Convergence rates for the moment-SoS hierarchy

Corbinian Schlosser, Matteo Tacchi, Alexey Lazarev

Convergence rates for the moment-SoS hierarchy

Corbinian Schlosser[1], Matteo Tacchi[2], Alexey Lazarev[3]

January 31, 2024

Abstract

We introduce a comprehensive framework for analyzing convergence rates for infinite dimensional linear programming problems (LPs) within the context of the moment-sum-of-squares hierarchy. Our primary focus is on extending the existing convergence rate analysis, initially developed for static polynomial optimization, to the more general and challenging domain of the generalized moment problem. We establish an easy-to-follow procedure for obtaining convergence rates. Our methodology is based on, firstly, a state-of-the-art degree bound for Putinar's Positivstellensatz, secondly, quantitative polynomial approximation bounds, and, thirdly, a geometric Slater condition on the infinite dimensional LP. We address a broad problem formulation that encompasses various applications, such as optimal control, volume computation, and exit location of stochastic processes. We illustrate the procedure at these three problems and, using a recent improvement on effective versions of Putinar's Positivstellensatz, we improve existing convergence rates.

Keywords

Real algebraic geometry; convex optimization; convergence rates; optimal control; stochastic processes, numerical methods for multivariate integration.

Acknowledgments

The work of Corbinian Schlosser was supported by the Polynomial Optimization, Efficiency through Moments and Algebra (POEMA) network and by the European Research Council (grant REAL 947908).

The work of M. Tacchi was funded by the Swiss National Science Foundation under the "NCCR Automation" grant n°51NF40_180545.

The work of Alexey Lazarev has benefited from the AI Interdisciplinary Institute ANITI. ANITI is funded by the French "Investing for the Future - PIA3" program under the Grant agreement n°ANR-19-PI3A-0004.

The authors are very grateful to Lorenzo Baldi for his comments and advice on this work.

Contents

[1] Inria Paris, France. **Corresponding author.**
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP (Institute of Engineering Univ. Grenoble Alpes), GIPSA-lab, France.
[3] Toulouse Institute of Mathematics, France.
[0]

1 Introduction

Context. In recent years various kinds of (nonlinear) problems have been formulated via a specific instance of infinite dimensional linear programming (LP), known as the generalized moment problem (GMP). GMPs appear in various problems coming from diverse fields wuch as geometry, where it has been used in volume computation, for semialgebraic sets [16, 31, 50] as well as sets defined with quantifiers [30], approximation of projections and polynomial images of semialgebraic sets [34], as well as set separation a la Urysohn [24]. Other examples arise from dynamical systems and include optimal control [32], stability analysis [20, 18, 40, 21, 17], localisation of global attractors [13, 44], and more generally calculus of variations [15], as well as partial differential equations [35]. Due to the natural role of Borel measures in the GMP, it is also widely used to study stochastic systems, with applications to exit location [14], infinite time averaging [11], computing invariant measures [23], peak value-at-risk [37] and probability of unsafety [36].

In the framework of the GMP approach the problems mentioned above can be represented by dual pairs of linear programming problems: the primal problem (measure LP) is defined on the infinite dimensional space of Borel measures and features (possibly infinitely many) linear equality constraints; the dual problem (function LP) is defined on a (possibly infinite dimensional) vector space of polynomials and features infinite dimensional (functional) linear inequality constraints. The moment-sum-of-squares (moment-SoS) hierarchy is a two steps procedure that provides a powerful tool for tackling such linear programming problems and

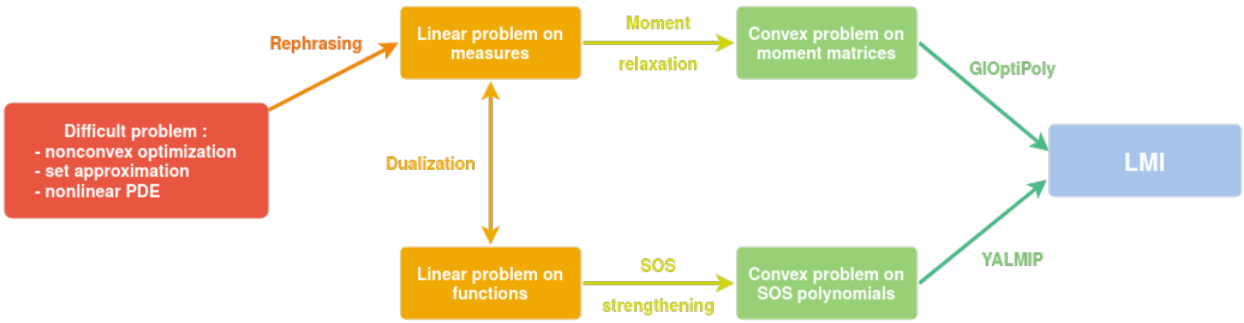

Figure 1: The standard application of the moment-SoS hierarchy.

has been applied widely. In the first step, the primal decision variables, i.e. Borel measures, are represented by their moments, that are characterized via linear matrix inequality (LMI) constraints on the so-called moment matrices. In the second step, the moment matrices are truncated, i.e. only moments up to a finite degree $\ell \in \mathbb{N}$ are considered and paired with the constraints in the LP; resulting in a hierarchy of semidefinite programs. This operation is referred to as *moment relaxation*. By duality, this procedure leads to a tightening in the function LP. There, the inequality constraints of the function LP are first strengthened to SoS constraints (by the aid of Putinar's Positivstellensatz [41]). In the second step, the degree of the SoS polynomials is truncated, thus obtaining a hierarchy of so-called *SoS strengthening*. The scheme of the moment-SoS hierarchy approach is summarized in Figure 1.

The moment-SoS hierarchy provides guaranteed convergence [28, 38, 29, 48] but, apart from the case of polynomial optimization, the speed of the convergence for the infinite dimensional problems has been rarely investigated. Two examples where explicit convergence rates were derived can be found in [22] and [19] where a slow convergence rate was presented based on [39]. Since then, the problem of providing bounds on the minimal truncation required to fully represent positive polynomials as sums of squares has been deeply studied, both in the generic case [2, 3] and in specific settings [46, 4].

Contribution. The speed of convergence can be derived from such bounds, and is determined by two main factors:

1. The regularity of optimal solutions to the function LP and, if they are not polynomial, their approximation with polynomials;
2. The degree $\ell \in \mathbb{N}$ needed for an SoS representation of the approximating polynomials.

We treat those two principal concepts in Section 3.2 and aim at describing an interplay between results on degree bounds in Putinar's Positivstellensatz, structural approximation properties for polynomials and compatibility conditions of the LPs concerning polynomials.

The main objective of this article is to provide a method for deriving convergence rates of the moment-SoS hierarchy when applied to a specific instance of the GMP, using the degree bounds provided in [2]. We provide examples of computing and improving the convergence rates of the hierarchy with the state-of-art versions of Putinar's Positivstellensatz. Namely, we use [2] to improve the convergence rates for the optimal control problem stated in [22] and for the standard volume problem [19]. Additionally, we derive an original convergence rate for the problem of exit location of stochastic processes [14]. Last but not least, we use our methodology to answer a long-standing question related to volume computation, namely: how much does the use of Stokes' theorem improve the moment-SoS hierarchy for volume computation? Indeed, the first application of the moment-SoS hierarchy to this problem in [16] exhibited a very slow convergence in practice, and was soon complemented with [31], yielding a sharp improvement in the numerical accuracy of the relaxations. To further understand this improvement, a qualitative study [49] showed that it was related to the Gibbs phenomenon and regularity of solutions in the function LP. In this work we complement the qualitative study with a first *quantitative* analysis of the two formulations, by computing and comparing the convergence rates in both cases.

As a consequence, in most examples covered in this work we get a convergence rate of $\mathcal{O}(\ell^{-1/c})$ for some constant $c > 0$ (the only exception being generic optimal control, for which the rate is $\mathcal{O}\left(1/\log \ell\right)$, see **Corollary 4.6**, although mild assumptions allowed us to bring back a polynomial convergence rate in **Theorem 4.10**), which is a significant improvement compared to the **double** log bounds obtained in [22, 19].

Outline. The paper is structured as follows: In Section 2, we fix the notation and focus on the central underlying concept of moment-SoS hierarchy for the generalized moment problem (GMP). Section 3 recalls the current state of an effective version of Putinar's Positivstellensatz and we introduce and motivate our general procedure for obtaining effective degree bounds for SoS tightening of the infinite dimensional function LPs. In Sections 4 and 5 we apply the procedure to establish convergence rates for old and new instances of the moment-SoS hierarchy and where strong improvements compared to the existing rates are demonstrated. Section 4 treats dynamic settings, where optimal control for deterministic systems and the exit location problem of stochastic processes are considered. Section 5 is concerned with volume computation with and without the aid of reinforcing Stokes constraints.

2 Preliminaries: the moment-SoS hierarchy

2.1 Basic notations

We work with the standard notations for usual sets $\mathbb{R}$ (real numbers), $\mathbb{Z}$ (integers), $\mathbb{N}$ (natural integers), for which the superscript * indicates that we remove the element 0. Real intervals are denoted $[a,b]$ when closed, (a,b) when open; integer intervals are denoted $[\![a,b]\!]$ (with particular case $[\![n]\!] = [\![1,n]\!]$ for $n \in \mathbb{N}^\star$). For $x \in \mathbb{R}$, $\lfloor x \rfloor := \max([x-1,x] \cap \mathbb{Z})$ denotes the floor and $\lceil x \rceil := \min([x,x+1] \cap \mathbb{Z})$ denotes the ceiling.

For a topological space $\mathcal{X}$, $\mathcal{C}(\mathcal{X})$ denotes the space of continuous functions from $\mathcal{X}$ to $\mathbb{R}$ equipped with the topology of uniform convergence. For two real vector spaces $\mathbb{V}$, $\mathbb{W}$, the set $\mathbb{L}(\mathbb{V},\mathbb{W})$ denotes the space of linear maps from $\mathbb{V}$ to $\mathbb{W}$. For a real Banach space $\mathbb{V}$, define the dual space $\mathbb{V}' := \mathbb{L}(\mathbb{V},\mathbb{R}) \cap \mathcal{C}(\mathbb{V})$, with duality $\langle v, v' \rangle \in \mathbb{R}$, $v \in \mathbb{V}, v' \in \mathbb{V}'$. In particular, for a compact Hausdorff space $\mathcal{X}$, the space of signed Radon measures $\mathcal{M}(\mathcal{X})$ is identified with $\mathcal{C}(\mathcal{X})'$.

For $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, $|\boldsymbol{\alpha}| := \alpha_1 + \ldots + \alpha_n$ is the range of $\boldsymbol{\alpha}$ and $(x_1, \ldots, x_n) = \mathbf{x} \mapsto \mathbf{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is the corresponding monomial. For $n, d \in \mathbb{N}$, $\mathbb{N}^n_d := \{\boldsymbol{\alpha} \in \mathbb{N}^n \,; |\boldsymbol{\alpha}| \leq d\}$ is the set of bounded multi-indices, $\mathbb{R}_d[\mathbf{x}] := \{\mathbf{x} \mapsto \sum_{|\boldsymbol{\alpha}| \leq d} c_{\boldsymbol{\alpha}}\, \mathbf{x}^{\boldsymbol{\alpha}} \,; (c_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}} \in \mathbb{R}^{\mathbb{N}^n_d}\}$ is the space of degree at most d polynomial functions, $\mathbb{R}[\mathbf{x}] := \cup_{d \in \mathbb{N}} \mathbb{R}_d[\mathbf{x}]$ is the space of polynomials. For $\boldsymbol{\Omega} \subset \mathbb{R}^n$, $\mathcal{P}_d(\boldsymbol{\Omega}) := \mathcal{C}(\boldsymbol{\Omega}) \cap \mathbb{R}_d[\mathbf{x}]$ and $\mathcal{P}(\boldsymbol{\Omega}) := \mathcal{C}(\boldsymbol{\Omega}) \cap \mathbb{R}[\mathbf{x}]$.

If $\mathbb{V}$ is equipped with a set of "entry" forms $\{\delta_i\}_{i \in \mathcal{I}} \subset \mathbb{V}'$ (e.g. dual basis $\{\mathbf{x} \mapsto x_i\}_{i \in [\![n]\!]}$ in $\mathbb{R}^n$, or evaluation functionals $\{f \mapsto f(\mathbf{x})\}_{\mathbf{x} \in \boldsymbol{\Omega}}$ in function spaces over a set $\boldsymbol{\Omega}$), then the nonnegative (resp. positive) cone of $\mathbb{V}$ is $\mathbb{V}_+ := \{v \in \mathbb{V} \,; \forall i \in \mathcal{I}, \langle v, \delta_i \rangle \geq 0\}$ (resp. $\mathbb{V}_\oplus := \{v \in \mathbb{V} \,; \forall i \in \mathcal{I}, \langle v, \delta_i \rangle > 0\}$), as in $\mathbb{R}_\oplus$, $\mathbb{R}^n_+$, $\mathcal{C}(\mathbf{X})_\oplus$. In particular, the dual cone of $\mathbb{K} \subset \mathbb{V}$ is $\mathbb{K}' := (\mathbb{V}')_+$ for entry forms $\{\langle \kappa, \bullet \rangle\}_{\kappa \in \mathbb{K}}$ (e.g. $\mathcal{M}(\mathbf{X})_+ = \mathcal{C}(\mathbf{X})'_+$).

$\mathbb{S}^n = \{\mathrm{M} \in \mathbb{R}^{n \times n} \,; \mathrm{M}^\top = \mathrm{M}\}$ is the vector space of symmetric real matrices of size n with entry forms $\{\mathrm{M} \mapsto \mathbf{x}^\top \mathrm{M}\, \mathbf{x}\}_{\mathbf{x} \in \mathbb{R}^n}$, so that $\mathbb{S}^n_+$ is the usual p.s.d. cone. For $\mathbf{h} = (h_1, \ldots, h_r) \in \mathbb{R}[\mathbf{x}]^r$, define the basic semialgebraic set $\mathbf{S}(\mathbf{h}) := \mathbf{h}^{-1}(\mathbb{R}^r_+)$.

We make the convention that the **Assumptions 1, 2** and **3** hold throughout the whole paper. **Conditions** hold only when explicitly stated.

2.2 Generalized Moment Problem

Let $M, N \in \mathbb{N}^\star$, $\mathbf{r} = (r_1, \ldots, r_M) \in (\mathbb{N}^\star)^M$, $\mathbf{m} = (m_1, \ldots, m_M) \in \mathbb{N}^M$, $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{N}^N$.
For $i \in [\![M]\!]$, let $\mathbf{X}_i := \mathbf{S}(\mathbf{h}_i) \Subset \mathbb{R}^{m_i}$ be a compact basic semialgebraic set with $\mathbf{h}_i \in \mathbb{R}[\mathbf{x}_i]^{r_i}$.

For $j \in [\![N]\!]$, let $\mathbf{Y}_j \Subset \mathbb{R}^{n_j}$ be a compact subset. Let:

$$\begin{aligned} \mathcal{X} &:= \mathcal{M}(\mathbf{X}_1) \times \ldots \times \mathcal{M}(\mathbf{X}_M), & \mathcal{Y} &= \mathcal{P}(\mathbf{Y}_1) \times \ldots \times \mathcal{P}(\mathbf{Y}_N), \\ \mathcal{X}' &:= \mathcal{C}(\mathbf{X}_1) \times \ldots \times \mathcal{C}(\mathbf{X}_M), & \mathcal{Y}' &:= \mathcal{P}(\mathbf{Y}_1)' \times \ldots \times \mathcal{P}(\mathbf{Y}_N)'. \end{aligned}$$

We equip $\mathcal{X}, \mathcal{X}'$ with the product topology. For $\mathcal{Y}$ we use a well-chosen norm $\|\cdot\|_{\mathcal{Y}}$ (see the discussion around equation (16) in Section 3.2) and $\mathcal{Y}'$ is its topological dual. For $\mathbf{v} = (v_1, \ldots, v_M) \in \mathcal{X}'$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M) \in \mathcal{X}$, we define the vector integral as

$$\int \mathbf{v} \cdot \mathrm{d}\boldsymbol{\mu} := \sum_{i=1}^{M} \int v_i \, \mathrm{d}\mu_i.$$

Let $\mathcal{A} : \mathcal{X} \longrightarrow \mathcal{Y}'$ be a continuous linear map, $\mathbf{T} \in \mathcal{Y}'$ be a vector of continuous linear forms (i.e. *moment* sequences), $\mathbf{g} \in \mathbb{R}[\mathbf{x}_1] \times \ldots \times \mathbb{R}[\mathbf{x}_M] \subset \mathcal{X}'$ be a vector of polynomials. We define the Generalized Moment Problem (GMP) as

$$\boxed{\begin{aligned} \mathrm{p}^\star_{\mathrm{GM}} = \quad & \sup_{\boldsymbol{\mu} \in \mathcal{X}} \quad \int \mathbf{g} \cdot \mathrm{d}\boldsymbol{\mu} \\ & \text{s.t.} \quad \forall i \in [\![M]\!], \quad \mu_i \in \mathcal{M}(\mathbf{X}_i)_+ \\ & \qquad\;\; \mathcal{A}\,\boldsymbol{\mu} = \mathbf{T} \end{aligned}} \tag{1}$$

Remark 2.1 (On the generality of (1))**.**

Note that the generic framework $\mathbf{X} = \mathbf{S}(\mathbf{h})$ allows for a finite (discrete) set $\mathbf{X}$, for which measures can be represented as vectors. In particular, binary optimization has a GMP formulation which is proved to be equivalent to semidefinite programming [26]. From this simple remark, one can observe that problem (1) can feature finite size PSD constraints on moments, as proposed in [37]. Hence, the tools displayed in this work can be used on GMPs featuring finite size LP, convex QP, SOCP, and SDP constraints on moments.

Throughout this text, we always make the following assumption:

Assumption 1 (Existence of feasible solutions)**.**

The feasible set of (1) is not empty.

Now that the problem has been introduced in full generality, we quickly drop some notational burden by setting $M = N = 1$, without loss of generality: now $\mathcal{X} = \mathcal{M}(\mathbf{X})$ with $\mathbf{X} = \mathbf{S}(\mathbf{h})$, $\mathcal{X}' = \mathcal{C}(\mathbf{X})$, $\mathcal{Y} = \mathcal{P}(\mathbf{Y})$, $\mathcal{Y}' = \mathcal{P}(\mathbf{Y})'$ so that the problem rewrites as a "simple" moment problem[1]:

$$\boxed{\begin{aligned} \mathrm{p}^\star_{\mathrm{M}} = \quad & \sup_{\mu \in \mathcal{M}(\mathbf{X})} \quad \int g \, \mathrm{d}\mu \\ & \text{s.t.} \quad \mu \in \mathcal{M}(\mathbf{X})_+ \\ & \qquad\;\; \mathcal{A}\,\mu = T. \end{aligned}} \tag{2}$$

Remark 2.2 (Existing GMPs)**.**

The framework of GMPs covers a large class of problems, notably static polynomial optimization [27], but also the LPs from [49, 32, 40, 21, 44, 23, 8, 17, 15], to name only a few, can all be represented in the form (1).

[1] The case where $g = 0$, $\mathbf{X} = \mathbf{Y}$ and $\mathcal{A}\,\mu = (\int \mathbf{x}^{\boldsymbol{\alpha}} \, \mathrm{d}\mu)_{\boldsymbol{\alpha} \in \mathbb{N}^m}$ is called the $\mathbf{X}$-moment problem.

The GMP is an infinite dimensional instance of conically constrained linear programs (CCLP), and as such it is subject to Lagrange duality. To write the dual problem, we introduce the Lagrange operator

$$\Lambda := \mathcal{M}(\mathbf{X}) \times \mathcal{P}(\mathbf{Y}) \ni (\mu, w) \longmapsto \int g \, \mathrm{d}\mu + \langle \underbrace{T - \mathcal{A}\,\mu}_{T_\mu}, w \rangle, \tag{3}$$

and it is straightforward (using the fact that $\inf_{w \in \mathcal{P}(\mathbf{Y})} \langle T_\mu, w \rangle = -\infty$ iff $T_\mu \neq 0$) that

$$\mathrm{p}^\star_{\mathrm{M}} = \sup \left\{ \inf \left\{ \Lambda(\mu, w) \,;\, w \in \mathcal{P}(\mathbf{Y}) \right\} \,;\, \mu \in \mathcal{M}(\mathbf{X})_+ \right\}.$$

Finally, the dual problem to (2) is obtained by swapping the sup and inf operators:

$$\mathrm{d}^\star_{\mathrm{M}} = \inf \left\{ \sup \left\{ \Lambda(\mu, w) \,;\, \mu \in \mathcal{M}(\mathbf{X})_+ \right\} \,;\, w \in \mathcal{P}(\mathbf{Y}) \right\}$$

i.e., defining the adjoint operator $\mathcal{A}' : \mathcal{P}(\mathbf{Y}) \longrightarrow \mathcal{C}(\mathbf{X})$ such that for $\mu \in \mathcal{M}(\mathbf{X})$ and $w \in \mathcal{P}(\mathbf{Y})$,

$$\int \mathcal{A}'\, w \, \mathrm{d}\mu = \langle \mathcal{A}\,\mu, w \rangle$$

so that $\Lambda(\mu, w) = \displaystyle\int (g - \mathcal{A}'\, w) \, \mathrm{d}\mu + \langle T, w \rangle$ (and again $\sup_{\mu \in \mathcal{M}(\mathbf{X})_+} \int (g - \mathcal{A}'\, w) \, \mathrm{d}\mu = +\infty$ iff $\mathcal{A}'\, w - g \notin \mathcal{C}(\mathbf{X})_+$):

$$\boxed{\begin{array}{lll} \mathrm{d}^\star_{\mathrm{M}} = & \displaystyle\inf_{w \in \mathcal{P}(\mathbf{Y})} & \langle T, w \rangle \\ & \text{s.t.} & \mathcal{A}'\, w - g \in \mathcal{C}(\mathbf{X})_+. \end{array}} \tag{2'}$$

From this, one can deduce the fully general dual problem by taking generic M and N:

$$\boxed{\begin{array}{lll} \mathrm{d}^\star_{\mathrm{GM}} = & \displaystyle\inf_{\mathbf{w} \in \mathcal{Y}'} & \langle \mathbf{T}, \mathbf{w} \rangle \\ & \text{s.t.} & \forall i \in [\![M]\!], \quad (\mathcal{A}'\, \mathbf{w})_i - g_i \in \mathcal{C}(\mathbf{X}_i)_+. \end{array}} \tag{1'}$$

This duality between (1) and (1′) comes with two interesting properties [5, 48], which we state next.

Proposition 2.3 (Weak duality)**.**

In all generality, with the above notations, one has $\mathrm{p}^\star_{\mathrm{GM}} \leq \mathrm{d}^\star_{\mathrm{GM}}$.

Proposition 2.4 (Strong duality)**.**

One has $\mathrm{p}^\star_{\mathrm{GM}} = \mathrm{d}^\star_{\mathrm{GM}}$ *if one of the following two* **Conditions** *is satisfied.*

Condition 1 (Slater [45])**.** $\exists \mathring{\mathbf{w}} \in \mathcal{Y}$ s.t. $\forall i \in [\![M]\!]$, $(\mathcal{A}'\, \mathring{\mathbf{w}})_i - g_i \in \mathcal{C}(\mathbf{X}_i)_\oplus$.

Condition 2 (Primal compactness [47])**.** $\exists B > 0$ s.t. $\forall \boldsymbol{\mu}$ feasible for (1), one has $\forall i \in [\![M]\!]$, $\displaystyle\int 1 \, \mathrm{d}\mu_i \leq B$.

Remark 2.5 (On the links between **Conditions 1** and **2**)**.**

Condition 1 is instrumental in numerically constructing approximate solutions of (1′), while **Condition 2** is used in [48] to prove a strong convergence result on the numerical approximation of (1). Ideally, one would like to deduce both conditions from one, stronger condition.

If for $i \in [\![M]\!]$, we have $g_i \in \mathcal{P}(\mathbf{X}_i)_+$, then **Condition 1** implies **Condition 2**. Indeed, assuming **Condition 1**, one gets a Slater point $\mathring{\mathbf{w}}$ such that $\forall i \in [\![M]\!]$, $(\mathcal{A}'\mathring{\mathbf{w}})_i > g_i \geq 0$ on $\mathbf{X}_i$. But then, denoting $\gamma_i^\star := \min_{\mathbf{X}_i} (\mathcal{A}'\mathring{\mathbf{w}})_i > 0$ and $\gamma^\star := \min_{1 \leq i \leq M} \gamma_i^\star > 0$, and taking a feasible $\boldsymbol{\mu}$ for (1), one gets

$$\begin{aligned} \sum_{i=1}^{M} \gamma_i^\star \int 1 \, \mathrm{d}\mu_i \leq \sum_{i=1}^{M} \int (\mathcal{A}'\mathring{\mathbf{w}})_i \, \mathrm{d}\mu_i &= \int \mathcal{A}'\mathring{\mathbf{w}} \cdot \mathrm{d}\boldsymbol{\mu} \\ &= \langle \mathcal{A}\,\boldsymbol{\mu}, \mathring{\mathbf{w}} \rangle = \langle \mathbf{T}, \mathring{\mathbf{w}} \rangle < \infty \end{aligned}$$

and, a sum of nonnegative terms being bigger than any of its terms, one deduces that for all $i \in [\![M]\!]$, it holds

$$\int 1 \, \mathrm{d}\mu_i \leq \frac{\langle \mathbf{T}, \mathring{\mathbf{w}} \rangle}{\gamma_i^\star} \leq \frac{\langle \mathbf{T}, \mathring{\mathbf{w}} \rangle}{\gamma^\star} =: B \in (0, \infty)$$

which is exactly **Condition 2**. Thus, up to a shift on $\mathbf{g}$, **Condition 1** implies **Condition 2**.

