Acta Numerica (2024), pp. 841–908 Printed in the United Kingdom
doi:10.1017/S0962492923000053

The Moment-SOS hierarchy: Applications and related topics

Jean B. Lasserre
LAAS-CNRS and Toulouse School of Economics (TSE), University of Toulouse,
7 Avenue du Colonel Roche, BP54200, 31031 Toulouse cédex 4, France
E-mail: lasserre@laas.fr

The Moment-SOS hierarchy, first introduced in optimization in 2000, is based on the theory of the S-moment problem and its dual counterpart: polynomials that are positive on S. It turns out that this methodology can also be used to solve problems with positivity constraints '$f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in S$' or linear constraints on Borel measures. Such problems can be viewed as specific instances of the *generalized moment problem* (GMP), whose list of important applications in various domains of science and engineering is almost endless. We describe this methodology in optimization and also in two other applications for illustration. Finally we also introduce the Christoffel function and reveal its links with the Moment-SOS hierarchy and positive polynomials.

2020 Mathematics Subject Classification: Primary 14Q30, 44A60, 90C22, 90C23
Secondary 42C05, 47A57, 49L25

CONTENTS

1. Introduction

The Moment-SOS hierarchy was initially designed to help solve *polynomial* optimization problems (POPs), that is, optimization problems of the form

$$\mathbf{P}\colon\ f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x})\colon \mathbf{x} \in S \}, \tag{1.1}$$

where f is a polynomial and $S \subset \mathbb{R}^d$ is a basic semi-algebraic set, that is,

$$S := \{ \mathbf{x} \in \mathbb{R}^d\colon g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m \}, \tag{1.2}$$

for some polynomials g_j, $j = 1, \ldots, m$. Crucially, the description of $\mathbf{P}$ is entirely *algebraic* via its polynomial data f, g_j, $j = 1, \ldots, m$. (However, semi-algebraic functions can also be tolerated to a certain extent.)

As $\mathbf{P}$ is a particular case of *non-linear programming* (NLP), what is so specific about $\mathbf{P}$ in (1.1)? The answer depends on the meaning of f^* in (1.1). If we are only interested in a *local* minimum, then the whole arsenal of efficient methods of NLP can be used to solve $\mathbf{P}$ and its algebraic features are not really exploited.

On the other hand, if f^* in (1.1) is understood as the *global* minimum of $\mathbf{P}$, then the situation is totally different. Why? First, to eliminate any ambiguity in the meaning of f^*, rewrite (1.1) as

$$\mathbf{P}\colon\ f^* = \sup\{ \lambda\colon\ f(\mathbf{x}) - \lambda \geq 0,\ \forall\, \mathbf{x} \in S \}, \tag{1.3}$$

because then indeed f^* is necessarily the *global minimum* of $\mathbf{P}$. In full generality $\mathbf{P}$ is very difficult to solve as it is NP-hard. The reason is:

Given $\lambda \in \mathbb{R}$, checking whether '$f(\mathbf{x}) - \lambda \geq 0$ for all $\mathbf{x} \in S$' is difficult.

Indeed, by its very nature this positivity constraint is *global* and therefore cannot be handled by standard NLP optimization algorithms, which use only local information around a current iterate $\mathbf{x} \in S$. Therefore, to compute f^* in (1.3) it is necessary to handle the positivity constraint '$f(\mathbf{x}) - \lambda \geq 0$ for all $\mathbf{x} \in S$' in some efficient manner. Fortunately, if the data are algebraic, then the following hold.

(1) Powerful *positivity certificates* from real algebraic geometry (*Positivstellensätze* in German) are available.

(2) Some of these positivity certificates have an efficient practical implementation via *linear programming* (LP) or *semidefinite programming* (SDP). In particular, and crucially, testing whether a given polynomial is a sum of squares (SOS) simply reduces to solving a single semidefinite program (which can be done in time polynomial in the input size of the polynomial, up to arbitrary fixed precision[1]).

After the pioneering works of Shor (1987) and Nesterov (2000), Lasserre (2000, 2001) and Parrilo (2000, 2003) were the first to provide a systematic use of these

[1] In fact see O'Donnell (2017) for an update and more details.

two key ingredients in optimization and control, with convergence guarantees. It is also worth mentioning another closely related pioneering work, namely the celebrated SDP relaxation of Goemans and Williamson (1995), which provides a 0.878 approximation guarantee for Max-Cut, a famous problem in non-convex combinatorial optimization (and probably the simplest one). In fact it is perhaps the first famous example of such a successful application of the powerful SDP convex optimization technique to provide *guaranteed* good approximations to a notoriously difficult non-convex optimization problem. It turns out that this SDP relaxation is precisely the first semidefinite relaxation in the Moment-SOS hierarchy (a.k.a. the Lasserre hierarchy) when applied to the Max-Cut problem. The spectacular success story of SDP relaxations has been at the origin of a flourishing research activity in combinatorial optimization and computational complexity. In particular, the study of LP and SDP relaxation techniques in hardness of approximation is at the core of a central topic in combinatorial optimization and computational complexity, namely proving/disproving Khot's famous Unique Games Conjecture[2] (UGC) in theoretical computer science (TCS).

Another (and equivalent) 'definition' of the global optimum f^* of $\mathbf{P}$ reads

$$f^* = \inf_{\phi \in \mathscr{M}(S)_+} \left\{ \int_S f \, \mathrm{d}\phi : \phi(S) = 1 \right\}, \tag{1.4}$$

where the 'inf' is over the set $\mathscr{M}(S)_+$ of (positive) Borel measures ϕ on S. Indeed, as $f \geq f^*$ on S and ϕ is a probability measure on S,

$$\int_S f \, \mathrm{d}\phi \geq \int_S f^* \, \mathrm{d}\phi = f^*,$$

so the infimum in (1.4) is not smaller than f^*. On the other hand, for an arbitrary $\mathbf{x} \in S$, its value $f(\mathbf{x})$ is also obtained via $\int_S f \, \mathrm{d}\phi$, where ϕ is the Dirac probability measure $\delta_{\{\mathbf{x}\}}$ at $\mathbf{x} \in S$, and hence the infimum in (1.4) is not larger than f^*. In particular, if $\mathbf{x}^* \in S$ is a global minimizer of $\mathbf{P}$, then the Dirac measure $\phi^* := \delta_{\{\mathbf{x}^*\}}$ at $\mathbf{x}^*$ is an optimal solution of (1.4).

In fact (1.3) is the LP dual of (1.4), where λ is the dual variable associated with the constraint $\phi(S) = 1$ in (1.4). In other words, standard LP duality between the two conic programs (1.3) and (1.4) nicely captures a convex duality between the 'S-moment problem' in real and functional analysis, and 'polynomials positive on S' in real algebraic geometry (more details are given later).

Moreover, problem (1.4) is a very particular instance (and even the simplest instance) of the more general *generalized moment problem* (GMP) defined by

$$\inf_{\phi_j \in \mathscr{M}(S_j)_+} \left\{ \sum_{j=1}^{p} \int_{S_j} f_j \, \mathrm{d}\phi_j : \sum_{j=1}^{p} f_{ij} \, \mathrm{d}\phi_j \geq b_i, \ i = 1, \ldots, s \right\}, \tag{1.5}$$

[2] For this conjecture and its theoretical and practical implications, Subhash Khot was awarded the prestigious Nevanlinna prize at ICM 2014 in Seoul (Khot 2010, 2014).

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

for some given functions f_j, $f_{ij}\colon \mathbb{R}^{d_j} \to \mathbb{R}$, $i = 1, \dots, s$, and sets $S_j \subset \mathbb{R}^{d_j}$, $j = 1, \dots, p$. The GMP (1.5) is an infinite-dimensional LP with dual

$$\sup_{\lambda_1,\dots,\lambda_s \geq 0} \left\{ \sum_{i=1}^{s} \lambda_i\, b_i \colon f_j - \sum_{i=1}^{s} \lambda_i\, f_{ij} \geq 0 \text{ on } S_j,\ j = 1, \dots, p \right\}. \tag{1.6}$$

Therefore it should be no surprise that the Moment-SOS hierarchy, initially developed for global optimization, also applies to solving the GMP. This is particularly interesting as moments and positive polynomials are at the intersection of several areas of mathematics (Landau 1987), and the list of important applications of the GMP is almost endless; see e.g. Landau (1987), Lasserre (2009*b*) and references therein, and see also Section 6, where for illustration we describe two particular applications.

Finally, since its birth in 2000 and in view of its many potential applications, the Moment-SOS hierarchy has gained attention from various research communities, with many different contributions, as follows.

(i) Its basic application in many (and diverse) areas after modelling the problem as an instance of the GMP. For illustration, two examples are described in Section 6; see also Henrion, Korda and Lasserre (2021), Korda, Henrion and Lasserre (2022) and references therein.

(ii) Its adaptation and extension to other domains, e.g. operations research (Parekh 2023), and entanglement, i.e. violation of Bell inequalities in quantum information, where its non-commutative version (the NPA hierarchy described in Navascués, Pironio and Acín (2012)) is also attracting a lot of attention; see also Burgdorf, Klep and Povh (2016), Parekh and Thompson (2021) and references therein.

(iii) Its detailed analysis by the TCS research community for hardness of approximation in combinatorial optimization (e.g. in relation to issues around the Unique Games Conjecture). See e.g. Barak and Steurer (2014), Raghavendra, Schramm and Steurer (2018) and Bafna *et al.* (2021).

(iv) The analysis of its rate of convergence with very interesting recent results on specific sets; see e.g. Slot and Laurent (2021), Slot (2022), Bach and Rudi (2023) and references therein.

(v) The development of algorithmic improvements to improve scalability of the standard Moment-SOS hierarchy. One direction is to take into account several types of sparsity or symmetries often present in large-scale optimization problems, as explained in Section 3.8. Another is to promote alternatives (e.g. first-order methods, second-order cone programming) to the costly interior point algorithm for semidefinite programming; see e.g. Ahmadi and Majumdar (2019), Yurtsever *et al.* (2021) and Ngoc Hoang Anh Mai *et al.* (2022).

Structure of the paper. For ease of exposition and clarity, we have not provided the proof of most results in the form of theorems and lemmas. However, at the end of each section we have included a subsection entitled 'Notes and sources', with pointers to articles for detailed proofs, and sometimes a discussion and comments on the results.

After introducing some notation and definitions, in Section 3 we describe the Moment-SOS hierarchy of *lower bounds* which converge to the global minimum in polynomial optimization. In Section 4 we provide a brief description of an alternative, the Moment-LP hierarchy. Section 5 describes the (less known) Moment-SOS hierarchy of *upper bounds* which also converge to the global minimum. Section 6 is devoted to other applications of the Moment-SOS hierarchy. For illustration, we describe how to apply the Moment-SOS hierarchy in two such applications and provide a (non-exhaustive) list of references to other applications in various fields. Finally, Section 7 is devoted to the Christoffel function (a classical tool from the theory of orthogonal polynomials and approximation) to reveal its connections to optimization and the Moment-SOS hierarchy.

