$, $F_\Pi \in L^1(\nu)$ and that the optimality conditions equation of (8), (14) and (15), can be written as:

$$\dot{\Pi} = \Theta - \widetilde{\nu}_\Pi, \; \Pi(1) = -\nabla g(\overline{y}(1)).$$

This implies that $\Pi \in \mathrm{BV}((0,1), \mathbb{R}^d)$ and that for all $t \in [0,1]$

$$\Pi(t) = -\nabla g(\overline{y}(1)) - \int_t^1 \Theta(s)\mathrm{d}s + \widetilde{\nu}_\Pi([t,1]). \tag{19}$$

We immediately deduce from (19) and (18) that Π is continuous except possibly on the set of atoms of ν, in particular Π is continuous as soon as ν has no atoms.

Lemma 4. *Let $\overline{u}$ be a solution of (8), $\overline{y} := y_{\overline{u}}$, and Π be defined by (12). For all $t \in [0,1]$ such that $\nu(\{t\}) = 0$, Π is continuous at t, hence Π has at most countably many discontinuity points. In particular, if $\nu(\{t\}) = 0$ for all $t \in [0,1]$ then Π is continuous on $[0,1]$.*

To deduce the continuity of an optimal control, let us use standard Fenchel duality:

$$L^*(t,x,p) := \sup_{v \in \mathbb{R}^d} \{p \cdot v - L(t,x,v)\}, \ \forall p \in \mathbb{R}^d.$$

Assuming that L^* is differentiable with respect to p (which is the case if L is strictly convex and superlinear in v), we then have:

$$\overline{u}(t) = \nabla_p L^*(t, \overline{y}(t), \Pi(t)).$$

We thus deduce the following:

Proposition 4. *Let us further assume that $L(t,x,.)$ is strictly convex for all $(t,x) \in [0,1] \times \mathbb{R}^d$ and that $\nabla_p L^*$ is continuous on $[0,1] \times \mathbb{R}^d \times \mathbb{R}^d$ and let $\overline{u}$ be any solution of (8). For every $t \in (0,1)$ such that $\nu(\{t\}) = 0$ then $\overline{u}$ is continuous at t. In particular, if ν has finitely many atoms then $\overline{u}$ is piecewise continuous.*

Proposition 4 states that optimal controls are continuous except possibly on the set of atoms of ν. Remarking that, for every $\tau \in [0,1]$,

$$\nu(\{\tau\}) = \int_0^1 \nu_t(\{\tau\})\mathrm{d}t,$$

we thus deduce that optimal controls are continuous at each point τ such that $\mathcal{L}^1(\{t \in [0,1] \ : \ \nu_t(\{\tau\}) > 0\}) = 0$. As a consequence, if for Lebesgue-*a.e.* t, ν_t is atomless then any optimal control is continuous. In the case of a deviating argument (*i.e.* $\nu_t = \delta_{\theta(t)}$), ν is the image of the Lebesgue measure by the deviation function θ and optimal controls are continuous at each point τ such that $\mathcal{L}^1(\theta^{-1}(\{\tau\})) = 0$. If ν has no atoms, optimal controls are continuous so that, if, in addition, we assume that $t \mapsto \langle g, \nu_t \rangle$ is continuous for every continuous g then the corresponding state is of class C^1. Under stronger regularity assumptions on ν, f, L and L^*, one can obtain higher regularity (C^1 or even C^∞). When ν has finitely many atoms, one can also obtain higher piecewise regularity (piecewise C^1 or even piecewise C^∞).

Example. To illustrate the previous results, let us consider the linear-quadratic case, namely:

$$L(t,x,v) = \frac{1}{2}(|x|^2 + |u|^2), \ g = 0, \ f(t,x) = A(t)x,$$

where A is a continuous $d \times d$-matrix-valued function. In this case, J is strictly convex and (8) admits a unique solution $\overline{u}$. If we denote by $\overline{y}$ the corresponding trajectory, the pair $(\overline{u}, \overline{y})$ is then characterized by the following linear system

$$\begin{cases} \dot{\overline{y}}(t) = \langle A(.)\overline{y}(.), \nu_t \rangle + \overline{u}(t), \ \overline{y}(0) = x_0, \\ \dot{\overline{u}} = \overline{y} - (A(.)^T \langle \overline{u}, \nu_.^* \rangle)\nu, \ \overline{u}(1) = 0. \end{cases}$$

It immediately follows that if ν is atomless (*i.e.* $\nu(\{s\}) = 0$ for all $s \in [0,1]$) and $t \mapsto \langle g, \nu_t \rangle$ is continuous for every continuous g, then $\overline{u}$ is continuous and $\overline{y}$ is of class C^1.

