Quadratic Optimal Control of Linear Complementarity Systems: First order necessary conditions and numerical analysis

Alexandre Vieira, Bernard Brogliato, Christophe Prieur

Quadratic Optimal Control of Linear Complementarity Systems: First order necessary conditions and numerical analysis

Alexandre Vieira* Bernard Brogliato† Christophe Prieur‡

18th April 2018

Abstract

This article is dedicated to the analysis of quadratic optimal control of linear complementarity systems (LCS), which are a class of strongly nonlinear and nonsmooth dynamical systems. Necessary first-order conditions are derived, that take the form of an LCS with inequality constraints, hence are numerically tractable. Then two numerical solvers are proposed, for the direct and the indirect approaches. They take advantage of MPEC solvers for computations. Numerical examples illustrate the theoretical developments and demonstrate the efficiency of the approach.

1 Introduction

The objective of this article is to analyse the quadratic optimal control of Linear Complementarity Systems (LCS). More precisely, we wish to investigate properties and numerical resolution of the problem:

$$\min \; J(x,u,v) = \int_0^T \left(x(t)^\intercal Q x(t) + u(t)^\intercal U u(t)\right) dt, \tag{1}$$

$$\text{subject to} \begin{cases} \dot{x}(t) = Ax(t) + Bv(t) + Fu(t), \\ w(t) = Cx(t) + Dv(t) + Eu(t), \\ 0 \leq v(t) \perp w(t) \geq 0, \\ Mx(0) + Nx(T) = x_b, \end{cases} \tag{2}$$

where $T > 0$, $A, Q \in \mathbb{R}^{n\times n}$, $U, D, E \in \mathbb{R}^{m\times m}$, $B, F \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{m\times n}$, $M, N \in \mathbb{R}^{2n\times n}$, $x : [0,T] \to \mathbb{R}^n$, $u, v : [0,T] \to \mathbb{R}^m$, $x_b \in \mathbb{R}^{2n}$. The notation $0 \leq v \perp w \geq 0$ means that each component v_i and w_i of the vectors v and w comply with: $v_i, w_i \geq 0$, $v_i w_i = 0$. In order to avoid trivial cases, we assume that $(C, E) \neq (0, 0)$ and x_b is in the image set of $(M\ N)$. Also, we choose U symmetric and positive-definite, and Q semi-positive definite. LCS as (2) find applications in several important fields such as Nash equilibrium games, genetic networks, contact mechanics and electrical circuits [1, 2, 4, 6, 7, 18, 25, 26, 27, 44, 53]. The analytical study of such nonlinear and nonsmooth dynamical systems is well developed, highlighting properties of existence and uniqueness of solutions, stability and stabilization, passivity, periodic oscillations, observer design, output regulation, or non-zenoness of solutions, see e.g., [10, 11, 14, 17, 19, 28, 32, 34, 40, 45, 47, 48, 50, 51]. However the optimal control issues remain unsolved to the best of the authors' knowledge. For fixed x and u, the complementarity problem of finding v such that $0 \leq v \perp Cx + Dv + Eu \geq 0$ may admit no solution, a unique solution (if D is a P-matrix [23]), or several solutions. When D is a P-matrix, then v is a piecewise linear function of x and u, and (2) can then be seen as a switching system with at most 2^m modes [26], where switches are defined by the complementarity conditions. When $D = 0$ and certain passivity-like input-output constraints are satisfied, LCS can be transformed into first order sweeping processes (a particular type of time-varying differential inclusions) [12]. The optimal control of the sweeping processes has been studied in [13, 21, 22]. However the problems tackled in these articles do not match with (1)(2) because the controller does not act at the same place in the constraint set. For further relations between LCS and other dynamical formalisms, see [9, 26]. The so-called Mathematical Programs with Equilibrium Constraints (MPEC), which are the finite-dimensional counterpart of (1)(2) (see for instance the monograph [39]),

*INRIA Grenoble Rhône-Alpes, Univ. Grenoble-Alpes, 655 avenue de l'Europe, 38334 Saint-Ismier, France. alexandre.vieira@inria.fr
†INRIA Grenoble Rhône-Alpes, Univ. Grenoble-Alpes, 655 avenue de l'Europe, 38334 Saint-Ismier, France. bernard.brogliato@inria.fr
‡Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France. christophe.prieur@gipsa-lab.fr

are at the core of the numerical solvers for the discretized version of (1)(2) that is used in the direct approach. Many definitions and properties presented in this paper are closely related to the tools developed for MPEC problems.
The problem of existence of solutions of the Optimal Control Problem (1)(2) is actually twofold. First, the existence of a trajectory for the LCS (2) is not straightforward, even if the system is expressed as an Initial Value Problem (IVP). Solutions sets depend strongly on the relative degree between w (seen as an output) and v (seen as an input): solutions are in general measures or even Schwarz' distributions of higher degree [3, 33]. Also the only available analysis about the controllability of (2) may be found in [15] (when D is a P-matrix), and in [8] (when $D = 0$) in a very particular case. Secondly and most importantly, the existence of solution for (1)(2) still is an open question. A famous result due to Filipov [20, Theorem 9.2i] states the existence of an optimal control under convexity of the so-called velocity set $\mathcal{V}(x)$. In our case, $\mathcal{V}(x) = \{(u, v) \in \mathbb{R}^{2m} | 0 \leq v \perp Cx + Dv + Eu \geq 0\}$ is clearly not convex, due to the complementarity. *Therefore troughout this article, we admit that an optimal solution exists (in the sense of Definition 2 below)*, and the focus is on necessary conditions this optimal solution must comply with (relying strongly on the seminal work in [30]), together with their numerical computation which relies on MPEC algorithms.

Our contributions are the following: in Section 2 some stationary results for (1)(2) are proved, stated as complementarity problems, hence getting rid of the index sets defining the complementarity modes. Secondly, we develop two numerical methods (and the corresponding on-line available codes) for solving this problem: the first one in Section 3 is a direct method using MPEC solvers, the second one in Section 4 is a hybrid method that uses the stationary results obtained in Section 2, which allows us to get fast and precise numerical solutions. Several examples are presented which demonstrate the validity of the method and of the developed codes. The numerical solutions also highlight some properties of the optimal trajectories, like (dis)continuity of the optimal control, or switching between different modes of the LCS. Conclusions are in Section 5 and useful results are presented in the Appendix.

Notation and definitions $\mathbb{N}$ is the set of non-negative integers. For $n \in \mathbb{N}$, we denote by $\overline{n}$ the set $\{1, ..., n\}$. Given a set of indices $I \subset \overline{n}$ and $z \in \mathbb{R}^n$, we denote $z_I = \{z_i,\ i \in I\}$. For two set of indices $I \subset \overline{n}$, $J \subset \overline{m}$ and a matrix $M \in \mathbb{R}^{n\times m}$, M_{IJ} is the matrix formed by the rows indexed by I and the columns indexed by J. If $I = \overline{n}$, then we simply write $M_{\bullet J}$ (the same holds when $J = \overline{m}$). $\mathbb{R}^n_+$ is the positive orthant of $\mathbb{R}^n$. For $z \in \mathbb{R}^n$, we denote by $z^\intercal$ its transpose (the same notation holds for matrices). For any set $\Omega \subseteq \mathbb{R}^n$, $\mathrm{cl}(\Omega)$ is the closure of B, $\mathrm{co}\Omega$ is the convex hull of Ω, and $\mathrm{dist}_\Omega(z)\ \inf_{y\in\Omega} \|y-z\|$ is the Euclidean distance from z to Ω, and for any $\delta > 0$, $B_\delta(z)$ denotes the open ball of radius δ and center z. For any matrix M, $\mathrm{im}(M)$ is the image set of M, $\ker(M)$ is the kernel of M. Given $x \in \mathrm{cl}\ \Omega$, the proximal normal cone to Ω at x is defined as: $\mathcal{N}^P_\Omega(x) = \{y \in \mathbb{R}^n : \exists \sigma > 0 \text{ s.t. } \langle y, z - x\rangle \leq \sigma\|z - x\|^2\ \forall z \in \Omega\}$. When Ω is a convex set, $\mathcal{N}^P_\Omega(x)$ just reduces to the normal cone of convex analysis. Then we simply denote it as $\mathcal{N}_\Omega(x)$. Given a lower semicontinuous function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point x with $\varphi(x)$ finite, then the Clarke subdifferential of φ at x is denoted as $\partial^C\varphi(x)$. For a set P included in a Hilbert space E, we denote by P^* its dual cone, defined by $P^* = \{y \in E : \forall x \in P,\ \langle y, x\rangle \geq 0\}$. For $D \in \mathbb{R}^{m\times m}$ and $q \in \mathbb{R}^m$, we call $\mathrm{LCP}(q, D)$ the problem of finding $\lambda \in \mathbb{R}^m$ such that $0 \leq \lambda \perp D\lambda + q \geq 0$ and $\mathrm{SOL}(q, D)$ the (possibly empty) set of such λ.

2 First-order necessary conditions for the optimal control problem (1)(2)

2.1 Preliminaries

Let us first recall some basic definitions which will be used throughout the article.

2.1.1 MPEC constraints qualification

As presented in the introduction, many results derived here are inspired by the MPEC literature. Simply speaking, MPECs are optimization programs of the form:

$$\begin{aligned} &\min\ f(z) \\ &\text{s.t. } 0 \leq G(z) \perp H(z) \geq 0, \end{aligned} \tag{3}$$

for some scalar function f and vector functions G and H. Usual Constraints Qualifications (CQ) for this kind of programs, as for instance the Mangasarian Fromovitz Constraint Qualification (MFCQ), are violated at any point satisfying the complementarity conditions. Using the piecewise programming approach, other CQ specific for MPEC can be derived. We present here some definitions and properties. Further results can be found in [54, 35, 43, 39].

Definition 1. *Let $n, m \in \mathbb{N}$. The* complementarity cone *is defined as $\mathcal{C}^\ell = \{(v, w) \in \mathbb{R}^m \times \mathbb{R}^m : 0 \leq v \perp w \geq 0\}$. Given a system of constraints $\Omega = \{z \in \mathcal{D} : (G(z), H(z)) \in \mathcal{C}^\ell\}$ where $\mathcal{D}$ is a closed subset in $\mathbb{R}^n$ and $G, H : \mathbb{R}^n \to$*

$\mathbb{R}^m$, *we say that the* local error bound condition *holds at* $\bar{z} \in \Omega$ *if there exist* $\tau > 0$ *and* $\delta > 0$ *such that*

$$dist_\Omega(z) \leq \tau dist_{\mathcal{C}^\ell}(G(z), H(z)), \ \forall z \in B_\delta(\bar{z}) \cap \mathcal{D}. \tag{4}$$

Three different index sets are defined from these constraints, called the active sets and the degenerate set: $I^{+0}(\bar{z}) = \{i \in \overline{m} : G_i(\bar{z}) > 0 = H_i(\bar{z})\}$, $I^{0+}(\bar{z}) = \{i \in \overline{m} : G_i(\bar{z}) = 0 < H_i(\bar{z})\}$, $I^{00}(\bar{z}) = \{i \in \overline{m} : G_i(\bar{z}) = 0 = H_i(\bar{z})\}$. *The sets* $I^{\bullet 0}(\bar{z})$ *and* $I^{0\bullet}(\bar{z})$ *are defined as* $I^{\bullet 0}(\bar{z}) = I^{+0}(\bar{z}) \cup I^{00}(\bar{z})$, $I^{0\bullet}(\bar{z}) = I^{0+}(\bar{z}) \cup I^{00}(\bar{z})$. *The* MPEC Linear Condition *holds if both functions* $G(\cdot)$, $H(\cdot)$ *are affine and* $\mathcal{D}$ *is a union of finitely many polyhedral sets. When* $\mathcal{D} = \mathbb{R}^n$, *the* MPEC Linear Independence Constraint Qualification (LICQ) *holds at* $\bar{z} \in \Omega$ *if the family of gradients* $\{\nabla G_i(\bar{z}) : i \in I^{0\bullet}(\bar{z})\} \cup \{\nabla H_i(\bar{z}) : i \in I^{\bullet 0}(\bar{z})\}$ *is linearly independent.*

Proposition 1. *[30] The local error bound condition (4) holds at* $\bar{z} \in \Omega$ *if the MPEC linear condition or the MPEC LICQ hold at* $\bar{z}$.

2.1.2 Non-smooth optimal control

Definition 2. *Let* $n, m \in \mathbb{N}$. *We refer to any absolutely continuous function as an arc. An* admissible pair *for (2) is a 3-tuple of functions* (x, u, v) *on* $[0, T]$ *for which* u, v *are controls and* x *is an arc, that satisfy all the constraints in (2). Let us define the constraint set* S, *by* $S = \left\{(x, u, v) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \ : \ (v, Cx + Dv + Eu) \in \mathcal{C}^\ell\right\}$. *Given a constant* $R > 0$, *we say that an admissible pair* (x^*, u^*, v^*) *is a* local minimizer of radius R *for (1)(2) if there exists* $\varepsilon > 0$ *such that for every pair* (x, u, v) *admissible for (2), which also satisfies* $\|x(t) - x^*(t)\| \leq \varepsilon$, $\left\|\binom{u}{v} - \binom{u^*(t)}{v^*(t)}\right\| \leq R$ *a.e.* $t \in [0, T]$ *and* $\int_0^T \|\dot{x}(t) - \dot{x}^*(t)\| dt \leq \varepsilon$, *we have* $J(x^*, u^*, v^*) \leq J(x, u, v)$. *For every given* $t \in [0, T]$, *and constant scalars* $\varepsilon > 0$ *and* $R > 0$, *we define the neighborhood of the point* $(x^*(t), u^*(t), v^*(t))$ *as* $S_*^{\varepsilon,R}(t) = \left\{(x, u, v) \in S : \ \|x - x^*(t)\| \leq \varepsilon, \ \left\|\binom{u}{v} - \binom{u^*(t)}{v^*(t)}\right\| \leq R\right\}$. *We also define* $C^{\varepsilon,R} = cl\,\{(t, x, u, v) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m : (x, u, v) \in S_*^{\varepsilon,R}(t)\}$. *The dependence on time of index sets is denoted as* $I_t^{0+}(x, u, v) = \{i \in \overline{m} : v_i(t) > 0 = (Cx(t) + Dv(t) + Eu(t))_i\}$. *The same definition follows for* $I_t^{+0}(x, u, v)$, $I_t^{00}(x, u, v)$, $I_t^{\bullet 0}(x, u, v)$, $I_t^{0\bullet}(x, u, v)$. *For a positive measurable function* k_S *defined for almost every* $t \in [0, T]$, the bounded slope condition *is defined as the following implication:*

$$(x, u, v) \in S^{\varepsilon,R}(t), (\alpha, \beta, \gamma) \in \mathcal{N}_S^P(x, u, v) \implies \|\alpha\| \leq k_S(t) \left\|\begin{pmatrix}\beta \\ \gamma\end{pmatrix}\right\|. \tag{5}$$

Proposition 2. *[30, Proposition 3.10] Define* $\Psi(x, u, v; \mu, \nu) = v^\intercal \mu + (Cx + Dv + Eu)^\intercal \nu$. *Assume that* $C^{\varepsilon,R}$ *is compact for some* $\varepsilon > 0$, *the local error bound holds, and that, for every* (t, x, u, v) *such that* $(t, x, u, v) \in C^{\varepsilon,R}$, *the system complies with the following implication:*

$$\left.\begin{aligned} 0 = \nabla_{u,v}\Psi(x, u, v, \mu, \nu) \\ \mu_i = 0, \ \forall i \in I_t^{+0}(x, u, v), \ \nu_i = 0, \ \forall i \in I_t^{0+}(x, u, v), \\ \mu_i > 0, \ \nu_i > 0, \ \text{or } \mu_i \nu_i = 0, \ \forall i \in I_t^{00}(x, u, v) \end{aligned}\right\} \implies \nabla_x \Psi(x, u, v; \mu, \nu) = 0.$$

Then there exists a certain positive constant k_S *such that for every* $t \in [0, T]$, *the bounded slope condition (5) holds with* $k_S(t) = k_S$.

2.2 Necessary first-order conditions

The necessary first order conditions for a very general optimal control problem containing complementarity constraints (see (32) in Appendix A.1) have been derived in [30]. In this section, our goal is to show how the results in [30] (which are briefly recalled in Appendix A) particularize when we consider the problem (1)(2), as stated in Theorem 2 below. Before going on, let us derive conditions which guarantee that (5) holds.

Lemma 1. *Suppose* $\mathrm{im}(C) \subseteq \mathrm{im}(E)$ *and that* $C^{\varepsilon,R}$ *is compact for some* $\varepsilon > 0$. *Then the bounded slope condition (5) for the system (2) holds.*

Proof. As stated in Proposition 1, since the system is linear, the local error bound condition (4) holds. Let us check that the implication in Proposition 2 holds. In our case, it sufficiently reads as:

$$\left.\begin{aligned} E^\intercal \nu = 0 \\ \mu + D^\intercal \nu = 0 \end{aligned}\right\} \implies C^\intercal \nu = 0.$$

The first line in the left-hand side of the implication implies that $\nu \in \ker(E^\intercal) = \mathrm{im}(E)^\perp$. But since $\mathrm{im}(C) \subseteq \mathrm{im}(E)$, it is equivalent to $\mathrm{im}(E)^\perp \subseteq \mathrm{im}(C)^\perp$. So $\nu \in \mathrm{im}(C)^\perp = \ker(C^\intercal)$. Then $C^\intercal \nu = 0$, and the implication holds. □

Let us now apply [30, Theorem 3.2], recalled in Theorem 4 in the Appendix, to the problem in (1)(2).

Proposition 3. *Let (x^*, u^*, v^*) be a local minimizer of constant radius $R > 0$ for (1)(2), and suppose that $C_*^{\varepsilon,R}$ is compact for some $\varepsilon > 0$. Suppose $\mathrm{im}(C) \subseteq \mathrm{im}(E)$. Then there exist an arc $p : [0,T] \to \mathbb{R}^n$, a scalar $\lambda_0 \leq 0$ and measurable functions $\lambda^G : \mathbb{R} \to \mathbb{R}^m$, $\lambda^H : \mathbb{R} \to \mathbb{R}^m$ such that the following conditions hold:*

1. *The non-triviality condition: $(\lambda_0, p(t)) \neq 0$, $\forall t \in [0,T]$.*
2. *The transversality condition: $\begin{pmatrix} p(0) \\ -p(T) \end{pmatrix} \in \mathrm{im} \begin{pmatrix} M^\intercal \\ N^\intercal \end{pmatrix}$.*
3. *The Euler adjoint equation: for almost every $t \in [0,T]$,*

$$\begin{aligned} \dot{p}(t) &= -A^\intercal p - 2\lambda_0 Q x^* - C^\intercal \lambda^H \\ 0 &= F^\intercal p + 2\lambda_0 U u^* + E^\intercal \lambda^H \\ 0 &= B^\intercal p + \lambda^G + D^\intercal \lambda^H \\ 0 &= \lambda_i^G(t), \qquad && \forall i \in I_t^{+0}(x^*, u^*, v^*) \\ 0 &= \lambda_i^H(t), && \forall i \in I_t^{0+}(x^*, u^*, v^*). \end{aligned} \tag{6}$$

4. *The Weierstrass condition for radius R: for almost every $t \in [t_0, t_1]$,*

$$\begin{aligned} &(x^*(t), u, v) \in S, \ \left\| \begin{pmatrix} u \\ v \end{pmatrix} - \begin{pmatrix} u^*(t) \\ v^*(t) \end{pmatrix} \right\| < R \\ \implies &\langle p(t), Ax^*(t) + Bv^*(t) + Fu^*(t)) \rangle + \lambda_0 \left(x^*(t)^\intercal Q x^*(t) + u^*(t)^\intercal U u^*(t) \right) \\ &\qquad \geq \langle p(t), Ax^*(t) + Bv + Fu) \rangle + \lambda_0 \left(x^*(t)^\intercal Q x^*(t) + u^\intercal U u \right). \end{aligned} \tag{7}$$

Proof. Let us check that the problem complies with [30, Assumption 3.1]. These assumptions are recalled in Assumption 1 in Section A.1.1.