2. Notation, definitions and some preliminaries

2.1. Notation and definitions

Let $\mathbb{R}[\mathbf{x}]$ denote the ring of polynomials in the variables $\mathbf{x} = (x_1, \dots, x_d)$ and let $\mathbb{R}[\mathbf{x}]_n$ be the vector space of polynomials of degree at most n (whose dimension is $s(d) := \binom{d+n}{d}$). For every $n \in \mathbb{N}$, let $\mathbb{N}^d_n := \{\boldsymbol{\alpha} \in \mathbb{N}^d : |\boldsymbol{\alpha}| \, (= \sum_i \alpha_i) \leq n\}$, and let $\mathbf{v}_n(\mathbf{x}) = (\mathbf{x}^{\alpha})$, $\boldsymbol{\alpha} \in \mathbb{N}^d_n$, be the vector of monomials of the canonical basis $(\mathbf{x}^{\alpha})$ of $\mathbb{R}[\mathbf{x}]_n$. Given a closed set $\mathcal{X} \subseteq \mathbb{R}^n$, let $\mathcal{P}(\mathcal{X}) \subset \mathbb{R}[\mathbf{x}]$ (resp. $\mathcal{P}_n(\mathcal{X}) \subset \mathbb{R}[\mathbf{x}]_n$) be the convex cone of polynomials (resp. polynomials of degree at most n) that are non-negative on $\mathcal{X}$. A polynomial $f \in \mathbb{R}[\mathbf{x}]_n$ is written

$$\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^d} f_\alpha \, \mathbf{x}^\alpha = \langle \mathbf{f}, \mathbf{v}_n(\mathbf{x}) \rangle,$$

with vector of coefficients $\mathbf{f} = (f_\alpha) \in \mathbb{R}^{s(n)}$ in the canonical basis of monomials $(\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}^d}$. For real symmetric matrices, let $\langle \mathbf{B}, \mathbf{C} \rangle := \operatorname{trace}(\mathbf{B}\,\mathbf{C})$. We write $\mathbf{B} \succeq 0$ when $\mathbf{B}$ is positive semidefinite and $\mathbf{B} \succ 0$ when $\mathbf{B}$ is positive definite. Let $\mathcal{S}^n$ denote the space of real $n \times n$ symmetric matrices and let $\mathcal{S}^n_+$ denote its subset of positive semidefinite matrices.

For a closed set $S \subset \mathbb{R}^d$, let $\mathscr{M}(S)$ denote the space of finite signed Borel measures on S, and let $\mathscr{M}(S)_+ \subset \mathscr{M}(S)$ (resp. $\mathscr{P}(S)$) denote the convex cone of finite non-negative Borel measures (resp. probability measures) on S. The support $\operatorname{supp}(\mu)$ of a Borel measure μ on $\mathbb{R}^d$ is the smallest closed set $\Omega \subset \mathbb{R}^d$ such that $\mu(\mathbb{R}^d \setminus \Omega) = 0$.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Riesz linear functional. Given a sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^d}$ (in bold), its associated Riesz linear functional is the linear mapping $\phi\colon \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ (not in bold) defined by

$$f \left(= \sum_\alpha f_\alpha\, \mathbf{x}^\alpha \right) \mapsto \phi(f) = \sum_{\alpha\in\mathbb{N}^d} f_\alpha\, \phi_\alpha = \langle \mathbf{f}, \boldsymbol{\phi}\rangle. \tag{2.1}$$

A sequence $\boldsymbol{\phi}$ has a *representing* measure if its associated Riesz linear functional ϕ is a (positive) Borel measure on $\mathbb{R}^d$, in which case

$$\phi_\alpha = \int_{\mathbb{R}^d} \mathbf{x}^\alpha\, \mathrm{d}\phi \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^d.$$

Given a sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^d}$ and a polynomial $g \in \mathbb{R}[\mathbf{x}]$, $\mathbf{x} \mapsto g(\mathbf{x}) = \sum_\gamma g_\gamma\, \mathbf{x}^\gamma$, define the new sequence $g \cdot \boldsymbol{\phi}$ by

$$(g\cdot\boldsymbol{\phi})_\alpha := \phi(\mathbf{x}^\alpha\, g) = \sum_{\gamma\in\mathbb{N}^d} g_\gamma\, \phi_{\alpha+\gamma} \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^d.$$

Therefore its associated Riesz linear functional, denoted by $g\cdot\phi$, satisfies

$$g\cdot\phi(f) = \phi(g\, f) \quad \text{for all } f \in \mathbb{R}[\mathbf{x}].$$

In particular, if $\boldsymbol{\phi}$ has a representing measure ϕ and g is non-negative, then the Riesz linear functional $g\cdot\phi$ is a representing measure, that is,

$$g\cdot\phi(f) = \phi(g\, f) = \int_{\mathbb{R}^d} f\, g\, \mathrm{d}\phi \quad \text{for all } f \in \mathbb{R}[\mathbf{x}].$$

Moment matrix. The (degree-n) *moment* matrix associated with a sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^d}$ (or equivalently with the Riesz linear functional ϕ) is the real symmetric matrix $\mathbf{M}_n(\boldsymbol{\phi})$ (or $\mathbf{M}_n(\phi)$) with rows and columns indexed by $\mathbb{N}^d_n$, and whose entry $(\boldsymbol{\alpha},\boldsymbol{\beta})$ is just $\phi_{\alpha+\beta}$, for every $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^d_n$. Thus $\mathbf{M}_n(\boldsymbol{\phi})$ depends only on moments ϕ_α of degree at most $2n$. Alternatively, if we introduce the real symmetric matrices $(\mathbf{B}^1_\alpha) \subset \mathcal{S}^{s(n)}$ defined by

$$\mathbf{v}_n(\mathbf{x})\, \mathbf{v}_n(\mathbf{x})^\top = \sum_{\alpha\in\mathbb{N}^d_{2n}} \mathbf{B}^1_\alpha\, \mathbf{x}^\alpha \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{2.2}$$

then $\mathbf{M}_n(\boldsymbol{\phi}) = \sum_{\alpha\in\mathbb{N}^d_{2n}} \phi_\alpha\, \mathbf{B}^1_\alpha$. Moreover, if $\boldsymbol{\phi}$ has a representing measure ϕ, then $\mathbf{M}_n(\boldsymbol{\phi}) \succeq 0$ because $\langle \mathbf{f}, \mathbf{M}_n(\boldsymbol{\phi})\,\mathbf{f}\rangle = \int f^2 \mathrm{d}\phi \geq 0$ for all $f \in \mathbb{R}[\mathbf{x}]_n$.

A measure whose moments are all finite is *moment determinate* if there is no other measure with the same moments.

Localizing matrix. With $\boldsymbol{\phi}$ as above and $g \in \mathbb{R}[\mathbf{x}]$ (with $g(\mathbf{x}) = \sum_\gamma g_\gamma \mathbf{x}^\gamma$), the localizing matrix associated with $\boldsymbol{\phi}$ and g is the moment matrix $\mathbf{M}_n(g\cdot\boldsymbol{\phi})$ associated with the sequence $g\cdot\boldsymbol{\phi}$. That is, $\mathbf{M}_n(g\cdot\boldsymbol{\phi})$ is the real symmetric matrix with rows and columns indexed by $\mathbb{N}^d_n$, and whose entry $(\boldsymbol{\alpha},\boldsymbol{\beta})$ is just $(g\cdot\phi)_{\alpha+\beta}$, that is, $\mathbf{M}_n(g\cdot\boldsymbol{\phi})(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_\gamma g_\gamma \phi_{\alpha+\beta+\gamma}$, for every $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{N}^d_n$.

Alternatively, letting $d_g := \lceil \deg(g)/2 \rceil$, and introducing the real symmetric matrices $\mathbf{B}_\alpha^g \in \mathcal{S}^{s(n)}$, $\boldsymbol{\alpha} \in \mathbb{N}^d$, defined by

$$g(\mathbf{x})\, \mathbf{v}_n(\mathbf{x})\, \mathbf{v}_n(\mathbf{x})^\top = \sum_{\alpha \in \mathbb{N}^d_{2(n+d_g)}} \mathbf{B}_\alpha^g\, \mathbf{x}^\alpha \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{2.3}$$

we obtain

$$\mathbf{M}_n(g \cdot \boldsymbol{\phi}) = \sum_{\alpha \in \mathbb{N}^d_{2(n+d_g)}} \phi_\alpha\, \mathbf{B}_\alpha^g.$$

If $\boldsymbol{\phi}$ has a representing measure ϕ whose support is contained in the set $\{\mathbf{x}\colon g(\mathbf{x}) \geq 0\}$, then $\mathbf{M}_n(g \cdot \boldsymbol{\phi}) \succeq 0$ for all n, because

$$\langle \mathbf{f}, \mathbf{M}_n(g \cdot \boldsymbol{\phi})\, \mathbf{f} \rangle = g \cdot \phi(f^2) = \phi(f^2 g) = \int f^2\, g\, \mathrm{d}\phi \geq 0 \quad \text{for all } f \in \mathbb{R}[\mathbf{x}]_n.$$

2.2. *SOS polynomials and quadratic modules*

A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is a sum of squares (SOS) if there exist $s \in \mathbb{N}$, and $f_1, \ldots, f_s \in \mathbb{R}[\mathbf{x}]$, such that $f(\mathbf{x}) = \sum_{k=1}^s f_k(\mathbf{x})^2$, for all $\mathbf{x} \in \mathbb{R}^d$. Let $\Sigma[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x}]_n$) denote the set of SOS polynomials (resp. SOS polynomials of degree at most $2n$). Of course, every SOS polynomial is non-negative. However, the converse is not true.

Membership in $\Sigma[\mathbf{x}]_n$. Checking whether a given polynomial f is non-negative on $\mathbb{R}^d$ is difficult, whereas – and this is crucial for the Moment-SOS hierarchy – checking whether f is SOS is much easier and can be done efficiently. Indeed, since the degree of f must be even for f to be SOS, we let $f \in \mathbb{R}[\mathbf{x}]_{2n}$ be defined by

$$\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^d_{2n}} f_\alpha\, \mathbf{x}^\alpha.$$

Then $f \in \mathbb{R}[\mathbf{x}]_{2n}$ is SOS if and only if there exists a real symmetric matrix $\mathbf{X}^\top = \mathbf{X}$ of size $s(n) = \binom{d+n}{d}$, such that

$$\mathbf{X} \succeq 0 \quad \text{and} \quad f_\alpha = \langle \mathbf{X}, \mathbf{B}_\alpha^1 \rangle \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^d_{2n}, \tag{2.4}$$

where the matrices $\mathbf{B}_\alpha^1$ were introduced in (2.2). It turns out that (2.4) defines the feasible set of what is called a *semidefinite program.*[3]

[3] A semidefinite program is a convex and conic optimization problem which can be solved (up to fixed arbitrary precision) in time polynomial in its input size; see e.g. Anjos and Lasserre (2012) and also O'Donnell (2017).

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Quadratic module. Introduce the constant polynomial $\mathbf{x} \mapsto g_0(\mathbf{x}) := 1$ for all $\mathbf{x} \in \mathbb{R}^d$ (also denoted $g_0 = \mathbf{1}$). Given a family $(g_1, \dots, g_m) \subset \mathbb{R}[\mathbf{x}]$, we associate the *quadratic module* $Q(g)\,(= Q(g_1, \dots, g_m)) \subset \mathbb{R}[\mathbf{x}]$ defined by

$$Q(g) := \left\{ \sum_{j=0}^{m} \sigma_j\, g_j : \sigma_j \in \Sigma[\mathbf{x}],\ j = 0, \dots, m \right\}, \tag{2.5}$$

and its degree-$2n$ *truncated* version

$$Q_n(g) := \left\{ \sum_{j=0}^{m} \sigma_j\, g_j : \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j},\ j = 0, \dots, m \right\}, \tag{2.6}$$

where $d_j := \lceil \deg(g_j)/2 \rceil$, $j = 0, \dots, m$. Observe that $Q_n(g) \subset \mathbb{R}[\mathbf{x}]_{2n}$, because indeed in (2.6), $\deg(\sigma_j\, g_j) \le 2n$, for all $j = 0, \dots, m$. Obviously both $Q(g)$ and its truncated version $Q_n(g)$ are convex cones of $\mathbb{R}[\mathbf{x}]$.