4.3. **A nonlocal version of Pontryagin principle**

In this section, we consider the case of a convex constraint on the control and prove that the optimality conditions take the form of a nonlocal version of the Pontryagin principle. More precisely, we consider the problem:

$$\inf_{u \in V} J(u) \text{ with } J(u) := \int_0^1 L(s, y_u(s), u(s))\mathrm{d}s + g(y_u(1)). \tag{20}$$

For the sake of simplicity we assume that the set of admissible controls V is defined by:

$$V := \{u \in L^p \ : \ u(t) \in K \text{ for } a.e. \ t\} \tag{21}$$

where $K := \{v \in \mathbb{R}^d \ : \ \Phi(v) \leq 0\}$ and Φ is a differentiable convex function such that $\inf_{\mathbb{R}^d} \Phi < 0$. In the remainder, we assume **(H1)**, **(H'1)**, **(H2)**, **(H3)**, **(H7)**, **(H8)** and **(H9)** and, in addition, that $L(t, x, .)$ is *strictly* convex on K for all $(t, x) \in [0, 1] \times \mathbb{R}^d$. This implies in particular that (20) possesses solutions and that the differentiability result of Lemma 3 holds. Under our assumptions, the optimality conditions of (20) may be expressed by the following multiplier rule:

Lemma 5. *Let $\overline{u}$ be a solution of* (20) *then there exists a nonnegative function β such that*

$$\beta(.)\nabla\Phi(\overline{u}(.)) \in L^{p'}, \ \beta(t)\Phi(\overline{u}(t)) = 0 \ a.e., \ and \tag{22}$$

$$\langle J'(\overline{u}), v \rangle = -\int_0^1 \beta(t)\nabla\Phi(\overline{u}(t)) \cdot v(t)\mathrm{d}t, \ \forall v \in L^p. \tag{23}$$

Proof. Let $F \in L^{p'}$ be such that $\langle J'(\overline{u}), v \rangle = \int_0^1 F \cdot v, \ \forall v \in L^p$. The variational inequalities of (20) read as:

$$\int_0^1 F \cdot v \geq 0, \ \forall v \in \overline{\mathbb{R}_+(V - \overline{u})} \tag{24}$$

where $\overline{\mathbb{R}_+(V - \overline{u})}$ denotes the L^p closure of the convex cone $\mathbb{R}_+(V - \overline{u})$. By Lusin's theorem, for all $n \in \mathbb{N}^*$, there exists a compact subset K_n of $[0, 1]$ whose complement has Lebesgue measure less than n^{-1} and $\overline{u}$ is continuous on K_n. Let us define then for all $\delta > 0$ and $n \in \mathbb{N}^*$:

$$I_0 := \{t \ : \ \Phi(\overline{u}(t)) = 0\}, \ I_\delta := \{t \ : \ \Phi(\overline{u}(t)) \leq -\delta\}, \ I_\delta^n = I_\delta \cap K_n, \ I_0^n = I_0 \cap K_n.$$

In what follows, for $A \subset [0, 1]$, $\mathbf{1}_A$ will denote the characteristic function of A. Let $v \in L^\infty$, there exists $\varepsilon > 0$ such that $\overline{u} + \varepsilon\mathbf{1}_{I_\delta^n}v$ and $\overline{u} - \varepsilon\mathbf{1}_{I_\delta^n}v$ both belong to V. Hence, by (24), we deduce that $\int_0^1 \mathbf{1}_{I_\delta^n}F \cdot v = 0$. Since v, $\delta > 0$ and n are arbitrary we get $F = 0$ *a.e.* on $[0, 1] \setminus I_0$.

Now let $v \in L^\infty$ be such that:

$$\nabla\Phi(\overline{u}(t)) \cdot v(t) \leq 0 \text{ for } a.e. \ t \in I_0. \tag{25}$$

It can be checked easily that for all $n \in \mathbb{N}^*$ and $\eta > 0$ there exists $\varepsilon > 0$ such that $\overline{u} + \varepsilon\mathbf{1}_{I_0^n}(v - \eta\nabla\Phi(\overline{u})) \in V$. Using (24) again, we get:

$$\int_0^1 \mathbf{1}_{I_0^n}F \cdot (v - \eta\nabla\Phi(\overline{u})) \geq 0$$

since n and η are arbitrary, we deduce that $\int_0^1 \mathbf{1}_{I_0}F \cdot v \geq 0$ for every $v \in L^\infty$ that satisfies (25). Since $F \in L^{p'}$, we deduce from Lebesgue's dominated convergence theorem that $\int_0^1 \mathbf{1}_{I_0}F \cdot v \geq 0$ for every $v \in L^p$ that satisfies (25). By a standard separation argument, we deduce that $\mathbf{1}_{I_0}F = -\mathbf{1}_{I_0}\beta\nabla\Phi(\overline{u})$ for some nonnegative function β. Together with $F = 0$ *a.e.* on $[0, 1] \setminus I_0$, this completes the proof. $\qquad\square$

