Convex Computation of Extremal Invariant Measures of Nonlinear Dynamical Systems and Markov Processes

Milan Korda\textsuperscript{1,2,3} · Didier Henrion\textsuperscript{1,2,3} · Igor Mezić\textsuperscript{4}

Received: 6 July 2018 / Accepted: 22 October 2020 / Published online: 7 January 2021
© Springer Science+Business Media, LLC, part of Springer Nature 2021

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, 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. As a by-product, we present a simple method to certify the nonexistence of an invariant measure, which is an important question in the theory of Markov processes. 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 with 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 and atomic measures. 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 (Lasserre 2001) 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 Korda et al. 2017). 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., Lasserre and Pauwels 2017; Pauwels and Lasserre 2016). For the latter, we construct a sequence of absolutely continuous measures with polynomial densities converging weakly to the target measure.

An interesting by product of the approach presented is the possibility to certify the nonexistence of an invariant measure, which is particularly pertinent for Markov processes. Such certification boils down to proving the emptiness of a spectrahedron defining the feasible set of the SDPs solved.

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.
This work is a continuation of the movement to apply convex optimization-based techniques to non-convex problems arising from dynamical systems theory and control. For example, the related problem of invariant set computation was addressed in Korda et al. (2014), whereas (Lasserre et al. 2008; Gaitsgory and Quincampoix 2009; Korda et al. 2016) addressed optimal control; Ozay et al. (2014), Ozay et al. (2015) addressed model validation and switching system identification, respectively. The problem addressed here, i.e., invariant measure computation, was also addressed by this approach in Henrion (2012) in one spatial dimension; this work can therefore be seen as a generalization of Henrion (2012) to multiple dimensions and with a far more detailed theoretical and computational analysis. In the concurrent work (Magron et al. 2018), 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 (Bollt 2005; Junge and Kevrekidis 2017). 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 (Jenkinson 2006, 2019; Bochi 2018), 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).

The presented approach based on optimization over Borel measures has a convex dual as an optimization over continuous functions that can be approximated by polynomial sum-of-squares. This line of research has been investigated independently for various problems from dynamical systems (e.g., Chernyshenko et al. 2014; Fantuzzi et al. 2016; Goluskin 2017). Of particular relevance to this work is Tobasco et al. (2017) which is dual to our approach in the continuous-time setting and when the objective functional in our approach is restricted to be linear. A by-product of our work is therefore an asymptotic convergence guarantee for the bounds obtained by Tobasco et al. (2017), provided that strong duality holds.

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 Sect. 7 to Markov processes (both in discrete and continuous time) as well as discusses how to certify nonexistence of invariant measures. Section 8 extends the method to eigenmeasures of the Perron–Frobenius operator, and Sect. 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 Sect. 7; the continuous-time cases are treated in Sects. 6 and 7.1. Consider therefore a deterministic discrete-time nonlinear dynamical system

\[ x^+ = T(x), \] (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 \( \mu \) satisfying the relation

\[ \mu(T^{-1}(A)) = \mu(A) \] (2)

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

\[ \int_X f \circ T \, d\mu = \int_X f \, d\mu \] (3)

for all \( f \in C(X) \).

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

\[ y_\alpha = \int_X x^\alpha \, d\mu(x), \] (4)

with \( x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n} \) 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

\[ \int_X T^\alpha(x) \, d\mu(x) = \int_X x^\alpha \, d\mu(x) \] (5)

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

\[ A(y) = 0, \] (6)

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

We remark that (6) characterizes all invariant measures associated with (1) with support in \( X \). 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

\[
\begin{align*}
\min_{y \in M(X)} & \quad F(y) \\
\text{s.t.} & \quad A(y) = 0 \\
& \quad y_0 = 1
\end{align*}
\] (7)
where the minimization is w.r.t. a sequence $y$ belonging to the convex cone
\[
M(X) := \left\{ y \in \mathbb{R}^\infty : y_\alpha = \int_X x^\alpha d\mu(x), \alpha \in \mathbb{N}^n, \mu \in \mathcal{M}(X)_+ \right\}
\]
of moments of nonnegative Borel measures on $X$, the objective functional $F : \mathbb{R}^\infty \to \mathbb{R}$ is convex, and the constraint $y_0 = 1$ is a normalization constraint enforcing that the measure is a probability measure.

We note that since $T$ is polynomial and hence continuous and $X$ 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 facilitates this. In particular, it can be extended-valued (i.e., equal to $+\infty$), therefore encoding any constraints of interest such as $\mu$ being absolutely continuous or singular w.r.t. to a given measure. See Sect. 5 for concrete choices of $F$.

A typical example encountered in practice for the choice of $F$ is
\[
F(y) = \sum_{|\alpha| \leq d} (y_\alpha - z_\alpha)^2,
\]
where $(z_\alpha)_{|\alpha| \leq d}$ is a given finite vector of moments of total degree no more than $d$. The moments $z_\alpha$ can be estimates of the first few moments of the invariant measure that we wish to compute obtained, e.g., from observed data or by analytical reasoning (e.g., based on the symmetries of the problem). The optimization problem (7) then seeks among all invariant measures $\mu$ the one which minimizes the discrepancy between the first moments $(y_\alpha)_{|\alpha| \leq d}$ of $\mu$ and the given moments $(z_\alpha)_{|\alpha| \leq d}$. 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 $F(y)$ in (7) depends only on finitely many moments $(y_\alpha)_{|\alpha| \leq d}$.

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 $f(x) = x^\alpha$, $|\alpha| \leq k$, 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 $M(X)$.

