Convex computation of extremal invariant measures of nonlinear dynamical systems and Markov processes

Milan Korda[1], Didier Henrion[2,3,4], Igor Mezić[1]

Draft of July 17, 2018

Abstract

We propose a convex-optimization-based framework for computation of invariant measures of polynomial dynamical systems and Markov processes, in discrete and continuous time. The set of all invariant measures is characterized as the feasible set of an infinite-dimensional linear program (LP). The objective functional of this LP is then used to single-out a specific measure (or a class of measures) extremal with respect to the selected functional such as physical measures, ergodic measures, atomic measures (corresponding to, e.g., periodic orbits) or measures absolutely continuous w.r.t. to a given measure. The infinite-dimensional LP is then approximated using a standard hierarchy of finite-dimensional semidefinite programming problems (SDPs), the solutions of which are truncated moment sequences, which are then used to reconstruct the measure. In particular, we show how to approximate the support of the measure as well as how to construct a sequence of weakly converging absolutely continuous approximations. The presented framework, where a convex functional is minimized or maximized among all invariant measures, can be seen as a generalization of and a computational method to carry out the so called ergodic optimization, where linear functionals are optimized over the set of invariant measures. Finally, we also describe how the presented framework can be adapted to compute eigenmeasures of the Perron-Frobenius operator.

Keywords: Invariant measure, convex optimization, ergodic optimization, physical measure.

1 Introduction

We propose a convex-optimization-based method for approximation of invariant measures. The method is based on the observation that the set of all invariant measures associated

---

[1]University of California, Santa Barbara, `milan.korda@engineering.ucsb.edu`, `mezic@engineering.ucsb.edu`

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

[3]Université de Toulouse; LAAS; F-31400 Toulouse; France.

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

to a deterministic nonlinear dynamical system or a stochastic Markov process is given by the set of solutions to a *linear* equation in the space of Borel measures. The problem of finding an invariant measure can therefore be formulated as the feasibility of an infinite-dimensional linear programming problem (LP). Adding an objective functional to this LP allows one to *target* a particular invariant measure (or a class of invariant measures) such as the physical measure, ergodic measures, absolutely continuous measures, atomic measures etc. The formulation is flexible in the sense that whenever a variational characterization of a given class of measures is known, then it can be used within the proposed framework. The approach is functional analytic in nature, by and large devoid of geometric or topological considerations. The only underlying assumption is that the dynamics is polynomial. This assumption is made for computational convenience even though the approach is far more general, applicable to any algebra of functions closed under function composition (in discrete-time) or differentiation (in continuous-time).

The infinite-dimensional LP in the space of Borel measures is subsequently approximated along the lines of the classical *Lasserre hierarchy* [16] using a sequence of finite-dimensional convex semidefinite programming problems (SDPs). The optimal values of the SDPs are proven to converge from below to the optimal value of the infinite-dimensional LP (for the analysis of the speed of convergence in a related setting of optimal control, see [13]). The outcome of the SDP is an approximate truncated moment sequence of the invariant measure targeted; this sequence is proven to converge weakly to the moment sequence of the target invariant measure, provided this measure is unique (otherwise every accumulation point of the sequence corresponds to an invariant measure).

As a secondary contribution we describe a numerical procedure to approximate the support and density of the invariant measure using the truncated moment sequence obtained from the SDP. For the former, we provide confidence intervals enclosing, in the limit, a prescribed portion of the support; this is achieved using the Christoffel polynomial, an interesting object constructed from the Christoffel-Darboux kernel, which has already been utilized for support approximation in machine learning applications (e.g., [18, 25]). For the latter, we construct a sequence of absolutely continuous measures with polynomial densities converging weakly to the target measure.

This work is a continuation of the movement to apply convex optimization-based techniques to nonconvex problems arising from dynamical systems theory and control. For example, the related problem of invariant set computation was addressed in [11] whereas [17, 4, 12] addressed optimal control; [24], [23] adressed model validation and switching system identification, respectively. The problem addressed here, i.e., invariant measure computation, was also addressed by this approach in [6] in one spatial dimension; this work can therefore be seen as a generalization of [6] to multiple dimensions and with a far more detailed theoretical and computational analysis. In the concurrent work [20] the authors are also applying the Lasserre hierarchy for approximately computing invariant measures for polynomial dynamical systems, but there is no convex functional to be minimized and the focus is on distinguishing measures with different regularity properties (singular vs absolutely continuous).

Let us also mention the optimization-based approaches to invariant measure computation [1, 10]. These approaches are based on non-convex optimization and therefore have to deal with

its inherent difficulties such as the existence of suboptimal local minimizers, saddle points or degeneracy. Therefore, contrary to the proposed convex-optimization based approach, these works do not provide convergence guarantees, despite being built on interesting ideas and showing promising practical performance.

The presented framework, where a convex user-specified functional is minimized among all invariant measures, can be seen as a generalization of and a computational method to carry out the so called *ergodic optimization* [9], where linear functionals are optimized among invariant measures (therefore leading to ergodic measures as the optimizers since these are the extreme points of the set of all invariant measures, hence the name ergodic optimization).

Finally, we also describe a generalization of the proposed approach to compute *eigenmeasures* of the *Perron-Frobenius* operator corresponding to a given (possibly complex) eigenvalue, with the invariant measures being a special case corresponding to eigenvalue one.

The paper is organized as follows. Section 2 formally states the problem of invariant measure computation. Section 3 describes the moment hierarchies and applies them to the invariant measure computation problem. Section 4 describes the reconstruction of the invariant measure from its moments. Section 5 discusses several concrete invariant measures to be targeted via the choice of objective functional to be optimized. Section 6 describes an extension to continuous time systems and Section 7 to Markov processes (both in discrete and continuous time). Section 8 extends the method to eigenmeasures of he Perron-Frobenius operator and Section 9 presents numerical examples.

2　Problem statement

For concreteness we present the approach for deterministic discrete-time dynamical systems. The case of stochastic Markov processes is treated in Section 7; the continuous time cases are treated in Section 6 and 7.1.

Consider therefore a deterministic discrete-time nonlinear dynamical system

$$x^+ = T(x), \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $x^+ \in \mathbb{R}^n$ is the successor state and each of the n components of the mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be a multivariate polynomial.

An invariant measure for the dynamical system (1) is any nonnegative Borel measure μ satisfying the relation

$$\mu(T^{-1}(\mathbf{A})) = \mu(\mathbf{A}) \tag{2}$$

for all Borel measurable $\mathbf{A} \subset \mathbb{R}^{\mathbf{n}}$. In this paper we restrict our attention to invariant measures with support included in some compact set $\boldsymbol{X} \subset \mathbb{R}^n$. With this assumption, the relation (2) reduces to

$$\int_{\boldsymbol{X}} f \circ T \, d\mu = \int_{\boldsymbol{X}} f \, d\mu \tag{3}$$

for all $f \in C(\boldsymbol{X})$.

When a measure μ is supported on a compact set $\boldsymbol{X}$, it follows from the Stone-Weierstrass Theorem that it is entirely characterized by its moment sequence $\boldsymbol{y} := (\boldsymbol{y}_\alpha)_{\alpha \in \mathbb{N}^n} \in \mathbb{R}^\infty$,

where

$$\boldsymbol{y}_\alpha = \int_{\boldsymbol{X}} x^\alpha \, d\mu(x), \tag{4}$$

with $x^\alpha := x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}$ and $\alpha \in \mathbb{N}^n$ running over all n-tuples of nonnegative integers and with $\mathbb{R}^\infty$ denoting the space of all real-valued sequences. In particular, for the choice $f(x) = x^\alpha$, relation (3) becomes

$$\int_{\boldsymbol{X}} T^\alpha(x) \, d\mu(x) = \int_{\boldsymbol{X}} x^\alpha \, d\mu(x) \tag{5}$$

for all $\alpha \in \mathbb{N}^n$, where $T^\alpha(x) := T_1(x)^{\alpha_1} \cdot \ldots \cdot T_n(x)^{\alpha_n}$. Since T is a polynomial, (5) is a *linear* constraint on the moments that can be written as

$$A(\boldsymbol{y}) = 0, \tag{6}$$

where $A : \mathbb{R}^\infty \to \mathbb{R}^\infty$ is a linear operator.