Combining the previous lemma, with the arguments of the proof of Theorem 1, we get:

Proposition 5. *Let $\overline{u}$ be a solution of* (20), *β be as in Lemma* 5, *Π, Θ and A be defined by* (12) *and set*

$$\overline{p}(t) := -\Pi(t) - \beta(t)\nabla\Phi(\overline{u}(t)), \ \forall t \in (0,1), \tag{26}$$

then $\overline{p} \in L^{p'} \cap \mathrm{BV}((0,1), \mathbb{R}^d)$, satisfies in the $(W^{1,p})'$ sense:

$$\dot{\overline{p}} = -\Theta - (A(.)^T \langle \overline{p}(.), \nu_.^* \rangle)\nu, \tag{27}$$

with the boundary condition $\overline{p}(1) = \nabla g(\overline{y}(1))$.

Proof. Let $x \in W^{1,p}$ be such that $x(0) = 0$ and define:

$$v(t) := \dot{x}(t) - \langle A(.)x(.), \nu_t \rangle.$$

By Lemmas 3 and 5, we get:

$$\int_0^1 (\Theta \cdot x - \overline{p} \cdot (\dot{x} - \langle A(.)x(.), \nu_t \rangle)\mathrm{d}t + \nabla g(\overline{y}(1)) \cdot x(1) = 0. \tag{28}$$

Arguing exactly as in the proof of Theorem 1, we get the desired result. $\square$

We may actually interpret $\overline{p}$ as a co-state variable and rewrite the previous conditions as a nonlocal version of the Pontryagin principle. To that end, let us first define:

$$H(t,x,p) := \min_{v \in K}\{L(t,x,v) + p \cdot v\}, \ \forall(t,x,p) \in [0,1] \times \mathbb{R}^d \times \mathbb{R}^d. \tag{29}$$

Our assumptions guarantee that for each $(t,x,p) \in [0,1] \times \mathbb{R}^d \times \mathbb{R}^d$, there exists a unique $U(t,x,p) \in K$ such that

$$H(t,x,p) := L(t,x,U(t,x,p)) + p \cdot U(t,x,p).$$

It is well-known (see for instance Cannarsa and Sinestrari [4]), under our assumptions, that U is continuous and the partial gradients $\nabla_x H$ and $\nabla_p H$ exist, are continuous and are given by:

$$\nabla_x H(t,x,p) = \nabla_x L(t,x,U(t,x,p)), \ \nabla_p H(t,x,p) = U(t,x,p). \tag{30}$$

Let $v \in V$, by convexity of Φ, we have:

$$0 \geq \beta(t)(\Phi(v) - \Phi(\overline{u}(t))) \geq \nabla\Phi(\overline{u}(t)) \cdot (v - \overline{u}(t)). \tag{31}$$

Using the convexity of $L(t,\overline{y}(t), .)$ and the definition (26) of the co-state, we also have:

$$L(t,\overline{y}(t),v) \geq L(t,\overline{y}(t),\overline{u}(t)) - (\beta(t)\nabla\Phi(\overline{u}(t)) + \overline{p}(t)) \cdot (v - \overline{u}(t)). \tag{32}$$

Using (31) and the arbitrarity of v in (32), we then get:

$$L(t,\overline{y}(t),\overline{u}(t)) + \overline{p}(t) \cdot \overline{u}(t) = H(t,\overline{y}(t),\overline{p}(t))$$

or, equivalently:

$$\overline{u}(t) = U(t,\overline{y}(t),\overline{p}(t)) \ a.e. \ \text{on} \ (0,1). \tag{33}$$

Let us then define for all $x \in W^{1,p}$ and $p \in L^{p'}$ the (*functional* or *nonlocal*) Hamiltonian of problem (20) by:

$$\widetilde{H}(x,p)(t) := H(t,x(t),p(t)) + p(t) \cdot \langle f(.,x(.)), \nu_t \rangle, \ t \in (0,1). \tag{34}$$

We then have the following result whose elementary proof is left to the reader:

Lemma 6. *Let* $x \in W^{1,p}$ *and* $p \in L^{p'}$.