In this text, we follow an established line of reasoning for solving the GMP (1) and its dual (1′) via moment relaxations and semidefinite programming. This technique is often called the moment-SoS hierarchy, which we recall in the following sections.

2.3 A motivating example

Consider the problem of computing the Lebesgue volume denoted $\lambda(\mathbf{X})$ of a bounded basic semi-algebraic set

$$\mathbf{X} := \mathbf{S}(\mathbf{h}) = \{\mathbf{x} \in \mathbb{R}^m \,; h_1(\mathbf{x}) \geq 0, \ldots, h_r(\mathbf{x}) \geq 0\}$$

with $m, r \geq 1$ integers and $h_1, \ldots, h_r \in \mathbb{R}[\mathbf{x}]$. For this, we suppose that $\mathbf{X}$ satisfies[2]

$$\mathbf{X} \subset \mathbf{B} := \{\mathbf{x} \in \mathbb{R}^m \,; \mathbf{x}^\top \mathbf{x} \leq 1\}.$$

The standard moment-SoS approach to numerically solve the volume problem is discussed in detail in [16]. The method consists of formulating a GMP whose optimal solution is $\lambda(\mathbf{X})$, after which one numerically approximates this optimal solution using the moment-SoS hierarchy that we are now going to introduce. We first notice that denoting $\lambda_\mathbf{B}$ the Lebesgue measure on $\mathbf{B}$, it holds $\lambda(\mathbf{X}) = \lambda_\mathbf{B}(\mathbf{X})$ (because $\mathbf{X} \subset \mathbf{B}$), and there exists a closed formula (in terms of Euler's Γ function) for the Lebesgue moments on the unit ball $\mathbf{B}$

$$\langle T, \mathbf{x}^{\boldsymbol{\alpha}} \rangle := \int \mathbf{x}^{\boldsymbol{\alpha}} \, \mathrm{d}\lambda_\mathbf{B}(\mathbf{x}).$$

Hence, one can write the following GMPs

$$\begin{aligned} \mathrm{p}^\star_\mathbf{X} := \max_{\substack{\mu \in \mathcal{M}(\mathbf{X})_+ \\ \nu \in \mathcal{M}(\mathbf{B})_+}} & \int 1 \, \mathrm{d}\mu \\ \text{s.t. } & \mu + \nu = \lambda_\mathbf{B} \end{aligned} \qquad (4)$$

$$\begin{aligned} \mathrm{d}^\star_\mathbf{X} := \inf_{w \in \mathcal{P}(\mathbf{B})} & \int w \, \mathrm{d}\lambda_\mathbf{B} \\ \text{s.t. } & w|_\mathbf{X} - 1 \in \mathcal{C}(\mathbf{X})_+ \\ & w \in \mathcal{C}(\mathbf{B})_+ \end{aligned} \qquad (4')$$

and it is straightforward to find that $\mathrm{p}^\star_\mathbf{X} = \lambda(\mathbf{X})$ with optimal solution $\mu^\star = \mathbb{1}_\mathbf{X} \, \lambda_\mathbf{B}$ the Lebesgue measure restricted to $\mathbf{X}$, $\nu^\star = (1 - \mathbb{1}_\mathbf{X}) \, \lambda_\mathbf{B}$, where $\mathbb{1}_\mathbf{X}$ stands for the indicator function of $\mathbf{X}$, which takes value 1 on $\mathbf{X}$ and 0 elsewhere. We are going to illustrate the moment-SoS hierarchy using (4′) as a motivating example. First, notice that the infimum in (4′) is not attained; a preliminary recasting can be done to tackle this issue:

$$\begin{aligned} \lambda(\mathbf{X}) = \max_{u, v \in L^\infty(\mathbf{B})_+} & \int u \, \mathrm{d}\lambda_\mathbf{B} \\ \text{s.t. } & u|_{\mathbf{B} \setminus \mathbf{X}} = 0 \\ & u + v = 1 \end{aligned} \qquad (5)$$

$$\begin{aligned} \lambda(\mathbf{X}) = \min_{w \in L^1(\mathbf{B})} & \int w \, \mathrm{d}\lambda_\mathbf{B} \\ \text{s.t. } & w|_\mathbf{X} - 1 \in L^1(\mathbf{X})_+ \\ & w \in L^1(\mathbf{B})_+ \end{aligned} \qquad (5')$$

where $L^\infty(\mathbf{B})$ is the Banach space of Lebesgue essentially bounded functions on $\mathbf{B}$ and, its pre-dual, $L^1(\mathbf{B})$ is the space of Lebesgue integrable functions on $\mathbf{B}$ (both factored by the equivalence $u \equiv v \iff u = v$ λ-a.e.) and for $s \in \{1, \infty\}$,

$$L^s(\mathbf{B})_+ = \{v \in L^s(\mathbf{B}) \,; v \geq 0 \ \lambda\text{-a.e.}\}.$$

[2] Up to a rescaling, the inclusion condition is equivalent to $\mathbf{X}$ being compact.

Now, both primal and dual have optimal solutions, $u^\star = w^\star = \mathbb{1}_{\mathbf{X}}$ and $v^\star = 1 - \mathbb{1}_{\mathbf{X}}$. Our preliminary step is formulated as follows:

Step 0: By density of $\mathcal{P}(\mathbf{B})$ in $L^1(\mathbf{B})$ with respect to the norm $\|u\|_{L^1(\mathbf{B})} := \int |u| \,\mathrm{d}\lambda_{\mathbf{B}}$, it holds $\mathrm{d}^\star_{\mathbf{X}} = \lambda(\mathbf{X})$ and minimizing sequences for (4′) approximate $\mathbb{1}_{\mathbf{X}}$ from above w.r.t. $\|\cdot\|_{L^1(\mathbf{B})}$:

$$\forall \varepsilon > 0, \exists w_\varepsilon \in \mathcal{P}(\mathbf{B}) \text{ feasible for (4′) and such that } \lambda(\mathbf{X}) < \int w_\varepsilon \,\mathrm{d}\lambda_{\mathbf{B}} < \lambda(\mathbf{X}) + \varepsilon.$$

In the following steps we recall how to recast (4′) as a hierarchy of finite dimensional convex problems that approximate $\lambda(\mathbf{X})$.

Step 1: Noticing that for $w \in \mathbb{R}[\mathbf{x}]$, $w|_{\mathbf{X}} - 1 \in \mathcal{P}(\mathbf{X})$, it is possible to replace $\mathcal{C}(\mathbf{X})_+$ and $\mathcal{C}(\mathbf{B})_+$ with $\mathcal{P}(\mathbf{X})_+$ and $\mathcal{P}(\mathbf{B})_+$ in (4′), without changing its optimal value, yielding

$$\begin{aligned} \lambda(\mathbf{X}) = \inf_{w \in \mathbb{R}[\mathbf{x}]} & \int w \,\mathrm{d}\lambda_{\mathbf{B}} = \langle T, w\rangle \\ \text{s.t. } & w|_{\mathbf{X}} - 1 \in \mathcal{P}(\mathbf{X})_+ \\ & w \in \mathcal{P}(\mathbf{B})_+. \end{aligned} \tag{6a}$$

Step 2: Second, we notice that Slater's **Condition 1** holds with $\mathring{w} := 2$, so that we can look for a minimizing sequence of *strictly feasible* polynomials w:

$$\begin{aligned} \lambda(\mathbf{X}) = \inf_{w \in \mathbb{R}[\mathbf{x}]} & \quad \langle T, w\rangle \\ \text{s.t. } & w|_{\mathbf{X}} - 1 \in \mathcal{P}(\mathbf{X})_\oplus \\ & w \in \mathcal{P}(\mathbf{B})_\oplus. \end{aligned} \tag{6b}$$

Step 3: Next, we use a real algebraic geometry theorem that we will introduce in the next section, to recast positivity constraints using sums of squares of polynomials:

$$\begin{aligned} \lambda(\mathbf{X}) = \inf_{\substack{w, \overline{\sigma_0}, \overline{\sigma_1} \in \mathbb{R}[\mathbf{x}] \\ \boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_{r+1})^T \in \mathbb{R}[\mathbf{x}]^{r+1}}} & \quad \langle T, w\rangle \\ \text{s.t. } & w = 1 + \begin{pmatrix} \mathbf{h}^\top & 1 \end{pmatrix} \boldsymbol{\sigma} \\ & w = \overline{\sigma_0} + (1 - \mathbf{x}^\top \mathbf{x}) \overline{\sigma_1} \\ & \overline{\sigma_0}, \overline{\sigma_1}, \sigma_1, \ldots, \sigma_{r+1} \text{ are sums of squares.} \end{aligned} \tag{6c}$$

Step 4: Eventually, we impose a bound $\ell \in \mathbb{N}$ on the degree of the involved polynomials to get a convex, SDP-representable optimization problem (see [28, Proposition 2.1] for SDP-representability of sum-of-squares constraints):

$$\begin{aligned} \mathrm{d}^\ell_{\mathbf{X}} := \inf_{\substack{w, \overline{\sigma_0}, \overline{\sigma_1} \in \mathbb{R}[\mathbf{x}] \\ \boldsymbol{\sigma} \in \mathbb{R}[\mathbf{x}]^{r+1}}} & \quad \langle T, w\rangle \\ \text{s.t. } & w = 1 + \begin{pmatrix} \mathbf{h}^\top & 1 \end{pmatrix} \boldsymbol{\sigma} = \overline{\sigma_0} + (1 - \mathbf{x}^\top \mathbf{x}) \overline{\sigma_1} \\ & \overline{\sigma_0}, \overline{\sigma_1}, \sigma_1, \ldots, \sigma_{r+1} \text{ are sums of squares} \\ & \max\{\deg(w), \deg(\overline{\sigma_0}), \deg(\overline{\sigma_1}) + 2\} \leq 2\ell \\ & \max\{\deg(h_1\,\sigma_1), \ldots, \deg(h_r\,\sigma_r), \deg(\sigma_{r+1})\} \leq 2\ell. \end{aligned} \tag{6d}$$

Note that (6d) is a *tigthening* of (6c) in the sense that we replaced the feasible set with a strictly smaller one (even finite dimensional): in this instance, it holds

$$\mathrm{d}^\ell_{\mathbf{X}} > \lambda(\mathbf{X}).$$

Moreover, increasing the degree bound ℓ increases the size of the feasible set for (6d), so that the following monotone convergence theorem holds for free:

$$\mathrm{d}_{\mathbf{X}}^{\ell} \underset{\ell\to\infty}{\searrow} \lambda(\mathbf{X}).$$

Step 5: It is possible to write a Lagrangian dual to problem (6d), which can be proved to be a finite dimensional *relaxation* of problem (4).

The moment-SoS hierarchy systematizes this process for generic instances of (1) and (1′).

2.4 The moment-SoS hierarchy

The moment-SoS hierarchy builds on real algebraic geometry results to formulate a sequence of finite dimensional convex optimization problems that approximate (1) and (1′):

Theorem 2.6 (Putinar's Positivstellensatz [41, Theorem 1.3 & Lemma 3.2])**.**

Let $r, m \in \mathbb{N}^\star$ be positive integers, $\mathbf{h} \in \mathbb{R}[\mathbf{x}]^r$ a family of r polynomials in m variables.

Introduce the closed semialgebraic set $\mathbf{S} := \left\{\mathbf{x} \in \mathbb{R}^m \,; \mathbf{h}(\mathbf{x}) \in \mathbb{R}_+^r\right\}$ as well as the **convex cones**

$$\begin{array}{ll} \mathcal{P}(\mathbf{S})_+ := \{p \in \mathbb{R}[\mathbf{x}] \,; p(\mathbf{S}) \subset \mathbb{R}_+\} & \Sigma[\mathbf{x}] := \left\{\sum_{k=1}^K p_k^2 \,; K \in \mathbb{N}^\star, p_1, \ldots, p_K \in \mathbb{R}[\mathbf{x}]\right\} \\ \mathcal{P}(\mathbf{S})_\oplus := \{p \in \mathbb{R}[\mathbf{x}] \,; p(\mathbf{S}) \subset \mathbb{R}_\oplus\} & \mathcal{Q}(\mathbf{h}) := \left\{\begin{pmatrix}\mathbf{h}^\top & 1\end{pmatrix} \boldsymbol{\sigma} \,; \boldsymbol{\sigma} \in \Sigma[\mathbf{x}]^{r+1}\right\} \subset \mathcal{P}(\mathbf{S})_+. \end{array}$$

If there exists $R > 0$ s.t. $R^2 - \mathbf{x}^\top\mathbf{x} \in \mathcal{Q}(\mathbf{h})$ (Archimedean property), then $\mathcal{P}(\mathbf{S})_\oplus \subset \mathcal{Q}(\mathbf{h})$.

Under the same Archimedean condition, the dual cones $\mathcal{Q}(\mathbf{h})'$ and $\mathcal{M}(\mathbf{S})_+$ are isomorphic.

Remark 2.7 (On the Archimedean condition)**.**

As $\mathcal{Q}(\mathbf{h}) \subset \mathcal{P}(\mathbf{S})_+$, the Archimedean property automatically yields that

$$\mathbf{S} \subset \mathbf{B}_R := \left\{\mathbf{x} \in \mathbb{R}^m \,; \mathbf{x}^\top\mathbf{x} \leq R^2\right\},$$

i.e. $\mathbf{S}$ is bounded (and thus compact as it is closed). Conversely, if $\mathbf{S} \subset \mathbf{B}_R$ for some $R > 0$, then adding a polynomial $h_{r+1} := R^2 - \mathbf{x}^\top\mathbf{x}$ to $\mathbf{h}$ does not change the geometry of $\mathbf{S}$, while it results in adding h_{r+1} to $\mathcal{Q}(\mathbf{h})$. Thus, in practice, the Archimedean condition is considered equivalent to compactness of $\mathbf{S}$.

Eventually, to properly formulate the moment-SoS hierarchy, we make a last assumption on $\mathcal{A}$.

Assumption 2 (Moment operator)**.**

$\mathcal{A}$ is a moment operator, which writes as follows in the setting $M = N = 1$ of (2):

There exists a sequence of polynomials $(\varphi_{\boldsymbol{\beta}})_{\boldsymbol{\beta}} \in \mathcal{P}(\mathbf{X})^{\mathbb{N}^n}$ such that, for all $\mu \in \mathcal{M}(\mathbf{X})$, $\mathcal{A}\,\mu$ is the linear operator defined on the monomial basis by

$$\langle \mathcal{A}\,\mu, \mathbf{y}^{\boldsymbol{\beta}} \rangle = \int \varphi_{\boldsymbol{\beta}} \,\mathrm{d}\mu. \tag{7}$$

Remark 2.8 (Generic moment operator)**.**

The generic case $M \neq 1$, $N \neq 1$ is identical to the above statement, where the sequence of polynomials lies in $\big(\mathcal{P}(\mathbf{X}_1) \times \ldots \times \mathcal{P}(\mathbf{X}_M)\big)^{\mathbb{N}^{n_1} \times \ldots \times \mathbb{N}^{n_N}}$ instead of $\mathcal{P}(\mathbf{X})^{\mathbb{N}^n}$.

Lemma 2.9. Assumption 2*, is equivalent to supposing that $\mathcal{A}'$ maps polynomials onto polynomials:*

$$\mathrm{Im}(\mathcal{A}') \subset \mathcal{P}(\mathbf{X}_1) \times \ldots \times \mathcal{P}(\mathbf{X}_M). \tag{8}$$

Proof. To avoid unnecessary technicalities and index notations, we restrict to the case $M = N = 1$ of (2). The generalization to (1) is straightforward. We first prove (7) $\Longrightarrow$ (8). By linearity of $\mathcal{A}'$, it is sufficient to prove that for all $\boldsymbol{\beta} \in \mathbb{N}^n$, $\mathcal{A}'[\mathbf{y}^{\boldsymbol{\beta}}] \in \mathcal{P}(\mathbf{X})$. Let $\boldsymbol{\beta} \in \mathbb{N}^n$, $\mathbf{x} \in \mathbf{X}$ and consider the Dirac measure $\mu = \delta_{\mathbf{x}}$. By **Assumption 2** and by construction of the adjoint operator, one has

$$\begin{aligned}\mathcal{A}'[\mathbf{y}^{\boldsymbol{\beta}}](\mathbf{x}) &= \int \mathcal{A}'[\mathbf{y}^{\boldsymbol{\beta}}]\,\mathrm{d}\mu \\ &= \langle \mathcal{A}\,\mu, \mathbf{y}^{\boldsymbol{\beta}}\rangle \\ &= \varphi_{\boldsymbol{\beta}}(\mathbf{x}).\end{aligned}$$

As this holds for all $\mathbf{x} \in \mathbf{X}$, we deduce that $\mathcal{A}'[\mathbf{y}^{\boldsymbol{\beta}}] = \varphi_{\boldsymbol{\beta}} \in \mathcal{P}(\mathbf{X})$.

We then prove (8) $\Longrightarrow$ (7). Let $\mu \in \mathcal{M}(\mathbf{X})$, $\boldsymbol{\beta} \in \mathbb{N}^n$. Then, one has

$$\langle \mathcal{A}\,\mu, \mathbf{y}^{\boldsymbol{\beta}}\rangle = \int \underbrace{\mathcal{A}'[\mathbf{y}^{\boldsymbol{\beta}}]}_{\varphi_{\boldsymbol{\beta}}}\,\mathrm{d}\mu$$

which concludes the proof as by (8), $\varphi_{\boldsymbol{\beta}} = \mathcal{A}'[\mathbf{y}^{\boldsymbol{\beta}}] \in \mathcal{P}(\mathbf{X})$ holds. □

Remark 2.10 (On the validity of **Assumption 2**)**.**

To our best knowledge, all existing formulations of the GMP (including those in [27, 49, 32, 40, 21, 44, 23, 8]) satisfy **Assumption 2**. Indeed, using **Lemma 2.9**, it is clear that all operations whose adjoint preserves the space of polynomials (such as summation, polynomial multiplication/pushforward/composition, differentiation) satisfy **Assumption 2**.

Corollary 2.11 (Action on bounded degree polynomials)**.**

Under **Assumption 2**, *for all $\ell \in \mathbb{N}$, there exists $d_\ell \in \mathbb{N}$ such that any $p \in \mathrm{Im}(\mathcal{A}') \cap \mathbb{R}_{2\ell}[\mathbf{x}]$ has an antecedent of degree at most d_ℓ:*

$$\exists w \in \mathbb{R}_{d_\ell}[\mathbf{y}] \quad ; \quad p = \mathcal{A}'\,w.$$

Proof. Let $\ell \in \mathbb{N}$ and define $\mathcal{B}_\ell := \{\boldsymbol{\beta} \in \mathbb{N}^n \,;\deg(\varphi_{\boldsymbol{\beta}}) \leq 2\ell\}$. By **Assumption 2** and **Lemma 2.9**,

$$\mathrm{Im}(\mathcal{A}') = \mathrm{span}\{\varphi_{\boldsymbol{\beta}} \,; \boldsymbol{\beta} \in \mathbb{N}^n\} \quad \text{and thus} \quad \mathrm{Im}(\mathcal{A}') \cap \mathbb{R}_{2\ell}[\mathbf{x}] = \mathrm{span}\{\varphi_{\boldsymbol{\beta}} \,; \boldsymbol{\beta} \in \mathcal{B}_\ell\}.$$

Let $p \in \mathrm{Im}(\mathcal{A}') \cap \mathbb{R}_{2\ell}[\mathbf{x}]$: there exists $(c_{\boldsymbol{\beta}})_{\boldsymbol{\beta}} \in \mathbb{R}^{\mathcal{B}_\ell}$ such that

$$p(\mathbf{x}) = \sum_{\boldsymbol{\beta} \in \mathcal{B}_\ell} c_{\boldsymbol{\beta}}\,\varphi_{\boldsymbol{\beta}}(\mathbf{x}).$$

If $\mathcal{B}_\ell$ is finite, then $d_\ell := \max\{\deg(\varphi_{\boldsymbol{\beta}}) \,; \boldsymbol{\beta} \in \mathcal{B}_\ell\}$ concludes the proof.

If $\mathcal{B}_\ell$ is infinite, then because $\dim(\mathrm{Im}(\mathcal{A}') \cap \mathbb{R}_{2\ell}[\mathbf{x}]) \leq \dim(\mathbb{R}_{2\ell}[\mathbf{x}]) = \binom{m+2\ell}{m} < \infty$, one can extract a finite subset $\hat{\mathcal{B}}_\ell \subset \mathcal{B}_\ell$ such that

$$\mathrm{Im}(\mathcal{A}') \cap \mathbb{R}_{2\ell}[\mathbf{x}] = \mathrm{span}\{\varphi_{\boldsymbol{\beta}} \,; \boldsymbol{\beta} \in \hat{\mathcal{B}}_\ell\},$$

and one is back to the case where $\mathcal{B}_\ell$ is finite. □

Then, it is possible to generalize steps **1** to **5** of the previous section to write a hierarchy of finite dimensional convex problems that approximate (1) and (1′). For simplicity of exposition, here we rather approximate (2) and (2′).

Step 1: First, by **Lemma 2.9**, under **Assumption 2** it is possible to replace $\mathcal{C}(\mathbf{X})_+$ with $\mathcal{P}(\mathbf{X})_+$ without changing the optimal value $\mathrm{d}^\star_{\mathrm{M}}$:

$$\begin{aligned}\mathrm{d}^\star_{\mathrm{M}} = \inf_{w \in \mathcal{P}(\mathbf{Y})} \quad & \langle T, w\rangle \\ \text{s.t.} \quad & \mathcal{A}'\,w - g \in \mathcal{P}(\mathbf{X})_+.\end{aligned} \tag{9a}$$

Step 2: Second, assuming that **Condition 1** holds, we look for a minimizing sequence of *strictly feasible* polynomials w:

$$\begin{aligned} \mathrm{d}_{\mathrm{M}} = \quad & \inf_{w\in\mathbb{R}[\mathbf{y}]} && \langle T, w\rangle \\ & \text{s.t.} && \mathcal{A}'\, w - g \in \mathcal{P}(\mathbf{X})_{\oplus}. \end{aligned} \tag{9b}$$

Step 3: Using **Theorem 2.6**, we recast the positivity constraint of (9b) as a quadratic module constraint:

$$\begin{aligned} \mathrm{d}_{\mathrm{M}} = \quad & \inf_{w\in\mathbb{R}[\mathbf{y}]} && \langle T, w\rangle \\ & \text{s.t.} && \mathcal{A}'\, w - g \in \mathcal{Q}(\mathbf{h}). \end{aligned} \tag{9c}$$

Step 4: Eventually, we *project* our infinite dimensional quadratic module onto the bounded degree quadratic module defined for $\ell \in \mathbb{N}$ by

$$\mathcal{Q}_\ell(\mathbf{h}) := \left\{ \begin{pmatrix}\mathbf{h}^\top & 1\end{pmatrix} \boldsymbol{\sigma} \in \mathcal{Q}(\mathbf{h})\,;\forall i \in [\![r]\!],\ \max(\deg(\sigma_i\, h_i), \deg(\sigma_{r+1})) \leq 2\ell \right\}$$

which happens to be a finite dimensional convex cone, obtaining the following SoS programming problem:

$$\boxed{\begin{aligned} \mathrm{d}_{\mathrm{M}}^\ell := \quad & \inf_{w\in\mathbb{R}_{d_\ell}[\mathbf{y}]} && \langle T, w\rangle \\ & \text{s.t.} && \mathcal{A}'\, w - g \in \mathcal{Q}_\ell(\mathbf{h}), \end{aligned}} \tag{9d}$$

where we used **Corollary 2.11** to bound the degree of w. Note that (9d) is a *tigthening* of (9c) in the sense that we replaced the feasible set with a strictly smaller one (even finite dimensional): in general,

$$\mathrm{d}_{\mathrm{M}}^\ell > \mathrm{d}_{\mathrm{M}}^\star.$$

Moreover, as $\mathcal{Q}(\mathbf{h}) = \cup_{\ell\in\mathbb{N}}\mathcal{Q}_\ell(\mathbf{h})$ and $\mathcal{Q}_L(\mathbf{h}) = \cup_{\ell\leq L}\mathcal{Q}_\ell(\mathbf{h})$ are clear, one also has the following monotone convergence theorem for free:

$$\mathrm{d}_{\mathrm{M}}^\ell \underset{\ell\to\infty}{\searrow} \mathrm{d}_{\mathrm{M}}^\star.$$

Step 5: Using Lagrange duality, from the SoS tightening (9d) we deduce the following moment *relaxation*:

$$\boxed{\begin{aligned} \mathrm{p}_{\mathrm{M}}^\ell := \quad & \sup_{Z\in\mathcal{Q}_\ell(\mathbf{h})'} && \langle Z, g\rangle \\ & \text{s.t.} && \mathcal{A}_\ell''\, Z = T\Big|_{\mathbb{R}_{d_\ell}[\mathbf{y}]} \end{aligned}} \tag{9e}$$

the "double adjoint" $\mathcal{A}_\ell'' := (\mathcal{A}'|_{\mathbb{R}_{d_\ell}[\mathbf{y}]})' : \mathbb{R}_{2\ell}[\mathbf{x}]' \to \mathbb{R}_{d_\ell}[\mathbf{y}]'$ coinciding with $\mathcal{A}$ on $\mathcal{M}(\mathbf{X}) \subset \mathbb{R}_{2\ell}[\mathbf{x}]'$:

$$\forall \boldsymbol{\beta} \in \mathbb{N}^n_{d_\ell}, \qquad \langle \mathbf{y}^{\boldsymbol{\beta}}, \mathcal{A}''\, Z\rangle = \langle \varphi_{\boldsymbol{\beta}}, Z\rangle$$

under **Assumption 2** and with $\deg(\varphi_{\boldsymbol{\beta}}) \leq 2\ell$ by **Corollary 2.11** and its proof.