We remark that (6) characterizes all invariant measures associated to (1) with support in $\boldsymbol{X}$. In order to single out one invariant measure of interest we propose to use optimization. In particular, we propose to solve the infinite-dimensional convex optimization problem

$$\begin{aligned} \min_{\boldsymbol{y} \in \boldsymbol{M}(\boldsymbol{X})} \quad & F(\boldsymbol{y}) \\ \text{s.t.} \quad & A(\boldsymbol{y}) = 0 \\ & \boldsymbol{y}_0 = 1 \end{aligned} \tag{7}$$

where the minimization is w.r.t. a sequence $\boldsymbol{y}$ belonging to the convex cone

$$\boldsymbol{M}(\boldsymbol{X}) := \left\{ \boldsymbol{y} \in \mathbb{R}^\infty \ : \ \boldsymbol{y}_\alpha = \int_{\boldsymbol{X}} x^\alpha d\mu(x), \ \alpha \in \mathbb{N}^n, \ \mu \in \mathcal{M}(\boldsymbol{X})_+ \right\}$$

of moments of non-negative Borel measures on $\boldsymbol{X}$, the objective functional $F : \mathbb{R}^\infty \to \mathbb{R}$ is convex, and the constraint $\boldsymbol{y}_0 = 1$ is a normalization constraint enforcing that the measure is a probability measure.

We note that, since T is polynomial and hence continuous and $\boldsymbol{X}$ is compact, the Krylov-Bogolyubov Theorem ensures that there exists at least one invariant measure for (1) and hence the optimization problem (7) is always feasible.

Remark 1 (Role of the objective function) *The role of F is to* target *or* single out a *specific invariant measure from the set of all invariant probability measures characterized by the constraints of (7). In principle, F can be any convex functional that facilities this. In particular, it can be extended-valued (i.e. equal to $+\infty$), therefore encoding any constraints of interest such as μ being absolutely continuous or singular w.r.t. to a given measure. See Section 5 for concrete choices of F.*

A typical example encountered in practice for the choice of F is

$$F(\boldsymbol{y}) = \sum_{|\alpha| \leq d} (\boldsymbol{y}_\alpha - \boldsymbol{z}_\alpha)^2, \tag{8}$$

4

where $(\boldsymbol{z}_\alpha)_{|\alpha|\le d}$ is a given *finite* vector of moments of total degree no more than d. The moments $\boldsymbol{z}_\alpha$ can be estimates of the first few moments of the invariant measure that we wish to compute obtained, e.g., from data or analytically. The optimization problem (7) then seeks among all invariant measures μ the one which minimizes the discrepancy between the first moments $(\boldsymbol{y}_\alpha)_{|\alpha|\le d}$ of μ and the given moments $(\boldsymbol{z}_\alpha)_{|\alpha|\le d}$. Of course, the least-squares criterion in (8) can be replaced by other convex metrics measuring the discrepancy between two truncated moment sequences.

3   Moment hierarchy

In this section we describe a hierarchy of finite dimensional convex optimization problems approximating the infinite-dimensional problem (7) and prove that the solutions obtained from these approximations converge to a solution of (7). For this we assume that the objective function $F(\boldsymbol{y})$ in (7) depends only on finitely many moments $(\boldsymbol{y}_\alpha)_{|\alpha|\le d}$.

The finite-dimensional approximations are derived from the so-called Lasserre hierarchy of approximations to the moment cone. In particular, we use semidefinite programming representable outer approximations to this cone and in addition we truncate the first equality constraint of (7) by imposing the linear constraint (5) only for $f(x) = x^\alpha$, $|\alpha| \le k$, i.e., for all monomials of degree no more than k. By linearity of the constraint, this implies that the constraint is satisfied for all polynomials of degree no more than k. The degree k is called relaxation degree.

Before writing down the finite-dimensional approximation of (7), we first describe the construction of the finite-dimensional outer approximations to $\boldsymbol{M}(\boldsymbol{X})$.

3.1   Finite-dimensional approximations of the moment cone

Here we describe the semidefinite-programming representable outer approximation to $\boldsymbol{M}(\boldsymbol{X})$. For this we assume that the compact set $\boldsymbol{X}$ is of the form[1]

$$\boldsymbol{X} := \{x \in \mathbb{R}^n \mid g_i(x) \ge 0, \ i = 1, \ldots, n_g\} \tag{9}$$

with g_i being multivariate polynomials. Let us denote the unit polynomial by $g_0(x) := 1$.

The outer approximation $\boldsymbol{M}_d(\boldsymbol{X})$ of degree d, d even, is

$$\boldsymbol{M}_d(\boldsymbol{X}) := \{\boldsymbol{y} \in \mathbb{R}^\infty \ : \ M_d(g_i\,\boldsymbol{y}) \succeq 0, \ i = 0, 1, \ldots, n_g\}, \tag{10}$$

where $\cdot \succeq 0$ denotes positive semidefiniteness of a matrix and $M_d(g_i\,\boldsymbol{y})$ are the so-called localizing moment matrices, to be defined below. The convex cone $\boldsymbol{M}_d(\boldsymbol{X})$ is an outer approximation to $\boldsymbol{M}(\boldsymbol{X})$ in the sense that for any non-negative measure μ on $\boldsymbol{X}$ the moment vector $\boldsymbol{y}$ of μ belongs to $\boldsymbol{M}_d(\boldsymbol{X})$.

---

[1]A set of the form (9) is called basic semialgebraic; this class of sets is very rich, including balls, boxes, ellipsoids, discrete sets and various convex and non-convex shapes.

The localizing moment matrices $M_d(g_i \, \boldsymbol{y})$ are defined by

$$M_d(g_i \, \boldsymbol{y}) = l_{\boldsymbol{y}}^d(g_i v_{d_i} v_{d_i}^\top), \tag{11}$$

where $d_i = \lfloor (d - \deg g_i)/2 \rfloor$,

$$v_d(x) := (x^\alpha)_{|\alpha| \le d}$$

and where the Riesz functional $\ell_{\boldsymbol{y}}^d : \mathbb{R}[x]_d \to \mathbb{R}$ is defined for any $f = \sum\limits_{|\alpha| \le d} \boldsymbol{f}_\alpha x^\alpha$ by

$$l_{\boldsymbol{y}}^d(f) = \sum_{|\alpha| \le d} \boldsymbol{f}_\alpha \boldsymbol{y}_\alpha. \tag{12}$$

This functional mimicks integration with respect to a measure; in particular when $\boldsymbol{y} \in \mathbb{R}^{\binom{n+d}{d}}$ is a truncated moment vector of a measure μ, then $l_{\boldsymbol{y}}^d(f) = \int f \, d\mu$ for any $f \in \mathbb{R}[x]_d$. In this case, the localizing matrices $M_d(g_i \, \boldsymbol{y})$ are necessarily positive semidefinite, corresponding to the fact that $\int g_i p^2 \, d\mu \ge 0$ for any polynomial p. Importantly, the following fundamental converse result states that if localizing moment matrices are positive semidefinite *for all* $d \in \mathbb{N}$, then necessarily $\boldsymbol{y}$ is a moment vector of a nonnegative measure supported on $\boldsymbol{X}$. In order for this to hold, the defining polynomials g_i need to satisfy the so-called Archimedian condition:

Assumption 1 *There exist polynomials σ_i and a constant $r \in \mathbb{R}$ such that*

$$r^2 - x^\top x = \sum_{i=0}^{n_g} \sigma_i g_i, \tag{13}$$

where $\sigma_i = \sum_j p_{i,j}^2$ with $p_{i,j}$ polynomial, i.e., each σ_i is a sum of squares of other polynomials.