Definition 2.1. The quadratic module $Q(g)$ is said to be *Archimedean* if there exists $M > 0$ such that the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ belongs to $Q(g)$ (i.e. belongs to $Q_n(g)$ for some n).

If $Q(g)$ is Archimedean, then necessarily the set

$$S := \{\mathbf{x} \in \mathbb{R}^d : g_j(\mathbf{x}) \ge 0,\ j = 1, \dots, m\} \tag{2.7}$$

is compact but the reverse is not true. The Archimedean condition depends on the *representation* of S and can be seen as an *algebraic certificate* that S is compact.

Dual cone. The dual cone $Q_n^*(g)$ of $Q_n(g)$ is the convex cone of $\mathbb{R}^{s(2n)}$ defined by

$$Q_n^*(g) = \{\boldsymbol{\phi} \in \mathbb{R}^{s(2n)} : \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \succeq 0,\ j = 0, \dots, m\}, \tag{2.8}$$

where $\mathbf{M}_n(g_j \cdot \boldsymbol{\phi})$ is the localizing matrix associated with the polynomial g_j and the sequence $\boldsymbol{\phi}$, defined in Section 2.1.

For more details on the above notions of moment and localizing matrix and the quadratic module, as well as their use in potential applications, the interested reader is referred to Laurent (2008) and Lasserre (2009*b*). As we will see, both convex cones $Q_n(g)$ and $Q_n^*(g)$ play a crucial role in the Moment-SOS hierarchy of lower bounds.

2.3. *Certificates of positivity (Positivstellensätze)*

Below we describe particular certificates of positivity that are important because they provide a theoretical justification (or rationale) for convergence of the so-called SDP and LP relaxations for global optimization. In particular, the one below in (2.9) is at the core of the Moment-SOS hierarchy of lower bounds.

Theorem 2.2 (Putinar 1993). Let $S \subset \mathbb{R}^d$ be as in (2.7) and assume that $Q(g)$ is Archimedean.

(i) If a polynomial $f \in \mathbb{R}[\mathbf{x}]$ is (strictly) positive on S, then $f \in Q(g)$, that is,

$$f = \sum_{j=0}^{m} \sigma_j \, g_j, \tag{2.9}$$

for some SOS polynomials $\sigma_j \in \Sigma[\mathbf{x}]$, $j = 0, \ldots, m$ (and thus $f \in Q_n(g)$ for some $2n \geq \deg(f)$).

(ii) A sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha \in \mathbb{N}^d} \subset \mathbb{R}$ has a representing Borel measure on S if and only if $\phi(f^2 \, g_j) \geq 0$ for all $f \in \mathbb{R}[\mathbf{x}]$, and all $j = 0, \ldots, m$, or equivalently if and only if $\mathbf{M}_n(g_j \cdot \boldsymbol{\phi}) \succeq 0$ for all $j = 0, \ldots, m$, and all $n \in \mathbb{N}$.

In fact Theorem 2.2 is a refinement of a theorem by Schmüdgen from two years earlier.

Theorem 2.3 (Schmüdgen 1991). Let the basic semi-algebraic set $S \subset \mathbb{R}^d$ in (2.7) be compact.

(i) If a polynomial $f \in \mathbb{R}[\mathbf{x}]$ is (strictly) positive on S, then

$$f = \sum_{\alpha \in \{0,1\}^m} \sigma_\alpha \, g_1^{\alpha_1} \cdots g_m^{\alpha_m}, \tag{2.10}$$

for some SOS polynomials $\sigma_\alpha \in \Sigma[\mathbf{x}]$, $\boldsymbol{\alpha} \in \{0, 1\}^m$.

(ii) A sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha \in \mathbb{N}^d} \subset \mathbb{R}$ has a representing Borel measure on S if and only if $\phi(f^2 \, g_1^{\alpha_1} \cdots g_m^{\alpha_m}) \geq 0$ for all $f \in \mathbb{R}[\mathbf{x}]$, and all $\boldsymbol{\alpha} \in \{0, 1\}^m$, or equivalently if and only if $\mathbf{M}_n(g_1^{\alpha_1} \cdots g_m^{\alpha_m} \cdot \boldsymbol{\phi}) \succeq 0$ for all $\boldsymbol{\alpha} \in \{0, 1\}^m$, and all $n \in \mathbb{N}$.

Observe that (2.9) is of the same flavour as (2.10) but much simpler, as it involves only $m + 1$ SOS polynomials $\sigma_j \in \Sigma[\mathbf{x}]$ (as opposed to 2^m SOS σ_α in (2.10)). On the other hand, in Theorem 2.3 the only condition is that the set S is compact, whereas in Theorem 2.2 the quadratic module $Q(g)$ should also be Archimedean (an additional condition on the representation of S).

The reader may have noticed that Theorems 2.2 and 2.3 have two facets, (i) and (ii): the former is the algebraic facet (certificate of positivity), while the latter with a real analysis flavour is related to the S-moment problem. Both facets are a nice illustration of the duality between moments and positive polynomials.

We next provide a Nichtnegativstellensatz (a theorem of non-negativity) of the author (Lasserre 2011), which is instrumental in proving convergence of the hierarchy of upper bounds in Section 5.

Theorem 2.4. Let $S \subset \mathbb{R}^d$ be a compact set (not necessarily basic semi-algebraic), and let ϕ be a Borel measure with $\mathrm{supp}(\phi) = S$, and with moment sequence

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

$\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^d}$. If $f \in \mathbb{R}[\mathbf{x}]$, then

$$f \geq 0 \text{ on } S \quad\Leftrightarrow\quad \mathbf{M}_n(f\cdot\boldsymbol{\phi}) \succeq 0 \quad \text{for all } n\in\mathbb{N}. \tag{2.11}$$

Theorem 2.4 states that to decide whether f is non-negative on S, one must check whether countably many linear matrix inequalities (LMIs) $\mathbf{M}_n(f\cdot\boldsymbol{\phi}) \succeq 0$, $n\in\mathbb{N}$, hold. Each constraint $\mathbf{M}_n(f\cdot\boldsymbol{\phi}) \succeq 0$ is indeed an LMI on the coefficients of the polynomial f because each entry of $\mathbf{M}_n(f\cdot\boldsymbol{\phi})$ is *linear* in the coefficients $\mathbf{f}$ of f. Therefore, identifying $f\in\mathbb{R}[\mathbf{x}]_k$ with its vector $\mathbf{f}\in\mathbb{R}^{s(k)}$ of coefficients, for each $n\in\mathbb{N}$, the convex cone $\Omega_n\subset\mathbb{R}[\mathbf{x}]_k$, defined by

$$\Omega_n := \{\mathbf{f}\in\mathbb{R}^{s(k)} : \mathbf{M}_n(f\cdot\boldsymbol{\phi}) \succeq 0\},$$

is a spectrahedron that contains the convex cone $\mathcal{P}_k(S)$ of polynomials of degree at most k that are non-negative on S. In addition, $\Omega_{n+1}\subset\Omega_n$ for all n, so that the sequence $(\Omega_n)_{n\in\mathbb{N}}$ of outer approximations of $\mathcal{P}_k(S)$ is monotone non-increasing. Moreover, it converges to $\mathcal{P}_k(S)$, i.e. $\bigcap_{n\in\mathbb{N}}\Omega_n = \mathcal{P}_k(S)$.

So in contrast to Theorems 2.2 and 2.3, Theorem 2.4 is valid for arbitrary compact sets $S\subset\mathbb{R}^d$ and non-negative (as opposed to positive) polynomials on S. On the other hand, its practical use in over-approximating the convex cone $\mathcal{P}_k(S)$ by Ω_n requires knowledge of moments $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^d}$ of a measure ϕ with $\operatorname{supp}(\phi) = S$. This is only possible for specific sets and measures. Examples of such special sets include the unit box, unit Euclidean ball, unit sphere, canonical simplex, discrete cube $\{-1,1\}^d$, and their image by an affine transformation.

LP-based certificate. We next introduce another certificate of positivity which does not use SOS. Given $g_1,\ldots,g_m\in\mathbb{R}[\mathbf{x}]$, introduce the notation $\mathbf{g}^\alpha\in\mathbb{R}[\mathbf{x}]$, and $(1-\mathbf{g}^\alpha)\in\mathbb{R}[\mathbf{x}]$, with

$$\mathbf{x}\mapsto\mathbf{g}^\alpha(\mathbf{x}) := g_1(\mathbf{x})^{\alpha_1}\cdots g_m(\mathbf{x})^{\alpha_m} \quad \text{for all } \mathbf{x}\in\mathbb{R}^d,$$
$$\mathbf{x}\mapsto(\mathbf{1}-\mathbf{g})^\alpha(\mathbf{x}) := (1-g_1(\mathbf{x}))^{\alpha_1}\cdots(1-g_m(\mathbf{x}))^{\alpha_m} \quad \text{for all } \mathbf{x}\in\mathbb{R}^d,$$

and the convex cone $\mathcal{L}_n(g)\subset\mathbb{R}[\mathbf{x}]$ defined by

$$\mathcal{L}_n(g) := \left\{ \sum_{(\alpha,\beta)\in\mathbb{N}^{2m}_n} c_{\alpha\beta}\,\mathbf{g}^\alpha\,(\mathbf{1}-\mathbf{g})^\beta : \mathbf{c} = (c_{\alpha\beta}) \geq 0 \right\}. \tag{2.12}$$

Theorem 2.5 (Krivine 1964*a*,*b*, Vasilescu 2003). Let $S\subset\mathbb{R}^d$ as in (2.7) be compact and such that (possibly after scaling) $0\leq g_j(\mathbf{x})\leq 1$ for all $\mathbf{x}\in S$, $j=1,\ldots,m$. Assume also that $[\mathbf{1},g_1,\ldots,g_m]$ generates $\mathbb{R}[\mathbf{x}]$.

(i) If a polynomial $f\in\mathbb{R}[\mathbf{x}]$ is (strictly) positive on S, then $f\in\mathcal{L}_n(g)$ for some n, that is,

$$f = \sum_{(\alpha,\beta)\in\mathbb{N}^{2m}} c_{\alpha\beta}\,\mathbf{g}^\alpha\,(\mathbf{1}-\mathbf{g})^\beta \tag{2.13}$$

for some non-negative vector $\mathbf{c} = (c_{\alpha\beta})_{(\alpha,\beta)\in\mathbb{N}^{2m}_n}$.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

(ii) A sequence $\boldsymbol{\phi} = (\phi_\gamma)_{\gamma\in\mathbb{N}^d} \subset \mathbb{R}$ has a representing Borel measure on S if and only if $\phi(\mathbf{g}^\alpha\,(\mathbf{1}-\mathbf{g})^\beta) \geq 0$ for all $\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^m$.

