Solving minimax control problems via nonsmooth optimization

Justina Gianatti[a,*], Laura S. Aragone[a], Pablo A. Lotito[b], Lisandro A. Parente[a]

[a] *CIFASIS-CONICET-UNR, Ocampo y Esmeralda, S2000EZP Rosario, Argentina*
[b] *PLADEMA-UNCPBA-CONICET, Paraje Arroyo Seco, B7000, Tandil, Argentina.*

Abstract

We address a minimax optimal control problems with linear dynamics. Under convexity assumptions, by using non-smooth optimization techniques, we derive a set of optimality conditions for the continuous-time case. We define an approximated discrete-time problem where analogous conditions hold. One of them allows us to design an easily implementable descent method. We analyze its convergence and we show some preliminary numerical results.

Keywords: Minimax control problems, Nonsmooth optimization, Optimality conditions, Numerical solutions

---

1. Introduction

We consider an optimal control problem whith linear dynamics and fixed initial state where the goal is to minimize a cost functional which is the essential supremum, over the time interval, of a function depending on the time, the state and the control. Studied in the last decades by several authors ([1], [2], [3], [4], [5], [6], [7], [8] and [9]) these problems differs from those with an accumulated cost criterium and arises naturally in many applications, as for instance, minimization of the maximum trajectory deviation from what is desired ([10], [11], [12]), or robust optimal control of uncertain systems ([13], [14]).

*Corresponding author

Email addresses: gianatti@cifasis-conicet.gov.ar (Justina Gianatti), aragone@cifasis-conicet.gov.ar (Laura S. Aragone), lotito@exa.unicen.edu.ar (Pablo A. Lotito), parente@cifasis-conicet.gov.ar (Lisandro A. Parente)

Certainly, by adding an auxiliary variable the minimax control problem can be written as a classical control problem with state constraints, in this framework, some authors ([2], [4], [5]) obtained necessary conditions as Pontryagin Maximum Principle [15]. Nevertheless, in this case the adjoint state involves Radon measures, and therefore it is not easily implementable. It is also relevant the dynamic programming approach ([1], [6], [7], [8], [9]), which usually requires the discretization of the state space for computational implementations, leading to large scale problems. Since in this work we consider fixed initial state, our approach only requires time discretization, avoiding dimensionality drawbacks.

The main idea of this paper is to consider the minimax control problem as a non-smooth optimization problem in a suitable space. We follow [16] but now we focus on the discrete-time approximation in order to develop a numerical scheme. Under suitable assumptions we prove the existence of the cost functional directional derivatives and we derive a set of first order optimality conditions from which we design a descent method following the Armijo's rule ([17]).

2. Continuous-Time Problem

2.1. Main Assumptions

We consider the dynamical system

$$\begin{cases} \dot{y}(t) = g(t, y(t), u(t)) \quad t \in [0, T], \\ y(0) = x \in \mathbb{R}^r, \end{cases} \tag{2.1}$$

where $g : [0, T] \times \mathbb{R}^r \times \mathbb{R}^m \to \mathbb{R}^r$ is a given function. In the notation above $y_u(t) \in \mathbb{R}^r$ denotes the state function and $u(t) \in \mathbb{R}^m$ the control. The optimal control problem consists in minimizing the functional $J : \mathcal{U} \mapsto \mathbb{R}$ defined as

$$J(u) := \operatorname{ess\,sup} \left\{ f(y_u(t)) : t \in [0, T] \right\}, \tag{2.2}$$

over the set of controls

$$\mathcal{U} = \{u : [0, T] \to U \subset \mathbb{R}^m : u(\cdot) \text{ measurable}\},$$

where U is a compact and convex set and $f : \mathbb{R}^r \to \mathbb{R}$ is given.

Let us now fix the standing assumptions that we will consider in this paper:

(H1) g is linear and has the form:

$$g(t, y(t), u(t)) = A(t)y(t) + B(t)u(t) + C(t)$$

where $A : [0, T] \to \mathbb{R}^{r\times r}$, $B : [0, T] \to \mathbb{R}^{r\times m}$ and $C : [0, T] \to \mathbb{R}^r$ are Lipschitz continuous functions.

(H2) f is convex, Lipschitz continuous and continuously differentiable.

Remark 2.1. *Under the above assumptions, for any $u \in \mathcal{U}$ the state equation* (2.1) *admits a unique solution y_u. Also, the function J is well defined, and the essential supremum is actually the maximum over $[0, T]$.*

2.2. Optimality Conditions

We consider the problem as a nonlinear optimization problem in $L^2[0, T]$. Note that if assumption **(H1)** holds and f is convex, then J is a convex function of u. If in addition f is a Lipschitz continuous function, then J is Lipschitz continuous on $\mathcal{U}$ endowed with the $L^2[0, T]$ norm. So the optimal control problem has solution, since we are minimizing a continuous and convex function over a convex, closed and bounded set of $L^2[0, T]$ (see [18]).

In order to obtain a necessary condition for u to be optimal, we would like to compute the gradient or, at least, a directional derivative of J for u along an admissible direction v. It is easy to see that because of the involved definition of J, it could not exist. Nevertheless, our assumptions on f guarantee the directional differentiability of J.

In the reminder we will note $J'(u, v)$ the directional derivative of the function J in u over the direction v, and by differentiable we understand Fréchet differentiable (see [19]). From now on, we suppose that assumptions **(H1)** and **(H2)** hold.

We denote by $T_{\mathcal{U}}(u)$ the tangent cone to $\mathcal{U}$ at u ([19]).