Assumption 1 is an algebraic certificate of compactness of $\boldsymbol{X}$ because this assumption implies that $\boldsymbol{X} \subset \{x \mid x^\top x \le r^2\}$. Since $\boldsymbol{X}$ is assumed compact, this is a non-restrictive condition because a redundant constraint of the form $g_i = r^2 - x^\top x$ can always be added to the definition of $\boldsymbol{X}$ for a sufficiently large r in which case Assumption 1 is satisfied trivially.

Theorem 1 (Putinar [27]) *Suppose that Assumption 1 holds and that $\boldsymbol{y} \in \boldsymbol{M}_d(\boldsymbol{X})$ for all $d \ge 0$. Then there exists a unique non-negative measure μ on $\boldsymbol{X}$ such that (4) holds for all $\alpha \in \mathbb{N}^n$.*

3.2 Approximation of the infinite-dimensional convex problem

Now we are ready to write down the finite-dimensional approximation to (7). The first step in the approximation is to impose the equality constraint of (7) only for all monomials of total degree no more than k instead of for all continuous functions. That is, we impose,

$$\int_{\boldsymbol{X}} T^\alpha(x) d\mu = \int_{\boldsymbol{X}} x^\alpha \, d\mu = 0, \quad |\alpha| \le k,$$

where $T^\alpha(x) = T_1(x)^{\alpha_1} \cdot \ldots \cdot T_n(x)^{\alpha_n}$. This is a set of k linear equations and since T is polynomial it can be re-written in terms of the truncated moment sequence of the measure μ of degree no more than

$$d_k := k \deg T = k \max_{i=1,\ldots,n} \deg T_i.$$

In terms of the Riesz functional, this set of equalities becomes

$$l_{\boldsymbol{y}}^{d_k}(T^\alpha(x)) = l_{\boldsymbol{y}}^{d_k}(x^\alpha), \quad |\alpha| \leq k. \tag{14}$$

This set of equalities can be re-written in a matrix form as

$$A_k \boldsymbol{y} = 0,$$

for some matrix $A_k \in \mathbb{R}^{\binom{n+k}{k} \times \binom{n+d_k}{d_k}}$ where $\boldsymbol{y}$ is the moment sequence of μ.

The equality constraint $\int_{\boldsymbol{X}} 1 \, d\mu$ translates to $\boldsymbol{y}_0 = 1$. The conic constraint $\boldsymbol{y} \in \boldsymbol{M}(\boldsymbol{X})$ is replaced, according to the previous section, by the constraint $\boldsymbol{y} \in \mathcal{M}_{d_k}(\boldsymbol{X})$. This leads to the following finite-dimensional relaxation of order k of the infinite-dimensional problem (7)

$$
\begin{aligned}
\min_{\boldsymbol{y} \in \mathbb{R}^{\binom{n+d_k}{d_k}}} \quad & F(\boldsymbol{y}) \\
\text{s.t.} \quad & A\boldsymbol{y} = 0 \\
& \boldsymbol{y}_0 = 1 \\
& M_{d_k}(g_i \, \boldsymbol{y}) \succeq 0, \quad \forall \, i = 0, 1, \ldots, n_g.
\end{aligned}
\tag{15}
$$

In optimization problem (7), a convex function is minimized over a convex semidefinite-programming representable set and hence (7) is a convex optimization problem. Provided that the objective functional $F(\boldsymbol{y})$ is also semidefinite programming representable (e.g., it is of the form (8)), then the problem (7) is a semidefinite programming problem and hence can be readily solved by off-the-shelf software (e.g., MOSEK or SeDuMi [28]). Importantly, the finite-dimensional relaxation (15) can be derived from the abstract form (7) and passed to a selected SDP solver automatically with the help of the modelling software Gloptipoly 3 [7] and Yalmip [19].

An immediate observation is that problem (15) is a relaxation of problem (7) in the sense that the moment sequence of any measure feasible in (7) truncated up to degree d_k is feasible in (15). Therefore in particular for any k, the optimal value of (15) provides a lower bound on the optimal value of (7). In the following section we study the convergence of these lower bounds to the optimal value of (7) as as well as convergence of the minimizers of (15) to a minimizer of (7).

3.3 Convergence of approximations

In this section we prove convergence of the finite-dimensional approximations (15) to a solution to the infinite-dimensional optimization problem (7).

Theorem 2 *Suppose that Assumption 1 holds, that the function F is lower semi-continuous[2] and let $\boldsymbol{y}^k \in \mathbb{R}^{\binom{n+d_k}{d_k}}$ denote an optimal solution to (15) and $p^\star$ the optimal value of (7). Then the following holds:*

1. *$\lim_{k\to\infty} F(\boldsymbol{y}^k) = p^\star$*

2. *There exists a subsequence $(k_i)_{i=1}^\infty$ such that $\boldsymbol{y}^{k_i}$ converges pointwise to a moment sequence of an invariant measure $\mu^\star$ attaining the minimum in (7).*

3. *In particular, if there is a unique invariant measure $\mu^\star$ attaining the minimum in (7), then $\boldsymbol{y}^k$ converges to the moment sequence of $\mu^\star$.*

Proof:

The proof follows a standard argument (see, e.g., [15]). By Assumption 1, for every $\bar{\alpha} \geq 0$ there exists a k_0 such that $|\boldsymbol{y}_\alpha| \leq M^{|\alpha|}$ for any vector $\boldsymbol{y}$ satisfying the constraints of (15) for $k \geq k_0$ and for any $\alpha \in \mathbb{N}^n$ satisfying $|\alpha| \leq \bar{\alpha}$, where M is the constant from (13). This statement implies that each component of $\boldsymbol{y}^k$ is bounded for sufficiently large k. To see this, let $f = \sigma_0 + \sum_i \sigma_i g_i$ be the right-hand-side polynomial from (13) and let k_0 be the smallest number $d \geq 1$ such that $d \deg T \geq \deg \sigma_0/2$ and $d \deg T \geq \lfloor (\sigma_i - \deg g_i)/2 \rfloor$. Then necessarily $\ell_{\boldsymbol{y}}^{d_k}(\sigma_0) \geq 0$ and $\ell_{\boldsymbol{y}}^{d_k}(\sigma_i g_i) \geq 0$ for any $k \geq k_0$. Therefore by linearity we have $\ell_{\boldsymbol{y}}^{d_k}(f) \geq 0$. But since $f = r^2 - x^\top x$ by (13) we get $\ell_{\boldsymbol{y}}^{d_k}(r^2) \geq \ell_{\boldsymbol{y}}^{d_k}(x^\top x) = \sum_{i=1}^n \ell_{\boldsymbol{y}}^{d_k}(x_i^2)$. Since $\ell_{\boldsymbol{y}}^{d_k}(r^2) = r^2 \ell_{\boldsymbol{y}}^{d_k}(1) = r^2 \boldsymbol{y}_0$ and $\boldsymbol{y}_0 = 1$ by the second constraint of (15) and since $\ell_{\boldsymbol{y}}^{d_k}(x_i^2) \geq 0$, we conclude that $\ell_{\boldsymbol{y}}^{d_k}(x_i^2) \in [0, r^2]$. Proceeding recursively, applying the same reasoning to $x^\alpha f$ with α_i even for all $\alpha \in \mathbb{N}^n$ satisfying $|\alpha| \leq \bar{\alpha} := \max\{\alpha \in \mathbb{N}^n \mid |\alpha|/2 + (\deg \sigma_i)/2 \leq \lfloor (d_k - \deg g)/2 \rfloor\}$, we conclude that all even moments $\boldsymbol{y}_\alpha$, $|\alpha| \leq \bar{\alpha}$ lie in $[0, r^{2|\alpha|}]$. Since even moments are on the diagonal of the matrix $M_d(\boldsymbol{y}) \succeq 0$ and since the off-diagonal elements of a positive semidefinite matrix are bounded in magnitude by the diagonal elements, the conclusion follows.

