CONVEX COMPUTATION OF THE MAXIMUM CONTROLLED INVARIANT SET FOR POLYNOMIAL CONTROL SYSTEMS*

MILAN KORDA†, DIDIER HENRION‡, AND COLIN N. JONES†

Abstract. We characterize the maximum controlled invariant (MCI) set for discrete- as well as continuous-time nonlinear dynamical systems as the solution of an infinite-dimensional linear programming problem. For systems with polynomial dynamics and compact semialgebraic state and control constraints, we describe a hierarchy of finite-dimensional linear matrix inequality (LMI) relaxations whose optimal values converge to the volume of the MCI set; dual to these LMI relaxations are sum-of-squares (SOS) problems providing a converging sequence of outer approximations to the MCI set. The approach is simple and readily applicable in the sense that the approximations are the outcome of a single semidefinite program with no additional input apart from the problem description. A number of numerical examples illustrate the approach.

Key words. maximum controlled invariant set, region of attraction, optimal control, nonlinear control, sum of squares, occupation measure, moments, semidefinite programming

AMS subject classifications. 49M, 49L, 90C22

DOI. 10.1137/130914565

1. Introduction. Given a controlled dynamical system described by a differential (continuous-time) or difference (discrete-time) equation, its maximum controlled invariant (MCI) set is the set of all initial states that can be kept within a given constraint set ad infinitum using admissible control inputs. This set goes by many other names in the literature, e.g., viability kernel in viability theory [5], or (A, B)-invariant set in the linear case [13].

Set invariance is a ubiquitous and essential concept in dynamical systems theory as far as both analysis and control synthesis is concerned. In particular, by its very definition, the MCI set determines fundamental limitations of a given control system with respect to constraint satisfaction. In addition, there is a very tight link between invariant sets and (control) Lyapunov functions. Indeed, sublevel sets of a Lyapunov function give rise to invariant sets. Conversely, at least in the linear case, any controlled invariant set gives rise to a control Lyapunov function, and therefore these sets can be readily used to design stabilizing control laws; see, e.g., [9] for a general treatment and, e.g., [17, 26] for applications in model predictive control design.

The problem of (maximum) controlled invariant set computation for discrete-time systems has been a topic of active research for more than four decades. The central tool in this effort has been the contractive algorithm of [7] and its expansive counterpart [18]. For an exhaustive survey and historical remarks see the survey [9] and the book [12].

Both algorithms, although conceptually applicable to any nonlinear system, have been predominantly applied in a linear setting where they boil down to a sequence of linear programs (LPs) and polyhedral projections. Finite termination of this sequence is a subtle problem, and sharp results are available only in the uncontrolled setting where no projections are required [16]; for discussion of finite-termination in the controlled case see [45]. The contractive and expansive algorithms were combined in [17] to design an algorithm terminating in a finite number of iterations and outputting an ϵ-accurate inner approximation of the MCI set (with the accuracy measured by the Hausdorff distance). Another line of research, culminating in [42], exploits the linearity of the system dynamics in a more systematic way and approximates the maximum (or minimum) *robust* controlled invariant set by the Minkowski sum of a parametrized family of sets. Very recently, in continuous time, [30] developed a parallel algorithm for ellipsoidal approximations of the robust MCI set scalable to very high dimensions. Computation of low-complexity polyhedral controlled invariant sets was investigated in [10] and [11].

In the nonlinear case, a common practice is to exploit the tight connection between invariance and Lyapunov functions and seek invariant sets as sublevel sets of a (control) Lyapunov function; see, e.g., [14, 51] and references therein for recent theoretical developments on the related problem of region of attraction (ROA) computation and, e.g., [34] for practical applications of these techniques. This, however, typically leads to nonconvex bilinear optimization problems which are notoriously hard to solve. Therefore, one often has to resort to ad hoc analysis of the specific system at hand, which is typically tractable only in small dimensions; see [46, 47] for concrete examples. Related in spirit is the localization technique of [25] for discrete-time *uncontrolled* systems, also requiring considerable effort in analyzing the system.

Recently, a general approach using a hierarchy of finite-dimensional LPs was used in [6] to design a controller ensuring invariance of a given candidate polyhedral set. In our opinion, although being the current state of the art, this work still suffers from the following drawbacks: (1) the sets obtained are convex polytopes (not general semialgebraic sets, a fact particularly limiting in the nonlinear case where nonconvex MCI sets are common); (2) the geometry of the candidate polytopic set must be given a priori; and (3) there are no convergence guarantees to the MCI set. In this paper, we explicitly address all these points.

Building upon our previous work [19] on the computation of the ROA for polynomial control systems, in this paper we characterize the MCI set for discrete- as well as continuous-time polynomial systems as the solution to an infinite-dimensional linear programming problem in the cone of nonnegative measures. The dual of this problem is an infinite-dimensional LP in the space of continuous functions. Finite-dimensional *relaxations* of the primal LP and finite-dimensional *approximations* of the dual LP turn out to be semidefinite programs (SDPs) also related by duality. The primal relaxations lead to a truncated moment problem, while the dual approximations lead to a sum-of-squares (SOS) problem. Superlevel sets of one of the polynomials appearing in the dual SOS problem then provide outer approximations to the MCI set with guaranteed convergence as the degree of the polynomial tends to infinity.

The main mathematical tool we use are the so-called occupation measures, which allow us to study the time evolution of the whole ensemble of initial conditions (described by a measure) rather than studying trajectories associated to each initial condition separately. The use of measures to study dynamical systems has a very

long tradition: see [44] for probably the first systematic treatment;[1] for a purely discrete-time treatment see [22, Chapter 6]. To the best of the authors' knowledge our paper is the first to use occupation measures for MCI set (approximate) *computation*. The MCI set was previously *characterized* using occupation measures in [15], but there the characterization is rather indirect and not straightforwardly amenable to computation. Apart from the authors' work [19], the related problem of ROA computation was tackled using measures in [52]. There, however, a very different approach was taken, not using *occupation* measures but rather analyzing convergence via discretization of the state-space and propagating the initial distribution by means of a discretized transfer operator. Here, instead, we employ the (discounted) occupation measure, which captures the behavior of the trajectories emanating from the initial distribution over the infinite time horizon. As a result, our approach requires no discretization and, contrary to [52], provides true guarantees (not in an "almost-everywhere" or "coarse" sense) and, more importantly, is applicable in a controlled setting. Closely related to the occupation measures used here is the Rantzer density [43], which was used in [41] to assess the stability of attractor sets of uncontrolled nonlinear systems. That approach, however, does not immediately yield approximations of the MCI set (or the ROA) and applies to uncontrolled systems only.

Similar in spirit to our approach, from the dual viewpoint of optimization over functions, are the Hamilton–Jacobi approaches (e.g., [35, 36]). However, contrary to these methods, our approach does not require state-space discretization and comes with convergence guarantees.

The contribution of our paper with respect to previous work on the topic can be summarized as follows:

- We deal with fully general continuous-time and discrete-time polynomial dynamics under semialgebraic state and control constraints.
- Our approximated MCI set is described by (the intersections of) polynomial superlevel sets, including more restrictive classes (e.g., polytopes and ellipsoids).
- We provide a convex infinite-dimensional linear programming characterization of the MCI set.
- We describe a hierarchy of convex finite-dimensional SDPs to solve the LP with convergence guarantees.
- Our approach is simple and readily applicable in the sense that the approximations are the result of a single SDP with *no additional data* required apart from the problem description.

The contribution with respect to our previous work [19] can be summarized as follows:

- In [19] we compute the ROA, which is a related although different object: it is the set of all of initial conditions that can be steered to a given target set while satisfying state and control constraints. In particular, the MCI set differs from the ROA in the sense that we do not try to hit any target set at a given time but rather try to keep the state within a given set forever. Therefore, we had to adapt our technique to deal explicitly with invariance;
- In [19] we dealt with continuous-time systems only, whereas here we address (with minor modifications) discrete-time systems as well; we choose to describe both the continuous-time and discrete-time setups in parallel precisely

---

[1]In [44], J. E. Rubio used Young measures [50] rather than occupation measures, but the basic idea of "linearizing" a nonlinear problem by going into an infinite-dimensional space of measures is the same.

to underline these common features.

- In [19] we considered ROA on a finite time-horizon, whereas here we characterize the MCI set, which is an inherently infinite-time object. This brought about additional technical difficulties not encountered on a finite time-horizon.

What can be considered a drawback of our approach is the fact that the approximations to the MCI set we obtain are from the outside and therefore not invariant. However, accurate outer approximations provide important information as to the performance limitations of the control system and are of practical interest, e.g., in collision avoidance. Therefore, we believe that our work bears both theoretical and practical value and naturally complements existing inner-approximation techniques.

The paper is organized as follows. The problem to be solved is described in section 2. Occupation measures are introduced in section 3. The infinite-dimensional primal and dual LPs are described in sections 4 and 5, respectively. The finite-dimensional relaxations with convergence results are presented in section 6. Numerical examples are in section 7. A reader interested only in the semialgebraic outer approximations of the MCI set can directly consult the infinite-dimensional dual LPs (8) and (9) and their finite-dimensional approximations (11) and (13) in discrete and continuous time, respectively.

1.1. Notation. Measures are understood as signed Borel measures on a Euclidean space, i.e., as countably additive maps from the Borel sets to the real numbers. From now on all subsets of a Euclidean space we refer to are automatically understood as Borel. The vector space of all signed Borel measures with its support contained in a set X is denoted by $M(X)$. The support (i.e., the smallest closed set whose complement has a zero measure) of a measure μ is denoted by $\operatorname{spt} \mu$. The space of continuous functions on X is denoted by $C(X)$, and likewise the space of once continuously differentiable functions is $C^1(X)$. The indicator function of a set X (i.e., a function equal to one on X and zero otherwise) is denoted by $I_X(\cdot)$. The symbol λ denotes the n-dimensional Lebesgue measure (i.e., the standard n-dimensional volume). The integral of a function v with respect to a measure μ over a set X is denoted by $\int_X v(x) \, d\mu(x)$. Sometimes for conciseness we use the shorter notation $\int v \, d\mu$, omitting the integration variable and also the set over which we integrate if they are obvious from the context. The ring of polynomials in (possibly vector) variables $x_1, \ldots, x_n$ is denoted by $\mathbb{R}[x_1, \ldots, x_n]$.

2. Problem statement. The approach is developed in parallel for discrete and continuous time.

2.1. Discrete time. Consider the discrete-time control system

$$(1) \qquad x_{t+1} = f(x_t, u_t), \quad x_t \in X, \quad u_t \in U, \quad t \in \{0, 1, \ldots\},$$

with a given polynomial vector field f with entries $f_i \in \mathbb{R}[x, u]$, $i = 1, \ldots, n$, and given compact[2] basic semialgebraic state and input constraints

$$x_t \in X := \{x \in \mathbb{R}^n \ : \ g_{X_i}(x) \geq 0, i = 1, 2, \ldots, n_X\},$$
$$u_t \in U := \{u \in \mathbb{R}^m \ : \ g_{U_i}(u) \geq 0, i = 1, 2, \ldots, n_U\}$$

---

[2] The assumption of compactness is crucial for the theoretical developments in the paper. The assumptions that the constraint sets are basic semialgebraic and the mapping $f(\cdot, \cdot)$ polynomial plays no role in the infinite-dimensional considerations and only facilitates the convergence results of the finite-dimensional relaxations. The above is valid for both discrete and continuous time.

with $g_{X\,i} \in \mathbb{R}[x]$, $g_{U\,i} \in \mathbb{R}[u]$.

The maximum controlled invariant (MCI) set is defined as

$$X_I := \Big\{ x_0 \in X \; : \exists \left(\{x_t\}_{t=1}^{\infty}, \{u_t\}_{t=1}^{\infty} \right) \text{ s.t. } x_{t+1} = f(x_t, u_t),$$
$$u_t \in U,\, x_t \in X \,\forall t \in \{0,1,\ldots\} \Big\}.$$

A control sequence $\{u_t\}_{t=0}^{\infty}$ is called *admissible* if $u_t \in U$ for all $t \in \{0,1,\ldots\}$.