Proposition 2.1. *Under the above assumptions, the function J is directionally differentiable at any $u \in \mathcal{U}$ and the directional derivative in a direction $v \in T_{\mathcal{U}}(u)$ is given by*

$$J'(u;v) = \sup_{t \in C_u} \langle \nabla f(y_u(t)), z_v(t) \rangle, \tag{2.3}$$

where C_u is the set of critical times

$$C_u = \arg\max_{t \in [0,T]} f(y_u(t)), \tag{2.4}$$

and z_v solves the following differential equation

$$\begin{cases} \dot{z}(t) = & A(t)z(t) + B(t)v(t), \qquad t \in [0,T] \\ z(0) = & 0. \end{cases} \tag{2.5}$$

Proof. By the convexity and differentiability of f, from [19] we know that J is directionally differentiable and for any direction v,

$$J'(u,v) = \sup_{t \in C_u} D_u f(y_u(t))v.$$

Let ϕ be the application $u \mapsto y_u$, then $D_u f(y_u(t))v = \langle \nabla f(y_u(t)), \phi'(u,v) \rangle$, and

$$\phi'(u,v) = z_v,$$

where z_v is the solution of (2.5). □

By classical continuous optimization, we know that a necessary optimality condition for u to be optimal is that every directional derivative is non-negative, for every direction in $T_{\mathcal{U}}(u)$ ([20]). This condition turns to be also sufficient in the convex case. The last assertion is equivalent to

$$\inf_{v \in T_{\mathcal{U}}(u)} \sup_{t \in C_u} \langle \nabla f(y_u(t)), z_v(t) \rangle \geq 0. \tag{2.6}$$

Let us explicit the linear operator $v \mapsto z_v$. By the variation of constants formula, the solution of (2.5) is given by

$$z_v(t) = \int_0^t S_{ts} B(s) v(s) \mathrm{d}s,$$

where the matrix S_{ts} is a solution of the system

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} S_{ts} = & A(t) S_{ts}, \quad t \in [s, T] \\ S_{ss} = & I. \end{cases} \tag{2.7}$$

Now, the directional derivative can be written as

$$J'(u; v) = \sup_{t \in C_u} \left\langle \nabla f(y_u(t)), \int_0^t S_{ts} B(s) v(s) \mathrm{d}s \right\rangle. \tag{2.8}$$

Defining for each $u \in \mathcal{U}$ and $t \in [0, T]$, the element of $L^2[0, T]$

$$q_{u,t}(s) := I_t(s) B^\top(s) S_{ts}^\top \nabla f(y_u(t)), \quad \forall s \in [0, T],$$

where $I_t(s)$ is equal to 1 if $s \leq t$ and 0 otherwise, we can rewrite (2.8) as

$$J'(u; v) = \sup_{t \in C_u} \langle q_{u,t}, v \rangle, \tag{2.9}$$

where the last scalar product is in $L^2[0, T]$.

Theorem 2.1. *Let $u \in \mathcal{U}$, then u is optimal if and only if*

$$\inf_{v \in \mathcal{U} - u} \sup_{t \in C_u} \langle q_{u,t}, v \rangle = 0. \tag{2.10}$$

Proof. If u is a minimizer of J, then $\inf_{v \in T_{\mathcal{U}}(u)} J'(u; v) \geq 0$. By (2.9), the last assertions is equivalent to

$$\inf_{v \in T_{\mathcal{U}}(u)} \sup_{t \in C_u} \langle q_{u,t}, v \rangle \geq 0. \tag{2.11}$$

By **(H1)-(H2)**, we can deduce that $q_{u,t}$ is bounded in $L^2[0, T]$ independently of t. Since $\mathcal{U}$ is convex, the infimum over $T_{\mathcal{U}}(u)$ in (2.11) coincides with the infimum over the set $\mathcal{U} - u$. Since $v = 0$ is an admissible direction, we have

$$\inf_{v \in \mathcal{U} - u} \sup_{t \in C_u} \langle q_{u,t}, v \rangle = 0.$$

Conversely, the sufficiency is straightforward from the convexity and directional differentiability of J (see [20]).

□

Condition (2.10) involves the computation of the set of critical times associated to u. The dependence of this set with respect to the control can cause some troubles in the aim of designing an algorithm based on that condition. In order to avoid this complication, we propose other necessary conditions where the supremum is taken over the whole interval $[0, T]$.

In the reminder, we denote $\mathcal{U}_u := \mathcal{U} - u$ the set of admissible directions.

Theorem 2.2. *Condition* (2.10) *implies*

$$\inf_{v \in \mathcal{U}_u} \sup_{t \in [0,T]} \{ f(y_u(t)) - J(u) + \langle q_{u,t}, v \rangle \} = 0. \tag{2.12}$$

Also, condition (2.12) *implies*

$$\inf_{v \in \mathcal{U}_u} \sup_{t \in [0,T]} \langle q_{u,t}, v \rangle = 0, \tag{2.13}$$

and for any $\rho > 0$*,*

$$\inf_{v \in \mathcal{U}_u} \sup_{t \in [0,T]} \{ f(y_u(t)) - J(u) + \langle q_{u,t}, v \rangle \} + \frac{\rho}{2} \|v\|^2 = 0. \tag{2.14}$$

Proof. Note that $v = 0$ is an admissible direction, so

$$0 = \inf_{v \in \mathcal{U}_u} \sup_{t \in C_u} \langle q_{u,t}, v \rangle \leq \inf_{v \in \mathcal{U}_u} \sup_{t \in [0,T]} \langle q_{u,t}, v \rangle \leq 0. \tag{2.15}$$