Having established that $\limsup_{k\to\infty} |\boldsymbol{y}_\alpha^k| < \infty$ for each $\alpha \in \mathbb{N}^n$, it follows using a standard diagonal argument that we can extract a subsequence $\boldsymbol{y}^{k_i}$ satisfying, for each $\alpha \in \mathbb{N}^n$, $\lim_{i\to\infty} \boldsymbol{y}_\alpha^{k_i} = \boldsymbol{y}_\alpha^\star$ with $\boldsymbol{y}_\alpha^\star \in \mathbb{R}$. To conclude the proof it remains to show that $\boldsymbol{y}^\star$ is a moment sequence of a measure attaining the minimum in (7). Using Theorem 1, it follows that $\boldsymbol{y}^\star$ is a moment sequence of a non-negative measure $\mu^\star$ on $\boldsymbol{X}$ since $M_d(g_i\, \boldsymbol{y}^\star) \succeq 0$ by continuity of the mapping $M \mapsto \lambda_{\min}(M)$, where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of a symmetric matrix (or equivalently by closedness of the cone of positive semidefinite matrices). In addition, $\boldsymbol{y}^\star$ satisfies the equality constraints of (7) by continuity since each row of the matrix A has only finitely many non-zero elements. Therefore $\mu^\star$ is an invariant measure. Finally, since (15) is a relaxation of (7) we have $F(\boldsymbol{y}^{k_i}) \leq p^\star$. By the lower semi-continuity of F we also have $F(\boldsymbol{y}^\star) \leq \lim_{i\to\infty} F(\boldsymbol{y}^{k_i}) \leq p^\star$ and hence necessarily $F(\boldsymbol{y}^\star) = p^\star$ since $\mu^\star$ is feasible in (7) and therefore $F(\boldsymbol{y}^\star) \geq p^\star$. $\qquad\square$

---

[2]More precisely, F is assumed to be lower semi-continuous with respect to the product topology on the space of sequences $\mathbb{R}^\infty$. This is in particular satisfied if F depends only on finitely many moments as, for example, in (8).

4 Reconstruction of measure from moments

In this section we show how the solutions to the finite dimensional relaxations (15) in the form of a truncated moment sequence can be used to approximately reconstruct the invariant measure. In particular we show how to approximate the support of the measure and how to construct a sequence of absolutely continuous measures converging weakly to the invariant measure. In this section, we assume that the optimal invariant measure is unique, which holds generically (in the Baire category sense) by [9, Theorem 3.2].

Assumption 2 (Unique invariant measure) *Convex problem (7) has a unique solution denoted by $\mu^\star$.*

4.1 Approximation of the support

In this section we show how the solutions to the finite dimensional relaxations (15) can be used to approximate the support of the invariant measure $\mu^\star$.

In order to construct the approximations we utilize a certain polynomial constructed for a vector of moments of a given measure. Assume that we are given the sequence of moments $\boldsymbol{y}$ of a non-negative measure μ on $\boldsymbol{X}$. Then from the truncated moments of degree up to d, d even, we define the Christoffel polynomial

$$q_d^{\boldsymbol{y}}(x) = v_{d/2}(x)^\top M_d(\boldsymbol{y})^{-1} v_{d/2}(x), \tag{16}$$

where $v_{d/2}(\cdot)$ is the basis vector of all monomials up to degree $d/2$ with the same ordering as the vector of moments $\boldsymbol{y}$. The polynomial $q_d^{\boldsymbol{y}}$ is well defined as long the moment matrix $M_d(\boldsymbol{y})$ is invertible, which is satisfied if and only if μ is not supported on the zero level set of a polynomial of degree $d/2$ or less.

The sublevel sets of the polynomial $q_d^{\boldsymbol{y}}(x)$ have a remarkable property of approximating the shape of the support of the measure μ. In the real multivariate domain, this was observed recently for empirical measures (sums of Dirac masses) in [25] and subsequently studied analytically for measures with certain regularity properties in [18]. In the complex domain, the theory is far more developed; see, e.g., [5] and references therein.

Here we use the following simple result which holds for arbitrary probability measures μ.

Lemma 1 *Let μ be a probability measure on $\boldsymbol{X}$ with moment sequence $\boldsymbol{y}$, let $\epsilon \in [0,1)$ be given and let $\gamma_\epsilon = \frac{1}{1-\epsilon}\binom{n+d/2}{n}$ and assume that $M_d(\boldsymbol{y})$ is invertible. Then*

$$\mu(\{x : q_d^{\boldsymbol{y}}(x) \leq \gamma_\epsilon\}) \geq \epsilon. \tag{17}$$

Proof: We bound the complementary event:

$$\mu(\{x : q_d^{\boldsymbol{y}}(x) > \gamma_\epsilon\}) = \int_{\{x : q_d^{\boldsymbol{y}}(x) > \gamma_\epsilon\}} 1 \, d\mu = \int_{\{x : \gamma_\epsilon^{-1} q_d^{\boldsymbol{y}}(x) > 1\}} 1 \, d\mu \leq \gamma_\epsilon^{-1} \int_{\boldsymbol{X}} q_d^{\boldsymbol{y}}(x) \, d\mu(x),$$

where we have used the fact that $q_d^{\boldsymbol{y}}$ is nonnegative. Using the definition of $q_d^{\boldsymbol{y}}$ we get

$$\int_{\boldsymbol{X}} q_d^{\boldsymbol{y}}(x)\, d\mu(x) = \int_{\boldsymbol{X}} v_{d/2}(x)^\top M_d(\boldsymbol{y})^{-1} v_{d/2}(x)\, d\mu(x) = \text{trace}\left\{ M_d(\boldsymbol{y})^{-1} \int_{\boldsymbol{X}} v_{d/2}(x) v_{d/2}(x)^\top\, d\mu(x) \right\}$$

$$= \text{trace}\{ M_d(\boldsymbol{y})^{-1} M_d(\boldsymbol{y}) \} = \text{trace}\left\{ \mathbb{I}_{\binom{n+d/2}{n}} \right\} = \binom{n+d/2}{n},$$

where $\mathbb{I}_k$ denotes the identity matrix of size k. Therefore

$$\mu(\{x : q_d^{\boldsymbol{y}}(x) > \gamma_\epsilon\}) \le \frac{1}{\gamma_\epsilon} \binom{n+d/2}{n}$$

and the result follows since $\mu(\{x : q_d^{\boldsymbol{y}}(x) \le \gamma_\epsilon\}) = 1 - \mu(\{x : q_d^{\boldsymbol{y}}(x) > \gamma_\epsilon\})$. $\qquad\square$

Lemma 1 can be readily used to construct approximations to the support of μ in the form

$$\left\{ x : q_d^{\boldsymbol{y}^k}(x) \le \gamma_\epsilon \right\}$$

where $\boldsymbol{y}^k \in \mathbb{R}^{\binom{n+d_k}{d_k}}$ is a solution to the kth order relaxation (15). We remark that $\boldsymbol{y}^k$ only approximates the moments of $\mu^\star$, with guaranteed convergence by Theorem 2, and hence the bound from Lemma 1 may be violated for finite k.