### 3.1 Finite-Dimensional Approximations of the Moment Cone

Here, we describe the semidefinite-programming representable outer approximation to $M(X)$. For this, we assume that the compact set $X$ is of the form

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

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

The outer approximation $M_d(X)$ of degree $d$, $d$ even, is

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

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

The localizing moment matrices $M_d(g_i y)$ are defined by

$$M_d(g_i y) = l^d_y(g_i v_d v_d^\top), \tag{11}$$

where $d_i = \lfloor (d - \deg g_i)/2 \rfloor$, $$v_d(x) := (x^\alpha)_{|\alpha| \leq d}$$

and where the Riesz functional $l^d_y : \mathbb{R}[[x]]_d \to \mathbb{R}$ is defined for any $f = \sum_{|\alpha| \leq d} f_\alpha x^\alpha$ by

$$l^d_y(f) = \sum_{|\alpha| \leq d} f_\alpha y_\alpha. \tag{12}$$

This functional mimics integration with respect to a measure; in particular when $y \in \mathbb{R}^{n+d}$ is a truncated moment vector of a measure $\mu$, then $l^d_y(f) = \int f \, d\mu$ for any $f \in \mathbb{R}[[x]]_d$. In this case, the localizing matrices $M_d(g_i y)$ are necessarily positive semidefinite, corresponding to the fact that $\int g_i p^2 \, d\mu \geq 0$ 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 $y$ is a moment vector of a nonnegative measure supported on $X$. In order for this to hold, the defining polynomials $g_i$ need to satisfy the so-called Archimedean condition:

---

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.
Assumption 1 There exist polynomials \( \sigma_i \) and a constant \( r \in \mathbb{R} \) such that

\[
r^2 - x^\top x = \sum_{i=0}^{n_g} \sigma_i g_i,
\]

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

Assumption 1 is an algebraic certificate of compactness of \( X \) because this assumption implies that \( X \subset \{x \mid x^\top x \leq r^2\} \). Since \( X \) 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 \( X \) for a sufficiently large \( r \) in which case Assumption 1 is satisfied trivially.

Theorem 1 (Putinar 1993) Suppose that Assumption 1 holds and that \( y \in M_d(X) \) for all \( d \geq 0 \). Then, there exists a unique nonnegative measure \( \mu \) on \( X \) 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,

\[
\int_X T^\alpha(x) \, d\mu = \int_X x^\alpha \, d\mu = 0, \quad |\alpha| \leq k,
\]

where \( T^\alpha(x) = T_1(x)^{\alpha_1} \cdots T_n(x)^{\alpha_n} \). 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 \( \mu \) 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

\[
l_y^{d_k}(T^\alpha(x)) = l_y^{d_k}(x^\alpha), \quad |\alpha| \leq k.
\]

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

\[
A_k y = 0,
\]

for some matrix \( A_k \in \mathbb{R}^{(n+k) \times (n+d_k)} \) where \( y \) is the moment sequence of \( \mu \). The equality constraint \( \int_X 1 \, d\mu \) translates to \( y_0 = 1 \). The conic constraint \( y \in M(X) \) is replaced, according to the previous section, by the constraint \( y \in M_{d_k}(X) \). This leads
to the following finite-dimensional relaxation of order $k$ of the infinite-dimensional problem (7)

$$
\min_{y \in \mathbb{R}^{(n+dk)}} F(y)
$$

s.t. \quad Ay = 0

$$
y_0 = 1
$$

$$
M_{dk}(g, y) \succeq 0, \quad \forall i = 0, 1, \ldots, n_g.
$$

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 $F(y)$ 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 Sturm 1999). 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 modeling software Gloptipoly 3 (Henrion et al. 2009) and Yalmip (Löfberg 2004).

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 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 and that the function $F$ is lower semi-continuous\(^2\) and let $y^k \in \mathbb{R}^{(n+dk)}$ denote an optimal solution to (15) and $p^*$ the optimal value of (7). Then, the following holds:

1. $\lim_{k \to \infty} F(y^k) = p^*$
2. There exists a subsequence $(k_i)_{i=1}^\infty$ such that $y^{k_i}$ converges pointwise to a moment sequence of an invariant measure $\mu^*$ attaining the minimum in (7).
3. In particular, if there is a unique invariant measure $\mu^*$ attaining the minimum in (7), then $y^k$ converges to the moment sequence of $\mu^*$.

**Proof** The proof follows a standard argument (see, e.g., Lasserre 2010). By Assumption 1, for every $\tilde{\alpha} \geq 0$ there exists a $k_0$ such that $|y_\alpha| \leq M|\alpha|$ for any vector $y$ satisfying the constraints of (15) for $k \geq k_0$ and for any $\alpha \in \mathbb{N}^n$ satisfying $|\alpha| \leq \tilde{\alpha}$, where $M$ is the constant from (13). This statement implies that each component of $y^k$