Since $f(y_u(t)) \leq J(u)$ for all $t \in [0, T]$, we obtain

$$\inf_{v \in \mathcal{U}_u} \sup_{t \in [0,T]} \{ f(y_u(t)) - J(u) + \langle q_{u,t}, v \rangle \} \leq \inf_{v \in \mathcal{U}_u} \sup_{t \in [0,T]} \langle q_{u,t}, v \rangle . \tag{2.16}$$

By (2.15) and (2.16), the left hand side of (2.12) is non positive. On the other hand, by the definition of C_u,

$$\begin{aligned} \sup_{t \in C_u} \langle q_{u,t}, v \rangle &= \sup_{t \in C_u} \{ f(y_u(t)) - J(u) + \langle q_{u,t}, v \rangle \} \\ &\leq \sup_{t \in [0,T]} \{ f(y_u(t)) - J(u) + \langle q_{u,t}, v \rangle \} , \end{aligned}$$

so condition (2.10) gives the opposite inequality in (2.12).

By (2.16), clearly (2.12) implies (2.13). Also, using the fact that

$$f(y_u(t)) - J(u) + \langle q_{u,t}, v \rangle \leq f(y_u(t)) - J(u) + \langle q_{u,t}, v \rangle + \frac{\rho}{2} \|v\|^2 ,$$

we deduce that (2.12) implies (2.14). $\square$

3. The Discrete-Time Approximated Problem

In this section we approximate the continuous problem by a discrete-time optimal control problem and we state optimality conditions which are suitable for designing an algorithm.

3.1. Description of the Discrete-Time Problem

We divide the interval $[0, T]$ into N subintervals with common length $h = T/N$ and we restrict the controls to be sectionally constant. So, the set of discrete controls is

$$\mathcal{U}^h = \{u \in \mathcal{U} : u \text{ is constant in } [kh, (k+1)h),\ k = 0, ..., N-1\}. \tag{3.1}$$

A discrete policy u is identified as $\{u^n\}_{n=0}^{N-1}$, $u^n \in U \subset \mathbb{R}^m$, so $\mathcal{U}^h$ can be identified as $U^N \subset \mathbb{R}^{m \times N}$.

Instead of dealing with the exact response of the system (2.1) to the controls (3.1), we introduce an approximated discrete-time system. For $u \in \mathcal{U}^h$ we define the response y_u of the discrete-time system by the recursive formula

$$\begin{cases} y_u^{n+1} = & y_u^n + hg(t_n, y_u^n, u^n), \quad n = 0, ..., N-1, \\ y_u^0 = & x, \end{cases} \tag{3.2}$$

where $g(t_n, y_u^n, u^n) = A(t_n)y_u^n + B(t_n)u^n + C(t_n)$. The discrete-time optimal control problem consists in minimizing the functional $J^h : \mathcal{U}^h \to \mathbb{R}$, given by

$$J^h(u) := \max \{f(y_u^n) : n = 0, ..., N\}$$

over the set of controls $\mathcal{U}^h$. Clearly, the minimization problem has a solution, since J^h is continuous over the compact set $\mathcal{U}^h$.

3.2. Discrete Optimality Conditions

In the same fashion that in the continuous case we obtain:

Proposition 3.1. *Given a discrete policy $u = \{u^n\}_{n=0}^{N-1}$, the functional J^h is directionally differentiable at u and for any $v \in T_{\mathcal{U}^h}(u)$ we have*

$$J^{h\prime}(u, v) = \max_{n \in C_u} \langle \nabla f(y_u^n), z_v^n \rangle$$

where $C_u = \arg\max \{f(y_u^n) : 0 \le n \le N\}$ *is the set of critical times, and* z_v *solves the following system of difference equations*

$$\begin{cases} z^{n+1} = & a^n z^n + b^n v^n \quad n = 0, \dots, N-1, \\ z^0 = & 0 \end{cases} \tag{3.3}$$

with $v^n := v(t_n)$, $a^n := I + hA(t_n)$ *and* $b^n := hB(t_n)$.

The solution of (3.3) can be written as a function of v, in fact

$$z_v^n = \sum_{j=0}^{n-1} S_{n-1,j} v^j \tag{3.4}$$

where S satisfies $S_{n+1,j} = a^{n+1} S_{n,j},\ 0 \le j \le n$ and $S_{jj} = b^j,\ \forall j \ge 0$.

Analogously to the continuous case, if we define

$$q_{u,n}^j := \begin{cases} 0 & \forall j \ge n, \\ S_{n-1,j}^\top \nabla f(y_u^n) & \forall j < n, \end{cases} \tag{3.5}$$

we can conclude

$$J^{h'}(u; v) = \max_{n \in C_u} \sum_{j=0}^{n-1} \left\langle q_{u,n}^j, v^j \right\rangle = \max_{n \in C_u} \sum_{j=0}^{N-1} \left\langle q_{u,n}^j, v^j \right\rangle = \max_{n \in C_u} \langle q_{u,n}, v \rangle ,$$

where $q_{u,n}$ is the matrix with columns $q_{u,n}^j$, v is identified with the matrix of columns v^j, and the last scalar product is defined as $\langle q_{u,n}, v\rangle := tr(q_{u,n}^\top v)$.

A straightforward consequence of the previous remarks is the following theorem.