1. Let $t \in [0,T]$ and $(x_1, u_1, v_1), (x_2, u_2, v_2) \in S_*^{\varepsilon,R}(t)$. First, let us check (34)(a):

$$\|(Ax_1 + Bv_1 + Fu_1) - (Ax_2 + Bv_2 + Fu_2)\| \leq \|A\|\|x_1 - x_2\| + \|B\|\|v_1 - v_2\| + \|F\|\|u_1 - u_2\|.$$

   Secondly, we must check the inequality concerning the cost in (34)(b). For that, remark first that:

$$\begin{aligned} |x_1^\intercal Q x_1 - x_2^\intercal Q x_2| &= |(x_1 + x_2)^\intercal Q (x_1 - x_2)| \\ &\leq \|x_1 - x^*(t) + x_2 - x^*(t) + 2x^*(t)\|\|Q\|\|x_1 - x_2\| \\ &\leq \left(\|x_1 - x^*(t)\| + \|x_2 - x^*(t)\| + 2\|x^*(t)\| \right) \|Q\|\|x_1 - x_2\| \\ &\leq 2\|Q\| \left(\|x^*(t)\| + \varepsilon \right) \|x_1 - x_2\|. \end{aligned}$$

   Similarly, one proves that $|u_1^\intercal U u_1 - u_2^\intercal U u_2| \leq 2\|U\|(\|u^*(t)\| + R)\|u_1 - u_2\|$. Therefore:

$$\begin{aligned} |(x_1^\intercal Q x_1 + u_1^\intercal U u_1) - (x_2^\intercal Q x_2 + u_2^\intercal U u_2)| &\leq |x_1^\intercal Q x_1 - x_2^\intercal Q x_2| + |u_1^\intercal U u_1 - u_2^\intercal U u_2| \\ &\leq k_x(t)\|x_1 - x_2\| + k_u(t)\|u_1 - u_2\|. \end{aligned}$$

   where $k_x(t) = 2\|Q\| \left(\|x^*(t)\| + \varepsilon \right)$ and $k_u(t) = 2\|U\|(\|u^*(t)\| + R)$. k_x, k_u are measurable functions of time, and $\|A\|$, $\|B\|$ and $\|F\|$ are all constants and therefore measurable functions. Thus (34) holds true.

2. Since $\mathrm{im}(C) \subseteq \mathrm{im}(E)$, and using Lemma 1, the bounded slope condition holds, with a positive constant k_S.

3. The terms$k_S[\|B\| + \|F\| + k_u]$, k_x and $\|A\|$ are all integrable on $[0,T]$, and there obviously exists a positive number η such that $R \geq \eta k_S$ on $[0,T]$ (just take $\eta = R/k_S$).

4. Since all involved functions are smooth, all conditions of measurability and differentiability are met.

Furthermore, since the MPEC Linear Condition holds (see Definition 1), the error bound condition for the system (2) holds at $(x^*(t), u^*(t), v^*(t))$ for all $t \in [0,T]$. Calculations of the non-triviality and Weierstrass conditions are straightforward. Since all functions are differentiable, the Clarke subdifferential in (38) contains only the gradient, i.e.
$\nabla_{x,u,v} \left(\langle p(t), Ax + Bv + Fu \rangle - \lambda_0 (x^\intercal Q x + u^\intercal U u) \right)$, and $U(\cdot)$ is in our case the whole space $\mathbb{R}^{2m}$, so the normal cone reduces to $\{0\}$. Simple calculations from (38) yield the Euler equation (6). Concerning the transversality condition (37), notice first that we did not impose any boundary cost. Denote $P_b = \left\{ \binom{x_0}{x_T} : Mx_0 + Nx_T = x_b \right\}$, then for $\binom{x_0}{x_T} \in P_b$, since P_b is an affine vector space,

$$\mathcal{N}_{P_b} \begin{pmatrix} x_0 \\ x_T \end{pmatrix} = - \left(P_b - \begin{pmatrix} x_0 \\ x_T \end{pmatrix} \right)^* = -\ker(M \ N)^* = \operatorname{im} \begin{pmatrix} M^\intercal \\ N^\intercal \end{pmatrix}.$$

□

Remark. *The tuple consisting of a trajectory and the associated multipliers solution of (6) is called an extremal. The case $\lambda_0 = 0$ is often called the abnormal case, and the corresponding extremal an abnormal extremal. In this case, no information can be derived from these necessary conditions. In other cases, we can choose this value most conveniently, since the adjoint state p is defined up to a multiplicative positive constant. In the rest of this paper, λ_0 will always be chosen as $-\frac{1}{2}$. The optimal trajectory is normal when, for instance, the initial point $x(0)$ or the final point $x(T)$ are free.*

The Weierstrass condition (7) can be re-expressed as searching a local maximizer of the following MPEC:

$$\begin{aligned} &\max_{u,v} \ \langle p(t), Ax^*(t) + Bv + Fu \rangle + \lambda_0 \left(x^*(t)^\intercal Q x^*(t) + u^\intercal U u \right) \\ &\text{s.t. } 0 \le v \perp Cx^*(t) + Dv + Eu \ge 0. \end{aligned} \tag{8}$$

For each $t \in [0,T]$, this is an MPEC, as presented in Section 2.1.1. These programs admit first-order conditions that are specific, called weak and strong stationarity: this motivates the next definition. More detailed results are exposed in Appendix B.

Definition 3. *Let (x^*, u^*, v^*) be an admissible pair for (2). Then:*

- *The FJ-type W(eak)-stationarity holds at (x^*, u^*, v^*) if there exist an arc p, a scalar $\lambda_0 \le 0$ and measurable functions λ^G, λ^H such that Proposition 3 (1)-(4) hold.*
- *The FJ-type S(trong)-stationarity holds at (x^*, u^*, v^*) if (x^*, u^*, v^*) is FJ-type W-stationary with arc p and there exist measurable functions η^G, η^H such that, for almost every $t \in [0,T]$,*

$$\begin{aligned} 0 &= F^\intercal p + 2\lambda_0 U u^* + E^\intercal \eta^H \\ 0 &= B^\intercal p + \eta^G + D^\intercal \eta^H \\ 0 &= \eta^G_i(t), && \forall i \in I_t^{+0}(x^*, u^*, v^*) \\ 0 &= \eta^H_i(t), && \forall i \in I_t^{0+}(x^*, u^*, v^*), \end{aligned}$$

$$\eta^G_i(t) \ge 0, \ \eta^H_i(t) \ge 0, \ \forall i \in I_t^{00}(x^*, u^*, v^*).$$

- *We simply call W-stationarity or S-stationarity the FJ-type W- or S-stationarity with $\lambda_0 = -\frac{1}{2}$.*

The multipliers η^G, η^H can be different in measure from the corresponding λ^G, λ^H in Proposition 3. The next theorem, whose proof follows directly from [30, Theorem 3.6] as recalled in Theorem 5 in Appendix A.1.1, addresses this problem:

Theorem 1. *Let (x^*, u^*, v^*) be a local minimizer of radius R for (1)(2). Suppose that for almost every $t \in [0,T]$, the MPEC LICQ holds at $(u^*(t), v^*(t))$ for problem (8), i.e., the family of gradients*

$$\left\{ \begin{pmatrix} 0 \\ e_i \end{pmatrix} : i \in I_t^{0\bullet}(x^*, u^*) \right\} \bigcup \left\{ \begin{pmatrix} (E_{i\bullet})^\intercal \\ (D_{i\bullet})^\intercal \end{pmatrix} : i \in I_t^{\bullet 0}(x^*, u^*) \right\}. \tag{9}$$

is linearly independent, where e_i is a vector such that its j-th component is equal to δ_i^j, the Kronecker delta. Then the S-stationarity holds at (x^*, u^*, v^*). Moreover, in this case, the multipliers η^G, η^H can be taken equal to λ^G, λ^H, respectively, almost everywhere.

We can now state the following result:

Corollary 1. *Suppose E is invertible. Then the local minimum (x^*, u^*, v^*) is S-stationary, and the multipliers η^G, η^H can be chosen equal to λ^G, λ^H almost everywhere.*

Proof. Suppose first that $\operatorname{rank}\begin{pmatrix} 0_m & E^\intercal \\ I_m & D^\intercal \end{pmatrix} = 2m$. This rank condition ensures the fact that the family $\left\{\begin{pmatrix} 0 \\ e_i \end{pmatrix} : 1 \leq i \leq m\right\} \bigcup \left\{\begin{pmatrix} (E_{i\bullet})^\intercal \\ (D_{i\bullet})^\intercal \end{pmatrix} : 1 \leq i \leq m\right\}$ is linearly independent. The family in (9) being a subfamily of this one, it is necessarily linearly independent. So the MPEC LICQ in Definition 1 holds, and (x^*, u^*, v^*) is S-stationary. Let us show now that this rank condition is equivalent to E being invertible. First notice that $2m = \operatorname{rank}\begin{pmatrix} 0_m & E^\intercal \\ I_m & D^\intercal \end{pmatrix} = \operatorname{rank}\begin{pmatrix} 0_m & I_m \\ E & D \end{pmatrix}$. Since $\binom{0}{E}$ and $\binom{I_m}{D}$ are linearly independent, we have:

$$2m = \operatorname{rank}\begin{pmatrix} 0_m & I_m \\ E & D \end{pmatrix} = \operatorname{rank}\begin{pmatrix} 0_m \\ E \end{pmatrix} + \operatorname{rank}\begin{pmatrix} I_m \\ D \end{pmatrix} = \operatorname{rank}(E) + m.$$

Thus, $\operatorname{rank}(E) = m$, so E is invertible. □

Let us now state a result that allows us to reformulate the S-stationarity conditions through a complementarity system, starting from Proposition 9, in order to remove the active sets. One can simply see it that way: for almost all $t \in [0, T]$, the conditions on the multipliers λ^H and λ^G are:

$$\begin{aligned} &\lambda_i^G(t) = 0, \ \forall i \in I_t^{+0}(x, u, v) \\ &\lambda_i^H(t) = 0, \ \forall i \in I_t^{0+}(x, u, v) \\ &\lambda_i^G(t) \geq 0, \ \lambda_i^H(t) \geq 0, \ \forall i \in I_t^{00}(x, u, v). \end{aligned} \tag{10}$$

The presence of the active and degenerate sets is bothersome, since they depend on the optimal solution, not in a useful way. Nonetheless, the conditions in (10) look almost like a linear complementarity problem. The only thing missing is the sign of λ_i^G for $i \in I_t^{0+}(x, u, v)$ (and the same thing with λ_i^H on $I_t^{+0}(x, u, v)$). On these index sets, the multipliers could be negative. But we could for instance create new variables, say α and β, that will both be non-negative and comply with these conditions. This is the purpose of the next Proposition.

Proposition 4. *Suppose (x, u, v) is an S-stationary trajectory. Then there exist measurable functions $\beta : [0, T] \to \mathbb{R}^m$, $\zeta : [0, T] \to \mathbb{R}$ such that:*

$$u(t) = U^{-1}\left(F^\intercal p(t) + E^\intercal \beta(t) - \zeta(t) E^\intercal v(t)\right)$$

and

$$\begin{aligned} \begin{pmatrix} \dot{x} \\ \dot{p} \end{pmatrix} &= \begin{pmatrix} A & FU^{-1}F^\intercal \\ Q & -A^\intercal \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} + \begin{pmatrix} B - \zeta FU^{-1}E^\intercal \\ \zeta C^\intercal \end{pmatrix} v + \begin{pmatrix} FU^{-1}E^\intercal \\ -C^\intercal \end{pmatrix}\beta \\ 0 \leq \begin{pmatrix} v \\ \beta \end{pmatrix} \perp &\begin{pmatrix} D - \zeta EU^{-1}E^\intercal & EU^{-1}E^\intercal \\ D - \zeta EU^{-1}E^\intercal & EU^{-1}E^\intercal \end{pmatrix}\begin{pmatrix} v \\ \beta \end{pmatrix} + \begin{pmatrix} C & EU^{-1}F^\intercal \\ C & EU^{-1}F^\intercal \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} \geq 0 \\ 0 \leq v \perp &\ \zeta(D + D^\intercal - \zeta EU^{-1}E^\intercal)v + (\zeta EU^{-1}E^\intercal - D^\intercal)\beta + (\zeta EU^{-1}F^\intercal - B^\intercal)p + \zeta Cx \geq 0. \end{aligned} \tag{11}$$

To prove this Proposition, we first need the following Lemma.

Lemma 2. *Let (x, u, v) be an S-stationary trajectory, and λ^G, λ^H be the associated multipliers. Then there exists a measurable function $\zeta : [0, T] \to \mathbb{R}$ such that $\begin{pmatrix} \lambda^G(t) + \zeta(t) w(t) \\ \lambda^H(t) + \zeta(t) v(t) \end{pmatrix} \geq 0$, where w is defined in (2).*

Proof. First, remark that, for all $t \in [0, T]$, a candidate $\zeta(t)$ has been defined in (43), Proposition 9 in Section B. Denote $F : [0, T] \times \mathbb{R} \to \mathbb{R}^{2m}$, $F(t, \zeta) = \begin{pmatrix} \lambda^G(t) + \zeta w(t) \\ \lambda^H(t) + \zeta v(t) \end{pmatrix}$. F is a Carathéodory mapping, since: λ^G, λ^H, v and w are measurable, so $F(\cdot, \zeta)$ is measurable for each fixed $\zeta \in \mathbb{R}$, and $F(t, \cdot)$ is an affine function, and as such it is continuous, for each fixed t. By the Implicit Measurable Function Theorem [42, Theorem 14.16], there exists a measurable function $\zeta : [0, T] \to \mathbb{R}$ such that $F(t, \zeta(t)) \in \mathbb{R}^{2m}_+$, which is the intended result. □

Proof of Proposition 4. As proved in Lemma 2, there exists a measurable function $\zeta : [0,T] \to \mathbb{R}$ such that:

$$\begin{pmatrix} \lambda^G(t) + \zeta(t)w(t) \\ \lambda^H(t) + \zeta(t)v(t) \end{pmatrix} \geq 0.$$

Define $\alpha, \beta : [0,T] \to \mathbb{R}^m$ as $\alpha = \lambda^G + \zeta w$, $\beta = \lambda^H + \zeta v$. The variables α and β are, by construction, measurable and non-negative. From the fact that (x,u,v) is an S-stationary trajectory, we also have that, for almost every $t \in [0,T]$, $\lambda_i^G(t)v_i(t) = 0$ and $\lambda_i^H(t)w_i(t) = 0$ for all $i \in \overline{m}$. Therefore, we can deduce that:

$$\begin{cases} \lambda^G = \alpha - \zeta w \\ \lambda^H = \beta - \zeta v \\ 0 \leq \alpha \perp v \geq 0 \\ 0 \leq \beta \perp w \geq 0. \end{cases} \tag{12}$$

In (6), let us isolate u, since we supposed U symmetric positive definite. Inserting the redefinition of λ^H yields:

$$u(t) = U^{-1}(F^\intercal p(t) + E^\intercal \lambda^H(t)) = U^{-1}\left(F^\intercal p(t) + E^\intercal \beta(t) - \zeta(t)E^\intercal v(t)\right). \tag{13}$$

Recall that $w = Cx + Dv + Eu$. Inserting this u in (12), we obtain:

$$\lambda^G = \alpha - \zeta(Cx + Dv + Eu) = \alpha - \zeta(Cx + (D - \zeta EU^{-1}E^\intercal)v + EU^{-1}F^\intercal p + EU^{-1}E^\intercal \beta)).$$

Inserting (12) and (13) into (2) and (6) allows us to rewrite the differential equations defining x and p as :

$$\begin{cases} \dot{x} = Ax + Bv + Fu = Ax + FU^{-1}F^\intercal p + (B - \zeta FU^{-1}E^\intercal)v + FU^{-1}E^\intercal \beta \\ \dot{p} = -A^\intercal p + Qx - C^\intercal \lambda^H = -A^\intercal p + Qx + \zeta C^\intercal v - C^\intercal \beta. \end{cases}$$

The only equation left is the third equation in (6). Replacing λ^G and λ^H with the expressions (12) yields:

$$B^\intercal p + \alpha - \zeta\left(Cx + (D - \zeta EU^{-1}E^\intercal)v + EU^{-1}F^\intercal p + EU^{-1}E^\intercal \beta\right) + D^\intercal(\beta - \zeta v) = 0$$

$$\implies \alpha = (\zeta EU^{-1}F^\intercal - B^\intercal)p + \zeta Cx + (\zeta EU^{-1}E^\intercal - D^\intercal)\beta + \zeta(D + D^\intercal - \zeta EU^{-1}E^\intercal)v.$$

Replacing α and u in the complementarity conditions appearing in (2) and in (12) yields the complementarity conditions in (11). □

Remark.

- *The decomposition of (λ^G, λ^H) into (α, β, ζ) proposed in (12) is not unique, and has a single degree of freedom. Indeed, if this decomposition works for (α, β, ζ), then for any $\rho \geq 0$, we can decompose (λ^G, λ^H) as $(\alpha + \rho w, \beta + \rho v, \zeta + \rho)$. Therefore, for a fixed $t \in [0,T]$, any scalar greater than $\zeta(t)$ is suitable. Thus, if we can find an upper-bounded function ζ decomposing (λ^G, λ^H) into (α, β, ζ), then (λ^G, λ^H) can be decomposed into $(\bar{\alpha}, \bar{\beta}, \bar{\zeta})$, where $\bar{\zeta}$ is a constant along $[0,T]$ greater or equal to the supremum of ζ.*
- *A second remark concerns the three complementarity conditions defining β and v in (11). It is not written as a classical Variational Inequality (VI), since it involves $2m$ unknowns but $3m$ complementarity problems. The next proposition addresses this problem.*

Proposition 5. *Let r be any given positive scalar. Denote (P) the complementarity conditions appearing in (11), and denote (P_r) the problem:*

$$\begin{cases} 0 \leq \beta + rv \perp (D - \zeta EU^{-1}E^\intercal)v + EU^{-1}E^\intercal \beta + EU^{-1}F^\intercal p + Cx \geq 0 \\ 0 \leq v \perp \zeta(D + D^\intercal - \zeta EU^{-1}E^\intercal)v + (\zeta EU^{-1}E^\intercal - D^\intercal)\beta + (\zeta EU^{-1}F^\intercal - B^\intercal)p + \zeta Cx \geq 0 \\ \beta \geq 0. \end{cases} \tag{P_r}$$

Then (v, β) is a solution of (P) if and only if it is a solution of (P_r).

Proof. We rewrite more simply the two problems as follows:

$$\begin{cases} 0 \le v \perp & \tilde{D}v + \tilde{U}\beta + q_1 \ge 0 \quad (P1) \\ 0 \le \beta \perp & \tilde{D}v + \tilde{U}\beta + q_1 \ge 0 \quad (P2) \\ 0 \le v \perp & \tilde{D}_2 v + \tilde{U}_2\beta + q_2 \ge 0 \quad (P3) \end{cases} \quad \begin{cases} 0 \le \beta + rv \perp & \tilde{D}v + \tilde{U}\beta + q_1 \ge 0 \quad (P_r1) \\ \beta \ge 0 & \quad (P_r2) \\ 0 \le v \perp & \tilde{D}_2 v + \tilde{U}_2\beta + q_2 \ge 0 \quad (P_r3) \end{cases}$$

where $\tilde{D} = (D - \zeta EU^{-1}E^\intercal)$, $\tilde{U} = EU^{-1}E^\intercal$, $\tilde{D}_2 = \zeta(D + D^\intercal - \zeta EU^{-1}E^\intercal)$, $\tilde{U}_2 = (\zeta EU^{-1}E^\intercal - D^\intercal)$, $q_1 = EU^{-1}F^\intercal p + Cx$, $q_2 = (\zeta EU^{-1}F^\intercal - B^\intercal)p + \zeta Cx$.

- Let (v, β) be a solution of (P). Denote:

$$I^{0+} = \{i : v_i = \beta_i = 0,\ (\tilde{D}v + \tilde{U}\beta + q_1)_i > 0\},$$

$$I^{\bullet 0} = \{i : (\tilde{D}v + \tilde{U}\beta + q_1)_i = 0\}.$$

  These two sets form a partition of $\{1, ..., m\}$. Since CP $(P3)$ and (P_r3) are the same problem, (v, β) is also solution of (P_r3). Using $(P2)$, we find that β complies with (P_r2). We are just left with (P_r1). By assumption it follows that $\forall i \in I^{0+}, \beta_i + rv_i = 0,\ (\tilde{D}v + \tilde{U}\beta + q_1)_i > 0$ and $\forall i \in I^{\bullet 0}, \beta_i + rv_i \ge 0,\ (\tilde{D}v + \tilde{U}\beta + q_1)_i = 0$. So (v, β) is also a solution of (P_r1). This proves that (v, β) is a solution of (P_r).