---

\(^2\) 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).
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 $d \deg T \geq [(\sigma_i - \deg g_i)/2]$. Then, necessarily $\ell^d_y (\sigma_0) \geq 0$ and $\ell^d_y (\sigma_i g_i) \geq 0$ for any $k \geq k_0$. Therefore, by linearity we have $\ell^d_y (f) \geq 0$. But since $f = r^2 - x^\top x$ by (13), we get $\ell^d_y (r^2) \geq \ell^d_y (x^\top x) = \sum_{i=1}^n \ell^d_y (x_i^2)$. Since $\ell^d_y (r^2) = r^2 \ell^d_y (1) = r^2 y_0$ and $y_0 = 1$ by the second constraint of (15) and since $\ell^d_y (x_i^2) \geq 0$, we conclude that $\ell^d_y (x_i^2) \in [0, r^2]$. Proceeding recursively, applying the same reasoning to $x^{a} f$ with $a_i$ even for all $a \in \mathbb{N}^n$ satisfying $|a| \leq \bar{a} := \max \{a \in \mathbb{N}^n \mid |a|/2 + (\deg g_i)/2 \leq [(dk - \deg g)/2]\}$, we conclude that all even moments $y_{a^\top}, |a| \leq \bar{a}$ lie in $[0, r^2|a|]$. Since even moments are on the diagonal of the matrix $M_d(y) \geq 0$ 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 $\limsup_{k \to \infty} |y^k_\alpha| < \infty$ for each $\alpha \in \mathbb{N}^n$, it follows using a standard diagonal argument that we can extract a subsequence $y^{k_i}$ satisfying, for each $\alpha \in \mathbb{N}^n$, $\lim_{i \to \infty} y^{k_i}_\alpha = y^*_\alpha$ with $y^*_\alpha \in \mathbb{R}$. To conclude the proof, it remains to show that $y^*$ is a moment sequence of a measure attaining the minimum in (7).

Using Theorem 1, it follows that $y^*$ is a moment sequence of a nonnegative measure $\mu^*$ on $X$ since $M_d(g_i, y^*) \geq 0$ 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, $y^*$ satisfies the equality constraints of (7) by continuity since each row of the matrix $A$ has only finitely many nonzero elements. Therefore, $\mu^*$ is an invariant measure. Finally, since (15) is a relaxation of (7) we have $F(y^{k_i}) \leq p^*$ . By the lower semi-continuity of $F$ we also have $F(y^*) \leq \lim_{i \to \infty} F(y^{k_i}) \leq p^*$ and hence necessarily $F(y^*) = p^*$ since $\mu^*$ is feasible in (7) and therefore $F(y^*) \geq p^*$. \hfill $\square$

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 Jenkinson (2006, Theorem 3.2).

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

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 $\mu^*$. The approximations constructed here aim at enclosing a certain prescribed portion of the support. Guaran-
teed outer approximates to the global attractor (on which certain invariant measures are supported) can be computed using the approach of Schlosser and Korda (2020). 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 \( y \) of a nonnegative measure \( \mu \) on \( X \). Then, from the truncated moments of degree up to \( d \), \( d \) even, we define the Christoffel polynomial

\[
q_d^y(x) = v_{d/2}(x)\top M_d(y)^{-1}v_{d/2}(x),
\]

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 \( y \). The polynomial \( q_d^y(x) \) is well defined as long the moment matrix \( M_d(y) \) is invertible, which is satisfied if and only if \( \mu \) is not supported on the zero level set of a polynomial of degree \( d/2 \) or less.

The sublevel sets of the polynomial \( q_d^y(x) \) have a remarkable property of approximating the shape of the support of the measure \( \mu \). In the real multivariate domain, this was observed recently for empirical measures (sums of Dirac masses) in Pauwels and Lasserre (2016) and subsequently studied analytically for measures with certain regularity properties in Lasserre and Pauwels (2017). In the complex domain, the theory is far more developed, see, e.g., Gustafsson et al. (2009) and references therein.

Here, we use the following simple result which holds for arbitrary probability measures \( \mu \).

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

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

**Proof** We bound the complementary event:

\[
\mu(\{x : q_d^y(x) > \gamma_{\epsilon} \}) = \int_X 1_{\{x : q_d^y(x) > \gamma_{\epsilon} \}} \, d\mu = \int_X 1_{\{x : q_d^y(x) > \gamma_{\epsilon} \}} \, d\mu \leq \gamma_{\epsilon}^{-1} \int_X q_d^y(x) \, d\mu(x),
\]

where we have used the fact that \( q_d^y(x) \) is nonnegative. Using the definition of \( q_d^y(x) \), we get

\[
\int_X q_d^y(x) \, d\mu(x) = \int_X v_{d/2}(x)\top M_d(y)^{-1}v_{d/2}(x) \, d\mu(x)
\]

\[
= \text{trace} \left\{ M_d(y)^{-1} \int_X v_{d/2}(x)v_{d/2}(x)\top \, d\mu(x) \right\}
\]

\[
= \text{trace} \{ M_d(y)^{-1}M_d(y) \} = \text{trace} \left\{ I_{\left(\frac{n+d/2}{n}\right)} \right\} = \left( \frac{n+d/2}{n} \right),
\]

where \( I_k \) denotes the identity matrix of size \( k \). Therefore,

\[
\mu(\{x : q_d^y(x) > \gamma_{\epsilon} \}) \leq \frac{1}{\gamma_{\epsilon}} \left( \frac{n+d/2}{n} \right).
\]

\( \Box \) Springer
and the result follows since \( \mu(\{ x : q_d^y(x) \leq \gamma \epsilon \}) = 1 - \mu(\{ x : q_d^y(x) > \gamma \epsilon \}) \).