Theorem 3.1. *Let* $u \in \mathcal{U}^h$ *and* $\mathcal{U}_u^h := \mathcal{U}^h - u$. *Then* u *is an optimal control for the discrete-time control problem if and only if*

$$\min_{v \in \mathcal{U}_u^h} \max_{n \in C_u} \langle q_{u,n}, v \rangle = 0. \tag{3.6}$$

We propose now an analogous version of the optimality condition (2.12) which is not only necessary, but also sufficient.

Theorem 3.2. *Condition* (3.6) *is equivalent to*

$$\min_{v \in \mathcal{U}_u^h} \max_{n=0,..,N} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, v \rangle \right\} = 0. \tag{3.7}$$

Proof. Similarly to the proof of Theorem 2.2 we can prove (3.6) implies (3.7).

Assume that (3.7) holds. If we suppose that (3.6) does not hold, then there exists a direction $w \in \mathcal{U}_u^h$ such that $\max_{n \in C_u} \ \langle q_{u,n}, w \rangle < 0$, and then for all $\lambda \in (0, 1)$,

$$\max_{n \in C_u} \ \langle q_{u,n}, \lambda w \rangle < 0. \tag{3.8}$$

Since $\mathcal{U}^h$ is convex, λw is an admissible direction for any $\lambda \in (0, 1)$.

By (3.7), we have

$$\max_{n=0,..,N} \left\{ f(y_u^h(t_n)) - J^h(u) + \langle q_{u,n}, \lambda w \rangle \right\} \geq 0, \quad \forall \lambda \in (0, 1).$$

By (3.8), necessarily we have

$$\max_{n \notin C_u} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, \lambda w \rangle \right\} \geq 0, \quad \forall \lambda \in (0, 1),$$

taking $\lambda \downarrow 0$ we obtain

$$\max_{n \notin C_u} \left\{ f(y_u^n) - J^h(u) \right\} \geq 0,$$

which is in contradiction with the definition of C_u. □

Again, we add a quadratic term in order to obtain a strongly convex function which has better numerical behavior.

Theorem 3.3. *Let $u \in \mathcal{U}^h$, then for any $\rho > 0$, condition* (3.6) *is equivalent to*

$$\min_{v \in \mathcal{U}_u^h} \max_{n=0,\ldots,N} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, v \rangle \right\} + \frac{\rho}{2} \|v\|^2 = 0. \tag{3.9}$$

Proof. Assume condition (3.6) holds. By the definition of C_u we have,

$$\begin{aligned} \max_{n \in C_u} \langle q_{u,n}, v \rangle \quad &= \quad \max_{n \in C_u} \ \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, v \rangle \right\} \\ &\leq \quad \max_{n=0,\ldots,N} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, v \rangle \right\} + \tfrac{\rho}{2} \|v\|^2 , \end{aligned} \tag{3.10}$$

since $v = 0$ is an admissible direction, it is easy to see that (3.6) implies (3.9).

Now, we consider that (3.9) holds true. If we suppose that (3.6) does not hold then, there exist $w \in \mathcal{U}_u^h$ such that $\max_{n \in C_u} \langle q_{u,n}, w \rangle < 0$. Let $\bar{\lambda} \in (0, 1)$ small enough such that for all $\lambda \in (0, \bar{\lambda})$ we have

$$\max_{n \in C_u} \langle q_{u,n}, w \rangle + \frac{\rho}{2} \lambda \|w\|^2 < 0. \tag{3.11}$$

Thus, for all $\lambda \in (0, \bar{\lambda})$ we obtain,

$$\lambda \left(\max_{n \in C_u} \langle q_{u,n}, w \rangle + \frac{\rho}{2} \lambda \|w\|^2 \right) = \max_{n \in C_u} \langle q_{u,n}, \lambda w \rangle + \frac{\rho}{2} \|\lambda w\|^2 < 0. \qquad (3.12)$$

Since $\mathcal{U}^h$ is convex, $\lambda w \in \mathcal{U}_u^h$ for any $\lambda \in (0, \bar{\lambda})$. By (3.9), we obtain

$$\max_{n=0,..,N} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, \lambda w \rangle \right\} + \frac{\rho}{2} \|\lambda w\|^2 \geq 0, \quad \forall \lambda \in (0, \bar{\lambda})$$

so, by using (3.11) we can conclude as in the previous theorem. □

Remark 3.1 (Convergence of Value)**.** *In order to analyze the relationship between the continuous and the discrete problems, we define, respectively, the value for the continuous-time problem and the value for the discrete-time problem as*

$$V := \inf_{u \in \mathcal{U}} J(u), \qquad V^h := \inf_{u \in \mathcal{U}^h} J^h(u).$$

It is proved in [9] that $\left|V - V^h\right| \leq M\sqrt{h}$, *for some constant* $M > 0$, *and this estimation is independent of the initial point* x. *Following the proof in [9], it is easy to see that for a minimizer of the discrete problem* $\bar{u}^h$, *we obtain*

$$\left|V - J(\bar{u}^h)\right| \leq M\sqrt{h}.$$

This implies that an optimal control for the discrete problem gives a good approximation of the value of the continuous problem.