In words, the MCI set is the set of all initial states which can be kept inside the constraint set X ad infinitum using admissible control inputs.

2.2. Continuous time. Consider the relaxed continuous-time control system

$$(2) \qquad\qquad \dot{x}(t) \in \operatorname{conv} f(x(t), U), \quad x(t) \in X, \quad t \in [0, \infty),$$

where conv denotes the convex hull, f is a polynomial vector field with entries $f_i \in \mathbb{R}[x, u]$, $i = 1, \ldots, n$, and compact basic semialgebraic state and input constraint sets are defined by

$$X := \{x \in \mathbb{R}^n \; : \; g_{X\,i}(x) \geq 0, i = 1, 2, \ldots, n_X\},$$
$$U := \{u \in \mathbb{R}^m \; : \; g_{U\,i}(u) \geq 0, i = 1, 2, \ldots, n_U\}$$

with $g_{X\,i} \in \mathbb{R}[x]$, $g_{U\,i} \in \mathbb{R}[u]$. The meaning of the convex differential inclusion (2) is as follows: for all time t, the state velocity $\dot{x}(t)$ is constrained to the convex hull of the set $f(x(t), U) := \{f(x(t), u) \; : \; u \in U\} \subset \mathbb{R}^n$. The connection of this convexified (or relaxed) control problem (2) and the classical control problem $\dot{x} = f(x, u)$ is the Filippov–Ważewski theorem [5], which shows that the trajectories of $\dot{x} = f(x, u)$ are dense (in the supremum norm) in the set of trajectories of the convexified inclusion[3] (2). Therefore, from a practical point of view, there is little difference between the two formulations for the purposes of MCI set computation; see section 3.2 and Appendices B and C of [19] for a detailed discussion on this subtle issue. The simplest assumption under which the MCI sets for both systems coincide is $f(x, U)$ being convex for all x, which is particularly true for input-affine systems of the form $\dot{x} = f(x) + g(x)u$ with U convex.

The MCI set is defined as

$$X_I := \Big\{ x_0 \in X \; : \; \exists\, x(\cdot) \text{ s.t. } \dot{x}(t) \in \operatorname{conv} f(x(t), U) \text{ a.e., } x(t) \in X \,\forall\, t \in [0, \infty) \Big\},$$

where $x(\cdot)$ is required to be absolutely continuous and a.e. stands for "almost everywhere" with respect to the Lebesgue measure on $[0, \infty)$.

In words, the MCI set is the set of all initial states for which there exists a trajectory of the convexified inclusion (2) which remains in X ad infinitum.

3. Occupation measures. In this section we introduce the concept of occupation measures, which is the centerpiece of our approach.

---

[3]Note that the set conv $f(x(t), U)$ is closed for every t since f is continuous and U is compact; therefore, there is no need to take closure of the convex hull in order to apply the Filippov–Ważewski theorem.

3.1. Discrete time. Given a discount factor $\alpha \in (0,1)$, an initial condition x_0, and an admissible control sequence $\{u_{t|x_0}\}_{t=0}^{\infty}$ such that the associated state sequence $\{x_{t|x_0}\}_{t=0}^{\infty}$ remains in X for all time, we define the *discounted occupation measure* $\mu(\cdot \mid x_0) \in M(X \times U)$ as

$$\mu(A \times B \mid x_0) := \sum_{t=0}^{\infty} \alpha^t I_{A \times B}(x_{t|x_0}, u_{t|x_0}) \tag{3}$$

for all sets $A \subset X$ and $B \subset U$.

In words, the discounted occupation measure measures the (discounted) number of visits of the state-control pair trajectory $(x(\cdot|x_0), \nu(\cdot|x_0))$ to subsets of $X \times U$. The discounting in the definition of the occupation measure ensures that $\mu(A \times B \mid x_0)$ is always finite; in fact, we have $\mu(X \times U \mid x_0) = (1-\alpha)^{-1}$.

Now suppose that the initial condition is not a single point but an *initial measure*[4] $\mu_0 \in M(X)$ and an admissible control sequence is associated to each initial condition from the support of μ_0 in such a way that the corresponding state sequence remains in X. Then we define the *average discounted occupation measure* $\mu \in M(X \times U)$ as

$$\mu(A \times B) := \int_X \mu(A \times B \mid x_0) \, d\mu_0(x_0).$$

The average discounted occupation measure measures the discounted average number of visits in subsets of $X \times U$ of trajectories starting from the initial distribution μ_0.

Now we derive an equation linking the measures μ_0 and μ. This equation will play a key role in subsequent development and in a sense replaces the dynamics equation (1). To derive this equation, fix an initial condition $x_0 \in X$ and a control sequence $\{u_{t|x_0}\}_{t=0}^{\infty}$ such that the associated state sequence $\{x_{t|x_0}\}_{t=0}^{\infty}$ stays in X. Then for any $v \in C(X)$ we have

$$\int_{X \times U} v(x) \, d\mu(x, u \,|\, x_0) = \sum_{t=0}^{\infty} \alpha^t v(x_{t|x_0}) = v(x_{0|x_0}) + \alpha \sum_{t=0}^{\infty} \alpha^t v(x_{t+1|x_0})$$

$$= v(x_{0|x_0}) + \alpha \sum_{t=0}^{\infty} \alpha^t v(f(x_{t|x_0}, u_{t|x_0}))$$

$$= v(x_{0|x_0}) + \alpha \int_{X \times U} v(f(x, u)) \, d\mu(x, u \,|\, x_0).$$

Integrating with respect to μ_0, we arrive at the sought equation

$$\int_{X \times U} v(x) \, d\mu(x, u) = \int_X v(x) \, d\mu_0(x) + \alpha \int_{X \times U} v(f(x, u)) \, d\mu(x, u) \quad \forall v \in C(X). \tag{4}$$

Note that this is an infinite-dimensional linear equation in variables (μ_0, μ).

The following crucial lemma establishes the connection between the support of any initial measure μ_0 solving (4) and the MCI set X_I.

LEMMA 1. *For any pair of measures (μ_0, μ) satisfying (4) with* spt $\mu_0 \subset X$ *and* spt $\mu \subset U \times X$ *we have* spt $\mu_0 \subset X_I$.

Proof. A detailed proof is in Appendix A. $\square$

---

[4]The initial measure μ_0 can be thought of as the probability distribution of the initial state, although we do not require the mass of μ_0 to be normalized to one.

3.2. Continuous time. Given an initial condition x_0 and a trajectory $x(\cdot\,|\,x_0)$ of the inclusion (2) that remains in X for all $t \geq 0$, there exists an admissible time-varying measure-valued relaxed control $\nu_t(\cdot\,|\,x_0) \in M(U)$, $\nu_t(U\,|\,x_0) = 1$, such that

$$\dot{x}(t) = \int_U f(x(t), u)\, d\nu_t(u\,|\,x_0)$$

a.e. with respect to the Lebesgue measure on $[0, \infty)$. This follows from the definition of the convex hull (in fact, for each t, $\nu_t(\cdot\,|\,x_0)$ can be taken to be a convex combination of finitely many Dirac measures).

Then, given a discount factor $\beta > 0$, we define the *discounted occupation measure* $\mu(\cdot\,|\,x_0) \in M(X \times U)$ as

$$\mu(A \times B \mid x_0) := \int_0^\infty \int_U e^{-\beta t} I_{A \times B}(x(t\,|\,x_0), u)\, d\nu_t(u\,|\,x_0)\, dt$$

for all sets $A \subset X$ and $B \subset U$.

In words, the discounted occupation measure measures the (discounted) time spent by the state-control pair trajectory $(x(\cdot\,|\,x_0), \nu(\cdot\,|\,x_0))$ in subsets of $X \times U$. The discounting in the definition of the occupation measure ensures that $\mu(A \times B \mid x_0)$ is always finite; in fact, we have $\mu(X \times U \mid x_0) = \beta^{-1}$.

Now suppose that the initial condition is not a single point but an *initial measure*[5] $\mu_0 \in M(X)$ and a state trajectory that remains in X along with an admissible relaxed control is associated to each initial condition from the support of μ_0. Then we define the *average discounted occupation measure* $\mu \in M(X \times U)$ as

$$\mu(A \times B) := \int_X \mu(A \times B \mid x_0)\, d\mu_0(x_0).$$

Now we derive an equation linking the measures μ_0 and μ. This equation will play a key role in subsequent development and in a sense replaces the dynamics equation (2). To derive the equation, fix an initial condition $x_0 \in X$ and a trajectory $x(\cdot\,|\,x_0)$ that remains in X with an associated admissible relaxed control $\nu_t(\cdot\,|\,x_0)$. Then for any $v \in C^1(X)$ integration by parts yields

$$
\begin{aligned}
\int_{X \times U} \operatorname{grad} v \cdot f(x, u)\, d\mu(x, u\,|\,x_0) &= \int_0^\infty \int_U e^{-\beta t} \operatorname{grad} v \cdot f(x(t\,|\,x_0), u)\, d\nu_t(u\,|\,x_0)\, dt \\
&= \int_0^\infty e^{-\beta t} \frac{d}{dt} v(x(t\,|\,x_0))\, dt \\
&= \beta \int_0^\infty e^{-\beta t} v(x(t\,|\,x_0))\, dt - v(x(0\,|\,x_0)) \\
&= \beta \int_{X \times U} v(x)\, d\mu(x, u\,|\,x_0) - v(x(0\,|\,x_0)),
\end{aligned}
$$

where the boundary term at infinity vanishes due to discounting and the fact that X is bounded. Integrating with respect to μ_0 then gives the sought equation
(5)
$$\beta \int_{X \times U} v(x)\, d\mu(x, u) = \int_X v(x)\, d\mu_0(x) + \int_{X \times U} \operatorname{grad} v \cdot f(x, u)\, d\mu(x, u) \quad \forall v \in C^1(X).$$

---

[5]The initial measure μ_0 can be thought of as the probability distribution of the initial state, although we do not require the mass of μ_0 to be normalized to one.

Note that this is an infinite-dimensional linear equation in variables (μ_0, μ).

The following crucial lemma establishes the connection between the support of any initial measure satisfying (5) and the MCI set X_I.

LEMMA 2. *For any pair of measures (μ_0, μ) satisfying (5) with* $\mathrm{spt}\,\mu_0 \subset X$ *and* $\mathrm{spt}\,\mu \subset U \times X$ *we have* $\lambda(\mathrm{spt}\,\mu_0) \leq \lambda(X_I)$.

Proof. A detailed proof is in Appendix B.　☐

4. Primal LP. In this section we show how the MCI set computation problem can be cast as an infinite-dimensional linear programming problem in the cone of nonnegative measures. As in [19], the basic idea is to maximize the mass of the initial measure μ_0 subject to the constraint that it be dominated by the Lebesgue measure, that is, $\mu_0 \leq \lambda$. System dynamics is captured by (4) and (5) for discrete and continuous times, respectively; state and input constraints are expressed through constraints on the supports of the initial and occupation measures. The constraint that $\mu_0 \leq \lambda$ can be equivalently rewritten as $\mu_0 + \hat{\mu}_0 = \lambda$ for some nonnegative slack measure $\hat{\mu}_0 \in M(X)$. This constraint is in turn equivalent to $\int_X w(x)\,d\mu_0(x) + \int_X w(x)\,d\hat{\mu}_0(x) = \int_X w(x)\,d\lambda(x)$ for all $w \in C(X)$. These considerations lead to the following primal LPs.