Lemma 1 can be readily used to construct approximations to the support of \( \mu \) in the form
\[
\{ x : q_d^{y_k}(x) \leq \gamma \epsilon \}
\]

where \( y_k^{\ast} \in \mathbb{R}^{(n+d_k)} \) is a solution to the \( k \)th order relaxation (15). We remark that \( y_k^{\ast} \) only approximates the moments of \( \mu^{\ast} \), 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 \( y \in \mathbb{R}^{(n+d)} \) (e.g., a truncated moment vector of a measure), we can always represent the Riesz functional \( \ell_d^y : \mathbb{R}[x]_d \to \mathbb{R} \) as
\[
\ell_d^y(p) = \int_X p(x)q(x)dx
\]
for some polynomial \( q \in \mathbb{R}[x]_d \), provided that the set \( X \) has a non-empty interior. Indeed, by linearity it suffices to satisfy (18) for \( p(x) = x^\alpha, |\alpha| \leq d \), which leads to a system of linear equations
\[
M_d^Lq = y,
\]
where \( q \) is the coefficient vector of polynomial \( q \) in the monomial basis with the same ordering as the vector of moments \( y \), and \( M_d^L \) is the moment matrix of the Lebesgue measure on \( X \) of degree \( 2d \), i.e.,
\[
M_d^L := \int_X v_d(x)v_d(x)^\top dx.
\]

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

**Theorem 3** Let Assumption 2 hold, suppose that the interior of \( X \) is non-empty, denote \( y_k^{\ast} \in \mathbb{R}^{(n+d_k)} \) any solution to (15), and let
\[
q_k := (M_d^L)^{-1} y_k^{\ast}.
\]

Then, the signed measures with densities \( q_k(x) := v_k^\top(x)q_k \in \mathbb{R}[x]_k \) with respect to the Lebesgue measure, converge weakly star on \( X \) to the invariant measure \( \mu^{\ast} \).
Proof Verifying weak star convergence means that \( \lim_{k \to \infty} \int_X f q_k \, dx = \int_X f \, d\mu^* \)
for all \( f \in C(X) \). Since \( X \) 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 \( C(X) \). By construction, we have \( \int_X x^\alpha q_k = y^k_\alpha \) for \( k \geq |\alpha| \) and hence
\[
\lim_{k \to \infty} \int_X x^\alpha q_k(x) \, dx = \lim_{k \to \infty} y^k_\alpha = y^*_\alpha = \int_X x^\alpha \, d\mu^*(x)
\]
by Theorem 2, part 3.

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
\[
(M^L_k)^{-1} \tilde{y}^k,
\]
where \( \tilde{y}^k \in \mathbb{R}^{\binom{n+k}{n}} \) is the vector of the first \( \binom{n+k}{n} \) elements of \( y^k \), rather than the full vector \( y^k \in \mathbb{R}^{\binom{n+k}{n}} \). 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 \( \mu \) with support included in \( X \). Therefore, for Lebesgue almost every \( x \in X \) for which the trajectory of (1) originating from \( x \) stays in \( X \) we have for any \( f \in C(X) \)
\[
\int_X f \, d\mu = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N f(T^i(x)).
\]
Selecting \( f(x) = x^\alpha \), we can approximately compute the moments of \( \mu \) as
\[
y^\text{num}_\alpha = \frac{1}{N} \sum_{i=1}^N f(T^i(x)) \approx \int_X x^\alpha \, d\mu
\]
with some $N \gg 1$ and $\alpha$ running over a selected subset of multi-indices $I$. 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 $\mu$, just from the information contained in $\{y_{\alpha}^{\text{num}} : \alpha \in I\}$. This is achieved by setting the objective function in (15) to

$$F(y) = \sum_{\alpha \in I} (y_{\alpha} - y_{\alpha}^{\text{num}})^2$$

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 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., Phelps 2001, Proposition 12.4; see also Cross et al. 1998 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 $y^k \in \mathbb{R}^{(n+d_k)}$ denote the optimal solution to (15). Then, $y^k$ converges to a moment sequence of an ergodic measure $\mu$, 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). \qed

A typical choice of the objective functional is $F(y) = \sum_{\alpha} c_{\alpha} y_{\alpha}$ with $c$ having only finitely many nonzero 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 $\mu$ absolutely continuous w.r.t. a given measure $\nu$ through the choice of the objective function $F$. In most practical applications, the measure $\nu$ will be the Lebesgue measure. The absolutely continuity of $\mu$ w.r.t. $\nu$ is equivalent to the existence of a density $\rho$ such that $\mu = \rho \, d\nu$. Assuming that $\rho(x) \leq \gamma < \infty$ for $\nu$-almost all $x \in X$, we can impose the absolute continuity constraint by choosing

$$F(\mu) = \mathbb{I}_{\{\mu \mid \mu \leq \gamma \nu\}},$$

where $\mathbb{I}$ is the indicator function of the set $\{\mu : \mu \leq \gamma \nu\}$. This ensures that $\mu$ is absolutely continuous with respect to $\nu$. The optimal solution $y^k$ satisfies $y^k = I(\mu) = \mathbb{I}_{\{\mu \mid \mu \leq \gamma \nu\}}$, which represents the moment sequence of an absolutely continuous measure $\mu$. The choice of the objective functional $F(y)$ is crucial for targeting absolutely continuous measures. A typical choice is $F(y) = \sum_{\alpha} c_{\alpha} y_{\alpha}$ with $c$ having only finitely many nonzero elements (i.e., $F$ is a linear combination of a finite number of moments).
where \( \mathbb{1}_A \) is the extended-value characteristic of a set \( A \) (i.e., \( \mathbb{1}_A(x) = 0 \) if \( x \in A \) and \( \mathbb{1}_A = +\infty \) if \( x \notin A \)). Then \( F(y) \) is a convex function of \( y \) and its domain is the set of moment vectors \( y \) such that

\[
M_d(yz - y) \succeq 0 \quad \forall d \in \mathbb{N},
\]

where \( z \) denotes the moment sequence of \( \nu \). 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 Sect. 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 Sect. 5.2.