- Conversely, let (v, β) be a solution of (P_r). Since it is a solution of (P_r1), denote $I_r^{0+} = \{i : \beta_i + rv_i = 0,\ (\tilde{D}v + \tilde{U}\beta + q_1)_i > 0\}$ and $I_r^{\bullet 0} = \{i : (\tilde{D}v + \tilde{U}\beta + q_1)_i = 0\}$. These two sets form a partition of $\{1, ..., m\}$. Since CP (P_r3) and $(P3)$ are the same problem, (v, β) is also solution of $(P3)$. For all $i \in I_r^{0+}$, $\beta_i + rv_i = 0$ and $(\tilde{D}v + \tilde{U}\beta + q_1)_i > 0$. Thanks to (P_r3) and (P_r2), we know that $\beta_i \ge 0$, $v_i \ge 0$. Since $r > 0$, we have a sum of positive terms that must equal 0, so $\beta_i = v_i = 0$. For all $i \in I_r^{\bullet 0}$, $(\tilde{D}v + \tilde{U}\beta + q_1)_i = 0$ and using (P_r3) and (P_r2), $\beta_i \ge 0$, $v_i \ge 0$. So (v, β) is also a solution of $(P1)$ and $(P2)$. It proves that (v, β) is a solution of (P).

□

Let us define $\tilde{\beta} = \beta + rv$ and replace β in (P_r). Thus we end up with the following LCP and inequality constraints:

$$\begin{cases} 0 \le \begin{pmatrix} \tilde{\beta} \\ v \end{pmatrix} \perp \begin{pmatrix} EU^{-1}E^\intercal & D - (\zeta + r)EU^{-1}E^\intercal \\ \zeta EU^{-1}E^\intercal - D^\intercal & \zeta D + (\zeta + r)\left(D^\intercal - \zeta EU^{-1}E^\intercal\right) \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ v \end{pmatrix} + \begin{pmatrix} EU^{-1}F^\intercal p + Cx \\ (\zeta EU^{-1}F^\intercal - B^\intercal)p + \zeta Cx \end{pmatrix} \ge 0 \\ \tilde{\beta} \ge rv. \end{cases}$$

To sum up, by Propositions 3, 4 and 5, the following theorem holds:

Theorem 2. *Let (x^*, u^*, v^*) be a local minimizer of constant radius $R > 0$ for (1)(2), and let $C^{\varepsilon,R}$ (in Definition 2) be compact for some $\varepsilon > 0$. Suppose E is invertible and (x^*, u^*, v^*) is not the projection of an abnormal extremal. Then there exist an arc $p : [0, T] \to \mathbb{R}^n$, and measurable functions $\tilde{\beta} : [0, T] \to \mathbb{R}^m$, $\zeta : [0, T] \to \mathbb{R}$ such that, for an arbitrary scalar $r > 0$:*

$$u^*(t) = U^{-1}\left(F^\intercal p(t) + E^\intercal \tilde{\beta}(t) - (\zeta(t) + r)E^\intercal v^*(t)\right)$$

and the following conditions hold:

1. *The transversality condition:* $\begin{pmatrix} p(0) \\ -p(T) \end{pmatrix} \in \operatorname{im} \begin{pmatrix} M^\intercal \\ N^\intercal \end{pmatrix}$.

2. *The Euler adjoint equation: for almost every $t \in [0, T]$,*

$$\begin{pmatrix} \dot{x}^* \\ \dot{p} \end{pmatrix} = \begin{pmatrix} A & FU^{-1}F^\intercal \\ Q & -A^\intercal \end{pmatrix} \begin{pmatrix} x^* \\ p \end{pmatrix} + \begin{pmatrix} FU^{-1}E^\intercal & B - (\zeta + r)FU^{-1}E^\intercal \\ -C^\intercal & (\zeta + r)C^\intercal \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ v^* \end{pmatrix} \tag{14}$$

$$\begin{cases} 0 \le \begin{pmatrix} \tilde{\beta} \\ v^* \end{pmatrix} \perp \begin{pmatrix} EU^{-1}E^\intercal & D - (\zeta + r)EU^{-1}E^\intercal \\ \zeta EU^{-1}E^\intercal - D^\intercal & \zeta D + (\zeta + r)\left(D^\intercal - \zeta EU^{-1}E^\intercal\right) \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ v^* \end{pmatrix} + \begin{pmatrix} C & EU^{-1}F^\intercal \\ \zeta C & \zeta EU^{-1}F^\intercal - B^\intercal \end{pmatrix} \begin{pmatrix} x^* \\ p \end{pmatrix} \ge 0 \\ \tilde{\beta} \ge rv^*. \end{cases}$$

3. The Weierstrass condition for radius R: for almost every $t \in [t_0, t_1]$,

$$(x^*(t), u, v) \in S, \ \left\| \begin{pmatrix} u \\ v \end{pmatrix} - \begin{pmatrix} u^*(t) \\ v^*(t) \end{pmatrix} \right\| < R$$

$$\begin{aligned} \implies \langle p(t), Ax^*(t) + Bv^*(t) + Fu^*(t)) \rangle - \frac{1}{2} \left(x^*(t)^\intercal Q x^*(t) + u^*(t)^\intercal U u^*(t) \right) \\ \geq \langle p(t), Ax^*(t) + Bv + Fu) \rangle - \frac{1}{2} \left(x^*(t)^\intercal Q x^*(t) + u^\intercal U u \right). \end{aligned}$$

The importance of this result is twofold. First, it gives a way to analyze the optimal trajectory using these necessary conditions. All results concerning the analysis of LCS can be used to prove some properties of possible trajectories of (14) and to derive results on continuity, jumps or sensitivity on parameters, and therefore to prove some properties of the optimal trajectory. The analysis of LCS relies heavily on the matrix appearing in front of $\begin{pmatrix} \tilde{\beta} \\ v \end{pmatrix}$ in the complementarity conditions of (14). However, with no more hypothesis on matrices appearing in (14), we were not able to derive sharper results. Secondly, in Section 4, this result will be used in an *indirect method* in order to compute numerically an approximate solution with high accuracy. However, this indirect method needs a first guess in order to converge, which is provided by the *direct method*, presented in Section 3. However, in order to be sure that the indirect method computes an optimum, one would also like that the necessary conditions were also sufficient. This is proved in the next Section.

2.3 Sufficiency of the W-stationarity

Surprisingly, the weakest form of stationarity for the problem (1)(2) turns to be also sufficient, in some sense. For this, we need to define trajectories with the same *history*. The development shown here is directly inspired by [24, Proposition 3.1] and by [49].

Definition 4. *Let (x, u, v) and (x^*, u^*, v^*) be two admissible trajectories for (2) (associated with $w = Cx + Dv + Eu$ and w^*, defined the same way). We say that they have the same history on $[0, T]$ if the following condition holds for almost every $t \in [0, T]$:*

$$[v_i = 0 \iff v_i^* = 0] \text{ and } [w_i = 0 \iff w_i^* = 0]$$

From the point of view of the switching systems, two trajectories have the same history on $[0, T]$ if they visit the same modes at the same time along $[0, T]$. The different trajectories compared in the following sufficient condition for optimality are done with respect to this history condition.

Theorem 3. *Suppose that (x^*, u^*, v^*) is an admissible W-stationary trajectory (with $\lambda_0 = -\frac{1}{2}$). Then, (x^*, u^*, v^*) minimizes (1)(2) among all admissible trajectories for (2) having the same history.*

Proof. Since (x^*, u^*, v^*) is a W-stationary trajectory, there exist an arc p and measurable functions λ^G and λ^H satisfying (6). Notice that (6) implies, for almost all $t \in [0, T]$ and all $i \in \overline{m}$,

$$\lambda_i^G(t) v_i^*(t) = 0 \text{ and } \lambda_i^H(t) w_i^*(t) = 0. \tag{15}$$

Let (x, u, v) be a second admissible trajectory for (2) with the same history as (x^*, u^*, v^*). Since they both have the same history, it also satisfies, for almost all $t \in [0, T]$ and all $i \in \overline{m}$:

$$\lambda_i^G(t) v_i(t) = 0 \text{ and } \lambda_i^H(t) w_i(t) = 0. \tag{16}$$

Denote $L(x, u, v) = \frac{1}{2} \left(x^\intercal Q x + u^\intercal U u + v^\intercal V v \right)$. The goal is to prove:

$$\Delta = \int_0^T \left(L(x(t), u(t), v(t)) - L(x^*(t), u^*(t), v^*(t)) \right) dt \geq 0$$

For this, let us first transform the expression of Δ.

$$\begin{aligned}
\Delta &= \int_0^T (L(x(t),u(t),v(t)) - L(x^*(t),u^*(t),v(t)))\,dt \\
&\quad + \int_0^T p(t)^\intercal(\dot{x}(t) - Ax(t) - Bv(t) - Fu(t) - (\dot{x}^*(t) - Ax^*(t) - Bv^*(t) - Fu^*(t)))dt \\
&= \int_0^T (L(x(t),u(t),v(t)) - L(x^*(t),u^*(t),v^*(t)))\,dt + \int_0^T \frac{d}{dt}\left[p(t)^\intercal(x(t)-x^*(t))\right]dt \\
&\quad - \int_0^T \dot{p}(t)^\intercal(x(t)-x^*(t))dt \\
&\quad - \int_0^T p(t)^\intercal\left(A(x(t)-x^*(t)) + B(v(t)-v^*(t)) + F(u(t)-u^*(t))\right)dt
\end{aligned}$$

The last equality is obtained by parts integration of $\int_0^T p^\intercal(\dot{x}-\dot{x}^*)$.

$$\begin{aligned}
\Delta &= \int_0^T (L(x(t),u(t),v(t)) - L(x^*(t),u^*(t),v^*(t)))\,dt + \int_0^T \frac{d}{dt}\left[p(t)^\intercal(x(t)-x^*(t))\right]dt \\
&\quad - \int_0^T (\dot{p}(t) + A^\intercal p(t) + C^\intercal\lambda^H(t))^\intercal(x(t)-x^*(t))dt + \int_0^T (C^\intercal\lambda^H(t))^\intercal(x(t)-x^*(t))dt \\
&\quad - \int_0^T (F^\intercal p(t) + E^\intercal\lambda^H(t))^\intercal(u(t)-u^*(t))dt + \int_0^T (E^\intercal\lambda^H(t))^\intercal(u(t)-u^*(t))dt \\
&\quad - \int_0^T (B^\intercal p(t) + \lambda^G(t) + D^\intercal\lambda^H(t))^\intercal(v(t)-v^*(t))dt \\
&\quad + \int_0^T (\lambda^G(t) + D^\intercal\lambda^H(t))^\intercal(v(t)-v^*(t))dt \\
&= \int_0^T \Bigg(L(x(t),u(t),v(t)) - L(x^*(t),u^*(t),v^*(t)) \\
&\qquad\qquad - \begin{pmatrix} \dot{p}(t) + A^\intercal p(t) + C^\intercal\lambda^H(t) \\ F^\intercal p(t) + E^\intercal\lambda^H(t) \\ B^\intercal p(t) + \lambda^G(t) + D^\intercal\lambda^H(t) \end{pmatrix}^\intercal \begin{pmatrix} x(t)-x^*(t) \\ u(t)-u^*(t) \\ v(t)-v^*(t) \end{pmatrix} \Bigg)\,dt \\
&\quad + \int_0^T \lambda^H(t)^\intercal(C(x(t)-x^*(t)) + D(v(t)-v^*(t)) + E(u(t)-u^*(t)))dt \\
&\quad + \int_0^T \lambda^G(t)^\intercal(v(t)-v^*(t))dt + \int_0^T \frac{d}{dt}\left[p(t)^\intercal(x(t)-x^*(t))\right]dt
\end{aligned}$$

As it is proved in Proposition 3, $\begin{pmatrix} \dot{p}(t) + A^\intercal p(t) + C^\intercal\lambda^H(t) \\ F^\intercal p(t) + E^\intercal\lambda^H(t) \\ B^\intercal p(t) + \lambda^G(t) + D^\intercal\lambda^H(t) \end{pmatrix} = \nabla L(x^*(t),u^*(t),v^*(t))$. Since L is a convex function, it yields for almost all t in $[0,T]$:

$$L(x(t),u(t),v(t)) - L(x^*(t),u^*(t),v^*(t)) - (\nabla L(x^*(t),u^*(t),v^*(t)))^\intercal \begin{pmatrix} x(t)-x^*(t) \\ u(t)-u^*(t) \\ v(t)-v^*(t) \end{pmatrix} \geq 0 \tag{17}$$

Therefore, this proves:

$$\begin{aligned}
\Delta &\geq \int_0^T \lambda^H(t)^\intercal(w(t)-w^*(t))dt + \int_0^T \lambda^G(t)^\intercal(v(t)-v^*(t))dt + \int_0^T \frac{d}{dt}\left[p(t)^\intercal(x(t)-x^*(t))\right]dt \\
&\geq \int_0^T \frac{d}{dt}\left[p(t)^\intercal(x(t)-x^*(t))\right]dt,
\end{aligned}$$

the last inequality being obtained with (15) and (16). On top:

$$\int_0^T \frac{d}{dt}\left[p(t)^\intercal(x(t)-x^*(t))\right]dt = -\begin{pmatrix} p(0) \\ -p(T)\end{pmatrix}^\intercal \begin{pmatrix} x(0)-x^*(0) \\ x(T)-x^*(T)\end{pmatrix}$$

But the boundary conditions in (2) yield $(M\ N)\begin{pmatrix} x(0)-x^*(0) \\ x(T)-x^*(T)\end{pmatrix} = 0$, such that

$$\begin{pmatrix} x(0)-x^*(0) \\ x(T)-x^*(T)\end{pmatrix} \in \ker(M\ N) = \left(\operatorname{im}\begin{pmatrix} M^\intercal \\ N^\intercal\end{pmatrix}\right)^\perp.$$

And since $\begin{pmatrix} p(0) \\ -p(T)\end{pmatrix} \in \operatorname{im}\begin{pmatrix} M^\intercal \\ N^\intercal\end{pmatrix}$, it proves $\int_0^T \frac{d}{dt}\left[p(t)^\intercal(x(t)-x^*(t))\right]dt = 0$. Finally, we conclude that $\Delta \geq 0$. □

Remark. *One could want to get rid of the history hypothesis, since it "fixes" the switching times and does not render optimality according to these times. Very formally, it is easy to see where the problem has some leeway. Without the history hypothesis, one still can prove:*

$$\begin{aligned}\Delta \quad = \quad & \int_0^T \Bigg(L(x(t),u(t),v(t)) - L(x^*(t),u^*(t),v^*(t)) \\ & \qquad - \begin{pmatrix} \dot{p}(t) + A^\intercal p(t) + C^\intercal \lambda^H(t) \\ F^\intercal p(t) + E^\intercal \lambda^H(t) \\ B^\intercal p(t) + \lambda^G(t) + D^\intercal \lambda^H(t)\end{pmatrix}^\intercal \begin{pmatrix} x(t)-x^*(t) \\ u(t)-u^*(t) \\ v(t)-v^*(t)\end{pmatrix}\Bigg) dt \\ & + \int_0^T (\lambda^H(t)^\intercal w(t) + \lambda^G(t)^\intercal v(t))dt\end{aligned}$$

Suppose that $u(t) \neq u^*(t)$ on a measurable subset J of $[0,T]$. Then, by strict convexity of L in variable u, for almost all t in J:

$$L(x(t),u(t),v(t)) - L(x^*(t),u^*(t),v^*(t)) - (\nabla L(x^*(t),u^*(t),v^*(t)))^\intercal \begin{pmatrix} x(t)-x^*(t) \\ u(t)-u^*(t) \\ v(t)-v^*(t)\end{pmatrix} > 0$$

Therefore, using again (17), one proves:

$$\Delta > \int_0^T \lambda^H(t)^\intercal w(t)dt + \int_0^T \lambda^G(t)^\intercal v(t)dt$$

In order to simplify the problem, suppose that v and w have a different history than v^* and w^* in the neighbourhood of a single switching point t^*. Then, the inequality becomes, for some $\varepsilon > 0$:

$$\Delta > \int_{t^*-\varepsilon}^{t^*+\varepsilon} \left(\lambda^H(t)^\intercal w(t) + \lambda^G(t)^\intercal v(t)\right) dt$$

If for some $\varepsilon > 0$ small enough, $\left|\int_{t^*-\varepsilon}^{t^*+\varepsilon} \left(\lambda^H(t)^\intercal w(t) + \lambda^G(t)^\intercal v(t)\right) dt\right|$ is small enough, then $\Delta \geq 0$. Therefore, the first order conditions also render optimality according to small variations of the switching times.

Remark. *All these considerations about sufficiency of the W-stationarity still hold true if L is replaced by any other convex function, possibly non differentiable. Also, Remark 2.3 also holds the same way as long as L is strictly convex in one of its variable.*

3 Direct method

The direct method consists in discretizing directly the problem (1)(2) in order to solve a finite-dimensional optimization problem. To this aim let us propose the following Euler discretization:

$$\begin{aligned} \min \ & \sum_{k=0}^{L} x_k^\intercal Q x_k + u_k^\intercal U u_k \\ \text{s.t.} \ & \begin{cases} \dfrac{x_{k+1}-x_k}{h} = Ax_k + Bv_k + Fu_k, \ k = 0, ..., L-1 \\ 0 \leq v_k \perp Cx_k + Dv_k + Eu_k \geq 0, \ k = 0, ..., L-1 \\ Mx_0 + Nx_L = x_b, \end{cases} \end{aligned} \tag{18}$$

where $h = \frac{T}{L}$ is the time-step, considered constant. By simple application of [39, Theorem 1.4.3], for all h, this problem admits a global minimum. The discretization of the complementarity appearing in (18) differs from the implicit Euler methods found in [2, 31, 37, 16, 1]. For this optimal control problem, the complementarity should not be seen as a way to express the variable v_k, but as a mixed constraint. Therefore, its discretization must hold at all discrete times t_k, and the trajectory, solution of this discretized LCS, will be computed not step by step but for all k at once. The problem is then to solve the program (18) numerically. To this end, we use one of the two algorithms found in [36] and [38]. These algorithms and some convergence results concerning them are recalled in Section B.2. The idea behind these algorithms is to relax the complementarity, creating a sequence of optimization problems converging to a stationary point. Roughly speaking, in [36], one replaces $v^\intercal w = 0$ by $v^\intercal w \leq \varepsilon$, with $\varepsilon > 0$ converging in a certain way to 0. In [38] one augments the cost with $v^\intercal w$. A scheme detailing how to numerically solve problem (18) is presented in Figure 1. The reason to use these algorithms are that, under some hypothesis, they converge to S-stationarity points. All the codes used in this paper are available for test at https://gitlab.inria.fr/avieira/optLCS. The codes were designed using the library CasAdi [5].

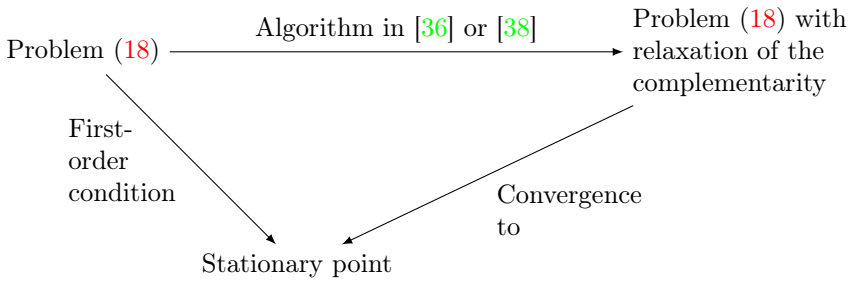


Figure 1: Sketch of the direct method for problem (1)(2).

3.1 Consistency of the scheme

Let us first compute the stationarity conditions for problem (18). Since the MPEC Linear Condition holds and if we suppose E invertible, according to [43, Theorem 2], a local minimizer must be S-stationary (see Definition 7 in Section B). We denote $\{p_i\}_{i=0}^{L-1}$ the multipliers for the discretized differential equations, $\{\theta_i\}_{i=1}^{L}$ and $\{\nu_i\}_{i=1}^{L}$ the multipliers for each side of the complementarity constraints. The stationarity conditions for the MPEC (18) read as:

$$\begin{aligned} \frac{x_{i+1}-x_i}{h} - Ax_i - Bv_i - Fu_i = \ & 0 \\ Qx_i - \left(A + \frac{1}{h}I\right)^\intercal p_i + \frac{1}{h}p_{i-1} - C^\intercal \nu_i = \ & 0 \\ Uu_i - F^\intercal p_i - E^\intercal \nu_i = \ & 0 \\ -B^\intercal p_i - \theta_i - D^\intercal \nu_i = \ & 0 \\ \theta_i = 0 \qquad & \forall i \in I^{+0}(x,u,v) \\ \nu_i = 0 \qquad & \forall i \in I^{0+}(x,u,v) \\ \nu_i \geq 0, \ \lambda_i \geq 0, \qquad & \forall i \in I^{00}(x,u,v), \end{aligned} \tag{19}$$

for all $i \in \{1, ..., L-1\}$, $h = \frac{T}{L}$, L being a fixed positive integer.