Remark 2.6. Interestingly, as for Theorems 2.2 and 2.3, Theorem 2.5 also has two facets. The algebraic facet (i) is concerned with representation of polynomials that are positive on S, while facet (ii) is concerned with the S-moment problem in real analysis. Hence Theorem 2.5 is another illustration of the duality between *polynomials positive on* S and the S-*moment problem*.

2.4. *Practical implementation of Positivstellensätze via SDP or LP*

In addition to being interesting in their own right, Theorems 2.2(i) and 2.5(i) have another distinguishing feature. Both have a *practical implementation* that allows us to perform interesting computations. Indeed:

- testing membership in $Q_n(g)$ is just solving a single semidefinite program, whereas
- testing membership in $\mathcal{L}_n(g)$ is just solving a single linear program.

Testing membership in $Q_n(g)$. This is crucial for a practical and efficient implementation of the Moment-SOS hierarchy. Fortunately it can be done by solving a semidefinite program. Namely, let $f \in \mathbb{R}[\mathbf{x}]_k$ and recall that $d_j = \lceil \deg(g_j)/2 \rceil$, $j = 0,\ldots,m$. Then testing whether $f \in Q_n(g)$ (where necessarily $2n \geq k$) reduces to solving

$$f_\alpha = \sum_{j=0}^{m} \langle \mathbf{X}_j, \mathbf{B}_\alpha^{g_j} \rangle \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^d_{2n},\ \mathbf{X}_j \in \mathcal{S}^{n-d_j}, \tag{2.14}$$

$$\mathbf{X}_j \succeq 0, \quad j = 0,\ldots,m, \tag{2.15}$$

where the real symmetric matrices $\mathbf{B}_\alpha^{g_j} \in \mathcal{S}^{s(n-d_j)}$ are defined in (2.3) (here with $n - d_j$ instead of n). Each real symmetric matrix $\mathbf{X}_j$ is a Gram matrix of a polynomial σ_j, $j = 0,\ldots,m$. Next, (2.14) are *linear* equality constraints on the unknown entries of $\mathbf{X}_j$, while (2.15) is a positive semidefinite constraint on $(\mathbf{X}_j)_{j=0}^m$ to ensure that every σ_j is an SOS. (In (2.14), $f_\alpha = 0$ whenever $k < |\boldsymbol{\alpha}| \leq 2n$ because $f \in \mathbb{R}[\mathbf{x}]_k$.) Observe that multiplying (2.14) by $\mathbf{x}^\alpha$ and summing up yields

$$\begin{aligned} f(\mathbf{x}) &= \sum_{\alpha\in\mathbb{N}^d_{2n}} f_\alpha\,\mathbf{x}^\alpha \\ &= \sum_{j=0}^{m} \left\langle \mathbf{X}_j, \sum_{\alpha\in\mathbf{B}^d_{2n}} \mathbf{B}_\alpha^{g_j}\mathbf{x}^\alpha \right\rangle \end{aligned}$$

$$= \sum_{j=0}^{m} \langle \mathbf{X}_j, \mathbf{v}_{n-d_j}(\mathbf{x})\mathbf{v}_{n-d_j}(\mathbf{x})^\top \rangle \, g_j(\mathbf{x}) \quad \text{(see (2.3))}$$

$$= \sum_{j=0}^{m} \underbrace{\mathbf{v}_{n-d_j}(\mathbf{x})^\top \mathbf{X}_j \mathbf{v}_{n-d_j}(\mathbf{x})}_{\sigma_j(\mathbf{x})} \, g_j(\mathbf{x}).$$

Hence checking whether (2.14)–(2.15) has a solution reduces to solving a semidefinite program.

Testing membership in $\mathcal{L}_n(g)$. Obviously testing whether some polynomial $f \in \mathbb{R}[\mathbf{x}]_k$ is in $\mathcal{L}_n(g)$ reduces to solving an LP problem. Indeed, with n such that $s := (\max_j \deg(g_j))^n \geq k$, it amounts to finding a non-negative vector $\mathbf{c} = (c_{\alpha,\beta})$, $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{N}_n^{2m}$, such that

$$f_\gamma = \left(\sum_{(\alpha,\beta)\in\mathbb{N}_n^{2m}} c_{\alpha\beta} \prod_{j=1}^{m} g_j(\mathbf{x})^{\alpha_j} \, (1 - g_j(\mathbf{x}))^{\beta_j} \right)_\gamma \quad \text{for all } \boldsymbol{\gamma} \in \mathbb{N}_s^d,$$

and with $f_\gamma = 0$ whenever $|\boldsymbol{\gamma}| > k$. Clearly, the above constraints are *linear* in the unknown coefficients $c_{\alpha\beta} \geq 0$, so checking the existence of such a vector $\mathbf{c} \geq 0$ reduces to solving an LP problem.

2.5. *Notes and sources*

Most of the material is from Lasserre (2009*b*, 2015). Good references for exhaustive results on positive polynomials and moment problems are Laurent (2008), Marshall (2008), Prestel and Denzel (2001), Schmüdgen (2017), Nie (2023) and Kočvara, Mourrain and Riener (2023); see also Blekherman, Parrilo and Thomas (2012) for related material on convex algebraic geometry.

Section 2.3. Theorem 2.4 is from Lasserre (2011, 2013). Interestingly, it is also valid for some non-compact sets, such as the positive orthant $\mathbb{R}_+^d$ or even the whole space $\mathbb{R}^d$. For the former, the measure ϕ can be chosen to be the exponential measure $\mathrm{d}\phi = \exp(-\sum_i x_i)\,\mathrm{d}\mathbf{x}$, while for the latter we may choose the Gaussian measure $\mathrm{d}\phi = \exp(-\|\mathbf{x}\|^2/2)\,\mathrm{d}\mathbf{x}$. In both cases we obtain a monotone sequence $(\boldsymbol{\Omega}_n)_{n\in\mathbb{N}}$ of outer approximations, which converge to $\mathcal{P}(\mathbb{R}_+^d)$ and $\mathcal{P}(\mathbb{R}^d)$ respectively.

Section 2.4. It is worth mentioning that other certificates of positivity (via convex cones of positive polynomials) have also been defined to overcome (or at least mitigate) the computational burden associated with testing membership in $\mathcal{Q}_n(g)$ in (2.14)–(2.15) (via semidefinite programming). For instance, membership in corresponding alternative convex cones can be checked by linear programming for DSOS and second-order cone programming for SDSOS (Ahmadi and Majumdar 2019); see also Majumdar, Hall and Ahmadi (2020). An alternative described in Section 3.8 is to consider a sparse version of Theorem 2.2 when $\mathbf{P}$ exhibits some sparsity pattern. As we will see, it yields a sparsity-adapted version of the

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Moment-SOS hierarchy which can handle non-convex POPs with more than a thousand variables; see also Magron and Wang (2023).

3. The Moment-SOS hierarchy in optimization

Consider the polynomial optimization problem (POP) **P** in (1.1), and assume that its associated feasible set $S \subset \mathbb{R}^d$ is compact.

3.1. The Moment-SOS hierarchy

The underlying principle behind the Moment-SOS hierarchy is quite simple and proceeds in two steps.

Viewing **P** *in its equivalent formulation* (1.3) *(real algebraic glasses).*

Step 1. We replace the hard constraint '$f - \lambda \geq 0$ on S' with the equivalent constraint $f \in Q(g)$ (where $Q(g)$ is the quadratic module defined in (2.5)). Indeed,

$$\begin{aligned} f^* &= \sup_\lambda \{\lambda\colon f - \lambda \geq 0 \text{ on } S\} \\ &= \sup_\lambda \{\lambda\colon f - \lambda > 0 \text{ on } S\} \\ &= \sup_\lambda \{\lambda\colon f - \lambda \in Q(g)\}, \end{aligned} \tag{3.1}$$

where the second equality follows from Theorem 2.2(i) if the quadratic module $Q(g)$ is Archimedean. However, (3.1) is still an infinite-dimensional problem.

Step 2. Next, with $n_0 := \max[\lceil \deg(f)/2\rceil, \max_j \lceil \deg(g_j)/2\rceil]$, and $n \geq n_0$, we replace (3.1) with the more restrictive constraint

$$\tau_n^* = \sup_\lambda \{\lambda\colon f - \lambda \in Q_n(g)\} \quad (n \geq n_0) \tag{3.2}$$

$$= \sup_{\lambda, \sigma_j} \left\{\lambda\colon f - \lambda = \sum_{j=0}^m \sigma_j\, g_j;\ \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j},\ \forall\, j\right\}, \tag{3.3}$$

so that $f^* \geq \tau_n^*$ for all $n \geq n_0$. A crucial feature of (3.3) is that it is a finite-dimensional convex optimization problem, and more precisely a *semidefinite program*. Therefore (3.3) can be solved (up to arbitrary fixed precision) in time polynomial in its input size.

So, solving (3.2) for increasing values of $n \in \mathbb{N}$, we obtain a monotone non-decreasing sequence $(\tau_n^*)_{n \geq n_0}$ of lower bounds on the global minimum of f^* of **P**.

Viewing **P** *in its equivalent formulation* (1.4) *(real analysis glasses).* First observe that since $f \in \mathbb{R}[\mathbf{x}]$,

$$\int_S f \,\mathrm{d}\phi = \sum_{\alpha \in \mathbb{N}^d} f_\alpha \int_S \mathbf{x}^\alpha \,\mathrm{d}\phi = \sum_{\alpha \in \mathbb{N}^d} f_\alpha\, \phi_\alpha.$$

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Therefore

$$f^* = \inf_{\phi\in\mathscr{M}(S)_+} \left\{\int f\,\mathrm{d}\phi \colon \phi(S) = 1\right\}$$
$$= \inf_{\boldsymbol{\phi}=(\phi_\alpha)_{\alpha\in\mathbb{N}^d}} \{\langle \mathbf{f}, \boldsymbol{\phi}\rangle \colon \phi_0 = 1;\ \boldsymbol{\phi} \text{ has a representing measure supported on } S\}.$$

Again we proceed in two steps.

Step 1. If $Q(g)$ is Archimedean, then by invoking Theorem 2.2(ii), we may replace the constraint '$\boldsymbol{\phi}$ has a representing measure supported on S' with the equivalent constraint '$\mathbf{M}_n(g_j\cdot\boldsymbol{\phi}) \succeq 0$ for all $j = 0, \dots, m$, and all $n \in \mathbb{N}$'. However, the optimization problem

$$f^* = \inf_{\boldsymbol{\phi}=(\phi_\alpha)} \{\langle \mathbf{f}, \boldsymbol{\phi}\rangle \colon \phi_0 = 1;\ \mathbf{M}_n(g_j\cdot\boldsymbol{\phi}) \succeq 0,\ j = 0, \dots, m,\ n \in \mathbb{N}\} \tag{3.4}$$

is still infinite-dimensional.

Step 2. Next we replace (3.4) with its truncated versions

$$\tau_n = \inf_{\boldsymbol{\phi}=(\phi_\alpha)} \{\langle \mathbf{f}, \boldsymbol{\phi}\rangle \colon \phi_0 = 1;\ \mathbf{M}_{n-d_j}(g_j\cdot\boldsymbol{\phi}) \succeq 0,\ j = 0, \dots, m\}, \tag{3.5}$$

where $n \geq n_0$ with $n_0 := \max[\lceil\deg(f)/2\rceil, \max_j\lceil\deg(g_j)/2\rceil]$. For each fixed n, (3.5) is a finite-dimensional semidefinite program.

