

Modal occupation measures and LMI relaxations for nonlinear switched systems control

Mathieu Claeys, Jamal Daafouz, Didier Henrion

▶ **To cite this version:**

Mathieu Claeys, Jamal Daafouz, Didier Henrion. Modal occupation measures and LMI relaxations for nonlinear switched systems control. Automatica, 2016, 64, pp.143-154. 10.1016/j.automatica.2015.11.003 . hal-00980355v2

HAL Id: hal-00980355
https://hal.science/hal-00980355v2

Submitted on 15 Dec 2014

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

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

Modal occupation measures and LMI relaxations for nonlinear switched systems control *

Mathieu Claeys[1], Jamal Daafouz[2], Didier Henrion[3,4,5]

Draft of December 15, 2014

Abstract

This paper presents a linear programming approach for the optimal control of nonlinear switched systems where the control is the switching sequence. This is done by introducing modal occupation measures, which allow to relax the problem as a primal linear programming (LP) problem. Its dual linear program of Hamilton-Jacobi-Bellman inequalities is also characterized. The LPs are then solved numerically with a converging hierarchy of primal-dual moment-sum-of-squares (SOS) linear matrix inequalities (LMI). Because of the special structure of switched systems, we obtain a much more efficient method than could be achieved by applying standard moment/SOS LMI hierarchies for general optimal control problems.

1 Introduction

A switched system is a particular class of a hybrid system that consists of a set of dynamical subsystems, one of which is active at any instant of time, and a policy for activating and deactivating the subsystems. These dynamical systems are encountered in a wide variety of application domains such as automotive industry, power systems, aircraft and traffic control, and more generally the area of embedded systems. Switched systems have been the concern of much research, and many results are available for stability analysis and control design. They make evident the importance of orchestrating the subsystems

[1]Avenue Edmond Cordier 19, 1160 Auderghem, Belgium. mathieu.claeys@gmail.com

[2]Université de Lorraine, CRAN, CNRS, IUF, 2 avenue de la forêt de Haye, 54516 Vandœuvre cedex, France. jamal.daafouz@univ-lorraine.fr

[3]CNRS; LAAS; 7 avenue du colonel Roche, F-31400 Toulouse; France. henrion@laas.fr

[4]Université de Toulouse; LAAS; F-31400 Toulouse; France

[5]Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic

*This work with partly supported by project 13-06894S of the Grant Agency of the Czech Republic and by the UK Engineering and Physical Sciences Research Council under Grant EP/G066477/1.

through an adequate switching strategy, in order to impose global stability and/or performance. Interested readers may refer to the survey papers [17, 31, 44, 33] and the books [32, 46] and the references therein.

In this context, these systems are generally controlled by intrinsically discontinuous control signals, whose switching rule must be carefully designed to guarantee stability and performance properties. As far as optimality is concerned, several results are available in two main different contexts.

The first category of methods exploits necessary optimality conditions, in the form of Pontryagin's maximum principle (the so-called indirect approaches), or through a large nonlinear discretization of the problem (the so-called direct approaches). The first contributions can be found in [8, 22, 36] where the problem has been formulated and partial solutions provided through generalized Hamilton-Jacobi-Bellman (HJB) equations or inequalities and convex optimization. In [38, 47, 43], the maximum principle is generalized to the case of general hybrid systems with nonlinear dynamics. The case of switched systems is discussed in [4, 42, 2]. For general nonlinear problems and hence for switched systems, only *local optimality* can be guaranteed , even when discretization can be properly controlled [51]. The subject is still largely open and we are far from a complete and numericaly tractable solution to the switched optimal control problem.

The second category collects extensions of the performance indices H_2 and H_∞ originally developed for linear time invariant systems without switching, and use the flexibility of Lyapunov's approach, see for instance [20, 16] and references therein. Even for linear switched systems, the proposed results are based on nonconvex optimization problems (e.g., bilinear matrix inequality conditions) difficult to solve directly. Sufficient convex Linear Matrix Inequality (LMI) design conditions may be obtained, yielding computed solutions which are *suboptimal*, but at the price of introducing a conservatism (pessimism) and a gap to optimality which is hard, if not impossible, to evaluate. Since the computation of this optimal strategy is a difficult task, a suboptimal solution is of interest only when it is proved to be consistent, meaning that it imposes to the switched system a performance not worse than the one produced by each isolated subsystem [21].

Despite the interest of these existing approaches, the optimal control problem is not globally solved for switched systems, and new strategies are more than welcome, as computationally viable design techniques are missing. The present paper aims at complementing this rich literature on switched systems with convex programming techniques from the optimal control literature. Indeed, it is a well-known fact [49, 50] that optimal control problems can be relaxed as linear programming (LP) problems over measure spaces. Despite some early numerical results [40], this approach was merely of theoretic interest, before the advent of tractable primal-dual moment-sum-of-squares (SOS) *LMI hierarchies*, also now called Lasserre hierarchies, see [29] for an introduction to the subject. Most notably, [28] explores in depth the application of such a numerical method for general polynomial optimal control problems, where the control is bounded. By modeling control and trajectory functions as *occupation measures*, one obtains a convex LP formulation of the optimal control problem. This infinite-dimensional LP can be solved numerically and approximately with a hierarchy of convex finite-dimensional LMIs.

In this paper, we consider the problem of designing optimal switching rules in the case

of polynomial switched dynamical systems. Classically, we formulate the optimal control switching problem as an optimal control problem with artificial controls being functions of time, each valued in the discrete set $\{0, 1\}$, and we relax it into a control problem with such controls now valued in the interval $[0, 1]$. In contrast with existing approaches following this relaxation strategy, see, e.g., [4, 37], relying on the *local* Pontryagin maximum principle, our aim is to apply the *global* approach of [28]. We show that by specializing the convex objects for switched systems, one obtains a numerical method for which the number of modes driving the system enters *linearly* into the complexity of the problem, instead of exponentially. On the one hand, our approach follows the optimal control modeling framework. On the other hand, it exploits the flexibility and computational efficiency of the convex LMI framework. In contrast with most of the existing work on LMI methods, we have a *guarantee of global optimality*, in the sense that we obtain an asympotically converging (i.e., with vanishing conservatism) hierarchy of lower bounds on the achievable performance.

Organization of the paper

The paper is organized as follows. §2 constructs the generalized moment problem associated to the optimal control of switched systems, and details some conic duality results. Theses results are used in §3 to obtain an efficient numerical method based on the moment/SOS LMI hierarchy. §4 then outlines a simple method to extract candidate optimal trajectories from moment data. §5 gathers several useful extensions to the problem, kept separate from the main body of the article so as to ease exposition and notations. Finally, §6 illustrates the method with several examples.

Contributions

The contributions of this paper are as follows. On a theoretical level, a novel way of relaxing optimal control problems of switched systems is proposed, following [23]. The equivalence of this convex program with existing results in the literature [49, 50] is now rigorously proven. In addition, a full characterization of the dual conic problem as a system of relaxed HJB inequalities is given.

On a practical level, we present an easy to implement numerical method to solve the control problem, based on the LMI relaxations for optimal control [28]. The specialized measure LP formulation allows for substantial computational gains over generic formulations, which we now characterize. To our knowledge, this is one of the few results, along with [13] and [14], exploiting problem structure for more efficient LMI relaxations of control problems.

Preliminaries

Let $\mathcal{M}(X)$ denote the vector space of finite, signed, Borel measures supported on an Euclidean subset $X \subset \mathbb{R}^n$, equipped with the weak-$*$ topology, see, e.g., [39] for background

material. Let $\mathcal{M}^+(X)$ denote the cone of non-negative measures in $\mathcal{M}(X)$. The support of $\mu \in \mathcal{M}^+(X)$, the closed set of all points x such that $\mu(A) > 0$ for every open neighborhood A of x, is denoted by $\operatorname{spt}\mu$. For $\mu \in \mathcal{M}(X)$, $\operatorname{spt}\mu$ is the union of the supports of the Jordan decomposition of μ. For a continuous function $f \in \mathcal{C}(X)$, we denote by $\int_X f(x)\,\mu(\mathrm{d}x)$ the integral of f w.r.t. the measure $\mu \in \mathcal{M}(X)$. When no confusion may arise, we note $\langle f, \mu\rangle = \int f\mu$ for the integral to simplify exposition and to insist on the duality relationship between $\mathcal{C}(X)$ and $\mathcal{M}(X)$. The Dirac measure supported at x^*, denoted by δ_{x^*}, is the measure for which $\langle f, \delta_{x^*}\rangle = f(x^*)$ for all $f \in \mathcal{C}(X)$. The indicator function of set A, denoted by $I_A(x)$, is equal to one if $x \in A$, and zero otherwise. The space of probability measures on X, viz., the subset of $\mathcal{M}^+(X)$ with mass $\langle 1, \mu\rangle = 1$, is denoted by $\mathcal{P}(X)$.

For multi-index $\alpha \in \mathbb{N}^n$, $\mathbb{N}$ being the set of natural numbers, and vector $x \in \mathbb{R}^n$, we use the notation $x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$. We denote by $\mathbb{N}^n_m$ the set of vectors $\alpha \in \mathbb{N}^n$ such that $\sum_{i=1}^n \alpha_i \leq m$. The moment of multi-index $\alpha \in \mathbb{N}^n$ of measure $\mu \in \mathcal{M}^+(X)$ is then defined as the real $y_\alpha = \langle x^\alpha, \mu\rangle$.

A multi-indexed sequence of reals $(y_\alpha)_{\alpha\in\mathbb{N}^n}$ is said to have a representing measure on X if there exists $\mu \in \mathcal{M}^+(X)$ such that $y_\alpha = \langle x^\alpha, \mu\rangle$ for all $\alpha \in \mathbb{N}^n$.

Let $\mathbb{R}[x]$ denote the ring of polynomials in the variables $x \in \mathbb{R}^n$, and let $\deg p$ denote the (total) degree of polynomial p.

A subset X of $\mathbb{R}^n$ is basic semi-algebraic if it is defined as the intersection of finitely many polynomial inequalities, namely $X = \{x \in \mathbb{R}^n : \ g_i(x) \geq 0, i = 1 \dots n_X\}$ with $g_i(x) \in \mathbb{R}[x]$, $i = 1 \dots n_X$.