4.1. Discrete time. The primal LP in discrete time reads

(6)
$$
\begin{aligned}
p^* \quad = \quad &\sup \quad \mu_0(X) \\
&\mathrm{s.t.} \quad \int v(x)\,d\mu(x,u) = \int v(x)\,d\mu_0(x) + \alpha \int v(f(x,u))\,d\mu(x,u) \ \forall v \in C(X), \\
&\qquad\quad \int w(x)\,d\mu_0(x) + \int w(x)\,d\hat{\mu}_0(x) = \int w(x)\,d\lambda(x) \qquad\qquad \forall w \in C(X), \\
&\qquad\quad \mu \geq 0, \ \mu_0 \geq 0, \ \hat{\mu}_0 \geq 0, \\
&\qquad\quad \mathrm{spt}\,\mu \subset X \times U, \ \ \mathrm{spt}\,\mu_0 \subset X, \ \ \mathrm{spt}\,\hat{\mu}_0 \subset X,
\end{aligned}
$$

where the supremum is over the vector of measures $(\mu, \mu_0, \hat{\mu}_0) \in M(X \times U) \times M(X) \times M(X)$.

This is an infinite-dimensional LP in the cone of nonnegative Borel measures. The following theorem, which is our main theoretical result, relates an optimal solution of this LP to the MCI set X_I.

THEOREM 1. *The optimal value of linear programming problem* (6) *is equal to the volume of the MCI set X_I, that is, $p^* = \lambda(X_I)$. Moreover, the supremum is attained by the restriction of the Lebesgue measure to the MCI set X_I.*

Proof. The proof follows from Lemma 1 by the same arguments as Theorem 1 in [19]. By definition of the MCI set X_I, for any initial condition $x_0 \in X_I$ there exists an admissible control sequence such that the associated state sequence remains in X. Therefore, for any initial measure $\mu_0 \leq \lambda$ with $\mathrm{spt}\,\mu_0 \subset X_I$ there exist a discounted occupation measure μ with $\mathrm{spt}\,\mu \subset X \times U$ and a slack measure $\hat{\mu}_0$ with $\mathrm{spt}\,\hat{\mu}_0 \subset X$ such that the constraints of (6) are satisfied. One such measure μ_0 is the restriction of the Lebesgue measure to X_I, and therefore $p^* \geq \lambda(X_I)$. The fact $p^* \leq \lambda(X_I)$ follows from Lemma 1.　☐

4.2. Continuous time. The primal LP in continuous time reads

(7)
$$
\begin{aligned}
p^* \quad = \quad &\sup \quad \mu_0(X) \\
&\mathrm{s.t.} \quad \beta \int v(x)\,d\mu(x,u) = \int v(x)\,d\mu_0(x) + \int \mathrm{grad}\,v \cdot f(x,u)\,d\mu(x,u) \ \forall v \in C^1(X), \\
&\qquad\quad \int w(x)\,d\mu_0(x) + \int w(x)\,d\hat{\mu}_0(x) = \int w(x)\,d\lambda(x) \qquad\qquad \forall w \in C(X), \\
&\qquad\quad \mu \geq 0, \ \mu_0 \geq 0, \ \hat{\mu}_0 \geq 0, \\
&\qquad\quad \mathrm{spt}\,\mu \subset X \times U, \ \ \mathrm{spt}\,\mu_0 \subset X, \ \ \mathrm{spt}\,\hat{\mu}_0 \subset X,
\end{aligned}
$$

where the infimum is over the vector of measures $(\mu, \mu_0, \hat{\mu}_0) \in M(X \times U) \times M(X) \times M(X)$.

This is an infinite-dimensional LP in the cone of nonnegative Borel measures. The following theorem, which is our main theoretical result, relates an optimal solution of this LP to the MCI set X_I.

THEOREM 2. *The optimal value of LP problem* (7) *is equal to the volume of the MCI set* X_I, *that is,* $p^* = \lambda(X_I)$. *Moreover, the supremum is attained by the restriction of the Lebesgue measure to the MCI set* X_I.

Proof. The fact that μ_0 is equal to the restriction of the Lebesgue measure to X_I is feasible in (7) (and therefore $p^* \geq \lambda(X_I)$) follows by the same arguments as in discrete time. The fact that $p^* \leq \lambda(X_I)$ follows from Lemma 2.   □

5. Dual LP. In this section we derive LPs dual to the primal LPs (6) and (7). Since the primal LPs are in the space of measures, the dual LPs will be on the space of continuous functions. Superlevel sets of feasible solutions to these LPs then provide outer approximations to the MCI sets, both in discrete and in continuous time. Both duals can be derived by standard infinite-dimensional linear programming duality theory; see [19] for a derivation in a similar setting or [3] for a general theory of infinite-dimensional linear programming.

5.1. Discrete time. The dual LP in discrete time reads

$$
\begin{aligned}
d^* \quad = \quad &\inf \quad \int_X w(x)\, d\lambda(x) \\
&\text{s.t.} \quad \alpha v(f(x,u)) \leq v(x) \quad \forall\, (x,u) \in X \times U, \\
&\qquad\quad w(x) \geq v(x) + 1 \qquad \forall\, x \in X, \\
&\qquad\quad w(x) \geq 0 \qquad\qquad\quad \forall\, x \in X,
\end{aligned}
$$
(8)

where the infimum is over the pair of functions $(v, w) \in C(X) \times C(X)$.

The following key observation shows that the unit superlevel set of any function w feasible in (8) provides an outer approximation to X_I.

LEMMA 3. *Any feasible solution to* (8) *satisfies* $v \geq 0$ *and* $w \geq 1$ *on* X_I.

Proof. Given any $x_0 \in X_I$, there exists a sequence $\{u_t\}_{t=0}^{\infty}$, $u_t \in U$, such that $x_t \in X$ for all t. The first constraint of problem (8) is equivalent to $\alpha v(x_{t+1}) \leq v(x_t)$, $t \in \{0, 1, \ldots\}$. By iterating this inequality we get

$$
v(x_0) \geq \alpha^t v(x_t) \to 0 \quad \text{as} \quad t \to \infty
$$

since $x_t \in X$ and X is bounded. Therefore, $v(x_0) \geq 0$ and $w(x_0) \geq 1$ for all $x_0 \in X_I$.   □

The following theorem is instrumental in proving the convergence results of section 6.

THEOREM 3. *There is no duality gap between primal linear programming problem* (6) *on measures and dual linear programming problem* (8) *on functions in the sense that* $p^* = d^*$.

Proof. The proof follows by the same arguments as does Theorem 2 in [19] using standard infinite-dimensional linear programming duality theory (see, e.g., [3]) and the fact that the feasible set of the primal LP is nonempty and bounded in the metric inducing the weak-* topology on $M(X) \times M(X \times U) \times M(X)$. To see nonemptiness, notice that the vector of measures $(\mu_0, \mu, \hat{\mu}_0) = (0, 0, \lambda)$ is trivially feasible. To see the boundedness, it suffices to evaluate the equality constraints of (6) for $v(x) = w(x) = 1$. This gives $\mu_0(X) + \hat{\mu}_0(X) = \lambda(X) < \infty$ and $\mu(X) = \mu_0(X)/(1 - \alpha)$, which, since $\alpha \in (0, 1)$ and all measures are nonnegative, proves the assertion.   □

5.2. Continuous time. The dual LP in continuous time reads

$$
\begin{aligned}
d^* \;=\; & \inf \; \int_X w(x)\,d\lambda(x) \\
(9) \qquad & \text{s.t.} \quad \operatorname{grad} v \cdot f(x,u) \leq \beta v(x) \qquad \forall\,(x,u) \in X \times U, \\
& \qquad\quad w(x) \geq v(x) + 1 \qquad\qquad\quad\; \forall\, x \in X, \\
& \qquad\quad w(x) \geq 0 \qquad\qquad\qquad\quad\;\; \forall\, x \in X,
\end{aligned}
$$

where the infimum is over the pair of functions $(v, w) \in C^1(X) \times C(X)$.

The following key observation shows that the unit superlevel set of any function w feasible in (9) provides an outer approximation to X_I.

LEMMA 4. *Any feasible solution to* (9) *satisfies* $v \geq 0$ *and* $w \geq 1$ *on* X_I.

Proof. Given any $x_0 \in X_I$, there exists an admissible relaxed control function $\nu_t(\cdot)$, $\nu_t(U) = 1$, such that $x(t) \in X$ for all t. For that $x(t)$ we have $\frac{d}{dt} v(x(t)) = \int_U \operatorname{grad} v \cdot f(x(t), u)\, d\nu_t(u) \leq \int_U \beta v(x(t))\, d\nu_t(u) = \nu_t(U)\beta v(x(t)) = \beta v(x(t))$. Then by Gronwall's inequality $v(x(t)) \leq e^{\beta t} v(x_0)$, and consequently

$$
v(x_0) \geq e^{-\beta t} v(x(t)) \to 0 \quad \text{as} \quad t \to \infty
$$

since $x(t) \in X$ and X is bounded. Therefore, $v(x_0) \geq 0$ and $w(x_0) \geq 1$ for all $x_0 \in X_I$. $\square$

The following theorem is instrumental in proving the convergence results of section 6.

THEOREM 4. *There is no duality gap between primal linear programming problem* (7) *on measures and dual linear programming problem* (9) *on functions in the sense that* $p^* = d^*$.

Proof. The proof follows by the same arguments as does Theorem 2 in [19] using standard infinite-dimensional linear programming duality theory (see, e.g., [3]) and the fact that the feasible set of the primal LP is nonempty and bounded in the metric inducing the weak-* topology on $M(X) \times M(X \times U) \times M(X)$. To see nonemptiness, notice that the vector of measures $(\mu_0, \mu, \hat\mu_0) = (0, 0, \lambda)$ is trivially feasible. To see the boundedness, it suffices to evaluate the equality constraints of (7) for $v(x) = w(x) = 1$. This gives $\mu_0(X) + \hat\mu_0(X) = \lambda(X) < \infty$ and $\mu(X) = \mu_0(X)/\beta$, which, since $\beta > 0$ and all measures are nonnegative, proves the assertion. $\square$

6. SDP relaxations. In this section we present finite-dimensional relaxations/ approximations of the infinite-dimensional primal/dual LPs. Both in continuous and discrete time, the *relaxations* of the primal LPs lead to a truncated moment problem which translates to an SDP that can be solved by freely available software, e.g., SeDuMi [39] or SDPA [48]. Dual to the primal SDP relaxation is an SOS problem that again translates to an SDP problem; this dual SDP problem can be interpreted as an *approximation* of the dual infinite-dimensional LPs in the sense that the space of functions searched over is restricted to polynomials of bounded degrees.

The following discussion, adapted from [27], highlights the main ideas behind the derivation of the finite-dimensional relaxations/approximations. The reader is referred to [19, section 5] or to the comprehensive reference [31] for details. First, since the supports of all measures feasible in (6) and (7) are compact, these measures are uniquely determined by their moments, i.e., by integrals of all monomials (which is a sequence of real numbers when indexed in, e.g., the canonical monomial basis). Therefore, it suffices to restrict the test functions $w(x)$ and $v(x)$ in (6) and (7) to all monomials, reducing the linear equality constraints on measures μ_0, μ, and $\hat\mu_0$ of (6)

and (7) to linear equality constraints on their moments. Next, by the Putinar Positivstellensatz (see [31, 40]), the constraint that the support of a measure be included in a given compact basic semialgebraic set is equivalent (modulo a nonrestrictive algebraic condition; see Assumption 1) to the feasibility of an infinite sequence of linear matrix inequalities (LMIs) involving the so-called moment and localizing matrices, which are linear in the coefficients of the moment sequence. By truncating the moment sequence and taking only the moments corresponding to monomials of total degree less than or equal to $2k$, where $k \in \{1, 2, \ldots\}$ is the relaxation order, we obtain a necessary condition for this truncated moment sequence to be the first part of a moment sequence corresponding to a measure with the desired support.

Notation. In what follows, $\mathbb{R}_k[\cdot]$ denotes the vector space of real multivariate polynomials of total degree less than or equal to k. Further, we let $d_{X\,i} := \lceil \deg(g_{X\,i})/2 \rceil$ and $d_{U\,i} := \lceil \deg(g_{U\,i})/2 \rceil$, where $\deg(\cdot)$ denotes the degree of a polynomial; we also use $\succeq 0$ to denote positive semidefiniteness of a matrix. In the primal SDP relaxations we use the notion of moment and localizing matrices $M_k(\cdot)$ and $M_k(\cdot, \cdot)$, following the notation of [31] or [19, section 6]. Finally, throughout the rest of this section we make the following standard assumption.