(1) *For every* $\varphi \in C^0$, *the following limit*

$$\left\langle \boldsymbol{D}_x \widetilde{H}(x,p), \varphi \right\rangle := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^1 \left(\widetilde{H}(x + \varepsilon\varphi, p)(t) - \widetilde{H}(x,p)(t) \right) \mathrm{d}t$$

exists and is equal to

$$\int_0^1 \left(\nabla_x H(t, x(t), p(t)) \cdot \varphi(t) + p(t) \cdot \langle D_x f(., x(.))\varphi(.), \nu_t \rangle \right) \mathrm{d}t =$$

$$\int_0^1 \nabla_x H(t, x(t), p(t)) \cdot \varphi(t)\mathrm{d}t + \int_0^1 \varphi(s) \cdot \left[D_x f(s, x(s))^T \langle p(.), \nu_s^* \rangle \right] \mathrm{d}\nu(s).$$

(2) *For every* $\varphi \in L^{p'}$, *the following limit*

$$\left\langle \boldsymbol{D}_p \widetilde{H}(x,p), \varphi \right\rangle := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^1 \left(\widetilde{H}(x, p + \varepsilon\varphi)(t) - \widetilde{H}(x,p)(t) \right) \mathrm{d}t$$

exists and is equal to

$$\left\langle \boldsymbol{D}_p \widetilde{H}(x,p), \varphi \right\rangle = \int_0^1 \left(\langle f(., x(.)), \nu_t \rangle + U(t, x(t), p(t)) \right) \cdot \varphi(t)\mathrm{d}t.$$

Let us remark that if $x \in W^{1,p}$ and $p \in L^{p'}$, then $\boldsymbol{D}_p \widetilde{H}(x,p) \in L^p$ and $\boldsymbol{D}_x \widetilde{H}(x,p)$ is a finite vector-valued measure. Slightly abusing notations, we may rewrite Lemma 6 in the more concise form:

$$\begin{cases} \boldsymbol{D}_x \widetilde{H}(x,p) = \nabla_x L(., x(.), U(., x(.), p(.))) + D_x f(., x(.))^T \langle p, \nu_.^* \rangle \nu, \\ \boldsymbol{D}_p \widetilde{H}(x,p)(t) = \langle f(., x(.)), \nu_t \rangle + U(t, x(t), p(t)). \end{cases}$$

Finally, we get the following version of the Pontryagin principle:

Theorem 2. *Let* $\overline{u}$ *be a solution of* (8) *and* $\overline{y} := y_{\overline{u}}$ *be the corresponding trajectory, then there exists* $\overline{p} \in L^{p'} \cap \mathrm{BV}((0,1), \mathbb{R}^d)$ *such that:*

(1) *For a.e.* $t \in (0,1)$,

$$L(t, \overline{y}(t), \overline{u}(t)) + \overline{p}(t) \cdot \overline{u}(t) = \min_{v \in K} \left\{ L(t, \overline{y}(t), v) + \overline{p}(t) \cdot v \right\};$$

(2) $(\overline{y}, \overline{p})$ *solves the nonlocal Hamiltonian system:*

$$\begin{cases} \dot{\overline{y}} = \boldsymbol{D}_p \widetilde{H}(\overline{y}, \overline{p}), \\ \dot{\overline{p}} = -\boldsymbol{D}_x \widetilde{H}(\overline{y}, \overline{p}) \end{cases}$$

together with the boundary conditions

$$\overline{y}(0) = x_0, \ \overline{p}(1) = \nabla g(\overline{y}(1)).$$

Proof. Let us define $\overline{p}$ as in Proposition 5, then by construction $\overline{u}(t) = U(t, \overline{y}(t), \overline{p}(t))$ for *a.e.* $t \in (0, 1)$, which means that

$$L(t, \overline{y}(t), \overline{u}(t)) + \overline{p}(t) \cdot \overline{u}(t) = \min_{v \in K} \{L(t, \overline{y}(t), v) + \overline{p}(t) \cdot v\}.$$

Using Lemma 6, the optimality condition (27) can be rewritten as:

$$\dot{\overline{p}} = -\mathbf{D}_x \widetilde{H}(\overline{y}, \overline{p}), \ \overline{p}(1) = \nabla g(\overline{y}(1)).$$

Finally, again using Lemma 6, the state equation (4) can be rewritten as

$$\dot{\overline{y}}(t) = \langle f(., \overline{y}(.), \nu_t) \rangle + U(t, \overline{y}(t), \overline{p}(t)) = \mathbf{D}_p \widetilde{H}(\overline{y}, \overline{p})(t). \qquad \square$$