2 Weak linear program

2.1 Problem statement

Consider the optimal control problem

$$\begin{array}{ll} \inf\limits_{\sigma} & \displaystyle\int_0^T l_{\sigma(t)}(t, x(t))\,\mathrm{d}t \\ \text{s.t.} & \dot{x}(t) = f_{\sigma(t)}(t, x(t)), \\ & x(0) = x_0,\ x(T) = x_T, \\ & (t, x(t)) \in [0, T] \times X, \\ & \sigma(t) \in \{1, 2, \dots, m\}, \end{array} \tag{1}$$

where the infimum is over the sequence σ. The state x is a vector of n elements and the control consists of choosing, for almost all times, which of the m possible dynamics drives the system. Throughout this section, the following assumptions hold:

Assumption 1 *Terminal time T, initial state x_0 and terminal state x_T are given. State constraint set X is given and compact. Let $K := [0, T] \times X$. Functions $f_j : K \mapsto \mathbb{R}^n$ (dynamics), and $l_j : K \mapsto \mathbb{R}$, $j = 1, \dots, m$ (Lagrangians) are given, continuous in all their variables, and Lipschitz continuous with respect to the state variable.*

The integrand in the cost is called a Lagrangian, following the standard optimal control terminology, see e.g., [12]. Note that the above assumptions can be significantly weakened. For instance, mere measurability of the dynamics w.r.t. time, lower semi-continuity of the Lagrangian or unbounded state spaces when the Lagrangian is coercive are often considered in the calculus of variation literature. However, as §3 explores in depth, the question of the numerical representation of problem data is essential for rigorous *global* optimization. Up to our knowledge, only polynomials can offer such guarantees, and Assumption 1 is general enough to cover such an essential case.

Obviously, optimal control problem (1) can be equivalently written as

$$\begin{aligned} \inf_{u} \quad & \int_0^T \sum_{j=1}^{m} l_j(t, x(t))\, u_j(t)\, \mathrm{d}t \\ \text{s.t.} \quad & \dot{x} = \sum_{j=1}^{m} f_j(t, x(t))\, u_j(t) \\ & x(0) = x_0,\ x(T) = x_T, \\ & (t, x(t)) \in K,\ u(t) \in U \end{aligned} \tag{2}$$

where the infimum is with respect to a time-varying vector $u(t)$ which belongs for all $t \in [0, T]$ to the (nonconvex) discrete set

$$U := \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\} \subset \mathbb{R}^m. \tag{3}$$

The next sections explore the various ways this problem can be lifted/relaxed as a linear program over so-called *occupation measures*.

Remark 1 *Section 5 gathers several extensions to problem* (1) *for which the present approach is applicable mutatis mutandis. This section is relegated at the end of the paper for ease of exposition.*

2.2 Occupation measures

Simple examples (see for instance at the end of the paper) show that when measurable functions of time are considered for controls such as in (2), optimal solutions might not exist. This section summarizes several well-known concepts to regain compactness in the control. These tools will then be specialized in the next section for switched systems.

Definition 1 (Young measure) *A Young measure is a time-parametrized family of probability measures $\omega(\mathrm{d}u|t) \in \mathcal{P}(U)$ defined almost everywhere, such that for all continuous functions*[1] *$v \in \mathcal{C}(U)$, the function $t \to \langle v, \omega \rangle$ is Lebesgue measurable.*

Young measures are also called generalized controls or relaxed controls, see for instance [18, Part III]. Note that classical controls $u(t)$, functions of time with values in U, are

[1] Considering measures as a subset of distributions, those may be referred to as test functions

a particular case of Young measures for the choice $\omega(\mathrm{d}u|t) = \delta_{u(t)}(\mathrm{d}u)$. Following the terminology of [50], it is natural to relax (2) into the following "strong form"[2]:

$$\begin{array}{rcll} p_S^* & = & \inf\limits_{\omega} & \displaystyle\int_0^T \sum_{j=1}^m l_j(t,x(t))u_j(t)\,\mathrm{d}t \\ & & \text{s.t.} & \dot{x} = \left\langle \displaystyle\sum_{j=1}^m f_j(t,x(t))u_j,\, \omega(\mathrm{d}u|t) \right\rangle \\ & & & x(0) = x_0,\, x(T) = x_T,\, (t,x(t)) \in K \end{array} \tag{4}$$

where the infimum is w.r.t. a Young measure $\omega(\mathrm{d}u|t) \in \mathcal{P}(U)$. Note that there might be a strict relaxation gap, viz., p_S^* is strictly less than the infimum in problem (1), for ill-posed instances. It is not the purpose of this article to explore such a non-generic case, and we refer to the discussion in, e.g., [24, App. C] and the references therein on such occurrences. There are sufficient conditions guaranteeing the absence of such a gap, referred to as inward-pointing conditions, see for instance [19].

An equivalent way of relaxing (2) is to relax the controls to belong, for all $t \in [0,T]$, to the (convex) simplex

$$\operatorname{conv} U = \{u \in \mathbb{R}^m : \sum_{j=1}^m u_j = 1,\ u_j \geq 0,\ j = 1,\ldots,m\}. \tag{5}$$

In other words, problem (4) on Young measures is equivalent to replacing U with $\operatorname{conv} U$ in problem (2). In [4], problem (4) is called the embedding of problem (1), and it is proved that the set of trajectories of problem (1) is dense (w.r.t. the uniform norm in the space of continuous functions of time) in the set of trajectories of embedded problem (4). Note however that these authors consider the more general problem of switching design in the presence of additional bounded controls in each individual dynamics. To cope with chattering effects due to the simultaneous presence of controls and (initial and terminal) state constraints, they have to introduce a further relaxation of the embedded control problem. In this paper, we do not have controls in the dynamics, and the only design parameter is the switching sequence.

Notice that once a Young measure $\omega(\mathrm{d}u|t)$ is given in problem (4), the corresponding state trajectory is uniquely determined by

$$x(t) = x_0 + \int_0^t \left\langle \sum_{j=1}^m f_j(s,x(s))u_j,\, \omega(\mathrm{d}u|s) \right\rangle \mathrm{d}s$$

and hence the following definition makes sense.

Definition 2 (Relaxed arc) *A pair (x,ω) which is feasible, or admissible, for problem (4) is called a relaxed arc.*

For optimization, strong problem (4) presents an essential first step, as sequential-compactness is regained, guaranteeing the existence of optimal solutions under the sole existence of a

[2] "Strong" is opposed here to the "weak" problem to be defined shortly after.

relaxed arc [50, Th. 1.2]. For *global* optimization though, the nonlinear dependence of (4) on trajectories $x(t)$ is problematic, and an additional generalization of Young measures must be introduced, capturing both controls *and* trajectories:

Definition 3 (Occupation measure) *Given a relaxed arc (x, ω), its corresponding occupation measure is defined by*

$$\mu(A \times B \times C) := \int_A I_B(x(t))\omega(C|t)\,\mathrm{d}t$$

for all Borel subsets A, B, C of $[0, T]$, X and U, resp.

That is, occupation measures assign to each set $A \times B \times C$ the total time "spent" on it by a relaxed arc (x, ω). An alternative, equivalent definition could be

$$\mathrm{d}\mu(t, x, u) = \delta_{x(t)}(\mathrm{d}x)\omega(\mathrm{d}u|t)\mathrm{d}t.$$

By construction, an occupation measure satisfies

$$\langle v, \mu \rangle = \int_0^T \left(\int_U v(t, x(t), u)\omega(\mathrm{d}u|t) \right) \mathrm{d}t \tag{6}$$

for all continuous test functions $v \in \mathcal{C}([0, T] \times X \times U)$. Indeed, by the Riesz representation theorem (see, e.g., [39, §21.5]), an occupation measure could alternatively be defined as the unique measure satisfying (6) for all continuous test functions (recall the hypothesis of compact supports).

Evaluating now a continuously differentiable test function along a trajectory and making use of (6) reveals the following lemma:

Lemma 1 *Given a relaxed arc (x, ω), its corresponding occupation measure satisfies the following* linear *constraints for any test function $v \in \mathcal{C}^1(\mathbb{R} \times \mathbb{R}^n)$:*

$$\left\langle \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \cdot \sum_{j=1}^m f_j\, u_j, \mu \right\rangle = v(T, x_T) - v(0, x_0). \tag{7}$$

Note the use of a scalar product, denoted by a dot, between the gradient vector $\frac{\partial v}{\partial x}$ and the controlled dynamics. In [40], it is proposed then to relax strong problem (4) by optimizing over all Borel measures satisfying (7) instead of simply occupation measures, yielding the following "weak" problem:

$$\begin{array}{rcll} p_W^* & = & \inf\limits_{\mu} & \left\langle \sum\limits_{j=1}^m l_j(t, x(t))u_j(t), \mu \right\rangle \\ & & \text{s.t.} & \forall v \in \mathcal{C}^1(K) : \; v(T, x_T) - v(0, x_0) = \\ & & & \qquad \left\langle \dfrac{\partial v}{\partial t} + \dfrac{\partial v}{\partial x} \cdot \sum\limits_{j=1}^m f_j\, u_j, \mu \right\rangle \\ & & & \mu \in \mathcal{M}^+(K \times U). \end{array} \tag{8}$$

Weak problem (8) is rigorously justified by the following essential result:

Theorem 1 *There is no relaxation gap between optimal control problem* (4) *and measure LP* (8)*, viz.,*

$$p_S^* = p_W^*.$$

Proof: This result is not new, so its proof is only sketched. Obviously, direction $p_S^* \geq p_W^*$ is a simple consequence of Lemma 1 and the continuity of the cost. The other direction, more involved, is the core of [49] and especially [50, Theorem 2.3]. Synthetically, the authors use specialized Hahn-Banach type separation arguments to show that the admissible set for weak problem (8) is the closure of the convex hull of the set of occupation measures of the admissible solutions of (4). That is, one could interpret those results as a Krein-Milman theorem for optimal control problems, see e.g. [39, Chapter 14] for background on the Krein-Milman theorem. □

Note that the use of relaxations and LP formulations of optimal control problems (on ordinary differential equations and partial differential equations) is classical, and can be traced back to the work by L. C. Young, Filippov, as well as Warga and Gamkrelidze, amongst many others. For more details and a historical survey, see, e.g., [18, Part III].

2.3 Modal occupation measures and primal LP

Occupation measures defined in §2.2 are supported on a subset of Euclidean space of dimension $n + m + 1$. As such, they prove to be challenging for practical numerical resolution when the size of the problem increases, a phenomenon known colloquially in the (dual) dynamic programming literature as the "curse of dimensionality".

For instance, the numerical approach of [40] proposes to consider a finite subset of the countably many linear constraints on μ given by (7). Then, as a consequence of Tchakaloff's theorem [29, Th. B.12], there exists an atomic measure with support included in the support of μ, and satisfying the truncated linear constraints exactly. By gridding the support of μ with Dirac measures of unknown masses, one can then approach weakly-$*$ this atomic measure as the spatial resolution of the grid gets finer. This results in a finite-dimensional LP to solve, with a number $\mathcal{O}\left(1/\Delta^{n+m+1}\right)$ of unknowns, with Δ the grid resolution. Therefore, for a fixed resolution, the size of the LP grows exponentially in the size of the state and control spaces.

Other examples suffering a similar fate are LMI relaxations as presented later on in §3, which the authors prefer over LP approximations for their greater rigor – most noticeably the guarantee of obtaining lower bounds and their better convergence properties, see [29, Section 5.4.2]. Indeed, as shown in §3.4, for a given relaxation order, the size of the semidefinite blocks in the LMIs also grows exponentially with the dimension of the underlying space.