Clearly, $\tau_n \leq \tau_{n+1} \leq f^*$ for all $n \geq n_0$, and therefore, solving (3.5) for increasing values of $n \in \mathbb{N}$, we obtain a monotone non-decreasing sequence $(\tau_n)_{n\geq n_0}$ of lower bounds on the global minimum of f^* of $\mathbf{P}$.

In fact the semidefinite program (3.3) is the dual of the semidefinite program (3.5), and by *weak duality* in convex optimization,

$$\tau_n^* \leq \tau_n \leq f^* \quad \text{for all } n \geq n_0. \tag{3.6}$$

As is clear from its formulation, (3.5) is a semidefinite *relaxation* of $\mathbf{P}$ as its constraints are only necessary conditions for $\boldsymbol{\phi}$ to have a representing measure ϕ on S. We call (3.5) a *Moment-relaxation* of $\mathbf{P}$.

On the other hand, its dual (3.3) is a *reinforcement* (or *strengthening*) of $\mathbf{P}$ (viewed as the maximization problem (1.3)) as we have replaced '$f - \lambda \geq 0$ on S' with the sufficient condition '$f - \lambda \in Q_n(g)$'. We call (3.3) an *SOS-reinforcement* (or *SOS-strengthening*) of $\mathbf{P}$, whence the name 'Moment-SOS hierarchy' for (3.3)–(3.5). In both cases we obtain a lower bound τ_n (resp. τ_n^*) on f^*.

When $\tau_n = f^*$ (resp. $\tau_n^* = f^*$) for some n, we say that the degree-n Moment-relaxation (3.5) (resp. the degree-n SOS-reinforcement (3.3)) of $\mathbf{P}$ is *exact*. In addition, if $\tau_n = f^*$ and an optimal solution $\boldsymbol{\phi}^*$ of (3.5) satisfies $\operatorname{rank}(\mathbf{M}_n(\boldsymbol{\phi}^*)) = 1$, then $\boldsymbol{\phi}^*$ is simply the vector of moments up to degree $2n$ of the Dirac measure $\delta_{\{\mathbf{x}^*\}}$ at a global minimizer $\mathbf{x}^* \in S$. In particular, the subvector $\phi^*(x_i)_{i=1,\dots,d}$ of first-order moments is just the vector $\mathbf{x}^*$.

In summary, the ultimate goal of the Moment-relaxation (3.5) is to obtain, at some step n, an optimal solution $\boldsymbol{\phi}^* \in \mathbb{R}^{s(2n)}$ which is the vector of moments (up to degree $2n$) of the Dirac measure $\delta_{\{\mathbf{x}^*\}}$ at a minimizer $\mathbf{x}^* \in S$.

When **P** has a unique global minimizer $\mathbf{x}^* \in S$, this happens generically; in the case of finitely many global minimizers it will happen generically that at some step n, $\boldsymbol{\phi}^*$ is the vector of moments of a convex combination of Dirac measures at such global minimizers; see Theorem 3.3 and Lemma 3.4 below.

Remark 3.1.

(a) As explained above, it is more appropriate to call (3.3) an SOS-reinforcement of **P** instead of an *SOS-relaxation* of **P**, as is sometimes written in the literature. Of course, it is also the dual of the Moment-relaxation (3.5) of **P**. (Relaxing in the primal is equivalent to reinforcing in the dual.)

(b) We call (1.3) (resp. (3.5)) a *primal* formulation (resp. a *primal* semidefinite relaxation) of **P** because in solving **P** we are mainly interested in an optimal solution $\mathbf{x}^* \in S$ (a global minimizer of **P**), so the primary variable of interest is $\mathbf{x} \in S \subset \mathbb{R}^d$. If $\mathbf{x}^* \in S$ is an optimal solution of **P** then $\boldsymbol{\phi}^* = ((\mathbf{x}^*)^\alpha)_{\alpha\in\mathbb{N}^d}$ is an optimal solution of (3.4).

Then (3.2) is the *dual* of the semidefinite relaxation (3.5), and we will see that indeed, when $\tau_n^* = f^*$ for some n and (3.2) has an optimal solution $\sigma_j^* \in \Sigma[\mathbf{x}]_{n-d_j}$, $j = 0,\dots,m$, then $\sigma_j^*(\mathbf{x}^*) = \lambda_j^* \geq 0$ for all $j = 1,\dots,m$, where the λ_j^* are optimal Karush–Kahn–Tucker (KKT) Lagrange multipliers associated with a global minimizer $\mathbf{x}^* \in S$.

Computational considerations. The Moment-relaxation (3.5) is a semidefinite program with

- $\binom{d+2n}{d}$ variables ϕ_α,
- $m+1$ moment-localizing matrices of size $\binom{d+n-d_j}{d}$, $j = 0,\dots,m$,

while the SOS-strengthening (3.3) is a semidefinite program with

- $1 + \sum_{j=0}^m \binom{d+2n-2d_j}{d}$ variables $(\lambda, (\sigma_0)_\alpha, \dots, (\sigma_m)_\alpha)$,
- $m+1$ semidefinite constraints for matrices of size $\binom{d+n-d_j}{d}$, $j = 0,\dots,m$,
- $\binom{d+2n}{d}$ equality constraints.

For fixed dimension d, the size of matrices and the number of variables in the primal and dual semidefinite programs are polynomial in n. Therefore in principle they can be solved efficiently (up to arbitrary fixed precision) in time polynomial in their input size.[4] However, in view of their non-modest size and the current status of semidefinite solvers, such semidefinite programs can be solved only for POPs

[4] More details can be found in O'Donnell (2017), for example.

of modest dimension d and small degree-n relaxation. So in its canonical form (3.3)–(3.5), the Moment-SOS hierarchy is limited to POPs of modest dimension. However, fortunately we find the following.

- Practice reveals that finite convergence often occurs at low degree n.
- As is often the case for many POPs of large dimension d, some *sparsity pattern* or *symmetries* are present. It turns out that they can be exploited to define appropriate Moment-relaxations (resp. SOS-strengthenings) of **P** whose size is still compatible with current SDP solvers; see Section 3.8 for more details.
- Another possibility is to neglect the costly interior-point methods of SDP solvers and solve the semidefinite programs (3.3) and (3.5) by first-order methods; see e.g. Yurtsever *et al.* (2021) and Ngoc Hoang Anh Mai, Lasserre and Magron (2023).

More details are provided in Section 3.8.

3.2. Convergence of the Moment-SOS hierarchy

Observe that if $S \subset \mathbb{R}^d$ is compact then S is contained in the Euclidean ball of radius M for some $M > 0$, and in many applications M is known. Therefore the quadratic constraint $M^2 - \|\mathbf{x}\|^2 \geq 0$ is redundant when $\mathbf{x} \in S$.

For a practical implementation of the Moment-SOS hierarchy, it is indeed always recommended to add the redundant constraint $g_1(\mathbf{x}) := M^2 - \|\mathbf{x}\|^2 \geq 0$ in the definition (2.7) of S. Moreover, to avoid possible numerical ill-conditioning if M is large, it is even recommended to scale problem **P** in such a manner that $S \subset \mathbf{B}_1 := \{\mathbf{x} \colon \|\mathbf{x}\| \leq 1\}$ so that $g_1(\mathbf{x}) = 1 - \|\mathbf{x}\|^2$. Hence we state this formally.

Assumption 3.2. $S \subset \mathbb{R}^d$ defined in (2.7) is compact with $\mathbf{x} \mapsto g_1(\mathbf{x}) = 1 - \|\mathbf{x}\|^2$ (so that $S \subset \mathbf{B}_1$).

The reason for doing this is because under Assumption 3.2, the quadratic module $Q(g)$ in (2.5) is guaranteed to be Archimedean, a crucial property for convergence of the Moment-SOS hierarchy. (In general, proving that $Q(g)$ is Archimedean may not be trivial.)

Theorem 3.3. With $S \subset \mathbb{R}^d$ as in (2.7), let Assumption 3.2 hold and let τ_n (resp. τ_n^*) be as in (3.5) (resp. (3.3)) for all $n \geq n_0$. Then the following hold.

(a) We have $\tau_n^* = \tau_n$ for all $n \geq n_0$, and for every $n \geq n_0$, the semidefinite relaxation (3.5) has an optimal solution $\boldsymbol{\phi}^n = (\phi_\alpha^n)_{\alpha \in \mathbb{N}_{2n}^d}$. Moreover, both sequences $(\tau_n^*)_{n \geq n_0}$ and $(\tau_n)_{n \geq n_0}$ are monotone non-decreasing, and

$$\lim_{n\to\infty} \tau_n^* = \lim_{n\to\infty} \tau_n = f^*. \tag{3.7}$$

(b) If $\boldsymbol{\phi}^n$ is an optimal solution of (3.5), let $v := \max_{j=1,\ldots,m} \lceil \deg(g_j)/2 \rceil$. If

$$\operatorname{rank} \mathbf{M}_t(\boldsymbol{\phi}^n) = \operatorname{rank} \mathbf{M}_{t-v}(\boldsymbol{\phi}^n) \ (=: s), \tag{3.8}$$

for some $v \le t \le n$, then $\tau_n^* = \tau_n = f^*$ and from the vector $\boldsymbol{\phi}^n$ one may extract $\mathbf{x}^*(\ell) \in S$, $\ell = 1, \dots, s$, where each $\mathbf{x}^*(\ell) \in S$ is a global minimizer of **P**, that is, $f(\mathbf{x}^*(\ell)) = f^*$, $\ell = 1, \dots, s$.

(c) If $\mathrm{int}(S) \neq \emptyset$, then for every $n \ge n_0$, the SOS-strengthening (3.3) of **P** has an optimal solution $(\tau_n^*, \sigma_0^*, \dots, \sigma_m^*)$.

Convergence of minimizers. Theorem 3.3 states that the sequence $(\tau_n)_{n \ge n_0}$ of optimal values converges to the global minimum f^* of **P** as the degree n increases, and moreover extraction of minimizers is also obtained if the degree-n Moment-relaxation is exact ($\tau_n = f^*$) and the flat extension condition (3.8) holds. But what about the sequence of minimizers $(\boldsymbol{\phi}^n)_{n \ge n_0}$ when the convergence is only asymptotic (as opposed to finite)?

Lemma 3.4. Let the sequence $(\boldsymbol{\phi}^n)_{n \ge n_0}$ with $\boldsymbol{\phi}^n = (\phi_\alpha^n)_{\alpha \in \mathbb{N}_{2n}^d}$ be as in Theorem 3.3(a). If $\mathbf{x}^* \in S$ is the unique global minimizer of **P**, then

$$\lim_{n \to \infty} \phi_\alpha^n \left(= \lim_{n \to \infty} \phi^n(\mathbf{x}^\alpha) \right) = (\mathbf{x}^*)^\alpha \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^d. \tag{3.9}$$

In particular, $\lim_{n \to \infty} \phi^n(x_i) = x_i^*$ for every $i = 1, \dots, d$.

So when **P** has a unique global minimizer and convergence of the Moment-relaxation (3.5) is only asymptotic (as opposed to finite), Lemma 3.4 states that the vector of degree-1 moments $(\phi^n(x_i))_{i=1,\dots,d}$ converges to the unique global minimizer $\mathbf{x}^* \in S$ as n increases.

Notice that (3.9) is also interesting even if finite convergence takes place at some n, because we may already obtain a good approximation of $\mathbf{x}^* \in S$ from the degree-1 moments of $\boldsymbol{\phi}^t$ for $t < n$.