ASSUMPTION 1. *One of the polynomials modeling the sets X, respectively, U, is equal to $g_{X\,i}(x) = R_X^2 - \|x\|_2^2$, respectively, $g_{U\,i}(u) = R_U^2 - \|u\|_2^2$, for some $R_X \geq 0$, $R_U \geq 0$.*

This assumption is made completely without loss of generality since redundant ball constraints can always be added to the description of the *compact* sets X and U.

6.1. Discrete time. The primal relaxation of order k in discrete time reads
(10)
$$
\begin{aligned}
p_k^* \quad = \quad &\max \quad (y_0)_0 \\
&\text{s.t.} \quad A_k(y, y_0, \hat{y}_0) = b_k, \\
&\qquad\quad M_k(y) \succeq 0, \qquad\quad M_{k-d_{X\,i}}(g_{X\,i}, y) \succeq 0, \qquad i = 1, 2, \ldots, n_X, \\
&\qquad\qquad\qquad\qquad\qquad\;\; M_{k-d_{U\,i}}(g_{U\,i}, y) \succeq 0, \qquad i = 1, 2, \ldots, n_U, \\
&\qquad\quad M_k(y_0) \succeq 0, \qquad\; M_{k-d_{X\,i}}(g_{X\,i}, y_0) \succeq 0, \qquad i = 1, 2, \ldots, n_X, \\
&\qquad\quad M_k(\hat{y}_0) \succeq 0, \qquad\; M_{k-d_{X\,i}}(g_{X\,i}, \hat{y}_0) \succeq 0, \qquad i = 1, 2, \ldots, n_X,
\end{aligned}
$$

where the maximum is over moment sequences $(y, y_0, \hat{y}_0)$ truncated to degree $2k$ corresponding to measures μ, μ_0, and $\hat{\mu}_0$ in (6). The linear equality constraint captures the two linear equality constraints of (6) with $v(x) \in \mathbb{R}_{2k}[x]$ and $w(x) \in \mathbb{R}_{2k}[x]$ being monomials of total degree less than or equal to $2k$. In (10), a linear objective is minimized subject to linear equality constraints and LMI constraints; therefore, (10) is an SDP.

The dual approximation of order k in discrete time reads
(11)
$$
\begin{aligned}
d_k^* \quad = \quad &\inf \quad \boldsymbol{w}'l \\
&\text{s.t.} \quad v(x) - \alpha v(f(x, u)) = q_0(x, u) + \sum_{i=1}^{n_X} q_i(x, u) g_{X\,i}(x) + \sum_{i=1}^{n_U} r_i(x, u) g_{U\,i}(u), \\
&\qquad\quad w(x) - v(x) - 1 = p_0(x) + \sum_{i=1}^{n_X} p_i(x) g_{X\,i}(x), \\
&\qquad\quad w(x) = s_0(x) + \sum_{i=1}^{n_X} s_i(x) g_{X\,i}(x),
\end{aligned}
$$

where l is the vector of Lebesgue moments over X indexed in the same basis in which the polynomial $w(x)$ with coefficients $\boldsymbol{w}$ is expressed. The infimum (which may not be attained) is over polynomials $v(x) \in \mathbb{R}_{2k}[x]$ and $w \in \mathbb{R}_{2k}[x]$, and polynomial SOS $q_i, p_i, s_i, i = 1, \ldots, n_X$, and $r_i, i = 1, \ldots, n_U$, of appropriate degrees. In (11), a linear objective function is minimized subject to SOS constraints; therefore, (11) is an SOS problem which can be readily cast as an SDP (see, e.g., [31]).

6.2. Continuous time. The primal relaxation of order k in continuous time reads

(12)
$$
\begin{aligned}
p_k^* \quad = \quad \max \quad & (y_0)_0 \\
\text{s.t.} \quad & A_k(y, y_0, \hat{y}_0) = b_k, \\
& M_k(y) \succeq 0, && M_{k-d_{X_i}}(g_{X_i}, y) \succeq 0, && i = 1, 2, \ldots, n_X, \\
& && M_{k-d_{U_i}}(g_{U_i}, y) \succeq 0, && i = 1, 2, \ldots, n_U, \\
& M_k(y_0) \succeq 0, && M_{k-d_{X_i}}(g_{X_i}, y_0) \succeq 0, && i = 1, 2, \ldots, n_X, \\
& M_k(\hat{y}_0) \succeq 0, && M_{k-d_{X_i}}(g_{X_i}, \hat{y}_0) \succeq 0, && i = 1, 2, \ldots, n_X,
\end{aligned}
$$

where the maximum is over moment sequences $(y, y_0, \hat{y}_0)$ truncated to degree $2k$ corresponding to measures μ, μ_0, and $\hat{\mu}_0$ in (7). The linear equality constraint captures the two linear equality constraints of (7) with $v(x) \in \mathbb{R}_{2k}[x]$ and $w(x) \in \mathbb{R}_{2k}[x]$ being monomials of total degree less than or equal to $2k$. In (12), a linear objective is minimized subject to linear equality constraints and LMI constraints; therefore, (12) is an SDP.

The dual approximation of order k in continuous time reads

(13)
$$
\begin{aligned}
d_k^* \quad = \quad \inf \quad & \boldsymbol{w}'l \\
\text{s.t.} \quad & \beta v(x) - \operatorname{grad} v \cdot f(x, u) = q_0(x, u) + \sum_{i=1}^{n_X} q_i(x, u) g_{X_i}(x) + \sum_{i=1}^{n_U} r_i(x, u) g_{U_i}(u), \\
& w(x) - v(x) - 1 = p_0(x) + \sum_{i=1}^{n_X} p_i(x) g_{X_i}(x), \\
& w(x) = s_0(x) + \sum_{i=1}^{n_X} s_i(x) g_{X_i}(x),
\end{aligned}
$$

where l is the vector of Lebesgue moments over X indexed in the same basis in which the polynomial $w(x)$ with coefficients $\boldsymbol{w}$ is expressed. The infimum (which may not be attained) is over polynomials $v(x) \in \mathbb{R}_{2k}[x]$ and $w \in \mathbb{R}_{2k}[x]$, and polynomial SOS $q_i, p_i, s_i, i = 1, \ldots, n_X$, and $r_i, i = 1, \ldots, n_U$, of appropriate degrees. In (13), a linear objective function is minimized subject to SOS constraints; therefore, (13) is an SOS problem which can be readily cast as an SDP (see, e.g., [31]).

6.3. Convergence results. In this section we state several convergence results for the finite-dimensional relaxations, respectively, approximations, (10), (12), respectively, (11), (13). Let w_k and v_k denote an optimal solution to the kth dual SDP approximation (11) or (13), and define

$$
X_{Ik} := \{x \in X \; : \; v_k(x) \geq 0\}.
$$

Then, in view of Lemmas 3 and 4, we know that w_k overapproximates the indicator function of the MCI set X_I on X, i.e., $w_k \geq I_{X_I}$ on X, and that the sets X_{Ik} approximate from outside the MCI set X_I, i.e., $X_{Ik} \supset X_I$. In what follows we prove the following:

- The optimal values of the finite-dimensional primal and dual problems p_k^* and d_k^* coincide and converge to the optimal values of the infinite-dimensional primal and dual LPs p^* and d^* which also coincide (in view of Theorems 3 and 4) and are equal to the volume of the MCI set.
- The sequence of functions w_k converges on X from above to the indicator function of the MCI set in the L_1 norm. In addition, the running minimum $\min_{i \leq k} w_i$ converges on X from above to the indicator function of the MCI set in the L_1 norm and almost uniformly.
- The sequence of sets X_{Ik} converges to the MCI set X_I in the sense that the volume discrepancy tends to zero, i.e., $\lim_{k \to \infty} \lambda(X_{Ik} \setminus X_I) = 0$.

The proofs of the results follow very similar reasoning to that of analogous results on ROA approximations in [19, section 6].

LEMMA 5. *There is no duality gap between primal LMI problems* (10) *and* (12) *and dual LMI problems* (11) *and* (13); *i.e.,* $p_k^* = d_k^*$.

Proof. The argument closely follows that in [19, Theorem 4], and therefore we only outline the key points of the proof. To prove the absence of duality gap, it is sufficient to show that the feasible sets of the primal SDPs (10) and (12) are nonempty and compact. The result then follows by standard SDP duality theory (see [19, Theorem 4] for a detailed argument). The nonemptiness follows trivially since the vector of measures $(\mu_0, \mu, \hat{\mu}) = (0, 0, \lambda)$ is feasible in the primal infinite-dimensional LPs (6) and (7), and therefore the truncated moment sequences corresponding to these measures are feasible in the primal SDP relaxations (10) and (12). To see the compactness observe that the first components (i.e., masses) of the truncated moment vectors y_0, y, and $\hat{y}$ are bounded. This follows by evaluating the equality constraints of (6) and (7) for $w(x) = v(x) = 1$. Indeed, in discrete time we get $(y)_0 = (y_0)_0/(1 - \alpha)$, and in continuous time we get $(y)_0 = (y_0)_0/\beta$; in addition, in both cases we have $(y_0)_0 + (\hat{y}_0)_0 = \lambda(X) < \infty$, and therefore the first components are indeed bounded (since they are trivially bounded from below and, in fact, nonnegative, due to the constraints on moment matrices). Boundedness of the even components of each truncated moment vector then follows from the structure of the localizing matrices corresponding to the functions from Assumption 1. Boundedness of the entire truncated moment vectors then follows since the even moments appear on the diagonal of the positive semidefinite moment matrices. ☐

The following result shows the convergence of the optimal values of the finite-dimensional relaxations/approximations to the optimal values of the infinite-dimensional LPs.

THEOREM 5. *The sequence of infima of LMI problems* (11) *and* (13) *converges monotonically from above to the supremum of the linear programming problems* (8) *and* (9), *i.e.,* $d^* \leq d_{k+1}^* \leq d_k^*$ *and* $\lim_{k \to \infty} d_k^* = d^* = p^*$. *Similarly, the sequence of maxima of LMI problems* (10) *and* (12) *converges monotonically from above to the maximum of the linear programming problems* (6) *and* (7), *i.e.,* $p^* \leq p_{k+1}^* \leq p_k^*$ *and* $\lim_{k \to \infty} p_k^* = p^* = d^*$.

Proof. The monotonicity of the optimal values of the relaxations p_k^*, respectively, approximations d_k^*, is evident from the structure of the feasible sets of the corresponding SDPs. The convergence of the primal relaxations p_k to p^* follows from the compactness of the feasible sets of the primal SDPs (10) and (12) (shown in the proof of Lemma 5) by standard arguments on the convergence of Lasserre's LMI hierarchy (see, e.g., [31]). The convergence of the optimal value of the dual approximations d_k^* to d^* then follows from Lemma 5. The equality between p^* and d^* is the statement of Theorems 3 and 4. ☐

The next theorem shows functional convergence from above to the indicator function of the MCI set.

THEOREM 6. *Let $w_k \in \mathbb{R}_{2k}[x]$ denote the w-component of a solution to the dual LMI problems* (11) *and* (13), *and let $\bar{w}_k(x) = \min_{i \leq k} w_i(x)$. Then w_k converges from above to I_{X_I} in the L^1 norm and $\bar{w}_k$ converges from above to I_{X_I} in the L^1 norm and almost uniformly.*

Proof. The convergence in the L_1 norm follows immediately from Theorem 5 and from the fact that $w_k \geq I_{X_I}$ by Lemmas 3 and 4. The convergence of the running minima follows from the fact that there exists a subsequence of $\{w_k\}_{k=0}^{\infty}$

which converges almost uniformly (by, e.g., [4, Theorems 2.5.2 and 2.5.3]). ☐

Our last theorem shows a setwise convergence of the outer approximations to the MCI set.