For finite-dimensional optimization problems, structural properties of the problem can be exploited to define equivalent LPs using more measures, but of smaller dimensions, see [30]. This allows for much faster numerical resolution. Motivated by these savings, we now introduce specialized occupation measures to satisfy much the same goal for the optimal control of switched systems:

Definition 4 (Modal occupation measures) *The j-th modal occupation measure $\mu_j(\mathrm{d}t, \mathrm{d}x) \in \mathcal{M}^+(K)$ associated with occupation measure μ as in Def. 3, is defined as*

$$\mu_j(A \times B) := \int_A \int_B \int_U u_j \, \mathrm{d}\mu(t, x, u),$$

for all Borel subsets A, B of $[0, T]$, X, resp.

In words, modal occupation measures are the *projections* on $K = [0, T] \times X$ of the first moments of the occupation measure defined in §2.2 w.r.t. the controls. For each Borel subset of K, each modal occupation measure therefore specifies the time spent by the pair $(t, x(t))$ on each mode. As a consequence, Def. 4 implies the following specialization of Lem. 1 for switched systems:

Corollary 1 *Given a relaxed arc (x, ω), its corresponding modal occupation measures satisfy the following* linear *constraints:*

$$\sum_{j=1}^{m} \left\langle \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \cdot f_j, \mu_j \right\rangle = v(T, x_T) - v(0, x_0) \tag{9}$$

for all test functions $v \in \mathcal{C}^1(K)$.

Proof: Notice that by definition of U, the occupation measure satisfies

$$\left\langle \frac{\partial v}{\partial t}, \mu \right\rangle = \left\langle \frac{\partial v}{\partial t} \cdot \sum_{j=1}^{m} u_j, \mu \right\rangle.$$

The corollary is then the immediate application of Def. 4 and Fubini's theorem. □

Corollary 1 naturally points to the following alternative weak problem for switched systems:

$$\begin{array}{rcll} p^* & = & \inf\limits_{(\mu_1, \ldots, \mu_m)} & \sum\limits_{j=1}^{m} \langle l_j, \mu_j \rangle \\ & & \text{s.t.} & \forall v \in \mathcal{C}^1(K) : \; v(T, x_T) - v(0, x_0) = \\ & & & \quad \sum\limits_{j=1}^{m} \left\langle \dfrac{\partial v}{\partial t} + \dfrac{\partial v}{\partial x} f_j, \mu_j \right\rangle \\ & & & \mu_j \in \mathcal{M}^+(K), \quad j = 1, \ldots, m \end{array} \tag{10}$$

where the minimization is w.r.t. the vector of modal occupation measures $(\mu_1, \ldots, \mu_m)$.

This is rigorously justified by the following:

Theorem 2 *There is no relaxation gap between measure LP* (8) *and measure LP* (10)*, viz.,*

$$p^* = p^*_W$$

Proof: We first show $p^* \leq p^*_W$. Consider an admissible measure μ for (10). Then, by Cor. 1, there exists an admissible vector $(\mu_1, \dots \mu_m)$ satisfying (9). As the cost can be treated in an analogous fashion, this shows that any admissible solution for (10) is also admissible for (8) with the same cost, hence $p^* \leq p^*_W$.

For the reverse inequality, consider now a vector $(\mu_1, \dots, \mu_m)$ admissible for (8), and define

$$\bar{\mu} := \sum_{j=1}^{m} \mu_j. \tag{11}$$

By construction, each μ_j is absolutely continuous with respect to $\bar{\mu}$, such that by the Radon-Nikodým Theorem [39, §18.4], there exists a non-negative, measurable function $\bar{u}_j(t,x)$ such that $\mu_j = \bar{u}_j\,\bar{\mu}$. Injecting this back in (11) shows that, in addition, $\sum_{j=1}^{m} \bar{u}_j = 1$, $\bar{\mu}$-almost everywhere. Define now $\mu \in \mathcal{M}^+(K \times U)$, with U as in (3), such that

$$\mathrm{d}\mu(t,x,u) := \left(\sum_{j=1}^{m} \bar{u}_j(t,x) \delta_1(\mathrm{d}u_j | t,x) \right) \mathrm{d}\bar{\mu}(t,x),$$

such that by definition

$$\int_U u_j \,\mathrm{d}\mu(t,x,u) = \bar{u}_j(t,x)\,\mathrm{d}\bar{\mu}(t,x),$$

where the last two equations are understood in a weak sense. Working out the weak dynamics of (10) leads then to:

$$\begin{aligned} \sum_{j=1}^{m} \left\langle \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \cdot f_j, \mu_j \right\rangle &= \sum_{j=1}^{m} \left\langle \left(\frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \cdot f_j \right) \bar{u}_j, \bar{\mu} \right\rangle \\ &= \left\langle \frac{\partial v}{\partial t} + \sum_{j=1}^{m} \frac{\partial v}{\partial x} \cdot f_j\, \bar{u}_j, \bar{\mu} \right\rangle \\ &= \left\langle \frac{\partial v}{\partial t} + \sum_{j=1}^{m} \frac{\partial v}{\partial x} \cdot f_j\, u_j, \mu \right\rangle. \end{aligned} \tag{12}$$

As the cost can be treated in a similar manner, this concludes $p^* \geq p^*_W$, hence the proof. □

Theorem (2) fulfills the objective laid at the beginning of this section. Indeed structured LP (10) involves m modal occupation measures supported on spaces of dimension $n+1$, instead of a single occupation measure on a space of dimension $n+m+1$ as in unstructured LP (8).

2.4 HJB inequalities and dual LP

Before investigating the practical implications on LMI relaxations of structured measure LP (10), we explore its conic dual. This is an interesting result in its own right for so-called "verification theorems" which supply necessary and sufficient conditions in the form

of more traditional HJB inequalities. In addition, practical numerical resolution of the LMI relaxations by primal/dual interior-point methods [35] implies that a strengthening of this dual will be solved as well, as shown later on in §3.3.

In this section, it is first shown that the solution of (10) is attained whenever an admissible solution exists. Then, the dual problem of (10), in the sense of conic duality, is presented. This leads directly to a system of subsolutions of HJB equations. Although, the cost of this problem might not be attained, it is however shown that there is no duality gap between the conic programs.

First of all, we establish that whenever the optimal cost of (10) is finite, there exists a vector of measures attaining the value of the problem:

Lemma 2 *If p^* is finite in problem (10), there exists an admissible $(\mu_1^*, \ldots, \mu_m^*)$ attaining the infimum, viz., such that*

$$\sum_{j=1}^{m} \langle l_j, \mu_j^* \rangle = p^*.$$

Proof: We first show that the mass of each measure is bounded. This is true for each μ_j, since test function $v(t,x) = t$ imposes $\sum_{j=1}^m \langle 1, \mu_j \rangle = T$. Then, following Alaoglu's theorem [39, §15.1], the unit ball in the vector space of compactly supported measures is compact in the weak-$*$ topology. Therefore, any sequence of admissible solutions for (10) possesses a converging subsequence. Since this must be true for any sequence, this is true for any minimizing sequence, which concludes the proof. □

Now, remark that (10) can be seen as an instance of a conic program, called hereafter the primal, in standard form (see for instance [3]):

$$\begin{array}{rcll} p^* & = & \inf\limits_{z_p} & \langle z_p, c \rangle_p \\ & & \text{s.t.} & \mathcal{A}\, z_p = b, \\ & & & x_p \in E_p^+ \end{array} \tag{13}$$

where

- decision variable $z_p := (\mu_1, \ldots, \mu_m)$ belongs to vector space $E_p := \mathcal{M}(K)^m$ equipped with the weak-$*$ topology;
- the cost $c := (l_1, \ldots, l_m)$ belongs to vector space $F_p := \mathcal{C}(K)^m$ equipped with the strong topology;
- $\langle \cdot, \cdot \rangle_p : \; E_p \times F_p \to \mathbb{R}$ is the duality bracket given by the integration of continuous functions with respect to Borel measures, since $E_p = F_p'$, with the prime denoting the topological dual;
- the cone E_p^+ is the non-negative orthant of E_p;
- the linear operator $\mathcal{A} : E_p \to \mathcal{C}^1(K)'$ is given by

$$z_p \mapsto \mathcal{A} z_p = \sum_{j=1}^{m} \mathcal{L}_j \mu_j$$

where $\mathcal{L}_j$ is the adjoint operator of $\mathcal{L}'_j : \mathcal{C}^1(K) \to \mathcal{C}(K)$ defined by

$$v \mapsto \mathcal{L}'_j v := \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \cdot f_j;$$

- right hand side $b := \delta_{(T,x_T)}(\mathrm{d}t, \mathrm{d}x) - \delta_{(0,x_0)}(\mathrm{d}t, \mathrm{d}x)$ belongs to vector space $E_d := \mathcal{C}^1(K)'$.

Lemma 3 *The conic dual of LP problem* (10) *is given by*

$$\begin{array}{rcll} d^* & = & \sup\limits_{v} & v(T,x_T) - v(0,x_0) \\ & & \text{s.t.} & l_j - \mathcal{L}'_j v \geq 0, \ \forall\, (t,x) \in K, \ j = 1 \ldots m, \\ & & & v \in \mathcal{C}^1(K). \end{array} \tag{14}$$

Proof: Following standard results of conic duality (see [1] or [3]), the conic dual of (13) is given by

$$\begin{array}{rcll} d^* & = & \sup\limits_{z_d} & \langle b, z_d \rangle_d \\ & & \text{s.t.} & c - \mathcal{A}' z_d \in \mathcal{C}^+(K)^m, \\ & & & z_d \in F_d \end{array} \tag{15}$$

where

- decision variable $z_d := v$ belongs to vector space $F_d := \mathcal{C}^1(K)$;
- the (pre-)dual cone is $\mathcal{C}^+(K)^m$, i.e., E_p^+ is dual to $\mathcal{C}^+(K)^m$;
- $\langle \cdot, \cdot \rangle_d : \ E_d \times F_d \to \mathbb{R}$ is the appropriate duality pairing between E_d and F_d.

The lemma just details this dual problem. □

Once the duality relationship established, the question arises of whether a duality gap occur between linear problems (10) and (14). The following theorem discards such a possibility:

Theorem 3 *There is no duality gap between primal LP* (10) *and dual LP* (14)*: if there is an admissible vector for* (10)*, then*

$$p^* = d^*.$$

Proof: One can show, following [1, Th. 3.10] (see also the exposition in [3, §4]), that the closure of the cone

$$R := \left\{ (\langle z_p, c \rangle_p, \mathcal{A}\, z_p) : \ z_p \in E_p^+ \right\}$$

belongs to $\mathbb{R} \times E_d$. Note that the weak-$*$ topology is implicit in the definition of E_d, and the closure of R should be understood in such a weak sense.