Several results, ranging from practical to theoretical, come with these tightenings and relaxations. First, as **Step 3** assumes Slater's **Condition 1**, strong duality holds between (9e) and (9d): $\mathrm{p}_{\mathrm{M}}^\ell = \mathrm{d}_{\mathrm{M}}^\ell$ (for all ℓ such that $d_\ell \geq \deg(\mathring{w})$).

Second, [28, Proposition 2.1] gives a representation of $\mathcal{Q}_\ell(\mathbf{h})$ with p.s.d. matrices (and hence by duality $\mathcal{Q}_\ell(\mathbf{h})'$ is represented by linear matrix inequalities), so that the moment relaxations and SoS strengthenings are equivalent to semidefinite programming (SDP) problems. Eventually, the following theorem ensures strong convergence guarantees of the corresponding numerical scheme:

Theorem 2.12 (Convergence of the moment-SoS hierarchy [48, Theorem 4 & Corollary 8])**.**

Suppose that $\exists B > 0$, $\ell_{\min} \in \mathbb{N}$ *s.t.* $\forall \ell \geq \ell_{\min}$, *any* Z *feasible for* (9e) *satisfies* $\langle 1, Z\rangle \leq B$, *and that* $1 - \mathbf{x}^\top \mathbf{x} \in \mathcal{Q}(\mathbf{h})$ *(up to rescaling, this second condition can be enforced if* $\mathbf{X}$ *is compact). Then, for* $\ell \geq \ell_{\min}$,

$$\mathrm{d}_{\mathrm{M}}^{\ell} = \mathrm{p}_{\mathrm{M}}^{\ell} \underset{\ell \to \infty}{\longrightarrow} \mathrm{p}_{\mathrm{M}}^{\star} = \mathrm{d}_{\mathrm{M}}^{\star}.$$

Moreover, if (2) *has a unique solution* $\mu^\star$, *then for* $\ell \geq \ell_{\min}$ (9e) *has a unique solution* Z_ℓ *and*

$$\forall \boldsymbol{\alpha} \in \mathbb{N}^m, \qquad \langle \mathbf{x}^{\boldsymbol{\alpha}}, Z_\ell \rangle \underset{\ell \to \infty}{\longrightarrow} \int \mathbf{x}^{\boldsymbol{\alpha}} \, \mathrm{d}\mu^\star(\mathbf{x}).$$

3 Method for convergence rates computation

The aim of this section (and more generally of this article) is to design methods for computing the rate of the optimal values convergence given in **Theorem 2.12**. From particular examples, we derive a generic method for computing such convergence rate, depending on the solutions of the infinite dimensional problem (9a).

Our strategy consists of the following steps:

1. **Construction of a suitable minimizing sequence of polynomials.** In this step it is important to control simultaneously the degree of those polynomials and the convergence of their cost towards the optimal value.
2. **Application of effective version of Positivstellensätze.** In this step, explicit convergence rates are derived. They are based on the convergence rates for Positivstellensätze and the minimizing sequence from the previous step.

There is an interplay between the two steps inherent to the choice of the minimizing sequence. We will see an adversarial behavior between, on the one hand, a good approximation of the optimal cost value via high degree polynomials and, on the other hand, degree bounds in the SDP relaxations.

Remark 3.1 (Focusing on the function LP)**.**

We will focus on the dual LP ($2'$) and not on the primal (2) simply because we will use the effective version of Putinar's Positivstellensatz **Theorem 3.7**, which is more adapted to the dual problem ($2'$) than to the primal (2). However, under Slater's **Condition 1**, strong duality holds in each level of the moment-SoS hierachy.

Remark 3.2 (Sparse and symmetric problems)**.**

The number of variables in the SDP for the ℓ-th level of the moment-SoS hierarchy grows combinatorial with $\ell \in \mathbb{N}$. Thus exploiting sparsity or symmetry, when present, is important in practice. Symmetry can be exploited without loss of accuracy, see [42], and therefore the convergence rates translate immediately from the full moment-SoS hierarchy to the symmetry-reduced one. By [25], correlation-sparsity allows to transfer convergence rates. For (correlation-)sparse dynamical systems the convergence rates can even be improved as long as the bounds in the effective version of Putinar's Positivstellensatz grow with increasing state dimension, see [43].

3.1 Example: Static optimization

The Polynomial Optimization Problem (POP) is at the root of the development of the moment-SoS hierarchy, and will serve as a fundamental example for our convergence rates computation. It consists in *globally* minimizing a polynomial $f \in \mathbb{R}[\mathbf{x}]$ on a nonempty, compact basic semi-algebraic set $\varnothing \neq \mathbf{X} := \mathbf{S}(\mathbf{h}) \subset \mathbb{R}^m$, where $\mathbf{h} \in \mathbb{R}[\mathbf{x}]^r$:

$$\begin{aligned} f_{\mathbf{X}}^\star := \min_{\mathbf{x} \in \mathbb{R}^m} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x}) \in \mathbb{R}_+^r. \end{aligned} \tag{10}$$

By definition of the minimum, it is straightforward that $f^\star_{\mathbf{X}} = \max\{w \in \mathbb{R} : f - w \geq 0 \text{ on } \mathbf{X}\}$, which can in turn be approximated by the moment-SoS hierarchy of SoS strengthenings:

$$f^\ell_{\mathbf{X}} := \max_{w \in \mathbb{R}} \quad w \qquad \text{s.t.} \quad f - w \in \mathcal{Q}_\ell(\mathbf{h}). \tag{11}$$

Here there are two possible cases: (i) $f - f^\star_{\mathbf{X}} \in \mathcal{Q}(\mathbf{h})$ or (ii) $f - f^\star_{\mathbf{X}} \in \mathcal{P}(\mathbf{X})_+ \setminus \mathcal{Q}(\mathbf{h})$.

In case (i), as $\mathcal{Q}(\mathbf{h}) = \cup_{\ell \in \mathbb{N}} \mathcal{Q}_\ell(\mathbf{h})$ and $\ell \leq \ell' \implies \mathcal{Q}_\ell(\mathbf{h}) \subset \mathcal{Q}_{\ell'}(\mathbf{h})$, there exists an $\ell^\star \in \mathbb{N}$ such that $f - f^\star_{\mathbf{X}} \in \mathcal{Q}_\ell(\mathbf{h}) \iff \ell \geq \ell^\star$ and one has finite convergence:

$$f^\star_{\mathbf{X}} \geq f^\ell_{\mathbf{X}} \qquad \text{with} \qquad f^\star_{\mathbf{X}} = f^\ell_{\mathbf{X}} \iff \ell \geq \ell^\star. \tag{12}$$

A sufficient condition for such finite convergence was given in [38], in the following setting:

Definition 3.3 (Constraint sets and Lagrange function)**.**

- $\mathbf{J}_+ := \{i \in [\![r]\!] \,;\, \forall j \in [\![r]\!], t < 0, h_i \neq t\, h_j\}$ is the set of *inequality constraints*,
- $\mathbf{J}_0 := {}^{([\![r]\!] \setminus \mathbf{J}_+)}/\leftrightarrow$, where $i \leftrightarrow j$ *iff* $-\frac{h_i}{h_j} \in \mathbb{R}_\oplus$, is the set of *equality constraints*,
- $\mathbf{J}_\star(\mathbf{x}) := \{i \in [\![r]\!] \,;\, h_i(\mathbf{x}) = 0\}$ is the set of *active constraints* in $\mathbf{x} \in \mathbf{X}$,
- $\Lambda(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) - \mathbf{y}^\top \mathbf{h}(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^r$, is the *Lagrange function.*

Condition 3.

$\mathcal{Q}(\mathbf{h})$ has the Archimedean property and, for each local minimum $\overline{\mathbf{x}} \in \mathbf{X}$ of f:

3.1 $\{\mathbf{grad}\, h_i(\overline{\mathbf{x}}) \,;\, i \in \mathbf{J}_\star(\overline{\mathbf{x}})\}$ are linearly independent *(constraint qualification)*,

3.2 $\begin{cases} \mathbf{grad}_{\mathbf{x}}\, \Lambda(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = 0 \\ \forall i \in \mathbf{J}_+,\ \overline{y}_i \geq 0 \text{ and } \overline{y}_i\, h_i(\overline{\mathbf{x}}) = 0 \end{cases} \implies \forall i \in \mathbf{J}_+, \overline{y}_i + h_i(\overline{\mathbf{x}}) > 0$ *(strict complementarity)*,

3.3 $\begin{cases} \mathbf{grad}_{\mathbf{x}}\, \Lambda(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = 0, \quad \mathbf{v} \neq \mathbf{0} \\ \forall i \in \mathbf{J}_+,\ \overline{y}_i \geq 0 \text{ and } \overline{y}_i\, h_i(\overline{\mathbf{x}}) = 0 \\ \forall i \in \mathbf{J}_\star(\overline{\mathbf{x}}),\ \mathbf{grad}\, h_i(\overline{\mathbf{x}})^\top \mathbf{v} = 0 \end{cases} \implies \mathbf{v}^\top \operatorname{Hess}_{\mathbf{x}} \Lambda(\overline{\mathbf{x}}, \overline{\mathbf{y}})\, \mathbf{v} > 0$ *(second order condition).*

Theorem 3.4 (Finite convergence of the moment-SoS hierarchy)**.**

Under **Condition 3**, $f - f^\star_{\mathbf{X}} \in \mathcal{Q}(\mathbf{h})$. *However, no bound on* ℓ *s.t.* $f - f^\star_{\mathbf{X}} \in \mathcal{Q}_\ell(\mathbf{h})$ *can be computed in general (see discussion in [38, Section 5]).*

Case (ii) is tackled in [2, Section 4] and [3]. We reproduce here the main results, which are based on the following **Theorem 3.5** and **Assumption 3**.

Theorem 3.5 (Łojasiewicz exponent [2, **Theorem 2.3**, **Definition 2.4**])**.**

For $\mathbf{x} \in [-1, 1]^m$, *let*

$$H(\mathbf{x}) := |\min(h_1(\mathbf{x}), \ldots, h_r(\mathbf{x}), 0)| \qquad D(\mathbf{x}) := \min\{|\mathbf{x} - \mathbf{x}'| \,;\, \mathbf{x}' \in \mathbf{S}(\mathbf{h})\},$$

where $|\mathbf{x}| := \sqrt{\mathbf{x}^\top \mathbf{x}}$ *is the Euclidean norm of* $\mathbf{x}$. *Then there exists* $Ł, \mathfrak{c} \in \mathbb{R}^\star_+$ *s.t. for* $\mathbf{x} \in [-1, 1]^m$

$$D(\mathbf{x})^{Ł} \leq \mathfrak{c}\, H(\mathbf{x}). \tag{13}$$

For the effective version of Putinar's Positivstellensatz we make the following assumption.

Assumption 3.

1. $1 - \mathbf{x}^\top \mathbf{x} \in \mathcal{Q}(\mathbf{h})$ *(normalized Archimedean property)*,
2. $\forall i \in [\![r]\!], \quad \|h_i\| := \max_{\mathbf{x}\in[-1,1]^m} h_i(\mathbf{x}) \leq \frac{1}{2}$

Remark 3.6 (On the validity of **Assumption 3**).

The normalized Archimedean property can be seen as a restatement of compactness of $\mathbf{X}$. For compact $\mathbf{X} = \mathbf{S}(\mathbf{h})$, up to rescaling, $\mathbf{S}(\mathbf{h})$ is included in the unit ball so that it is possible to add the redundant inequality constraint $1 - \mathbf{x}^\top \mathbf{x} \geq 0$ to the description of $\mathbf{S}(\mathbf{h})$. This is the practical approach for guaranteeing **Assumption 3.1**. The second condition in **Assumption 3** is only of technical nature and can be obtained by scaling $\mathbf{h}$.

In this text, we will use the below effective version of Putinar's Positivstellensatz. For the statement of this theorem we recall the notations

$$p^\star_{\mathbf{X}} := \inf_{\mathbf{x}\in\mathbf{X}} p(\mathbf{x}), \quad \|p\| := \max_{\mathbf{x}\in[-1,1]^m} p(\mathbf{x}).$$

Theorem 3.7 (Effective Putinar Positivstellensatz [2, **Theorem 1.7**]).

For $m \geq 2$, $p \in \mathcal{P}(\mathbf{X})_\oplus$, under **Assumption 3***, one has*

$$\ell \geq \gamma(m, \mathbf{h}) \deg(p)^{3.5m\text{Ł}} \left(\|p\|/p^\star_{\mathbf{X}}\right)^{2.5m\text{Ł}} \quad \Longrightarrow \quad p \in \mathcal{Q}_\ell(\mathbf{h}) \tag{14}$$

where $1 \leq \gamma(m, \mathbf{h}) \leq \Gamma\, m^3\, 2^{5\text{Ł}-1}\, r^m\, \mathfrak{c}^{2m} \deg(\mathbf{h})^m$ and $\Gamma > 0$ does not depend on $m, p, \mathbf{h}$. In the rest of this paper, we will consider fixed m and $\mathbf{h}$, so that we simplify the notation $\gamma(m, \mathbf{h})$ into γ.

Remark 3.8 (Farkas Lemma).

Regardless of m, if $\deg(p) = \deg(\mathbf{h}) = 1$ (affine forms), then $p \in \mathcal{P}(\mathbf{X})_+ \iff p \in \mathcal{Q}_1(\mathbf{h})$. In such case, both (10) and its order 0 SoS strengthening (11) are equivalent to the same linear program (LP).

Corollary 3.9 (Convergence rate for POP [2, **Theorems 2.11, 4.2 & 4.3**]).

For $m \geq 2$, under **Assumption 3***, one has*

$$0 < \varepsilon \leq \|f\| \quad \textit{and} \quad \ell \geq \gamma \deg(f)^{3.5m\text{Ł}} \left(3\|f\|/\varepsilon\right)^{2.5m\text{Ł}} \quad \Longrightarrow \quad 0 \leq f^\star_{\mathbf{X}} - f^\ell_{\mathbf{X}} \leq \varepsilon \tag{15a}$$

$$0 \leq f^\star_{\mathbf{X}} - f^\ell_{\mathbf{X}} \leq (\gamma/\ell)^{\frac{1}{2.5m\text{Ł}}}\, 3\|f\| \deg(f)^{\frac{7}{5}} \in \mathcal{O}\left(\ell^{-1/2.5m\text{Ł}}\right) \tag{15b}$$

If constraint qualification **Condition 3.1** *holds for any $\overline{\mathbf{x}} \in \mathbf{X}$, then Ł$= 1$.*

Proof. (15a) is (14) with $p = f - f^\star_{\mathbf{X}} + \varepsilon$, so that $p^\star_{\mathbf{X}} = \varepsilon$ and, as $\varepsilon \leq \|f\|$ and $|f^\star_{\mathbf{X}}| \leq \|f\|$, $\|p\| \leq 3\|f\|$. Thus, by **Theorem 3.7**, $p \in \mathcal{Q}_\ell(\mathbf{h})$, which means that $w = f^\star_{\mathbf{X}} - \varepsilon$ is feasible for (11), so that $f^\ell_{\mathbf{X}} \geq f^\star_{\mathbf{X}} - \varepsilon$, which is the announced inequality. □

Notice that the key idea here consists in perturbating the optimal $f - f^\star_{\mathbf{X}}$ (which is nonnegative by design but in case (ii) does not belong to $\mathcal{Q}(\mathbf{h})$) with some ε to obtain a positive polynomial p (which is then guaranteed to be in $\mathcal{Q}(\mathbf{h})$, using Putinar's **Theorem 2.6**) to which we apply **Theorem 3.7** to get an effective order ℓ quadratic module representation. This in turn allows us to derive a bound on the rate of the convergence $f^\ell_{\mathbf{X}} \underset{\ell\to\infty}{\longrightarrow} f^\star_{\mathbf{X}}$.

In the next section, we will derive generic methods for constructing the right polynomial p, depending on the solutions of (9a).

3.2 General method and function approximation

In this section, we specify the procedure that we have indicated at the beginning of Section 3. Let $\varepsilon > 0$. Supposing that (9a) has an optimal solution $w^\star$ and noting $d := \deg(w^\star)$, we want to perturbate it with some

$\tilde{w} \in \mathbb{R}_d[\mathbf{y}]$ such that $\hat{w} := w^\star + \tilde{w}$ is still feasible for (9a) but also $\langle T, \hat{w} \rangle \leq \mathrm{d}_{\mathrm{M}}^\star + \varepsilon$ and $p := \mathcal{A}'\,\hat{w} - g \in \mathcal{P}(\mathbf{X})_\oplus$, *i.e.* $\hat{w}$ is feasible for (9b) and thus for (9c) by **Theorem 2.6**. Then, **Theorem 3.7** will give us a lower bound on ℓ such that $\hat{w}$ is feasible for our SoS strenghtening (9d), which will prove that $\mathrm{d}_{\mathrm{M}}^\star \leq \mathrm{d}_{\mathrm{M}}^\ell \leq \mathrm{d}_{\mathrm{M}}^\star + \varepsilon$.

This is what we did in the previous section:

$$\begin{aligned} -f_{\mathbf{X}}^\star := \min_{w \in \mathbb{R}} \quad & -w = \langle T, w \rangle \\ \text{s.t.} \quad & \underbrace{(-w)}_{\mathcal{A}'\,w} - \underbrace{(-f)}_{g} \in \mathcal{P}(\mathbf{X})_+. \end{aligned}$$

with $n = 0$ (so that $\mathcal{Y} = \mathcal{P}(\{\mathbf{0}\}) = \mathbb{R}$), $w^\star = f_{\mathbf{X}}^\star$ and $\tilde{w} = -\varepsilon$.

If (9a) has no optimal solution, then we look for minimizing sequences. One way of doing so is to relax (2′) into

$$\begin{aligned} \mathrm{d}_{\mathrm{F}}^\star = \inf_{w \in \overline{\mathcal{Y}}} \quad & \langle \overline{T}, w \rangle \\ \text{s.t.} \quad & \overline{\mathcal{A}'}\,w - g \in \overline{\mathcal{X}'_+} \end{aligned} \tag{16}$$

where $\overline{\mathcal{Y}} \supset \mathcal{Y}$ is the closure of $\mathcal{Y}$ for its "well-chosen" topology, $\overline{T}$ (resp. $\overline{\mathcal{A}'}$) is the unique continuous linear extension of T (resp. $\mathcal{A}'$) to $\overline{\mathcal{Y}}$ and $\overline{\mathcal{X}'_+} = \mathrm{Im}(\overline{\mathcal{A}'}) \cap \mathbb{R}_+^{\mathbf{X}}$ is a cone of nonnegative functions. To our best knowledge, in relevant applications, a good choice for the topology on $\mathcal{Y}$ often results in (16) having an optimal solution $w^\star \in \overline{\mathcal{Y}}$ with $\langle \overline{T}, w^\star \rangle = \mathrm{d}_{\mathrm{F}}^\star = \mathrm{d}_{\mathrm{M}}^\star$. Then, by continuity of $\overline{T}$ and density of $\mathcal{Y}$ in $\overline{\mathcal{Y}}$ one can find $w_\varepsilon \in \mathcal{Y}$ feasible for (9a) such that $\langle T, w_\varepsilon \rangle \leq \mathrm{d}_{\mathrm{M}}^\star + \frac{\varepsilon}{2}$, after which one only needs to repeat the process of previous paragraph, looking for $\tilde{w} \in \mathbb{R}_d[\mathbf{y}]$ (where $d = \deg(w_\varepsilon)$) such that $\hat{w} = w_\varepsilon + \tilde{w}$ is feasible for (9b) and (9c) and $\langle T, \hat{w} \rangle \leq \langle T, w_\varepsilon \rangle + \frac{\varepsilon}{2} \leq \mathrm{d}_{\mathrm{M}}^\star + \varepsilon$, and using **Theorem 3.7** to find a lower bound on ℓ s.t. $\hat{w}$ is feasible for (9d), proving again that $\mathrm{d}_{\mathrm{M}}^\star \leq \mathrm{d}_{\mathrm{M}}^\ell \leq \mathrm{d}_{\mathrm{M}}^\star + \varepsilon$.

Hence, the general process for computing the degree ℓ needed for a given $\varepsilon > 0$ accuracy and the corresponding convergence rate is summarized as follows:

1. Take a minimizer $w^\star$ of the LP (9a) (or of the extended LP (16) for which a minimizer exists).
2. If $w^\star$ is not a polynomial, then approximate it with a feasible polynomial w_ε with $\langle T, w_\varepsilon \rangle \leq \mathrm{d}_{\mathrm{M}}^\star + {}^{\varepsilon}/_{2}$.
3. Perturb the polynomial $w^\star$ (resp. w_ε) into a strictly feasible polynomial $\hat{w}$ with $\langle T, \hat{w} \rangle \leq \mathrm{d}_{\mathrm{M}}^\star + \varepsilon$.
4. Apply effective Positivstellensätze to show that $\hat{w}$ is feasible for the SDP hierarchy at some level $\ell \in \mathbb{N}$.
5. Relate the approximation error ε and the hierarchy level ℓ to derive a convergence rate.

In conclusion, in the case where only (16) has an optimal solution, only one additional step is needed, and the convergence rate is obtained by combining **Theorem 3.7** with an approximation theorem that gives a lower bound on the degree d required for $w_\varepsilon \in \mathbb{R}_d[\mathbf{y}]$ to be as defined above. Thus, the rest of this section will be devoted to proving general results on the degree needed to find proper w_ε and $\tilde{w}$ polynomials.

3.3 Polynomial approximation (finding w_ε)

This section introduces polynomial approximation results that depend on the regularity of the function to be approximated. For $\mathbf{Y} \subset \mathbb{R}^n$ open or compact with nonempty interior, we need to define the vector spaces $C_b^k(\mathbf{Y}), k \in \mathbb{N}$ by induction: $C_b^0(\mathbf{Y}) := \mathcal{C}_b(\mathbf{Y}) := \{f \in \mathcal{C}(\mathbf{Y})\,; f \text{ is bounded on } \mathbf{Y}\}$ and

$$C_b^{k+1}(\mathbf{Y}) := \left\{ f \in \mathcal{C}_b(\mathbf{Y})\,; \forall j \in [\![n]\!],\ \frac{\partial f}{\partial y_j} \in C_b^k(\mathbf{Y}) \right\}.$$

These vector spaces are equipped with the norms $\|f\|_{C_b^0(\mathbf{Y})} := \|f\|_\infty^{\mathbf{Y}} = \sup_{\mathbf{Y}} |f|$ and, again by induction,

$$\|f\|_{C_b^{k+1}(\mathbf{Y})} := \|f\|_\infty^{\mathbf{Y}} + \sum_{j=1}^n \left\| \frac{\partial f}{\partial y_j} \right\|_{C_b^k(\mathbf{Y})}.$$

When $\mathbf{Y}$ is compact, the subscript b is omitted as continuous functions are bounded on compact sets.

An important object is the modulus of continuity $\omega_{f,k,\mathbf{y}}(\rho)$ of a function $f \in C_b^k(\mathbf{Y})$ of order k at a point $\mathbf{y} \in \mathbf{Y} \subset \mathbb{R}^n$ for the radius $\rho > 0$, defined as

$$\omega_{f,k}(\mathbf{y},\rho) := \sup_{\boldsymbol{\alpha}\in\mathbb{N}_k^n} \left(\sup_{\|\mathbf{y}-\mathbf{y}'\|\leq\rho} |\partial_{\boldsymbol{\alpha}} f(\mathbf{y}) - \partial_{\boldsymbol{\alpha}} f(\mathbf{y}')| \right) \tag{17a}$$

$$\text{where} \qquad \partial_{\boldsymbol{\alpha}} f = \frac{\partial^{\alpha_1}}{\partial y_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial y_n^{\alpha_n}} f,$$

as well as the global modulus of continuity

$$\omega_{f,k}^{L^\infty}(\mathbf{Y},\rho) := \|\omega_{f,k}(\cdot,\rho)\|_\infty^{\mathbf{Y}} = \sup_{\mathbf{y}\in\mathbf{Y}} \omega_{f,k}(\mathbf{y},\rho) \leq \infty \tag{17b}$$

and, for $\mu \in \mathcal{M}(\mathbf{Y})_+$ and $s \geq 1$, the $L^s(\mu)$-averaged modulus of continuity

$$\omega_{f,k}^{L^s}(\mu,\rho) := \|\omega_{f,k}(\cdot,\rho)\|_{L^s(\mu)} := \left(\int \omega_{f,k}(\cdot,\rho)^s \, d\mu \right)^{1/s} \tag{17c}$$

Notice that $\mathbf{Y} \subset \mathbf{Y}' \Longrightarrow \omega_{f,k}^{L^\infty}(\mathbf{Y},\cdot) \leq \omega_{f,k}^{L^\infty}(\mathbf{Y}',\cdot)$ and $\mu_1 - \mu_2 \in \mathcal{M}(\mathbf{Y})_+ \Longrightarrow \omega_{f,k}^{L^s}(\mu_1,\cdot) \geq \omega_{f,k}^{L^s}(\mu_2,\cdot)$. With the notion of modulus of continuity we can state the following theorem from [1] concerning convergence speed for polynomial approximation of regular functions.