THEOREM 7. *Let $(v_k, w_k) \in \mathbb{R}_{2k}[x] \times \mathbb{R}_{2k}[x]$ denote an optimal solution to the dual LMI problem* (11) *or* (13), *and let $X_{Ik} := \{x \in \mathbb{R}^n : v_k(x) \geq 0\}$. Then $X_I \subset X_{Ik}$,*

$$\lim_{k \to \infty} \lambda(X_{Ik} \setminus X_I) = 0, \quad and \quad \lambda(\cap_{k=1}^\infty X_{Ik} \setminus X_I) = 0.$$

Proof. From Lemmas 3 and 4 we have $X_{Ik} \supset X_I$ and $w_k \geq I_{X_I}$; therefore, since $w \geq v + 1$ and $w \geq 0$ on X, we have $w_k \geq I_{X_{Ik}} \geq I_{X_I}$ and $\{x : w_k(x) \geq 1\} \supset X_{Ik} \supset X_0$. From Theorem 6, we have $w_k \to I_{X_I}$ in the L^1 norm on X. Consequently,

$$\lambda(X_I) = \int_X I_{X_I} \, d\lambda = \lim_{k \to \infty} \int_X w_k \, d\lambda \geq \lim_{k \to \infty} \int_X I_{X_{Ik}} \, d\lambda$$
$$= \lim_{k \to \infty} \lambda(X_{Ik}) \geq \lim_{k \to \infty} \lambda(\cap_{i=1}^k X_{Ii}) = \lambda(\cap_{k=1}^\infty X_{Ik}).$$

But since $X_I \subset X_{Ik}$ for all k, we must have

$$\lim_{k \to \infty} \lambda(X_{Ik}) = \lambda(X_I) \quad and \quad \lambda(\cap_{k=1}^\infty X_{Ik}) = \lambda(X_I),$$

and the theorem follows. ☐

7. Numerical examples. In this section we present numerical examples that illustrate our results. The primal SDP relaxations were modeled using Gloptipoly 3 [20] and the dual SOS problems using Yalmip [33]. The resulting SDP problems were solved using SeDuMi [39] (which, in the case of primal relaxations, also returns the dual solution providing the outer approximations). For numerical computation (especially for higher relaxation orders), the problem data should be scaled such that the constraint sets are (within) unit boxes or unit balls; for ease of reproduction, most of the numerical problems shown are already scaled. On our problem class we observed only marginal sensitivity to the values of the discrete- and continuous-time discount factors α and β and report results with $\alpha = 0.9$ and $\beta = 1$ for all examples presented.

For a discussion on the scalability of our approach and the performance of alternative SDP solvers see the conclusion and the acrobot-on-a-cart example below.

7.1. Discrete time.

7.1.1. Double integrator. Consider the discrete-time double integrator:

$$x_1^+ = x_1 + 0.1x_2,$$
$$x_2^+ = x_2 + 0.05u$$

with the state constraint set $X = [-1, 1]^2$ and input constraint set $U = [-0.5, 0.5]$. The resulting MCI set outer approximations of degrees 8 and 12 are shown in Figure 1; the approximation is fairly tight for modest degrees. The true MCI set was computed using the standard algorithm based on polyhedral projections [9].

7.1.2. Cathala system. Consider the Cathala system borrowed from [28]:

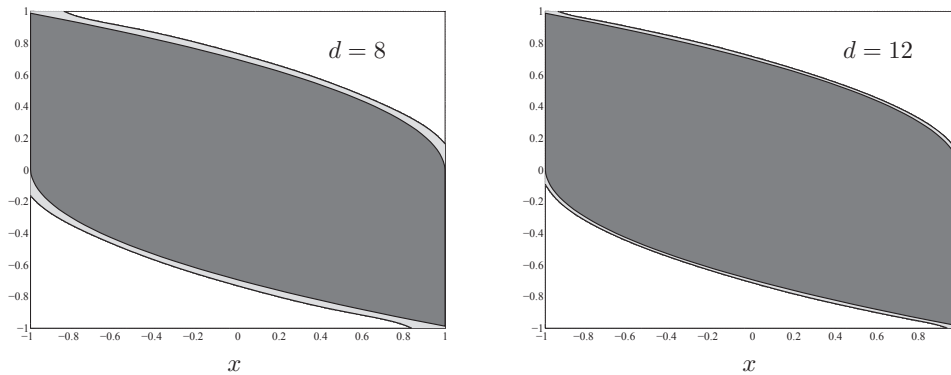$$x_1^+ = x_1 + x_2,$$
$$x_2^+ = -0.5952 + x_2 + x_1^2.$$

FIG. 1. *Discrete time double integrator; polynomial outer approximations (light gray) to the MCI set (dark gray) for degrees $d \in \{8, 12\}$.*

The chaotic attractor of this system is contained in the set $X = [-1.6, 1.6]^2$. MCI set outer approximations are shown in Figure 2; again, the approximations are relatively tight for small relaxation orders. The true MCI set was (approximately) computed by gridding.
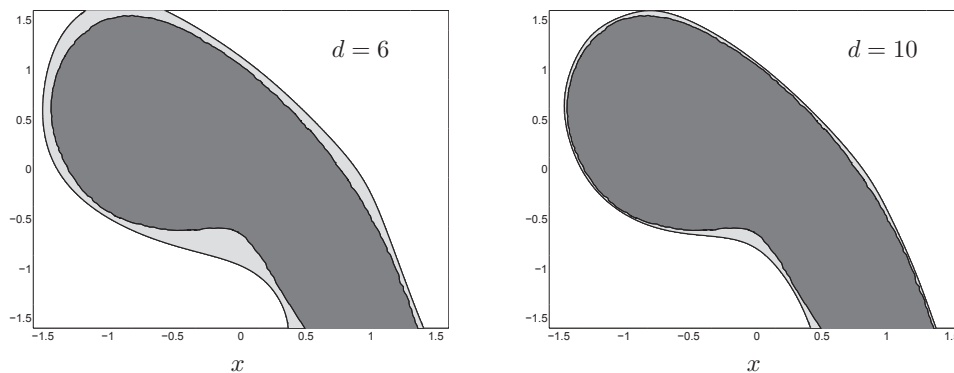

FIG. 2. *Cathala system; polynomial outer approximations (light gray) to the MCI set (dark gray) for degrees $d \in \{6, 10\}$.*

7.1.3. Julia sets. Consider over $z \in \mathbb{C}$, or, equivalently, over $x \in \mathbb{R}^2$, with $z := x_1 + ix_2$, the quadratic recurrence

$$z^+ = z^2 + c$$

with $c \in \mathbb{C}$ a given complex number and i the imaginary unit. The filled Julia set is the set of all initial conditions of the above recurrence for which the trajectories remain bounded. The shape of the Julia set depends strongly on the parameter c. If c lies inside the Mandelbrot set, then the Julia set is connected; otherwise the set is disconnected. In both cases the boundary of the set has a very complicated (in fact, fractal) structure. Here we shall compute outer approximations of the filled Julia set intersected with the unit ball. To this end we set $X = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$. Figure 3 shows outer approximations of degree 12 for parameter values $c = -0.7 + i0.2$ (inside
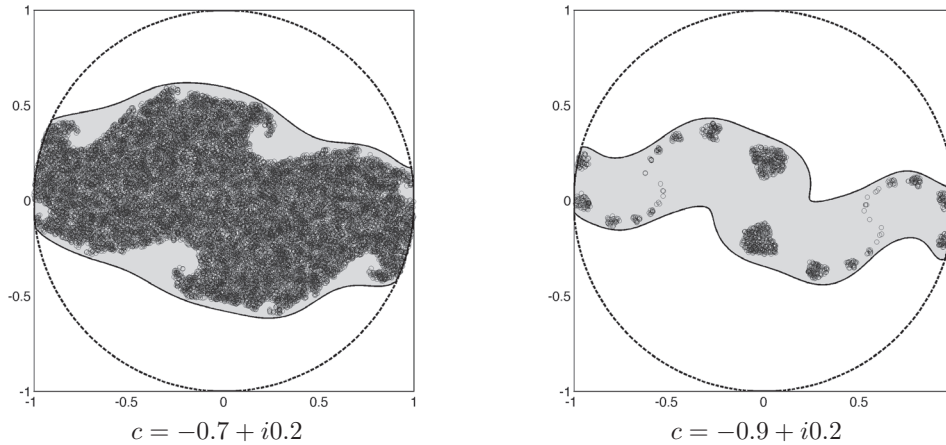
$$c = -0.7 + i0.2 \qquad\qquad c = -0.9 + i0.2$$

Fig. 3. *Filled Julia set; polynomial outer approximation of degree* 12 *(light gray) and (an approximation of) the "true" set (dark grey) represented as an ensemble of initial conditions randomly sampled within the state-constraint set. The dashed line shows the boundary of the unit-ball state-constraint set.*

the Mandelbrot set) and $c = -0.9 + i0.2$ (outside the Mandelbrot set). The "true" filled Julia set was (approximately) obtained by randomly sampling initial conditions within the unit ball and iterating the recurrence for one hundred steps. Taking higher degrees of the approximating polynomials does not result in significant improvement due to our choice of the monomial basis to represent polynomials. An alternative basis (e.g., Chebyshev polynomials; see the related discussions in [21] and [19]) would allow us to further improve the outer estimates and better capture the intricate structure of the filled Julia set's boundary.

7.1.4. Hénon map. Consider the modified controlled Hénon map

$$
\begin{aligned}
x_1^+ &= 0.44 - 0.1x_3 - 4x_2^2 + 0.25u, \\
x_2^+ &= x_1 - 4x_1x_2, \\
x_3^+ &= x_2,
\end{aligned}
$$

adapted from [32] with $X = [-1, 1]^3$ and $U = [-u_{\max}, u_{\max}]$. We investigate two cases: uncontrolled (i.e., $u_{\max} = 0$) and controlled (with $u_{\max} = 1$). Figure 4 shows outer approximations to the MCI set of degree eight for both settings and the "true" MCI set in the uncontrolled setting (approximately) obtained by random sampling of initial conditions inside the constraint set X. The outer approximations suggest that, as expected, allowing for control leads to a larger MCI set.