To prove closure, one may show that, for any sequence of admissible solutions $(z_p^{(n)})_{n\in\mathbb{N}}$, all accumulation points of $(\langle z_p^{(n)}, c \rangle_p, \mathcal{A}\, z_p^{(n)})_{n\in\mathbb{N}}$ belong to R. Note that Lemma 2 establishes

that any sequence $(z_p^{(n)})_{n\in\mathbb{N}}$ has a converging subsequence. Therefore, all that is left to show is the weak-$*$ continuity of $\mathcal{A}$, hence of each $\mathcal{L}_j$. Following [28], this can be shown by noticing that each $\mathcal{L}_j'$ is continuous for the strong topology of $\mathcal{C}^1(K)$, hence for its associated weak topologies. Operators $\mathcal{L}_j'$, hence $\mathcal{A}$, are therefore weakly-$*$ continuous, and each sequence $(\langle x_p^{(n)}, c\rangle_p, \mathcal{A}x_p^{(n)})_{n\in\mathbb{N}}$ converges in R, which concludes the proof. □

Note that what is not asserted in Th. 3 is the existence of a continuous function for which the optimal cost is attained in dual problem (14). Indeed, it is a well-known fact that value functions of optimal control problems, to which v is closely related, may not be continuous, let alone continuously differentiable. However, there does exist an admissible vector of measures for which the optimal cost of primal (10) is attained (whenever there exists an admissible solution), following Lemma 2. This gives practical motivations for working in a purely primal approach on measures, as the dual problem will be solved anyway as a side product, see in the later sections of this article.

We close this section by detailing how dual problem (14) can nonetheless be used analytically as a "verification theorem", following the discussion of [49] for the unstructured case, that is following the dual of (10) instead.

Denote by τ_j the time subset for which the j-th mode is "active":

$$\tau_j := \mathrm{spt}\pi_t\mu_j, \tag{16}$$

where $\pi_t\mu_j$ is the projection, or marginal of μ_j on the time axis, viz., the unique measure such that

$$\langle v(t), \pi_t\mu_j\rangle = \int_K v(t)\,\mathrm{d}\mu_j(t,x)$$

for all continuous test functions v and $j = 1,\ldots,m$. If an admissible control $u(t)$ is piecewise-continuous, then obviously $\tau_j := \int_0^T I_{\{e_j\}}(u(t))\,\mathrm{d}t$, where e_j is the j-th unit vector of $\mathbb{R}^m$. Note however that the intersection of several of the τ_j might be of positive Lebesgue measure. In this case, relaxed controls can only be weakly approximated by fast oscillating sequences of classical controls. In the language of differential inclusions, this consists of realizing a control in the convexified vector field but not belonging to the original vector field.

Corollary 2 *Let (x,ω) be a relaxed arc, and let the intervals τ_j be computed as in* (16) *from their occupation measures. Then there exists a sequence $(v_n)_{n\in\mathbb{N}} \in \mathcal{C}^1(K)$ admissible for* (14)*, such that*

$$\lim_{n\to\infty} \int_{\tau_j} \big(l_j(t,x(t)) - \mathcal{L}_j' v_n(t,x(t))\big)\,\mathrm{d}t = 0,\ j = 1\ldots m,$$

if and only if (x,ω) is a globally optimal solution of optimal control problem (4)*.*

Proof: This corollary explicits weak duality (via the complementarity condition $\langle z_p^*, c - \mathcal{A}' z_d^*\rangle_p = 0$) if z_p^* and z_d^* are optimal for their respective problems, and exploits Lem. 2 guaranteeing the existence of an optimal z_p^*. □

The advantage of such specialized optimality certificates for switched systems over the general case is that the test needs only to be checked on a mode-by-mode basis. This is particularly true if mode-dependent constraints are introduced, see §5.2.

3 LMI relaxations

This section outlines the numerical method to solve (10) in practice, by means of primal-dual moment-SOS LMI relaxations. This is done by restricting the problem data to be polynomial, so that occupation measures can be equivalently manipulated by their moments for the problem at hand, see §3.1. Then, this new infinite-dimensional problem is truncated so as to obtain in §3.2 a convex, finite-dimensional relaxation, with dual described in §3.3. Finally, §3.4 shows how the structural properties of (10) improves the computational load of each of these relaxations in comparison to unstructured relaxations, as could be derived from (8).

In the remainder of the paper, the standing Assumption 1 is therefore strengthened as follows:

Assumption 2 *All functions are polynomial in their arguments:*

$$l_j, f_j \in \mathbb{R}[t,x], \quad j = 1, \ldots, m.$$

In addition, set $K = [0,T] \times X$ is basic semi-algebraic, viz., it is defined as the conjunction of finitely many polynomial inequalities

$$K = \{(t,x) \in \mathbb{R} \times \mathbb{R}^n : \, g_i(t,x) \geq 0, i = 1, \ldots, n_K\}$$

with $g_i \in \mathbb{R}[t,x]$, $i = 1, \ldots, n_K$. Let $g_0(t,x)$ denote the polynomial identically equal to one. In addition, it is assumed that one of the g_i enforces a ball constraint on all variables, which is possible w.l.g. since K is assumed compact (see [29, Th. 2.15] for slightly weaker conditions).

3.1 Moment LP

Given a sequence $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$, let $L_y : \mathbb{R}[x] \to \mathbb{R}$ be the Riesz linear functional

$$f(x) = \sum_\alpha f_\alpha x^\alpha \in \mathbb{R}[x] \quad \mapsto \quad L_y(f) \, = \sum_\alpha f_\alpha y_\alpha.$$

Define the *localizing matrix* of order d associated with sequence y and polynomial $g(x) = \sum_\gamma g_\gamma x^\gamma \in \mathbb{R}[x]$ as the real symmetric matrix $M_d(g\,y)$ whose entry (α, β) reads

$$[M_d(g\,y)]_{\alpha,\beta} = L_y\left(g(x)\, x^{\alpha+\beta}\right) \tag{17}$$

$$= \sum_\gamma g_\gamma \, y_{\alpha+\beta+\gamma}, \quad \forall \alpha, \beta \in \mathbb{N}^n_d. \tag{18}$$

In the particular case that $g(x) = 1$, the localizing matrix is called the *moment matrix*.

The construction of the LMI associated with problem (10) can now be stated. Its decision variable is the aggregate sequence of moments $y = (y_1, y_2, \ldots, y_m)$ where each $y_j = (y_{j,\beta})_{\beta \in \mathbb{N}^{n+1}}$ is the moment sequence of measure $\mathrm{d}\mu_j(t,x)$ for $j = 1, 2, \ldots, m$, i.e.,

$$y_{j,\beta} = \langle v_\beta(t,x), \mu_j \rangle \tag{19}$$

for the particular choice of test function

$$v_\beta(t,x) := t^{\beta_0} x_1^{\beta_1} x_2^{\beta_2} \cdots x_n^{\beta_n}, \;\; \beta \in \mathbb{N}^{n+1}. \tag{20}$$

As each term of the cost of (10) is polynomial by assumption, they can be rewritten as

$$\sum_{j=1}^{m} \sum_{\beta \in \mathbb{N}^{n+1}} c_{j,\beta} y_{j,\beta} = \sum_{j=1}^{m} L_{y_j}(l_j).$$

That is, vector $(c_{j,\beta})_\beta$ contains the coefficients of polynomials l_j expressed in the monomial basis. Similarly, the weak dynamics constraints of (10) need only be satisfied for countably many polynomial test functions, since the measures are supported on compact subsets of $\mathbb{R}^{n+1}$. Hence the weak dynamics defines a linear constraint between moments of the form

$$\sum_{j=1}^{m} \sum_{\beta \in \mathbb{N}^{n+1}} a_{j,\alpha,\beta} y_{j,\beta} = b_\alpha = v_\alpha(T, x_T) - v_\alpha(0, x_0) \tag{21}$$

where coefficients $a_{j,\alpha,\beta}$ can be deduced by identification from

$$\sum_{\beta \in \mathbb{N}^{n+1}} a_{j,\alpha,\beta} y_{j,\beta} = L_{y_j}(\mathcal{L}_j' v_\alpha), \;\; \alpha \in \mathbb{N}^{n+1}. \tag{22}$$

Finally, the only nonlinear constraints are the convex LMI constraints for measure representativeness, i.e., constraints on y such that (19) holds for all $j = 1, \ldots, m$ and $\beta \in \mathbb{N}^{n+1}$. Indeed, it follows from Putinar's theorem [29, Theorem 3.8] that the sequence of moments y has a representing measure defined on a set satisfying Assumption 2 if and only if $M_d(g_i\, y) \succeq 0$ for all $d \in \mathbb{N}$ and for all polynomials g_i defining the set, $i = 0, 1, \ldots, n_K$, and where $A \succeq 0$ stands for matrix A positive semidefinite, i.e., with nonnegative real eigenvalues.

This leads to the problem:

$$\begin{aligned} p^*_\infty \;=\; \inf_{y} \;& \sum_{j=1}^{m} \sum_{\beta \in \mathbb{N}^{n+1}} c_{j,\beta} y_{j,\beta} \\ \text{s.t.} \;& \sum_{j=1}^{m} \sum_{\beta \in \mathbb{N}^{n+1}} a_{j,\alpha,\beta} y_{j,\beta} = b_\alpha, \; \alpha \in \mathbb{N}^{n+1}, \\ & M_d(g_i\, y_j) \succeq 0, \; i = 0 \ldots n_K, \; j = 1 \ldots m, \\ & \qquad d \in \mathbb{N}. \end{aligned} \tag{23}$$

Theorem 4 *Measure LP* (10) *and infinite-dimensional LMI problem* (23) *share the same optimum:*

$$p^* = p^*_\infty.$$

For the rest of the paper, we will therefore use p^* to denote the cost of measure LP (10) or LMI problem (23) indifferently.

3.2 Moment LMI hierarchy

The final step to reach a tractable problem is relatively obvious: infinite-dimensional LMI problem (23) is truncated to its first few moments.

To streamline exposition, first notice in LMI (23) that $M_{d+1}(\cdot) \succeq 0$ implies $M_d(\cdot) \succeq 0$, such that when truncated, only the constraints of highest order must be taken into account. Now, let $d_0 \in \mathbb{N}$ be the smallest integer such that all Lagrangians, dynamics and constraint monomials belong to $\mathbb{N}^{n+1}_{2d_0}$. This is the degree of the so-called *first relaxation*. For any relaxation order $d \geq d_0$, the decision variable is now the vector $(y_{j,\alpha})_\alpha$ with $\alpha \in \mathbb{N}^{n+1}_{2d}$, made of the first $\binom{n+2d+1}{n+1}$ moments of each measure μ_j. Then, define the index set

$$\bar{\mathbb{N}}^{n+1}_{2d} := \left\{\alpha \in \mathbb{N}^{n+1} : \deg \mathcal{L}'_j v_\alpha \leq 2d, j = 1, \ldots, m\right\},$$

viz., the set of monomials for which test functions of the form (20) lead to linear constraints of appropriate degree. By assumption, this set is finite and not empty – the constant monomial always being a member.