Remark. For the sake of simplicity, we have assumed here that the set of controls K is convex. It is actually unnecessary for a Pontryagin principle to hold but is a sufficient condition for the existence of an optimal control. For a nonconvex K, if $\overline{u}$ is a solution of (8) and $\overline{y} := y_{\overline{u}}$ is the corresponding trajectory, then $\overline{u}$ also solves the relaxation of (8) which consists in minimizing $J(u)$ subject to the relaxed constraint $u \in \overline{\mathrm{co}}(K)$. In this case, $(\overline{y}, \overline{u})$ satisfies the optimality conditions of the relaxed problem for which the set of admissible controls is convex.

5. The case of a state equation with memory in the control

This final section is devoted to optimality conditions in the case of a state equation of the form (2) where the control also enters the dynamics in a nonlocal way.

5.1. **Assumptions and preliminaries**

In addition to f and the family $(\nu_t)_t$ (that, throughout are assumed to satisfy **(H1)**, **(H'1)**, **(H2)** and **(H3)**), we are now also given a measurable family of nonnegative measures $(\mu_t)_t$ such that $t \mapsto \mu_t([0, 1])$ belongs to L^1. For $u \in L^p$ and $x_0 \in \mathbb{R}^d$, let us consider the Cauchy problem:

$$\begin{cases} \dot{x}(t) = \langle f(., x(.)), \nu_t \rangle + \langle u, \mu_t \rangle, \ t \in (0, 1), \\ x(0) = x_0. \end{cases} \tag{35}$$

In order to be able to apply the results of Section 2 (with $t \mapsto \langle u, \mu_t \rangle$ as new control), it is enough that the linear map $u \mapsto (t \mapsto \langle u, \mu_t \rangle)$ is well-defined and continuous from L^p to L^1. Let us define $\eta := \mu_t \otimes \mathcal{L}^1$ and define μ as the second marginal of η, *i.e.*:

$$\int_0^1 \phi(s) \mathrm{d}\mu(s) = \int_0^1 \langle \phi, \mu_t \rangle \, \mathrm{d}t, \ \forall \phi \in C^0([0, 1], \mathbb{R}).$$

By the disintegration theorem (see [5] or [1]), we may also write $\eta = \mu \otimes \mu_s^*$ for some measurable family of probability measures $(\mu_s^*)_s$ on $[0, 1]$. Again, this means:

$$\int_0^1 \phi(t, s) \mathrm{d}\eta(t, s) = \int_0^1 \langle \phi(t, .), \mu_t \rangle \, \mathrm{d}t$$

$$= \int_0^1 \langle \phi(., s), \mu_s^* \rangle \, \mathrm{d}\mu(s), \ \forall \phi \in C^0([0, 1]^2, \mathbb{R}).$$

Now let $p \in (1, +\infty)$ and let p' be its conjugate exponent and let us assume that $\mu \in L^{p'}$ then for every $u \in C^0$ one has:

$$\int_0^1 |\langle u, \mu_t \rangle| \mathrm{d}t \le \int_0^1 |u| \mathrm{d}\mu \le \|\mu\|_{L^{p'}} \|u\|_{L^p}.$$

Thus, if $\mu \in L^{p'}$, $u \mapsto (t \mapsto \langle u, \mu_t \rangle)$ is a well-defined and continuous mapping from L^p to L^1. In the sequel we will always assume $\mu \in L^{p'}$ and, slightly abusing notations, we will also denote by μ the density of μ with respect to $\mathcal{L}^1$. In particular this implies that the Cauchy problem (35) admits a unique solution that will be denoted $z_u \in W^{1,1}$ in the sequel.

The next paragraphs then deal with the optimal control problem

$$\inf_{u \in V} I(u) \text{ with } I(u) := \int_0^1 L(s, z_u(s), u(s)) \mathrm{d}s + g(z_u(1)). \tag{36}$$

To derive optimality conditions, we further assume that f, L and g satisfy assumptions **(H7)**, **(H8)** and **(H9)** and that V is a closed convex subset of L^p.

Finally, for the sake of simplicity, we further assume that ν is atomless (*i.e.* $\nu(\{t\}) = 0$ for every $t \in [0,1]$) where we recall that ν is the second marginal of $\gamma := \nu_t \otimes \mathcal{L}^1$ as defined in Section 4.1. Still using the notations of Section 4.1, we shall also write $\gamma = \nu \otimes \nu_s^*$.