We use the observation that if a measure \( \mu \) consists of \( K \) atoms, then the associated moment matrix \( M_d(y) \) 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

\[
F(y) = \mathbb{1}_{\{y \mid \text{rank } M_d(y) = K\}}.
\]

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

\[
F(y) = \mathbb{1}_{\{y \mid \text{trace } M_d(y) \leq \gamma\}},
\]

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

\[
\text{trace } M_d(y) \leq \gamma
\]

which is added to the constraints of (7) and (15). We note that since \( M_d(y) \) is positive semidefinite, its trace coincides with its nuclear norm, which is a standard proxy for rank minimization (Fazel 2002) (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

\[ \dot{x} = b(x), \]  

(25)
each component of the vector field \( b \) is assumed to be a multivariate polynomial. We are seeking a nonnegative measure \( \mu \) on \( X \) invariant under the flow of dynamical system (25), which is equivalent to the condition

\[ \int_X \nabla f \cdot b \, d\mu = 0 \]  

(26)
for all \( f \in C^1(X) \).

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

\[ \min_{\mu \in M_+(X)} \quad F(\mu) \]

s.t. \[ \int_X \nabla f \cdot b \, d\mu = 0 \quad \forall f \in C(X) \]

(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 Sect. 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 \( X \). 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 \( W \) 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 \( A \subset X \), \( P_w(A) \) is the probability that \( w_k \in A \).

The condition for a probability measure \( \mu \) to be invariant then reads

\[ \int_X \int_W f(T(x, w)) \, dP_w(w) \, d\mu(x) = \int_X f(x) \, d\mu(x) \]  

(28)

\(^3\) For a relation of this form of a Markov process to the one specified by the transition kernel, see, e.g., Hernández-Lerma and Lasserre (1996).
for all \( f \in C(X) \). This equation is linear in \( \mu \) and can therefore be used in (7) instead of the first equality constraint. The approach then proceeds along the steps of Sect. 3.2, i.e., we set \( f(x) = x^\alpha \) and enforce (28) for all such \( f \) with \( |\alpha| \leq k \), leading to

\[
\int_X \int_W T^\alpha(x, w) dP_w(w) d\mu(x) = \int_X x^\alpha d\mu, \quad |\alpha| \leq k. \tag{29}
\]

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

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

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

\[
\int_W T^\alpha(x, w)dP_w(w) = \sum_{\beta, \gamma} t_{\alpha,\beta,\gamma} x^\beta \int_W w^\gamma dP_w(w) = \sum_{\beta, \gamma} t_{\alpha,\beta,\gamma} x^\beta m_\gamma,
\]

where

\[
m_\gamma = \int_W w^\gamma dP_w(w)
\]

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 rewritten as

\[
\sum_{\beta, \gamma} t_{\alpha,\beta,\gamma} m_\gamma \int_X x^\beta d\mu(x) = \int_X x^\alpha d\mu(x), \quad |\alpha| \leq k,
\]

or, in terms of the moments of \( \mu \) (4),

\[
\sum_{\beta, \gamma} t_{\alpha,\beta,\gamma} m_\gamma y_\beta = y_\alpha, \quad |\alpha| \leq k.
\]

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

\[
Ay = 0.
\]

Adding the normalization constraint \( y_0 = 1 \) and the positive-semidefiniteness constraints \( M_{dk}(y) \succeq 0 \) and \( M_{dk}(y, g_i) \succeq 0 \) leads to the optimization problem (15). The objective functional \( F(y) \) 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 (Meyn and Tweedie 2012, 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

\[
\int_{X} A f \, d\mu = 0, \tag{30}
\]

where \(A\) 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

\[
Af = \sum_i b_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} [\sigma \sigma^\top]_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j},
\]

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 \(\mu\) (4). The approach then proceeds in exactly the same fashion as described in Sect. 3.2, leading to a finite-dimensional relaxation of the same form as (15).

Certifying Nonexistence of an Invariant Measure

Contrary to the deterministic case, stochastic processes driven by a Wiener process typically evolve on non-compact domains, thereby rendering the question of the existence of an invariant measure much more subtle. For example, the Wiener process itself does not admit an invariant measure but the Ornstein–Uhlenbeck process \(dx(t) = -ax(t)dt + \sigma dW(t)\) does for any \(a > 0\). Interestingly, the proposed approach provides a means to numerically certify that no invariant measure exists. Indeed, since the feasible set of the SDP relaxation (15) contains the truncated moment sequences of all invariant measures, proving the emptiness of this feasible set (which is a finite-dimensional spectrahedron) implies the nonexistence of an invariant measure. Indication of such infeasibility is detected during the solution of SDP (15) by most existing solvers, although its rigorous certification is more involved (Henrion et al. 2016). The proposed method thereby complements methods based on sum-of-squares programming that can be used to prove the existence of an invariant measure using Foster–Lyapunov conditions (Meyn and Tweedie 2012).