Then, the LMI relaxation of order d is given by

$$\begin{array}{rcll} p^*_d & = & \inf\limits_{y} & \sum\limits_{j=1}^{m} \sum\limits_{\beta \in \mathbb{N}^{n+1}_{2d}} c_{j,\beta} y_{j,\beta} \\ & & \text{s.t.} & \sum\limits_{j=1}^{m} \sum\limits_{\beta \in \mathbb{N}^{n+1}_{2d}} a_{j,\alpha,\beta} y_{j,\beta} = b_\alpha,\ \alpha \in \bar{\mathbb{N}}^{n+1}_{2d} \\ & & & M_d(g_i\, y_j) \succeq 0,\ i = 0, \ldots, n_K,\ j = 1, \ldots, m. \end{array} \tag{24}$$

Notice that for each relaxation, we get a standard LMI problem that can be solved numerically by off-the-shelf software. In addition, the relaxations converge asymptotically to the cost of the moment LP:

Theorem 5 *For increasing relaxation orders d, the hierarchy of finite-dimensional LMI relaxations (24) yields an asymptotically converging monotonically non-decreasing sequence of lower bounds on the values of measure LP* (10) *and infinite-dimensional LMI problem* (23) *, i.e.*

$$p^*_{d_0} \leq p^*_{d_0+1} \leq \cdots \leq p^*_\infty = p^*.$$

Proof: By construction, observe that $j > i$ implies $p^*_{d_0+j} \geq p^*_{d_0+i}$, viz., the sequence p^*_d is monotonically non-decreasing. Asymptotic convergence to p^* follows from [29, Theorem 3.8] as in the proof of Theorem 4. □

Therefore, by solving the truncated problem for ever greater relaxation orders, we obtain a monotonically non-decreasing sequence of lower bounds to the true optimal cost.

3.3 SOS LMI hierarchy

As for the measure LP of §2, the moment LMI relaxations detailed in the previous section possess a conic dual. In this section, we show that this dual problem can be interpreted as

a polynomial SOS strengthening of the dual outlined in §2.4. The exact form of the dual problem is an essential aspect of the numerical method, since it will be solved implicitly whenever primal-dual interior-point algorithms are used for solving the moment LMI hierarchy of §3.2.

Let $\mathbb{S}^n$ be the space of symmetric $n \times n$ real matrices. One can show (see, e.g., [29, §C]) that, for $A, B \in \mathbb{S}^n$, $\langle A, B\rangle := \operatorname{trace} AB$ is a duality bracket $\mathbb{S}^n \times \mathbb{S}^n \mapsto \mathbb{R}$, and that $\{A \in \mathbb{S}^n : A \succeq 0\}$ defines a convex cone of $\mathbb{S}^n$. In problem (24), let us define the matrices $A_{i,\beta} \in \mathbb{S}^{\binom{n+d+1}{n+1}}$ satisfying the identity

$$M_d(g_i\, y) = \sum_{\beta} A_{i,\beta} y_\beta$$

for every sequence $(y_\beta)_\beta$ and $i = 0, 1, \ldots, n_K$.

Proposition 1 *The conic dual of moment LMI problem* (24) *is given by the SOS LMI problem*

$$\begin{array}{ll} \sup\limits_{z,Z} & \sum\limits_{\alpha \in \bar{\mathbb{N}}^{n+1}_{2d}} b_\alpha z_\alpha \\ s.t. & \sum\limits_{\alpha \in \bar{\mathbb{N}}^{n+1}_{2d}} a_{j,\alpha,\beta} z_\alpha + \sum\limits_{i=0}^{n_K} \langle A_{i,\beta}, Z_{i,j}\rangle = c_{j,\beta},\ \beta \in \mathbb{N}^{n+1}_{2d}, \\ & Z_{i,j} \succeq 0, \quad i = 0, 1, \ldots, n_K,\ j = 1, \ldots, m. \end{array} \tag{25}$$

Proof: Replacing the equality constraints in (24) as two inequalities, it is easy to see that the moment relaxation can be written as an instance of LP (15), whose dual is given symbolically by (13). Working out the details leads to the desired result, using for semi-definite constraints the duality bracket as explained earlier. □

The relationship between (25) and (14) might not be obvious at a first glance. Denote by $\Sigma[z]$ the subset of $\mathbb{R}[z]$ that can be expressed as a finite sum of squares of polynomials. Then a standard interpretation (see, e.g., [29]) of (25) in terms of such objects is given by the next proposition.

Proposition 2 *LMI problem* (25) *can be stated as the following polynomial SOS strengthening of problem* (14)*:*

$$\begin{array}{ll} \sup & v(T, x_T) - v(0, x_0) \\ s.t. & l_j - \mathcal{L}'_j v = \sum\limits_{i=0}^{n_K} g_i s_{i,j}, \quad j = 1, \ldots, m \end{array} \tag{26}$$

where the maximization is w.r.t. the vector of coefficients z of polynomial $v(t,x) = \sum_{\alpha \in \bar{\mathbb{N}}^{n+1}_{2d}} z_\alpha v_\alpha(t,x)$ and the vectors of coefficients of polynomials $s_{i,j} \in \Sigma[t,x]$, $\deg g_i\, s_{i,j} \leq 2d$.

Proof: By (21), the cost of (26) is equivalent to (25). Then multiply each scalar constraint (indexed by α) by v_α as given by (20), and sum them up. By definition, $\sum_\alpha c_{j,\alpha} v_\alpha = l_j$,

and similarly, $\sum_\beta a_{j,\alpha,\beta} z_\alpha v_\alpha = \mathcal{L}'_j v_\alpha$. The conversion from the semi-definite terms to SOS exploits their well-known relationship (e.g., [29, §4.2]) to obtain the desired result, by definition (17) of the localizing matrices. □

Prop. 2 specifies in which sense "polynomial SOS strengthenings" must be understood: positivity constraints of (14) are enforced by SOS certificates, and the decision variable of (14) is now limited to polynomials of appropriate degrees.

Finally, the following result states that no numerical troubles are expected when using classical interior-point algorithms on the primal-dual LMI pair (24-25).

Proposition 3 *The infimum in primal LMI problem (24) is equal to the supremum in dual LMI problem (25), i.e., there is no duality gap.*

Proof: By Assumption 2, one of the polynomials g_i in the description of set K enforces a ball constraint. The corresponding localizing constraint $M_d(g_i\, y_j) \succeq 0$ then implies that the vector of moments y is bounded in LMI problem (24). Then to prove the absence of duality gap, we use the same arguments as in the proof of Theorem 4 in Appendix D of [24]. □

3.4 Computational complexity

This section details rough estimates on the numerical gains expected by employing modal occupation measures in LMI relaxations as opposed to the more generic method developed in [28].

Following the standing assumptions, the computational complexity of solving an LMI relaxation of order d with $\bar{p}$ occupation measures supported on a space of dimension $n + m + 1$ (time, states, controls) is dominated by $\bar{p}$ LMI constraints of size $\bar{m}$ (the moment matrices of the occupation measures) in $\bar{n}$ variables (the moment vectors of the occupation measures), with

$$\bar{n} = \binom{n+m+2d+1}{n+m+1}, \quad \bar{m} = \binom{n+m+d+1}{n+m+1}.$$

A standard primal-dual interior-point algorithm to solve this LMI at given relative accuracy $\epsilon > 0$ (duality gap threshold) requires a number of iterations (Newton steps) growing as $\mathcal{O}(\bar{p}^{\frac{1}{2}}\bar{m}^{\frac{1}{2}}) \log \epsilon$, see, e.g., [5, Section 6.6.3]. In the real model of computation (for which each addition, subtraction, multiplication, division of real numbers has unit cost), each Newton iteration requires no more than $\mathcal{O}(\bar{n}^3) + \mathcal{O}(\bar{p}\bar{n}^2\bar{m}^2) + \mathcal{O}(\bar{p}\bar{n}\bar{m}^3)$ operations. When solving a hierarchy of simple LMI relaxations as described above, the accuracy ϵ is fixed, the number of states n and controls m are fixed the number $\bar{p}$ of measures is fixed, and the relaxation order d varies, such that $\mathcal{O}(\bar{n}) = \mathcal{O}(\bar{m}) = \mathcal{O}(d^{4(n+m+1)})$ and the dominating term in the complexity estimate grows in $\mathcal{O}(\bar{p}^{\frac{3}{2}} d^{\frac{9}{2}(n+m+1)})$, which clearly shows a strong dependence on the number of state and control variables.

Therefore, an LMI relaxation of an optimal switching control problem with n states and m controls will be solved in $\mathcal{O}(d^{\frac{9}{2}\,(n+m+1)})$ operations when solved with the general

formulation of [28] and $\mathcal{O}(m^{\frac{3}{2}}\, d^{\frac{9}{2}\,(n+1)})$ when solved with structured formulation (10). That is a $\frac{n+1}{n+m+1}$ reduction of the polynomial rate at which the CPU time grows with relaxation orders. This back-of-the-envelope estimation is slightly conservative in practice, as shown in §6.2. Note that the same analysis as performed with the reported computation times of [7] or [41] reveals a similar underestimation of actual computation gains by applying the method described in this paper. As seen in these examples, those gains are substantial and justify the practical use of the modal occupation measures for switched systems.

4 Trajectory extraction

This section presents a method for extracting approximate optimal trajectories from moment data.

As presented so far, the method provides a converging sequence of lower bounds to the value of the control problem. This already provides a valuable tool to the user, allowing to certify as *globally* optimal local solutions found by other means such as a local solver or heuristics – or even reject unsound candidate solutions, see §6.3. In addition, comparing lower bounds with the value of an admissible solution, when they do not match, allows to easily quantify the maximal gain that may be achieved by investing more numerical efforts for that particular problem.

As useful as these may prove, it seems intuitive that much more than the cost could be extracted from moment data, especially at sufficiently high relaxation orders. In this section, we show that optimal controls and trajectories can indeed be inferred. The focus here is not on obtaining very precise estimates of the optimal solution, as this would require an impractically large number of test functions in the weak ODE constraints. In addition, estimates of arbitrary high precision can already be obtained routinely from local optimal control solvers (see, e.g., `BOCOP` [6]), *if suitably initialized*. Therefore, what is only needed from the moment data is a sensible approximation of the globally optimal solution(s), to be then simply refined by local methods.

Consider the trajectory $x(t)$ which must be reconstructed on $[0,T]$ from partial moment sequence $\left\{L_{y^j}(t^\alpha x^\beta)\right\}_{\alpha+|\beta|\leq 2r}$, $j = 1 \dots m$. As a consequence of (1), see [50, Cor. 1.4], optimal solutions $\{\mu_j\}$ are supported on the graph of optimal trajectories. Therefore, one possible strategy for reconstructing $x(t)$ is to construct an approximate support on a discretization of $[0,T] \times X$ from the partial moment data. We propose a specialization for switched systems - needed for recovering the modal duty cycles – of the method developed in [15]. As in this reference, we assume that the optimal pair $(u^*(t), x^*(t))$ is unique for ease of exposition. We also assume for the same reason that the state space is unidimensional, as the outlined procedure could be trivially repeated for each component of a multidimensional state space.