7.2. Continuous time.

7.2.1. Double integrator. Consider the continuous-time double integrator

$$
\begin{aligned}
\dot{x}_1 &= x_2, \\
\dot{x}_2 &= u,
\end{aligned}
$$

with state constraint set $X = [-1, 1]^2$ and input constraint set $U = [-1, 1]$. The resulting MCI set outer approximations for degrees 8 and 12 are in Figure 5. The
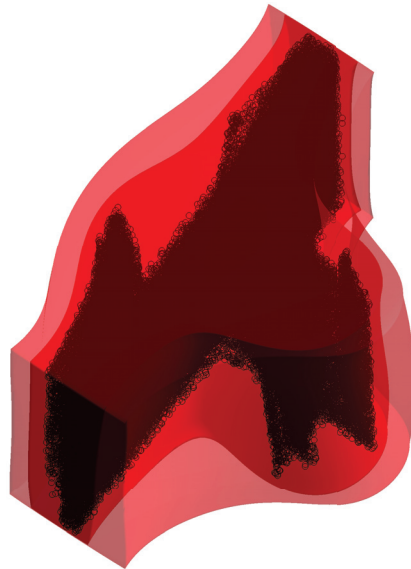
FIG. 4. *Controlled Hénon map; polynomial outer approximation of degree eight in the uncontrolled setting (darker red, smaller) and in the controlled setting (lighter red, larger). The (approximation of) the "true" set (black) in the uncontrolled setting is represented as an ensemble of initial conditions randomly sampled within the state-constraint set.*

approximations are fairly tight even for relatively low relaxation orders. The true MCI set was (approximately) computed as in section 7.1.1 by methods of [9] after dense time discretization.
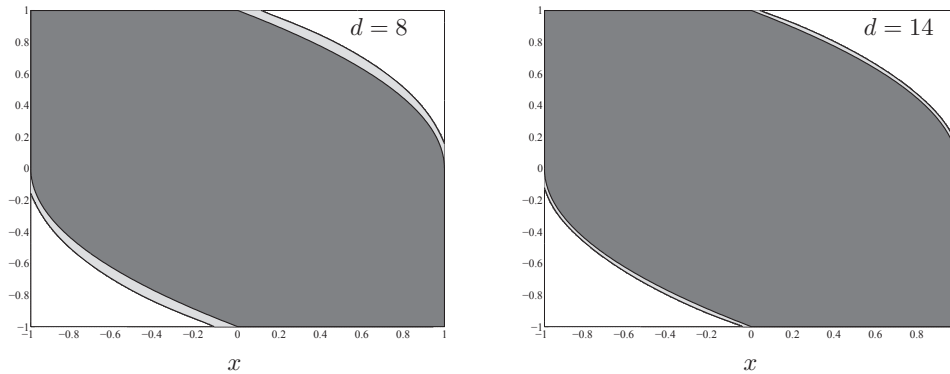

FIG. 5. *Continuous-time double integrator; polynomial outer approximations (light gray) to the MCI set (dark gray) for degrees $d \in \{8, 14\}$.*

7.2.2. Spider-web system. As our second example we take the spider-web system from [1] given by equations

$$\dot{x}_1 = -0.15x_1^7 + 200x_1^6 x_2 - 10.5x_1^5 x_2^2 - 807x_1^4 x_2^3 + 14x_1^3 x_2^4 + 600x_1^2 x_2^5 - 3.5x_1 x_2^6 + 9x_2^7,$$
$$\dot{x}_2 = -9x_1^7 - 3.5x_1^6 x_2 - 600x_1^5 x_2^2 + 14x_1^4 x_2^3 + 807x_1^3 x_2^4 - 10.5x_1^2 x_2^5 - 200x_1 x_2^6 - 0.15x_2^7$$

with the constraint set $X = [-1,1]^2$. Here we exploit the fact that the system dynamics are captured by constraints on v only, whereas w is merely overapproximating $v + 1$, and the fact that outer approximations to the MCI set are given not only by $\{x : v(x) \geq 0\}$ but also by $\{x : w(x) \geq 1\}$. Therefore, if low-complexity outer approximations are desired, it is reasonable to choose different degrees of v and w in (13) (high for v and lower for w) and use the set $\{x : w(x) \geq 1\}$ as the outer approximation. That way, we expect to obtain relatively tight low-order approximations. This is confirmed by numerical results shown in Figure 6. The degree of v is equal to 16 for both figures, whereas $\deg w = 8$ for the left figure and $\deg w = 16$ for the right figure. We observe no significant loss in tightness by choosing a smaller degree of w. The true MCI set was (approximately) computed by gridding.
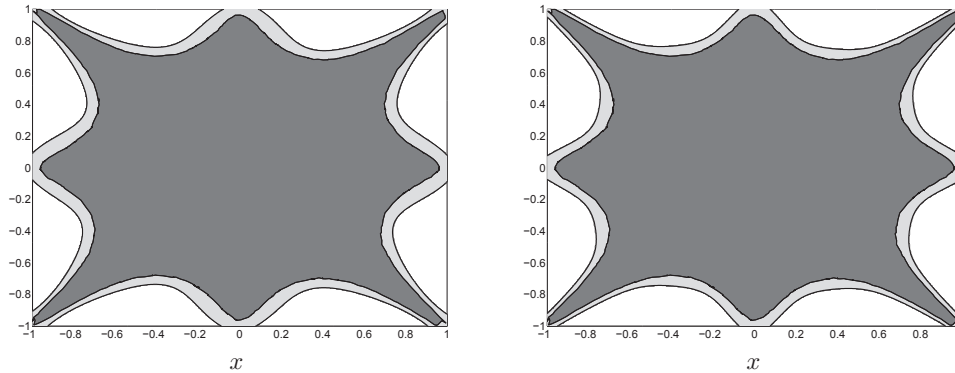

FIG. 6. *Spider-web system; polynomial outer approximations (light gray) to the MCI set (dark gray) for degrees* $\deg v = 16$ *and* $\deg w = 8$ *on the left and* $\deg w = 16$ *on the right.*

7.2.3. Acrobot on a cart. As our last example we consider the acrobot-on-a-cart system adapted from [23], which is essentially a double pendulum on a cart where the inputs are the force acting on the cart and the torque in the middle joint of the double pendulum. The system is sketched in Figure 7. It is a sixth order system with two control inputs; the dynamic equation is given by

$$\dot{x} = \begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ M(x)^{-1}N(x,u) \end{bmatrix} \in \mathbb{R}^6,$$

where

$$M(x) = \begin{bmatrix} a_1 & a_2 \cos x_2 & a_3 \cos x_3 \\ a_2 \cos x_2 & a_4 & a_5 \cos(x_2 - x_3) \\ a_3 \cos x_3 & a_5 \cos(x_2 - x_3) & a_6 \end{bmatrix}$$

and

$$N(x,u) = \begin{bmatrix} u_1 + a_2 x_5^2 \sin x_2 + a_3 x_6^2 \sin x_3 - \delta_0 x_4 \\ -a_5 x_6^2 \sin(x_2 - x_3) + \delta_2 x_6 + a_7 \sin x_2 - x_5(\delta_1 + \delta_2) \\ u_2 + a_5 \sin(x_2 - x_3)x_5^2 + \delta_2 x_5 - \delta_2 x_6 + a_8 \sin x_3 \end{bmatrix}.$$

The states x_1, x_2, x_3 represent, respectively, the position of the cart (in meters), the angle of the lower rod, and the angle of the upper rod of the double pendulum (both in

radians); the states x_4, x_5, and x_6 are then the corresponding velocities in meters per second for the cart and radians per second for the pendulum rods. The constants are given by $a_1 = 0.85$, $a_2 = 0.2063$, $a_3 = 0.0688$, $a_4 = 0.0917$, $a_5 = 0.0344$, $a_6 = 0.0229$, $a_7 = 2.0233$, $a_8 = 0.6744$, $\delta_0 = 0.3$, $\delta_1 = 0.1$, and $\delta_2 = 0.1$. We are interested in computing the maximum controlled invariant subset of the state constraint set

$$X = [-1, 1] \times [-\pi/3, \pi/3] \times [-\pi/3, \pi/3] \times [-0.5, 0.5] \times [-5, 5] \times [-5, 5].$$

We investigate two cases. First, we consider the situation where only the middle joint is actuated and there is no force on the cart; therefore, we impose the constraint $(u_1, u_2) \in U = \{0\} \times [-1, 1]$. Second, we consider the situation where we can also exert a force on the cart; in this case we impose $(u_1, u_2) \in U = [-1, 1] \times [-1, 1]$. Naturally, the MCI set for the second case is larger (or at least the same) as for the first case. This is confirmed[6] by outer approximations of degree four whose section for $x_1 = 0$, $x_4 = 0$, $x_5 = 0$ is shown in Figure 8. In order to compute the outer approximations, we took a third order Taylor expansion of the nonpolynomial dynamics even though exact treatment would be possible via a coordinate transformation leading to rational dynamics to which our methods can be readily extended; this extension is, however, not treated in this paper, and therefore we opted for the simpler (and non-exact) approach using Taylor expansion. Before solving the problem we made a linear coordinate transform so that the state constraint set becomes the unit hypercube $[-1, 1]^6$.

   This example, which is the largest of those considered in this paper, took 110 seconds to solve[7] with SeDuMi for $d = 4$; the corresponding time with the MOSEK SDP solver was 10 seconds. Using MOSEK, we could also solve this example for $d = 6$ (in 420 seconds), although there the solver converged to a solution with a rather poor accuracy,[8] and therefore we do not report the results.
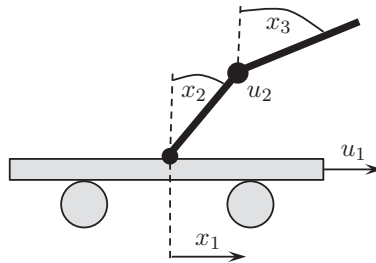

Fig. 7. *Acrobot on a cart; sketch.*

8. Conclusion. We derived an infinite-dimensional convex characterization of the maximum controlled invariant (MCI) set, finite-dimensional approximations of (the dual of) which provide a converging sequence of semialgebraic outer approximations to this set. The outer approximations are the outcome of a single semidefinite

---

[6]There is no a priori guarantee on setwise ordering of the outer approximations; what is guaranteed is the ordering of optimal values of the optimization problem (12) or (13).

[7]All examples were run on an Apple iMac with 3.4 GHz Intel Core i7, 8 GB RAM, and Mac OS X 10.8.2. The time reported is the pure solver time, not including the Yalmip preprocessing time.

[8]Note that the MOSEK SDP solver is still being developed, and its accuracy is likely to improve in the future.
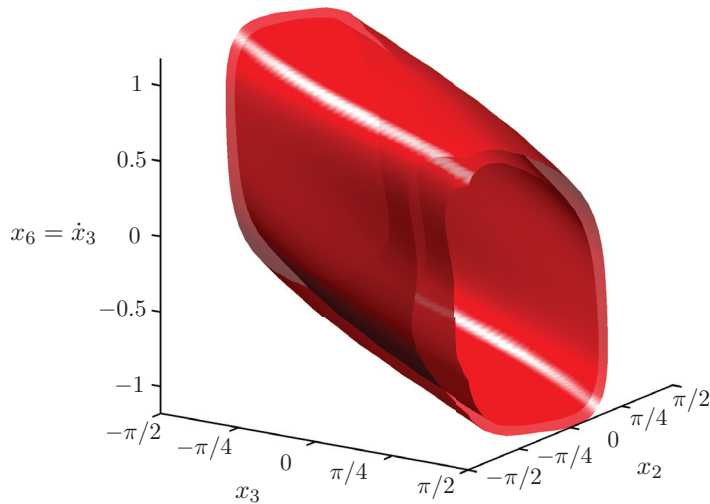
FIG. 8. *Acrobot on a cart; section of the polynomial outer approximations of degree four for* $(x_1, x_4, x_5) = (0, 0, 0)$. *Only the middle joint actuated (darker, smaller); the middle joint and the cart actuated (lighter, larger). The displayed states* x_2, x_3, *and* x_6 *are, respectively, the lower pendulum angle, the upper pendulum angle, and the upper pendulum angular velocity.*

program (SDP) with no additional data required other than the problem description. Therefore, the approach is readily applicable using freely available modeling tools such Gloptipoly 3 [20] or YALMIP [33] with no hand-tuning involved.

The cost to pay for this comfort is the relatively unfavorable scalability of the SDPs solved—the number of variables grows as $O((n + m)^d)$, where n and m are the state and control dimensions and d is the degree of the approximating polynomial. Therefore, in order for this approach to scale to medium dimensions (say, more than $m + n = 6$), one has to either trade off accuracy by taking small d or go beyond the standard freely available solvers such as SeDuMi or SDPA. One possibility is parallelization; for instance, the free parallel solver SDPARA [49] allows for scaling to larger dimensions. Alternatively, one can utilize one of the (few) commercial SDP solvers; in particular, the recently released MOSEK SDP solver seems to show far superior performance on our problem class, and therefore this may allow scaling to larger dimensions (see also the discussion following the acrobot-on-a-cart example in section 7.2.3). In addition, first-order (e.g., DSA-BD [37]) or augmented Lagrangian methods (e.g., SDPNAL [53]) should allow scaling much further provided that proper conditioning of the problem data is ensured; positive results in this direction have been reported in [38] and are currently investigated further by the authors. Finally, customized structure-exploiting solutions can be developed; this is a promising direction of future research also investigated by the authors. At this point it should be emphasized that, to the best of the authors' knowledge, all of the existing approaches providing approximations of similar quality experience similar or worse scalability properties.