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 (Mezić and Banaszuk
2004). For concreteness, we work with the discrete-time dynamics (1), although analogous results can be obtained for continuous time, following the developments of Sect. 6. The Perron–Frobenius operator $\mathcal{P} : M_c(X) \to M_c(X)$ is defined by

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

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

$$\int f \circ T \, d\mu = \lambda \int f \, d\mu$$

for all $f \in C(X)$. Since the moment-based approach developed previously applies to nonnegative 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 $\mu_R^+ \in M(X)$, $\mu_R^- \in M(X)$, $\mu_I^+ \in M(X)$, $\mu_I^- \in M(X)$. Similarly, we write

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

Then, the condition (31) is equivalent to

$$\int (f \circ T - \lambda_R f) \, d\mu_R^+ - \int (f \circ T - \lambda_R f) \, d\mu_R^-$$

$$+ \lambda_I \int f \, d\mu_I^+ - \lambda_I \int f \, d\mu_I^- = 0,$$

$$\int (f \circ T - \lambda_R f) \, d\mu_I^+ - \int (f \circ T - \lambda_R f) \, d\mu_I^-$$

$$+ \lambda_I \int f \, d\mu_R^+ - \lambda_I \int f \, d\mu_R^- = 0$$

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

$$A(y_R^+, y_R^-, y_I^+, y_I^-) = 0,$$

where $y_R^+$ etc are the moment sequences of the respective measures and $A$ is a linear operator. Coupled with the normalization constraint

$$(y_R^+)_0 + (y_R^-)_0 + (y_I^+)_0 + (y_I^-)_0 = 1$$
and objective functional $F(y_R^+, y_R^-, y_I^+, y_I^-)$, we arrive at an infinite dimensional linear programming problem

\[
\begin{align*}
\text{min} & \quad F(y_R^+, y_R^-, y_I^+, y_I^-) \\
\text{s.t.} & \quad A(y_R^+, y_R^-, y_I^+, y_I^-) = 0 \\
& \quad (y_R^+)_0 + (y_R^-)_0 + (y_I^+)_0 + (y_I^-)_0 = 1 \\
& \quad y_R^+ \in M(X), \ y_R^- \in M(X), \ y_I^+ \in M(X), \ y_I^- \in M(X),
\end{align*}
\]

which is then approximated by a sequence of finite-dimensional SDPs in exactly the same fashion as described in Sect. 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 $X = [-1, 1] = \{x \in \mathbb{R} : (x + 1)(1 - x) \geq 0\}$.

9.1.1 Physical Measure

First, we compute the moments of the unique physical measure $\mu$ on $X$. For this example, the density of the physical measure is given by Lasota and Mackey (1994)

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

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 $F(y) = (y_1 - 0)^2$ (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 Henrion (2012), 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 $x^\alpha$ and replace the monomial basis vectors $v_{d}^{\alpha}$ by the vectors of Chebyshev polynomials up to degree $d_{i}$ in (11)); see (Henrion 2012) for more details on the use of Chebyshev polynomials in this context. In Fig. 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

\[\text{Springer}\]
Fig. 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

| 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 |

Only the first moment was given as input to (15)

Table 2 Hénon map: comparison of moments computed by the SDP relaxation (15) and using (22)

| 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 |

Only the first moment corresponding to \( x_1 \) was given as input to (15)

approximation theory. The computed moments (transformed to the monomial basis) are compared in Table 1; we see a very good match.

9.2 Hénon Map

Our second example is the Hénon map:

\[
\begin{align*}
x_1^+ &= 1 - 1.4x_1^2 + x_2 \\
x_2^+ &= 0.3x_1.
\end{align*}
\]

The set \( X \) is the box \([-1.5, 1.5] \times [-0.4, 0.4] \) expressed as

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

The goal is to compute the moments of the physical measure \( \mu \) on \( X \). The available information is only the first moment \( y_{(1,0)}^{\text{num}} \) (i.e., the expectation of the first coordinate) of \( \mu \) approximately computed using (22) with \( f(x) = x_1 \). The objective function is

\[
F(y) = (y_{(1,0)} - y_{(1,0)}^{\text{num}})^2.
\]

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 the Christoffel polynomial (17) with $y = y^k$.

### 9.3 Lorenz System

Next, we consider the classical Lorenz system

$$\begin{align*}
\dot{x}_1 &= 10(x_2 - x_1) \\
\dot{x}_2 &= x_1(28 - x_3) - x_2 \\
\dot{x}_3 &= x_1x_2 - \frac{8}{3}x_3
\end{align*}$$

scaled by the linear coordinate transformation $\hat{x} = \text{diag}(1/25, 1/30, 1/50)x$. The goal is to compute the moments of the physical measure $\mu$ which is supported on the Lorenz attractor. Here, we investigate the effect of the amount of information available for the computation in terms of the number of moments used in the objective function $F(y) = \sum_{i=1}^{N} (y_i - y_i^{\text{num}})^2$, where $y_i^{\text{num}}$ are approximations of the true moments computed using simulation. We investigate as well the effect of the accuracy of the approximate moments provided by varying the simulation length $M$. The accuracy is measured in terms of the percentage root mean square error on the sequence of the first 56 moments (i.e., moments up to degree 5) with respect the numerical approximations from a simulation of length $10^6$. Figure 3 shows the results for $M = 10^2$ and $M = 10^6$ with $N \in \{1, \ldots, 10\}$. We observe a rather small impact of the accuracy of the moments provided (determined by $M$); on the other hand, the number of moments provided $N$ plays a significant role, resulting in an error of roughly 15% with one moment provided and error of roughly 0.05% with seven or more moments.
Table 3 Ornstein–Uhlenbeck process: comparison of moments computed by the SDP relaxation (15) with moments computed analytically

| Moments | \( x \) | \( x^2 \) | \( x^3 \) | \( x^4 \) | \( x^5 \) | \( x^6 \) | \( x^7 \) | \( x^8 \) | \( x^9 \) | \( x^{10} \) |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| SDP     | 0     | 0.5   | 0     | 0.75  | 0     | 1.8750| 0     | 6.5625| 0     | 29.5312|
| True    | 0     | 0.5   | 0     | 0.75  | 0     | 1.8750| 0     | 6.5625| 0     | 29.5312|