Proposition 6. *The stationarity conditions (19) define a scheme consistent with the Euler adjoint equation of an S-stationary trajectory.*

Proof. Let us check that the consistency error goes to 0 when h goes to 0. For this, we take the solutions $(x, u, v, p, \lambda^H, \lambda^G)$ at discretisation times t_i. For $k = 1, .., L-1$, let us denote ε_k^h the consistency error at time t_k :

$$\begin{aligned}\varepsilon_k^h &= \begin{pmatrix} \frac{-x(t_{k+1})}{h} + \left(A + \frac{1}{h}I\right) x(t_k) + Bv(t_k) + Fu(t_k) \\ Qx(t_k) - \left(A + \frac{1}{h}I\right)^\intercal p(t_k) + \frac{1}{h}p(t_{k-1}) - C^\intercal \lambda^H(t_k) \\ Uu(t_k) - F^\intercal p(t_k) - E^\intercal \lambda^H(t_k) \\ -B^\intercal p(t_k) - \lambda^G(t_k) - D^\intercal \lambda^H(t_k) \end{pmatrix} \\ &= \begin{pmatrix} Ax(t_k) + \dot{x}(t_k) + Bv(t_{k-1}) + Fu(t_{k-1}) + O(h) \\ Qx(t_k) - A^\intercal p(t_k) - \dot{p}(t_k) - C^\intercal \lambda^H(t_k) + O(h) \\ Uu(t_k) - F^\intercal p(t_k) - E^\intercal \lambda^H(t_k) \\ -B^\intercal p(t_k) - \lambda^G(t_k) - D^\intercal \lambda^H(t_k) \end{pmatrix} = \begin{pmatrix} O(h) \\ O(h) \\ 0 \\ 0 \end{pmatrix}.\end{aligned} \tag{20}$$

It follows that $\lim_{h\to 0} \left(\max_{k=1,...,L} \|\varepsilon_k^h\|_\infty\right) = 0$. On top of that, discrete multipliers ν and θ respect the same equality and inequality conditions as the multipliers λ^H and λ^G of a S-stationary trajectory at discrete times t_k. □

3.2 Numerical examples

This section is devoted to illustrate the computation of optimal trajectories with the direct method (18).

3.2.1 Analytical 1D example

Let us apply the direct method on the following example:

Example 1.

$$\begin{aligned} &\textit{minimize} \int_0^T \left(x(t)^2 + u(t)^2\right) dt, \\ &\textit{such that:} \begin{cases} \dot{x}(t) = ax(t) + bv(t) + fu(t), \\ 0 \le v(t) \perp dv(t) + eu(t) \ge 0, \ \textit{a.e. on } [0,T] \\ x(0) = x_0, \ x(T) \textit{ free}, \end{cases} \end{aligned} \tag{21}$$

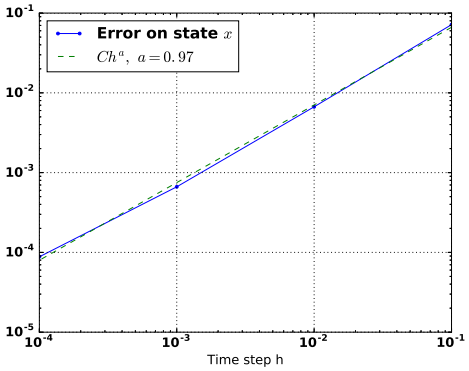
where all variables are scalars, $d > 0$, $b, e \neq 0$. Using the results in [15], the constants in the LCS are chosen such that the system is completely controllable. As proved in Section C, the only stationary trajectory is given by:

$$\begin{cases} x^*(t) = \dot{p}(t) + ap(t) \\ u^*(t) = \begin{cases} fp(t) & \textit{if } efx(0) \le 0, \\ \left(f - \frac{eb}{d}\right) p(t) & \textit{if } efx(0) \ge 0. \end{cases} \\ v^*(t) = \frac{1}{d} \max(0, -eu^*(t)) \end{cases} \tag{22}$$

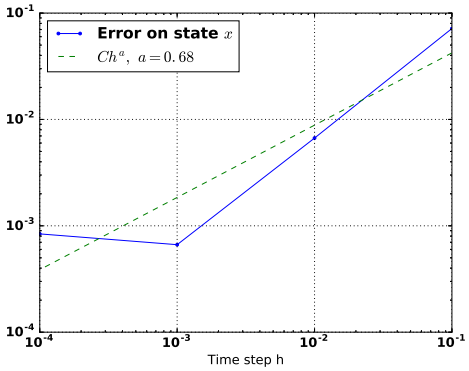
where:

$$p(t) = \frac{1}{2\sqrt{\gamma}} \left[\left((\sqrt{\gamma} - a)e^{\sqrt{\gamma}t} + (\sqrt{\gamma} + a)e^{-\sqrt{\gamma}t} \right) p(0) + \left(e^{\sqrt{\gamma}t} - e^{-\sqrt{\gamma}t} \right) x(0) \right]$$

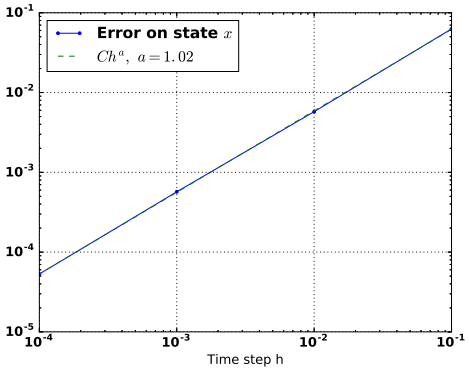
$p(0) = -\frac{x(0)\left(e^{2\sqrt{\gamma}T} - 1\right)}{(\sqrt{\gamma} - a)e^{2\sqrt{\gamma}T} + \sqrt{\gamma} + a}$ *and* $\gamma = \begin{cases} \left(a^2 + f^2\right) & \textit{if } efx(0) \le 0, \\ \left(a^2 + \left(f - \frac{be}{d}\right)^2\right) & \textit{if } efx(0) \ge 0. \end{cases}$. *Figures 2-4 show the evolution of error with time-step in log-log scales, using the two different algorithms, with the parameters $a = 3$, $b = -0.5$, $d = 1$, $e = -2$, $f = 3$, $T = 1$, and either $x_0 = 1$ or $x_0 = -1$. In these examples, we clearly see convergence of both*



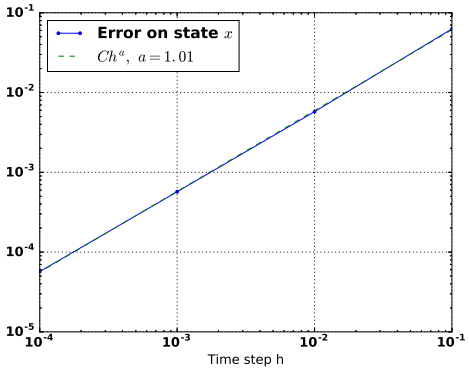
(a) Algorithm in [36], $x_0 = -1$



(b) Algorithm in [38], $x_0 = -1$



(c) Algorithm in [36], $x_0 = 1$



(d) Algorithm in [38], $x_0 = 1$

Figure 2: Error on state x when using Algorithms in [36] or in [38] with $a = 3$, $b = -0.5$, $d = 1$, $e = -2$, $f = 3$, $T = 1$ and $x_0 = -1$ or $x_0 = 1$ in Example 1.

algorithms, with an order close to 1. However, Figures 2b and 4c suggest that in some cases, the algorithms face difficulties when the time-step is decreasing. This is actually something known with direct methods: often they fail to be precise. We can simply understand it, since decreasing the time-step increases the dimension of the optimization problem to solve. In order to tackle such problems, one has to choose a different method presented in Section 4.

3.2.2 Example with $D = 0$

As alluded to in the introduction, a crucial parameter in LCS is the relative degree between w and v, which determines the solution set as a subclass of Schwarz' distributions [3]. Let us consider now a case with $D = 0$ and relative degree one.

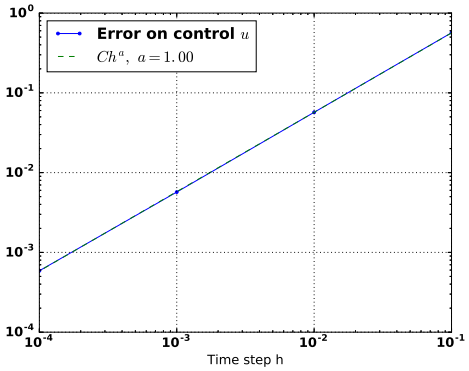
Example 2.

$$\begin{aligned} &\textit{minimize} \int_0^T \left(\|x(t)\|_2^2 + u(t)^2\right) dt \\ &\textit{such that:} \begin{cases} \dot{x}(t) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x(t) + \begin{pmatrix} -1 \\ 1 \end{pmatrix} v(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \\ 0 \le v(t) \perp \begin{pmatrix} -1 & 1 \end{pmatrix} x(t) + u(t) \ge 0, \\ x(0) = (-0.5, 1), \ x(T) \textit{ free.} \end{cases} \end{aligned} \tag{23}$$

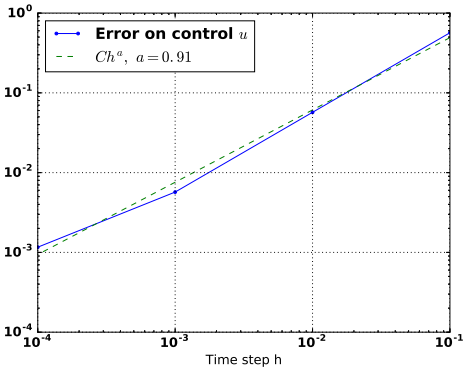
The numerical results for Example 2 are shown in Figure 5. They demonstrate that the direct method can also succeed when D is not a P-matrix.

3.2.3 Higher dimensional examples

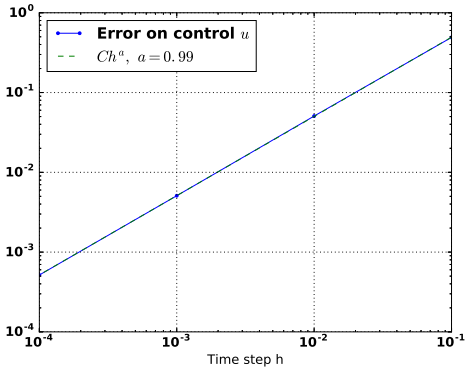
For higher dimension or when $C \neq 0$, we do not have an analytical solution to compare with the numerical one, but still we can check if the multipliers comply with an S-stationary trajectory. For this purpose, let us test them



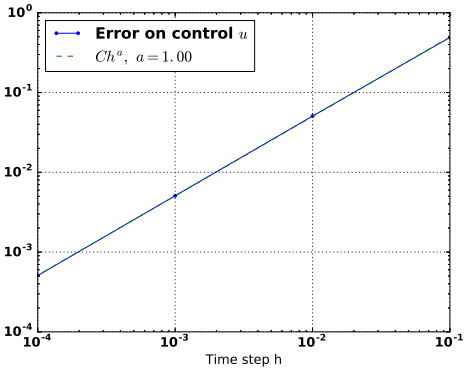
(a) Algorithm in [36], $x_0 = -1$



(b) Algorithm in [38], $x_0 = -1$



(c) Algorithm in [36], $x_0 = 1$



(d) Algorithm in [38], $x_0 = 1$

Figure 3: Error on control u when using Algorithms in [36] or in [38] with $a = 3$, $b = -0.5$, $d = 1$, $e = -2$, $f = 3$, $T = 1$ and $x_0 = -1$ or $x_0 = 1$ in Example 1.

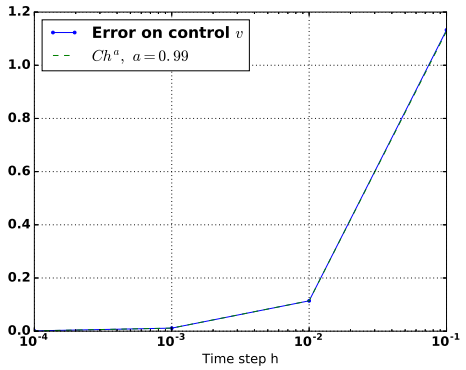
on a third example:

Example 3.

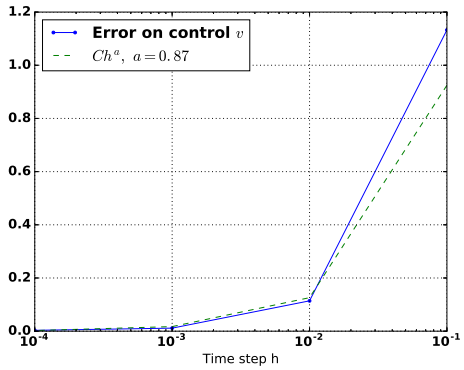
$$
\begin{aligned}
&\textit{minimize} \int_0^1 \left(\|x(t)\|_2^2 + 25\|u(t)\|_2^2\right) dt, \\
&\textit{such that:} \begin{cases}
\dot{x}(t) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} x(t) + \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} v(t) + \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} u(t), \\
0 \le v(t) \perp \begin{pmatrix} 3 & -1 \\ -2 & 0 \end{pmatrix} x(t) + v(t) + \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} u(t) \ge 0, \quad \textit{a.e. on } [0,1] \\
x(0) = \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}, \ x(T) \textit{ free},
\end{cases}
\end{aligned}
\tag{24}
$$

where x, u and v are functions $[0,1] \to \mathbb{R}^2$. As shown in Figure 6, the Algorithm in [38] seems to fail to respect the complementarity condition between v_2 and w_2 at the beginning. The Algorithm in [36] seems to behave better. Comparing first Figure 7b and Figure 6a, then Figure 7c and Figure 6c, results suggest that we retrieve an S-stationary trajectory (according to the sign of the multipliers, and their complementarity with v and w), as expected.

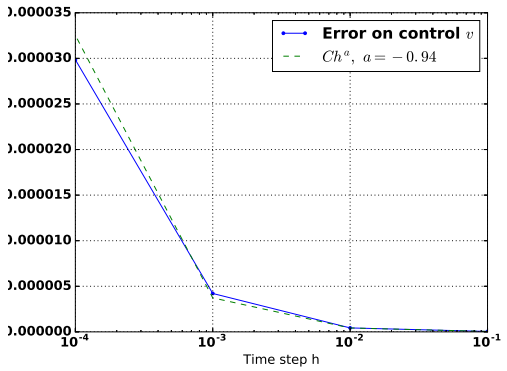
Since v is not upper-bounded nor present in the running cost in (1), the optimal trajectory may present big variations due to v. It is the case for the following Example 4, where x takes values in $\mathbb{R}^2$, u and v take values in $\mathbb{R}$.



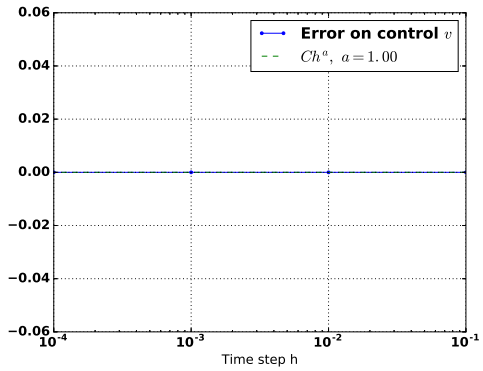
(a) Algorithm in [36], $x_0 = -1$



(b) Algorithm in [38], $x_0 = -1$



(c) Algorithm in [36], $x_0 = 1$



(d) Algorithm in [38], $x_0 = 1$

Figure 4: Error on v when using Algorithms in [36] or in [38] with $a = 3$, $b = -0.5$, $d = 1$, $e = -2$, $f = 3$, $T = 1$ and $x_0 = -1$ or $x_0 = 1$ in Example 1.

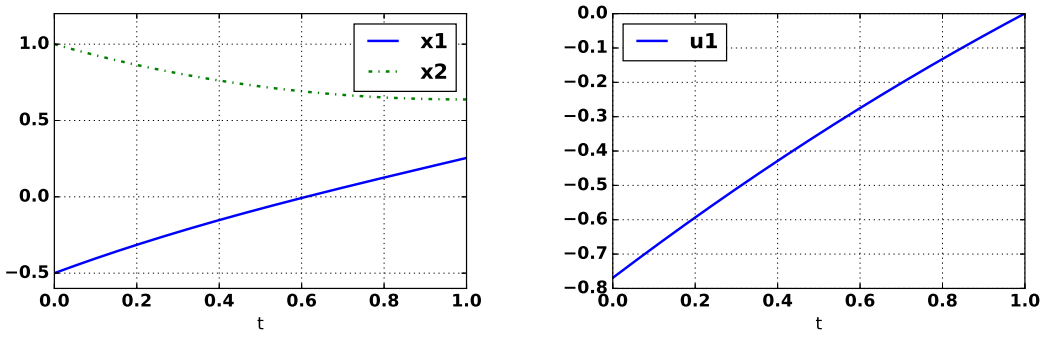
Example 4.

$$\begin{aligned} &\textit{minimize} \int_0^1 \left(\|x(t)\|_2^2 + u(t)^2\right) dt, \\ &\textit{such that:} \begin{cases} \dot{x}(t) = \begin{pmatrix} 5 & -6 \\ 3 & 9 \end{pmatrix} x(t) + \begin{pmatrix} 4 \\ 5 \end{pmatrix} v(t) + \begin{pmatrix} 0 \\ -4 \end{pmatrix} u(t), \\ 0 \le v(t) \perp \begin{pmatrix} -1 & 5 \end{pmatrix} x(t) + v(t) + 6u(t) \ge 0, \qquad \textit{a.e. on } [0,1] \\ x(0) = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}, \ x(T) \textit{ free}, \end{cases} \end{aligned} \tag{25}$$

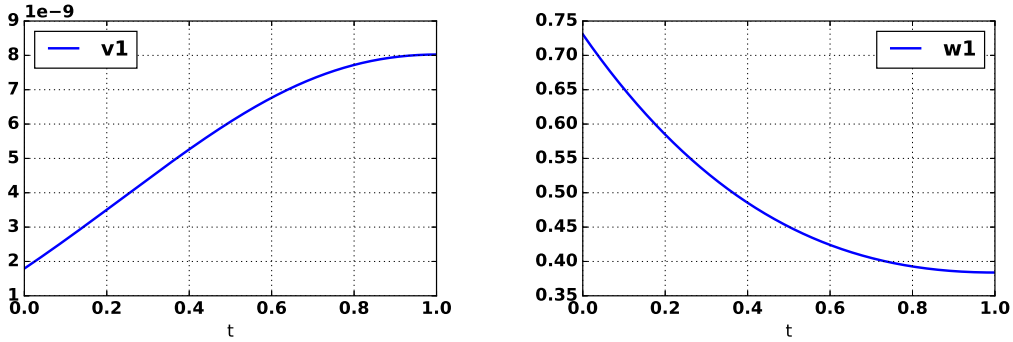
As shown in Figure 8, the optimal solution admits a peak on v at the very beginning of the interval. One could think that the state x admits a jump, which could mean that the solution of the LCS is distributional (in which case the dynamics in (2) has to be recast into measure differential inclusions), but shrinking the time-step does not change this peak on v, which is always positive on a non-shrinking interval whatever the time-step h.

One could wonder what happens in Example 4 if a quadratic cost $v^\intercal V v$ (with V symmetric positive definite) is added in the running cost. This could prevent the initial huge peak on v. This is the investigation of Example 5. The code has been slightly changed in order to add a quadratic cost in v. Concerning the first order conditions in Theorem 2, simple calculations show that adding this quadratic cost just turns the optimal control into $u^*(t) = U^{-1}\left(F^\intercal p(t) + Vv^*(t) + E^\intercal \tilde{\beta}(t) - (\zeta(t) + r)E^\intercal v^*(t)\right)$ and it changes the matrix of the LCP appearing in the Euler equation (14), becoming:

$$\begin{pmatrix} EU^{-1}E^\intercal & D - (\zeta + r)EU^{-1}E^\intercal \\ \zeta EU^{-1}E^\intercal - D^\intercal & \zeta(D+V) + (\zeta + r)\left(D^\intercal - \zeta EU^{-1}E^\intercal\right) \end{pmatrix}$$



(a) State x (b) Control u



(c) Control v (d) Variable w

Figure 5: Numerical results for Example 2 using Algorithm in [36], $h = 10^{-3}$.

Example 5.