Equality constraints. Of course, in (2.7) we may tolerate equality constraints $g_j(\mathbf{x}) \ge 0$ and $g_{j+1}(\mathbf{x}) \ge 0$ with $g_{j+1} = -g_j$, $j \in J$, for some subset $J \subset \{1, \dots, m\}$, in which case we simply write $g_j(\mathbf{x}) = 0$, $j \in J$ (and remove the constraint $g_{j+1} \ge 0$). The resulting modifications are as follows.

- In (3.3) the unknown SOS weight $\sigma_j \in \Sigma[\mathbf{x}]_{n-d_j}$ is now a polynomial in $\mathbb{R}[\mathbf{x}]_{2(n-d_j)}$ and no longer an SOS.
- In (3.5) the positive semidefinite constraint $\mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \succeq 0$ now reads as the equality constraints $\mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) = 0$ on the variables (ϕ_α).
- Theorem 3.3(a,b) remains valid, whereas Theorem 3.3(c) needs some adjustment since $\mathrm{int}(S) = \emptyset$. For instance, if the ideal $\langle g_j \rangle_{j \in J} \subset \mathbb{R}[\mathbf{x}]$ generated by the polynomials g_j associated with the equality constraints is real radical, then for n sufficiently large, the SOS-strengthening (3.3) of **P** has an optimal solution $(\tau_n^*, \sigma_0^*, \dots, \sigma_m^*)$.

Pseudo-Boolean case. An important case is when $S \subset \{-1,1\}^d$ (or equivalently $\{0,1\}^d$ after a simple linear transformation), that is, $J = \{1,\dots,d\}$ and

$$S = \left\{\mathbf{x} \in \mathbb{R}^d : x_j^2 - 1 = 0, j \in J;\ g_j(\mathbf{x}) \geq 0,\ j \notin J\right\}, \tag{3.10}$$

of which the celebrated Max-Cut problem is a particular case (no inequality constraint). Then the ideal

$$\left\langle x_j^2 - 1\right\rangle_{j\in J} = \left\langle x_1^2 - 1,\dots,x_d^2 - 1\right\rangle$$

is indeed real radical and Theorem 3.3 applies. Of course, in this case it follows that $\tau_n = f^*$ whenever $n \geq d + \max_j d_j$, and therefore the semidefinite relaxation (3.5) is not interesting as it contains 2^d variables ϕ_α. But the interest of Theorem 3.3 is that (3.8) may take place for $n \ll d$. For instance, in most random instances of Max-Cut problems with $d = 11$, we observe $f^* = \tau_2$ (and even $f^* = \tau_1$ in several cases).

3.3. *A global optimality condition for polynomial optimization*

Theorem 3.3(a) guarantees that asymptotically as n increases, we recover the global optimum f^*, and moreover, by Theorem 3.3(b), *finite* convergence takes place whenever the so-called *flatness condition* (3.8) holds at some degree n. In the latter case we can say more. Indeed, and remarkably, one can provide a global optimality condition for non-convex POPs, of the same flavour as the celebrated KKT optimality conditions for convex optimization, and under the same second-order sufficiency condition.

We first recall the well-known standard first-order necessary and second-order sufficient KKT optimality conditions in non-linear programming (NLP).

First-order necessary KKT optimality conditions. With S as in (2.7), let $\mathbf{x}^* \in S$ be a local minimizer of $\mathbf{P}$, and let $I(\mathbf{x}^*) := \{j \in \{1,\dots,m\} : g_j(\mathbf{x}^*) = 0\}$ be the set of *active* constraints at $\mathbf{x}^* \in S$. With $\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$, let

$$(\mathbf{x}^*)^\perp := \{\mathbf{u} \in \mathbb{S}^{d-1} : \langle \mathbf{u}, \nabla g_j(\mathbf{x}^*)\rangle = 0,\ \forall\, j \in I(\mathbf{x}^*)\}.$$

If the gradients $\nabla g_j(\mathbf{x}^*)$, $j \in I(\mathbf{x}^*)$, are linearly independent, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m_+$ such that

$$\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*) = 0, \quad \lambda_j^* g_j(\mathbf{x}^*) = 0,\ j = 1,\dots,m. \tag{3.11}$$

Moreover, *strict complementarity* holds if $\lambda_j^* > 0$ whenever $g_j(\mathbf{x}^*) = 0$. Next, observe that if in addition f and $-g_j$ are convex, then the Lagrangian

$$\mathbf{x} \mapsto L(\mathbf{x}) := f(\mathbf{x}) - f^* - \sum_{j=1}^m \lambda_j^* g_j(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d$$

is convex, non-negative, and satisfies $\nabla L(\mathbf{x}^*) = 0$. Hence $\mathbf{x}^*$ is also a global

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

minimizer of L on the whole space $\mathbb{R}^d$, and $L(\mathbf{x}) \geq L(\mathbf{x}^*) = 0$ for all $\mathbf{x} \in \mathbb{R}^d$. This is a very strong property of the convex case.

Second-order sufficient KKT optimality condition holds at a local minimizer $\mathbf{x}^* \in S$ *of* $\mathbf{P}$. If (i) (3.11) and strict complementarity hold at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, and (ii) in addition,

$$\left\langle \mathbf{u}, \left(\nabla^2 f(\mathbf{x}^*) - \sum_{j=1}^{m} \lambda_j^* \nabla^2 g_j(\mathbf{x}^*) \right) \mathbf{u} \right\rangle > 0 \quad \text{for all } \mathbf{u} \in (\mathbf{x}^*)^{\perp}. \tag{3.12}$$

If (3.11), strict complementarity and (3.12) hold at a global minimizer $\mathbf{x}^* \in S$ of $\mathbf{P}$, then we obtain a remarkable certificate of global optimality.

Theorem 3.5 (certificate of global optimality). With S as in (2.7), let $\mathbf{x}^* \in S$ be a global minimizer of $\mathbf{P}$, and assume the following.

(i) The gradients $\nabla g_j(\mathbf{x}^*)$, $j \in I(\mathbf{x}^*)$, are linearly independent (so that (3.11) holds for some $\boldsymbol{\lambda}^* \in \mathbb{R}_+^m$) and strict complementarity holds at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$.

(ii) Second-order sufficient condition (3.12) holds at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$.

Then there exists $n \in \mathbb{N}$ such that the SOS-strengthening (3.3) is exact, that is,

$$f(\mathbf{x}) - f^* = \sum_{j=0}^{m} \sigma_j^*(\mathbf{x})\, g_j(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{3.13}$$

$$\sigma_j^*(\mathbf{x}^*)\, g_j(\mathbf{x}^*) = 0, \quad j = 0, \ldots, m, \tag{3.14}$$

for some SOS polynomials $\sigma_j^* \in \Sigma[\mathbf{x}]_{n-d_j}$. Moreover, let $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_j) \in \mathbb{R}_+^m$ with $\hat{\lambda}_j := \sigma_j^*(\mathbf{x}^*)$, $j = 1, \ldots, m$. Then the pair $(\mathbf{x}^*, \hat{\boldsymbol{\lambda}}) \in S \times \mathbb{R}_+^m$ satisfies (3.11) and (3.12).

As an immediate consequence of Theorem 3.5, finite convergence of the Moment-SOS hierarchy takes place at degree n, i.e. $\tau_n^* = \tau_n = f^*$.

Remark 3.6. We claim that Theorem 3.5, which provides an algebraic certificate of global optimality, is the perfect analogue for *non-convex* polynomial optimization of the KKT optimality conditions for *convex* optimization. Indeed, if f and $-g_j$ are all convex, then any local optimizer $\mathbf{x}^* \in S$ is a global minimizer and then (3.11) implies

$$\mathbf{x} \mapsto L(\mathbf{x}) = f(\mathbf{x}) - f^* - \sum_{j=1}^{m} \lambda_j^* g_j(\mathbf{x}) \geq 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{3.15}$$

$$L(\mathbf{x}^*) = 0. \tag{3.16}$$

But this of course is valid because f and $-g_j$ are convex, a very specific case. In general a global minimizer $\mathbf{x}^* \in S$ is *not* a global minimizer of the Lagrangian L on $\mathbb{R}^d$.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

On the other hand, Theorem 3.5 states that if $\mathbf{x}^* \in S$ is a global minimizer, then under the standard second-order sufficient KKT optimality condition in NLP, $\mathbf{x}^*$ is also a global minimizer of the *extended Lagrangian*

$$\hat{L} := f - f^* - \sum_{j=1}^{m} \sigma_j^* \, g_j$$

over the whole space $\mathbb{R}^d$. Indeed,

$$\hat{L}(\mathbf{x}) = f(\mathbf{x}) - f^* - \sum_{j=1}^{m} \sigma_j^*(\mathbf{x}) \, g_j(\mathbf{x}) = \sigma_0^*(\mathbf{x}) \; (\geq 0) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{3.17}$$

$$\hat{L}(\mathbf{x}^*) = 0. \tag{3.18}$$

So in (3.17) the extended Lagrangian $\hat{L}$ looks like the standard Lagrangian L in (3.15) except that the scalar weight λ_j^* is now replaced by the SOS polynomial weight σ_j^*. Moreover, the scalar $\hat{\lambda}_j = \sigma_j^*(\mathbf{x}^*)$ is a standard Lagrange–KKT multiplier associated with the constraint $g_j \geq 0$ (like λ_j^* in (3.15)).

Interestingly, if the constraint $g_j \geq 0$ is not active at $\mathbf{x}^* \in S$ (i.e. $g_j(\mathbf{x}^*) > 0$), then $\hat{\lambda}_j = \sigma_j^*(\mathbf{x}^*) = 0$, but in general the SOS polynomial σ_j^* is *not* the trivial polynomial equal to zero. In fact, suppose that the constraint $g_j \geq 0$ is important even if it is not active at a global minimizer $\mathbf{x}^*$, meaning that if we remove that constraint in the definition (2.7) of S, then the new global minimum of the modified problem $\mathbf{P}$ is strictly smaller than f^*. Then, quite remarkably, $\sigma_j^* \neq 0$. In other words, a non-trivial SOS multiplier σ_j^* in Putinar's certificate of global optimality (3.13) identifies $g_j \geq 0$ as an important constraint, even if it is not active at global minimizers.

3.4. Genericity

In view of the remarkable form of Theorem 3.5, we may wonder how 'generic' the results of Theorem 3.5 might be. It turns out that Theorem 3.5 holds generically in a rigorous sense.

More precisely, let $m \in \mathbb{N}$ and $r_j \in \mathbb{N}$, $j = 0, \ldots, m$, be fixed, and consider the family of POPs whose (possibly empty) feasible set $S \subset \mathbb{R}^d$ is as in (2.7) for some polynomials $g_j \in \mathbb{R}[\mathbf{x}]_{r_j}$, and its criterion is some polynomial $f \in \mathbb{R}[\mathbf{x}]_{r_0}$.

Recall that $s(t) := \binom{d+t}{t}$. So a vector $\boldsymbol{\theta} \in \mathbb{R}^a$ with $a := \sum_{j=0}^{m} s(r_j)$ completely specifies an instance $\mathbf{P}(\boldsymbol{\theta})$ of such a problem $\mathbf{P}$. The next result is due to Nie (2014).