5.2. Euler-Lagrange equations

In this paragraph, we consider the unconstrained case, *i.e.* problem (36) when $V = L^p$. Let us assume that $\overline{u}$ solves (36) and let us set $\overline{z} := z_{\overline{u}}$, and define:

$$\Pi(t) := \nabla_v L(t, \overline{z}(t), \overline{u}(t)), \ A(t) := D_x f(t, \overline{z}(t)),$$
$$\Theta(t) := \nabla_x L(t, \overline{z}(t), \overline{u}(t)), \ \beta := \nabla g(\overline{z}(1)). \tag{37}$$

Using the continuity of $u \mapsto (t \mapsto \langle u, \mu_t \rangle)$ and Proposition 2, we easily obtain:

$$\int_0^1 (\Theta \cdot k_v + \Pi \cdot v) + \beta \cdot k_v(1) = 0, \ \forall v \in L^p \tag{38}$$

where $k_v \in W^{1,1}$ denotes the solution of the linearized equation:

$$\begin{cases} \dot{k}(t) = \langle A(.)k(.), \nu_t \rangle + \langle v(.), \mu_t \rangle, \ t \in (0,1), \\ k(0) = 0. \end{cases} \tag{39}$$

Now, let us consider the adjoint equation:

$$\begin{cases} \dot{p} = -A^T(.) \langle p, \nu_.^* \rangle \nu - \Theta, \\ p(1) = \beta \end{cases} \tag{40}$$

which can equivalently be rewritten as:

$$p(t) = \beta + \int_t^1 \Theta + \int_t^1 A^T(s) \langle p, \nu_s^* \rangle \mathrm{d}\nu(s).$$

Arguing as in Section 2, and using the fact that ν is atomless, the adjoint equation (40) admits a unique solution $\overline{p} \in \mathrm{BV} \cap C^0$. For every $h \in W^{1,1}$ such that $h(0) = 0$, let us also remark that the following integration by parts formula holds:

$$\langle h, \dot{\overline{p}} \rangle = -\int_0^1 \dot{h} \cdot \overline{p} + \beta \cdot h(1). \tag{41}$$

The Euler-Lagrange equations of (36) are then given by:

Theorem 3. *Let $\overline{u}$ be a solution of* (36) *and* $\overline{z} := z_{\overline{u}}$, *then for a.e.* t, *one has:*

$$\nabla_v L(t, \overline{z}(t), \overline{u}(t)) = - \langle \overline{p}(.), \mu_t^* \rangle \, \mu(t),$$

where $\overline{p} \in \mathrm{BV} \cap C^0$ *is the solution of the adjoint equation:*

$$\begin{cases} \dot{p} & = -\Lambda_p \nu - \nabla_x L(., \overline{z}(.), \overline{u}(.)), \\ p(1) = \nabla g(\overline{z}(1)) \end{cases} \tag{42}$$

and $\Lambda_p(t) := D_x f(t, \overline{z}(t))^T \langle p, \nu_t^* \rangle$.

Proof. Let $v \in L^p$, and let us use the notations of (37). Using (39) and (40), we then have:

$$\begin{aligned}
\langle k_v, \dot{\overline{p}} \rangle &= - \int_0^1 (A^T(s) \langle \overline{p}, \nu_s^* \rangle) \cdot k_v(s) \mathrm{d}\nu(s) - \int_0^1 \Theta \cdot k_v \\
&= - \int_0^1 (\langle A k_v, \nu_t \rangle) \cdot \overline{p}(t) \mathrm{d}t - \int_0^1 \Theta \cdot k_v \\
&= - \int_0^1 \dot{k}_v \cdot \overline{p} + \int_0^1 \langle v, \mu_t \rangle \cdot \overline{p}(t) \mathrm{d}t - \int_0^1 \Theta \cdot k_v \\
&= - \int_0^1 \dot{k}_v \cdot \overline{p} + \int_0^1 v(s) \langle \overline{p}, \mu_s^* \rangle \mu(s) \mathrm{d}s - \int_0^1 \Theta \cdot k_v
\end{aligned}$$

with (41), we then get:

$$\int_0^1 \Theta \cdot k_v + \beta \cdot k_v(1) = \int_0^1 v(s) \langle \overline{p}, \mu_s^* \rangle \mu(s) \mathrm{d}s, \ \forall v \in L^p.$$

Finally, (38) exactly yields $\Pi(t) = \nabla_v L(t, \overline{z}(t), \overline{u}(t)) = - \langle \overline{p}(.), \mu_t^* \rangle \mu(t)$ for *a.e.* t. □

5.3. **Pontryagin principle**

In this paragraph, we consider the constrained case where V is defined by (21) as in Section 4.3. We recall that $K := \{v \in \mathbb{R}^d \ : \ \Phi(v) \leq 0\}$ where Φ is a differentiable convex function such that $\inf_{\mathbb{R}^d} \Phi < 0$. As in Section 4.3, in addition to the general assumptions of this section, we assume that $L(t, x, .)$ is strictly convex on K for every $(t, x) \in [0, 1] \times \mathbb{R}^d$.