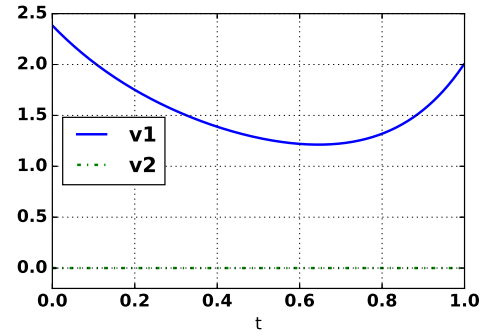
$$\begin{array}{l} \textit{minimize} \displaystyle\int_0^1 \left(\|x(t)\|_2^2 + u(t)^2 + \alpha v(t)^2\right) dt, \\ \textit{such that:} \begin{cases} \dot{x}(t) = \begin{pmatrix} 5 & -6 \\ 3 & 9 \end{pmatrix} x(t) + \begin{pmatrix} 4 \\ 5 \end{pmatrix} v(t) + \begin{pmatrix} 0 \\ -4 \end{pmatrix} u(t), \\ 0 \le v(t) \perp \begin{pmatrix} -1 & 5 \end{pmatrix} x(t) + v(t) + 6u(t) \ge 0, \qquad \textit{a.e. on } [0,1] \\ x(0) = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}, \ x(T) \textit{ free}, \end{cases} \end{array} \tag{26}$$

where $\alpha > 0$. The numerical results are shown Figure 9, and a special focus on v for $\alpha \in \{10, 5, 10^{-1}, 10^{-3}, 0\}$ is shown Figure 10 for $t \in [0, 0.1]$. We clearly see a continuity property of the solution with respect to α when it shrinks to 0. Adding this quadratic cost on v may then be a way to smoothen the solution, getting rid of the initial huge peak.

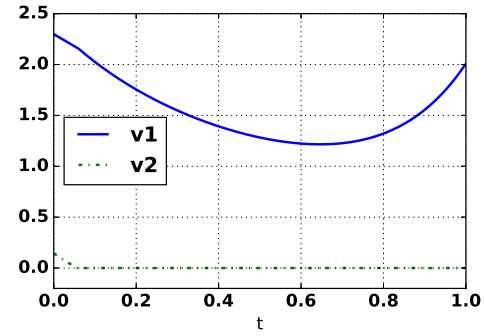
In the previous numerical simulations, the optimal control seems always continuous. The next example suggests that the optimal control may jump. Let us consider the following problem, where for all t in $[0, 1]$, $x(t) \in \mathbb{R}^2$ and $u(t) \in \mathbb{R}$.

Example 6. *[Discontinuous optimal controller]*

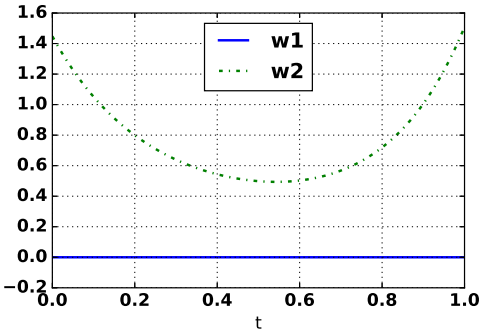
$$\begin{array}{l} \textit{minimize} \displaystyle\int_0^1 \left(\|x(t)\|_2^2 + u(t)^2\right) dt, \\ \textit{such that:} \begin{cases} \dot{x}(t) = \begin{pmatrix} 1 & -3 \\ -8 & 10 \end{pmatrix} x(t) + \begin{pmatrix} -3 \\ -1 \end{pmatrix} v(t) + \begin{pmatrix} 4 \\ 8 \end{pmatrix} u(t), \\ 0 \le v(t) \perp \begin{pmatrix} 1 & -3 \end{pmatrix} x(t) + 5v(t) + 3u(t) \ge 0, \qquad \textit{a.e. on } [0,1] \\ x(0) = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}, \ x(T) \textit{ free}, \end{cases} \end{array} \tag{27}$$



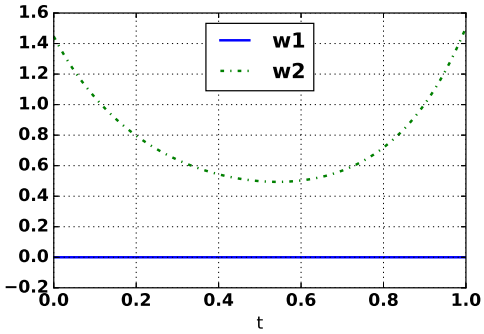
(a) v, Algorithm in [36]



(b) v, Algorithm in [38]

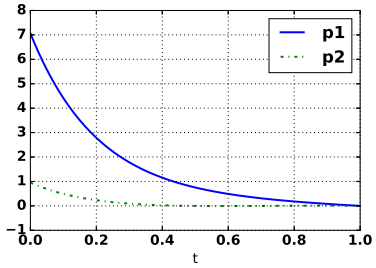


(c) w, Algorithm in [36]

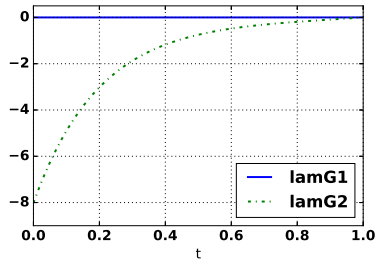


(d) w, Algorithm in [38]

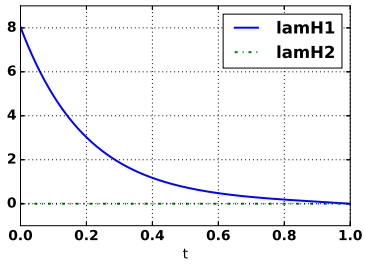
Figure 6: Numerical results for Example 3 using Algorithms in [36] and in [38], for comparison concerning complementarity. $h = 10^{-3}$.



(a) Adjoint state p



(b) Multiplier λ^G

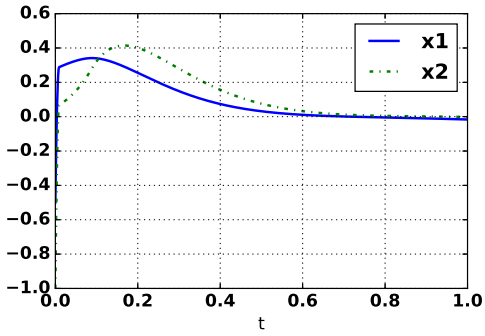


(c) Multiplier λ^H

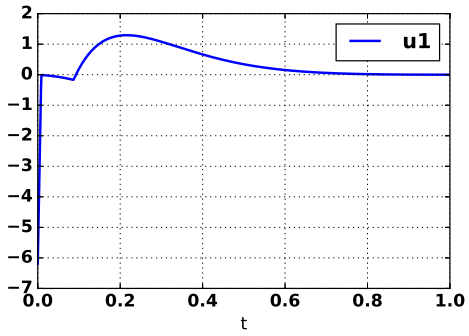
Figure 7: Computed multipliers for Example 3 using Algorithm in [36]. $h = 10^{-3}$.

The numerical results are shown in Figure 11. The associated multipliers and adjoint state, retrieved from these calculations, are shown in Figure 12. The complementarity constraint is satisfied, and the associated multipliers suggest that the trajectory indeed is an S-stationary trajectory. It is clear that u admits a switch around $t_1 = 0.112$ and is not continuous (see Figure 11b). It is noteworthy to take a look at the different modes activated along the solution. In this case where the complementarity constraint is of dimension 1, we have three possible cases : $v = 0 < w$ (happening on $[0, t_1]$), $v > 0 = w$ (happening on $[t_1, 0.87]$ approximately), $v = 0 = w$ (happening on $[0.87, 1]$). It shows that, compared with some other methods for optimal control of switching systems (see for instance [41, 46]), this method does not require to guess a priori the number of switches nor the times of commutation in order to approximate the solution. The tracking of the switches is taken care of by the MPEC solver. This is a major advantage of the complementarity approach over event-driven, hybrid-like approaches.

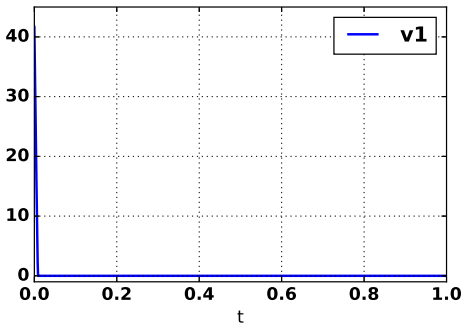
Eventually, the class of solution considered may actually be too small, and the direct method may converge to a solution with the state admitting jumps. This is the main focus of the Example 7.



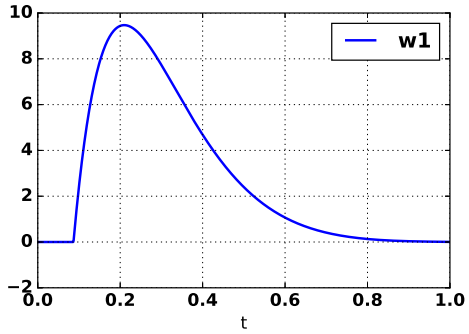
(a) State x



(b) Control u

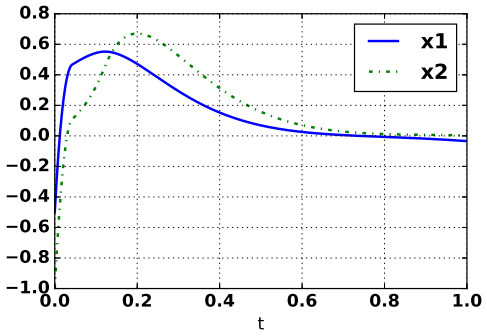


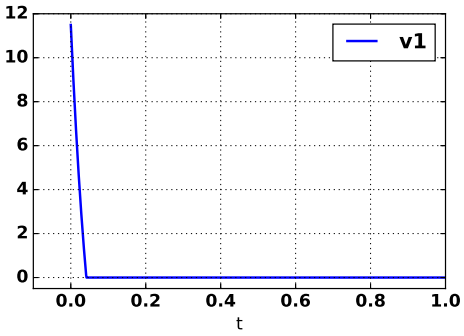
(c) Control v



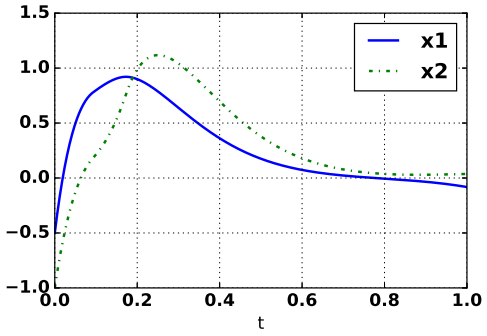
(d) Variable w

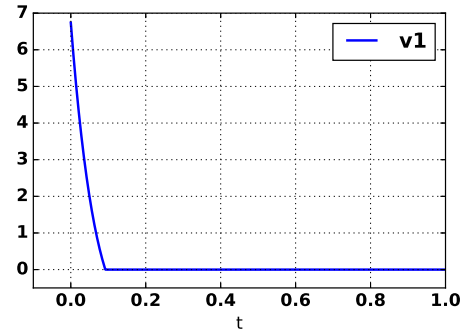
Figure 8: Numerical results for Example 4 using Algorithm in [36], $h = 10^{-3}$.





(a) $\alpha = 1$





(b) $\alpha = 10$

Figure 9: Numerical x and v found for Example 5 using Algorithm in [36], $h = 10^{-3}$, and different values of α.

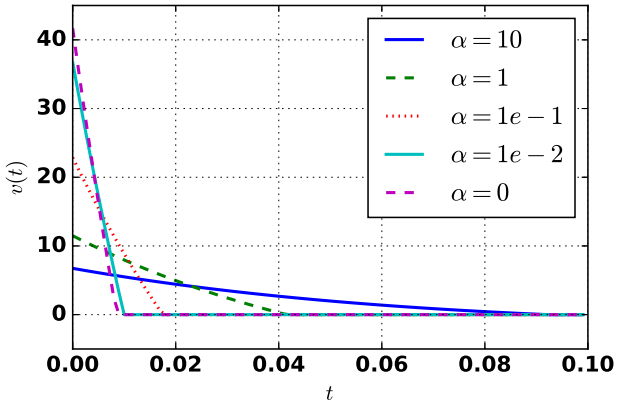
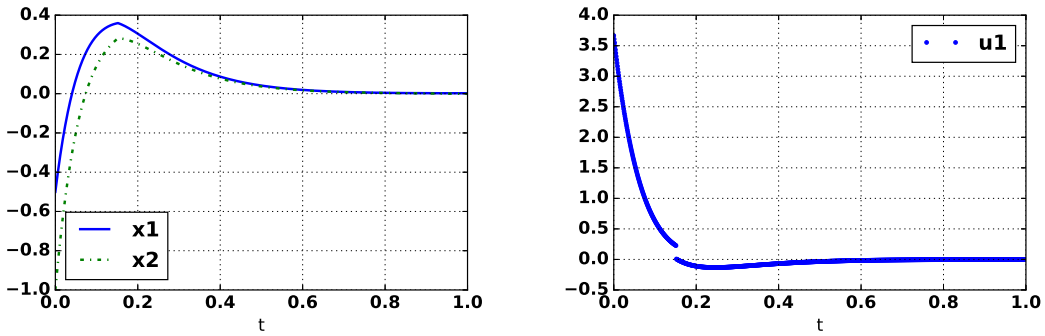
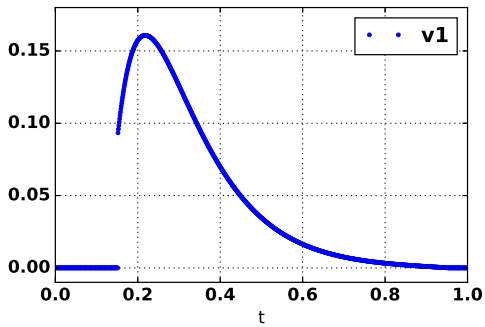


Figure 10: Numerical v on $[0, 0.1]$ found for Example 5 using Algorithm in [36], $h = 10^{-3}$, and different values of α.

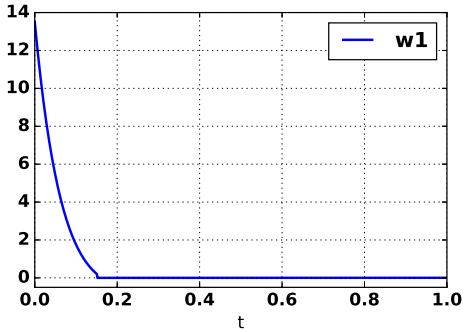


(a) State x

(b) Control u



(c) Control v

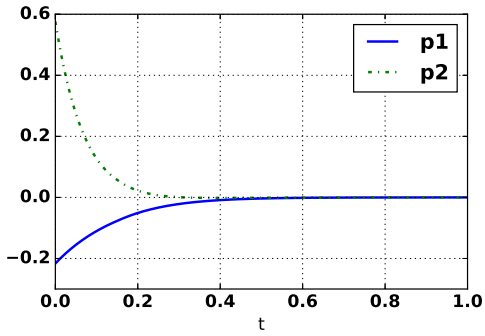


(d) Variable w

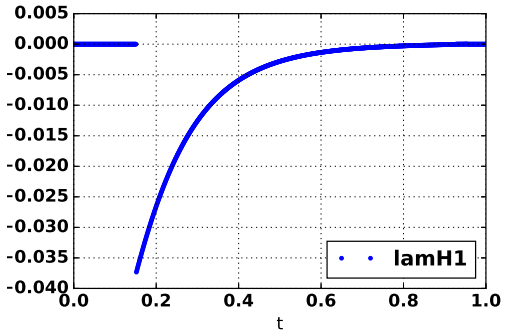
Figure 11: Numerical results for Example 6 using Algorithm in [36], $h = 10^{-3}$.

Example 7.

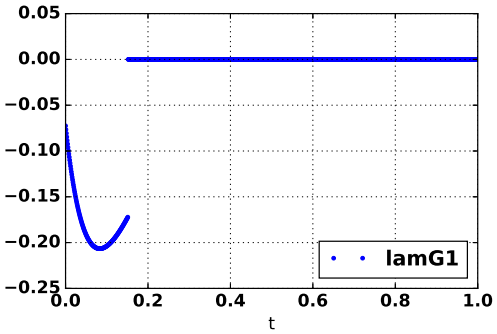
$$
\begin{aligned}
&\textit{minimize} \int_0^{10} \left(\|x(t)\|_2^2 + u(t)^2 + \alpha v(t)^2 \right) dt, \\
&\textit{such that:} \begin{cases} \dot{x}(t) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} x(t) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} v(t) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} u(t), \\ 0 \le v(t) \perp \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} x(t) + u(t) \ge 0, \qquad \textit{a.e. on } [0, 10] \\ x(0) = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}, \ x(T) \textit{ free}, \end{cases}
\end{aligned} \tag{28}
$$



(a) Adjoint state p



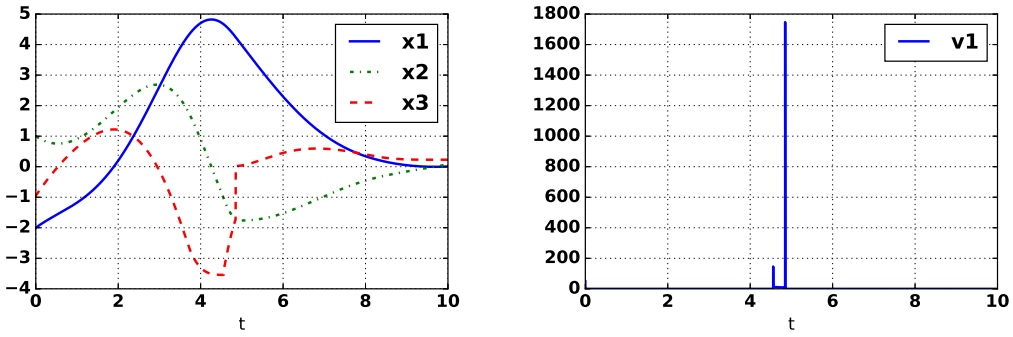
(b) Multiplier λ^H



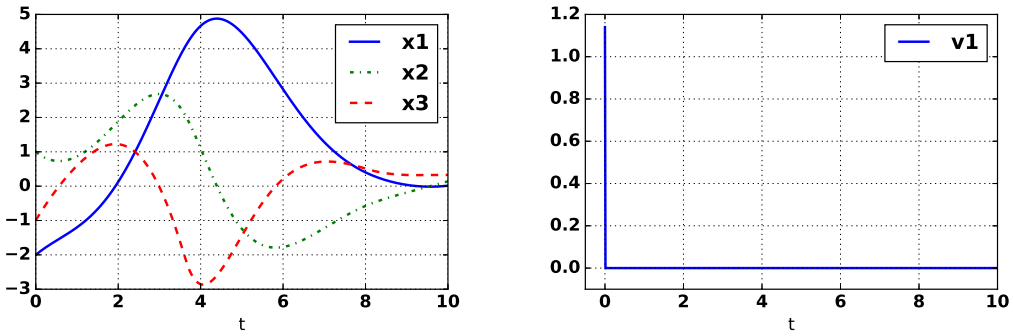
(c) Multiplier λ^G

Figure 12: Numerical results for Example 6 using Algorithm in [36], $h = 10^{-3}$.

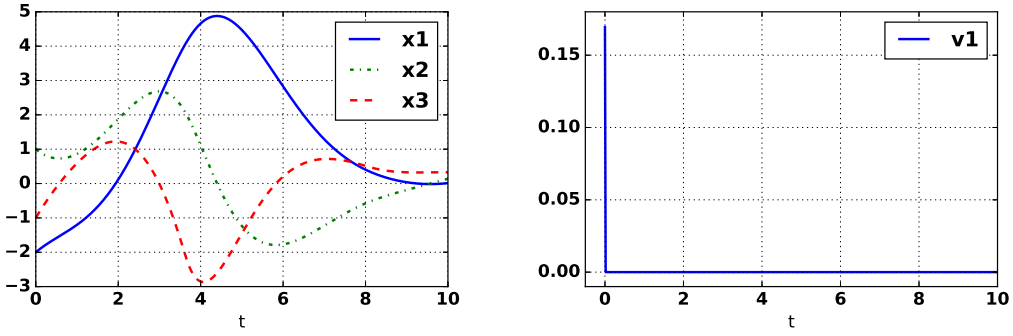
with $\alpha \in \{0, 1, 10\}$. *As shown Figure 13, the solution with* $\alpha = 0$ *admits a huge peak around* $t = 4.85$, *that yields a jump on* x_3. *When* $\alpha > 0$, *this peak disappears, but a smaller at* $t = 0$ *is recovered. Even though adding* v *in the running cost smoothen the solution, it shows that the optimal solution still admits huge variation.*



(a) $\alpha = 0$



(b) $\alpha = 1$



(c) $\alpha = 10$

Figure 13: Numerical results for x and v for Example 7 using Algorithm in [36], $h = 10^{-3}$, using different values for α.