Theorem 3.10 (An extended Jackson inequality [1])**.**

Let $\mathbf{Y} \subset \mathbb{R}^n$ be open and bounded, $f \in C_b^k(\mathbf{Y})$. For $d \in \mathbb{N}$ there is a polynomial $p_d \in \mathbb{R}_d[\mathbf{y}]$ such that for each $\boldsymbol{\alpha} \in \mathbb{N}^n$ with $|\boldsymbol{\alpha}| \leq \min(k,d)$ we have

$$\|\partial_{\boldsymbol{\alpha}}(f - p_d)\|_\infty^{\mathbf{Y}} \leq \frac{c}{d^{k-|\boldsymbol{\alpha}|}} \, \omega_{f,k}^{L^\infty}(\mathbf{Y}, 1/d) \,. \tag{18a}$$

where c is a positive constant depending only on n, k and $\mathbf{Y}$.

Corollary 3.11 (Approximating smooth functions)**.**

Let $\mathbf{Y} \subset \mathbb{R}^n$ be open and bounded, $f \in C^k(\overline{\mathbf{Y}})$. For $d \geq k$ there is a polynomial $p_d \in \mathbb{R}_d[\mathbf{y}]$ such that

$$\|f - p_d\|_{C^k(\overline{\mathbf{Y}})} \leq c \left(1 + \frac{1}{d}\right)^k \omega_{f,k}^{L^\infty}\left(\overline{\mathbf{Y}}, 1/d\right) \leq c_0 \, \omega_{f,k}^{L^\infty}\left(\overline{\mathbf{Y}}, 1/d\right) \tag{18b}$$

where $c, c_0 = \mathrm{e} \cdot c$ are positive constants depending only on n, k and $\overline{\mathbf{Y}}$.

Proof. Consider the polynomial p_d given by **Theorem 3.10**. Then, one has

$$\begin{aligned}
\|f - p_d\|_{C^k(\overline{\mathbf{X}})} &\overset{*}{=} \textstyle\sum_{|\boldsymbol{\alpha}|\leq k} \|\partial_{\boldsymbol{\alpha}}(f-p_d)\|_\infty^{\mathbf{Y}} \\
&\overset{(18a)}{\leq} \textstyle\sum_{|\boldsymbol{\alpha}|\leq k} \frac{c}{d^{k-|\boldsymbol{\alpha}|}} \, \omega_{f,k}^{L^\infty}(\mathbf{Y}, 1/d) \\
&= \tfrac{c}{d^k} \, \omega_{f,k}^{L^\infty}(\mathbf{Y}, 1/d) \textstyle\sum_{|\boldsymbol{\alpha}|\leq k} d^{|\boldsymbol{\alpha}|} \\
&= \tfrac{c}{d^k} \, \omega_{f,k}^{L^\infty}(\mathbf{Y}, 1/d) \, (1+d)^k \\
&= c \left(1 + \tfrac{1}{d}\right)^k \omega_{f,k}^{L^\infty}(\mathbf{Y}, 1/d) \\
&\leq c \left(1 + \tfrac{1}{d}\right)^k \omega_{f,k}^{L^\infty}\left(\overline{\mathbf{Y}}, 1/d\right)
\end{aligned}$$

where the first equality at $*$ follows from continuity of $\partial_{\boldsymbol{\alpha}}(f - p_d)$. We conclude by noticing $(1 + 1/d)^k \leq (1 + 1/k)^k \leq \mathrm{e}$. $\square$

With a view toward the problem of computing the volume of a semialgebraic set from Section 2.3, the following one-sided approximation result is useful.

Theorem 3.12 (One-sided polynomial approximation [6])**.**

Let $\mathbf{Y} \subset [-1,1]^n$, $\lambda_{\mathbf{Y}}$ be the Lebesgue measure on $\mathbf{Y}$ and $f : \mathbf{Y} \longrightarrow \mathbb{R}$ be bounded and measurable.

For all $s \in [1, \infty)$ and $d \in \mathbb{N}$ there is a polynomial $p_d \in \mathbb{R}_d[\mathbf{y}]$ such that $p_d \geq f$ on $\mathbf{Y}$ and

$$\int \left(p_d - f\right)^s \, \mathrm{d}\lambda_{\mathbf{Y}} \leq \bar{c}\, \omega_{f,0}^{L^s}\left(\lambda, {}^1\!/\!_d\right) \tag{18c}$$

for some constant $\bar{c}$ depending only on n and s.

Moreover, for all $d \in \mathbb{N}$ there is a polynomial $p_d \in \mathbb{R}_d[\mathbf{y}]$ such that $p_d \geq f$ on $\mathbf{Y}$ and

$$\lambda\left(\left\{\mathbf{y} \in \mathbf{Y} \;;\; p_d(\mathbf{y}) > f(\mathbf{y}) + \hat{c}\, \omega_{f,0}^{L^\infty}\left(\mathbf{Y}, {}^1\!/\!_d\right)\right\}\right) = 0 \tag{18d}$$

for some constant $\hat{c}$ depending only on n.

3.4 Inward pointing condition (finding $\tilde{w}$)

In this section, we complement polynomial approximations from the previous section with conditions that assure feasibility for those approximations. We will see in **Lemma 3.13** that the following condition is sufficient for guaranteeing the existence of a minimizing sequence of strictly feasible polynomials.

Condition 4 (Inward-pointing condition)**.**

We say the LP (16) satisfies the inward-pointing condition if for each feasible point w for (16) there exists $\phi \in \overline{\mathcal{Y}}$ such that

$$\overline{\mathcal{A}'}(w + \theta\phi) - g > 0 \quad \text{on } \mathbf{X} \tag{19}$$

for all $\theta \in [0, 1]$.

Lemma 3.13. *Under* **Condition 4***, there exists a minimizing sequence of strictly feasible polynomials.*

Proof. Let $\varepsilon > 0$ and $w_\varepsilon \in \overline{\mathcal{Y}}$ with

$$\langle \overline{T}, w_\varepsilon \rangle < \mathrm{d}_{\mathrm{F}}^\star + \frac{\varepsilon}{3}. \tag{20}$$

Let ϕ_ε be as in the inward-pointing **Condition 4**, such that $w_\varepsilon + \theta\phi_\varepsilon$ is strictly feasible for all $\theta \in [0, 1]$. By continuity of $\overline{T}$ let $\theta = \theta_\varepsilon$ be small enough such that

$$\left|\langle \overline{T}, \theta_\varepsilon \phi_\varepsilon \rangle\right| = \theta_\varepsilon \left|\langle \overline{T}, \phi_\varepsilon \rangle\right| < \frac{\varepsilon}{3}. \tag{21}$$

From compactness of $\mathbf{X}$ it follows that there exists $\rho > 0$ with $\overline{\mathcal{A}'}(w_\varepsilon + \theta_\varepsilon\phi_\varepsilon) - g \geq \rho$ on $\mathbf{X}$. By density of $\mathcal{Y}$ in $\overline{\mathcal{Y}}$, and continuity of $\overline{\mathcal{A}}$ and $\overline{T}$, there exists a polynomial $p_\varepsilon \in \mathcal{Y}$ (close enough to $w_\varepsilon + \theta_\varepsilon\phi_\varepsilon$) with $\mathcal{A}p_\varepsilon - g \geq \frac{\rho}{2}$ on $\mathbf{X}$, i.e. p_ε is strictly feasible, and

$$\left|\langle T, p_\varepsilon \rangle - \langle \overline{T}, w_\varepsilon + \theta_\varepsilon\phi_\varepsilon \rangle\right| < \frac{\varepsilon}{3}. \tag{22}$$

Putting together (20), (21) and (22) we get for p_ε

$$\langle T, p_\varepsilon \rangle \leq \left|\langle T, p_\varepsilon \rangle - \langle \overline{T}, w_\varepsilon + \theta_\varepsilon\phi_\varepsilon \rangle\right| + \left|\langle \overline{T}, \theta_\varepsilon\phi_\varepsilon \rangle\right| + \langle \overline{T}, w_\varepsilon \rangle < \mathrm{d}_{\mathrm{F}}^\star + \varepsilon.$$

Letting ε go to zero shows the statement. ◻

A simpler version of **Condition 4** is that there exists $\phi \in \overline{\mathcal{Y}}$ with $\overline{\mathcal{A}}\phi > 0$.

Lemma 3.14. *Assume there exists $\phi \in \overline{\mathcal{Y}}$ with $\overline{\mathcal{A}}\phi > 0$. Then* **Condition 4** *is satisfied.*

Proof. Let $w \in \overline{\mathcal{Y}}$ be feasible. Then for all $\theta > 0$ it holds $\overline{\mathcal{A}}(w + \theta\phi) - g = \overline{\mathcal{A}}(w) - g + \theta\overline{\mathcal{A}}(\phi) > 0$ on $\mathbf{X}$. □

The inward-pointing **Condition 4** is closely related to the Slater **Condition 1**. We address this shortly in the following proposition.

Proposition 3.15 (Inward-pointing and Slater conditions)**.**

The inward-pointing **Condition 4** *and the Slater* **Condition 1** *are equivalent.*

Proof. If **Condition 4** is satisfied, then **Condition 1** is satisfied. This follows immediately because for feasible $w \in \overline{\mathcal{A}}$ and ε and ϕ as in **Condition 4** the point $w + \varepsilon\phi$ is strictly feasible, i.e. (by continuity of $\overline{\mathcal{A}}$) lies in the interior of the feasible set. On the other hand, if **Condition 1** is satisfied with strictly feasible point $\mathring{w} \in \mathcal{Y}$ then, by convexity of the feasible set, **Condition 4** is satisfied. Indeed, for w feasible let $\phi := \mathring{w} - w$; then for all $\theta \in [0, 1]$ it holds

$$\overline{\mathcal{A}}'(w + \theta\phi) - g = \overline{\mathcal{A}}'((1-\theta)w + \theta\mathring{w}) - g = (1-\theta)\underbrace{(\overline{\mathcal{A}}'w - g)}_{\geq 0} + \theta\underbrace{(\overline{\mathcal{A}}'\mathring{w} - g)}_{>0} > 0.$$

□

Remark 3.16 (On the relevance of **Condition 4**)**.**

Upon reading **Proposition 3.15**, one could wonder why the inward-pointing condition is important, as it is equivalent to the better-known Slater condition. The reason is that this condition is a quantitative version of Slater's condition, in the sense that checking it gives effective values for ϕ, which will be instrumental in the computation of the convergence rates. Hence, the inward-pointing **Condition 4** is to Slater's **Condition 1** what the effective Putinar **Theorem 3.7** is to the original Putinar **Theorem 2.6**. In practice, most of the time the inner pointing condition and corresponding ϕ are deduced from **Lemma 3.14**.

3.5 Obtaining the convergence rates

Here, we put together the steps we discussed in this section. We consider the GMP (2′) and formulate the following (quantitative) conditions.

Condition 5.

5.1 Effective version of Putinar's P-satz: There is an effective degree bound $\ell_b(m, \mathbf{h}, p^\star_{\mathbf{X}}, \|p\|^{\mathbf{X}}_\infty, \deg(p))$ for Putinar's Positivstellensatz. That is, for $\mathbf{X} = \mathbf{S}(\mathbf{h}) \subset \mathbb{R}^m$ and $p \in \mathcal{P}(\mathbf{X})$, it holds

$$p^\star_{\mathbf{X}} := \min\{p(\mathbf{x})\ ;\ \mathbf{x} \in \mathbf{X}\} > 0 \quad \text{and} \quad \ell \geq \ell_b(m, \mathbf{h}, p^\star_{\mathbf{X}}, \|p\|^{\mathbf{X}}_\infty, \deg(p)) \quad \Longrightarrow p \in \mathcal{Q}_\ell(\mathbf{h}). \tag{23}$$

5.2 Existence and regularity of minimizer: There exists a minimizer $w^\star$ of the GMP (2′). To be precise, we generalize the GMP (2′) to allow for a larger class of functions than only polynomials, such that a (smooth) minimizer with desired optimal cost exists.

5.2 Quantitative inward-pointing condition: We have access to quantitative estimates of "how much the inward-pointing direction is pointing inward". That is, we can bound from below the function $\psi(\theta)$ given by

$$\psi(\theta) := \inf_{\mathbf{x}\in\mathbf{X}} \overline{\mathcal{A}}'(w + \theta\phi)(\mathbf{x}) - g(\mathbf{x}) > 0.$$

Remark 3.17 (A need for embeddings and smooth extensions)**.**

The effective degree bounds for Putinar's Positivstellensatz **Theorem 3.7** works with $\|p\|^{\mathbf{K}}_\infty$ for the ambient set $\mathbf{K} = [-1, 1]^m \supset \mathbf{X}$. This comes with the disadvantage that we need to bound the polynomials of interest on a larger set $\mathbf{K}$ for which we have less or even no a-priori knowledge of the behavior of those polynomials. This may lead to additional terms in the degree bound that depend exponentially on the degree d, see **Lemma 4.5**. Thus, we propose to (smoothly) extend the minimizer to the larger set $\mathbf{K}$ and approximate the minimizer on this set; for this reason, we recall an extension theorem for smooth functions in appendix. In **Theorem 4.10** and **Theorem 5.6** we demonstrate that an extension argument, indeed, avoids the appearance of exponential degree bounds.

Conditions 5.2 and **5.3** are specifically formulated to keep simultaneous control of $\deg(p)$, $p^\star_{\mathbf{X}}$ and $\|p\|^{\mathbf{X}}_\infty$ (ideally; in the present case, we rather have to control $\|p\|^{\mathbf{K}}_\infty$), allowing for a simple use of the effective Positivstellensatz. Hence, the "only" remaining difficulty lies in effectively verifying those two conditions. Examples on how the above three concepts work together for obtaining convergence rates are demonstrated in the following sections, in which we focus on optimal control problems and volume computation of semi-algebraic sets.

4 Application: dynamical systems

We are now going to demonstrate the methodology on instances of the moment-SoS hierarchy related to the study of dynamical systems, such as optimal control [32] or stochastic differential equations [14].

4.1 Optimal control

In this section, we consider the infinite horizon optimal control problem as presented in [22]:

$$\begin{aligned} V^\star(\mathbf{y}_0) := \inf_{\mathbf{y}(\cdot),\mathbf{u}(\cdot)} & \int_0^\infty \mathrm{e}^{-\beta t}\; g(\mathbf{y}(t),\mathbf{u}(t))\;\mathrm{d}t \\ \text{s.t.}\;\; & \mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}(\mathbf{y}(s),\mathbf{u}(s))\;\mathrm{d}s \\ & \mathbf{y}(t)\in\mathbf{Y},\quad \mathbf{u}(t)\in\mathbf{U} \end{aligned} \tag{24}$$

with discount factor $\beta > 0$, $\mathbf{f} \in \mathbb{R}[\mathbf{y},\mathbf{u}]^n$, $g \in \mathbb{R}[\mathbf{y},\mathbf{u}]$ and compact basic semi-algebraic sets $\mathbf{Y} := \mathbf{S}(\mathbf{h}_{\mathbf{Y}}) \subset \mathbb{R}^n$, $\mathbf{U} := \mathbf{S}(\mathbf{h}_{\mathbf{U}}) \subset \mathbb{R}^{m_{\mathbf{u}}}$, $\mathbf{h}_{\mathbf{Y}} \in \mathbb{R}[\mathbf{y}]^{r_{\mathbf{y}}}$, $\mathbf{h}_{\mathbf{U}} \in \mathbb{R}[\mathbf{u}]^{r_{\mathbf{u}}}$.

If the function $\mathbf{y}_0 \longmapsto V^\star(\mathbf{y}_0)$ is continuously differentiable, then it satisfies the Hamilton-Jacobi-Bellman inequality

$$g - \beta\, V - \mathbf{f}\cdot\mathbf{grad}\, V \geq 0 \text{ on } \mathbf{X} := \mathbf{Y}\times\mathbf{U} \tag{25}$$

and for any V satisfying (25) it holds $V \leq V^\star$ on $\mathbf{Y}$, see for instance [22]. Hence, for any probability measure $\mu_0 \in \mathcal{M}(\mathbf{Y})_+$ (i.e. s.t. $\mu_0(\mathbf{Y}) = 1$) defining a random initial condition Y_0[3]:

$$\begin{aligned} \mathbb{E}_{\mu_0}[V^\star(Y_0)] \geq \sup_{V\in C^1(\mathbf{Y})} & \int V(\mathbf{y}_0)\;\mathrm{d}\mu_0(\mathbf{y}_0) =: \mathbb{E}_{\mu_0}[V(Y_0)] \\ \text{s.t.}\quad & g - \beta\, V - \mathbf{f}\cdot\mathbf{grad}\, V \geq 0 \text{ on } \mathbf{X} := \mathbf{Y}\times\mathbf{U} \end{aligned} \tag{26a}$$

which can in turn be approximated, defining $r := r_{\mathbf{y}} + r_{\mathbf{u}}$, $\mathbf{h} := (\mathbf{h}_{\mathbf{Y}},\mathbf{h}_{\mathbf{U}}) \in \mathbb{R}[\mathbf{y},\mathbf{u}]^r$, by the hierarchy of SoS strengthenings:

$$\begin{aligned} \overline{V_\ell}(\mu_0) := \sup_{V\in\mathbb{R}_{d_\ell}[\mathbf{y}]} & \int V(\mathbf{y}_0)\;\mathrm{d}\mu_0(\mathbf{y}_0) \\ \text{s.t.}\quad & g - \beta\, V - \mathbf{f}\cdot\mathbf{grad}\, V \in \mathcal{Q}_\ell(\mathbf{h}). \end{aligned} \tag{26b}$$

The approximation scheme and its convergence rely on the following Condition (see [22, **Assumption 1**]):

Condition 6. The following conditions hold:

1. $\mathbf{X} = \mathbf{Y}\times\mathbf{U} \subset \mathbf{B} = \{\mathbf{x}\in\mathbb{R}^m\,;\mathbf{x}^\top\mathbf{x} \leq 1\}$ where $m = n + m_{\mathbf{u}}$

[3] This includes the deterministic setting under the form $\mu_0 = \delta_{\mathbf{y}_0}$, where $\mathbf{y}_0 \in \mathbf{Y}$ and $\delta_{\mathbf{y}_0}$ is the Dirac measure in $\mathbf{y}_0$ s.t. for all Borel measurable $\mathbf{A} \subset \mathbf{Y}$, $\delta_{\mathbf{y}_0}(\mathbf{A}) = 1$ if $\mathbf{y}_0 \in \mathbf{A}$, 0 else. Then, $\mathbb{P}(Y_0 = \mathbf{y}_0) = \mu_0(\{\mathbf{y}_0\}) = 1$: Y_0 is deterministic, and $\mathbb{E}_{\mu_0}[V^\star(Y_0)] := \int V^\star\;\mathrm{d}\mu_0 = V^\star(\mathbf{y}_0)$.

2. $\mathbf{h}(\mathbf{0}) \in \mathbb{R}^r_\oplus$ (i.e. the interior of $\mathbf{X}$ contains $\mathbf{0}$).
3. $V^\star \in C^{1,1}(\mathbf{X})$, that is $V^\star$ is differentiable and $\mathbf{grad}\, V^\star$ is Lipschitz continuous on $\mathbf{X}$ (thus the sup in (26a) is attained).
4. For all $\mathbf{y} \in \mathbf{Y}$, the set $\mathbf{f}(\mathbf{y}, \mathbf{U})$ and the map $\mathbf{v} \mapsto \inf\{g(\mathbf{y}, \mathbf{u}) \,;\, \mathbf{u} \in \mathbf{U} \text{ and } \mathbf{f}(\mathbf{x}, \mathbf{u}) = \mathbf{v}\}$ are convex.

Using **Condition 6**.1 one can complement the description of $\mathbf{X} = \mathbf{S}(\mathbf{h})$ with the polynomial

$$h_{r+1}(\mathbf{x}) := 1 - \mathbf{x}^\top \mathbf{x},$$

so that **Assumption 3** always holds.

Theorem 4.1 ([22, **Theorem 1**]).

Let $\mu_0 \in \mathcal{M}(\mathbf{Y})_+$ be a probability measure. Then, under **Condition 6**,

1. *There is $\ell_0 \in \mathbb{N}$ s.t. $\forall \ell \geq \ell_0$, $\overline{V_\ell}(\mu_0) > -\infty$, i.e. (26b) is feasible.*
2. *Any V feasible for (26b) satisfies $V \leq V^\star$ (lower bounds on the value function).*
3. *$\forall \ell \geq \ell_0$, (26b) has an optimal solution $V_\ell^\star \in \mathbb{R}_{d_\ell}[\mathbf{x}]$ s.t. $\overline{V_\ell}(\mu_0) = \mathbb{E}_{\mu_0}[V_\ell^\star]$.*
4. *$\mathbb{E}_{\mu_0}[V^\star(X_0)] - \overline{V_\ell}(\mu_0) = \mathbb{E}_{\mu_0}[V^\star(X_0) - V_\ell^\star(X_0)] \underset{\ell\to\infty}{\longrightarrow} 0$ i.e. $V_\ell^\star$ converges to $V^\star$ in $L^1(\mu_0)$.*

In [22], the authors give an upper bound on the convergence rate for item *4.* in **Theorem 4.1**, which we will now improve using our function approximation results and [2].

Lemma 4.2. *For $d \in \mathbb{N}$ let $V_d \in \mathbb{R}_d[\mathbf{y}]$ with $\|V_d - V^\star\|_{C^1(\mathbf{Y})} \leq \frac{c_1}{d}$, where the constant c_1 is deduced from* **Corollary 3.11**. *For any $\eta > 0$ let $V_{d,\eta} := V_d - \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) - \eta \in \mathbb{R}_d[\mathbf{y}]$. Then $V_{d,\eta}$ satisfies*

$$\|V_{d,\eta} - V^\star\|_{C^1(\mathbf{Y})} \leq \left(2 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right)\frac{c_1}{d} + \eta \tag{27a}$$

and

$$g - \beta V_{d,\eta} + \mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta} \geq \beta\eta \quad \text{on } \mathbf{X} = \mathbf{Y} \times \mathbf{U}. \tag{27b}$$

Proof. The arguments can be found in [19, Lemma 3] and [19, Lemma 2]. Since the arguments are short, we will state the proof here as well. First, we compute c_1 using **Corollary 3.11**: There exists $V_d \in \mathbb{R}_d[\mathbf{y}]$ such that

$$\|V_d - V^\star\|_{C^1(\mathbf{Y})} \leq c_0\, \omega^{L^\infty}_{V^\star,1}(\mathbf{Y}, 1/d)$$

with, using the Lipschitz condition on $\mathbf{grad}\, V^\star$ given in **Condition 6**, $\omega^{L^\infty}_{V^\star,1}(\mathbf{Y}, r) \propto r$, yielding the constant c_1 such that $\omega^{L^\infty}_{V^\star,k}(\mathbf{Y}, r) \leq c_1 r$. Then, we have

$$\|V_{d,\eta} - V^\star\|_{C^1(\mathbf{Y})} \leq \|V_d - V^\star\|_{C^1(\mathbf{Y})} + |V_{d,\eta} - V_d\|_{C^1(\mathbf{Y})} \leq \frac{c_1}{d} + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) + \eta.$$

This is (27a). For (27b) we compute

$$\begin{aligned}
g - \beta V_{d,\eta} + \mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta} &= \underbrace{g - \beta V^\star + \mathbf{f} \cdot \mathbf{grad}\, V^\star}_{\geq 0} + \beta(V^\star - V_{d,\eta}) + \mathbf{f} \cdot \mathbf{grad}(V_{d,\eta} - V^\star) \\
&\geq 0 + \beta\left(V^\star - V_d + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) + \eta\right) + \mathbf{f} \cdot \mathbf{grad}(V_d - V^\star) \\
&\geq \beta\left(-\|V^\star - V_d\|_\infty^{\mathbf{Y}} + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) + \eta\right) - \|\mathbf{grad}(V^\star - V_d)\|_\infty^{\mathbf{Y}} \|\mathbf{f}\|_\infty^{\mathbf{X}} \\
&\geq \beta\left(-\|V^\star - V_d\|_{C^1(\mathbf{Y})} + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) + \eta\right) - \|V^\star - V_d\|_{C^1(\mathbf{Y})} \|\mathbf{f}\|_\infty^{\mathbf{X}} \\
&\geq \beta\left(-\frac{c_1}{d} + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) + \eta\right) - \frac{c_1}{d}\|\mathbf{f}\|_\infty^{\mathbf{X}} \\
&= \beta\eta.
\end{aligned}$$

$\square$

Remark 4.3 (Inward pointing condition for smooth control)**.**

Notice that **Lemma 4.2** is essentially a statement of satisfaction of the inward pointing **Condition 4** while keeping control on the objective, in the special case of the optimal control problem with continuously differentiable value function.