4. Approximation Scheme

In this section we present an approximation scheme based on condition (3.9). An admissible control satisfying this condition is optimal; otherwise, the minimizer in (3.9) gives a descent direction for the minimization of functional J^h. Indeed, let $\theta : U^N \to \mathbb{R}$ and $\eta : U^N \to \mathbb{R}^{m \times N}$ (where we identify $\mathcal{U}^h \equiv U^N$, $\mathcal{U}_u^h \equiv U_u^N$) be given by

$$\theta(u) := \min_{v \in U_u^N} \max_{n=0,\dots,N} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, v \rangle \right\} + \frac{\rho}{2} \|v\|^2, \qquad (4.1)$$

$$\eta(u) := \operatorname*{arg\,min}_{v \in U_u^N} \max_{n=0,\dots,N} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, v \rangle \right\} + \frac{\rho}{2} \|v\|^2. \qquad (4.2)$$

Then

$$
\begin{aligned}
J^{h\prime}(u;\eta(u)) &= \max_{n\in C_u} \langle q_{u,n}, \eta(u)\rangle \\
&\le \max_{n=0,\ldots,N} \left\{ f(y_u^n) - J^h(u) + \langle q_{u,n}, \eta(u)\rangle \right\} + \frac{\rho}{2}\left\|\eta(u)\right\|^2 \\
&= \theta(u) < 0.
\end{aligned}
\tag{4.3}
$$

First, we propose a conceptual descent algorithm, where it is assumed that all the computations can be done exactly and all the subproblems can be solved without error, and we give a theoretical convergence result. Secondly, we consider an implementable version of this algorithm, taking into account the inaccuracies, and we prove that a natural stopping criterion can be satisfied in finite time if the subproblems are solved with enough precision.

Basically, Algorithm 4.1 computes at each step a descent direction solving (3.9) and performs an Armijo line search. Using condition (3.9) has two main advantages. On the one hand, the supremum is computed over the whole set of times. The application $u \mapsto C_u$ is not always continuous as a set-valued function, which is a drawback in the aim to obtain convergence properties. On the other hand, the quadratic term in (3.9) regularizes the operator to be minimized, which turns to be strongly convex so it has unique solution. These facts imply that the functions θ and η are continuous (see [17]).

Algorithm 4.1. (Conceptual Algorithm)

Step 1: *Choose the parameters* $\alpha, \beta \in (0,1)$ *and* $\rho > 0$. *Set* $k = 1$ *and choose the initial point* $u_1 \in U^N$.

Step 2: *Compute:*

$$
\begin{aligned}
&y_{u_k}^n,\ f(y_{u_k}^n),\ \ n = 0, ..., N, \\
&J^h(u_k) = \max_{n=0,..,N} f(y_{u_k}^n).
\end{aligned}
$$

Step 3: *Compute* $\theta(u_k)$ *and* $\eta(u_k)$ *given by* (4.1) *and* (4.2), *respectively.*

Step 4: *If* $\theta(u_k) = 0$, *Stop* (u_k *satisfies the optimality condition). Else, find the maximum* $\lambda_k = \beta^j,\ \ j \in \mathbb{N}_0$, *such that*

$$
J^h(u_k + \lambda_k\, \eta(u_k)) < J^h(u_k) + \alpha\lambda_k\, \theta(u_k).
$$

Step 5: *Set* $u_{k+1} = u_k + \lambda_k\, \eta(u_k)$, $k = k+1$ *and restart Step 2.*

In practice, only a finite number of arithmetical operations and functions evaluations can be done (inexactly in most of the cases) and we can only achieve approximate solutions of the problems (4.1) and (4.2). So the stopping criterion in Step 4 of algorithm 4.1 is no longer meaningful and a natural replacement is $|\theta(u_k)| \le \varepsilon$ for some positive tolerance ε. Nevertheless, the nature of the problems (4.1) and (4.2) allows to obtain very accurate solutions. Note that, by introducing an auxiliary variable $\xi \in \mathbb{R}$, we obtain the equivalent quadratic program

$$\begin{aligned} & \min \tfrac{\rho}{2} \|v\|^2 + \xi \\ \text{s.t.} \quad & v \in U_u^N, \xi \in \mathbb{R}, \\ & \xi \ge f(y_u^n) + \langle q_{u,n}, v \rangle, \quad n = 0, ..., N, \end{aligned} \tag{4.4}$$

which can be solved efficiently by standard algorithms.

Algorithm 4.2. (Implementable Algorithm)

Step 1: *Choose the parameters* $\alpha, \beta \in (0,1)$, $\rho > 0$ *and the tolerance* $\varepsilon > 0$. *Set* $k = 1$ *and choose the initial point* $u_1 \in U^N$.

Step 2: *Compute:*

$$\begin{aligned} & y_{u_k}^n,\ f(y_{u_k}^n), \quad n = 0, ..., N, \\ & J^h(u_k) = \max_{n=0,..,N} f(y_{u_k}^n). \end{aligned}$$

Step 3: *Obtain approximations* $\hat{\theta}(u_k)$ *and* $\hat{\eta}(u_k)$ *of* (4.1) *and* (4.2), *respectively, i.e.,*

$$\hat{\theta}(u_k) = \max_{n=0,...,N} \left\{ f(y_{u_k}^n) - J^h(u_k) + \langle q_{u_k,n}, \hat{\eta}(u_k) \rangle \right\} + \frac{\rho}{2} \|\hat{\eta}(u_k)\|^2 ,$$

with

$$\hat{\theta}(u_k) = \theta(u_k) + e_\theta^k, \qquad \hat{\eta}(u_k) = \eta(u_k) + e_\eta^k, \tag{4.5}$$

where $e_\theta^k \in \mathbb{R}$ *and* $e_\eta^k \in \mathbb{R}^N$ *are unknown errors.*