Fix a mesh resolution ε. Denote then by X_ε a mesh supported on finitely many points, such that for all $x \in X$, there exists $x_i \in X_\varepsilon$ such that $|x_i - x| \leq \varepsilon$. Let T_ε denote the

analogous discretization of $[0,T]$. It is proposed in [15, Th. 3] to solve the program

$$\begin{aligned} &\min_{\tilde{\mu}_j} \|L_{y^j}(t^\alpha x^\beta) - \langle t^\alpha x^\beta, \tilde{\mu}_j\rangle\| \\ &\text{s.t. } \tilde{\mu}_j \in \mathcal{M}^+(T_\varepsilon \times X_\varepsilon), \end{aligned} \tag{27}$$

to obtain the best discrete approximation $\tilde{\mu}_j$ of the partial moment sequence $\{y^j\}$. Indeed, this can be recast as a simple finite-dimensional, linear program, since measures in $\mathcal{M}^+(T_\varepsilon \times X_\varepsilon)$ can be parametrized by a discrete set of weights, each assigned to an atom supported on $T_\varepsilon \times X_\varepsilon$. The approximate trajectory way points are simply deduced from the support of the optimal solution of (27), viz., from the non-zero atomic weights.

This simple procedure could be repeated for each modal measure μ_j in order to reconstruct the full trajectory. Active modes along the trajectory are then deduced by inspecting the relative modal weights at each atom. However, this naive approach proves to be very imprecise in practice, since modes are deduced from solutions of independent optimization problems (one for each measure). In particular, the sum of all modal measures must disintegrate as (as plugging test functions of time only in (8) reveals):

$$\sum_{j=1}^{m} \mu_j = \zeta(\mathrm{d}x|t)\,\mathrm{d}t, \tag{28}$$

where ζ is a probability measure on X define almost-everywhere on $[0,T]$. However, the discrete version of this property is not guaranteed with independent solving of (27), so that duty cycles can hardly be reconstructed. This observation leads directly to the following alternative program:

$$\begin{aligned} \min_{\tilde{\mu}_1,\ldots,\tilde{\mu}_m} \; & \sum_{j=1}^{m} L_{y^j}(t^\alpha x^\beta) - \langle t^\alpha x^\beta, \tilde{\mu}_j\rangle \\ \text{s.t. } & \sum_{j=1}^{m}\sum_{k=1}^{|X_\epsilon|} \tilde{\mu}_j\{(t_i,x_k)\} = (\Delta t)_i, \quad i = 1,\ldots,|T_\epsilon|, \\ & \tilde{\mu}_j \in \mathcal{M}^+(T_\varepsilon \times X_\varepsilon), \end{aligned} \tag{29}$$

where $(\Delta t)_i$ is the length of the ith cell, containing $t_i \in T_\epsilon$, of a given partition of $[0,T]$. In (29), we have done nothing else that to couple the modal reconstruction by constraining the discretization of $\zeta(\mathrm{d}x|t_i)$ to be a probability measure. When the solution of the optimal control problem is unique, it is then possible to approximate the value of $x^*(t)$ at the time nodes via

$$x^*(t_i) \approx \frac{\sum_{j=1}^{m}\sum_{k=1}^{|X_\epsilon|} \tilde{\mu}_j\{(t_i,x_k)\}\, x_k}{(\Delta t)_i},$$

and the duty cycles via

$$d_j^*(t_k) \approx \frac{\sum_{k=1}^{|X_\epsilon|} \tilde{\mu}_j\{(t_i,x_k)\}}{(\Delta t)_i}.$$

To show that a relaxation has converged, we use those way-points to hot start a numerical method, see the examples in §6. If the cost of the local method agrees with some tolerance

to the relaxation cost, we then terminate the moment hierarchy and validate the local solution as globally optimal. In practice, and in accordance with the analysis of §3.4, problems with state-space smaller than 6 can be usually solved with a globality gap of a few percent at most. The method may still be applied to larger problems; For some instances, one may still solve the problem, whereas on some others, the method simply provide a lower bound on the cost.

5 Extensions

Several standard extensions of our results are now presented.

5.1 Free/distributed initial state

The approach can easily take into account a free initial state and/or time, by introducing an initial occupation measure $\mu_0 \in \mathcal{M}^+(\{0\} \times X_0)$, with $X_0 \subset \mathbb{R}^n$ a given compact set, such that for an admissible starting point $(t_0, x(t_0))$,

$$\langle v, \mu_0 \rangle = v(t_0, x(t_0)), \tag{30}$$

for all continuous test functions $v(t, x)$ of time and space. That is, in weak problem (10), one replaces some of the boundary conditions by injecting (30) and making μ_0 an additional decision variable. In dual (14), this adds a constraint on the initial value and modifies the cost.

Similarly, a terminal occupation measure μ_T can be introduced for free terminal states and/or time. Note that injecting both (30) and its terminal counterpart in (10) requires the introduction of an additional affine constraint to exclude trivial solutions, e.g., $\langle 1, \mu_0 \rangle = 1$ so that μ_0 is a probability measure. In dual problem (14), this introduces an additional decision variable.

More interestingly, the initial time may be fixed w.l.g. to $t_0 = 0$, but only the spatial probabilistic distribution of initial states is known. Let $\xi_0(\mathrm{d}x) \in \mathcal{P}(X_0)$ be the measure whose law describes such a distribution. Then, the additional constraints for (10) are

$$\langle v, \mu_0 \rangle = \int_{X_0} v(0, x)\, \xi_0(\mathrm{d}x).$$

As remarked in [28], this changes the interpretation of LP (10) to the minimization of the expected value of the cost given the initial distribution. See also [24] for the Liouville interpretation of the LP as transporting measure ξ_0 along the optimal flow.

5.2 Additional constraints

In measure LP (10), each modal occupation measure is supported on the same set K. This simply translates the fact that state constraints are mode-independent. However, nothing prevents the use of specific mode constraints by defining m sets $\{K_j\}_{j=1,\ldots,m}$.

In unstructured LP (8), these would be specified by constraining the support of μ to $\prod_{j=1}^{m}\{e_j\} \times K_j$, where e_j is the unit vector in $\mathbb{R}^m$ whose j-th entry is 1. Then, in the proof of Th. 2, for direction $p^* \leq p^*_W$, it is easy to see that this implies that each μ_j is supported on K_j. For the reverse direction, one can define w.l.g. the support of $\bar{\mu}$ to be $K := \cup_{j=1}^{m} K_j$ by extending that of each μ_j to K and setting $\mu_j(K \setminus K_j) = 0$. Then μ has the expected support.

A second class of constraints that can be easily accounted for by the method are a finite number of integral constraints of the form

$$\int_0^T h^k_{\sigma(t)}(t, x(t))\,\mathrm{d}t \leq e_k, \qquad k = 1, \ldots, n_i \tag{31}$$

with given $h^k_j \in \mathbb{R}[t, x]$, $j = 1, \ldots, m$ and $e_k \in \mathbb{R}$, $k = 1, \ldots, n_i$. Indeed, rewriting the constraint in the equivalent formalism of (2) and using modal occupation measures as in Def. 4, (31) becomes

$$\sum_{j=1}^{m} \langle h^k_j, \mu_j \rangle \leq e_k. \tag{32}$$

That is, in moment problem (24), whose first relaxation order may need to be updated in function of the degrees of each h^k_j, each of the n_i integral constraints simply appears as an additional linear constraint. The interest of treating these integral constraints explicitly (instead of rewriting them as additional states to fit in the rigid formalism of (1)) is obvious from §3.4 in terms of computational load. Conversely, some translational-invariant problems might strongly benefit in replacing selected states by integral constraints, see, e.g., [41].

5.3 Wider class of dynamical systems

First note that additional drift terms, viz., dynamics of the form $\dot{x} = f_0 + \sum_{j=1}^{m} f_j\, u_j$, fit easily within the framework of problem (1) by adding the drift term to each mode dynamics.

Secondly, in a related fashion, affine-in-the-control problems with polytopic control sets more general that the box described in (5) can also be handled by the formalism presented in this paper. Indeed, the problem is equivalent to controlling (with relaxed objects – see section §2) the system with modes associated with vertices of the control polytope.

Finally, dynamics of the form $f_j(t, x, u_j)$, parametrized by a mode-dependent measurable control $u_j(t)$, can be incorporated into the presented formalism by combining the approach in this paper with that of [28]. However, mode-dependent states, where the size of the state-space varies for each given mode change and/or whose meaning is mode-dependent, must be incorporated by extending those states to all modes with a null vector field. It is an open question whether a more elegant approach can be incorporated into the framework of this paper.

5.4 Open-loop versus closed-loop

In problem (1), the control signal is the switching sequence $\sigma(t)$ which is a function of time: this is an open-loop control, similarly to what was proposed in [13] for impulsive control design. One could also want to constrain the switching sequence to be an explicit or implicit function of the state, i.e., $\sigma(x(t))$, a closed-loop control signal. In this case, each occupation measure, explicitly depending on time, state and control, may be disintegrated as $\mathrm{d}\mu(t,x,u) = \xi(\mathrm{d}t \mid t,u)\omega(\mathrm{d}u \mid t)\mathrm{d}t$, and we should follow the framework described originally in [28].

5.5 Switching and impulsive control

We may also combine switching control and impulsive control if we extend the system dynamics to

$$\mathrm{d}x(t) = \sum_{k=1}^{m} f_k(x(t))\, u(t)\, \mathrm{d}t + \sum_{j=1}^{p} g_j(t)\nu_j(\mathrm{d}t),$$

which must be understood in a weak sense. Here, g_j are given continuous vector functions of time and ν_j are signed measures to be found, jointly with the switching signal $u(t)$. Whereas modal occupation measures are restricted to be absolutely continuous w.r.t. the Lebesgue measure of time, impulsive control measures ν_j can concentrate in time. For example, for a dynamical system $\mathrm{d}x(t) = g(t)\nu(\mathrm{d}t)$, a Dirac measure $\nu(\mathrm{d}t) = \delta_s$ enforces at time $t = s$ a state jump $x^+(s) = x^-(s) + g(s)$. In this case, to avoid trivial solutions, the objective function should penalize the total variation of the impulsive control measures, see [13].

6 Examples

In this section, we show several examples to illustrate the method. The first example illustrates the fact that the infimum is not attained even in the simplest instances of optimal control problem (1). The second example is the constrained double integrator example of [28]. The final two examples are taken from [48], where they are solved by a local optimization method dedicated for switched systems. These last examples highlight how our method can complement more traditional approaches.

All moment relaxations were built with the `GloptiPoly` toolbox [26] and solved via the `Mosek` semi-definite solver [35], interfaced through `Yalmip` [34].