Let $(x, p) \in C^0 \times C^0$ and define

$$\mathcal{H}(x, p) := \inf_{u \in V} \int_0^1 (L(s, x(s), u(s)) + p(s) \cdot \langle u, \mu_s \rangle) \mathrm{d}s$$

and let us remark that this Hamiltonian can also be written as:

$$\begin{aligned}
\mathcal{H}(x, p) &= \inf_{u \in V} \int_0^1 (L(t, x(t), u(t)) + u(t) \cdot \langle p, \mu_t^* \rangle \mu(t)) \mathrm{d}t \\
&= \int_0^1 \min_{v \in K} \{L(t, x(t), v) + v \cdot \langle p, \mu_t^* \rangle \mu(t)\} \mathrm{d}t \\
&= \int_0^1 H(t, x(t), \langle p, \mu_t^* \rangle \mu(t)) \mathrm{d}t
\end{aligned}$$

where H is defined by (29). Define also:

$$\widetilde{\mathcal{H}}(x, p) := \mathcal{H}(x, p) + \int_0^1 p(t) \cdot \langle f(., x(.)), \nu_t \rangle \, \mathrm{d}t. \tag{43}$$

For fixed $(x,p) \in C^0 \times C^0$, our assumptions ensure that there exists a unique $u = U(x,p)$ that solves the minimization problem in the definition of $\mathcal{H}$ *i.e.* $U(x,p)(t) \in K$ and:

$$L(t, x(t), U(x,p)(t)) + U(x,p)(t) \cdot \langle p, \mu_t^* \rangle \mu(t) = H(t, x(t), \langle p, \mu_t^* \rangle \mu(t)) \text{ for } a.e. \ t.$$

Moreover, $\widetilde{\mathcal{H}}$ is Gâteaux-differentiable over $C^0 \times C^0$, and for every $(x,p) \in C^0 \times C^0$ and $(y,q) \in C^0 \times C^0$, one can easily establish:

$$\left\langle \mathbf{D}_x \widetilde{\mathcal{H}}(x,p), y \right\rangle = \int_0^1 \nabla_x L(., x(.), U(x,p)(.)) \cdot y + \int_0^1 p(t) \cdot \langle D_x f(., x(.))y(.), \nu_t \rangle \, \mathrm{d}t, \tag{44}$$

and

$$\left\langle \mathbf{D}_p \widetilde{\mathcal{H}}(x,p), q \right\rangle = \int_0^1 (\langle f(., x(.)), \nu_t \rangle + \langle U(x,p), \mu_t \rangle) \cdot q(t) \mathrm{d}t \tag{45}$$

which, slightly abusing notations, can be rewritten in the more concise form:

$$\begin{cases} \mathbf{D}_x \widetilde{\mathcal{H}}(x,p) = \nabla_x L(., x(.), U(x,p)(.)) + D_x f(., x(.))^T \langle p, \nu_.^* \rangle \nu, \\ \mathbf{D}_p \widetilde{\mathcal{H}}(x,p)(t) = \langle f(., x(.)), \nu_t \rangle + \langle U(x,p)(.), \mu_t \rangle. \end{cases}$$

Optimality conditions for (36) when V is defined by (21) are given by the following variant of the Pontryagin principle.

Theorem 4. *Let $\overline{u}$ be a solution of (36) and $\overline{z} := z_{\overline{u}}$ be the corresponding trajectory, then there exists $\overline{p} \in C^0 \cap \mathrm{BV}$ such that:*

(1) *For a.e. $t \in (0,1)$,*

$$L(t, \overline{z}(t), \overline{u}(t)) + (\langle \overline{p}, \mu_t^* \rangle \mu(t)) \cdot \overline{u}(t) = \min_{v \in K} \{ L(t, \overline{z}(t), v) + (\langle \overline{p}, \mu_t^* \rangle \mu(t)) \cdot v \};$$