4 Combining direct and indirect methods: the hybrid approach

The indirect method consists in solving the first-order necessary conditions derived in Section 2.2 in order to solve the optimal control problem. As pointed out in [52], it has the advantage that the numerical solutions are usually very precise, but the method suffers from a huge sensitivity on the initial guess. Indeed, if the initial guess is not close enough to a solution, then the method may fail to converge.
A natural approach is then to use both the direct and the indirect methods in order to obtain a very precise solution, taking advantage of both methods: this is called the hybrid approach. In our framework, we have to face two problems. First, the active index sets appearing in the Euler equations (6), used to impose conditions on the multipliers, are not useful as they are. This problem has been tackled by re-expressing these equations in Theorem 2. Secondly, we often have to solve a Boundary Value Problem (BVP). This is the case for instance in Example 1. Since $x(T)$ is free, the transversality conditions impose in that case $p(T) = 0$. The problem is then to find a solution (x, p) of (14) such that $x(0) = x_0$ and $p(T) = 0$. Finding such a solution is not trivial, especially in this case since the dynamical system is an LCS.

4.1 BVP solver for the indirect method

Theorem 2 implies that the optimal solution is the projection of an extremal, which is a solution of (14) with boundary values. Denote by $z = \binom{x}{p}$ and suppose we can rewrite boundary values on z as a linear equation $\tilde{M}z(0) + \tilde{N}z(T) = \tilde{x}_b$. Then Theorem 2 implies that the extremal is a solution of the Boundary Value Problem (BVP):

$$\begin{cases} (a)\ \dot{z} = \mathcal{A}z + \mathcal{B}\lambda \\ (b)\ 0 \leq \lambda \perp \mathcal{D}\lambda + \mathcal{C}z \geq 0 \\ (c)\ \mathcal{E}^\intercal \lambda \geq 0 \\ (d)\ \tilde{M}z(0) + \tilde{N}z(T) = \tilde{x}_b, \end{cases} \tag{29}$$

where the matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ are easily identifiable from (14), and $\lambda = \binom{\tilde{\beta}}{v}$. This is a Boundary Value Problem (BVP) formulated for an LCS with constraint (29)(c). The shooting method is usually employed to solve such a problem: roughly speaking, given $z_0 \in \mathbb{R}^{2n}$, we compute the solution $z(\cdot; z_0)$ of (29)(a)(b)(c) with initial data $z(0) = z_0$. Letting $F(z_0) = \tilde{M}z_0 + \tilde{N}z(T; z_0) - \tilde{x}_b$, the BVP becomes a root-search of F. In practice, we employ multiple shooting : we also take into account in F shooting nodes inside the interval $[0, T]$, where we make sure that $z(\cdot, z_0)$ is continuous. In the smooth case, we would use a Newton method, which needs the Jacobian $F'(z_0)$ to compute each iteration. In our case, the dependence on z_0 of $z(T; z_0)$ is not smooth. Some properties concerning such dependence for LCS have been derived in [40]. Therein, the authors built a linear Newton Approximation, which allow them to design a non-smooth Newton method for solving BVP for LCS. However, their result can not be directly applied here for two reasons. First, on top of the complementarity conditions, we also have to take into account the inequality condition (29)(c). Secondly, their result relies on the fact that $\mathcal{B}\ \mathrm{SOL}(\mathcal{C}z(t), \mathcal{D})$ is a singleton for all $t \in [0, T]$. However, this method still could work for (29), since the research will only be local. Section 4.2 shows numerical results where this non-smooth Newton method has been used successfully.

4.2 An efficient method to solve the LCS

In the first simulations we ran, we noticed that the integration step by step of the LCS IVP (29)(a)(b)(c), $z(0) = z_0$, admitted some numerical instability that multiple shooting could not solve. This problem was solved using the following proposition:

Proposition 7. *Let t_i, $t_{i+1} \in [0, T]$, $t_i < t_{i+1}$. (z, λ) is a solution of*

$$\begin{cases} \dot{z} = \mathcal{A}z + \mathcal{B}\lambda \\ 0 \leq \lambda \perp \mathcal{D}\lambda + \mathcal{C}z \geq 0 \\ \mathcal{E}^\intercal \lambda \geq 0 \\ z(t_i) = z_i, \end{cases} \tag{30}$$

on $[t_i, t_{i+1}]$*, if and only if it is a global minimum of the optimal control problem:*

$$\begin{aligned} \min \quad & \int_{t_i}^{t_{i+1}} \lambda(t)^\intercal (\mathcal{D}\lambda(t) + \mathcal{C}z(t))dt \\ s.t. \quad & \left.\begin{array}{l} \dot{z}(t) = \mathcal{A}z(t) + \mathcal{B}\lambda(t) \\ \lambda(t) \geq 0 \\ \mathcal{D}\lambda(t) + \mathcal{C}z(t) \geq 0 \\ \mathcal{E}\lambda(t) \geq 0 \end{array}\right\} \text{a.e. on } [t_i, t_{i+1}] \\ & z(t_i) = z_i \end{aligned} \tag{31}$$

with minimum equal to 0.

Proof. ($\Longrightarrow$) Suppose (z, λ) is a solution of (30), then (z, λ) is obviously admissible for (31), and $\int_{t_i}^{t_{i+1}} \lambda(t)^\intercal (\mathcal{D}\lambda(t) + \mathcal{C}z(t))dt = 0$. Suppose there exists an admissible solution $(\tilde{z}, \tilde{\lambda})$ of (31) such that $\int_{t_i}^{t_{i+1}} \tilde{\lambda}(t)^\intercal (\mathcal{D}\tilde{\lambda}(t) + \mathcal{C}\tilde{z}(t))dt < 0$. It then means that there exists $\tau_1, \tau_2 \in [t_i, t_{i+1}]$ such that $[\tau_1, \tau_2]$ is of positive measure and $\tilde{\lambda}(t)^\intercal (\mathcal{D}\tilde{\lambda}(t) + \mathcal{C}\tilde{z}(t)) < 0$ a.e. on $[\tau_1, \tau_2]$. This contradicts the fact that $\tilde{\lambda} \geq 0$ and $\mathcal{D}\tilde{\lambda} + \mathcal{C}\tilde{z} \geq 0$ a.e. on $[t_i, t_{i+1}]$. Then the minimum is non-negative, and (λ, z) is a global minimum.
($\Longleftarrow$) Suppose (z, λ) is a solution of (31). Notice that $\lambda(t)^\intercal (\mathcal{D}\lambda(t) + \mathcal{C}z(t)) \geq 0$ a.e. on $[t_i, t_{i+1}]$, so $\int_{t_i}^{t_{i+1}} \lambda(t)^\intercal (\mathcal{D}\lambda(t) + \mathcal{C}z(t))dt = 0$ implies that $\lambda(t)^\intercal (\mathcal{D}\lambda(t) + \mathcal{C}z(t)) = 0$ a.e. on $[t_i, t_{i+1}]$. So (z, λ) is a solution of (30). □

Numerically, problem (31) will be solved for each interval $[t_i, t_{i+1}]$ using a classical direct method, where t_i is a node for the Multiple Shooting method. One could ask why this formulation is more stable than just discretizing directly equation (30). Intuitively, we can explain it as follows:

- Using for instance an implicit Euler a discretization of (30), one solves at each step the problem:

$$\begin{aligned} & z_{k+1} - z_k = h\left(\mathcal{A}z_{k+1} + \mathcal{B}\lambda_{k+1}\right) \\ & 0 \leq \lambda_{k+1} \perp \mathcal{D}\lambda_{k+1} + \mathcal{C}z_{k+1} \geq 0 \\ & \mathcal{E}^\intercal \lambda_{k+1} \geq 0, \end{aligned}$$

  which takes the form of an LCP with unknown λ_{k+1} and an inequality constraint. But the exact solution $(z^*_{k+1}, \lambda^*_{k+1})$ will not be found. Instead, an approximated solution $(z_{k+1}, \lambda_{k+1}) = (z^*_{k+1} + \varepsilon_k, \lambda^*_{k+1} + \varepsilon_\lambda)$ will be sought. Then the error will propagate along the solution on $[0, T]$, causing instabilities.

- However, if one solves (31), all errors will appear under the integral sign. Since this integral is minimized (and we expect the result to be 0), the errors on the whole interval can also be expected to be minimized.

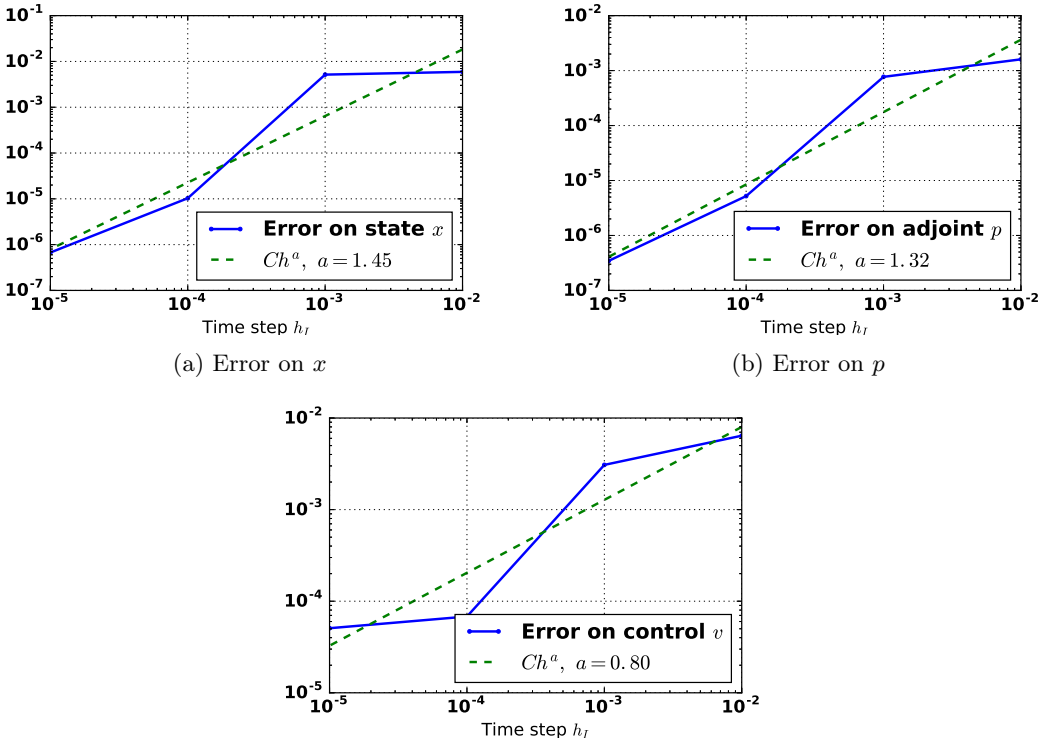
4.3 Numerical results

4.3.1 Analytical example revisited

First, let us check the convergence of the method of Section 4.1 and 4.2 on the 1D Example 1. Since the Direct Method achieved to reach a satisfactory precision, one can expect also the Indirect method to converge. The results are presented in Figures 14 and 15. Overall, the method reaches the precision of the time step, even for very small precision. Concerning the state x and the adjoint state p, the convergence is even faster. Concerning λ^H, λ^G and v, it seems however harder to converge. But still, the desired precision is met, and it is often more precise than the Direct Method.

4.3.2 Example 3 revisited

In order to compare the Hybrid Approach with the raw direct method, we ran simulations on Example 3 using different time-steps, and comparing the time spent for solving it at the desired precision. The results are shown in Tables 1 and 2. It appears clearly that, even though the Indirect method is not that interesting for rough precisions, it becomes necessary for really high precisions. The Newton Method developed in this context is also satisfying, as shown in Figure 16, which shows the maximum gap left on x and p at shooting nodes. The program assumes to reach convergence as soon as the continuity on $\binom{x}{p}$ is met with precision h. As shown in this example, convergence is achieved in two iterations.



(a) Error on x

(b) Error on p

(c) Error on v

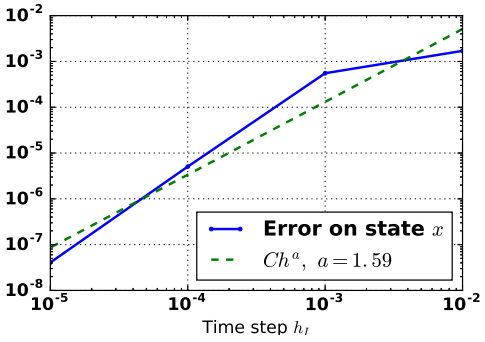
(d) Error on λ^H

(e) Error on λ^G

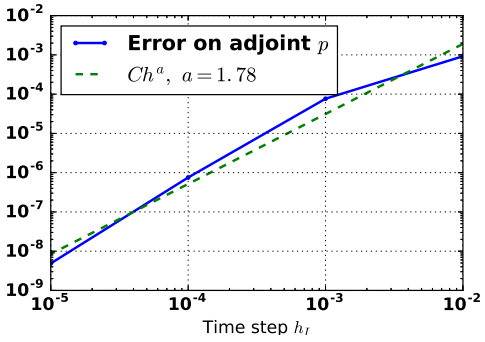
Figure 14: Errors with the Hybrid Approach for Example 1 with $a = 3$, $b = -0.5$, $d = 1$, $e = -2$, $f = 3$, $T = 1$ and $x_0 = -1$.

5 Conclusion

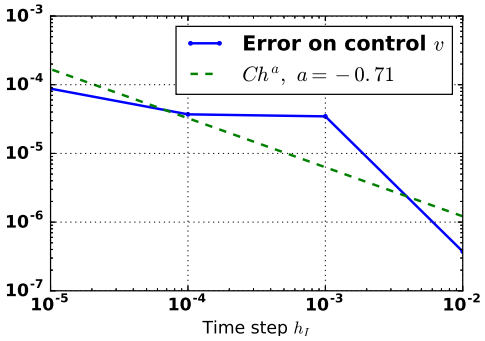
This paper focuses on the quadratic optimal control of Linear Complementarity Systems. Necessary first-order conditions are presented, and two numerical algorithms providing fast and accurate numerical approximation are developed and proposed. Several examples prove the efficiency of the approach. Future investigations should concern several aspects: (i) The addition of a quadratic term in v (the multiplier) in the cost, in order to cope with the fact that this signal is not necessarily bounded. This creates instability in the computations, and could be related with possible state jumps (which are outside the scope of this work); (ii) The optimal solution found here is an open loop solution. Trajectory tracking algorithms for LCS should be studied to stabilize the optimal solution. Closed-loop optimal control via the Hamilton-Jacobi equations might also be analysed. (iii) All the results developed in this article assume that the state x is absolutely continuous. Extension towards larger solutions sets is another topic of future research.



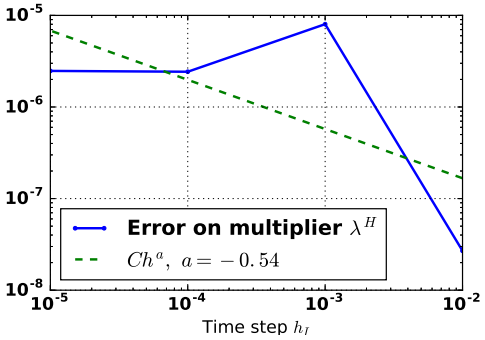
(a) Error on x



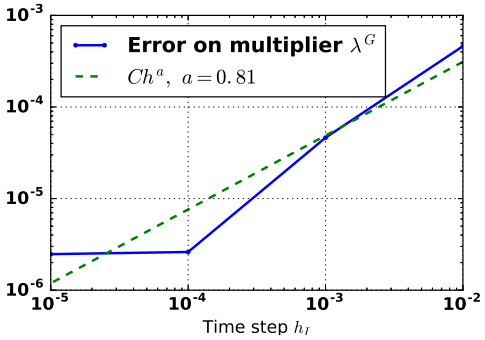
(b) Error on p



(c) Error on v



(d) Error on λ^H



(e) Error on λ^G

Figure 15: Errors with the Hybrid Approach for Example 1 with $a = 3$, $b = -0.5$, $d = 1$, $e = -2$, $f = 3$, $T = 1$ and $x_0 = 1$.

References

[1] V. Acary, O. Bonnefon, and B. Brogliato. Time-stepping numerical simulation of switched circuits within the nonsmooth dynamical systems approach. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 29(7):1042–1055, 2010.

[2] V. Acary, O. Bonnefon, and B. Brogliato. *Nonsmooth Modeling and Simulation for Switched Circuits*, volume 69 of *Lecture Notes in Electrical Engineering*. Springer Science & Business Media, 2011.

[3] V. Acary, B. Brogliato, and D. Goeleven. Higher order Moreau's sweeping process: Mathematical formulation and numerical simulation. *Mathematical Programming A*, 113:133–217, 2008.

[4] V. Acary, H. de Jong, and B. Brogliato. Numerical simulation of piecewise-linear models of gene regulatory networks using complementarity systems. *Physica D: Nonlinear Phenomena*, 269:103–119, 2014.

h_D	Time spent (s)
10^{-2}	1.31
10^{-3}	37.50
10^{-4}	400.65
10^{-5}	∞
10^{-6}	∞

Table 1: Time spent for computing an approximate solution of Example 3 using the direct method, with different time steps h_D. ∞ means that the calculations did not end (segmentation fault).

Parameters	Time spent (s)
$h_D = 10^{-1}$, $h_I = 10^{-2}$, $n_S = 5$	1.39
$h_D = 10^{-1}$, $h_I = 10^{-3}$, $n_S = 10$	11.26
$h_D = 10^{-2}$, $h_I = 10^{-4}$, $n_S = 20$	97.56
$h_D = 10^{-3}$, $h_I = 10^{-5}$, $n_S = 50$	1 298.62
$h_D = 10^{-4}$, $h_I = 10^{-6}$, $n_S = 100$	32 163.36

Table 2: Time spent for computing an approximate solution of Example 3 using the Hybrid approach, with different time steps: h_D for the first guess, and final solution with h_I, using n_S intervals of shooting.

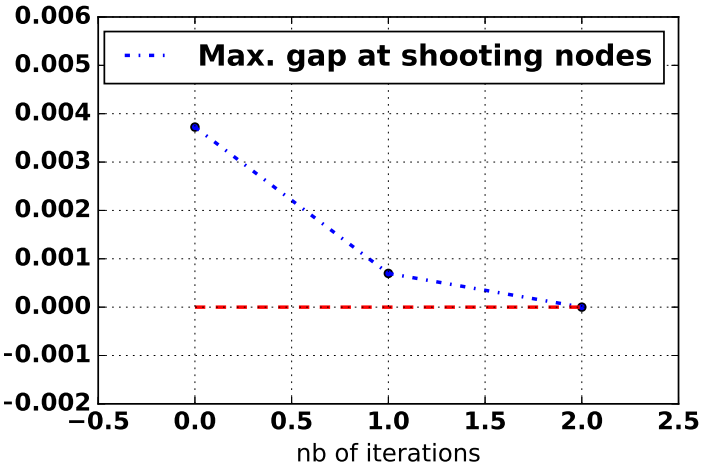


Figure 16: Maximum gaps on (x, p) at each iteration of the Newton Method used for the Indirect Method for computation of Example 3. $h_I = 10^{-5}$.

[5] J. Andersson. *A General-Purpose Software Framework for Dynamic Optimization*. PhD thesis, Arenberg Doctoral School, KU Leuven, Department of Electrical Engineering (ESAT/SCD) and Optimization in Engineering Center, Kasteelpark Arenberg 10, 3001-Heverlee, Belgium, October 2013.

[6] C. Batlle, E. Fossas, I. Merillas, and A. Miralles. Generalized discontinuous conduction modes in the complementarity formalism. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 52(8):447–451, 2005.

[7] B. Brogliato. Some perspectives on the analysis and control of complementarity systems. *IEEE Transactions on Automatic Control*, 48(6):918–935, 2003.

[8] B. Brogliato. Some results on the controllability of planar evolution variational inequalities. *Systems and Control Letters*, 54(1):65–71, 2005.

[9] B. Brogliato, A. Daniilidis, C. Lemaréchal, and V. Acary. On the equivalence between complementarity systems, projected systems and differential inclusions. *Systems and Control Letters*, 55:45–51, 2006.

[10] B. Brogliato and D. Goeleven. Well-posedness, stability and invariance results for a class of multivalued Lur'e dynamical systems. *Nonlinear Analysis: Theory, Methods and Applications*, 74:195–212, 2011.

[11] B. Brogliato and D. Goeleven. Existence, uniqueness of solutions and stability of nonsmooth multivalued Lur'e dynamical systems. *Journal of Convex Analysis*, 20(3):881–900, 2013.

[12] B. Brogliato and L. Thibault. Existence and uniqueness of solutions for non-autonomous complementarity dynamical systems. *Journal of Convex Analysis*, 17(3):961–990, 2010.

[13] T. H. Cao and B. S. Mordukhovich. Optimal control of a perturbed sweeping process via discrete approximations. *Disc. Cont. Dyn. Syst. Ser. B*, 21(10):3331–3358, 2015.

[14] M K. Çamlibel, J.-S. Pang, and J. Shen. Lyapunov stability of complementarity and extended systems. *SIAM Journal on Optimization*, 17(4):1056–1101, 2006.