6.1 Chattering

Consider the scalar ($n = 1$) optimal control problem (1):

$$\begin{array}{rcll} p^* & = & \inf & \displaystyle\int_0^1 x^2(t)\mathrm{d}t \\ & & \text{s.t.} & \dot{x}(t) = a_{\sigma(t)}x(t), \\ & & & x(0) = \frac{1}{2}, \\ & & & x(t) \in [-1,\ 1], \quad \forall t \in [0,1] \end{array}$$

where the infimum is w.r.t. a switching sequence $\sigma : [0,1] \mapsto \{1,\ 2\}$ and

$$a_1 := -1, \quad a_2 := 1.$$

In Table 1 we report the lower bounds p_d^* on the optimal value p^* obtained by solving LMI relaxations (24) of increasing orders d, rounded to 5 significant digits. We also indicate the number of variables $\bar{n}$ (i.e., total number of moments) of each LMI problem, as well as the zeroth order moment of each occupation measure (recall that these are approximations of the time spent on each mode). We observe that the values of the lower bounds and the masses stabilize quickly.

d	p_d^*	$\bar{n}$	$y_{1,0}$	$y_{2,0}$
1	$-5.9672 \cdot 10^{-9}$	18	0.74056	0.25944
2	$4.1001 \cdot 10^{-2}$	45	0.75170	0.24830
3	$4.1649 \cdot 10^{-2}$	84	0.74632	0.25368
4	$4.1666 \cdot 10^{-2}$	135	0.74918	0.25082
5	$4.1667 \cdot 10^{-2}$	198	0.74974	0.25026
6	$4.1667 \cdot 10^{-2}$	273	0.74990	0.25010
7	$4.1667 \cdot 10^{-2}$	360	0.74996	0.25004

Table 1: Lower bounds p_d^* on the optimal value p^* obtained by solving LMI relaxations of increasing orders d; $\bar{n}$ is the number of variables in the LMI problem; $y_{j,0}$ is the approximate time spent on each mode $j = 1, 2$.

In this simple case, it is easy to obtain analytically the optimal switching sequence: it consists of driving the state from $x(0) = \frac{1}{2}$ to $x(\frac{1}{2}) = 0$ with the first mode, i.e., $u_1(t) = 1, u_2(t) = 0$ for $t \in [0, \frac{1}{2}[$, and then chattering between the first and second mode with equal proportion so as to keep $x(t) = 0$, i.e., $u_1(t) = \frac{1}{2}, u_2(t) = \frac{1}{2}$ for $t \in]\frac{1}{2}, 1]$. It follows that the infimum is equal to

$$p^* = \int_0^{1/2} \left(\frac{1}{2} - t\right)^2 \mathrm{d}t = \frac{1}{24} \approx 4.1667 \cdot 10^{-2}.$$

Because of chattering, the infimum in problem (1) is not attained by an admissible switching sequence. It is however attained in the convexified problem (4).

The optimal moments can be obtained analytically

$$y_{1,\alpha} = \int_0^{\frac{1}{2}} t^\alpha \mathrm{d}t + \frac{1}{2}\int_{\frac{1}{2}}^1 t^\alpha \mathrm{d}t = \frac{2 + 2^{-\alpha}}{4 + 4\alpha},$$

$$y_{2,\alpha} = \frac{1}{2}\int_{\frac{1}{2}}^{1} t^\alpha \mathrm{d}t = \frac{2-2^{-\alpha}}{4+4\alpha}$$

and they can be compared with the following moment vectors obtained numerically at the 7th LMI relaxation:

$$\begin{array}{rcl} \text{computed } y_1 & = & (0.74996\ 0.31246\ 0.18746\ 0.13277\cdots), \\ \text{exact } y_1 & = & (0.75000\ 0.31250\ 0.18750\ 0.13281\cdots), \\ \text{computed } y_2 & = & (0.25004\ 0.18754\ 0.14588\ 0.11723\cdots), \\ \text{exact } y_2 & = & (0.25000\ 0.18750\ 0.14583\ 0.11719\cdots). \end{array}$$

We observe that the computed moments closely match the exact optimal moments, so that the approximate control law extracted from the computed moments will be almost optimal.

6.2 Double integrator

We revisit the double integrator example with state constraint studied in [28], formulated as the following optimal switching problem:

$$\begin{array}{rcll} p^* & = & \inf & T \\ & & \text{s.t.} & \dot{x}(t) = f_{\sigma(t)}(x(t)), \\ & & & x(0) = (1,\ 1), \quad x(T) = (0,\ 0) \\ & & & x_2(t) \geq -1, \quad \forall t \in [0,T] \end{array}$$

where the infimum is w.r.t. a switching sequence $\sigma : [0,T] \mapsto \{1,\ 2\}$ with free terminal time $T \geq 0$ and affine dynamics

$$f_1 := \begin{pmatrix} x_2 \\ -1 \end{pmatrix}, \quad f_2 := \begin{pmatrix} x_2 \\ 1 \end{pmatrix}.$$

We know from [28] that the optimal sequence consists of starting with mode 1, i.e., $u_1(t) = 1$, $u_2(t) = 0$ for $t \in [0,2]$, then chattering with equal proportion between mode 1 and 2, i.e., $u_1(t) = u_2(t) = \frac{1}{2}$ for $t \in [2, \frac{5}{2}]$ and then eventually driving the state to the origin with mode 2, i.e., $u_1(t) = 0$, $u_2(t) = 1$ for $t \in [\frac{5}{2}, \frac{7}{2}]$. Here too the infimum $p^* = \frac{7}{2}$ is not attained for problem (1), whereas it is attained with the above controls for problem (4).

In Table 2 we report the lower bounds p_d^* on the optimal value p^* obtained by solving LMI relaxations (24) of increasing orders d, rounded to 5 significant digits. We also indicate the number of variables $\bar{n}$ (i.e., total number of moments) of each LMI problem, as well as the zeroth order moment of each occupation measure (recall that these are approximations of the time spent on each mode). We observe that the values of the lower bounds and the masses stabilize quickly to the optimal values $p^* = \frac{7}{2}$, $y_{1,0} = \frac{5}{2}$, $y_{2,0} = \frac{5}{4}$.

We carried out a comparison of computational times required for solving the unstructured generic LMI relaxations of [28] versus their structured modal counterpart, for relaxation orders 1 to 7. The computational gains induced by using modal occupation measures follow the expected exponential improvement, with an exponent of about 1.7, better than the $\frac{4}{3}$ predicted by the asymptotic analysis of §3.4.

d	p_d^*	$\bar{n}$	$y_{1,0}$	$y_{2,0}$
1	2.5000	30	1.7500	0.75000
2	3.2015	105	2.1008	1.1008
3	3.4876	252	2.2438	1.2438
4	3.4967	495	2.2484	1.2484
5	3.4988	858	2.2494	1.2494
6	3.4993	1365	2.2496	1.2497
7	3.4996	2040	2.2498	1.2498

Table 2: Lower bounds p_d^* on the optimal value p^* obtained by solving LMI relaxations of increasing orders d; $\bar{n}$ is the number of variables in the LMI problem; $y_{j,0}$ is the approximate time spent on each mode $j = 1, 2$.

6.3 Switched LQR

We consider the problem

$$\begin{aligned} p^* = \quad & \inf \int_0^T 0.01\, u^2(t)\, \mathrm{d}t + \sum_{i=1}^{3} (x_i(T) - 1)^2 \\ & \text{s.t. } \dot{x}(t) = A\, x(t) + B_{\sigma(t)}\, u(t), \\ & \quad x(0) = (0, 0, 0), \\ & \quad u(t) \in U = [-20, 20], \\ & \quad \sigma \in \{1, 2, 3\}, \\ & \quad T = 2, \end{aligned}$$

with problem data

$$A = \begin{pmatrix} 1.0979 & -0.0105 & 0.0167 \\ -0.0105 & 1.0481 & 0.0825 \\ 0.0167 & 0.0825 & 1.1540 \end{pmatrix},$$

$$B_1 = \begin{pmatrix} 0.9801 \\ -0.1987 \\ 0 \end{pmatrix},\ B_2 = \begin{pmatrix} 0.1743 \\ 0.8601 \\ -0.4794 \end{pmatrix},\ B_3 = \begin{pmatrix} 0.0952 \\ 0.4699 \\ 0.8776 \end{pmatrix}$$

and where the infimum is w.r.t. a switching sequence $\sigma(t)$ but also a classical control $u(t)$, consistently with the extension sketched in §5.3. The lower bounds and computation times are given in Tab. 3. We also report the cost of a direct, local, optimal control methods as implemented in `BOCOP` [6] via embedding (4), and hot started with the method of §4. As one can see, the costs are in close agreement, and we can numerically certify the local solutions as globally optimal. We also used `VSDP` [27] to strengthen the semi-definite bounds, so as to compute verified lower bounds on the cost of a relaxation via rigorous interval arithmetic. The lower bound for the third order relaxation could be certified to the significant figures presented in Tab. 3. The disagreement with the lower cost reported in [48] is due to numerical round-off issues in the execution of their local method. In fact, integrating finely the solution reported in [48] leads the a candidate arc whose cost is *above* our certified lower bound.

d	p_d^*	$\bar{n}$
1	$0.071\,10^{-3}$	93
2	$1.823\,10^{-3}$	518
3	$1.829\,10^{-3}$	1806
...		
BOCOP	$1.831\,10^{-3}$	

Table 3: Lower bounds p_d^* and number of variables $\bar{n}$ of the LMI problem in function of relaxation order d for the problem of §6.3. The value found by hot starting BOCOP with our method is also shown.

d	p_d^*	$\bar{n}$
1	0.0000	62
2	4.4886	322
3	4.7265	1092
4	4.7298	2904
5	4.7304	6578
...		
BOCOP	4.7304	

Table 4: Lower bounds p_d^* and number of variables $\bar{n}$ of the LMI problem in function of relaxation order d for the problem of §6.4. The value found by hot starting BOCOP with our method is also shown.

6.4 Double tank problem

We consider the double tank problem of [48]:

$$\begin{aligned} p^* = \quad & \inf \int_0^T 2(x_2(t) - 3)\,\mathrm{d}t \\ & \text{s.t. } \dot{x}_1(t) = c_{\sigma(t)} - \sqrt{x_1(t)} \\ & \quad \dot{x}_2(t) = \sqrt{x_1(t)} - \sqrt{x_2(t)}, \\ & \quad x(0) = (2,\ 2)' \\ & \quad \sigma \in \{1, 2\}, \\ & \quad T = 10, \end{aligned}$$

with data $c_1 = 1, c_2 = 2$ and where the infimum is w.r.t. a switching sequence $\sigma(t)$. By introducing the lifts $\ell_i = \sqrt{x_i}$, algebraically constrained as $\ell_i^2 = x_i$, $\ell_i \geq 0$, the problem is easily recast with polynomial data. Table 4 reports the relaxation costs of our method along with the cost of BOCOP, hot-started by the moment data. Again, agreement is reached quickly. Note that the cost is here strictly better than the one reported in [48], which is stuck in a local but not global minimum. This adverse behavior is of course inherent to any local optimization scheme, and shows the interest of methods which allow to obtain decent starting point to initialize them. Obviously, such a pass with a local method is still necessary to obtain very precise numeric solutions, and our general framework allows to combine the best of the global and local worlds.

d	p_d^*	$\bar{n}$
2	$9.0255\,10^{-3}$	3135
3	$9.4229\,10^{-2}$	21021
...		
BOCOP	$9.5754\,10^{-2}$	

Table 5: Lower bounds p_d^* and number of variables $\bar{n}$ of the LMI problem in function of relaxation order d for the problem of §6.5. The value found by hot starting BOCOP with our method is also shown.