Step 4: *If* $|\hat{\theta}(u_k)| < \varepsilon$, *Stop. Else, find the maximum* $\lambda_k = \beta^j$, $j \in \mathbb{N}_0$, *such that*

$$J^h(u_k + \lambda_k\, \hat{\eta}(u_k)) < J^h(u_k) + \alpha \lambda_k\, \hat{\theta}(u_k).$$

Step 5: *Set* $u_{k+1} = u_k + \lambda_k \hat{\eta}(u_k)$*,* $k = k+1$ *and restart Step 2.*

Algorithm 4.2 is an inexact version of Algorithm 4.1 with a reachable stopping criterion. In this first approach, for simplicity we consider fixed parameters, in particular the regularization parameter ρ, but a variable one could be introduced without significant changes in the analysis. The importance of the quadratic term involving ρ is to prove theoretical convergence results. Probably, a second order analysis would give a more precise idea on how to choose ρ in order to improve the efficiency of the method, but this is out of the scope of the present work.

4.1. Convergence Results

Theorem 4.1. *Let* $\{u_k\}$ *be the sequence generated by the Algorithm 4.1. Then, either* $\{u_k\}$ *finishes at a minimizer or it is infinite and every accumulation point of* $\{u_k\}$ *is optimal.*

Proof. Suppose $\{u_k\}$ is infinite, so it has a subsequence $\{u_{k_n}\}$ converging to some point $\bar{u}$. The continuity of the functional J^h implies

$$J^h(u_{k_n}) \to J^h(\bar{u}).$$

Since $\{u_k\}$ is infinite, we know that $\theta(u_k) < 0$ for all $k \in \mathbb{N}$ and

$$J^h(u_{k+1}) < J^h(u_k) + \alpha \lambda_k \, \theta(u_k), \quad \forall k \in \mathbb{N}. \tag{4.6}$$

Therefore the sequence $\left\{ J^h(u_k) \right\}_{k \in \mathbb{N}}$ is monotone decreasing so the whole sequence converges to $J^h(\bar{u})$, and by (4.6) and the fact that $\lambda_k \theta(u_k) < 0$ for all $k \in \mathbb{N}$, we obtain

$$\lim_{k \to \infty} \lambda_k \theta(u_k) = 0. \tag{4.7}$$

By Theorem 3.3, to prove that $\bar{u}$ is optimal is equivalent to prove that $\theta(\bar{u}) = 0$. If we suppose that $\theta(\bar{u}) < 0$, then there exists $\bar{\lambda}$ the maximum of the form β^j such that

$$J^h(\bar{u} + \bar{\lambda} \eta(\bar{u})) < J^h(\bar{u}) + \alpha \bar{\lambda} \theta(\bar{u}). \tag{4.8}$$

Since J^h, η and θ are continuous, there exists $N_0 \in \mathbb{N}$ such that

$$J^h(u_{k_n} + \bar{\lambda}\eta(u_{k_n})) - J^h(u_{k_n}) - \alpha\bar{\lambda}\theta(u_{k_n}) < 0, \quad \forall n \geq N_0.$$

We can conclude that $\lambda_{k_n} \geq \bar{\lambda}$ for all $n \geq N_0$. Then by (4.7) we obtain

$$\lim_{n\to\infty} \theta(u_{k_n}) = 0.$$

But, since θ is continuous, that implies $\theta(\bar{u}) = 0$, which contradicts our assumption. So, we can deduce that $\theta(\bar{u}) = 0$, i.e. $\bar{u}$ is optimal. $\square$

Remark 4.1. *By the previous theorems, if $\mathcal{S}$ is the set of solutions, i.e.*

$$\mathcal{S} := \left\{ \bar{u} \in \mathcal{U}^h : J^h(\bar{u}) = \inf_{u\in\mathcal{U}^h} J^h(u) \right\},$$

and $\{u_k\}$ is the sequence generated by the algorithm, then

$$d(u_k, \mathcal{S}) \to 0, \tag{4.9}$$

where d is the Euclidean distance in $\mathbb{R}^{mN}$. In particular, if the set of solution is a singleton, as for instance in the strictly convex case, then the whole sequences converges to the minimizer.

Theorem 4.2. *Let $\{u_k\}$ be the sequence generated by the Algorithm 4.2 and suppose that in every iteration the subproblems* (4.1) *and* (4.2) *are solved with enough accuracy such that the errors verify*

$$|e_\theta^k| < \frac{(1-\alpha)\varepsilon}{4}, \qquad \|e_\eta^k\| < \frac{(1-\alpha)\varepsilon}{4L}, \tag{4.10}$$

where L is the Lipschitz constant of J^h. Then, $\{u_k\}$ is finite, i.e., there exists $K \in \mathbb{N}$ such that $|\hat{\theta}(u_K)| < \varepsilon$ and the algorithm stops.

Proof. Suppose that the sequence $\{u_k\}$ is infinite. Then,

$$\hat{\theta}(u_k) \leq -\varepsilon < 0 \tag{4.11}$$

so $\hat{\eta}(u_k)$ is a descent direction for all $k \in \mathbb{N}$ and the Armijo search in Step 4 of algorithm 4.2 is well defined.