Now we are already in the position to apply an effective version of Putinar's Positivstellensatz.

Theorem 4.4 (Effective Putinar for optimal control)**.**

Let $d \in \mathbb{N}$ and $\eta > 0$. Under **Condition 6**, *if d is large enough and $d_{\mathbf{f}} := \deg(f)+d$, there exist $A, B, C \in \mathbb{R}_{\oplus}$ so that $V_{d,\eta}$ is feasible for (26b) for any*

$$\ell \geq \gamma d_{\mathbf{f}}^{3.5m\text{Ł}} \left(\frac{A}{\eta} + \frac{B}{\eta\, d} + 1\right)^{2.5m\text{Ł}} \left(1 + C^{d_{\mathbf{f}}+1} \cdot d_{\mathbf{f}}^2/4\right)^{2.5m\text{Ł}}. \tag{28a}$$

Moreover, if

$$\varepsilon > \left(2 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right)\frac{c_1}{d} \tag{28b}$$

and

$$\eta \leq \varepsilon - \left(2 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right)\frac{c_1}{d}, \tag{28c}$$

then it holds

$$\mathbb{E}_{\mu_0}[V_{d,\eta}(Y_0)] \geq \mathbb{E}_{\mu_0}[V^\star(Y_0)] - \varepsilon, \tag{28d}$$

hence proving that $0 \leq \mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0) \leq \varepsilon$.

Proof. Notice that (28b) is only to ensure (28c) has at least one solution η. We use **Theorem 3.7**: Denoting $p := g - \beta\, V_{d,\eta} - \mathbf{f}\cdot \mathbf{grad}\, V_{d,\eta} \geq \beta\,\eta > 0$, we know that $p \in \mathcal{Q}_\ell(\mathbf{h})$ (i.e. $V_{d,\eta}$ is feasible for (26b)) for

$$\ell \geq \gamma \deg(p)^{3.5m\text{Ł}} \left(\|p\|/p^\star_{\mathbf{X}}\right)^{2.5m\text{Ł}}. \tag{29a}$$

Hence we only need to estimate bounds on $\|p\|$, $\deg(p)$ and $p^\star_{\mathbf{X}}$.

- $p^\star_{\mathbf{X}} \geq \beta\,\eta$ is a direct consequence of how we constructed $V_{d,\eta}$ in **Lemma 4.2**.
- $\deg(p) \leq \max(\deg(g), \deg(V_{d,\eta}), \deg(\mathbf{f}\cdot\mathbf{grad}\, V_{d,\eta})) \leq d_{\mathbf{f}} := \deg(\mathbf{f}) + d$.
- Estimating $\|p\|$ is the difficult part. First, we have, denoting $q := g - \beta\, V^\star - \mathbf{f}\cdot\mathbf{grad}\, V^\star \geq 0$:

$$\begin{aligned}
\|p\|_\infty^{\mathbf{X}} - \|q\|_\infty^{\mathbf{X}} &\leq \|p-q\|_\infty^{\mathbf{X}} \\
&= \|\beta(V^\star - V_{d,\eta}) + \mathbf{f}\cdot\mathbf{grad}(V^\star - V_{d,\eta})\|_\infty^{\mathbf{X}} \\
&= \left\|\beta\left(V^\star - V_d + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) + \eta\right) + \mathbf{f}\cdot\mathbf{grad}(V^\star - V_d)\right\|_\infty^{\mathbf{X}} \\
&\leq \beta\left(\|V^\star - V_d\|_\infty^{\mathbf{X}} + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right) + \eta\right) + \|\mathbf{f}\|_\infty^{\mathbf{X}}\,\|\mathbf{grad}(V^\star - V_d)\|_\infty^{\mathbf{X}} \\
&\leq (\beta + \|\mathbf{f}\|_\infty^{\mathbf{X}})\left(\|V^\star - V_d\|_{C^1(\mathbf{Y})} + \frac{c_1}{d}\right) + \eta\,\beta \\
&\leq 2(\beta + \|\mathbf{f}\|_\infty^{\mathbf{X}})\frac{c_1}{d} + \eta\,\beta
\end{aligned}$$

which gives the upper bound

$$\|p\|_\infty^{\mathbf{X}} \leq \|q\|_\infty^{\mathbf{X}} + 2(\beta + \|\mathbf{f}\|_\infty^{\mathbf{X}})\frac{c_1}{d} + \eta\,\beta. \tag{29b}$$

However, what we are looking for is an upper bound on $\|p\| \geq \|p\|_\infty^{\mathbf{X}}$ (since $\mathbf{X} \subset \mathbf{K} := [-1,1]^m$ by **Condition 6**.1). By equivalence of norms in finite dimensional spaces there is a constant c such that $\|p\| \leq c\|p\|_\infty^{\mathbf{X}}$. We will now compute such a bound using results in [4].

Lemma 4.5. *For any nonnegative polynomial $p \in \mathcal{P}(\mathbf{X})_+$, it holds*

$$\|p\| \leq \left(1 + \frac{\deg(p)^2}{4} \, (2/b)^{\deg(p)+1}\right) \|p\|_\infty^{\mathbf{X}}, \tag{30}$$

where $b \in (0,1)$ is such that $[-b,b]^m \subset \mathbf{X}$ (whose existence is guaranteed by **Condition 6***.2).*

Proof. See **Appendix A**. □

We finally get our bound by reinjecting the bounds on $p_{\mathbf{X}}^\star$, $\deg(p)$ and equations (29b) and (30) in (29a):

$$\ell \geq \gamma d_{\mathbf{f}}^{3.5m\text{Ł}} \underbrace{\left(\frac{\|q\|_\infty^{\mathbf{X}}}{\beta\,\eta} + 2\frac{\beta + \|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta\,\eta}\frac{c_1}{d} + 1\right)^{2.5m\text{Ł}}}_{\geq \|p\|_\infty^{\mathbf{X}}/p_{\mathbf{X}}^\star} \underbrace{\left(1 + \frac{d_{\mathbf{f}}^2}{4}\,(2/b)^{d_{\mathbf{f}}+1}\right)^{2.5m\text{Ł}}}_{\geq \|p\|/\|p\|_\infty^{\mathbf{X}}}$$

which is exactly (28a). Eventually, we compute

$$\begin{aligned} 0 &\leq \mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0) \\ &\leq \mathbb{E}_{\mu_0}[V^\star(Y_0) - V_{d,\eta}(Y_0)] \\ &\leq \|V^\star - V_{d,\eta}\|_\infty^{\mathbf{Y}} \\ &\leq \|V^\star - V_{d,\eta}\|_{C^1(\mathbf{Y})} \\ &\overset{(27a)}{\leq} \left(2 + \frac{\|\mathbf{f}\|_\infty^{\mathbf{X}}}{\beta}\right)\frac{c_1}{d} + \eta \\ &\overset{(28c)}{\leq} \varepsilon. \end{aligned}$$

□

Corollary 4.6 (Convergence rate for optimal control)**.**
Under **Condition 6***, for $\ell \in \mathbb{N}$ large enough it holds*

$$|\mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0)| \in \mathcal{O}(1/\log \ell) \quad \text{as } \ell \to \infty. \tag{31}$$

Proof. We are going to use **Theorem 4.4**. Let $d \in \mathbb{N}$ and take

$$\varepsilon_d := \frac{1}{d}\left(\left(2 + \frac{\|f\|_\infty^{\mathbf{X}}}{\beta}\right)c_1 + 1\right) > \left(2 + \frac{\|f\|_\infty^{\mathbf{X}}}{\beta}\right)\frac{c_1}{d}$$

so that (28b) holds and we can saturate (28c) with

$$\eta := \varepsilon_d - \left(2 + \frac{\|f\|_\infty^{\mathbf{X}}}{\beta}\right)\frac{c_1}{d} = \frac{1}{d}.$$

Then, **Theorem 4.4** ensures that $0 \leq \mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0) \leq \varepsilon_d$ for

$$\ell \geq \gamma d_{\mathbf{f}}^{3.5m\text{Ł}} \,(A\,d + B + 1)^{2.5m\text{Ł}}\,(1 + C^{d_{\mathbf{f}}+1} \cdot d_{\mathbf{f}}^2/4)^{2.5m\text{Ł}} =: \ell_b.$$

Now, an asymptotic equivalent is given by

$$\ell_b \underset{d\to\infty}{\sim} \gamma \frac{A}{4} d^{11m\text{Ł}} C^{(d+\deg(\mathbf{f})+1)2.5m\text{Ł}} \in \underset{d\to\infty}{\mathcal{O}} \left(d^{11m\text{Ł}} C^{2.5m\text{Ł}d}\right)$$

and taking the log yields

$$\log(\ell_b) \in \mathcal{O}(d) = \mathcal{O}(1/\varepsilon_d) \quad \text{as } d \to \infty.$$

Finally, remembering that $0 \leq \mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_{\lceil \ell_b \rceil}}(\mu_0) \leq \varepsilon_d$ and inverting the above asymptotic expression gives the announced result. □

Remark 4.7 (Comparison with [22]).

In [22], using a previous effective version of Putinar's Positivstellensatz, the authors came up with a much worse convergence rate of $1/\log\log \ell$. Using the new effective Putinar Positivstellensatz from [2], we could remove an exponential dependence in the degree and hence sharply improve the convergence rate. Moreover, at the price of some additional assumptions on the state set $\mathbf{Y}$, it is even possible to derive a *polynomial* convergence rate, as we will show now. Indeed, the remaining exponential dependence is an artifact coming from **Lemma 4.5**, but the effective Positivstellensatz actually gives polynomial dependence.

4.2 A polynomial convergence rate

In this paragraph, we derive a convergence rate for the GMP (26b) in which the level ℓ of the hierarchy is polynomial in $\frac{1}{\varepsilon}$ for the relaxation gap $\mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0) \leq \varepsilon$. The idea is to side-step the exponential growth of the degree bound in **Lemma 4.5** that arises from bounding the supremum of a polynomial on the hypercube by its supremum on a smaller cube. Here, we will extend the optimal cost function $V^\star$ to the whole hypercube $[-1,1]^n$ and, only then, approximate it on the full hypercube $[-1,1]^n$ by a polynomial V_d. This allows us to bound $\|V_d\|$ simply by $\|V^\star\|+1$ (for d large enough) instead of $\|V_d\| \leq C^d \|V_d\|_\infty^{\mathbf{Y}}$. For this to work we need to guarantee that there exists an extension of $V^\star$ to $[-1,1]^n$ with sufficient regularity.

To extend $V^\star$ we need to introduce the Hölder spaces and norms: We say a function w belongs to the Hölder space $C^{k,a}(\mathbf{Y})$ for $k \in \mathbb{N}$ and $a \in (0,1]$ if $w \in C^k(\mathbf{Y})$ and its k-th derivative is a-Hölder-continuous, i.e. its a-Hölder coefficient is finite:

$$\zeta_{k,a}^{\mathbf{Y}}(w) := \max_{|\boldsymbol{\alpha}|=k} \sup_{\mathbf{y}\neq\mathbf{y}'\in\mathbf{Y}} \frac{|\partial_{\boldsymbol{\alpha}} w(\mathbf{y}) - \partial_{\boldsymbol{\alpha}} w(\mathbf{y}')|}{\|\mathbf{y}-\mathbf{y}'\|^a} < \infty.$$

For bounded $\mathbf{Y}$, we equip the space $C^{k,a}(\mathbf{Y})$ with the norm

$$\|w\|_{C^{k,a}(\mathbf{Y})} := \|w\|_{C^k(\mathbf{Y})} + \zeta_{k,a}^{\mathbf{Y}}(w).$$

The notion of Hölder regularity is used to state the following condition, which is instrumental in ensuring a polynomial convergence rate instead of the logarithmic one given in **Corollary 4.6**.

Condition 7. The set $\mathbf{Y}$ has $C^{1,1}$ boundary, that is the boundary $\partial\mathbf{Y}$ is locally the graph of a $C^{1,1}$ function in the above sense of having a finite Hölder coefficient.

Next, we provide an extension result for Hölder functions from [12].

Lemma 4.8 (Extension Lemma; [12, Lemma 6.37]).

Let $k \geq 1$ be an integer and $a \in (0,1]$. Let $\mathbf{Y} \subset \mathbb{R}^n$ be compact with $C^{k,a}$ boundary. Let $\boldsymbol{\Omega}$ be an open and bounded set containing $\mathbf{Y}$. Then for every function $w \in C^{k,a}(\mathbf{Y})$ there exists an extension $\bar{w} \in C^{k,a}(\overline{\boldsymbol{\Omega}})$ with $w(\mathbf{y}) = \bar{w}(\mathbf{y})$ for all $\mathbf{y} \in \mathbf{Y}$ and

$$\|\bar{w}\|_{C^{k,a}(\overline{\boldsymbol{\Omega}})} \leq c_2 \, \|w\|_{C^{k,a}(\mathbf{Y})} \tag{33}$$

for some constant $c_2 = c_2(n,k,a,\mathbf{Y},\boldsymbol{\Omega})$ independent of w.

Under **Conditions 7 and 6**.3, **Lemma 4.8** ensures that there exists an extension $V \in C^1([-1,1]^n)$ of $V^\star$ such that **grad** V is Lipschitz continuous and there exists a constant $c_2 = c_2(n,\mathbf{Y})$ such that

$$\|V\|_{C^{1,1}([-1,1]^n)} \leq c_2 \|V^\star\|_{C^{1,1}(\mathbf{Y})}. \tag{34a}$$

For the rest of this paragraph, we follow the same path as previously in this section. That is, by **Corollary 3.11**, let $V_d \in \mathbb{R}_d[\mathbf{y}]$ be a polynomial and c_1 be a constant (independent of V) with

$$\|V - V_d\|_{C^1([-1,1]^n)} \leq \frac{c_1}{d}. \tag{34b}$$

As in **Lemma 4.2**, for $\eta > 0$, we define

$$V_{d,\eta} := V_d - \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|}{\beta}\right) - \eta \in \mathbb{R}_d[\mathbf{y}], \tag{34c}$$

where we recall that $\|\mathbf{f}\| = \max\{|\mathbf{f}(\mathbf{y},\mathbf{u})| \; ; \; (\mathbf{y},\mathbf{u}) \in [-1,1]^m\}$ with $m = n + m_{\mathbf{u}}$. In the following lemma, we show that $V_{d,\eta}$ is strictly feasible.

Lemma 4.9. *Let* **Condition 7** *hold and $c_1 = c_1(n, \mathbf{Y}, V^\star)$, $c_2 = c_2(n, \mathbf{Y})$ be the constants from (34a) and (34b). For $d \in \mathbb{N}$, $\eta > 0$ the function $V_{d,\eta}$ satisfies*

$$\|V_{d,\eta} - V^\star\|_{C^1(\mathbf{Y})} \leq \left(2 + \frac{\|\mathbf{f}\|}{\beta}\right)\frac{c_1}{d} + \eta. \tag{35a}$$

Further, for the polynomial function $p := g - \beta V_{d,\eta} + \mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta} \in \mathcal{P}(\mathbf{X})$ it holds

$$p \geq \beta\eta \quad \text{on } \mathbf{X} = \mathbf{Y} \times \mathbf{U} \tag{35b}$$

and, recalling that $\|p\| = \max\{|p(\mathbf{x})| \; ; \; \mathbf{x} \in [-1,1]^m\}$,

$$\|p\| \leq \|g\| + c_2 \|V^\star\|_{C^{1,1}(\mathbf{Y})}(\beta + \|\mathbf{f}\|) + 2\frac{c_1}{d}(\beta + \|\mathbf{f}\|) + \beta\eta. \tag{35c}$$

Proof. The statements (35a) and (35b) follow similarly to (27a) and (27b) in **Lemma 4.2**. To show (35c), we simply apply the triangle inequality

$$\begin{aligned} \|p\| &= \|g - \beta V_{d,\eta} + \mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta}\| \\ &\leq \|g\| + \beta\|V_{d,\eta}\| + \|\mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta}\| \end{aligned} \tag{36a}$$

and separately bound $\beta V_{d,\eta}$ and $\mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta}$. We begin with $\beta V_{d,\eta}$

$$\begin{aligned} \|\beta V_{d,\eta}\| &\leq \beta\left(\|V\| + \|V_d - V\| + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|}{\beta}\right) + \eta\right) \\ &\leq \beta\left(c_2 \|V^\star\|_{C^{1,1}(\mathbf{Y})} + \frac{c_1}{d} + \frac{c_1}{d}\left(1 + \frac{\|\mathbf{f}\|}{\beta}\right) + \eta\right) \\ &\leq \beta\left(c_2 \|V^\star\|_{C^{1,1}(\mathbf{Y})} + \frac{c_1}{d}\left(2 + \frac{\|\mathbf{f}\|}{\beta}\right) + \eta\right) \end{aligned}$$

where in the last inequality we used (34a) and (34b). For bounding $\mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta}$ we use $\mathbf{grad}\, V_{d,\eta} = \mathbf{grad}\, V_d$ and we have

$$\begin{aligned} \|\mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta}\| &= \|\mathbf{f} \cdot \mathbf{grad}\, V_d\| \\ &\leq \|\mathbf{f}\|\left(\|V\| + \|V_d - V\|\right) \\ &\leq \|\mathbf{f}\|\left(c_2 \|V^\star\|_{C^{1,1}(\mathbf{Y})} + \frac{c_1}{d}\right) \end{aligned}$$

Putting together in (36a) gives (35c). ∎

As in the previous section, **Lemma 4.9** ensures that $V_{d,\eta}$ is an inward pointing perturbation of $V^\star$. Now, all that remains is to apply an effective version of Putinar's Positivstellensatz from **Theorem 3.7**.

Theorem 4.10 (Polynomial rate for optimal control)**.**

Under **Conditions 6 and 7**, *for $\ell \in \mathbb{N}$ large enough it holds*

$$0 \leq \mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0) \in \mathcal{O}\left(\ell^{-\frac{1}{6m\L}}\right) \quad \text{as } \ell \to \infty. \tag{37}$$

Proof. We use the notation and constants from **Lemma 4.9**. Let $V_{d,\eta}$ be as in (34c). For $d \in \mathbb{N}$ we choose $\eta = \eta_d := \frac{1}{d}$. Let $d_0 \in \mathbb{N}$ such that

$$\left(2 + \frac{\|\mathbf{f}\|}{\beta}\right) \frac{c_1}{d_0} + \eta_{d_0} \quad , \quad 2\frac{c_1}{d_0}(\beta + \|\mathbf{f}\|) + \beta \eta_{d_0} \leq 1. \tag{38a}$$

By monotony in d, both terms on the left-hand side in (38a) are bounded by 1 for all $d \geq d_0$. From **Lemma 4.9**, we have

$$\|V_{d,\eta} - V^\star\|_{C^1(\mathbf{Y})} \in \mathcal{O}\left(\frac{1}{d}\right) \quad \text{as } d \to \infty \tag{38b}$$

and for $p := g - \beta V_{d,\eta} + \mathbf{f} \cdot \mathbf{grad}\, V_{d,\eta}$ it holds

$$p \geq \frac{\beta}{d} \quad \text{on } \mathbf{Y} \times \mathbf{U} \tag{38c}$$

and

$$\|p\| \leq \|g\| + c\|V^\star\|_{C^{1,1}(\mathbf{Y})}(\beta + \|\mathbf{f}\|) + 1 =: c_3. \tag{38d}$$

Note that the constant c_3 is independent of d and the choice of extension V. Inserting (38c) and (38d) into **Corollary 3.9**, we get that V_{d,η_d} is feasible for (26b) for any $\ell \in \mathbb{N}$ with

$$\ell \geq \gamma(m, \mathbf{h}) \deg(p)^{3.5m\text{Ł}} \left(\frac{c_3 d}{\beta}\right)^{2.5m\text{Ł}}. \tag{38e}$$

To finalize the proof, recall that $\deg(p) \leq d + \deg(\mathbf{f}) \in \mathcal{O}(d)$. Thus, for given $\ell \in \mathbb{N}$ (large enough), we choose the largest $d = d_\ell \in \mathbb{N}$ (with $d \geq d_0$) such that (38e) is satisfied. By (38e), such d_ℓ is of order $\ell^{\frac{1}{6m\text{Ł}}}$ and we get

$$\begin{aligned}
|\mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0)| &\leq \int |V^\star(\mathbf{y}_0) - V_{d_\ell,\eta_{d_\ell}}(\mathbf{y}_0)| \,\mathrm{d}\mu_0(\mathbf{y}_0) \\
&\leq \|V_{d_\ell,\eta_{d_\ell}} - V^\star\|_\infty^{\mathbf{Y}} \\
&= \|V_{d_\ell,\eta_{d_\ell}} - V^\star\|_{C^1(\mathbf{Y})} \in \mathcal{O}\left(\frac{1}{d_\ell}\right) \in \mathcal{O}\left(\ell^{-\frac{1}{6m\text{Ł}}}\right) \quad \text{as } \ell \to \infty.
\end{aligned}$$

This shows the statement. ☐

Remark 4.11 (Relaxing the regularity assumption on $V^\star$)**.**

The same arguments in the proof of **Theorem 4.10** work still for $V^\star$ with slightly less regularity, namely, for $V^\star \in C^{1,a}(\mathbf{Y})$ and $\mathbf{Y}$ with $C^{1,a}$ boundary for some $a \in (0,1)$. The convergence rate then takes the form

$$\mathbb{E}_{\mu_0}[V^\star(Y_0)] - \overline{V_\ell}(\mu_0) \in \mathcal{O}\left(\ell^{-\frac{1}{2.5m\text{Ł}+3.5m\text{Ł}/a}}\right) \quad \text{as } \ell \to \infty$$

4.3 Exit location of stochastic processes

In this example, we apply our framework to [14], in which the exit location of stochastic processes is computed by a GMP. We recall the setting from [14]. Consider a stochastic differential equation

$$dX_t = \mathbf{f}_0(X_t)\, dt + \mathrm{F}(X_t)\, dB_t, \quad X_0 = \mathbf{x}_0 \tag{39}$$

for $\mathbf{f}_0 = (f_{0i})_i : \mathbb{R}^m \to \mathbb{R}^m$, $\mathrm{F} = (f_{ij})_{i,j} : \mathbb{R}^m \to \mathbb{R}^{m\times n}$, a deterministic initial condition $\mathbf{x}_0$ and $(B_t)_{t\geq 0}$ a n-dimensional Brownian motion. The SDE (39) is equipped with an open, bounded constraint set $\mathbf{X}$ and for a given function $g : \partial\mathbf{X} \to \mathbb{R}$, the expected exit value for $\mathbf{x} \in \mathbf{X}$ is given by

$$v^\star(\mathbf{x}_0) := \mathbb{E}(g(X_\tau)) \tag{40}$$

where $\tau = \inf\{t \geq 0\,;\, X_t \in \partial\mathbf{X}\}$ is the first time at which the process $(X_t)_t$ starting at $X_0 = \mathbf{x}_0$ hits $\partial\mathbf{X}$.

In [14], the following assumptions were made

Condition 8.

8.1 It holds $\overline{\mathbf{X}} = \mathbf{S}(\mathbf{h}) \subset \mathbf{K} = [-1,1]^m$ for some $\mathbf{h} \in \mathbb{R}[\mathbf{x}]^r$.

8.2 The boundary $\partial\mathbf{X}$ is smooth and is represented by $\partial\mathbf{X} = \mathbf{S}(\mathbf{h}_\partial)$ for some $\mathbf{h}_\partial \in \mathbb{R}[\mathbf{x}]^{r_\partial}$.

8.3 We assume $g, f_{0i}, f_{ij} \in \mathbb{R}[\mathbf{x}]$ for $i = 1, \dots, m$ and $j = 1, \dots, n$.

8.4 The matrix $\mathrm{F}(\mathbf{x})\mathrm{F}(\mathbf{x})^\top$ is positive definite for all $\mathbf{x} \in \overline{\mathbf{X}}$.

Remark 4.12. The case when the boundary $\partial\mathbf{X}$ decomposes into several disjoint components $\partial\mathbf{X} = \mathbf{X}_1^\partial \cup \ldots \cup \mathbf{X}_l^\partial$ is treated similarly in [14]. By **Condition 8.2**, we restrict to the notationally simpler case of $\partial\mathbf{X} = \mathbf{S}(\mathbf{h}_\partial)$.

Under the above conditions there exists a unique solution X_t of (39) for $t \leq \tau$, see [14, 9].