Other directions of future research include the extension of the presented approach to inner approximations of MCI sets, to stochastic systems, and to uncertain systems. Partial results on the inner approximations for the related problem of ROA computation already exist [27], albeit in an uncontrolled setting only.

Appendix A. We start by embedding our problem in the setting of discrete-time Markov control processes; terminology and notation is borrowed from the classical reference [22]. Let us define a stochastic kernel on U given X as a map $\nu(\cdot \,|\, \cdot)$ such that $\nu(\cdot \,|\, x)$ is a probability measure on U for all $x \in X$ and $\nu(B \,|\, \cdot)$ is a measurable function on X for all $B \subset U$. Any such stochastic kernel gives rise to a discrete-time Markov process when applied to system (1) as a stationary randomized control policy (a policy which, given x, chooses the control action randomly based on the probability distribution $\nu(\cdot \,|\, x)$, i.e., $\mathrm{Prob}(u \in B \,|\, x) = \nu(B \,|\, x)$ for all $B \subset U$). The transition kernel $Q_\nu(\cdot \,|\, \cdot)$ of this stationary Markov process is then given by

$$Q_\nu(A \,|\, x) = \int_U I_A(f(x,u)) \, d\nu(u \,|\, x) = \mathrm{Prob}(x^+ \in A \,|\, x) \quad \forall A \subset \mathbb{R}^n,$$

where x is the current state and x^+ is the successor state. The t-step transition kernel is then defined by induction as

$$Q_\nu^t(A \,|\, x) := \int_{\mathbb{R}^n} Q(A \,|\, y) \, dQ_\nu^{t-1}(y \,|\, x), \quad t \in \{2, 3, \ldots\},$$

with $Q_\nu^1 := Q_\nu$. Given an initial distribution μ_0, the distribution of the Markov chain at time t, $\tilde{\mu}_t$ is given by

$$\tilde{\mu}_t(A) = \int_X Q_\nu^t(A \,|\, x) \, d\mu_0(x) = \mathrm{Prob}(x_t \in A).$$

The joint distribution of state and control is then

$$\mu_t(A \times B) = \int_A \nu(B \,|\, x) \, d\tilde{\mu}_t(x).$$

The discounted occupation measure associated to the Markov process is defined by

$$\mu(A \times B) = \sum_{t=0}^{\infty} \alpha^t \mu_t(A \times B).$$

Note that this relation reduces to (3) when $\mu_t = \delta_{(x_t, u_t)}$.

In order to prove Lemma 1 we need the following result, which can be found in [22].

LEMMA 6. *For any pair of measures (μ_0, μ) satisfying (4) there exists a stationary randomized control policy $\nu(\cdot \mid x)$ such that the Markov chain obtained by applying this control policy to the difference (1) starting from initial distribution μ_0 has the discounted occupation measure equal to μ.*

Proof. Disintegrate μ as $d\mu(x, u) = d\nu(u \mid x) d\tilde{\mu}(x)$, where $\tilde{\mu}$ denotes the x-marginal of μ and ν is a stochastic kernel on U given X. According to the discussion preceding Lemma 6, applying ν to (1) gives rise to a stationary discrete-time Markov process with the transition kernel Q_ν starting form the initial distribution μ_0.

With this notation, (4) can be equivalently rewritten as

$$(14) \qquad \int_X v(x) \, d\tilde{\mu}(x) = \int_X v(x) \, d\mu_0(x) + \alpha \int_X \int_{\mathbb{R}^n} v(y) \, dQ_\nu(y \,|\, x) \, d\tilde{\mu}(x)$$

for any measurable $v(x)$ (derivation of (4) did not depend on the continuity of v). Taking $v(x) := I_A(x)$, we obtain

$$(15) \qquad \tilde{\mu}(A) = \mu_0(A) + \alpha \int_X Q_\nu(A \,|\, x) \, d\tilde{\mu}(x) \quad \forall A \subset X.$$

Using relation (14) with $v(x) := Q_\nu(A\,|\,x)$ to evaluate the integral with respect to $\tilde{\mu}$ on the right-hand side of (15), we get

$$\tilde{\mu}(A) = \mu_0(A) + \alpha \int_X Q_\nu(A\,|\,x)\,d\mu_0(x) + \alpha^2 \int_X Q_\nu^2(A\,|\,x)\,d\tilde{\mu}(x).$$

By iterating this procedure we obtain

$$(16) \qquad \tilde{\mu}(A) = \mu_0(A) + \sum_{i=1}^{t} \alpha^i \underbrace{\int_X Q_\nu^i(A\,|\,x)\,d\mu_0(x)}_{\mu_i(A)} + \underbrace{\alpha^{t+1}\int_X Q_\nu^{t+1}(A\,|\,x)\,d\tilde{\mu}(x)}_{\to 0},$$

and taking the limit as $t \to \infty$ gives

$$\tilde{\mu}(A) = \sum_{t=0}^{\infty} \alpha^t \tilde{\mu}_t(A),$$

where the third term in (16) converges to zero because $\alpha \in (0,1)$, $Q_\nu^{t+1}(A\,|\,x) \le 1$, and $\tilde{\mu}$ is a finite measure. Hence the x-marginal of the discounted occupation measure of the Markov chain coincides with the x-marginal of μ.

Finally, to establish equality of the whole measures, observe that

$$\sum_{t=0}^{\infty} \alpha^t \mu_t(A \times B) = \sum_{t=0}^{\infty} \alpha^t \int_A \nu(B\,|\,x)\,d\tilde{\mu}_t(x) = \int_A \nu(B\,|\,x)\,d\tilde{\mu}(x) = \mu(A \times B). \qquad \square$$

Proof of Lemma 1. Disintegrate μ to $d\mu(x,u) = d\nu(u\,|\,x)d\tilde{\mu}(x)$ as in the proof of Lemma 6. Then for any $x \in S := \mathrm{spt}\,\tilde{\mu}$ we have

$$\int_U I_S(f(x,u))\,\nu(u\,|\,x) = 1.$$

This relation says that the support of $\tilde{\mu}$ is invariant under ν and follows from Lemma 6, from the definition of the occupation measure μ, from the definition of the support, and from the fact that $\nu(\cdot\,|\,x)$ is a probability measure for all x.

Define an admissible stationary deterministic control policy by taking any measurable selection $u(x) \in \mathrm{spt}\,\nu(\cdot\,|\,x) \subset U$. Define further the sequence of probability measures

$$\nu_n(A\,|\,x) = \frac{\nu(B_{1/n}(u(x)) \cap A\,|\,x)}{\nu(B_{1/n}(u(x)) \cap U\,|\,x)} \quad \forall n \in \{1,2,\ldots\}, \ A \subset U,$$

where $B_{1/n}(u(x))$ is a closed ball of radius $1/n$ centered at $u(x)$. Then $\nu_n(\cdot\,|\,x)$ converges weakly-* (or weakly or narrowly) to $\delta_{u(x)}$ and

$$\int_U I_S(f(x,u))\,\nu_n(u\,|\,x) = 1 \quad \forall n \in \{1,2,\ldots\}.$$

Therefore,

$$1 = \limsup_{n\to\infty} \int_U I_S(f(x,u))\,\nu_n(u\,|\,x) \le \int_U I_S(f(x,u))\,\delta_{u(x)}(u) = I_S(f(x,u(x))),$$

where the inequality follows by the Portmanteau lemma since the set $\{u\,|\,f(x,u) \in S \cap B_{1/n}(u(x))\}$ is closed for all x by continuity of f. Therefore, in fact $I_S(f(x,u(x))) = 1$, and so $f(x,u(x)) \in \mathrm{spt}\,\tilde{\mu}$ for all $x \in \mathrm{spt}\,\tilde{\mu}$. Therefore, $\mathrm{spt}\,\tilde{\mu} \subset X$ is invariant for the closed loop system $x_{t+1} = f(x_t, u(x_t))$, where $u(x)$ is an admissible deterministic control policy. Therefore, necessarily $\mathrm{spt}\,\tilde{\mu} \subset X_I$. Finally, from (4) clearly $\mathrm{spt}\,\mu_0 \subset \mathrm{spt}\,\tilde{\mu}$, and so $\mathrm{spt}\,\mu_0 \subset X_I$. $\quad\square$

Appendix B.

LEMMA 7. *For any pair of measures (μ_0, μ) solving (5), there exists a family of trajectories of the convexified inclusion (2) starting from μ_0 such that the x-marginal of its discounted occupation measure is equal to the x-marginal of μ.*

Proof. The proof is based on fundamental results of [2] and [8] and on the compactification procedure discussed in [29].

We begin by embedding the problem in a stochastic setting. To this end, define the extended state space E as the one-point compactification of $\mathbb{R}^n$, i.e., $E = \mathbb{R}^n \cup \{\Delta\}$, where Δ is the point compactifying $\mathbb{R}^n$. Define also the linear operator $A : \mathcal{D}(A) \to C(E \times U)$ by

$$w \mapsto Aw := \operatorname{grad} w \cdot f,$$

where the domain of A, $\mathcal{D}(A)$ is defined as

$$\mathcal{D}(A) := \{w : E \to \mathbb{R} \mid w \in C^1(\mathbb{R}^n),\ w(\Delta) = 0,\ \lim_{x \to \Delta} w(x) = 0,$$
$$\lim_{x \to \Delta} \operatorname{grad} w \cdot f(x, u) = 0 \ \forall\, u \in U\}.$$

In words, $\mathcal{D}(A)$ is the space all continuously differentiable functions vanishing at infinity such that $\operatorname{grad} w \cdot f$ also vanishes at infinity for all $u \in U$. Now consider the relaxed martingale problem [8]: find a stochastic process $Y : [0, \infty] \times \Omega \to E$ defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ and a stochastic kernel $\nu(\cdot \mid \cdot)$ (stationary relaxed Markov control) on U given E such that

- $P(Y(0) \in A) = \mu_0(A)$ for all $A \subset E$; and
- for all $w \in \mathcal{D}(A)$ the stochastic process

$$(17) \qquad w(Y(t)) - \int_0^t \int_U Aw(Y(\tau), u)\, \nu(du \mid Y(\tau))\, d\tau$$

is an $\mathcal{F}_t$-martingale (see, e.g., [24] for a definition).