4.2   Weakly converging approximations

In this section we show how the to construct a sequence of absolutely continuous measures (w.r.t. the Lebesgue measure) converging weakly to the invariant measure. This is especially useful if in fact the invariant measure possesses a density with respect to the Lebesgue measure although the approach is general and always provides a sequence of signed measures with polynomial densities that converges weakly to the invariant measure. The idea is simple: given a vector $\boldsymbol{y} \in \mathbb{R}^{\binom{n+d}{n}}$ (e.g., a truncated moment vector of a measure), we can always represent the Riesz functional $\ell_{\boldsymbol{y}}^d : \mathbb{R}[x]_d \to \mathbb{R}$ as

$$\ell_{\boldsymbol{y}}^d(p) = \int_{\boldsymbol{X}} p(x) q(x) dx \tag{18}$$

for some polynomial $q \in \mathbb{R}[x]_d$, provided that the set $\boldsymbol{X}$ has a nonempty interior. Indeed, by linearity it suffices to satisfy (18) for $p(x) = x^\alpha$, $|\alpha| \le d$, which leads to a system of linear equations

$$M_d^L \boldsymbol{q} = \boldsymbol{y}, \tag{19}$$

where $\boldsymbol{q}$ is the coefficient vector of polynomial q in the monomial basis with the same ordering as the vector of moments $\boldsymbol{y}$, and M_d^L is the moment matrix of the Lebesgue measure on $\boldsymbol{X}$ of degree $2d$, i.e.,

$$M_d^L := \int_{\boldsymbol{X}} v_d(x) v_d(x)^\top\, dx.$$

Provided that the interior of $\boldsymbol{X}$ is nonempty, matrix M_d^L is invertible and hence the linear system of equations (19) has a unique solution $\boldsymbol{q} = (M_d^L)^{-1} \boldsymbol{y}$. This approach applied to the solutions of (15) leads to the following result:

Theorem 3 *Let Assumption 2 hold, suppose that the interior of $\boldsymbol{X}$ is nonempty, denote $\boldsymbol{y}^k \in \mathbb{R}^{\binom{n+d_k}{n}}$ any solution to (15), and let*

$$\boldsymbol{q}^k := (M_{d_k}^L)^{-1}\boldsymbol{y}^k. \tag{20}$$

Then the signed measures with densities $q_k(x) := v_k^\top(x)\boldsymbol{q}^k \in \mathbb{R}[x]_k$ with respect to the Lebesgue measure, converge weakly star on $\boldsymbol{X}$ to the invariant measure $\mu^\star$.

Proof: Verifying weak star convergence means that $\lim_{k\to\infty} \int_{\boldsymbol{X}} f q_k \, dx = \int_{\boldsymbol{X}} f \, d\mu^\star$ for all $f \in C(\boldsymbol{X})$. Since $\boldsymbol{X}$ is compact, it is enough to verify this relationship for all f of the form $f = x^\alpha$, $\alpha \in \mathbb{N}^n$, which forms a basis of the space of all polynomials, which is a dense subspace of $C(\boldsymbol{X})$. By construction we have $\int_{\boldsymbol{X}} x^\alpha q_k = \boldsymbol{y}_\alpha^k$ for $k \geq |\alpha|$ and hence

$$\lim_{k\to\infty} \int_{\boldsymbol{X}} x^\alpha q_k(x) \, dx = \lim_{k\to\infty} \boldsymbol{y}_\alpha^k = \boldsymbol{y}_\alpha^\star = \int_{\boldsymbol{X}} x^\alpha \, d\mu^\star(x)$$

by Theorem 2, part 3. $\qquad\square$

Theorem 3 says that the density approximations constructed from the solutions to (15) using (20) converge in the weak star topology to the invariant measure optimal in (7). From a practical point of view, e.g., for the purpose of visualization, we recommend using polynomial densities with coefficients

$$(M_k^L)^{-1}\bar{\boldsymbol{y}}^k, \tag{21}$$

where $\bar{\boldsymbol{y}}^k \in \mathbb{R}^{\binom{n+k}{n}}$ is the vector of the first $\binom{n+k}{n}$ elements of $\boldsymbol{y}^k$, rather than the full vector $\boldsymbol{y}^k \in \mathbb{R}^{\binom{n+d_k}{n}}$. This is because the invariance constraint (14) is imposed only for all monomials up to degree k and hence moments of degrees higher than k are less constrained in (15) and hence are likely to be less accurate approximations to the true moments.

5 Applications - choosing the objective function

In this section we list a several classes of measures that can be targeted through the choice of the objective function F of convex problem (7).

5.1 Physical measures

Here we describe how the proposed methodology can be used to compute the moments of physical measures. Let Assumption 2 hold so that there is a unique physical measure μ with support included in $\boldsymbol{X}$. Therefore for Lebesgue almost every $x \in \boldsymbol{X}$ for which the trajectory of (1) originating from x stays in $\boldsymbol{X}$ we have for any $f \in C(\boldsymbol{X})$

$$\int_{\boldsymbol{X}} f \, d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(T^i(x)).$$

Selecting $f(x) = x^\alpha$, we can approximately compute the moments of μ as

$$\boldsymbol{y}_\alpha^{\mathrm{num}} = \frac{1}{N} \sum_{i=1}^{N} f(T^i(x)) \approx \int_{\boldsymbol{X}} x^\alpha \, d\mu \tag{22}$$

with some $N \gg 1$ and α running over a selected subset of multiindices $\mathcal{I}$. The idea is that the number of moments we (inaccurately) compute using (22) is very small and then we use the optimization problem (15) to compute a much larger number of moments of μ, just from the information contained in $\{\boldsymbol{y}_\alpha^{\mathrm{num}} : \alpha \in \mathcal{I}\}$. This is achieved by setting the objective function in (15) to

$$F(\boldsymbol{y}) = \sum_{\alpha \in \mathcal{I}} (\boldsymbol{y}_\alpha - \boldsymbol{y}_\alpha^{\mathrm{num}})^2 \tag{23}$$

or any other metric measuring the discrepancy among moments.

5.2 Ergodic measures

In this section we describe how one can target ergodic measures through the choice of the objective function F. In particular, these can be use used to locate embedded unstable fixed points or periodic orbits. The starting point is the well-known fact that ergodic measures are precisely the extreme points of the set of all invariant probability measures (see, e.g., [26, Proposition 12.4]; see also [2] for general results on the behavior of extreme points under projection in infinite-dimensional vector spaces). Therefore, we can target ergodic measures by selecting a *linear* objective functional in (7) since the minimum of a linear program is attained at an extreme point, provided the minimizer is unique. This leads to the following immediate result:

Theorem 4 *Let F be a continuous linear functional such that Assumption 2 holds and let $\boldsymbol{y}^k \in \mathbb{R}^{\binom{n+d_k}{d_k}}$ denote the optimal solution to (15). Then $\boldsymbol{y}^k$ converges to a moment sequence of an ergodic measure μ, provided Assumption 1 holds.*

Proof: The result follows from Theorem 2 and the fact that the minimizer is unique, therefore necessarily an extreme point of the feasible set of (7). $\square$

A typical choice of the objective functional is $F(\boldsymbol{y}) = \sum_\alpha c_\alpha \boldsymbol{y}_\alpha$ with c having only finitely many non-zero elements (i.e., F is a linear combination of a finite number of moments).

5.3 Absolutely continuous measures

In this section we describe how to target measures μ absolutely continuous w.r.t. a given measure ν through the choice of the objective function F. In most practical applications the measure ν will be the Lebesgue measure. The absolutely continuity of μ w.r.t. ν is equivalent to the existence of a density ρ such that $\mu = \rho d\nu$. Assuming that $\rho(x) \le \gamma < \infty$ for ν-almost all $x \in X$, we can impose the absolute continuity constraint by choosing

$$F(\mu) = \mathbb{I}_{\{\mu \,|\, \mu \le \gamma\nu\}},$$

12

where $\mathbb{I}_A$ is the extended-value characteristic of a set A (i.e., $\mathbb{I}_A(x) = 0$ if $x \in A$ and $\mathbb{I}_A = +\infty$ if $x \notin A$). Then $F(\boldsymbol{y})$ is a convex function of $\boldsymbol{y}$ and its domain is the set of moment vectors $\boldsymbol{y}$ such that

$$M_d(\gamma \boldsymbol{z} - \boldsymbol{y}) \succeq 0 \quad \forall d \in \mathbb{N}, \tag{24}$$

where $\boldsymbol{z}$ denotes the moment sequence of ν. This is added to the constraints of (7) and (15) (for a particular fixed d in the latter case).