(2) *$(\overline{z}, \overline{p})$ solves the nonlocal Hamiltonian system:*

$$\begin{cases} \dot{\overline{z}} = \mathbf{D}_p \widetilde{\mathcal{H}}(\overline{z}, \overline{p}), \\ \dot{\overline{p}} = -\mathbf{D}_x \widetilde{\mathcal{H}}(\overline{z}, \overline{p}) \end{cases}$$

(with $\widetilde{\mathcal{H}}$ defined by (43)) together with the boundary conditions

$$\overline{z}(0) = x_0, \ \overline{p}(1) = \nabla g(\overline{z}(1)).$$

Proof. Let us again use the notations (37). Arguing as in Lemma 5, there exists a nonnegative function β such that, $\beta(.)\nabla \Phi(\overline{u}(.)) \in L^{p'}$, $\beta(.)\Phi(\overline{u}(.)) = 0$ *a.e.* and

$$\int_0^1 (\Theta \cdot k_v + \Pi \cdot v) + \beta \cdot k_v(1) = -\int_0^1 \beta(t) \nabla \Phi(\overline{u}(t)) \cdot v(t) \mathrm{d}t, \ \forall v \in L^p \tag{46}$$

where k_v denotes the solutions of the linearized equation (39). Now let, $\overline{p}$ be the solution of the adjoint equation (40). Arguing as in the proof of Theorem 3, we obtain:

$$\Pi(t) = \nabla_v L(t, \overline{z}(t), \overline{u}(t)) = -\langle \overline{p}, \mu_t^* \rangle \mu(t) - \beta(t) \nabla \Phi(\overline{u}(t)), \ a.e. \tag{47}$$

which by our convexity assumptions is equivalent to $\overline{u}(t)$ being for *a.e.* t, a solution of

$$\min_{v \in K} \{ L(t, \overline{z}(t), v) + (\langle \overline{p}, \mu_t^* \rangle \mu(t)) \cdot v \}.$$

We thus have $\overline{u} = U(\overline{z}, \overline{p})$, so that using (44) and (45), the state equation and the adjoint equations exactly take the form of the Hamiltonian system:

$$\begin{cases} \dot{\overline{z}} = \mathbf{D}_p \widetilde{\mathcal{H}}(\overline{z}, \overline{p}), \\ \dot{\overline{p}} = -\mathbf{D}_x \widetilde{\mathcal{H}}(\overline{z}, \overline{p}). \end{cases} \qquad\qquad \square$$

REFERENCES

[1] L. Ambrosio, *Lecture Notes on Optimal Transport Problems, Mathematical aspects of evolving interfaces, CIME Summer School in Madeira* **1812**. Springer (2003).

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

[3] R. Boucekkine, O. Licandro, L. Puch and F. del Rio, Vintage capital and the dynamics of the AK model. *J. Economic Theory* **120** (2005) 39–72.

[4] P. Cannarsa and C. Sinestrari, *Semiconcave Functions, Hamilton-Jacobi Equations and Optimal Control.* Birkhäuser (2004).

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

[6] M.E. Drakhlin and E. Stepanov, On weak lower-semi continuity for a class of functionals with deviating arguments. *Nonlinear Anal. TMA* **28** (1997) 2005–2015.

[7] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems, Classics in Mathematics.* Society for Industrial and Applied Mathematics, Philadelphia (1999).

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

[9] L. El'sgol'ts, *Introduction to the Theory of Differential Equations with Deviating Arguments.* Holden-Day, San Francisco (1966).

[10] F. Gozzi and C. Marinelli, Stochastic optimal control of delay equations arising in advertising models, in *Stochastic partial differential equations and applications* **VII**, Chapman & Hall, Boca Raton, *Lect. Notes Pure Appl. Math.* **245** (2006) 133–148.

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

[12] E. Jouini, P.-F. Koehl and N. Touzi, Optimal investment with taxes: an existence result. *J. Math. Econom.* **33** (2000) 373–388.

[13] M.N. Oguztöreli, *Time-Lag Control Systems.* Academic Press, New-York (1966).

[14] F.P. Ramsey, A mathematical theory of saving. *Economic J.* **38** (1928) 543–559.

[15] L. Samassi, *Calcul des variations des fonctionelles à arguments déviés.* Ph.D. thesis, University of Paris Dauphine, France (2004).

[16] L. Samassi and R. Tahraoui, Comment établir des conditions nécessaires d'optimalité dans les problèmes de contrôle dont certains arguments sont déviés? *C. R. Math. Acad. Sci. Paris* **338** (2004) 611–616.

[17] L. Samassi and R. Tahraoui, How to state necessary optimality conditions for control problems with deviating arguments? *ESAIM: COCV* (2007) e-first, doi: 10.1051/cocv:2007058.