Observe that there exists a countable subset of $\mathcal{D}(A)$ (e.g., all polynomials with rational coefficients attenuated near infinity) dense in $\mathcal{D}(A)$ in the supremum norm. Next, $\mathcal{D}(A)$ is clearly an algebra that separates points of E and $A1 = 0$. Finally, since $f(x, u)$ is polynomial and hence locally Lipschitz, the ODE $\dot{x} = f(x, u)$ has a solution on $[0, \infty)$ for any $x_0 \in E$ and any fixed $u \in U$ in the sense that if there is a finite escape time t_e, then we define $x(t) = \Delta$ for all $t \geq t_e$. Each such solution satisfies the martingale relation (17) (with a trivial probability space). Therefore, A satisfies Conditions 1–3 of [8], and it follows from Theorem 2.2 and Corollary 2.2 therein that for any pair of measures satisfying the discounted Liouville equation (5), there exists a solution to the above martingale problem whose discounted occupation measure is equal to μ, that is,

$$\mu(A \times B) = \mathbf{E}\Big\{ \int_0^\infty e^{-\beta t} I_{A \times B}(Y(t), u)\, \nu(du \mid Y(t))\, dt \Big\}, \quad P(Y(0) \in A) = \mu_0(A),$$

where $\mathbf{E}$ denotes the expectation with respect to the probability measure P. From the martingale property of (17) and the definition of A we get

$$\mathbf{E}\{w(Y(t))\} - \mathbf{E}\Big\{ \int_0^t \int_U \operatorname{grad} w \cdot f(Y(\tau), u)\, \nu(du \mid Y(\tau))\, d\tau \Big\} = \mathbf{E}\{Y(0)\}.$$

Now let μ_t denote the marginal distribution of $Y(t)$ at time t; that is,

$$\mu_t(A) := P(Y(t) \in A) = \mathbf{E}\{I_A(Y(t))\} \quad \forall A \subset X.$$

Then the above relation becomes

$$\int_X w(x) \, d\mu_t(x) - \int_0^t \int_X \int_U \mathrm{grad}\, w(x) \cdot f(x,u) \, \nu(du \,|\, x) \, d\mu_\tau(x) \, d\tau = \int w(x) \, d\mu_0(x),$$

where we have used Fubini's thorem to interchange the expectation operator and integration with respect to time. Defining the relaxed vector field

$$\bar{f}(x) = \int_U f(x,u) \, \nu(du \,|\, x) \in \mathrm{conv}\, f(x, U)$$

and rearranging, we obtain

(18) $$\int_X w(x) \, d\mu_t(x) = \int w(x) \, d\mu_0(x) + \int_0^t \int_X \mathrm{grad}\, w(x) \cdot \bar{f}(x) \, d\mu_\tau(x) \, d\tau,$$

where the equation holds for all $w \in C^1(X)$ a.e. with respect to the Lebesgue measure on $[0, \infty)$. The lemma then follows from Ambrosio's superposition principle [2, Theorem 3.2] using the same arguments as in the proof of Lemma 4 in [19]. ☐

Proof of Lemma 2. Suppose that a pair of measures (μ_0, μ) satisfies (5) and that $\lambda(\mathrm{spt}\, \mu_0 \setminus X_I) > 0$. From Lemma 7 there is a family of trajectories of (2) starting from μ_0 with discounted occupation measure whose x-marginal coincides with the x-marginal of μ. However, this is a contradiction since no trajectory starting from $\mathrm{spt}\, \mu_0 \setminus X_I$ remains in X for all times and $\mathrm{spt}\, \mu \subset X$. Thus, $\lambda(\mathrm{spt}\, \mu_0 \setminus X_I) = 0$ and so $\lambda(\mathrm{spt}\, \mu_0) \le \lambda(X_I)$. ☐

REFERENCES

[1] A. A. AHMADI, *Non-monotonic Lyapunov Functions for Stability of Nonlinear and Switched Systems: Theory and Computation*, Master's Thesis, MIT, Cambridge, MA, 2008.

[2] L. AMBROSIO, *Transport equation and Cauchy problem for non-smooth vector fields*, in Calculus of Variations and Nonlinear Partial Differential Equations, L. Ambrosio et al., eds., Lecture Notes in Math. 1927, Springer-Verlag, Berlin, 2008, pp. 1–41.

[3] E. J. ANDERSON AND P. NASH, *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications*, Wiley, New York, 1987.

[4] R. B. ASH, *Real Analysis and Probability*, Academic Press, San Diego, CA, 1972.

[5] J. P. AUBIN AND H. FRANKOWSKA, *Set-Valued Analysis*, Springer-Verlag, Berlin, 1990.

[6] M. A. BEN SASSI AND A. GIRARD, *Controller synthesis for robust invariance of polynomial dynamical systems using linear programming*, Systems Control Lett., 61 (2012), pp. 506–512.

[7] D. BERTSEKAS, *Infinite time reachability of state-space regions by using feedback control*, IEEE Trans. Automat. Control, 17 (1972), pp. 604–613.

[8] A. G. BHATT AND V. S. BORKAR, *Occupation measures for controlled Markov processes: Characterization and optimality*, Ann. Probab., 24 (1996), pp. 1531–1562.

[9] F. BLANCHINI, *Set invariance in control*, Automatica J. IFAC, 35 (1999), pp. 1747–1767.

[10] F. BLANCHINI, S. MIANI, AND C. SAVORGNAN, *Dynamic augmentation and complexity reduction of set-based constrained control*, in Proceedings of the IFAC World Congress on Automatic Control, Seoul, South Korea, 2008.

[11] T. B. Blanco, M. Cannon, and B. De Moor, *On efficient computation of low-complexity controlled invariant sets for uncertain linear systems*, Internat. J. Control, 83 (2010), pp. 1339–1346.

[12] F. Blanchini and S. Miani, *Set-theoretic methods in control*, Birkhäuser Boston, Boston, 2007.

[13] C. E. T. Dorea and J. C. Hennet, *(A, B)-invariant polyhedral sets of linear discrete-time systems*, J. Optim. Theory Appl., 103 (1999), pp. 521–542.

[14] G. Chesi, *Domain of Attraction: Analysis and Control via SOS Programming*, Lecture Notes in Control and Inform. Sci. 415, Springer-Verlag, Berlin, 2011.

[15] V. Gaitsgory and M. Quincampoix, *Linear programming approach to deterministic infinite horizon optimal control problems with discounting*, SIAM J. Control Optim., 48 (2009), pp. 2480–2512.

[16] E. G. Gilbert and K. T. Tan, *Linear systems with state and control constraints: The theory and application of maximal output admissible sets*, IEEE Trans. Automat. Control, 36 (1991), pp. 1008–1020.

[17] R. Gondhalekar, J. Imura, and K. Kashima, *Controlled invariant feasibility—a general approach to enforcing strong feasibility in MPC applied to move-blocking*, Automatica J. IFAC, 45 (2009), pp. 2869–2875.

[18] P. O. Gutman and M. Cwikel, *An algorithm to find maximal state constraint sets for discrete-time linear dynamical systems with bounded controls and states*, IEEE Trans. Automat. Control, 32 (1987), pp. 251–254.

[19] D. Henrion and M. Korda, *Convex computation of the region of attraction of polynomial control systems*, IEEE Trans. Automat. Control, to appear; also available online at http://arxiv.org/abs/1208.1751.

[20] D. Henrion, J. B. Lasserre, and J. Löfberg, *Gloptipoly 3: Moments, optimization and semidefinite programming*, Optim. Methods Softw., 24 (2009), pp. 761–779.

[21] D. Henrion, J. B. Lasserre, and C. Savorgnan, *Approximate volume and integration for basic semialgebraic sets*, SIAM Rev., 51 (2009), pp. 722–743.

[22] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, Berlin, 1996.

[23] S. Jadlovská and A. Jadlovská, *Inverted pendula simulation and modeling—a generalized approach*, in Proceedings of the International Conference on Process Control, Kouty nad Desnou, Czech Republic, 2010.

[24] O. Kallenberg, *Foundations of Modern Probability*, Springer-Verlag, Berlin, 2010.

[25] A. N. Kanatnikov and A. P. Krishchenko, *Localization of compact invariant sets of discrete-time nonlinear systems*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 21 (2011), pp. 2057–2065.

[26] E. C. Kerrigan, *Robust Constraint Satisfaction: Invariant Sets and Predictive Control*, Ph.D. thesis, University of Cambridge, Cambridge, UK, 2000.

[27] M. Korda, D. Henrion, and C. N. Jones, *Inner approximations of the region of attraction for polynomial dynamical systems*, in Proceedings of the 9th IFAC Symposium on Nonlinear Control Systems (NOLCOS), Toulouse, France, 2013.

[28] A. P. Krishchenko and A. N. Kanatnikov, *Maximal compact positively invariant sets of discrete-time nonlinear systems*, in Proceedings of the IFAC World Congress on Automatic Control, Milano, Italy, 2011.

[29] T. G. Kurtz, *Equivalence of stochastic equations and martingale problems*, in Stochastic Analysis 2010, Springer, Heidelberg, 2011, pp. 113–130.

[30] A. N. Daryin and A. B. Kurzhanski, *Parallel algorithm for calculating the invariant sets of high-dimensional linear systems under uncertainty*, Comput. Math. Math. Phys., 53 (2013), pp. 34–43.

[31] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.

[32] M. Liu, S. Zhang, Z. Fan, and M. Qiu, H_∞ *state estimation for discrete-time chaotic systems based on a unified model*, IEEE Trans. Systems Man Cybernet. B, 42 (2012), pp. 1053–1063.

[33] J. Löfberg, *YALMIP: A toolbox for modeling and optimization in MATLAB*, in Proceedings of the IEEE CCA/ISIC/CACSD Conference, Taipei, Taiwan, 2004.

[34] A. Majumdar, A. A. Ahmadi, and R. Tedrake, *Control design along trajectories with sums of squares programming*, in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany, 2013.

[35] K. Margellos and J. Lygeros, *Hamilton-Jacobi formulation for reach-avoid differential games*, IEEE Trans. Automat. Control, 56 (2011), pp. 1849–1861.

[36] I. Mitchell and C. Tomlin, *Overapproximating reachable sets by Hamilton-Jacobi projections*, J. Sci. Comput., 19 (2003), pp. 323–346.

[37] R. D. C. MONTEIRO, C. ORTIZ, AND B. F. SVAITER, *Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems*, Comput. Optim. Appl., 57, pp. 45–69.

[38] J. NIE AND L. WANG, *Regularization methods for SDP relaxations in large-scale polynomial optimization*, SIAM J. Optim., 22 (2012), pp. 408–428.

[39] I. PÓLIK, T. TERLAKY, AND Y. ZINCHENKO, *SeDuMi: A package for conic optimization*, in IMA Workshop on Optimization and Control, University of Minnesota, Minneapolis, MN, 2007.

[40] M. PUTINAR, *Positive polynomials on compact semi-algebraic sets*, Indiana Univ. Math. J., 42 (1993), pp. 969–984.

[41] R. RAJARAMA, U. VAIDYA, M. FARDADC, AND B. GANAPATHYSUBRAMANIAN, *Stability in the almost everywhere sense: A linear transfer operator approach*, J. Math. Anal. Appl., 368 (2010), pp. 144–156.

[42] S. V. RAKOVIĆ, *Parameterized robust control invariant sets for linear systems: Theoretical advances and computational remarks*, IEEE Trans. Automat. Control, 55 (2010), pp. 1599–1614.

[43] A. RANTZER, *A dual to Lyapunov's stability theorem*, Systems Control Lett., 42 (2001), pp. 161–168.

[44] J. E. RUBIO, *Control and Optimization: The Linear Treatment of Nonlinear Problems*, Manchester University Press, Manchester, UK, 1985.

[45] R. VIDAL, S. SCHAERT, J. LYGEROS, AND S. SASTRY, *Controlled invariance of discrete time systems*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 1790, Springer-Verlag, Berlin, 2000, pp. 437–451.

[46] K. STARKOV, *Bounds for compact invariant sets of the system describing dynamics of the nuclear spin generator*, Commun. Nonlinear Sci. Numer. Simul., 14 (2009), pp. 2565–2570.

[47] K. STARKOV, *Estimation of the domain containing all compact invariant sets of the optically injected laser system*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 17 (2007), pp. 4213–4217.

[48] M. YAMASHITA, K. FUJISAWA, M. FUKUDA, K. KOBAYASHI, K. NAKTA, AND M. NAKATA, *Latest developments in the SDPA family for solving large-scale SDPs*, in Handbook on Semidefinite, Conic and Polynomial Optimization, M. Anjos and J. B. Lasserre, eds., Internat. Ser. Oper. Res. Management Sci. 166, Springer, New York, 2012, pp. 687–713.

[49] M. YAMASHITA, K. FUJISAWA, AND M. KOJIMA, *SDPARA: SemiDefinite Programming Algorithm paRAllel version*, Parallel Comput., 29 (2003), pp. 1053–1067.

[50] L. C. YOUNG, *Calculus of Variations and Optimal Control Theory*, Saunders, Philadelphia, 1969.

[51] U. TOPCU, A. K. PACKARD, P. SEILER, AND G. J. BALAS, *Robust region-of-attraction estimation*, IEEE Trans. Automat. Control, 55 (2010), pp. 137–142.

[52] K. WANG AND U. VAIDYA, *Transfer operator approach for computing domain of attraction*, in Proceedings of the IEEE Conference on Decision and Control (CDC), Atlanta, GA, 2010.

[53] X.-Y. ZHAO, D. SUN, AND K.-C. TOH, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM J. Optim, 20 (2010), pp. 1737–1765.