In [14], from Dynkin's formula, the following GMP for the value $v^\star(\mathbf{x}_0)$ of the exit value (40) is derived

$$v^\star(\mathbf{x}_0) = \max_v v(\mathbf{x}_0) \quad \text{s.t.} \quad \mathcal{L}v \leq 0 \text{ on } \mathbf{X}, \quad v \leq g \text{ on } \partial\mathbf{X} \tag{41}$$

where $\mathcal{L}$ is the second-order partial differential operator

$$\mathcal{L}v(\mathbf{x}) := -\sum_{i,j=1}^m a_{ij}(\mathbf{x}) \frac{\partial^2 v}{\partial \mathbf{x}_i \partial \mathbf{x}_j}(\mathbf{x}) + \sum_{i=1}^m f_{0i}(\mathbf{x}) \frac{\partial v}{\partial \mathbf{x}_i}(\mathbf{x}) \tag{42}$$

for $(a_{ij}(\mathbf{x}))_{i,j=1,\dots,m} = \mathrm{F}(\mathbf{x})\mathrm{F}(\mathbf{x})^\top$. The moment-SoS hierarchy for the LP (41) reads

$$\begin{aligned} v_\ell^\star(\mathbf{x}_0) := \quad & \sup_{v \in \mathbb{R}[\mathbf{x}]} && v(\mathbf{x}_0) \\ & \text{s.t.} && -\mathcal{L}v \in \mathcal{Q}_\ell(\mathbf{h}) \\ & && g - v \in \mathcal{Q}_\ell(\mathbf{h}_\partial). \end{aligned} \tag{43}$$

The function $v^\star$ is the solution of the following boundary value problem

$$\begin{array}{rcll} \mathcal{L}v & = & 0 & \text{on} \quad \mathbf{X} \\ v & = & g & \text{on} \quad \partial\mathbf{X}. \end{array} \tag{44}$$

Thus, the question about existence and regularity of minimizers of (41) is transferred to the question about existence and regularity of solutions of (44). Fortunately, the answer here is positive, see [12, 51, 14]. Namely, under **Condition 8**, there exists a unique solution $v \in C^\infty(\mathbf{X})$ of (44). Next, we investigate an inward-pointing direction. Therefore, we recall that for $\phi \in C^\infty(\mathbf{X})$ there exists a unique solution $u_\phi \in C^\infty(\overline{\mathbf{X}})$ of

$$\begin{array}{rcll} \mathcal{L}u_\phi & = & \phi & \text{in} \quad \mathbf{X} \\ u_\phi & = & 0 & \text{on} \quad \partial\mathbf{X}. \end{array} \tag{45}$$

To construct an inward-pointing direction let $\phi(\mathbf{x}) := -1$ for all $\mathbf{x} \in \overline{\mathbf{X}}$ and u_ϕ be a corresponding solution to (45). Let $0 < \eta, \theta \in \mathbb{R}$, we define the function

$$v := v^\star + \theta(u_\phi - \eta). \tag{46}$$

In the following lemma we show that v is indeed strictly feasible for all $\theta > 0$; in other words $u_\phi - \eta$ is an inward-pointing direction for $v^\star$.

Lemma 4.13. *For all $\theta > 0$, the function v from (46) is strictly feasible for (41).*

Proof. On $\mathbf{X}$ it holds

$$\mathcal{L}v \;=\; \mathcal{L}(v^\star + \theta(u_\phi - \eta)) = \mathcal{L}v^\star + \theta\mathcal{L}(u_\phi - \eta) = \mathcal{L}v^\star + \theta\mathcal{L}u_\phi = 0 + \theta(-1) = -\theta < 0.$$

For $\mathbf{x} \in \partial\mathbf{X}$ we have

$$v(\mathbf{x}) = v^\star(\mathbf{x}) + \theta(u_\phi(\mathbf{x}) - \eta) = g(\mathbf{x}) + \theta(0 - \eta) = g(\mathbf{x}) - \theta\eta < g(\mathbf{x}).$$

This shows that v is strictly feasible for (41). $\square$

The cost of v for the infinite dimensional LP (41) is given by $v(\mathbf{x}_0) = v^\star(\mathbf{x}_0) + \theta(u_\phi(\mathbf{x}_0) - \eta)$.

Following our procedure from Section 3.2, we obtain the following convergence rate for the moment-SoS hierarchy for (41) from [14].

Theorem 4.14 (Convergence rate for exit location of stochastic systems)**.**

Let **Assumption 3** *hold for* $\mathbf{X}$ *and* $\partial\mathbf{X}$*, and let* $Ł$ *(resp.* $Ł_\partial$*) be the Łojaciewicz exponent of* $\mathbf{h}$ *(resp.* $\mathbf{h}_\partial$*). Then, defining* $\widehat{Ł} := \max(Ł, Ł_\partial)$*, under* **Condition 8***, it holds for* $\ell \in \mathbb{N}$ *large enough that*

$$v^\star(\mathbf{x}_0) - v^\star_\ell(\mathbf{x}_0) \in \mathcal{O}\left(\ell^{-\frac{1}{(2.5+s)m\widehat{Ł}}}\right) \text{ for any } s > 0. \tag{47}$$

Proof. Let $s > 0$, $k := \lceil 3.5/s \rceil$, $v, u \in C^{k+2}(\mathbf{K})$ be extensions of $v^\star$ and u_ϕ according to Lemma 4.8 with

$$\|v\|_{C^{k+2}(\mathbf{K})} \leq c_2 \|v\|_{C^{k+2}(\mathbf{X})} \quad \text{and} \quad \|u\|_{C^{k+2}(\mathbf{K})} \leq c_2 \|u_\phi\|_{C^{k+2}(\mathbf{X})} \tag{48a}$$

for some constant $c_2 \in \mathbb{R}$. For $d \in \mathbb{N}$, by Theorem 3.10, let $p_q, q_d \in \mathbb{R}_d[\mathbf{x}]$ with

$$\|v - p_d\|_{C^{k+2}(\mathbf{K})}, \|u - q_d\|_{C^{k+2}(\mathbf{K})} \leq \frac{c_1}{d^k} \tag{48b}$$

for some constant $c_1 \in \mathbb{R}$. Further, we set

$$A := \sup_{\mathbf{x}\in\mathbf{K}} \sum_{i,j=1}^m |a_{ij}(\mathbf{x})| + \sum_{i=1}^m |f_{0i}(\mathbf{x})|.$$

We define $\theta_d, \eta_d > 0$ for large enough $d \in \mathbb{N}$ as

$$\theta_d := \frac{2c_2 A}{d^k(1 - c_1 A/d^k)} \in \mathcal{O}(d^{-k}) \quad \text{and} \quad \eta_d := \frac{c_2}{d^k}(1 + 2\theta_d^{-1}) \in \mathcal{O}(1). \tag{48c}$$

Motivated by (46), we define v_d by

$$v_d := p_d + \theta_d(q_d - \eta_d) \tag{48d}$$

and verify that v_d is feasible for (43) for $\ell \in \mathbb{N}$ to be determined. We first bound v_d on $\mathbf{K}$. On $\mathbf{K}$ we have for large enough $d \in \mathbb{N}$

$$\begin{aligned}
|v_d| &\leq |v| + |v - p_d| + \theta_d(|u| + |u - q_d| + \eta_d) \\
&\leq \|v\|^{\mathbf{K}}_\infty + \|v - p_d\|^{\mathbf{K}}_\infty + \theta_d(\|u\|^{\mathbf{K}}_\infty + \|u - q_d\|^{\mathbf{K}}_\infty + \eta_d) \\
&\leq \|v\|_{C^2(\mathbf{K})} + \|v - p_d\|_{C^2(\mathbf{K})} + \theta_d(\|u\|_{C^2(\mathbf{K})} + \|u - q_d\|_{C^2(\mathbf{K})} + \eta_d) \\
&\leq c_2\|v^\star\|_{C^{k+2}(\mathbf{X})} + \frac{c_1}{d^k} + \theta_d c_2 \|u_\phi\|_{C^{k+2}(\mathbf{X})} + \theta_d c_2 \|v^\star\|_{C^{k+2}(\mathbf{X})} + \theta_d \frac{c_1}{d^k} + \theta_d \eta_d \\
&\leq c_2\|v^\star\|_{C^{k+2}(\mathbf{X})} + 1 + \theta_d c_2 \|u_\phi\|_{C^{k+2}(\mathbf{X})} + 3 =: C_1.
\end{aligned}$$

Similarly, we can bound $\mathcal{L}v_d$ on $\mathbf{K}$. Note first that for all $\mathbf{x} \in \mathbf{K}$ we have for all $w \in C^2(\mathbf{K})$

$$\begin{aligned}
|\mathcal{L}w(\mathbf{x})| &= \left| \sum_{i,j=1}^m a_{ij}(\mathbf{x})\partial_i\partial_j w(\mathbf{x}) + \sum_{i=1}^m f_{0i}(\mathbf{x})\partial_i w(\mathbf{x}) \right| \\
&\leq \left(\sum_{i,j=1}^m |a_{ij}(\mathbf{x})| + \sum_{i=1}^m |f_{0i}(\mathbf{x})| \right) \|w\|_{C^2(\mathbf{K})} \leq A\|w\|_{C^2(\mathbf{K})}.
\end{aligned}$$

For large enough $d \in \mathbb{N}$ we get on $\mathbf{K}$

$$\begin{aligned}
|\mathcal{L}v_d| &\leq |\mathcal{L}v| + |\mathcal{L}(v - p_d)| + \theta(|\mathcal{L}u| + |\mathcal{L}(u - q_d)|) \\
&\leq A\|v\|_{C^2(\mathbf{K})} + A\|v - p_d\|_{C^2(\mathbf{K})} + \theta A \|u\|_{C^2(\mathbf{K})} + \theta A \|u - q_d\|_{C^2(\mathbf{K})} \\
&\leq Ac_2\|v^\star\|_{C^{k+2}(\mathbf{X})} + A\frac{c_1}{d^k} + \theta A \|u_\phi\|_{C^{k+2}(\mathbf{X})} + \theta A \frac{c_1}{d^k} \\
&\leq Ac_2\|v^\star\|_{C^{k+2}(\mathbf{X})} + 3 =: C_2.
\end{aligned}$$

Next, we verify strict feasibility of v_d for (43). It holds on $\mathbf{X}$

$$\begin{aligned}
\mathcal{L}v_d &= \underbrace{\mathcal{L}v^\star}_{=0} + \mathcal{L}(p_d - v^\star) + \theta_d(\underbrace{\mathcal{L}u_\phi}_{=-1} + \mathcal{L}(q_d - u_\phi) + \eta_d) \\
&\leq A\|v^\star - p_d\|_{C^{k+2}(\mathbf{X})} - \theta + \theta A\|u_\phi - q_d\|_{C^{k+2}(\mathbf{X})} \\
&\leq A\|v - p_d\|_{C^{k+2}(\mathbf{X})} - \theta + \theta A\|u - q_d\|_{C^{k+2}(\mathbf{X})} \\
&\leq \theta_d(-1 + A\frac{c_1}{d^k}) + \frac{Ac_1}{d^k} \\
&\overset{(48c)}{=} -\frac{Ac_1}{d^k}
\end{aligned} \tag{48e}$$

and on $\partial\mathbf{X}$ it holds

$$\begin{aligned}
v_d &= \underbrace{v^\star}_{=g} + p_d - v^\star + \theta_d(\underbrace{u_\phi}_{=0} + q_d - u_\phi - \eta_d) \\
&\leq g + \|v^* - p_d\|_\infty^{(\partial\mathbf{X})} + \theta\|u_\phi - q_d\|_\infty^{(\partial\mathbf{X})} - \theta_d\eta_d \\
&\leq g + \theta_d(1 + \frac{c_1}{d^k} - \eta_d) \\
&\overset{(48c)}{=} g - \frac{c_2}{d^k}.
\end{aligned} \tag{48f}$$

Applying Theorem 3.7, we get that v_d is feasible for $\ell \in \mathbb{N}$ with

$$\ell \geq d^{3.5m\widehat{Ł}} \max\left\{\gamma(m,\mathbf{h})\left(\frac{C_2 d^k}{Ac_1}\right)^{2.5mŁ}, \gamma(m,\mathbf{h}_\partial)\left(\frac{C_1 d^k}{c_2}\right)^{2.5mŁ_\partial}\right\} \in \mathcal{O}\left(d^{(3.5+2.5k)m\widehat{Ł}}\right). \tag{48g}$$

For such ℓ the optimal value $v_\ell^\star$ is at least $v_d(\mathbf{x}_0)$; hence we get

$$\begin{aligned}
v^\star(\mathbf{x}_0) - v_\ell^\star(\mathbf{x}_0) &\leq (v^\star - v^\star - (p_d - v^\star) - \theta_d(u_\phi + q_d - u_\phi - \eta_d))(\mathbf{x}_0) \\
&\leq \frac{c_1}{d^k} + \theta_d(\|u_\phi\|_\infty^{\mathbf{X}} + \frac{c_1}{d^k} + \eta_d) \in \mathcal{O}\left(d^{-k}\right).
\end{aligned} \tag{48h}$$

Defining $\varepsilon_d := d^{-k}$, (48g) yields that v_d is feasible for $\ell \geq \ell_b \in \mathcal{O}\left(\varepsilon_d^{-\frac{1}{(3.5/k+2.5)m\widehat{Ł}}}\right)$ so that (48h) ensures that $|v^\star(\mathbf{x}_0) - v_{\ell_b}^\star(\mathbf{x}_0)| \leq \varepsilon_d \in \mathcal{O}\left(\ell_b^{-\frac{1}{(3.5/k+2.5)m\widehat{Ł}}}\right) \subset \mathcal{O}\left(\ell_b^{-\frac{1}{(s+2.5)m\widehat{Ł}}}\right)$, which is the announced result. $\square$

5 Application: Volume computation

In this section, we analyze the moment-SoS hierarchy for computing the volume $\lambda(\mathbf{X})$ of a bounded basic semi-algebraic set

$$\mathbf{X} := \mathbf{S}(\mathbf{h}) = \{\mathbf{x} \in \mathbb{R}^m \,;\, h_1(\mathbf{x}) \geq 0, \ldots, h_r(\mathbf{x}) \geq 0\} \subset \mathbf{B}$$

with $r \geq 1$ integer and $h_1, \ldots, h_r \in \mathbb{R}[\mathbf{x}]$.

A standard moment-SoS hierarchy method was proposed in [16], but with a bad convergence behavior both in practice and in theory (see [19] for convergence rates), due to a Gibbs phenomenon occurring in the SoS approximations. An alternative formulation was proposed in [31], with much better numerical behavior, which was recently supported by a qualitative analysis in [49], showing that no Gibbs phenomenon occurs in this improved formulation. In this section we complement the existing qualitative analysis with a first quantitative analysis of how much better the convergence rate is in the improved formulation.

5.1 The standard approach

The standard moment-SoS approach to numerically solve the volume problem is discussed in detail in [16]. The method consists in formulating a GMP whose optimal solution is $\lambda(\mathbf{X})$, after what one numerically approximates this optimal solution using the moment-SoS hierarchy. We have discussed this example in Section 2.3 and recall its moment-SoS hierarchy:

$$\begin{aligned} \lambda(\mathbf{X}) = \max_{\mu \in \mathcal{M}(\mathbf{X})} & \int 1 \, \mathrm{d}\mu \\ \text{s.t. } & \mu \in \mathcal{M}(\mathbf{X})_+ \\ & \lambda_{\mathbf{Y}} - \mu \in \mathcal{M}(\mathbf{Y})_+ \end{aligned} \tag{4}$$

$$\begin{aligned} \lambda(\mathbf{X}) = \inf_{w \in \mathcal{P}(\mathbf{Y})} & \int w \, \mathrm{d}\lambda_{\mathbf{Y}} \\ \text{s.t. } & w|_{\mathbf{X}} - 1 \in \mathcal{C}(\mathbf{X})_+ \\ & w \in \mathcal{C}(\mathbf{Y})_+ \end{aligned} \tag{4'}$$

where $\mathbf{Y}$ contains $\mathbf{X}$ and is an Archimedean basic-semialgebraic set, $\lambda_{\mathbf{Y}}$ denotes the Lebesgue measure on $\mathbf{Y}$ and these are such that the numbers $\int \mathbf{y}^{\boldsymbol{\beta}} \, \mathrm{d}\lambda_{\mathbf{Y}}(\mathbf{y})$, $\boldsymbol{\beta} \in \mathbb{N}^m$, are known. Here it is straightforward to show that the dual problem (4′) neither has a polynomial nor even continuous optimal solution w.

In this subsection, the convergence rate for the hierarchy of SoS programs obtained in [22] for the volume problem considered in [16] is improved with the help of **Theorem 3.7**, with convention

$$\mathbf{Y} = \mathbf{K} = [-1,1]^m = \mathbf{S}(\mathbf{f}) \quad \text{with} \quad \mathbf{f} = (1 - x_1^2, \ldots, 1 - x_m^2) \tag{50}$$

(which is the best choice for computing the convergence rate; notice that $\mathbf{Y}$ can be chosen arbitrarily here without changing the optimal value $\lambda(\mathbf{X})$). Let us consider a hierarchy of problems from [22] (which can be regarded as SoS strengthenings of Problem (4′)):

$$\begin{aligned} \mathrm{d}_{\mathbf{X}}^{\ell} := \inf_{w \in \mathbb{R}[\mathbf{y}]} \quad & \int w \, \mathrm{d}\lambda_{\mathbf{Y}} \\ \text{s.t.} \quad & w - 1 \in \mathcal{Q}_{\ell}(\mathbf{h}), \\ & w \in \mathcal{Q}_{\ell}(\mathbf{f}). \end{aligned} \tag{51}$$

To compute the rate of convergence of (51) we need to estimate the dependence of the degree ℓ on $\varepsilon > 0$ for which it holds $|\mathrm{d}_{\mathbf{X}}^{\ell} - \lambda(\mathbf{X})| < \varepsilon$.

We shall use the standard condition from [22]:

Condition 9 (Finite one-sided Gibbs phenomenon)**.**

There exists a constant $c_G \geq 0$ depending only on $\mathbf{X}$ such that problem (4′) admits a minimizing sequence $\{w_d\}_{d \in \mathbb{N}}$ with for all $d \in \mathbb{N}$, $\deg(w_d) \leq d$ and $\max\{w_d(\mathbf{x}) \; ; \; \mathbf{x} \in \mathbf{K}\} \leq c_G$.

Remark 5.1 (On finite Gibbs phenomena)**.**

It is well known in Fourier analysis that the Gibbs phenomenon that occurs when approximating a discontinuous periodic function ϕ with trigonometric polynomials induces an overshoot of approximately 9%, and thus the polynomial approximation is uniformly bounded by some constant c_ϕ that only depends on ϕ. This is also the case for generic L^1 approximation of discontinuous functions with algebraic polynomials [7]. However, to our best knowledge, these results have not been extended to *one-sided* polynomial approximations, as it is the case here. Following [19], we conjecture (which is supported by the numerical experiments displayed in [16, 31, 50, 49]) that **Condition 9** also holds generically.

Theorem 5.2 (Effective Putinar for volume computation)**.**

Define $\gamma(m, \mathbf{f}, \mathbf{h}) := \max(\gamma(m, \mathbf{f}), \gamma(m, \mathbf{h}))$ and $\widehat{Ł} := \max(Ł, 1)$ where $Ł$ is the Łojaciewicz exponent of $\mathbf{h}$*. Then, under* **Condition 9**, *there exists $C > 0$ such that, for all $\varepsilon \in (0,1)$ it holds $\mathrm{d}_{\mathbf{X}}^{\ell} - \lambda(\mathbf{X}) < \varepsilon$ for any*

$$\ell \geq \gamma(m, \mathbf{f}, \mathbf{h}) \; (C/\varepsilon)^{3.5m\widehat{Ł}} \left(1 + 2^{m+1} \frac{c_G}{\varepsilon}\right)^{2.5m\widehat{Ł}} \in \mathcal{O}\left(\frac{1}{\varepsilon^{6m\widehat{Ł}}}\right) \tag{52}$$

Proof. We first notice that for any $d, \ell \in \mathbb{N}$ and any $w \in \mathbb{R}_d[\mathbf{x}]$ feasible for (51) at order ℓ, it holds

$$0 \leq \mathrm{d}_{\mathbf{X}}^{\ell} - \lambda(\mathbf{X}) = \mathrm{d}_{\mathbf{X}}^{\ell} - \int \mathbb{1}_{\mathbf{X}} \, \mathrm{d}\lambda_{\mathbf{K}} \leq \int (w - \mathbb{1}_{\mathbf{X}}) \, \mathrm{d}\lambda_{\mathbf{K}} \leq \int (w_d - \mathbb{1}_{\mathbf{X}}) \, \mathrm{d}\lambda_{\mathbf{K}} + \int |w - w_d| \, \mathrm{d}\lambda_{\mathbf{K}}, \tag{53a}$$

where w_d comes from **Condition 9**. Let $\varepsilon > 0$. From **Theorem 3.12**, we know that

$$\int (w_d - \mathbb{1}_{\mathbf{X}})\, \mathrm{d}\lambda_{\mathbf{K}} \leq \overline{c}\, \omega^{L^1}_{\mathbb{1}_{\mathbf{X}},0}(\lambda_{\mathbf{K}}, {}^1\!/_d), \tag{53b}$$

with

$$\omega^{L^1}_{\mathbb{1}_{\mathbf{X}},0}(\lambda_{\mathbf{K}}, {}^1\!/_d) = \int \sup\left\{ |\mathbb{1}_{\mathbf{X}}(\mathbf{x}) - \mathbb{1}_{\mathbf{X}}(\mathbf{y})| \;;\; \mathbf{x} \in \mathbf{K},\ \|\mathbf{x}-\mathbf{y}\| \leq {}^1\!/_d \right\} \mathrm{d}\lambda_{\mathbf{K}}(\mathbf{y}).$$

From [19, **Lemma 1**], there exists a $c_4 \geq 0$ depending only on $\mathbf{X}$ such that

$$\omega^{L^1}_{\mathbb{1}_{\mathbf{X}},0}(\lambda_{\mathbf{K}}, {}^1\!/_d) \leq \frac{c_4}{d},$$

which we reinject into (53b) to get (introducing $C := 2\,\overline{c}\,c_4$)

$$\int (w_d - \mathbb{1}_{\mathbf{X}})\, \mathrm{d}\lambda_{\mathbf{K}} \leq \frac{C}{2d}. \tag{53c}$$

It remains to specify w so that we get a good bound on the second term in (53a). Defining $w := w_d + {}^C\!/_{2d\lambda(\mathbf{K})}$, we automatically get

$$\int |w - w_d|\, \mathrm{d}\lambda_{\mathbf{K}} = \frac{C}{2d},$$

which we can reinject into (53a) to get

$$0 \leq \mathrm{d}^\ell_{\mathbf{X}} - \lambda(\mathbf{X}) \leq \frac{C}{d},$$

for any ℓ such that w is feasible in (51) at order ℓ. The last remaining piece is a value for such ℓ, which we compute using again **Theorem 3.7** on both $\mathbf{K}$ and $\mathbf{X}$. We first work on $\mathbf{X}$: we want a lower bound on $\ell \in \mathbb{N}$ such that $p := w - 1 \in \mathcal{Q}_\ell(\mathbf{h})$; denoting Ł the Łojaciewicz exponent of $\mathbf{h}$, it is given by the effective Putinar P-satz as

$$\ell \geq \gamma(m, \mathbf{h})\ \deg(p)^{3.5m\text{Ł}}\, ({}^{\|p\|}\!/_{p^\star_{\mathbf{X}}})^{2.5m\text{Ł}} \qquad \text{with:}$$

- $\deg(p) = \deg(w-1) = \deg(w_d + {}^C\!/_{(2^{m+1}d)} - 1) \leq d$ (as $\lambda(\mathbf{K}) = \lambda[-1,1]^m = 2^m$)
- $\|p\| = \max\{p(\mathbf{x}) \;;\; \mathbf{x} \in [-1,1]^m\} \leq c_G + {}^C\!/_{(2^{m+1}d)} - 1$ (because $\mathbf{K} = [-1,1]^m$)
- $p^\star_{\mathbf{X}} = \min\{p(\mathbf{x}) \;;\; \mathbf{x} \in \mathbf{X}\} \geq {}^C\!/_{(2^{m+1}d)}$

so that we get

$$\ell \geq \gamma(m,\mathbf{h})\, d^{3.5m\text{Ł}} \left(\frac{c_G + {}^C\!/_{(2^{m+1}d)} - 1}{{}^C\!/_{(2^{m+1}d)}} \right)^{2.5m\text{Ł}} = \gamma(m,\mathbf{h})\, d^{3.5m\text{Ł}} \left(1 + 2^{m+1}\frac{c_G - 1}{C} d \right)^{2.5m\text{Ł}}. \tag{53d}$$

Next, we work on $\mathbf{K} = [-1,1]^m = \mathbf{S}(\mathbf{f})$, for which the Łojaciewicz exponent is 1, and we want a lower bound on ℓ such that $w \in \mathcal{Q}_\ell(\mathbf{f})$, which is again given by the effective Putinar P-satz as

$$\ell \geq \gamma(m, \mathbf{f})\ \deg(w)^{3.5m}\, ({}^{\|w\|}\!/_{w^\star_{\mathbf{K}}})^{2.5m} \qquad \text{with:}$$

- $\deg(w) = \deg(w_d + {}^C\!/_{(2^{m+1}d)}) \leq d$
- $\|w\| = \max\{w(\mathbf{x}) \;;, \mathbf{x} \in \mathbf{K}\} \leq c_G + {}^C\!/_{(2^{m+1}d)}$
- $w^\star_{\mathbf{K}} = \min\{w(\mathbf{x}) \;;\; \mathbf{x} \in \mathbf{K}\} \geq {}^C\!/_{(2^{m+1}d)}$

so that we get

$$\ell \geq \gamma(m,\mathbf{f})\, d^{3.5m} \left(1 + 2^{m+1}\frac{c_G}{C} d \right)^{2.5m} \tag{53e}$$

Eventually, taking ℓ larger than the maximum between the right hand sides of (53d) and (53e) with $d := \lceil {}^C\!/_\varepsilon \rceil$ yields the announced bound.