Theorem 3.7. There exists an integer L and finitely many real polynomials $\varphi_1, \ldots, \varphi_L \in \mathbb{R}[\boldsymbol{\theta}]$ in the coefficients $\boldsymbol{\theta}$ of the polynomials $f, g_1, \ldots, g_m$, such that if $\varphi_\ell(\boldsymbol{\theta}) \neq 0$ for all $\ell = 1, \ldots, L$, then (3.11), strict complementarity and second-order sufficient KKT optimality condition (3.12) hold at any global minimizer of problem $\mathbf{P}(\boldsymbol{\theta})$.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

As a result, there exists $n \in \mathbb{N}$ such that (3.13) and (3.14) hold at every global minimizer $\mathbf{x}^* \in S$ of $\mathbf{P}(\boldsymbol{\theta})$, that is, finite convergence of the Moment-SOS hierarchy is generic.

3.5. The convex case

In this section we consider the particular case when **P** is a convex program,[5] that is, when the polynomials f and $-g_j$ are all convex, and thus the set S in (2.7) is convex. This class of problems is very important as they are considered 'easy' or at least 'easier' than non-convex problems. Indeed, as any local optimum of **P** is a global optimum, then **P** can be solved by several powerful local optimization algorithms.

So as the Moment-SOS hierarchy is able to solve difficult non-convex problems **P**, a natural question is: How does the Moment-SOS hierarchy behave when **P** is a convex program?

The reason why such a question is relevant is because if the Moment-SOS hierarchy were not efficient in solving a convex problem, then one might raise reasonable doubts as to its efficiency in solving more difficult problems!

SOS-convex programs. Let us first consider the class of SOS-convex polynomials.

Definition 3.8. A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is SOS-convex if its Hessian $\nabla^2 f$ is an SOS-matrix polynomial, that is, $\nabla^2 f = L L^\top$ for some real matrix polynomial $L \in \mathbb{R}[\mathbf{x}]^{d\times s}$ (for some integer s). In particular, every SOS-convex polynomial is convex and all quadratic convex polynomials are SOS-convex.

We have the following characterizations of SOS-convexity.

Theorem 3.9. Let $f \in \mathbb{R}[\mathbf{x}]$. The following four propositions are equivalent:

(i) f is SOS-convex,

(ii) $\nabla^2 f$ is SOS,

(iii) $(\mathbf{x}, \mathbf{y}) \mapsto f(\mathbf{x})/2 + f(\mathbf{y})/2 - f((\mathbf{x}+\mathbf{y})/2)$ is SOS,

(iv) $(\mathbf{x}, \mathbf{y}) \mapsto f(\mathbf{x}) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), (\mathbf{x}-\mathbf{y}\rangle$ is SOS.

Notice that if f and $-g_j$ are convex, then necessarily their degree is either one or even. Importantly, SOS-convexity can be checked numerically by solving a semidefinite program (e.g. following Theorem 3.9(iii) and Section 2.2). The next result states that the Moment-SOS hierarchy somehow 'recognizes' easy SOS-convex problems.

[5] The set S in (2.7) may be convex even if the $-g_j$ are not convex (e.g. they can be quasi-convex). *Convex programming* usually refers to the case where f and the $-g_j$ are all convex.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Theorem 3.10. Let S be as in (2.7) and let Slater's condition hold (i.e. there exists $\mathbf{x}_0 \in S$ such that $g_j(\mathbf{x}_0) > 0$ for all j). If f and $-g_j$ are all SOS-convex, then with $n' := \max[\deg(f)/2, \max_j[\deg(g_j)/2]]$,

$$f - f^* = \sigma_0^* + \sum_{j=1}^{m} \lambda_j^* g_j \tag{3.19}$$

for some scalars $\lambda_j^* \geq 0$ and some $\sigma_0^* \in \Sigma[\mathbf{x}]_{n'}$. In addition,

$$f^* = \min_{\boldsymbol{\phi}}\{\phi(f)\colon \phi(\mathbf{1}) = 1;\ \mathbf{M}_{n'}(\boldsymbol{\phi}) \succeq 0;\ \phi(g_j) \geq 0,\ j = 1, \dots, m\}. \tag{3.20}$$

Moreover, $\phi^*(x_i) = x_i^*$, for all $i = 1, \dots, d$, where $\boldsymbol{\phi}^*$ is an optimal solution of (3.20), and $\mathbf{x}^* \in S$ is a local (hence global) minimizer of $\mathbf{P}$.

It is also fairly straightforward to check that in (3.19), $\boldsymbol{\lambda}^* = (\lambda_j^*)_{1 \leq j \leq m}$ are Lagrange–KKT multipliers at an optimal solution $\mathbf{x}^* \in S$ of $\mathbf{P}$.

Next, observe that the semidefinite Moment-relaxation (3.20) is a particular case of (3.5) where the constraints $\mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \succeq 0$ are replaced by the simpler $\mathbf{M}_0(g_j \cdot \boldsymbol{\phi}) \succeq 0$ (i.e. the scalar linear inequality constraint $\phi(g_j) \geq 0$).

However, if we were unaware that f and $-g_j$ were SOS-convex and we solved (3.5) as a general POP, then we would still obtain $f^* = \tau_{n_0}$, that is, the first Moment-relaxation of the hierarchy would be exact. In other words, the Moment-SOS hierarchy has recognized that $\mathbf{P}$ was a convex (easy) problem.

The reason why the Moment-relaxation (3.5) can be replaced by the simpler (3.20) is because linear functionals $\phi \in \mathbb{R}[\mathbf{x}]_{2n}^*$ such that $\mathbf{M}_n(\boldsymbol{\phi}) \succeq 0$ have a nice property when acting on SOS-convex polynomials.

Lemma 3.11 (Jensen's inequality for linear functionals). Let $\phi \in \mathbb{R}[\mathbf{x}]_{2n}^*$ be such that $\mathbf{M}_n(\boldsymbol{\phi}) \succeq 0$, $\phi(\mathbf{1}) = 1$, and let $\mathbf{x}^* := (\phi(x_1), \dots, \phi(x_d)) \in \mathbb{R}^d$. Then

$$\phi(f) \geq f(\mathbf{x}^*) \quad \text{for every SOS-convex polynomial } f \in \mathbb{R}[\mathbf{x}]_{2n}. \tag{3.21}$$

So let $\boldsymbol{\phi}^*$ be an optimal solution of

$$\tau_n' = \min_{\boldsymbol{\phi}}\{\phi(f)\colon \phi(\mathbf{1}) = 1;\ \mathbf{M}_{n'}(\boldsymbol{\phi}) \succeq 0;\ \phi(g_j) \geq 0,\ j = 1, \dots, m\}.$$

Of course, $\tau_n' \leq f^*$, as τ_n' is the optimal value of a relaxation of $\mathbf{P}$. As f and $-g_j$ are SOS-convex, and with $\mathbf{x}^* := (\phi^*(x_1), \dots, \phi^*(x_d)) \in \mathbb{R}^d$,

$$\tau_n' = \phi^*(f) \geq f(\mathbf{x}^*), \quad 0 \leq \phi^*(g_j) \leq g_j(\mathbf{x}^*), \quad j = 1, \dots, m,$$

which implies $\mathbf{x}^* \in S$ and $f(\mathbf{x}^*) \leq \tau_n' \leq f^*$, so that $\mathbf{x}^*$ is a global minimizer of $\mathbf{P}$.

General convex POPs. In the more general case of convex POPs, we also obtain finite convergence under some strict convexity assumption at every global minimizer $\mathbf{x}^* \in S$.

Theorem 3.12. With $S \subset \mathbb{R}^d$ as in (2.7), assume that $Q(g)$ is Archimedean, Slater's condition holds, and f and $-g_j$ are convex, $j = 1, \ldots, m$. If $\nabla^2 f(\mathbf{x}^*) \succ 0$ at every global minimizer $\mathbf{x}^* \in S$ (assumed to be finitely many), then finite convergence takes place, that is, the Moment-relaxation (3.5) of **P** is exact at some degree n. Moreover, the SOS-strengthening (3.3) of **P** is also exact, and both (3.5) and (3.3) have an optimal solution $\boldsymbol{\phi}^n$ and $(\lambda^*, \sigma_0^*, \ldots, \sigma_m^*)$, respectively.

So again without specifying that **P** is convex, the Moment-SOS hierarchy will converge in finitely many steps. However, in contrast to Theorem 3.9, in Theorem 3.12 we do not specify at which step n finite convergence takes place.

3.6. *General versus* ad hoc

We would like to emphasize that Theorem 3.5 is a fairly general global optimality condition that holds *generically* for POPs, hence with non-convex criterion and non-convex (and possibly disconnected) feasible sets S, and even with mixed-integer variables. The only requirement is to be able to translate all constraints of the problem into polynomial inequality and equality constraints.

Usually generality is at the price of reduced efficiency, and the usual algorithmic practice of optimization suggests using *ad hoc* algorithms, i.e. algorithms tailored to the type of problem to be solved. Indeed, for instance, if $x_i \in \{0, 1\}$, $i \in I$, for some I, it is not a good idea to model this constraint with the equality constraints $x_i^2 - x_i = 0$, $i \in I$, and then use standard first-order or second-order methods to obtain an (only local) optimum. We typically use Branch and Bound (or Branch and Cut) methods.

Remarkably, the Moment-SOS hierarchy does not suffer from its generality in just describing any POP by a set of polynomial inequality and equality constraints. (Of course, some descriptions may be more interesting than others.) Indeed, for instance, for SOS-convex programs and in particular convex quadratically constrained quadratic programs (convex QCQPs), Theorem 3.9 ensures that finite convergence takes place at the first step of the hierarchy, without the need to specify that the POP is SOS-convex. Similarly, if f and $-g_j$ are all convex, and $\nabla^2 f(\mathbf{x}) \succ 0$ at all global minimizers $\mathbf{x} \in S$, then finite convergence also takes place.

Of course, again, we do not claim that the Moment-SOS hierarchy is the most efficient algorithm for solving such convex problems, and indeed other efficient algorithms exist. But this remark simply emphasizes that, in some way, the Moment-SOS hierarchy recognizes easy problems (convex programs are usually considered easier to solve) as finite convergence takes place quickly. On the other hand, the Moment-SOS hierarchy has also been recognized by the theoretical computer science research community as a meta-algorithm which provides the best lower bounds for many combinatorial optimization problems, and in particular problems with $\{0, 1\}$ (or $\{-1, 1\}$) variables such as Max-Cut and its variants, which are notoriously difficult NP-hard problems. It is now considered an important tool for proving/disproving Khot's celebrated Unique Games Conjecture.

Table 3.1. Rates of convergence for the hierarchy of lower bounds (Slot 2022).

S	Error $f^* - \tau_n$	Certificate	Reference
Archimedean	$O(1/n^c)$	Putinar	Baldi and Mourrain (2022)
Compact	$O(1/n^c)$	Schmüdgen	Schweighofer (2005)
$\mathbb{S}^{d-1}$	$O(1/n^2)$	Putinar	Fang and Fawzi (2021)
$\mathbf{B}^d$	$O(1/n^2)$	Putinar	Slot (2022)
$[-1,1]^d$	$O(1/n^2)$	Schmüdgen	Laurent and Slot (2023)
Δ^d	$O(1/n^2)$	Schmüdgen	Slot (2022)

3.7. Rates of convergence

In this section we provide rates of convergence for the Moment-SOS hierarchy of lower bounds on general compact basic semi-algebraic sets. Those rates have been refined for specific sets like the unit sphere $\mathbb{S}^{d-1}$, the unit ball $\mathbf{B}^d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|^2 \leq 1\}$, the box $[-1,1]^d$, and the simplex $\Delta^d = \{\mathbf{x} \in \mathbb{R}^d_+ : 1 - \mathbf{e}^\top \mathbf{x} \geq 0\}$.