[15] M.K. Çamlibel. Popov–Belevitch–Hautus type controllability tests for linear complementarity systems. *Systems & Control Letters*, 56(5):381–387, 2007.

[16] M.K. Çamlibel, W.P.M.H. Heemels, and J.M. Schumacher. Consistency of a time-stepping method for a class of piece-wise networks. *IEEE Transactions on Circuits and Systems I*, 49:349–357, 2002.

[17] M.K. Çamlibel, W.P.M.H. Heemels, and J.M. Schumacher. On linear passive complementarity systems. *European Journal of Control*, 8(3):220 – 237, 2002.

[18] M.K. Çamlibel, W.P.M.H Heemels, A.J. van der Schaft, and J.M. Schumacher. Switched networks and complementarity. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 50(8):1036–1046, 2003.

[19] M.K. Çamlibel and L. Iannelli. Passivity and complementarity. *Mathematical Programming A*, 145(1-2):531–563, 2014.

[20] L. Cesari. *Optimization-Theory and Applications: Problems with Ordinary Differential Equations*, volume 17. Springer Science & Business Media, 2012.

[21] G. Colombo, R. Henrion, N. D. Hoang, and B. S. Mordukhovich. Optimal control of the sweeping process. *Dyn. Contin. Discrete Impuls. Syst.-Ser. B*, 19:117–159, 2012.

[22] G. Colombo, R. Henrion, D. H. Nguyen, and B. S. Mordukhovich. Optimal control of the sweeping process over polyhedral controlled sets. *Journal of Differential Equations*, 260(4):3397–3447, 2016.

[23] R. Cottle, J.-S. Pang, and R. Stone. *The linear complementarity problem*. SIAM, 2009.

[24] M. d. R. de Pinho. Mixed constrained control problems. *Journal of mathematical analysis and applications*, 278(2):293–307, 2003.

[25] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.

[26] C. Georgescu, B. Brogliato, and V. Acary. Switching, relay and complementarity systems: A tutorial on their well-posedness and relationships. *Physica D: Nonlinear Phenomena*, 241(22):1985 – 2002, 2012.

[27] D. Goeleven. Existence and uniqueness for a linear mixed variational inequality arising in electrical circuits with transistors. *journal of Optimization Theory and Applications*, 138(3):397–406, 2008.

[28] D. Goeleven and B. Brogliato. Stability and instability matrices for linear evolution variational inequalities. *IEEE Transactions on Automatic Control*, 49(4):521–534, 2004.

[29] L. Guo, G.-H. Lin, and J. J. Ye. Solving mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 166(1):234–256, 2015.

[30] L. Guo and J. J. Ye. Necessary optimality conditions for optimal control problems with equilibrium constraints. *SIAM Journal on Control and Optimization*, 54(5):2710–2733, 2016.

[31] L. Han, A. Tiwari, M.K. Çamlibel, and J.-S. Pang. Convergence of time-stepping schemes for passive and extended linear complementarity systems. *SIAM Journal on Numerical Analysis*, 47(5):3768–3796, 2009.

[32] W.P.M.H. Heemels, M.K. Çamlibel, J. Schumacher, and B. Brogliato. Observer-based control of linear complementarity systems. *International Journal of Robust and Nonlinear Control*, 21(10):1193–1218, 2011.

[33] W.P.M.H. Heemels, J.M. Schumacher, and S. Weiland. Linear complementarity systems. *SIAM Journal of Applied Mathematics*, 60(4):1234–1269, 2000.

[34] L. Iannelli, F. Vasca, and G. Angelone. Computation of steady-state oscillations in power converters through complementarity. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58(6):1421–1432, 2011.

[35] C. Kanzow and A. Schwartz. Mathematical programs with equilibrium constraints: enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results. *SIAM Journal on Optimization*, 20(5):2730–2753, 2010.

[36] C. Kanzow and A. Schwartz. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM Journal on Optimization*, 23(2):770–798, 2013.

[37] D. Leenaerts. On linear dynamic complementarity systems. *IEEE Transactions on Circuits and Systems I*, 46(8):1022–1026, 1999.

[38] S. Leyffer, G. López-Calva, and J. Nocedal. Interior methods for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 17(1):52–77, 2006.

[39] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[40] J-S. Pang and D. E. Stewart. Solution dependence on initial conditions in differential variational inequalities. *Mathematical Programming*, 116(1):429–460, 2009.

[41] B. Passenberg, P.E. Caines, M. Leibold, O. Stursberg, and M. Buss. Optimal control for hybrid systems with partitioned state space. *IEEE Transactions on Automatic Control*, 58(8):2131–2136, 2013.

[42] R. Rockafellar and R.J-B Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.

[43] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[44] V. Sessa, L. Iannelli, and F. Vasca. A complementarity model for closed-loop power converters. *IEEE Transactions on Power Electronics*, 29(12):6821–6835, 2014.

[45] V. Sessa, L. Iannelli, F. Vasca, and V. Acary. A complementarity approach for the computation of periodic oscillations in piecewise linear systems. *Nonlinear Dynamics*, 85(2):1255–1273, 2016.

[46] M.S. Shaikh and P.E. Caines. On the hybrid optimal control problem: theory and algorithms. *IEEE Transactions on Automatic Control*, 52(9):1587–1603, 2007.

[47] J. Shen. Robust non-zenoness of piecewise affine systems with applications to linear complementarity systems. *SIAM Journal on Optimization*, 24(4):2023–2056, 2014.

[48] J. Shen and J.-S. Pang. Linear complementarity systems: Zeno states. *SIAM Journal on Control and Optimization*, 44(3):1040–1066, 2005.

[49] H. J. Sussmann. A maximum principle for hybrid optimal control problems. volume 1, pages 425–430. IEEE, 1999.

[50] A. Tanwani, B. Brogliato, and C. Prieur. Stability and observer design for Lur'e systems with multivalued, non-monotone, time-varying nonlinearities and state jumps. *SIAM Journal on Control and Optimization*, 56(2):3639–3672, 2014.

[51] A. Tanwani, B. Brogliato, and C. Prieur. Well-posedness and output regulation for implicit time-varying evolution variational inequalities. *SIAM Journal on Control and Optimization*, 2018.

[52] E. Trélat. Optimal control and applications to aerospace: some results and challenges. *Journal of Optimization Theory and Applications*, 154(3):713–758, 2012.

[53] F. Vasca, L. Iannelli, M.K. Çamlibel, and R. Frasca. A new perspective for modelling power electronics converters: complementarity framework. *IEEE Transactions on Power Electronics*, 24(2):456–468, 2009.

[54] J.J. Ye. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 307(1):350–369, 2005.

A Results on stationarity for Optimal Control with Complementarity Conditions

A.1 Stationarity results

We recall here some results stated in [30] concerning the following optimal control problem:

$$\begin{array}{ll} \min\ J(x,w) = \displaystyle\int_{t_0}^{t_1} F(t,x(t),\tilde{u}(t))dt & \\ \text{s.t.} \quad \left.\begin{array}{l} \dot{x}(t) = \phi(t,x(t),\tilde{u}(t)) \\ 0 \le G(t,x(t),\tilde{u}(t)) \perp H(t,x(t),\tilde{u}(t)) \ge 0 \end{array}\right\} & \text{a.e. } t \in [t_0,t_1] \\ (x(t_0),x(t_1)) \in \mathscr{E}. & \end{array} \tag{32}$$

with $F : [t_0,t_1]\times\mathbb{R}^n\times\mathbb{R}^m \to \mathbb{R}$, $G,H : [t_0,t_1]\times\mathbb{R}^n\times\mathbb{R}^m \to \mathbb{R}^l$, $\mathscr{E}$ a closed convex subset of $\mathbb{R}^{2n}$. In our settings, $\tilde{u} = \binom{u}{v}$, $F(t,x,\tilde{u}) = x^\intercal Q x + u^\intercal U u$, $\phi(t,x,\tilde{u}) = Ax + Bv + Fu$, $G(t,x,\tilde{u}) = v$, $H(t,x,\tilde{u}) = w = Cx + Dv + Eu$, $\mathscr{E} = \{(x_0,x_T) : Mx_0 + Nx_T = x_b\}$. The article [30] actually tackles a broader problem (with inequality and equality constraints, and possible non-smoothness). We present only the material useful for our purpose.

A.1.1 Definitions related to non-smooth optimal control

We suppose that F and ϕ are $\mathcal{L}\times\mathcal{B}$-measurable, where $\mathcal{L}\times\mathcal{B}$ denotes the σ-algebra of subsets of appropriate spaces generated by product sets $M\times N$, where M is a Lebesgue ($\mathcal{L}$) measurable subset in $\mathbb{R}$, and N is a Borel ($\mathcal{B}$) measurable subset in $\mathbb{R}^n\times\mathbb{R}^m$.

Definition 5. *We refer to any absolutely continuous function as an arc. An admissible pair for (32) is a pair of functions (x,w) on $[t_0,t_1]$ for which w is a control and x is an arc, that satisfy all the constraints in (32). We define the set constraint at time $t\in[t_0,t_1]$, $S(t)$, by:*

$$S(t) = \left\{(x,\tilde{u}) \in \mathbb{R}^n\times\mathbb{R}^m : (G(t,x,\tilde{u}),H(t,x,\tilde{u})) \in \mathcal{C}^l\right\}.$$

For every given $t\in[t_0,t_1]$ and two positive constants R and ε, we define a neighbourhood of the point $(x^*(t),w^*(t))$ as:

$$S_*^{\varepsilon,R}(t) = \{(x,w)\in S(t) : \|x - x^*(t)\| \le \varepsilon,\ \|w - w^*(t)\| \le R\}. \tag{33}$$

Assumption 1. *1. There exist measurable functions k_x^ϕ, k_x^F, k_w^ϕ, k_w^F such that for almost every $t\in[t_0,t_1]$ and for every (x^1,w^1), $(x^2,w^2)\in S_*^{\varepsilon,R}(t)$, we have:*

$$\begin{array}{l} (a)\|\phi(t,x^1,w^1) - \phi(t,x^2,w^2)\| \le k_x^\phi(t)\|x^1 - x^2\| + k_w^\phi(t)\|w^1 - w^2\| \\ (b)|F(t,x^1,w^1) - F(t,x^2,w^2)| \le k_x^F(t)\|x^1 - x^2\| + k_w^F(t)\|w^1 - w^2\|. \end{array} \tag{34}$$

2. There exists a positive measurable function k_S such that for almost every $t\in[t_0,t_1]$, the bounded slope condition holds:

$$(x,w)\in S_*^{\varepsilon,R}(t), (\alpha,\beta)\in\mathcal{N}_{S(t)}^P(x,w) \implies \|\alpha\| \le k_S(t)\|\beta\|. \tag{35}$$

3. The functions k_x^ϕ, k_x^F and $k_S[k_w^\phi + k_w^F]$ are integrable, and there exists a positive number η such that $R(t)\ge \eta k_S(t)$ a.e. $t\in[t_0,t_1]$.

4. F and ϕ are $\mathcal{L}\times\mathcal{B}$-measurable, G and H are $\mathcal{L}$-measurable in variable t and strictly differentiable in variable (x,w), f is locally Lipschitz continuous, and $\mathscr{E}$ is a closed subset in $\mathbb{R}^{2n}$.

Define the sets $I_t^{+0}(x,\tilde{u}) = \{i : G_i(t,x,\tilde{u}) > 0 = H_i(t,x,\tilde{u})\}$, $I_t^{0+}(x,\tilde{u}) = \{i : G_i(t,x,\tilde{u}) = 0 < H_i(t,x,\tilde{u})\}$, $I_t^{00}(x,\tilde{u}) = \{i : G_i(t,x,\tilde{u}) = 0 = H_i(t,x,\tilde{u})\}$, and for any $(\mu,\nu)\in\mathbb{R}^l\times\mathbb{R}^l$, denote:

$$\Psi(t,x,\tilde{u};\mu,\nu) = -G(t,x,\tilde{u})^\intercal\mu - H(t,x,\tilde{u})^\intercal\nu. \tag{36}$$

Theorem 4. *[30] Let $(x^*,\tilde{u}^*)$ be a local minimizer of radius R for (32) and let Assumption 1 hold. If for almost every $t\in[t_0,t_1]$ the local error bound condition for the system representing $S(t)$ holds at $(x^*(t),\tilde{u}^*(t))$, then there exist a number $\lambda_0\le 0$, an arc p and measurable functions $\lambda^G : \mathbb{R}\to\mathbb{R}^l$, $\lambda^H : \mathbb{R}\to\mathbb{R}^l$ such that the following conditions hold:*

1. *the non-triviality condition:* $(\lambda_0, p(t)) \neq 0,\ \forall t \in [t_0, t_1]$

2. *the transversality condition:*
$$(p(t_0), -p(t_1)) \in \mathcal{N}_{\mathscr{E}}(x^*(t_0), x(t_1)) \tag{37}$$

3. *the Euler adjoint inclusion: for almost every* $t \in [t_0, t_1]$,
$$\begin{aligned}(\dot{p}(t), 0) \in \partial^C \left\{\langle -p(t), \phi(t, \cdot, \cdot)\rangle - \lambda_0 F(t, \cdot, \cdot)\right\}(x^*(t), \tilde{u}^*(t)) \\ + \nabla_{x,\tilde{u}} \Psi(t, x^*(t), \tilde{u}^*(t), \lambda^G(t), \lambda^H(t))\end{aligned} \tag{38}$$
$$\lambda_i^G(t) = 0,\ \forall i \in I_t^{+0}(x^*(t), \tilde{u}^*(t)),\ \lambda_i^H(t) = 0,\ \forall i \in I_t^{0+}(x^*(t), \tilde{u}^*(t))$$

4. *the Weierstrass condition for radius* R*: for almost every* $t \in [t_0, t_1]$,
$$(x^*(t), \tilde{u}) \in S(t),\ \|\tilde{u} - \tilde{u}^*(t)\| < R(t)$$
$$\implies \langle p(t), \phi(t, x^*(t), \tilde{u})\rangle + \lambda_0 F(t, x^*(t), \tilde{u}) \leq \langle p(t), \phi(t, x^*(t), \tilde{u}^*(t))\rangle + \lambda_0 F(t, x^*(t), \tilde{u}^*(t))$$

The Weierstrass condition can be re-expressed as searching a local minimizer of the following MPEC:
$$\begin{aligned}\max\ & \langle p(t), \phi(t, x^*(t), \tilde{u})\rangle + \lambda_0 F(t, x^*(t), \tilde{u}) \\ \text{s.t.}\ & (G(t, x^*(t), \tilde{u}), H(t, x^*(t), \tilde{u})) \in \mathcal{C}^l\end{aligned} \tag{39}$$

where $\mathcal{C}^l$ is defined in Definition 1. For each $t \in [t_0, t_1]$, this is an MPEC which admits stationarity conditions as exposed in Section B.

Definition 6. *Let* $(x^*, \tilde{u}^*)$ *be an admissible pair for* (32).

- *The Fritz-John (FJ) type W(eak)-stationarity holds at* (x^*, w^*) *if there exist a number* $\lambda_0 \leq 0$*, an arc* p *and measurable functions* λ^G*,* λ^H *such that Theorem 4 (1)-(4) hold.*
- *The FJ-type S(trong)-stationarity holds at* $(x^*, \tilde{u}^*)$ *if* $(x^*, \tilde{u}^*)$ *is W-stationarity with arc* p *and there exist measurable functions* η^G*,* η^H *such that, for almost every* $t \in [t_0, t_1]$,
$$\begin{aligned}0 \in \partial^C \left\{\langle -p(t), \phi(t, x^*(t), \cdot) - \lambda_0 F(t, x^*(t), \cdot)\rangle\right\}(\tilde{u}^*(t)) \\ + \nabla_{\tilde{u}} \Psi(t, x^*(t), \tilde{u}^*(t), \eta^G(t), \eta^H(t))\end{aligned}$$
$$\begin{aligned}\eta_i^G(t) = 0,\ \forall i \in I_t^{+0}(x^*(t), \tilde{u}^*(t)),\ \eta_i^H(t) = 0,\ \forall i \in I_t^{0+}(x^*(t), \tilde{u}^*(t)) \\ \eta_i^G(t) \geq 0,\ \eta_i^H(t) \geq 0,\ \forall i \in I_t^{00}(x^*(t), \tilde{u}^*(t)).\end{aligned}$$

We refer to the FJ-type W-, and S-stationarities as the W-, and S-stationarities, respectively, if $\lambda_0 = -1$.

However, as shown in [30], these new multipliers η^G, η^H can be different in measure from the corresponding λ^G, λ^H. These instabilities were coped using the following proposition:

Theorem 5. *[30] Let* $(x^*, \tilde{u}^*)$ *be a local minimizer of radius* R *for* (32)*, and let Assumption 1 hold. If for almost every* $t \in [t_0, t_1]$*, the functions* $F(t, \cdot, \cdot)$ *and* $\phi(t, \cdot, \cdot)$ *are strictly differentiable at* $(x^*(t), \tilde{u}^*(t))$*, and the MPEC LICQ holds at* $\tilde{u}^*(t)$ *for problem* (39)*, i.e., the family of gradients*
$$\{\nabla_u G_i(t, x^*(t), \tilde{u}^*(t)) : i \in I_t^{0\bullet}(x^*(t), \tilde{u}^*(t))\} \cup \{\nabla_u H_i(t, x^*(t), \tilde{u}^*(t)) : i \in I_t^{\bullet 0}(x^*(t), \tilde{u}^*(t))\}$$
is linearly independent, where
$$I_t^{0\bullet}(x^*(t), \tilde{u}^*(t)) = I_t^{0+}(x^*(t), \tilde{u}^*(t)) \cup I_t^{00}(x^*(t), \tilde{u}^*(t)),$$
$$I_t^{\bullet 0}(x^*(t), \tilde{u}^*(t)) = I_t^{+0}(x^*(t), \tilde{u}^*(t)) \cup I_t^{00}(x^*(t), \tilde{u}^*(t)),$$
then the FJ-type S-stationarity holds at $(x^*, \tilde{u}^*)$*. Moreover, the multipliers* η^G*,* η^H *can be taken as equal to* λ^G*,* λ^H*, respectively, almost everywhere.*

A.1.2 Sufficient condition for the Bounded Slope Condition (35)

Let us define the following set:

$$C_*^{\varepsilon,R} = cl\{(t,x,\tilde{u}) \in [t_0,t_1] \times \mathbb{R}^n \times \mathbb{R}^m : (x,\tilde{u}) \in S_*^{\varepsilon,R}(t)\}$$

Proposition 8. *[30] Let the mappings G, H be autonomous. Assume that $C_*^{\varepsilon,R}$ is compact for some $\varepsilon > 0$, the local error bound holds, and that, for every $(x,\tilde{u})$ such that $(t,x,\tilde{u}) \in C^{\varepsilon,R}$, the system complies with the following implication:*

$$\left.\begin{array}{r} 0 = \nabla_w \Psi(t,x,w,\mu,\nu) \\ \mu_i = 0,\ \forall i \in I_t^{+0}(x,\tilde{u}),\ \nu_i = 0,\ \forall i \in I_t^{0+}(x,\tilde{u}), \\ \mu_i > 0,,\ \nu_i > 0,\ \text{or } \mu_i\nu_i = 0,\ \forall i \in I_t^{00}(x,\tilde{u}) \end{array}\right\} \implies \nabla_x \Psi(t,x,\tilde{u},\mu,\nu) = 0,$$

where Ψ is defined in (36). Then there exists a positive constant k_S such that for every $t \in [t_0,t_1]$, the bounded slope condition (35) holds with $k_S(t) = k_S$.

B Finite-dimensional MPEC

A first way to solve numerically the problem (1)(2) is by discretizing directly the problem. This, as it was made clear in Section 3, leads to a problem called MPEC which is finite-dimensional. We present here some results linked with this problem.

B.1 Definitions and properties

Let us give some definitions and results on the MPEC in (3). Recall the sets $I^{+0}(z) = \{i : G_i(z) > 0 = H_i(z)\}$, $I^{0+}(z) = \{i : G_i(z) = 0 < H_i(z)\}$, $I^{00}(z) = \{i : G_i(z) = 0 = H_i(z)\}$.

Definition 7. *1. The W-stationarity holds at z^* if there exist (θ,ν) such that*

$$\nabla f(z^*) - \nabla G(z^*)\theta - \nabla H(z^*)\nu = 0 \ \text{ and } \ \theta_{I^{+0}(z^*)} = 0,\ \nu_{I^{0+}(z^*)} = 0 \tag{40}$$

2. The M-stationarity holds at z^* if it is W-stationary and furthermore

$$\text{either } \theta_i\nu_i = 0 \text{ or } \theta_i > 0, \nu_i > 0,\ \forall i \in I^{00}(z^*). \tag{41}$$

3. The S-stationarity holds at z^* if it is W-stationary and furthermore

$$\theta_i \geq 0,\ \nu_i \geq 0,\ \forall i \in I^{00}(z^*). \tag{42}$$

The next Proposition, that re-expresses S-stationarity conditions, is used in the proof of Lemma 2.