□

Corollary 5.3. *Using the notations in* **Proposition 5.2** *and under* **Condition 9**, *it holds*

$$\mathrm{d}_{\mathbf{X}}^{\ell} - \lambda(\mathbf{X}) \in \mathcal{O}\left(\ell^{-\frac{1}{6m\hat{L}}}\right) \quad \textit{as} \quad \ell \to \infty.$$

Proof. Simply inverting the expression in (52). □

5.2 Stokes constraints

In this subsection, we investigate the effect of smoothness of optimal solutions of the infinite dimensional LP. We consider the case of only one defining polynomial inequality, i.e. $r = 1$. This means we compute the volume of the *open set*

$$\mathbf{X} := \{\mathbf{x} \in \mathbb{R}^m \,; h(\mathbf{x}) > 0\}$$

and, as in (50), let $\mathbf{K}$ be given by $\mathbf{K} = [-1,1]^m = \mathbf{S}(\mathbf{f})$. Moreover, we add the following condition

Condition 10. It holds $\mathbf{grad}\, h(x) \neq 0$ for all $\mathbf{x} \in \partial\mathbf{X}$, in particular the boundary $\partial\mathbf{X}$ is smooth.

Remark 5.4 (No more Gibbs phenomenon)**.**

Note that now, we do not assume the finite Gibbs phenomenon from **Condition 9**. As we will show, this is because in the following formulations, optimal solutions cease to be discontinuous and thus the Gibbs phenomenon does not occur anymore, see [49] for a more in-depth discussion on that topic.

In [49], a new formulation is designed to cope with the slow convergence of the moment-SoS hierarchy corresponding to (4) and (4′) using the divergence theorem:

$$\begin{aligned}
\lambda(\mathbf{X}) = \max_{\substack{\mu \in \mathcal{M}(\overline{\mathbf{X}}) \\ \nu \in \mathcal{M}(\partial\mathbf{X})}} & \int 1 \,\mathrm{d}\mu \\
\text{s.t. } & \mu \in \mathcal{M}(\overline{\mathbf{X}})_+ \\
& \nu \in \mathcal{M}(\partial\mathbf{X})_+ \\
& \lambda_{\mathbf{K}} - \mu \in \mathcal{M}(\mathbf{K})_+ \\
& \mathbf{grad}\ \mu = (\mathbf{grad}\ h)\ \nu
\end{aligned} \tag{54a}$$

$$\begin{aligned}
\lambda(\mathbf{X}) = \inf_{\substack{w \in C^0(\mathbf{K}) \\ \mathbf{u} \in C^1(\overline{\mathbf{X}})^r}} & \int w \,\mathrm{d}\lambda_{\mathbf{K}} \\
\text{s.t. } & w|_{\overline{\mathbf{X}}} - \mathrm{div}\ \mathbf{u} - 1 \in \mathcal{C}(\overline{\mathbf{X}})_+ \\
& -(\mathbf{u} \cdot \mathbf{grad}\ h)|_{\partial\mathbf{X}} \in \mathcal{C}(\partial\mathbf{X})_+ \\
& w \in \mathcal{C}(\mathbf{K})_+
\end{aligned} \tag{54b}$$

It has been proved in [49] that the existence of an optimal solution to (54b) can be deduced from the existence of a solution to a Poisson PDE with Neumann boundary condition:

$$\begin{array}{rclcc}
-\Delta u & = & \phi & \text{in} & \mathbf{X} \\
\partial_{\mathbf{n}} u & = & 0 & \text{on} & \partial\mathbf{X} \\
\phi & \leq & 1 & \text{in} & \mathbf{X} \\
\phi & = & 1 & \text{on} & \partial\mathbf{X}
\end{array} \tag{55}$$

Namely, for a pair (u, ϕ) satisfying (55), set

$$\mathbf{u} := \mathbf{grad}\, u \quad \text{and} \quad w(\mathbf{x}) := \begin{cases} 1 - \phi(\mathbf{x}), & \mathbf{x} \in \mathbf{X} \\ 0, & \text{else,} \end{cases} \tag{56}$$

then $(w, \mathbf{u})$ is optimal for (54b).

In [49], ϕ is proposed under the form

$$\phi(\mathbf{x}) = 1 - h(\mathbf{x}) \sum_{i=1}^{N} \frac{\lambda(\mathbf{X}_i)}{\int g \; d\lambda_{\mathbf{X}_i}} \mathbb{1}_{\mathbf{X}_i}(\mathbf{x})$$

where the $\mathbf{X}_i$ are the connected components of $\mathbf{X}$. As a result, ϕ was proved to be only Lipschitz continuous, so that the optimal function $w = 1 - \phi$ was also only Lipschitz continuous. However, another, smooth optimal function can be designed.

Theorem 5.5 (Existence of smooth solutions)**.**

There exist smooth functions $u, \phi \in C^\infty(\overline{\mathbf{X}})$ solutions to (55). Further, $u, \phi \in C^\infty(\overline{\mathbf{X}})$ can be chosen such that $\mathbf{u}, w$ given by (56) are smooth and optimal for (54b), i.e. it holds $\int w \, \mathrm{d}\lambda_{\mathbf{K}} = \lambda(\mathbf{X})$.

Proof. See Appendix C. □

The regularity result in **Theorem 5.5** allows us to incorporate higher order approximation rates via the Jackson-inequality **Theorem 3.10**. Its effect on the convergence rate of the moment-SoS hierarchy for the GMP (54b) is stated in the following theorem. For $\ell \in \mathbb{N}$, let us denote by Vol_ℓ the optimal value in the ℓ-th level of the moment-SoS hierarchy for (54b).

Theorem 5.6 (Rate for Stokes-augmented volume computation)**.**

Under **Condition 10** *it holds, for $\ell \in \mathbb{N}$ large enough and with $\widehat{Ⱡ} := \max\{1, Ⱡ\}$, that*

$$0 \leq \mathrm{Vol}_\ell - \lambda(\mathbf{X}) \in \mathcal{O}\left(\ell^{-\frac{1}{(2.5+s)m\widehat{Ⱡ}}}\right) \quad \text{as } \ell \to \infty \quad \text{for any } s > 0. \tag{57}$$

Proof. Recall that we assume $\mathbf{K} = [-1,1]^m$. By **Theorem 5.5**, let u, ϕ be smooth solutions of (55) such that $\mathbf{u} = \mathbf{grad}\ u$ and $w = (1-\phi)\big|_{\mathbf{X}}$ from (56) are smooth and optimal for (54b). Let $k \in \mathbb{N}$ and $\bar{w}$ and $\bar{u}$ be C^{k+1} respectively C^{k+3} extensions of w respectively u from **Theorem 4.8**, i.e. $\bar{w} \in C^{k+1}(\mathbf{K}), \bar{u} \in C^{k+3}(\mathbf{K})$ with

$$\|\bar{w}\|_{C^{k+1}(\mathbf{K})} \leq c\|w\|_{C^{k+1}(\mathbf{X})}, \quad \|\bar{u}\|_{C^{k+3}(\mathbf{K})} \leq c\|u\|_{C^{k+3}(\mathbf{X})}$$

for some constant $c = c(k, \mathbf{X})$. We denote by $W, U \in \mathbb{R}$ the following constants

$$W := \|\bar{w}\|_\infty^{\mathbf{K}} \leq c\|w\|_{C^{k+1}(\mathbf{X})}, \quad U := \|\bar{u}\|_{C^2(\mathbf{K})} \leq c\|u\|_{C^{k+3}(\mathbf{K})}. \tag{i}$$

In the rest of the proof, we will also use the following constants:

$$a_1 := \|\Delta h\|_\infty^{\overline{\mathbf{X}}}, \quad , a_2 := \inf_{\mathbf{x} \in \partial\mathbf{X}} \|\mathbf{grad}\ h(\mathbf{x})\|^2, \quad a_3 := \|h\|_{C^2(\overline{\mathbf{K}})} \tag{ii}$$

Note that $a_2 > 0$ by **Condition 10**. We define an inward-pointing direction, namely, for $\theta > 0$ it holds

$$(w_\theta, \mathbf{u}_\theta) := (w + 2a_1\theta, \mathbf{u} - \theta\ \mathbf{grad}\ h) \text{ is strictly feasible.}$$

To verify this, note first that by feasibility of w it holds $w \geq 0$ on $\mathbf{K}$ and thus

$$0 < 2a_1\theta \leq w + 2a_1\theta = w_\theta \leq W + 2a_1\theta \quad \text{on } \mathbf{K}. \tag{iii}$$

In particular, this shows feasibility for the last constraint in (54b). For the second constraint in (54b) let $\mathbf{x} \in \partial\mathbf{X}$; we have

$$-(\mathbf{u}_\theta \cdot \mathbf{grad}\ h)(\mathbf{x}) = -(\mathbf{u} \cdot \mathbf{grad}\ h)(\mathbf{x}) + \theta\|\mathbf{grad}\ h(\mathbf{x})\|^2 = \theta\|\mathbf{grad}\ h(\mathbf{x})\|^2 = \theta a_2 > 0 \tag{iv}$$

with the constant a_2 from (ii); i.e. feasibility for the second constraint in (54b). We further have, for $\mathbf{x} \in \mathbf{K}$

$$|(\mathbf{u}_\theta \cdot \mathbf{grad}\ h)(\mathbf{x})| = |(\mathbf{u} - \theta\ \mathbf{grad}\ h) \cdot \mathbf{grad}\ h)(\mathbf{x})| \leq (U + \theta a_3) \cdot a_3. \tag{v}$$

Now, let us verify strict feasibility in the first constraint in (54b). On $\mathbf{X}$ we have

$$w_\theta - \mathrm{div}\ \mathbf{u}_\theta - 1 \quad = \quad w - \mathrm{div}\ \mathbf{u} - 1 + 2\theta a_1 + \theta\Delta h = 2\theta a_1 + \theta\Delta h \geq 2\theta a_1 - \theta a_1 = \theta a_1 \tag{vi}$$

with a_1 from (ii).

Further, on $\mathbf{K}$ we have

$$|w_\theta - \text{div } \mathbf{u}_\theta - 1| = |w + 2a_1\theta - \text{div } (u - \theta\mathbf{grad}\ h) - 1| \leq W + 2\theta a_1 + U + \theta a_3 + 1 \tag{vii}$$

for the constants W, U from (i) and a_1, a_3 from (ii). The cost for $(w_\theta, \mathbf{u}_\theta)$ is simply

$$\int w_\theta \ \mathrm{d}\lambda_{\mathbf{K}} = \int w \ \mathrm{d}\lambda_{\mathbf{K}} + 2a_1\theta\lambda(\mathbf{K}) = \lambda(\mathbf{X}) + 2^{m+1}a_1\theta. \tag{viii}$$

In the next step, we approximate the pair $(w_\theta, \mathbf{u}_\theta)$ by feasible polynomials. In this step we use smoothness of $w_\theta, \mathbf{u}_\theta$ and **Theorem 3.10**. Let $c = c_k \in \mathbb{R}$ be the constant from **Theorem 3.10**. That is, there exist polynomials $p_d, q_d \in \mathbb{R}_d[\mathbf{x}]$ with

$$\|w - p_d\|_\infty^{\mathbf{K}}\ ,\ \|u - q_d\|_{C^2(\mathbf{K})} \leq \frac{c_k}{d^k}. \tag{ix}$$

Recall that $\mathbf{u} = \mathbf{grad}\ u$, i.e. $\mathbf{grad}\ q_d$ is an approximation of $\mathbf{u}$. For $\theta > 0$ and $d \geq \deg(h)$ we define

$$p_{d,\theta} := p_d + 2\theta a_1 \in \mathbb{R}_d[\mathbf{x}], \quad \mathbf{q}_{d,\theta} := \mathbf{grad}\ q_d - \theta\ \mathbf{grad}\ h \in \mathbb{R}_d[\mathbf{x}]^m.$$

We have $\|w_\theta - p_{d,\theta}\|_\infty^{\mathbf{K}} = \|w - p_d\|_\infty^{\mathbf{K}}$ and $\|\mathbf{u}_\theta - \mathbf{q}_{d,\theta}\|_{C^1(\mathbf{K})^m} = \|\mathbf{grad}\ (u - q_d)\|_{C^1(\mathbf{K})^m} \leq \|u - q_d\|_{C^2(\mathbf{K})}$, thus, by (ix), we get

$$\|w_\theta - p_{d,\theta}\|_\infty^{\mathbf{K}}\ ,\ \|\mathbf{u}_\theta - \mathbf{q}_{d,\theta}\|_{C^1(\mathbf{K})} \leq \frac{c_k}{d^k}. \tag{x}$$

On $\mathbf{X}$ we have, for the first constraint in (54b),

$$p_{d,\theta} - \text{div } \mathbf{q}_{d,\theta} - 1 \quad = \quad w_\theta - \text{div } \mathbf{u}_\theta - 1 + p_{d,\theta} - w_\theta + \text{div } (\mathbf{q}_{d,\theta} - \mathbf{u}_\theta),$$

and hence, from (vi) and (x), we get

$$p_{d,\theta} - \text{div } \mathbf{q}_{d,\theta} - 1 \geq a_1\theta - 2\frac{c_k}{d^k}. \tag{xi}$$

Further, on $\mathbf{K}$ we have by (vii) and (x)

$$\begin{aligned} |p_{d,\theta} - \text{div } \mathbf{q}_{d,\theta} - 1| \quad &\leq \quad |w_\theta - \text{div } \mathbf{u}_\theta - 1| + |p_{d,\theta} - w_\theta| + |\text{div } \mathbf{q}_{d,\theta} - \text{div } \mathbf{u}_\theta| \\ &\leq \quad W + 2\theta a_1 + U + \theta a_3 + 1 + 2\frac{c_k}{d^k}. \end{aligned} \tag{xii}$$

For the second constraint in (54b) we have on $\partial\mathbf{X}$, by (iv) and (ix),

$$\begin{aligned} -\mathbf{q}_{d,\theta} \cdot \mathbf{grad}\ h \quad &= \quad -\mathbf{u}_\theta \cdot \mathbf{grad}\ h + (\mathbf{u}_\theta - \mathbf{q}_{d,\theta}) \cdot\ \mathbf{grad}\ h \\ &\geq \quad \theta a_2 + (\mathbf{u}_\theta - \mathbf{q}_{d,\theta}) \cdot\ \mathbf{grad}\ h \geq \theta a_2 - \frac{c_k}{d^k}\sqrt{a_2}. \end{aligned} \tag{xiii}$$

Further, on $\mathbf{K}$ we have, by (v) and (ix),

$$\begin{aligned} |\mathbf{grad}\ q_{d,\theta} \cdot \mathbf{grad}\ h| \quad &\leq \quad |\mathbf{u}_\theta \cdot \mathbf{grad}\ h| + |\mathbf{grad}\ (u_\theta - q_{d,\theta}) \cdot\ \mathbf{grad}\ h| \\ &\leq \quad (U + \theta a_3) \cdot a_3 + \frac{c_k}{d^k}a_3. \end{aligned} \tag{xiv}$$

And for the third constraint in (54b), we have by (iii)

$$2a_1\theta - \frac{c_k}{d^k} \leq w_\theta - \frac{c_k}{d^k} \leq p_{d,\theta} \leq W + 2a_1\theta + \frac{c_k}{d^k}. \tag{xv}$$

Before invoking the effective version of Putinar's Positivstellensatz, **Theorem 3.7**, we make the choice

$$\theta := \theta_d := \frac{c_k}{d^k} \max\left\{\frac{3}{a_1}, \frac{1 + \sqrt{2}}{a_2}\right\} \in \mathcal{O}(d^{-k}).$$

For this choice of θ we have for (xi) and (xiii) on $\mathbf{X}$ that

$$p_{d,\theta_d} - \text{div } \mathbf{q}_{d,\theta_d} - 1,\ -\mathbf{q}_{d,\theta_d} \cdot \mathbf{grad}\ h \geq \frac{c_k}{d^k} > 0. \tag{xvi}$$

On $\mathbf{K}$ it holds
$$p_{d,\theta_d} \geq 6\frac{c_k}{d^k} - \frac{c_k}{d^k} = 5\frac{c_k}{d^k} > 0, \tag{xvii}$$
in particular $(p_{d,\theta}, \mathbf{q}_{d,\theta})$ is feasible for (54b). Further, for the upper bounds (xii),(xiv) and (xv), we have on $\mathbf{K}$ for d large enough (such that $\theta_d \leq 1, {}^{c_k}/_{d^k} \leq 1$)
$$p_{d,\theta},\ p_{d,\theta} - \operatorname{div}\, \mathbf{q}_{d,\theta} - 1,\ -\mathbf{q}_{d,\theta} \cdot \mathbf{grad}\ h \leq K \tag{xviii}$$
for the constant $K := \max\{W + 2a_1 + U + a_3 + 3, (U + a_3 + 1) \cdot a_3\}$. Now, by **Theorem 3.7** and inserting (xvi) and (xviii), the pair $(p_{d,\theta}, q_{d,\theta})$ is feasible for the first two constraints in the ℓ-th level of the moment-SoS hierarchy for (54b) for
$$\ell \geq \gamma(m,h) d^{3.5n\text{Ł}} \left(\frac{K}{{}^{c_k}/_{d^k}}\right)^{2.5m\text{Ł}}.$$
Similarly, by inserting (xvii) and (xviii) into **Theorem 3.7** for $\mathbf{K}$ (note that $\text{Ł} = 1$ in that case), we get that the pair $(p_{d,\theta}, q_{d,\theta})$ is feasible for the third constraint in the ℓ-th level of the moment-SoS hierarchy for (54b) for
$$\ell \geq \gamma(m,\mathbf{f}) d^{3.5m} \left(\frac{K}{{}^{5c_k}/_{d^k}}\right)^{2.5m}$$
Taking the maximum of the just obtained two bounds for ℓ we get that $(p_{d,\theta}, \mathbf{q}_{d,\theta})$ for the optimization problem (54b) for
$$\ell \leq \max\{\gamma(m,h), \gamma(m,\mathbf{f})\} \left(\frac{K}{c_k}\right)^{2.5m\widehat{\text{Ł}}} d^{(3.5+2.5k)m\widehat{\text{Ł}}}. \tag{xix}$$
The cost of $(p_{d,\theta}, \mathbf{q}_{d,\theta})$ for the optimization problem (54b) is bounded by
$$\begin{aligned}\int p_{d,\theta}\ \mathrm{d}\lambda_{\mathbf{K}} &\leq \int w_\theta\ \mathrm{d}\lambda_{\mathbf{K}} + \frac{c_k}{d^k}\lambda(\mathbf{K}) \\ &\overset{\text{(viii)}}{=} \lambda(\mathbf{X}) + 2^m \left(2a_1\theta(d) + \frac{c_k}{d^k}\right) \\ &= \lambda(\mathbf{X}) + \frac{c_k}{d^k} 2^m \max\left\{\frac{3}{a_1}, \frac{1+\sqrt{2}}{a_2}\right\}.\end{aligned}$$
This shows (57); namely, for $\varepsilon > 0$ take the smallest $d \in \mathbb{N}$ with $d \geq \left(2^m \frac{c_k}{\varepsilon} \max\{\frac{3}{a_1}, \frac{1+\sqrt{2}}{a_2}\}\right)^{\frac{1}{k}}$. Then, from (xix), $\mathrm{Vol}_\ell - \lambda(\mathbf{X}) \leq \varepsilon$ for $\ell \in \mathbb{N}$ with
$$\ell \geq \max\{\gamma(m,h), \gamma(m,\mathbf{f})\} \left(\frac{K}{c_k}\right)^{2.5m\widehat{\text{Ł}}} d^{(3.5+2.5k)m\widehat{\text{Ł}}} \in \mathcal{O}\left(\left(\frac{1}{\varepsilon}\right)^{2.5m\widehat{\text{Ł}} + \frac{3.5m\widehat{\text{Ł}}}{k}}\right).$$
In other words, $\mathrm{Vol}_\ell - \lambda(\mathbf{X}) \in \mathcal{O}\left(\ell^{-\frac{1}{2.5m + 3.5m/k}}\right)$. Taking $k \in \mathbb{N}$ arbitrarily large proves the claim. □

Remark 5.7 (Quantifying the efficiency of Stokes constraints).

The convergence rate in **Theorem 5.6** improves the rate in **Theorem 5.2** by more than the power of two. This improvement originates from the smoothness of solutions $(w, \mathbf{u})$ of (54b).

Remark 5.8 (Room for improvement in effective Positivstellensätze).

The proof of **Theorems 5.6, 4.4** and **4.14** show that an effective version of a Positivstellensatz for polynomials p on $\mathbf{X} \subsetneq [-1,1]^m$, taking into account only the maximum of p on $\mathbf{X}$ and not on $[-1,1]^m$, is desirable to obtain stronger rates. The reason why we would obtain better rates is that we obtained an upper bound of the function of interest on $\mathbf{K}$ by first extending it from $\mathbf{X}$ to a function on $[-1,1]^m$ which heavily increased the upper bound. Thus, an effective Positivstellensatz that only takes into account $\|p\|_\infty^{\mathbf{X}}$ (without inducing large degree-dependent bounds) would further improve the convergence rates in the Stokes-augmented case, highlighting that the actual difference between **Theorems 5.2** and **5.6** is in fact much sharper, as **Theorem 5.2** would not benefit from such improvement of the effective Positivstellensatz.

Remark 5.9 (On specialized Positivstellensätze)**.**

In this section we applied **Theorem 3.7** to obtain degree bounds for quadratic module representations over the hypercube $\mathbf{K} = [-1, 1]$. Nevertheless, there exists specialized (and probably tighter) versions of effective Putinar Positivstellensätze on a variety of sets, such as the unit ball [46], the unit sphere [10] and, more recently, the hypercube [4]. However, these effective P-sätze do not come with explicit bounds depending on $\deg(p)$ in addition to $\max_{\mathbf{X}} p$ and $\min_{\mathbf{X}} p$; more precisely, they include constants similar to the $\gamma(m, \mathbf{h})$ displayed in the current work, but that also depend on $\deg(p)$, i.e. under the form $\gamma(m, \deg(p))$. As the volume computing hierarchy involves polynomials p with varying degrees $d \to \infty$, these bounds could not be directly plugged into our analyses, and would require to be specified into more explicit expressions to be useful in all applications of the moment-SoS hierarchy.

6 Conclusion

We state a structured approach to obtaining convergence rates for the moment-SoS hierarchy for the generalized moment problem. For the analysis of the convergence rates, we distinguish three important objects and properties. Namely, the existence and regularity of minimizers, an effective version of Putinar's Positivstellensatz, and a geometric feasibility condition (see the inward-pointing condition in Section 3.4). Our proposed procedure points out how those properties interact and is demonstrated to obtain upper bounds on the convergence rate for certain instances of the moment-SoS hierarchy: Using recent improvements on an effective version of Putinar's Positivstellensatz, we build up on and strongly improve existing convergence rates for the optimal control and the volume computation of a semialgebraic set; and we give an original convergence rate for a moment-SoS hierarchy of exit location computation for stochastic differential equations. We hope our work provides a guideline and the necessary tools for computing convergence rates of the moment-SoS hierarchy for various generalized moment problems that are actively formulated in the field in recent and following years.

Future work and improvement of effective Positivstellensätze can be integrated within our work simply by applying the most suited available convergence rate for Putinar's Positivstellensatz. Furthermore, we observe in our analysis that a well-suited effective Positivstellensatz could strongly further improve the convergence rate. As mentioned in **Remark 5.8**, particularly advantageous for our method would be an effective Positivstellensatz – for a polynomial p on a semialgebraic set $\mathbf{X}$ – that only takes into account the values of p on $\mathbf{X}$ without the need of bounding its value on an ambient set (such as the hypercube in **Theorem 3.7**). Similarly, specialized Positivstellensätze could be improved by expliciting all the terms in their degree bounds. Considering the recent improvement and active work on degree bounds for Positivstellensätze, we see here a very interesting, exciting, and promising development for further improvements of existing convergence rates for the moment-SoS hierarchy for generalized moment problems, as well as quantitative analysis of many other moment-SoS-based methodologies that will appear in the future.

We think it is important to mention that the asymptotic analysis of the moment-SoS hierarchy for generalized moment problems might not transfer to practical applications. The reason is twofold. Firstly, current computational capacities restrict the computation of the moment-SoS hierarchy already for medium-sized problems to low-degree instances. Secondly, the conditioning of the ℓ-th level of the moment-SoS hierarchy gets worse with increasing $\ell \in \mathbb{N}$, hampering the convergence in practice. In other words, this work essentially addressed the recasting from the infinite dimensional GMP into SoS programming problems, while future works will shift the focus onto the translation from SoS programming to actual SDP, involving deeper investigations on what polynomial basis to choose in that process (the usual one being the numerically ill-behaved basis of monomials).