9.4 Stochastic Processes

**Ornstein–Uhlenbeck** To demonstrate the approach of Sect. 7.1, we first choose the classical Ornstein–Uhlenbeck process given by

\[
\text{d}x(t) = -ax(t)\text{d}t + \sigma \text{d}W(t),
\]

where \( W(t) \) is the Wiener process. For \( a > 0 \), this process has a unique invariant measure equal to \( \mathcal{N}(0, \frac{\sigma^2}{2a}) \), i.e., the normal distribution with zero mean and standard deviation \( \frac{\sigma}{\sqrt{2a}} \). In order to compute the moments, we solved (15) without the objective function (i.e., as a feasibility problem). The results for \( a = 1 \) and \( \sigma = 1 \) and \( d = 10 \) are shown in Table 3; we observe a precise match.
Table 4 Nonlinear drift: comparison of moments computed by the SDP relaxation (15) with moments computed using a Monte Carlo simulation

| Moments | $x$ | $x^2$ | $x^3$ | $x^4$ | $x^5$ | $x^6$ | $x^7$ | $x^8$ | $x^9$ | $x^{10}$ |
|----------|-----|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| SDP      | $-0.0000$ | $0.4758$ | $0.0000$ | $0.5000$ | $-0.0000$ | $0.7138$ | $0.0000$ | $1.2500$ | $-0.0000$ | $2.4982$ |
| Mont Carlo | $0.0001$ | $0.4776$ | $0.0011$ | $0.4997$ | $0.0039$ | $0.7192$ | $0.0130$ | $1.2637$ | $0.0413$ | $2.5683$ |

Table 5 Detecting nonexistence of invariant measure: minimum degree required to render the SDP relaxation (15) infeasible versus the value of the drift coefficient $a$

| Drift coefficient $a$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|---|---|---|---|---|---|---|---|---|---|----|
| Minimum degree for infeasibility of (15) | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |

**Nonlinear Drift** Next, we test the approach on a stochastic differential equation with a nonlinear drift

$$dx(t) = -ax^3(t)dt + \sigma dW(t).$$

For $a > 0$, this stochastic process is uniformly ergodic and hence admits a unique invariant measure (Meyn and Tweedie 2012). Table 4 shows the results of solving (15) for $d = 10$ in comparison with Monte Carlo simulation using the Euler forward discretization and $10^5$ sample trajectories.

**Detecting Nonexistence of Invariant Measures** In this section, we demonstrate the ability of the approach to numerically prove the nonexistence of an invariant measure as described in Sect. 7.1. For this, we consider the process

$$dx(t) = -ax^3(t)dt + (1 + x^2(t))dW(t).$$

Clearly, the process does not admit an invariant measure for $a < 0$. The situation is more interesting for $a > 0$. We carried out computations for $a \in \{0, 1, \ldots, 10\}$, solving the SDP relaxation (15). The results indicated that no invariant measure exists for this range of values of $a$, although the degree to certify it (i.e., to render (15) infeasible) increases with $a$. This is intuitive since, vaguely speaking, the larger the value of $a$, the “more stable” the process is. Table 5 shows this dependence; it appears that the minimum degree obeys the linear relation $d = 2a + 2$, although we did not attempt to prove this analytically. It should be noted that the infeasibility of (15) was decided by the interior point solver SeDuMi; in particular we did not use rigorous certification tools such as Henrion et al. (2016) and therefore these results should be understood as a strong numerical evidence of nonexistence rather than a rigorous proof.
10 Conclusion

This work presented a convex-optimization-based method for computation of invariant measures for continuous- and discrete-time deterministic and stochastic systems. We described how to cast the problem of invariant set computation as an infinite-dimensional LP in the space of Borel measures and how to target particular invariant measures by the choice of an objective functional. We showed how this infinite-dimensional LP can be approximated by a sequence of finite-dimensional SDPs with a guaranteed asymptotic convergence and how the results of this SDP can be used for support approximation of the invariant measure using the Christoffel–Darboux kernel. Interesting by-products of the approach are a method to certify nonexistence of invariant measures and a method to compute eigenmeasures of the Perron–Frobenius operator.

Future work should focus on improving the scalability of the approach by exploit sparsity or symmetries of the problem at hand or on developing a data-driven counterpart of the approach where the model is unknown and only finite collection of observations is available, in the spirit of Korda (2020).

Acknowledgements This work benefited from discussions with Victor Magron. This research was supported in part by the ARO-MURI grant W911NF-17-1-0306. The research of M. Korda was also supported by the Swiss National Science Foundation under grant P2ELP2_165166 and by the Czech Science Foundation (GACR) under contract No. 20-11626Y.