6.5 Quadrotor

For our last example, we take the quadrotor example considered in [48]:

$$\begin{aligned} p^* = \quad & \inf \int_0^T 5\,u^2(t)\,\mathrm{d}t + \left\| \begin{smallmatrix} \sqrt{5}(x_1(T)-6) \\ \sqrt{5}(x_2(T)-1) \\ \sin(\frac{x_3(T)}{2}) \end{smallmatrix} \right\|_2^2 \\ , \quad & \text{s.t. } \ddot{x} = f_{\sigma(t)}, \\ & x(0) = (0,1,0), \dot{x}(0) = (0,0,0), \\ & u(t) \in U = [0, 10^{-3}], \\ & \sigma(t) \in \{1,2,3\}, \\ & T = 7.5, \end{aligned} \tag{33}$$

with data $M = 1.3$, $L = 0.3050$, $I = 0.0605$, $g = 9.8$, and

$$f_1 = \begin{pmatrix} \frac{\sin x_3(t)}{M}(u(t)+Mg) \\ \frac{\cos x_3(t)}{M}(u(t)+Mg)-g \\ 0 \end{pmatrix}, \tag{34}$$

$$f_2 = \begin{pmatrix} g\sin x_3(t) \\ g\cos x_3(t)-g \\ \frac{-Lu(t)}{I} \end{pmatrix}, \quad f_3 = \begin{pmatrix} g\sin x_3(t) \\ g\cos x_3(t)-g \\ \frac{Lu(t)}{I} \end{pmatrix} \tag{35}$$

and where the infimum is w.r.t. a switching sequence $\sigma(t)$ but also a classical control $u(t)$, consistently with the extension sketched in §5.3. We turn this problem into one with polynomial data by considering another parametrization of the attitude x_3 belonging to $SO(2)$, namely by replacing x_3 with $\ell_1 := \cos x_3$ and $\ell_2 := \sin x_3$, who must then satisfy the ODE

$$\begin{aligned} \dot{\ell}_1 &= -\ell_2 \dot{x}_3, \\ \dot{\ell}_2 &= \ell_1 \dot{x}_3. \end{aligned} \tag{36}$$

The problem has however too high of a dimension to hope to obtain tight bounds systematically, with 7 states/lifts, 1 control and time (see the analysis in §3.4). Nonetheless, for this specific example, the bounds presented in Tab. 5 show convergence in two relaxations only, since the local minimum obtained with BOCOP is very close to the bound obtained at the relaxation of order $d = 3$, solved in about 7 minutes on our desktop PC. Note that the relaxation of order $d = 4$, involving $\bar{n} = 98670$ moments, would still be attainable with modern machines, at the expense of a full day of computations. Finally, also note that the solution is again better than the local minimum reported in [48].

Acknowledgments:

This work benefited from discussions with Milan Korda, Jean-Bernard Lasserre and Luca Zaccarian.

References

[1] E.J. Anderson and P. Nash. Linear programming in infinite-dimensional spaces: theory and applications. Wiley, NY, 1987.

[2] H. Axelsson, Y. Y. Wardi, M. Egerstedt, E. I. Verriest. Gradient descent approach to optimal mode scheduling in hybrid dynamical systems. J. Optim. Theory Appl., 136:167–186, 2008.

[3] A. Barvinok. A Course in Convexity. American Mathematical Society, Providence, NJ, 2002.

[4] S. C. Bengea, R. A. DeCarlo. Optimal control of switching systems. Automatica, 41:11-27, 2005.

[5] A. Ben-Tal, A. Nemirovski. Lectures on modern convex optimization: analysis, algorithms, and engineering applications. SIAM, Philadelphia, PA, 2001.

[6] F. J. Bonnans, P. Martinon, V. Grélard. Bocop - A collection of examples. INRIA technical report RR-8053, 2012.

[7] B. Bonnard, M. Claeys, O. Cots, P. Martinon. Geometric and numerical methods in the contrast imaging problem in nuclear magnetic resonance. To appear in Acta Applicandae Mathematicae, 2014.

[8] M. S. Branicky, V. S. Borkar, S. K. Mitter. A unified framework for hybrid control: model and optimal control theory. IEEE Trans. Autom. Control, 43:31–45, 1998.

[9] A. Bressan, B. Piccoli. Introduction to the mathematical theory of control. Amer. Inst. Math. Sci., Springfield, MO, 2007.

[10] E. J. Candès, C. Fernández-Granda. Towards a mathematical theory of super-resolution. To appear in Communications on Pure and Applied Mathematics, 2014.

[11] C. Cassandras, D. L. Pepyne, Y. Wardi. Optimal control of a class of hybrid systems. IEEE Trans. Autom. Control, 46:398–415, 2001.

[12] F. Clarke. Functional analysis, calculus of variations and optimal control. Springer, London, 2013.

[13] M. Claeys, D. Arzelier, D. Henrion, J. B. Lasserre. Measures and LMIs for Impulsive Nonlinear Optimal Control. IEEE Trans. Autom. Control, 59:1374-1379, 2014.

[14] M. Claeys, D. Arzelier, D. Henrion, J. B. Lasserre. Moment LMI approach to LTV impulsive control. Proc. IEEE Conf. Decision and Control, Florence, Italy, 2013.

[15] M. Claeys, R. Sepulchre. Reconstructing trajectories from the moments of occupation measures. Proc. IEEE Conf. Decision and Control, Los Angeles, California, 2014.

[16] G. S. Deaecto, J. C. Geromel, J. Daafouz. Dynamic output feedback H_∞ control of switched linear systems. Automatica, 47:1713–1720, 2011.

[17] R. A. DeCarlo, M. S. Branicky, S. Pettersson, B. Lennartson. Perspectives and results on the stability and stabilizability of hybrid systems. Proc. of the IEEE, 88:1069-1082, 2000.

[18] H. O. Fattorini. Infinite dimensional optimization and control theory. Cambridge Univ. Press, Cambridge, UK, 1999.

[19] H. Frankowska, F. Rampazzo. Filippov's and Filippov-Wazewski's Theorems on Closed Domains. Journal of Differential Equations, 161(2), 449-478, 2000.

[20] J. C. Geromel, P. Colaneri, P. Bolzern. Dynamic output feedback control of switched linear systems. IEEE Trans. Autom. Control, 53:720-733, 2008.

[21] J.C. Geromel, G. Deaecto, J. Daafouz. Suboptimal switching control consistency analysis for switched linear systems. IEEE Trans. Autom. Control, 58(7):1857-1861, 2013.

[22] S. Hedlund, A. Rantzer. Optimal control of hybrid systems. Proc. IEEE Conf. Decision and Control, Pheonix, Arizona, 1999.

[23] D. Henrion, J. Daafouz, M. Claeys. Optimal switching control design for polynomial systems: an LMI approach. Proc. IEEE Conf. Decision and Control, Florence, Italy, 2013.

[24] D. Henrion, M. Korda. Convex computation of the region of attraction of polynomial control systems. IEEE Trans. Autom. Control, 59(2):297-312, 2014.

[25] D. Henrion, J. B. Lasserre, M. Mevissen. Mean squared error minimization for inverse moment problems. Applied Mathematics and Optimization,70(1):83-110, 2014.

[26] D. Henrion, J. B. Lasserre, J. Löfberg. Gloptipoly 3: Moments, optimization and semidefinite programming. Optim. Methods and Software, 24(4-5):761–779, 2009.

[27] C. Jansson. VSDP: a Matlab software package for verified semidefinite programming. Nonlinear Theory and its Applications, 327–330, 2006.

[28] J. B. Lasserre, D. Henrion, C. Prieur, E. Trélat. Nonlinear optimal control via occupation measures and LMI relaxations. SIAM J. Control Opt., 47:1643-1666, 2008.

[29] J. B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press, London, UK, 2010.

[30] J. B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM Journal on Optimization, 17(3), 822-843, 2006.

[31] D. Liberzon, A. S. Morse. Basic problems in stability and design of switched systems. IEEE Control Systems Mag. 19:59–70, 1999.

[32] D. Liberzon. Switching in systems and control. Birkhaeuser, Basel, 2003.

[33] H. Lin, P. J. Antsaklis. Stability and stabilizability of switched linear systems: A survey of recent results. IEEE Trans. Autom. Control, 54:308–322, 2009.

[34] J. Löfberg. Yalmip : A toolbox for modeling and optimization in Matlab. Proc. IEEE Conf. CACSD, Taipei, Taiwan, 2004.

[35] MOSEK ApS. The MOSEK optimization software. 2014.

[36] B. Piccoli. Hybrid systems and optimal control. Proc. IEEE Conf. Decision and Control, Tampa, Florida, 1998.

[37] P. Riedinger, C. Zanne, F. Kratz. Time optimal control of hybrid systems. Proc. Amer. Control Conf., San Diego, California, 1999.

[38] P. Riedinger, C. Iung, F. Kratz. An optimal control approach for hybrid systems. Europ. J. Control, 9:449–458, 2003.

[39] H.L. Royden and P. Fitzpatrick. Real analysis. Prentice Hall, NJ, 2010.

[40] J. E. Rubio. Control and Optimization: the linear treatment of nonlinear problems. Manchester University Press, UK, 1986.

[41] S. Sager, M. Claeys, F. Messine. Efficient upper and lower bounds for global mixed-integer optimal control. To appear in J. Global Optim., 2014.

[42] C. Seatzu, D. Corona, A. Giua, A. Bemporad. Optimal control of continuous-time switched affine systems. IEEE Trans. Autom. Control, 51:726–741, 2006.

[43] M. S. Shaikh, P. E. Caines. On the hybrid optimal control problem: theory and algorithms. IEEE Trans. Autom. Control, 52:1587–1603, 2007.

[44] R. Shorten, F. Wirth, O. Mason, K. Wulff, C. King, Stability criteria for switched and hybrid systems. SIAM Review, 49:545–592, 2007.

[45] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optimization methods and software, 11(1-4), 625-653, 1999.

[46] Z. Sun, S. S. Ge. Switched linear systems: control and design. Springer, London, 2006.

[47] H. J. Sussmann. A maximum principle for hybrid optimal control problems. Proc. IEEE Conf. Decision and Control, Phoenix, Arizona, 1999.

[48] R. Vasudevan, H. Gonzalez, R. Bajcsy, S. Sastry. Consistent approximations for the optimal control of constrained switched systems—part 2: An implementable algorithm. SIAM Journal on Control and Optimization, 51:4484–4503, 2013.

[49] R. B. Vinter, R. M. Lewis. The equivalence of strong and weak formulations for certain problems in optimal control. SIAM Journal on Control and Optimization, 16(4), 546-570, 1978.

[50] R. B. Vinter. Convex duality and nonlinear optimal control. SIAM journal on control and optimization, 31(2), 518-538, 1993.

[51] X. Xu, P. J. Antsaklis. Results and perspectives on computational methods for optimal control of switched systems. In: O. Maler, A. Pnueli, A. (eds.), HSCC 2003, Lecture Notes Comput. Sci., 2623:540-555, Springer, Heidelberg, 2003.