By the same argument used in the proof of Theorem 4.1, we have a subsequence $\{u_{k_n}\}$ converging to some point $\bar{u}$ and the whole sequence satisfies

$$J^h(u_k) \to J^h(\bar{u}) \text{ and } \lambda_k \hat{\theta}(u_k) \to 0.$$

Hence, (4.11) implies

$$\lambda_k \to 0. \tag{4.12}$$

Also, from (4.10) and (4.11) we have that, for all $k \in \mathbb{N}$, $\theta(u_k) \le -\frac{3}{4}\varepsilon$, so

$$\theta(\bar{u}) \le -\frac{3}{4}\varepsilon < 0 \tag{4.13}$$

and $\eta(\bar{u})$ is a descent direction. Thus, for $\bar{\alpha} = \frac{1+\alpha}{2} \in (0,1)$, there exists $\bar{\lambda} = \beta^{\bar{j}} > 0$ such that

$$J^h(\bar{u} + \bar{\lambda}\eta(\bar{u})) < J^h(\bar{u}) + \bar{\alpha}\bar{\lambda}\theta(\bar{u}),$$

and the continuity of J^h, η and θ gives

$$J^h(u_{k_n} + \bar{\lambda}\eta(u_{k_n})) - J^h(u_{k_n}) - \bar{\alpha}\bar{\lambda}\theta(u_{k_n}) < 0, \quad \forall n \ge N_0. \tag{4.14}$$

By (4.14), (4.5), (4.10) and the Lipschitz continuity of J^h, we obtain

$$J^h(u_{k_n} + \bar{\lambda}\hat{\eta}(u_{k_n})) - J^h(u_{k_n}) - \bar{\alpha}\bar{\lambda}\hat{\theta}(u_{k_n}) \le \bar{\lambda}\left(L\|e_\eta^k\| + \bar{\alpha}|e_\theta^k|\right) < \frac{(1-\alpha)\bar{\lambda}\varepsilon}{2},$$

and by the definition of $\bar{\alpha}$ and (4.11), we have

$$J^h(u_{k_n} + \bar{\lambda}\hat{\eta}(u_{k_n})) - J^h(u_{k_n}) - \alpha\bar{\lambda}\hat{\theta}(u_{k_n}) < \frac{(1-\alpha)\bar{\lambda}\varepsilon}{2} + (\bar{\alpha} - \alpha)\bar{\lambda}\hat{\theta}(u_{k_n}) \le 0.$$

Therefore, $\lambda_{k_n} \ge \bar{\lambda}$ for all $n \ge N_0$, which contradicts (4.12), so the sequence $\{u_k\}$ must be finite. □

5. Numerical Results

In this section, we illustrate the implementation of Algorithm 4.1 on a simple academic example studied in [4]. Specifically, consider the problem

$$\min_u \left\{ \max_{t \in [0,6]} \{y_1(t) + y_2(t)\} \right\},$$

subject to $(\dot{y}_1, \dot{y}_2)^\top = (y_2, u)^\top$, $(y_1(0), y_2(0))^\top = (2,2)^\top$ and $|u| \le 1$.

This problem admits infinitely many optimal controls, all of them satisfying $u(t) \equiv -1$ for $t \in [0,1)$, with optimal value 4.5, attained at $t = 1$ (see [4]). Our scheme reproduced these results in the performed numerical trials. Algorithm 4.2 has been coded by using Scilab 5.4.1 (INRIA-ENPC, see www.scilab.org) and ran on an Intel i7 2.67GHz with 8Gb of RAM. Each iteration comprises a quadratic program (solved using `quapro` toolbox) and an Armijo line search.

Table 1: Discrete value function and error

N	h	V^h	$\lvert V^h - V\rvert$	$\lvert\theta(u_k)\rvert$	Iter.	Time (s)
60	0.100000	4.550010	0.050010	9.85e-06	136	3.73
120	0.050000	4.525000	0.025000	3.42e-07	1	0.10
240	0.025000	4.512500	0.012500	1.71e-07	1	0.48
480	0.012500	4.506250	0.006250	8.56e-08	1	2.37
960	0.006250	4.503125	0.003125	4.28e-08	1	24.13
1920	0.003125	4.501563	0.001563	2.14e-08	1	235.25

Table 1 shows the results for 6 successive partitions of the interval $[0, 6]$, starting with $N = 60$ (so $h = 0.1$) and duplicating it in the following partitions. The used stopping test was $|\theta(u_k)| < \varepsilon_N$, with $\varepsilon_N = \frac{60}{N}10^{-5}$. We report on each case the optimal value, the error and the value of $|\theta(u_k)|$, as well as the required number of iterations and the computational time. The first trial (for $N = 60$) was initialized with an arbitrary constant and for the successive partitions the initial control was the linear interpolation of the control given by the previous trial. This choice significantly reduces the computational time, since the number of iterations is large only for the first trial, where the quadratic programs to be solved are small scale problems.

Figure 1 illustrates the evolution of the algorithm by showing the graphics of the function f on some iterations y_{u_k}. We can see how its maximum descends on successive iterations (because it is a descent method).

Figure 2 shows the obtained optimal control u and the function f. As

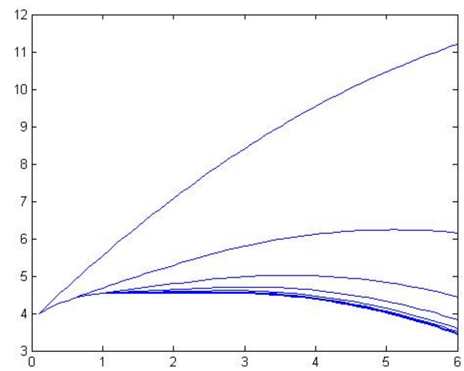

Figure 1: Iterations f

expected, the optimal control verifies $u(t) = 1$ for $t \in [0, 1)$, and f attains a maximum in $t = 1$ with value near 4.5.