We remark that, in general, there may be multiple absolutely continuous measures in which case the constraint (24) can be combined with an additional choice of the objective functional F in order to target a specific measure (e.g., the physical measure as in Section 5.1).

5.4   Singular measures

In this section we describe how to target singular measures through the choice of the objective function F. In this work, we focus on atomic measures only. Whether there exists and effective variational characterization of the elusive singular-continuous measures (e.g., the Cantor measure) remains an open problem, to the best of our knowledge.

Atomic measures are of interested because they may correspond to unstable periodic orbits or fixed points, which are difficult to obtain using simulation-based techniques. These structures are typically ergodic and therefore the methods of this section can be combined with those of Section 5.2.

We use the observation that if a measure μ consists of K atoms, then the associated moment matrix $M_d(\boldsymbol{y})$ is of rank at most K for all d and of rank exactly K for sufficiently large d. Therefore, in order to seek an invariant measure consisting of K atoms (e.g., a K-period orbit) one would in principle want to choose

$$F(\boldsymbol{y}) = \mathbb{I}_{\{\boldsymbol{y} \,|\, \operatorname{rank} M_d(\boldsymbol{y}) = K\}}.$$

Unfortunately, such F is not convex and therefore we propose to use a convex relaxation

$$F(\boldsymbol{y}) = \mathbb{I}_{\{\boldsymbol{y} \,|\, \operatorname{trace} M_d(\boldsymbol{y}) \leq \gamma\}},$$

where $\gamma \geq 0$ is a regularization parameter. This translates to the constraint

$$\operatorname{trace} M_d(\boldsymbol{y}) \leq \gamma$$

which is added to the constraints of (7) and (15). We note that since $M_d(\boldsymbol{y})$ is positive semidefinite, its trace coincides with its nuclear norm, which is a standard proxy for rank minimization [3] (since the convex hull of the rank is the trace on the unit ball of symmetric matrices).

6   Continuous time version

In this section we briefly outline how the presented approach extends to continuous time. Assume therefore that we are dealing with the dynamical system of the form

$$\dot{x} = b(x), \tag{25}$$

each component of the vector field b is assumed to be a multivariate polynomial. We are seeking a non-negative measure μ on $\boldsymbol{X}$ invariant under the flow of dynamical system (25), which is equivalent to the condition

$$\int_{\boldsymbol{X}} \mathrm{grad} f \cdot b \, d\mu = 0 \tag{26}$$

for all $f \in C^1(\boldsymbol{X})$.

The infinite-dimensional convex optimization problem (7) then becomes

$$\begin{aligned} \min_{\mu \in \mathcal{M}_+(\mathbf{X})} \quad & F(\mu) \\ \text{s.t.} \quad & \int_{\boldsymbol{X}} \mathrm{grad} f \cdot b \, d\mu = 0 \quad \forall f \in C(\mathbf{X}) \\ & \int_{\mathbf{X}} d\mu = 1. \end{aligned} \tag{27}$$

This optimization problem is then approximated by taking $f = x^\alpha$, $\alpha \in \mathbb{N}^n$, and proceeding in exactly the same way as described in Section 3.2, leading to a finite-dimensional relaxation of the same form as (15), with the same convergence results of Theorems 2 and 3.

7 Markov processes

In this section we describe a generalization to Markov processes evolving on the state-space $\mathbf{X}$. We assume a Markov chain in the state-space form[3]

$$x_{k+1} = T(x_k, w_k),$$

where $(w_k)_{k=0}^\infty$ is a sequence of independent identically distributed random variables with values in a given set $\boldsymbol{W}$ and the mapping T is assumed to be a polynomial in (x, w). The distribution of the random variables w_k is denoted by P_w, i.e., for all Borel $\mathbf{A} \subset \mathbf{X}$, $P_w(\mathbf{A})$ is the probability that $w_k \in \mathbf{A}$.

The condition for a probability measure μ to be invariant then reads

$$\int_{\mathbf{X}} \int_{\boldsymbol{W}} f(T(x, w)) \, dP_w(w) \, d\mu(x) = \int_{\mathbf{X}} f(x) \, d\mu(x) \tag{28}$$

for all $f \in C(\mathbf{X})$. This equation is *linear* in μ and can therefore be used in (7) instead of the first equality constraint. The approach then proceeds along the steps of Section 3.2, i.e., we set $f(x) = x^\alpha$ and enforce (28) for all such f with $|\alpha| \leq k$, leading to

$$\int_{\boldsymbol{X}} \int_{\boldsymbol{W}} T^\alpha(x, w) \, dP_w(w) \, d\mu(x) = \int_{\boldsymbol{X}} x^\alpha \, d\mu, \quad |\alpha| \leq k. \tag{29}$$

Since T is a polynomial in (x, w), T^α can be written as

$$T^\alpha(x, w) = \sum_{\beta, \gamma} t_{\alpha, \beta, \gamma} \, x^\beta w^\gamma$$

---

[3]For a relation of this form of a Markov process to the one specified by the transition kernel, see, e.g., [8].

for some coefficients $t_{\alpha,\beta,\gamma}$. Therefore,

$$\int_{\boldsymbol{W}} T^\alpha(x,w) dP_w(w) = \sum_{\beta,\gamma} t_{\alpha,\beta,\gamma}\, x^\beta \int_{\boldsymbol{W}} w^\gamma\, dP_w(w) = \sum_{\beta,\gamma} t_{\alpha,\beta,\gamma}\, x^\beta m_\gamma,$$

where

$$m_\gamma = \int_{\boldsymbol{W}} w^\gamma\, dP_w(w)$$

are the moments of w_k, which are fixed numbers that can be either precomputed analytically or using sampling techniques. The equation (29) can therefore be re-written as

$$\sum_{\beta,\gamma} t_{\alpha,\beta,\gamma} m_\gamma \int_{\boldsymbol{X}} x^\beta d\mu(x) = \int_{\boldsymbol{X}} x^\alpha\, d\mu(x), d\mu, \quad |\alpha| \le k,$$

or, in terms of the moments of μ (4),

$$\sum_{\beta,\gamma} t_{\alpha,\beta,\gamma} m_\gamma \boldsymbol{y}_\beta = \boldsymbol{y}_\alpha, \quad |\alpha| \le k.$$

This is a finite-dimensional system of linear equations of the form

$$A\boldsymbol{y} = 0.$$

Adding the normalization constraint $\boldsymbol{y}_0 = 1$ and the positive-semidefiniteness constraints $M_{d_k}(\boldsymbol{y}) \succeq 0$ and $M_{d_k}(\boldsymbol{y}, g_i) \succeq 0$ leads to the optimization problem (15). The objective functional $F(\boldsymbol{y})$ of (15) is again chosen in order to target a particular class of invariant measures. The same convergence guarantees of Theorems 2 and 3 hold.

Remark 2 (Uniqueness) *It is interesting to note that, in the presence of randomness, it is much more common for a* unique *invariant measure to exist. For example, a sufficient condition for this is the recurrence of the Markov chain; see [21, Chapter 10] for more details.*

7.1 Continuous-time Markov processes

The extension to continuous-time stochastic processes is straightforward. The invariance condition (26) is simply replaced by

$$\int_{\mathbf{X}} \mathcal{A}f\, d\mu = 0, \tag{30}$$

where $\mathcal{A}$ is the infinite-dimensional generator of the process. For concreteness, let us consider the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

where $b : \mathbb{R}^n \to \mathbb{R}^n$ is the drift, $\sigma : \mathbb{R}^n \to \mathbb{R}^{n \times k}$ the diffusion matrix and W_t is a vector-valued Wiener process. Then we have

$$\mathcal{A}f = \sum_i b_i \frac{\partial f}{\partial x_i} + \sum_{i,j} [\sigma\sigma^\top]_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j},$$

which is a polynomial provided that $b(x)$ and $\sigma(x)$ are polynomial in x and we set $f(x) = x^\alpha$. The invariance condition (30) can therefore be expressed solely in terms of the moments of μ (4). The approach then proceeds in exactly the same fashion as described in Section 3.2, leading to a finite-dimensional relaxation of the same form as (15), with the same convergence results of Theorems 2 and 3, provided that the set of invariant measures of the stochastic process considered is non-empty.