References

Bochi, J.: Ergodic optimization of Birkhoff averages and Lyapunov exponents. In: Proceedings of the International Congress of Mathematicians (2018)
Bollt, E.M.: The path towards a longer life: on invariant sets and the escape time landscape. Int. J. Bifurc. Chaos 15(05), 1615–1624 (2005)
Chernyshenko, S.I., Goulart, P., Huang, D., Papachristodoulou, A.: Polynomial sum of squares in fluid dynamics: a review with a look ahead. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 372(2020), 20130350 (2014)
Cross, W.P., Romeijn, H.E., Smith, R.L.: Approximating extreme points of infinite dimensional convex sets. Math. Oper. Res. 23(2), 433–442 (1998)
Fantuzzi, G., Goluskin, D., Huang, D., Chernyshenko, S.I.: Bounds for deterministic and stochastic dynamical systems using sum-of-squares optimization. SIAM J. Appl. Dyn. Syst. 15(4), 1962–1988 (2016)
Fazel, M.: Matrix rank minimization with applications. Ph.D. Thesis, Electrical Engineering Department, Stanford University (2002)
Gaitsgory, V., Quincampoix, M.: Linear programming approach to deterministic infinite horizon optimal control problems with discounting. SIAM J. Control Optim. 48(4), 2480–2512 (2009)
Goluskin, D.: Bounding averages rigorously using semidefinite programming: mean moments of the Lorenz system. J. Nonlinear Sci. 28(2), 621–651 (2017)
Gustafsson, B., Putinar, M., Saff, E.B., Stylianopoulos, N.: Bergman polynomials on an archipelago: estimates, zeros and shape reconstruction. Adv. Math. 222(4), 1405–1460 (2009)
Henrion, D.: Semidefinite characterisation of invariant measures for one-dimensional discrete dynamical systems. Kybernetika 48(6), 1089–1099 (2012)
Henrion, D., Lasserre, J.B., Löfberg, J.: Gloptipoly 3: moments, optimization and semidefinite programming. Optim. Methods Softw. 24, 761–779 (2009)
Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, Berlin (1996)
Jenkinson, O.: Ergodic optimization. Discrete Contin. Dyn. Syst. 15(1), 197 (2006)
Jenkinson, O.: Ergodic optimization in dynamical systems. Ergod. Theory Dyn. Syst. 39(10), 2593–2618 (2019)
Junge, O., Kevrekidis, I.G.: On the sighting of unicorns: a variational approach to computing invariant sets in dynamical systems. Chaos: Interdiscip. J. Nonlinear Sci. 27(6), 063102 (2017)
Korda, M.: Computing controlled invariant sets from data using convex optimization. SIAM J. Control Optim. 58(5), 2871–2899 (2020)
Korda, M., Henrion, D., Jones, C.N.: Convex computation of the maximum controlled invariant set for polynomial control systems. SIAM J. Control Optim. 52(5), 2944–2969 (2014)
Korda, M., Henrion, D., Jones, C.N.: Controller design and value function approximation for nonlinear dynamical systems. Automatica 67, 54–66 (2016)
Korda, M., Henrion, D., Jones, C.N.: Convergence rates of moment-sum-of-squares hierarchies for optimal control problems. Syst. Control Lett. 100, 1–5 (2017)
Lasota, A., Mackey, M.C.: Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Springer, Berlin (1994)
Henrion, Didier, Naldi, Simone, Din, Mohab Safey El: Exact algorithms for linear matrix inequalities. SIAM J. Optim. 26(4), 2512–2539 (2016)
Lasserre, J.B.: Moments, Positive Polynomials and their Applications. Imperial College Press (2010)
Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. 11(3), 796–817 (2001)
Lasserre, J.B., Henrion, D., Prieur, C., Trélat, E.: Nonlinear optimal control via occupation measures and LMI relaxations. SIAM J. Control Optim. 47(4), 1643–1666 (2008)
Lasserre, J.B., Pauwels, E.: The empirical christoffel function in statistics and machine learning. arXiv:1701.02886 (2017)
Löfberg, J.: Yalmip: a toolbox for modeling and optimization in Matlab. In: Proceedings of the IEEE CACSD Conference, Taipei, Taiwan (2004)
Magron, V., Henrion, D., Forets, M.: Semidefinite characterization of invariant measures for polynomial systems. Submitted for publication (2018)
Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Springer, Berlin (2012)
Mezić, Igor, Banaszuk, Andrzej: Comparison of systems with complex behavior. Physica D 197, 101–133 (2004)
Ozay, N., Lagoa, C., Sznaier, M.: Set membership identification of switched linear systems with known number of subsystems. Automatica 51, 180–191 (2015)
Ozay, N., Sznaier, M., Lagoa, C.: Convex certificates for model (in) validation of switched affine systems with unknown switches. IEEE Trans. Autom. Control 59(11), 2921–2932 (2014)
Pauwels, E., Lasserre, J. B.: Sorting out typicality with the inverse moment matrix sos polynomial. In: Advances in Neural Information Processing Systems (NIPS) (2016)
Phelps, R.R.: Lectures on Choquet’s Theorem. Springer, Berlin (2001)
Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. 42, 969–984 (1993)
Schlosser, C., Korda, M.: Converging outer approximations to global attractors using semidefinite programming. arXiv preprint arXiv:2005.03346 (2020)
Sturm, J.: Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. Optim. Methods Softw. 11:625–653 (1999)
Tobasco, I., Goluskin, D., Doering, C.: Optimal bounds and extremal trajectories for time averages in dynamical systems. APS M1–002 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Milan Korda\textsuperscript{1,2,3} · Didier Henrion\textsuperscript{1,2,3} · Igor Mezić\textsuperscript{4}

\textsuperscript{1} Milan Korda
korda@laas.fr
Didier Henrion
henrion@laas.fr

Igor Mezić
mezic@engineering.ucsb.edu

1 CNRS; LAAS, 7 Avenue du Colonel Roche, 31400 Toulouse, France
2 Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 16626 Prague, Czech Republic
3 Université de Toulouse, LAAS, 31400 Toulouse, France
4 University of California, Santa Barbara, Santa Barbara, USA