A On norm equivalence in polynomial spaces

Lemma A.1. *Let* $\mathbf{X} \subset \mathbf{K} := [-1,1]^m$ *satisfy* **Condition 6.2**. *For any nonnegative polynomial* $p \in \mathcal{P}(\mathbf{X})_+$, *it holds*

$$\|p\| = \max_{\mathbf{x}\in\mathbf{K}}\{p(\mathbf{x})\} \leq \left(1 + \frac{\deg(p)^2}{4}\,({}^2\!/\!{}_b)^{\deg(p)+1}\right)\|p\|_\infty^{\mathbf{X}}, \tag{30}$$

where $b \in (0,1)$ *is such that* $[-b,b]^m \subset \mathbf{X}$ *(whose existence is guaranteed by* **Condition 6.2***).*

Proof. The proof of [4, Lemma 28] actually shows that, for $\varphi \in \mathcal{P}(\mathbf{K})_+$ of degree k and $\rho > 0$, defining $\mathbf{K}_\rho := [-1-\rho, 1+\rho]^m$, one has

$$\varphi^\star_{\mathbf{K}_\rho} \geq \varphi^\star_{\mathbf{K}} - T_k(1+\rho)\cdot\rho\cdot k^2\cdot\max_{\mathbf{K}}\varphi \tag{†}$$

Where T_k denotes the degree k Chebyshov polynomial of the first kind. We apply this result to

$$\varphi(\mathbf{x}) := \max_{\mathbf{X}} p - p(b\,\mathbf{x}),$$

so that $k = \deg(\varphi) = \deg(p)$. Let $\rho := \frac{1-b}{b}$ so that $b\cdot\mathbf{K}_\rho = \mathbf{K}$. Then, one has

$$\begin{aligned}
\varphi^\star_{\mathbf{K}_\rho} &= \max_{\mathbf{X}} p - \max_{\mathbf{K}} p\\
\varphi^\star_{\mathbf{K}} &= \max_{\mathbf{X}} p - \max_{b\cdot\mathbf{K}} p\\
\max_{\mathbf{K}}\varphi &= \max_{\mathbf{X}} p - \min_{b\cdot\mathbf{K}} p
\end{aligned}$$

which can be reinjected in (†) to get

$$\cancel{\max_{\mathbf{X}} p} - \max_{\mathbf{K}} p \geq \cancel{\max_{\mathbf{X}} p} - \max_{b\cdot\mathbf{K}} p - T_k(\underbrace{1+\rho}_{1/b})\cdot\rho\cdot k^2\left(\max_{\mathbf{X}} p - \min_{b\cdot\mathbf{K}} p\right).$$

This expression in turn rephrases, accounting for inequalities $\min_{b\cdot\mathbf{K}} p \geq \min_{\mathbf{X}} p \geq 0$ and $\max_{b\cdot\mathbf{K}} p \leq \max_{\mathbf{X}} p = \|p\|_\infty^{\mathbf{X}}$ (because $b\cdot\mathbf{K} \subset \mathbf{X}$), as

$$\max_{\mathbf{K}} p \leq \max_{b\cdot\mathbf{K}} p + T_k({}^1\!/\!{}_b)\cdot\frac{1-b}{b}\cdot k^2\cdot\max_{\mathbf{X}} p \leq \left(1 + T_k({}^1\!/\!{}_b)\cdot\frac{1-b}{b}\cdot k^2\right)\|p\|_\infty^{\mathbf{X}}.$$

It remains to compute an upper bound of $T_k({}^1\!/\!{}_b)$. Here we recall that the Chebyshov polynomials are defined by $T_0(s) = 1$, $T_1(s) = s$ and the recurrence formula $T_{k+1}(s) = 2s\,T_k(s) - T_{k-1}(s)$ (for $k \geq 1$). Moreover, $T_k(s)$ oscillates between -1 and 1 for $s \in [-1,1]$ (because $T_k(\cos\theta) = \cos(k\cdot\theta)$), and is always strictly increasing on $(1,+\infty)$, so that for $s > 1$ one has $T_k(s) > T_k(1) = 1 > 0$. Eventually, we can prove that, for all $k \geq 1$, it holds

$$T_k({}^1\!/\!{}_b) \leq \frac{1}{2}\left(\frac{2}{b}\right)^k. \tag{‡}$$

First, we check that this holds for $k = 1$: $T_1({}^1\!/\!{}_b) = {}^1\!/\!{}_b \leq {}^1\!/\!{}_2\cdot{}^2\!/\!{}_b$. Second, defining $s = {}^1\!/\!{}_b$, we notice that by design of $b \in (0,1)$, $s > 1$. Then we simply use the recurrence formula to get, for $k \geq 1$:

$$T_{k+1}(s) = 2s\,T_k(s) - \underbrace{T_{k-1}(s)}_{>0} \leq 2s\,T_k(s)$$

so that, if $T_k(s) \leq {}^1\!/\!{}_2(2s)^k$ (which holds for $k = 1$) then $T_{k+1}(s) \leq {}^1\!/\!{}_2(2s)^{k+1}$. This way, we get

$$\max_{\mathbf{K}} p \leq \left(1 + \frac{1}{2}\left(\frac{2}{b}\right)^k\cdot\frac{1-b}{b}\cdot k^2\right)\|p\|_\infty^{\mathbf{X}}.$$

Finally, let us not forget that it is $\max_{\mathbf{K}} |p|$ that we want to upper bound, and not only $\max_{\mathbf{K}} p$, so we still have to upper bound $\max_{\mathbf{K}}(-p) = -\min_{\mathbf{K}} p = -p^\star_{\mathbf{K}}$. Again we use (†) with $\varphi(\mathbf{x}) = p(b\mathbf{x})$ to get

$$\begin{aligned}
p^\star_{\mathbf{K}} &= \varphi^\star_{\mathbf{K}_\rho} \\
&\geq \varphi^\star_{\mathbf{K}} - T_k(1+\rho)\cdot\rho\cdot k^2\cdot\max_{\mathbf{K}}\varphi \\
&\geq \min_{b\cdot\mathbf{K}} p - \frac{1}{2}\left(\frac{2}{b}\right)^k\cdot\frac{1-b}{b}\cdot k^2\cdot\max_{b\cdot\mathbf{K}} p \\
&\geq \min_{\mathbf{X}} p - \frac{1}{2}\left(\frac{2}{b}\right)^k\cdot\frac{1-b}{b}\cdot k^2\cdot\max_{\mathbf{X}} p \\
&\geq -\frac{1}{2}\left(\frac{2}{b}\right)^k\cdot\frac{1-b}{b}\cdot k^2\cdot\|p\|_\infty^{\mathbf{X}}
\end{aligned}$$

and hence

$$\max_{\mathbf{K}}(-p) = -p^\star_{\mathbf{K}} \leq \frac{1}{2}\left(\frac{2}{b}\right)^k\cdot\frac{1-b}{b}\cdot k^2\cdot\|p\|_\infty^{\mathbf{X}}$$

leading to

$$\|p\| \leq \left(1+\frac{1}{2}\left(\frac{2}{b}\right)^{\deg(p)}\cdot\frac{1-b}{b}\cdot\deg(p)^2\right)\|p\|_\infty^{\mathbf{X}}.$$

Finally we deduce the announced inequality by observing that $(1-b)\in(0,1)$, so that

$$\frac{1}{2}\left(\frac{2}{b}\right)^{\deg(p)}\cdot\frac{1-b}{b} \leq \frac{1}{4}\left(\frac{2}{b}\right)^{\deg(p)+1}$$

□

B Extension of Hölder continuous functions

Lemma B.1 (Extension Lemma; [12, Lemma 6.37])**.**

Let $k \geq 1$ be an integer and $a \in (0,1]$. Let $\mathbf{Y} \subset \mathbb{R}^n$ be a compact with $C^{k,a}$ boundary. Let $\mathbf{\Omega}$ be an open and bounded set containing $\mathbf{Y}$. Then every function $w \in C^{k,a}(\mathbf{Y})$ there exists an extension $\bar{w} \in C^{k,a}(\mathbf{\Omega})$ with $w(\mathbf{y}) = \bar{w}(\mathbf{y})$ for all $\mathbf{y} \in \mathbf{Y}$ and

$$\|\bar{w}\|_{C^{k,a}(\mathbf{\Omega})} \leq c\|w\|_{C^{k,a}(\mathbf{Y})} \tag{33}$$

for some constant $c = c(n,k,a,\mathbf{Y},\mathbf{\Omega})$ independent of w.

As a corollary, we obtain the following extension result that aims at preserving the maximum value for the extension.

Corollary B.2. *Let $\mathbf{Y} \subset \mathbb{R}^m$ be a compact with C^∞ boundary and $f \in C^\infty(\mathbf{Y})$. Then, for any $k \in \mathbb{N}$ and $\varepsilon > 0$, there exists an extension $\bar{f} \in C^k(\mathbb{R}^n)$ of f with*

$$\|\bar{f}\|_\infty^{\mathbb{R}^n} \leq \|f\|_\infty^{\mathbf{Y}} + \varepsilon. \tag{$\diamond$}$$

Proof. Let $\varepsilon > 0$ and $k \in \mathbb{N}$ and let $R > 0$ with $B_{0.5R}(0) \supset \mathbf{Y}$. By the extension Theorem 4.8 there exists an extension $\tilde{f} \in C^k(B_R(0))$ of f with $F := \|\tilde{f}\|_{C^1(B_R(0))} < \infty$. Without loss of generality we assume $\frac{\varepsilon}{F} < \frac{R}{2}$, otherwise we take a smaller ε. Set $\mathbf{\Omega} := \{\mathbf{x}+\mathbf{y} : \mathbf{x}\in\mathbf{Y}, \|\mathbf{y}\| < \frac{\varepsilon}{F}\}$ and let $\phi \in C^\infty(\mathbb{R}^n)$ with $0 \leq \phi \leq 1$ with $\phi = 1$ on $\mathbf{Y}$ and $\phi = 0$ on $\mathbb{R}^n \setminus \mathbf{\Omega}$. We claim that the function $\bar{f} := \phi\cdot\tilde{f}$ (and extended by zero on $\mathbb{R}^n\setminus\mathbf{\Omega}$) is C^k and satisfies ($\diamond$). From the construction, it follows that $\tilde{f}$ is C^k. For $\mathbf{x}\in\mathbf{Y}$ it holds $\tilde{f}(\mathbf{x}) = f(\mathbf{x})$ and for $\mathbf{x}\in\mathbb{R}^n\setminus\mathbf{\Omega}$ we have $\tilde{f}(\mathbf{x}) = 0$. It remains to bound $\tilde{f}(\mathbf{x})$ for $\mathbf{x}\in\mathbf{\Omega}\setminus\mathbf{Y}$. Let $\mathbf{x}\in\mathbf{\Omega}\setminus\mathbf{Y}$ and $\mathbf{y}$ with $\|y\| < \frac{\varepsilon}{F}$ such that $\mathbf{x}-\mathbf{y}\in\mathbf{Y}$. We have

$$|\tilde{f}(\mathbf{x})| \leq \left|\tilde{f}(\mathbf{x}-\mathbf{y}) + \|\mathbf{grad}\tilde{f}\|_\infty^{B_R(0)}\|y\|\right| \leq \|f\|_\infty^{\mathbf{Y}} + F\frac{\varepsilon}{F} = \|f\|_\infty^{\mathbf{Y}} + \varepsilon.$$

□

C Smooth solutions to the Poisson PDE

Here we state the proof of **Theorem 5.5** for existence of smooth solutions of (54b) and (55).

Proof. Of ***Theorem 5.5***. We prove this Theorem using two lemmas.

Lemma C.1 (Existence of smooth source term)**.**

There exists a smooth $\phi \in C^\infty(\mathbb{R}^n)$ *such that*

1. ϕ *satisfies conditions* (55.c), (55.d) *and* $\phi = 1$ *on* $\mathbb{R}^n \setminus \mathbf{X}$.
2. $\displaystyle\int_{\mathbf{X}_i} \phi \, d\lambda = 0$ *for all* $i \in \{1, \ldots, \Omega\}$, *where*

$$\mathbf{X} = \bigsqcup_{i=1}^{\Omega} \mathbf{X}_i$$

is the partition of $\mathbf{X}$ *into its connected components.*

Proof. We work on a connected component $\mathbf{X}_i$, $i \in \{1, \ldots, \Omega\}$. As $\mathbf{X}_i$ is an open set, there exists $\boldsymbol{\omega}_i \in \mathbf{X}_i$, $R_i > 0$ such that

$$\mathbf{B}_i := \{\mathbf{x} \in \mathbb{R}^n \, ; |\mathbf{x} - \boldsymbol{\omega}_i| \leq R_i\} \subset \mathbf{X}_i.$$

According to [33, **Proposition 2.25**], there exists a smooth bump function $\varphi_i \in C^\infty(\mathbb{R}^n)$ such that:

- $\forall \mathbf{x} \in \mathbb{R}^n \setminus \mathbf{X}_i$, $\varphi_i(\mathbf{x}) = 0$
- $\forall \mathbf{x} \in \mathbf{B}_i$, $\varphi_i(\mathbf{x}) = 1$
- $\forall \mathbf{x} \in \mathbf{X}_i \setminus \mathbf{B}_i$, $0 \leq \varphi_i(\mathbf{x}) \leq 1$.

In particular, $\varphi_i \geq 0$ in $\mathbf{X}_i$ and $\varphi_i = 0$ on $\partial \mathbf{X}_i$. Next, we define for $\mathbf{x} \in \mathbb{R}^n$,

$$\psi_i(\mathbf{x}) := \frac{\lambda(\mathbf{X}_i)}{\displaystyle\int_{\mathbf{X}_i} \varphi_i \, d\lambda} \varphi_i(\mathbf{x}),$$

with $0 < \lambda(\mathbf{B}_i) < \displaystyle\int_{\mathbf{X}_i} \varphi_i \, d\lambda$ by design of φ_i.

Again, $\psi_i \geq 0$ in $\mathbf{X}_i$ and $\psi_i = 0$ on $\partial \mathbf{X}_i$. Moreover, now

$$\int_{\mathbf{X}_i} \psi_i \, d\lambda = \lambda(\mathbf{X}_i). \tag{$*$}$$

Eventually, we construct, for $\mathbf{x} \in \mathbb{R}^n$,

$$\phi(\mathbf{x}) := 1 - \sum_{i=1}^{\Omega} \psi_i,$$

so that condition *1.* is trivially satisfied, and smoothness of ϕ follows from smoothness of the φ_is. We conclude by checking condition *2.*: for $i \in \{1, \ldots, \Omega\}$,

$$\int_{\mathbf{X}_i} \phi \, d\lambda = \int_{\mathbf{X}_i} \left(1 - \sum_{j=1}^{\Omega} \psi_j\right) d\lambda = \int_{\mathbf{X}_i} 1 \, d\lambda - \sum_{j=1}^{\Omega} \int_{\mathbf{X}_i} \psi_j \, d\lambda$$

$$= \lambda(\mathbf{X}_i) - \int_{\mathbf{X}_i} \psi_i \, d\lambda - \sum_{\substack{j=1 \\ j \neq i}}^{\Omega} \cancel{\int_{\mathbf{X}_i} \psi_j \, d\lambda} \overset{(*)}{=} 0$$

$\square$

Lemma C.2 (Existence of smooth PDE solution)**.**

Let $\phi \in C^\infty(\overline{\mathbf{X}})$ be given by **Lemma C.1**. *Then, there exists a solution $u \in C^\infty(\overline{\mathbf{X}})$ to the Poisson PDE with Neumann boundary condition:*

$$\begin{cases} -\Delta u &= \phi \quad \text{in} \quad \mathbf{X} \\ \partial_{\mathbf{n}} u &= 0 \quad \text{on} \quad \partial\mathbf{X} \end{cases} \tag{55}$$

Proof. If $\Omega = 1$ (*i.e.* $\mathbf{X}$ is connected), then this is a classical result, see e.g. [12][4]. Else, we just solve the problem separately on each connected component and glue the resulting solutions u_i together into

$$u = \sum_{i=1}^{\Omega} u_i \, \mathbb{1}_{\mathbf{X}_i} \in C^\infty(\overline{\mathbf{X}})$$

because by construction of $\mathbf{X}$ (with smooth boundary) the $\overline{\mathbf{X}}_i$ are disjoint. □

The ϕ and u given by **Lemmas C.1** and **C.2** are a valid solution to (55). By construction, they also have the required smoothness. It remains to show that we can choose ϕ, u such that the functions $\mathbf{u}, w$ given by (56) are optimal for (54b). Let us take ϕ as in Lemma C.1 and u the corresponding solution of (55). By the above, the functions $\mathbf{u}, w$ from (56) are smooth. Further, as shown in [49], $\mathbf{u}, w$ are feasible (and optimal) for (54b). Here, we only recall optimality, i.e.

$$\int w \, \mathrm{d}\lambda = \lambda(\mathbf{X}). \tag{$\triangledown$}$$

To show ($\triangledown$), we use condition 2. in Lemma C.1 and simply integrate $w = 1 - \phi \geq 0$ on $\mathbf{X}$. This gives

$$\int w \, \mathrm{d}\lambda_{\mathbf{K}} = \int 1 - \phi \, \mathrm{d}\lambda_{\mathbf{K}} = \lambda(\mathbf{X}) - \int \phi \, \mathrm{d}\lambda_{\mathbf{K}} = \lambda(\mathbf{X}).$$

□

References

[1] Thomas Bagby, Len Bos, and Norman Levenberg. Multivariate simultaneous approximation. Constructive approximation, 18(4):569–577, 2002.

[2] Lorenzo Baldi and Bernard Mourrain. On the effective Putinar's Positivstellensatz and moment approximation. Mathematical Programming, 200(1):71–103, 2023.

[3] Lorenzo Baldi, Bernard Mourrain, and Adam Parusiński. On Łojasiewicz inequalities and the effective Putinar's Positivstellensatz. arXiv: 2212.09551 [math.AC], 2022.

[4] Lorenzo Baldi and Lucas Slot. Degree bounds for Putinar's Positivstellensatz on the hypercube. SIAM Journal on Applied Algebra and Geometry, 8(1):1–25, 2024.

[5] Alexander Barvinok. A Course in Convexity, volume 54 of Graduate Studies in Mathematics. American Mathematical Society, 2002.

[6] Eman Samir Bhaya. Interpolating operators for multiapproximation. Journal of Mathematics and Statistics, 6(3):240–245, 2010.

[7] John M. Davis and Paul Hagelstein. Gibbs phenomena for some classical orthogonal polynomials. Journal of Mathematical Analysis and Applications, 505(1):125574, 2022.

[8] Yohann De Castro, Fabrice Gamboa, Didier Henrion, and Jean B. Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. IEEE Transactions on Information Theory, 63:621–630, 2017.

[4] Smoothness of $\partial\mathbf{X}$ is necessary here. This is the reason for **Assumption 10**.

[9] Lawrence C Evans. An introduction to stochastic differential equations, volume 82. American Mathematical Soc., 2012.

[10] Kun Fang and Hamza Fawzi. The Sum-of-Squares hierarchy on the sphere and applications in quantum information theory. Mathematical Programming, 190:331–360, 2021.

[11] Giovanni Fantuzzi, David Goluskin, Deqing Huang, and Sergei I Chernyshenko. Bounds for deterministic and stochastic dynamical systems using sum-of-squares optimization. SIAM Journal on Applied Dynamical Systems, 15(4):1962–1988, 2016.

[12] David Gilbarg, Neil S Trudinger, David Gilbarg, and NS Trudinger. Elliptic partial differential equations of second order, volume 224. Springer, 1977.

[13] David Goluskin. Bounding extrema over global attractors using polynomial optimisation. Nonlinearity, 33(9):4878, 2020.

[14] Didier Henrion, Mauricio Junca, and Mauricio Velasco. Moment-SOS hierarchy and exit location of stochastic processes. HAL preprint: hal-03110452, 2023.

[15] Didier Henrion, Milan Korda, Martin Kružík, and Rodolfo Rios-Zertuche. Occupation measure relaxations in variational problems: the role of convexity. arXiv preprint arXiv:2303.02434, 2023.

[16] Didier Henrion, Jean B. Lasserre, and Carlo Savorgnan. Approximate volume and integration for basic semialgebraic sets. SIAM Review, 51:722–743, 2009.

[17] Morgan Jones and Matthew M Peet. Converse Lyapunov functions and converging inner approximations to maximal regions of attraction of nonlinear systems. In 2021 60th IEEE Conference on Decision and Control (CDC), pages 5312–5319. IEEE, 2021.

[18] Milan Korda. Computing controlled invariant sets from data using convex optimization. SIAM Journal on Control and Optimization, 58(5):2871–2899, 2020.

[19] Milan Korda and Didier Henrion. Convergence rates of moment-sum-of-squares hierarchies for volume approximation of semialgebraic sets. Optimization Letters, 12(3):435–442, 2018.

[20] Milan Korda, Didier Henrion, and Colin N Jones. Inner approximations of the region of attraction for polynomial dynamical systems. IFAC Proceedings Volumes, 46(23):534–539, 2013.

[21] Milan Korda, Didier Henrion, and Colin 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.

[22] Milan Korda, Didier Henrion, and Colin N Jones. Convergence rates of moment-sum-of-squares hierarchies for optimal control problems. Systems & Control Letters, 100:1–5, 2017.

[23] Milan Korda, Didier Henrion, and Igor Mezić. Convex computation of extremal invariant measures of nonlinear dynamical systems and Markov processes. Journal of Nonlinear Science, 31(1):1–26, 2021.

[24] Milan Korda, Jean-Bernard Lasserre, Alexey Lazarev, Victor Magron, and Simone Naldi. Urysohn in action: separating semialgebraic sets by polynomials. arXiv preprint arXiv:2207.00570, 2022.

[25] Milan Korda, Victor Magron, and Rodolfo Rios-Zertuche. Convergence rates for sums-of-squares hierarchies with correlative sparsity. arXiv preprint arXiv:2303.14824, 2023.

[26] Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In Integer Programming and Combinatorial Optimization conference, pages 293–303, 2001.

[27] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization, 11:796–817, 2001.

[28] Jean B. Lasserre. Moments, positive polynomials and their applications, volume 1. World Scientific, 2009.

[29] Jean B. Lasserre. An introduction to polynomial and semi-algebraic optimization, volume 52. Cambridge University Press, 2015.

[30] Jean B Lasserre. Tractable approximations of sets defined with quantifiers. Mathematical Programming, 151:507–527, 2015.

[31] Jean B. Lasserre. Computing Gaussian and exponential measures of semi-algebraic sets. Advances in Applied Mathematics, 91:137–163, 2017.

[32] Jean B. Lasserre, Didier Henrion, Christophe Prieur, and Emmanuel Trélat. Nonlinear optimal control via occupation measures and LMI-relaxations. SIAM journal on control and optimization, 47(4):1643–1666, 2008.

[33] John M. Lee. Introduction to smooth manifolds. Graduate Texts in Mathematics. Springer, second edition, 2013.

[34] Victor Magron, Didier Henrion, and Jean-Bernard Lasserre. Semidefinite approximations of projections and polynomial images of semialgebraic sets. SIAM Journal on Optimization, 25(4):2143–2164, 2015.

[35] Swann Marx, Tillmann Weisser, Didier Henrion, and Jean B. Lasserre. A moment approach for entropy solutions to nonlinear hyperbolic PDEs. Mathematical Control and Related Fields, 10(1):113–140, 2020.

[36] Jared Miller, Matteo Tacchi, Didier Henrion, and Mario Sznaier. Unsafe probabilities and risk contours for stochastic processes using convex optimization. arXiv preprint arXiv:2401.00815, 2023.

[37] Jared Miller, Matteo Tacchi, Mario Znaier, and Ashkan Jasour. Peak Value-at-Risk estomation for stochastic differential equations using occupation measures. In IEEE Conference on Decision and Control, 2023.

[38] Jiawang Nie. Optimality conditions and finite convergence in Lasserre's hierarchy. Mathematical Programming A, 146:97–121, 2014.

[39] Jiawang Nie and Markus Schweighofer. On the complexity of Putinar's Positivstellensatz. Journal of Complexity, 23(1):135–150, 2007.

[40] Antoine Oustry, Matteo Tacchi, and Didier Henrion. Inner approximations of the maximal positively invariant set for polynomial dynamical systems. IEEE Control Systems Letters, 3(3):733–738, 2019.

[41] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal, 42(3):969–984, 1993.

[42] Cordian Riener, Thorsten Theobald, Lina Jansson Andrén, and Jean B Lasserre. Exploiting symmetries in SDP-relaxations for polynomial optimization. Mathematics of Operations Research, 38(1):122–141, 2013.

[43] Corbinian Schlosser and Milan Korda. Sparse moment-sum-of-squares relaxations for nonlinear dynamical systems with guaranteed convergence. arXiv preprint arXiv:2012.05572, 2020.

[44] Corbinian Schlosser and Milan Korda. Converging outer approximations to global attractors using semidefinite programming. Automatica, 134:109900, 2021.

[45] Morton L. Slater. Lagrange multipliers revisited. Technical report, Cowles Commission, 1950. Discussion Paper No. 403.

[46] Lucas Slot. Sum-of-Squares hierarchies for polynomial optimization and the Christoffel-Darboux kernel. arXiv preprint arXiv: 2111.04610, 2021.

[47] Matteo Tacchi. Moment-SoS hierarchy for large scale set approximation. Application to power systems transient stability analysis. PhD thesis, Toulouse University, 2021.

[48] Matteo Tacchi. Convergence of Lasserre's hierarchy: the general case. Optimization Letters, 16:1015–1033, 2022.

[49] Matteo Tacchi, Jean B. Lasserre, and Didier Henrion. Stokes, Gibbs and volume computation of semi-algebraic sets. Discrete and Computational Geometry, 69:260–283, 2023.

[50] Matteo Tacchi, Tillmann Weisser, Jean Bernard Lasserre, and Didier Henrion. Exploiting sparsity for semi-algebraic set volume computation. Foundations of Computational Mathematics, pages 1–49, 2022.

[51] John Urbas. Lecture on second order linear partial differential equations. In Instructional Workshop on Analysis and Geometry, Part 1, volume 34, pages 39–76. Australian National University, Mathematical Sciences Institute, 1996.