8   Eigenmeasures of Perron-Frobenius

In this section we briefly describe how the presented approach can be extended to computation of the eigenmeasures of the Perron-Frobenius operator [22]. For concreteness we work with the discrete-time dynamics (1), although analogous results can be obtained for continuous time, following the developments of Section 6. The Perron-Frobenius operator $\mathcal{P} : \boldsymbol{M}_c(\boldsymbol{X}) \to \boldsymbol{M}_c(\boldsymbol{X})$ is defined by

$$(\mathcal{P}\mu)(A) = \mu(T^{-1}(A))$$

for every Borel measurable set $A \subset \boldsymbol{X}$. Here $M_c(\boldsymbol{X})$ stands for the vector space of all complex-valued measures on $\boldsymbol{X}$. Given $\lambda \in \mathbb{C}$, a complex-valued measure μ is an eigenmeasure of $\mathcal{P}$ if

$$\int f \circ T \, d\mu = \lambda \int f \, d\mu \tag{31}$$

for all $f \in C(\boldsymbol{X})$. Since the moment-based approach developed previously applies to non-negative measures, we use the Jordan decomposition

$$\mu = \mu_R^+ - \mu_R^- + i(\mu_I^+ - \mu_I^-),$$

where $i = \sqrt{-1}$ is the imaginary unit and $\mu_R^+ \in \boldsymbol{M}(\boldsymbol{X})$, $\mu_R^- \in \boldsymbol{M}(\boldsymbol{X})$, $\mu_I^+ \in \boldsymbol{M}(\boldsymbol{X})$, $\mu_I^- \in \boldsymbol{M}(\boldsymbol{X})$. Similarly, we write

$$\lambda = \lambda_R + i\lambda_I.$$

Then the condition (31) is equivalent to

$$\int (f \circ T - \lambda_R f) \, d\mu_R^+ - \int (f \circ T - \lambda_R f) \, d\mu_R^- + \lambda_I \int f \, d\mu_I^- - \lambda_I \int f \, d\mu_I^+ = 0, \tag{32a}$$

$$\int (f \circ T - \lambda_R f) \, d\mu_I^+ - \int (f \circ T - \lambda_R f) \, d\mu_I^- + \lambda_I \int f \, d\mu_R^- - \lambda_I \int f \, d\mu_R^+ = 0 \tag{32b}$$

for all $f \in C(\boldsymbol{X})$. When expressed with $f = x^\alpha$, $\alpha \in \mathbb{N}^n$, this is equivalent to

$$A(\boldsymbol{y}_R^+, \boldsymbol{y}_R^-, \boldsymbol{y}_I^+, \boldsymbol{y}_I^-) = 0,$$

where $\boldsymbol{y}_R^+$ etc are the moment sequences of the respective measures and A is a linear operator. Coupled with the normalization constraint

$$(\boldsymbol{y}_R^+)_0 + (\boldsymbol{y}_R^-)_0 + (\boldsymbol{y}_I^+)_0 + (\boldsymbol{y}_I^-)_0 = 1$$

and and objective functional $F(\boldsymbol{y}_R^+, \boldsymbol{y}_R^-, \boldsymbol{y}_I^+, \boldsymbol{y}_I^-)$, we arrive at an infinite dimensional linear programming problem

$$
\begin{aligned}
\min_{\boldsymbol{y}_R^+, \boldsymbol{y}_R^-, \boldsymbol{y}_I^+, \boldsymbol{y}_I^-} \quad & F(\boldsymbol{y}_R^+, \boldsymbol{y}_R^-, \boldsymbol{y}_I^+, \boldsymbol{y}_I^-) \\
\text{s.t.} \quad & A(\boldsymbol{y}_R^+, \boldsymbol{y}_R^-, \boldsymbol{y}_I^+, \boldsymbol{y}_I^-) = 0 \\
& (\boldsymbol{y}_R^+)_0 + (\boldsymbol{y}_R^-)_0 + (\boldsymbol{y}_I^+)_0 + (\boldsymbol{y}_I^-)_0 = 1 \\
& y_R^+ \in \boldsymbol{M}(\boldsymbol{X}), \; y_R^- \in \boldsymbol{M}(\boldsymbol{X}), \; y_I^+ \in \boldsymbol{M}(\boldsymbol{X}), \; y_I^- \in \boldsymbol{M}(\boldsymbol{X}),
\end{aligned}
\tag{33}
$$

which is then approximated by a sequence of finite-dimensional SDPs in exactly the same fashion as described in Section 3.2, with the convergence results of Theorem 2 also holding in this setting.

9 Numerical examples

9.1 Logistic map

As our first example we consider the Logistic map:

$$
x^+ = 2x^2 - 1
$$

on the set $\boldsymbol{X} = [-1, 1] = \{x \in \mathbb{R} : (x+1)(1-x) \geq 0\}$.

9.1.1 Physical measure

First, we compute the moments of the unique physical measure μ on $\boldsymbol{X}$. For this example, the density of the physical measure is given by [14]

$$
\rho(x) = \frac{1}{\pi} \frac{1}{\sqrt{1 - x^2}}.
$$

We used the first moment $\frac{1}{\pi} \int_{-1}^1 \frac{x}{\sqrt{1-x^2}} \, dx = 0$ as data input for the objective function of the form (23), i.e., we set $F(\boldsymbol{y}) = (\boldsymbol{y}_1 - 0)^2$ (note that instead of an exactly computed value we could have used an imprecise value of the moment from a simulation or determine this value based on symmetry without analytically integrating the density). Then we solve (15) with $k \in \{5, 10, 100\}$ and compute a degree k polynomial approximation to the density using (21); as in [6], for numerical stability reasons, we work in the Chebyshev basis rather than the monomial basis (i.e., we express the constraint (14) using the Chebyshev basis polynomials instead of monomials x^α and replace the monomial basis vectors v_{d_i} by the vectors of Chebyshev polynomials up to degree d_i in (11)); see [6] for more details on the use of Chebyshev polynomials in this context. In Figure 1 we compare the true density and the polynomial approximations. We observe a very good fit even for low-degree approximations and oscillations of the sign of the approximation error, akin to classical results from approximation theory. The computed moments (transformed to the monomial basis) are compared in Table 1; we see a very good match.
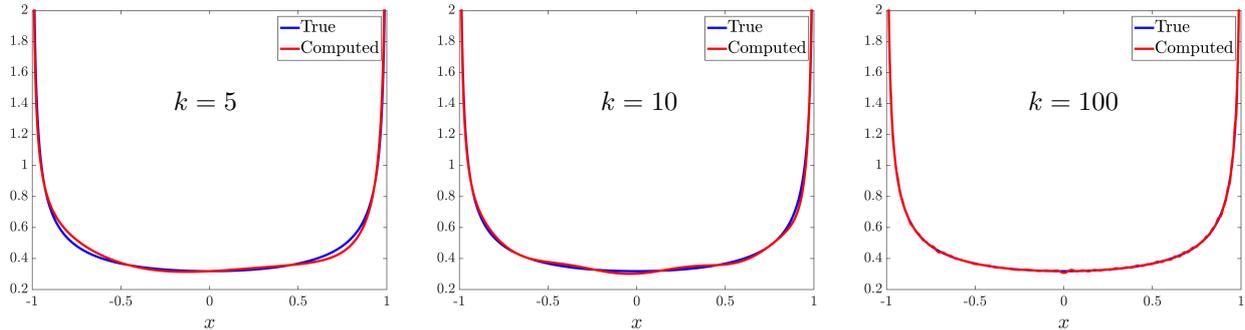
Figure 1: Logistic map: approximation of the density. Only the first moment was given as input to optimization problem (15).