Proposition 9. *[29] Conditions (40)-(42) hold if and only if there exist $\alpha,\beta \in \mathbb{R}^m$, $\zeta \in \mathbb{R}$ such that:*

$$\begin{array}{c} \alpha_i G_i(z^*) = \beta_i H_i(z^*) = 0,\ \alpha_i \geq 0,\ \beta_i \geq 0, \\ \theta_i = \alpha_i - \zeta H_i(z^*),\ \nu_i = \beta_i - \zeta G_i(z^*),\ (i = 1,...,m) \end{array} \tag{43}$$

Remark. *The first line of equation (43) can be re-expressed as the complementarity conditions:*

$$0 \leq \alpha \perp G_i(z^*) \geq 0, \qquad 0 \leq \beta \perp H_i(z^*) \geq 0$$

Also, we can see that this decomposition of θ and ν is not unique. If we denote by (α,β,ζ) a decomposition of (θ,ν) as written before, we see that $(\alpha+\rho H(z^*), \beta+\rho G_i(z^*), \zeta+\rho)$ is also a decomposition of (θ,ν) for all strictly positive scalars ρ. It then becomes clear that, for a fixed ζ big enough, the decomposition exists and is then unique.

B.2 Numerical treatments of MPEC

In general, the numerical treatment of MPEC is not an easy task. We present here two schemes, converging to M-stationary points.

B.2.1 Complementarity relaxation

The idea is to relax the complementarity condition using the following NCP function: define $\varphi : \mathbb{R} \to \mathbb{R}$ by:

$$\varphi(a,b) = \begin{cases} ab & \text{if } a+b \geq 0 \\ -\frac{1}{2}(a^2+b^2) & \text{if } a+b < 0 \end{cases}$$

and then relax the problem as: let define $\Phi : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^q$ be defined component-wise by:

$$\Phi_i(z,\tau) = \varphi(G_i(z) - \tau, H_i(z) - \tau)$$

With this function, we define the relaxed problem NLP(τ) for $\tau \geq 0$ as:

$$\begin{aligned} \min\ & f(z) \\ \text{s.t. } & g(z) \leq 0,\ h(z) = 0 \\ & G(z) \geq 0,\ H(z) \geq 0 \\ & \Phi(z,\tau) \leq 0 \end{aligned} \tag{NLP(τ)}$$

Algorithm 1: Relaxation algorithm $(z^0, \tau_0, \sigma, \tau_{min}, \varepsilon)$.

Input: A starting vector z^0, an initial relaxation parameter τ_0, and parameters $\sigma \in (0,1[$, $\tau_{min} > 0$, and $\varepsilon > 0$

1 Set k:=0

2 **while** *($\tau_k \geq \tau_{min}$ and compVio(z^k) $> \varepsilon$) or k=0* **do**

3 Find an approximate solution z^{k+1} of NLP(τ_k). To solve NLP(τ_k), use z^k as starting vector. If NLP(τ_k) is not feasible, terminate the algorithm

4 Let $\tau_{k+1} \leftarrow \sigma \min\{t_k, \text{compVio}(z^{k+1})\}$ and $k \leftarrow k+1$

5 **end**

Output: The final iterate $z_{opt} = z^k$, the corresponding function value $f(z_{opt})$, and the maximum constraint violation $maxVio(z_{opt})$.

In this algorithm, we denote:

$$\text{compVio}(z) = \max\{\min\{G_i(z), H_i(z)\}, i = 1, \ldots, l\}$$

$$\text{maxVio}(z) = \max\{\max\{0, g_j(z)\}, |h_k(z)|, |\min\{G_i(z), H_i(z)\}|, j = 1, \ldots, l_1,\ k = 1, \ldots, l_2,\ i = 1, \ldots, l\}$$

Theorem 6. *[36] Let $\{\tau_k\} \downarrow 0$ and $\{(z^k, \lambda^k, \mu^k, \gamma^k, \nu^k, \delta^k\}$ be a sequence of KKT points of NLP(τ_k) with $z^k \to z^*$. If MPEC LICQ holds in z^*, then z^* is an M-stationary point of the MPEC (3).*
Furthermore, if there is a subsequence $K \subseteq \mathbb{N}$ such that

$$G_i(z^k) \leq \tau_k,\ H_i(z^k) \leq \tau_k, \forall k \in K,\ \forall i \in I_{00}(z^*)$$

then z^* is a S-stationary point of (3).

However, it should be noted that this convergence is actually sensitive to instabilities.

Definition 8. *Let $\varepsilon > 0$. We say that z^* is an ε-stationary point of the problem*

$$\min f(z)\ \text{ s.t. } g(z) \leq 0,\ h(z) = 0$$

if there are multipliers λ and μ such that:

$$\|\nabla f(z^*) + (\nabla g(z^*))^\intercal \lambda + (\nabla h(z^*))^\intercal \mu\|_\infty \leq \varepsilon$$

$$g(z^*) \leq 0,\ \lambda \geq 0,\ \lambda_i g_i(z^*) \geq -\varepsilon,\ \forall i$$

$$|h_i(z^*)| \leq \varepsilon,\ \forall i$$

Theorem 7. *[36] Let $\{\tau_k\} \downarrow 0$, $\varepsilon_k = o(\tau_k)$, and z^k be a sequence of ε_k-stationary points of NLP(τ_k) with multipliers $(\lambda^k, \mu^k, \gamma^k, \nu^k, \delta^k)$. Assume that $z^k \to z^*$. If MPEC LICQ holds at z^*, then z^* is a W-stationary point of MPEC (3).*

B.2.2 Cost penalization

The technique used here is the penalization of the objective function. The complementarity is moved to the objective function in the form of an ℓ_1-penalty term, so that the objective becomes:

$$f(z) + \pi G(z)^\intercal H(z)$$

The associated barrier problem is defined as:

$$\begin{aligned} \min\ & f(z) + \pi G(z)^\intercal H(z) - \mu \left(\sum_i \log s_i + \sum_i \log G_i(z) + \sum_i \log H_i(z) \right) \\ \text{s.t. } & h(z) = 0 \\ & g(z) - s = 0 \end{aligned} \tag{44}$$

The Lagrangian of this barrier problem is given by:

$$\begin{aligned} \mathcal{L}_{\mu,\pi}(z, s, \lambda, \theta) = f(z) + \pi G(z)^\intercal H(z) - \mu \left(\sum_i \log s_i + \sum_i \log G_i(z) + \sum_i \log H_i(z) \right) \\ - \sum_i \theta_i h_i(z) - \sum_i \lambda_i (g_i(z) - s_i) \end{aligned}$$

Algorithm 2: Classic: A Practical Interior-Penalty Method for MPECs.

Input: Let $z^0, s^0, \lambda^0, \theta^0$ be the initial value of the primal and dual variables.

1 Set $k = 1$.

2 **repeat**

3 Choose a barrier parameter μ^k, a stopping tolerance ε^k_{pen} and ε^k_{comp}

4 Find π^k and an approximate solution $(x^k, s^k, \lambda^k, \theta^k)$ of problem (44) with parameter μ^k and π^k that satisfy $G(z^k) > 0$, $H(z^k) > 0$, $s^k > 0$, $\lambda^k > 0$ and the following conditions:

$$\|\nabla_z \mathcal{L}_{\mu^k,\pi^k}(z^k, s^k, \lambda^k, \theta^k)\| \leq \varepsilon^k_{pen}$$

$$\|s^k_i \lambda^k_i - \mu^k\| \leq \varepsilon^k_{pen},\ \forall i$$

$$\left\| \begin{matrix} h(z^k) \\ g(z^k) - s^k \end{matrix} \right\| \leq \varepsilon^k_{pen}$$

$$\|\min\{G(z^k), H(z^k)\}\| \leq \varepsilon^k_{comp}$$

5 Let $k \leftarrow k + 1$

6 **until** *a stopping test for the MPEC is satisfied*

Theorem 8. *[38] Suppose that Algorithm 2 generates an infinite sequence of iterates $\{z^k, s^k, \lambda^k, \theta^k\}$ and parameters $\{\pi^k, \mu^k\}$, for sequences $\{\varepsilon^k_{pen}\}$, $\{\varepsilon^k_{comp}\}$, $\{\mu^k\}$ converging to zero. If z^* is a limit point of the sequence $\{z^k\}$, and f, g and h are continuously differentiable in an open neighborhood $\mathcal{N}(z^*)$ of z^*, then z^* is feasible for (3). If in addition, MPEC LICQ holds at z^*, then z^* is C-stationary. Moreover, if $\{\pi^k\}$ is bounded, then z^* is a strongly stationary point of (3).*

C Analytical solution of Example 1

We give the details of the analytical solution found for the optimal control problem in (21).

C.1 Complete controllability

In order to compute an optimal control, the system should obviously be controllable between the initial and final states. For Example 1, we will only focus on completely controllable systems, relying on [15]. We recall here this result:

Theorem 9. *[15] Assume (2) satisfies the following conditions:*

1. *The matrix D is a P-matrix ; e.g., all its principal minors are positive.*
2. *The transfer matrix $E + C(sI - A)^{-1}F$ is invertible as a rational matrix.*

Then, the LCS in (2) is completely controllable if, and only if, the following two conditions hold:

1. *The pair $(A, [F\ B])$ is controllable.*
2. *The system of inequalities*

$$\eta \geq 0 \tag{45a}$$

$$\begin{pmatrix} \zeta^\intercal & \eta^\intercal \end{pmatrix} \begin{pmatrix} A - \lambda I & F \\ C & E \end{pmatrix} = 0 \tag{45b}$$

$$\begin{pmatrix} \zeta^\intercal & \eta^\intercal \end{pmatrix} \begin{pmatrix} B \\ D \end{pmatrix} \leq 0 \tag{45c}$$

admits no solution $\lambda \in \mathbb{R}$ and $0 \neq (\zeta, \eta) \in \mathbb{R}^{n+m}$.

The complete controllability conditions for the 1D system (21) thus boils down to check that the following system:

$$\begin{aligned} (a - \lambda)\zeta + c\eta &= 0, & (46)\\ f\zeta + e\eta &= 0, & (47)\\ \eta &\geq 0, & (48)\\ b\zeta + d\eta &\leq 0, & (49) \end{aligned}$$

has no solution $\lambda \in \mathbb{R}$ and $(\zeta, \eta) \neq 0$

If $e > 0$: we deduce through (47): $\eta = -\frac{f\zeta}{e}$.

1. If $f = 0$, then $\eta = 0$. In (46), we can take $\lambda = a$. However, with (49), we have that $\zeta b \leq 0$. Let us take $\zeta = -\text{sign}(b)$. Then we found a solution with $\zeta \neq 0$: the system is not completely controllable.
2. If $f < 0$, then with (48), we have that $\zeta \geq 0$. Through (46), we take $\lambda = a + \frac{cf}{e}$.
   - If $b \geq 0$, then (49) is a sum of positive terms which must be nonpositive, so $\eta = 0$ and $\zeta = 0$: the system is completely controllable.
   - If $b < 0$, then (49) becomes $\zeta(b - \frac{fd}{e}) \leq 0$ with $\zeta \geq 0$.
     - If $b - \frac{fd}{e} \leq 0$ then we can take any $\zeta \geq 0$: the system is not completely controllable.
     - Otherwise, only $\zeta = 0$ suits, so $\eta = 0$, and then the system is completely controllable.
3. If $f > 0$, then in (46), we take $\lambda = a + \frac{cf}{e}$. Through (48), we have that $\zeta \leq 0$.
   - If $b \leq 0$, then (49) is a positive terms sum which must be nonpositive, so $\eta = 0$ and $\zeta = 0$: the system is completely controllable.
   - If $b > 0$, then (49) becomes $\zeta(b - \frac{fd}{e}) \leq 0$ with $\zeta \leq 0$.
     - If $b - \frac{fd}{e} \geq 0$ then then we can take any $\zeta \leq 0$: the system is not completely controllable.
     - Otherwise, only $\zeta = 0$ suits, so $\eta = 0$, and then the system is completely controllable.

If $e < 0$: we have the same cases as with $e > 0$ by inverting the sign of f.

C.2 Necessary first-order conditions

C.2.1 Adjoint equation

The dynamic system in (21) can be rewritten as $\dot{x} = ax + fu + \frac{b}{d}\Pi_{\mathbb{R}_+}(-eu)$, where $\Pi_{\mathbb{R}_+}$ is the orthogonal projection on $\mathbb{R}_+$. Therefore, the Hamiltonian function is written as: $H(x, p, u) = p\left(ax + fu + \frac{b}{d}\Pi_{\mathbb{R}_+}(-eu)\right) - \frac{1}{2}\left(x^2 + u^2\right)$. As stated in Theorem 3, if the trajectory is optimal, then there exist an absolutely continuous function $p : [0, T] \to \mathbb{R}$ such that it satisfies the differential equation $\dot{p}(t) = -ap(t) + x(t)$. Furthermore, the maximum condition on the

Hamiltonian holds: $H(x(t), p(t), u(t)) = \max_{v\in\mathbb{R}} H(x(t), p(t), v)$. Since $x(T)$ is free, there is a terminal condition on p: $p(T) = 0$. We can even differentiate twice p, and obtain the following second-order differential equation:

$$\begin{aligned}\ddot{p} &= -a\dot{p} + \dot{x} \\ &= a^2 p + fu + \frac{b}{d}\Pi_{\mathbb{R}_+}(-eu) \\ &= \begin{cases} a^2 p + fu & \text{if } eu \geq 0 \\ a^2 p + \left(f - \frac{be}{d}\right) u & \text{if } eu \leq 0. \end{cases}\end{aligned} \tag{50}$$

C.2.2 Maximization of the Hamiltonian function

We now search for an expression of the optimal control u^*, function of x and p, maximizing the Hamiltonian function $H(x, p, \cdot)$. To that aim, let us the Clarke subdifferential of H with respect to u, written $\partial_u^C H(x, p, u)$, and the fact that if u^* maximizes H, then $0 \in \partial_u^C H(x, p, u^*)$. In our problem, the subdifferential is written as

$$\partial_u^C H(x, p, u) = \begin{cases} \{fp - u\} & \text{if } eu > 0 \\ \left\{\left(f - \frac{eb}{d}\right) p - u\right\} & \text{if } eu < 0 \\ -\left[f - \frac{eb}{d}, f\right] p & \text{if } eu = 0. \end{cases}$$

Let us focus only on the completely controllable cases in order to find a control u maximizing this function:

If $e > 0$: In this case, $\text{sgn}(eu) = \text{sgn}(u)$.

1. We consider first $f < 0$.
   - If $b > 0$, then if $p \leq 0$, then $fp \geq 0$, $\left(f - \frac{eb}{d}\right) p \geq 0$, and if $p \geq 0$, then $fp \leq 0$, $\left(f - \frac{eb}{d}\right) p \leq 0$. We also notice that $0 \notin \left[f, f - \frac{eb}{d}\right]$. So we have:
   $$u^* = \begin{cases} fp & \text{if } p \leq 0, \\ \left(f - \frac{eb}{d}\right) p & \text{if } p \geq 0. \end{cases}$$
   - If $b < 0$, then we must make sure that $f < \frac{eb}{d}$. We notice that in this case, $0 \notin \left[f, f - \frac{eb}{d}\right]$. We are then in the exact same case as the previous one, and therefore, the control is expressed the same way:
   $$u^* = \begin{cases} fp & \text{if } p \leq 0, \\ \left(f - \frac{eb}{d}\right) p & \text{if } p \geq 0. \end{cases}$$
2. We consider now $f > 0$.
   - If $b < 0$, then if $p \leq 0$, then $fp \leq 0$, $\left(f - \frac{eb}{d}\right) p \leq 0$, and if $p \geq 0$, then $fp \geq 0$, $\left(f - \frac{eb}{d}\right) p \geq 0$. We also notice that $0 \notin \left[f, f - \frac{eb}{d}\right]$. So we have:
   $$u^* = \begin{cases} fp & \text{if } p \geq 0, \\ \left(f - \frac{eb}{d}\right) p & \text{if } p \leq 0. \end{cases}$$
   - If $b > 0$, then we must make sure that $f > \frac{eb}{d}$. We notice that in this case, $0 \notin \left[f, f - \frac{eb}{d}\right]$. We are then in the exact same case as the previous one, and therefore, the control is expressed the same way:
   $$u^* = \begin{cases} fp & \text{if } p \geq 0, \\ \left(f - \frac{eb}{d}\right) p & \text{if } p \leq 0. \end{cases}$$

If $e < 0$: we have the same cases as with $e > 0$ by inverting the sign of f.

Therefore, we can summarize this result as follows:

$$u^* = \begin{cases} fp & \text{if } efp \geq 0 \\ \left(f - \frac{eb}{d}\right) p & \text{if } efp \leq 0. \end{cases} \tag{51}$$

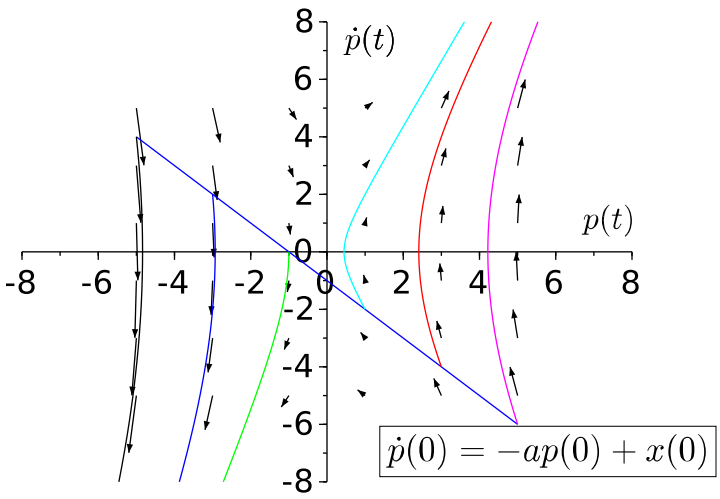


Figure 17: Phase portrait of (52) - $a = 1$, $b = -0.5$, $d = 1$, $e = -2$, $f = 3$, $x(0) = -1$,

C.2.3 Final adjoint equation

Finally, let us use the optimal control found in (51) in (50), which yields:

$$\ddot{p} = \begin{cases} \left(a^2 + f^2\right) p & \text{if } efp \geq 0 \\ \left(a^2 + \left(f - \frac{be}{d}\right)^2\right) p & \text{if } efp \leq 0, \end{cases}$$

which we rewrite in the simpler form:

$$\ddot{p} = \gamma(p)p, \tag{52}$$

with $\gamma(p) > 0$ and piecewise constant.

C.2.4 Initial conditions

It is now needed to find $p(0)$ such that $p(T) = 0$ (since $x(T)$ is free, according to the maximum principle). First, we know that the initial value for the derivative $\dot{p}$ is given by $\dot{p}(0) = x(0) - ap(0)$. The phase portrait is depicted in Figure 17. It is clear that, in order to have $p(T) = 0$, the sign of $p(0)$ is determined by the sign of the constants in the model:

- If $a > 0$, $x(0) > 0$, then $p(T) = 0 \implies p(0) < 0$,
- If $a > 0$, $x(0) < 0$, then $p(T) = 0 \implies p(0) > 0$,
- If $a < 0$, $x(0) > 0$, then $p(T) = 0 \implies p(0) < 0$,
- If $a < 0$, $x(0) < 0$, then $p(T) = 0 \implies p(0) > 0$.

We can summarize this by $sgn(p(0)) = -sgn(x(0))$. Moreover p will always have the same sign on $[0, T]$, so the optimal control u^* given in equation (51) always has the same sign on $[0, T]$, and is smooth (since p is smooth). Furthermore, γ will be constant on $[0, T]$. Consequently, the solution $p(t)$ on $[0, T]$ is known explicitly , namely:

$$p(t) = \frac{1}{\sqrt{\gamma}} \left[(\sqrt{\gamma} \cosh(\sqrt{\gamma}t) - a \sinh(\sqrt{\gamma}t)) p(0) + \sinh\left(\sqrt{\gamma}t\right) x(0) \right],$$

where cosh and sinh are the hyperbolic cosine and sine functions. In order to have $p(T) = 0$, we must take $p(0) = -\frac{\sinh\left(\sqrt{\gamma}T\right)x(0)}{\left(\sqrt{\gamma}\cosh(\sqrt{\gamma}T) - a\sinh(\sqrt{\gamma}T)\right)}$. From that, it is easy to obtain the expression of the optimal trajectory x, using the fact that $x(t) = \dot{p}(t) + ap(t)$.