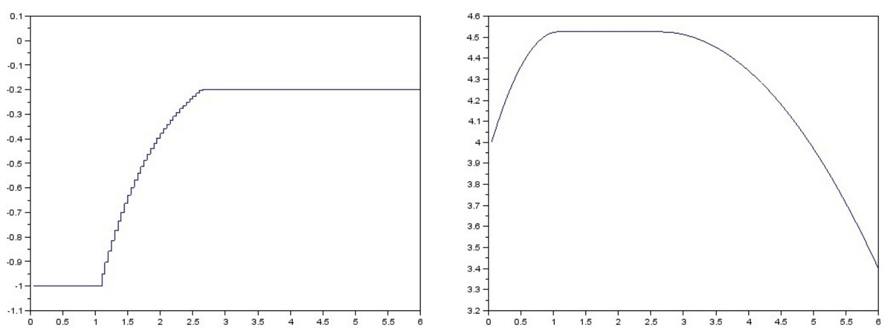

Figure 2: Optimal control u and function f

We point that in [4] the algorithm is based on an optimality condition derived from Pontryagin Maximum Principle [15], whose implementation requires to assume that the objective function has an isolated maximum. An advantage of our approach is the avoidance of such assumptions.

Acknowledgments

We thank the anonymous referee for her/his useful comments and suggestions.

This work has been partially supported by the following projects: PICT-2012-2212 ANPCyT, PIP 112-201101-286 CONICET and ING 475 UNR.

References

[1] E. N. Barron, H. Ishi, The Bellman equation for minimizing the maximum cost, Nonlinear Analysis, Theory, Methods and Applications 13 (9) (1989) 1067–1090. `doi:10.1016/0362-546X(89)90096-5`.

[2] E. N. Barron, The Pontryagin maximum principle for minimax problems of optimal control, Nonlinear Analysis, Theory, Methods and Applications 15 (12) (1990) 1155–1165. `doi:10.1016/0362-546X(90)90051-H`.

[3] E. N. Barron, R. Jensen, Relaxed minimax control, SIAM Journal on Control and Optimization 33 (4) (1995) 1028–1039. `doi:10.1137/S0363012993250530`.

[4] P. Lu, N. Vinh, Optimal control problems with maximum functional,, J. Guidance, 14 (6) (1991) 1215–1223. `doi:10.2514/3.20777`.

[5] R. B. Vinter, Minimax optimal control, sicon 44 (3) (2005) 939–968. `doi:10.1137/S0363012902415244`.

[6] L. S. Aragone, R. L. V. González, A Bellman's equation for minimizing the maximum cost, Indian J. Pure Appl. Math 31 (12) (2000) 1621–1632.

[7] L. S. Aragone, S. Di Marco, R. L. V. González, Numerical analysis of a minimax optimal control problem with an additive final cost, Mathematical Models and Methods in Applied Sciences 12 (2) (2002) 183–203. `doi:10.1142/S021820250200160X`.

[8] L. S. Aragone, S. Di Marco, R. L. V. González, Solution of a minimax problem with additive final cost, Mathematica Notae 43 (2005) 7–27.

[9] S. Di Marco, R. L. V. González, Minimax optimal control problems. Numerical analysis of the finite horizon case, ESAIM: Mathematical Modelling and Numerical Analysis 33 (1) (1999) 23–54. `doi:10.1051/m2an:1999103`.

[10] R. Bulirsch, F. Montrone, H. J. Pesch, Abort landing in the presence of windshear as a minimax optimal control problem, part 1: Necessary conditions, Journal of Optimization Theory and Applications 70 (1) (1991) 1–23. doi:10.1007/BF00940502.

[11] R. Bulirsch, F. Montrone, H. J. Pesch, Abort landing in the presence of windshear as a minimax optimal control problem, part 2: Multiple shooting and homotopy, Journal of Optimization Theory and Applications 70 (2) (1991) 223–254. doi:10.1007/BF00940625.

[12] H. J. Oberle, Numerical solution of minimax optimal control problems by multiple shooting technique, Journal of Optimization Theory and Applications 50 (2) (1986) 331–357. doi:10.1007/BF00939277.

[13] I. R. Petersen, M. R. James, P. Dupuis, Minimax optimal control of stochastic uncertain systems with relative entropy constraints, IEEE Transactions on Automatic Control 45 (3) (2000) 398–412. doi:10.1109/9.847720.

[14] V. Ugrinovskii, I. Petersen, Absolute stabilization and minimax optimal control of uncertain systems with stochastic uncertainty, SIAM journal on control and optimization 37 (4) (1999) 1089–1122. doi:10.1137/S0363012996309964.

[15] L. Pontryagin, V. Boltyanski, R. Gamkrelidze, E. Mishchenko, The Mathematical Theory of Optimal Processes, Gordon and Breach Science Publishers, New York, 1979.

[16] L. S. Aragone, P. A. Lotito, Numerical solution of a minimax problem using a specially designed necessary optimal condition, MACI 3 (2011) 675–678.

[17] E. Polak, Computational Methods in Optimization. A Unified Approach, Academic Press, New York, 1986.

[18] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, Springer Verlag, New York, 2010.

[19] J. F. Bonnans, A. Shapiro, Perturbation Analysis of Optimization Problems, Springer–Verlag, New York, 2000.

[20] J. F. Bonnans, Optimisation Continue, Dunod, Paris, 2006.