Table 1: Logistic map: comparison of moments computed by the SDP relaxation (15) with the true moments computed analytically. Only the first moment was given as data input to (15).

Moments	x	x^2	x^3	x^4	x^5	x^6	x^7	x^8	x^9	x^{10}
SDP	0.0000	0.5000	0.0047	0.3750	0.0062	0.3131	0.0068	0.2746	0.0071	0.2477
True	0.0000	0.5000	0.0000	0.3750	0.0000	0.3125	0.0000	0.2734	0.0000	0.2461

9.2   Hénon map

Our second example is the Hénon map:

$$x_1^+ = 1 - 1.4x_1^2 + x_2$$
$$x_2^+ = 0.3x_1.$$

The set $\boldsymbol{X}$ is the box $[-1.5, 1.5] \times [-0.4, 0.4]$ expressed as

$$\boldsymbol{X} = \{x \mid (x_1 - 1.5)(1.5 - x_1) \geq 0, \ (x_2 - 0.4)(0.4 - x_2) \geq 0\}.$$

The goal is to compute the moments of the physical measure μ on $\boldsymbol{X}$. The available information is only the first moment $\boldsymbol{y}_{(1,0)}^{\text{num}}$ (i.e., the expectation of the first coordinate) of μ approximately computed using (22) with $f(x) = x_1$. The objective function is $F(\boldsymbol{y}) = (\boldsymbol{y}_{(1,0)} - \boldsymbol{y}_{(1,0)}^{\text{num}})^2$. Table 2 compares the moments returned by relaxation (15) of order $k = 10$ with moments computed numerically using (22); we observe a good agreement. Figure 2 then compares the support approximations computed using using the Christoffel polynomial (17) with $\boldsymbol{y} = \boldsymbol{y}^k$.

Acknowledgment

Table 2: Hénon map: comparison of moments computed by the SDP relaxation (15) and using (22). Only the first moment corresponding to x_1 was given as data input to (15).

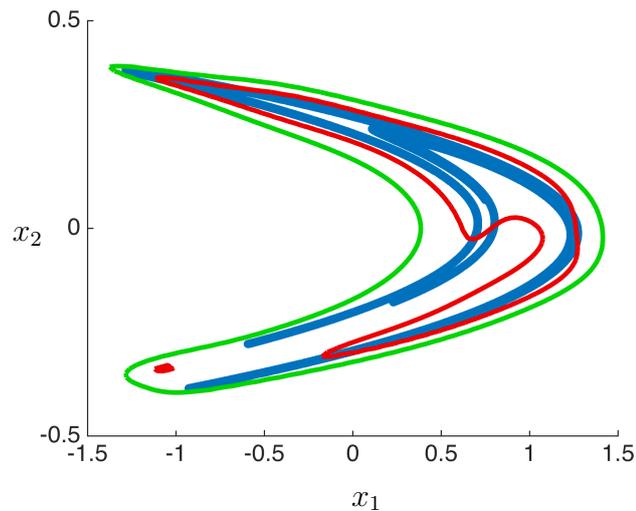| Moments	x_1	x_2	x_1^2	$x_1 x_2$	x_2^2	x_1^3	$x_1^2 x_2$	$x_1 x_2^2$	x_2^3
SDP	0.2570	0.0771	0.5858	-0.0379	0.0527	0.2468	0.0131	-0.0140	0.0067
Numeric	0.2570	0.0771	0.5858	-0.0291	0.0527	0.2320	0.0510	-0.0174	0.0063

Figure 2: Approximation of the support of the invariant measure of the Hénon map, confidence regions of the Christoffel polynomial built from 20 moments : green 90 %, red 50 %, blue simulated trajectories.

References

[1] E. M. Bollt. The path towards a longer life: On invariant sets and the escape time landscape. International Journal of Bifurcation and Chaos, 15(05):1615–1624, 2005.

[2] W. P. Cross, H. E. Romeijn, R. L. Smith. Approximating extreme points of infinite dimensional convex sets. Mathematics of Operations Research, 23(2):433–442, 1998.

[3] M. Fazel. Matrix rank minimization with applications. PhD Thesis, Elec. Eng. Dept, Stanford University, 2002.

[4] V. Gaitsgory, M. Quincampoix. Linear programming approach to deterministic infinite horizon optimal control problems with discounting. SIAM Journal on Control and Optimization, 48(4):2480–2512, 2009.

[5] B. Gustafsson, M. Putinar, E. B. Saff, N. Stylianopoulos. Bergman polynomials on an archipelago: estimates, zeros and shape reconstruction. Advances in Mathematics, 222(4):1405–1460, 2009.

[6] D. Henrion. Semidefinite characterisation of invariant measures for one-dimensional discrete dynamical systems. Kybernetika, 48(6):1089–1099, 2012.

[7] D. Henrion, J. B. Lasserre, J. Löfberg. Gloptipoly 3: moments, optimization and semidefinite programming. Optimization Methods and Software, 24:761–779, 2009.

[8] O. Hernández-Lerma, J. B. Lasserre. Discrete-time Markov control processes: basic optimality criteria. Springer, 1996.

[9] O. Jenkinson. Ergodic optimization. Discrete and Continuous Dynamical Systems, 15(1):197, 2006.

[10] O. Junge, I. G. Kevrekidis. On the sighting of unicorns: A variational approach to computing invariant sets in dynamical systems. Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(6):063102, 2017.

[11] M. Korda, D. Henrion, C. N. Jones. Convex computation of the maximum controlled invariant set for polynomial control systems. SIAM Journal on Control and Optimization, 52(5):2944–2969, 2014.

[12] M. Korda, D. Henrion, C. N. Jones. Controller design and value function approximation for nonlinear dynamical systems. Automatica, 67:54–66, 2016.

[13] M. Korda, D. Henrion, C. N. Jones. Convergence rates of moment-sum-of-squares hierarchies for optimal control problems. *Systems & Control Letters*, 100:1–5, 2017.

[14] A. Lasota, M. C. Mackey. Chaos, fractals, and noise - Stochastic aspects of dynamics. Springer, 1994.

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

[16] J. B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization, 11(3):796–817, 2001.

[17] J. B. Lasserre, D. Henrion, C. Prieur, E. Trélat. Nonlinear optimal control via occupation measures and LMI relaxations. SIAM Journal on Control and Optimization, 47(4):1643–1666, 2008.

[18] J. B. Lasserre, E. Pauwels. The empirical christoffel function in statistics and machine learning. `arXiv:1701.02886`, 2017.

[19] J. Löfberg. Yalmip : A toolbox for modeling and optimization in Matlab. Proceedings of the IEEE CACSD Conference, Taipei, Taiwan, 2004.

[20] V. Magron, D. Henrion, M. Forets. Semidefinite characterization of invariant measures for polynomial systems. Submitted for publication, 2018.

[21] S. P. Meyn, R. L. Tweedie. Markov chains and stochastic stability. Springer, 2012.

[22] Mezić, Igor, Banaszuk, Andrzej. Comparison of systems with complex behavior. Physica D: Nonlinear Phenomena, 197:101–133, 2004.

[23] N. Ozay, C. Lagoa, M. Sznaier. Set membership identification of switched linear systems with known number of subsystems. Automatica, 51:180–191, 2015.

[24] N. Ozay, M. Sznaier, C. Lagoa. Convex certificates for model (in) validation of switched affine systems with unknown switches. IEEE Transactions on Automatic Control, 59(11):2921–2932, 2014.

[25] E. Pauwels, J. B. Lasserre. Sorting out typicality with the inverse moment matrix sos polynomial. *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[26] R. R. Phelps. Lectures on Choquet's theorem. Springer, 2001.

[27] M. Putinar. Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal,, 42:969–984, 1993.

[28] J. Sturm. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. Optimization Methods and Software, 11:625–653, 1999.