Table 3.1 is taken from Slot (2022). The first column is related to the set $S \subset \mathbb{R}^d$, and 'Archimedean' means that the quadratic module $Q(g)$ in (2.5) associated with S, is Archimedean (an algebraic certificate of its compactness used in Putinar's Positivstellensatz). The third column, 'Certificate', specifies whether in the Moment-SOS hierarchy we use Putinar's certificate (Theorem 2.2) or the more costly Schmüdgen's certificate (Theorem 2.3) for the semidefinite relaxation (3.3). In the notation $O(1/n^c)$, n is the order (or degree) of the semidefinite relaxations (3.3)–(3.5), and c is some positive constant. The rate $O(1/n^c)$ means that $f^* - \tau_n \leq M/n^c$ for some constant $M > 0$, where f^* is the global minimum and $\tau_n\,(\leq f^*)$ is as in (3.5). Finally, for optimization of trigonometric polynomials (and hence for POPs on the box $[0,1]^d$ as well) and under some condition on global minimizers (isolated and with positive definite Hessian), an exponential rate of convergence has been provided in Bach and Rudi (2023). This shows that beyond general results such as those in Table 3.1, there is hope for even faster rate convergence, at the price of some additional conditions on the minimizers or the set S.

3.8. Handling sparsity

As already mentioned, in its canonical form (3.3)–(3.5), the Moment-SOS hierarchy is limited to problems **P** of modest dimension, even though for fixed dimension d the size parameters of each Moment-relaxation (3.5) are polynomial in the degree n. This is because, (3.5) being a semidefinite program, efficient algorithms based on interior-point methods are still very time-consuming. Fortunately, practice reveals that Moment-relaxations (3.5) of low degree n already provide tight lower bounds on f^*, and are sometimes *exact*. In addition, large-scale problems **P** usually exhibit some *sparsity* pattern or symmetries. For instance, as is typical for applications in large dimension d, (i) each constraint $g_j \geq 0$ in (2.7) sees only some small subset of variables $\{x_i : i \in I_k\}$ with $I_k \subset \{1, \dots, d\}$, and (ii) the criterion f of **P** is very often a sum $\sum_k f_k$ of (low degree) polynomials f_k, where each polynomial f_k only sees variables $\{x_i : i \in I_k\}$. This type of sparsity is called *correlative sparsity*. Also, another type of sparsity called *term sparsity* occurs when all polynomials f and g_j in the description (2.7) of **P** contain a few monomials only. It turns out that correlative and term sparsity can be exploited so as to yield a new sparsity-adapted Moment-SOS hierarchy of semidefinite relaxations of **P**. These two types of sparsity can even be combined for further efficiency; see e.g. Wang, Magron, Lasserre and Ngoc Hoang Anh Mai (2022) and Magron and Wang (2023).

For the sake of completeness, we now briefly describe how correlative sparsity allows us to define an appropriate sparsity-adapted Moment-SOS hierarchy that can handle large-scale POPs.

Assumption 3.13. With **P** as in (1.1) with $S \subset \mathbb{R}^d$ as in (2.7):

- $I_0 := \{1, \dots, d\} = \bigcup_{k=1}^p I_k$ (with possible overlaps),
- $\mathbb{R}[\mathbf{x}; I_k]$ is the ring of polynomials in the variables $\{x_i : i \in I_k\}$,
- $f = \sum_{k=1}^p f_k$ with $f_k \in \mathbb{R}[\mathbf{x}; I_k]$, $k = 1, \dots, p$,
- for each $j = 1, \dots, m$, $g_j \in \mathbb{R}[\mathbf{x}; I_k]$ for some $k \in \{1, \dots, p\}$, so let

$$J_k := \{j : g_j \in \mathbb{R}[\mathbf{x}; I_k]\}, \quad k = 1, \dots, p.$$

Of course, as we will see next, Assumption 3.13 is interesting when the cardinal $\#I_k$ of I_k is small for every $k = 1, \dots, p$.

Observe that if $g \in \mathbb{R}[\mathbf{x}; I_k]$, then in the expansion $g(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^d} g_\alpha \mathbf{x}^\alpha$, $\alpha_i = 0$ if $i \notin I_k$, whenever $g_\alpha \neq 0$. So let

$$\mathbb{N}^{(k)} := \{\boldsymbol{\alpha} \in \mathbb{N}^d : \alpha_i = 0, \ \forall i \notin I_k\}, \quad k = 1, \dots, p,$$
$$\mathbb{N}^{(k)}_n := \left\{\boldsymbol{\alpha} \in \mathbb{N}^{(k)} : \sum_i \alpha_i \leq n\right\}, \quad k = 1, \dots, p.$$

Next, given $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha \in \mathbb{N}^d}$, define $\mathbf{M}_n(\boldsymbol{\phi}; I_k)$ to be the submatrix of $\mathbf{M}_n(\boldsymbol{\phi})$ whose rows and columns are associated with monomials $(\mathbf{x}^\alpha)$, $\boldsymbol{\alpha} \in \mathbb{N}^{(k)}_n$. Similarly, when $j \in J_k$, the localizing matrix $\mathbf{M}_n(g_j \cdot \boldsymbol{\phi}; I_k)$ is the submatrix of $\mathbf{M}_n(g_j \cdot \boldsymbol{\phi})$ whose rows and columns are associated with monomials $(\mathbf{x}^\alpha)$, $\boldsymbol{\alpha} \in \mathbb{N}^{(k)}_{n-d_j}$.

When Assumption 3.13 holds, it is quite natural to define Moment-relaxations

$$\tau_n^{\mathrm{sparse}} = \inf_{\boldsymbol{\phi}=(\phi_\alpha)} \Big\{ \sum_{k=1}^{p} \phi(f_k) \colon \phi_0 = 1; \\ \mathbf{M}_n(\boldsymbol{\phi}; I_k) \succeq 0,\ k = 1,\dots,p; \\ \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}; I_k) \succeq 0,\ \forall j \in J_k;\ k = 1,\dots,p \Big\}. \tag{3.22}$$

The reason why (3.22) is appealing is because the size of the matrix $\mathbf{M}_n(\boldsymbol{\phi}; I_k)$ is $s_k(n) := \binom{\#I_k+n}{n}$ while that of $\mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi})$ is $s(n-d_j)$. Therefore, if $\#I_k \ll d$, then $s_k(n) \ll s(n)$ and $s_k(n-d_j) \ll s(n-d_j)$. So for a semidefinite solver it is always definitely better to have several (and perhaps even many) small-size matrices rather than a single large-size matrix constrained to be positive semidefinite.

It is quite straightforward to obtain that $\tau_n^{\mathrm{sparse}} \le f^*$ for all $n \ge n_0$. Indeed, $\tau_n^{\mathrm{sparse}} \le \tau_n \le f^*$ (where τ_n is the optimal value of the degree-n Moment-relaxation (3.5)). Moreover, the sequence $(\tau_n^{\mathrm{sparse}})_{n\ge k_0}$ is monotone non-decreasing and, being bounded above, converges to some $\gamma \le f^*$.

Again assume (possibly after scaling) that the quadratic polynomial $1 - \|\mathbf{x}\|^2$ is in the quadratic module $Q_1(g)$, and therefore we can and do add the p redundant constraints

$$1 - \sum_{i\in I_k} x_i^2 \ge 0, \quad k = 1,\dots,p, \tag{3.23}$$

to the definition of S.

Theorem 3.14 (Lasserre 2006). Let $S \subset \mathbb{R}^d$ be as in (2.7) with constraints (3.23) in its definition (2.7), and consider the hierarchy of semidefinite relaxations (3.22) with optimal value τ_n^{sparse}. Then $\tau_n^{\mathrm{sparse}} \uparrow \gamma \le f^*$, as n increases. Moreover, if for every $k = 2,\dots,p$,

$$I_k \cap \Bigg(\bigcup_{j=1}^{k-1} I_j \Bigg) \subseteq I_\ell, \tag{3.24}$$

for some $\ell \in \{1,\dots,k-1\}$, then $\gamma = f^*$.

The condition (3.24) is called the *running intersection property*. It has the following important property: suppose that we are given p probability measures ϕ_j on $\mathbb{R}^{\#I_j}$, $j = 1,\dots,p$, which are compatible, that is, such that for all pairs (i,j) with $I_i \cap I_j \neq \emptyset$,

$$\int \mathbf{x}^\alpha\, \mathrm{d}\phi_i(\mathbf{x} \in I_i \cap I_j) = \int \mathbf{x}^\alpha\, \mathrm{d}\phi_j(\mathbf{x} \in I_i \cap I_j) \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^{\#I_i \cap I_j}.$$

If (3.24) holds then there exists a probability measure ϕ on $\mathbb{R}^d$ such that

$$\int \mathbf{x}^\alpha\, \mathrm{d}\phi = \int \mathbf{x}^\alpha\, \mathrm{d}\phi_j \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^{\#I_j}, \quad j = 1,\dots,p.$$

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

That is, given *local* measures ϕ_j on $\mathbb{R}^{\#I_j}$ that are compatible, we can reconstruct a *global* measure on $\mathbb{R}^d$ whose marginal on $\mathbb{R}^{\#I_j}$ is ϕ_j, for all $j = 1, \ldots, p$. Therefore the local information provided by the ϕ_j corresponds to partial (but consistent) knowledge of global information that we do not necessarily need.

On the dual side of positive polynomials we have the following sparse version of Putinar's Positivstellensatz.

Theorem 3.15 (sparse Positivstellensatz (Lasserre 2006)). Let $S \subset \mathbb{R}^d$ be as in (2.7) with constraints (3.23) in its definition (2.7), and let $f = \sum_{k=1}^{p} f_k$ with $f_k \in \mathbb{R}[\mathbf{x}; I_k]$, $k = 1, \ldots, p$. If (3.24) holds and $f > 0$ on S, then

$$f(\mathbf{x}) = \sum_{k=1}^{p} \left(\sigma_{0,k}(\mathbf{x}) + \sum_{j \in J_k} \sigma_{j,k}(\mathbf{x})\, g_j(\mathbf{x}) \right) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{3.25}$$

for some SOS polynomials $\sigma_{0,k}, \sigma_{j,k} \in \Sigma[\mathbf{x}, I_k]$, $j \in J_k$, $k = 1, \ldots, p$.

Theorem 3.15 provides a sparsity-adapted certificate of positivity *à la* Putinar where the SOS weight σ_j associated with a constraint $g_j \geq 0$, $j \in J_k$, 'sees' only the variables $\{x_i : i \in I_k\}$.

To cite a few examples, such sparsity-adapted semidefinite relaxations have been implemented for solving the optimal power flow (OPF) problem in the management of large-scale electricity networks (Molzhan and Hiskens 2015, Molzahn and Josz 2018), in geometric perception (Yang and Carlone 2023), robotics (Rosen, Carlone, Bandera and Leonard 2019) and sensor network localization (Nie 2009), as well as in Nie and Demmel (2008).

3.9. Notes and sources

Section 3 is mainly based on Lasserre (2009*b*, 2015), where the reader can find all proofs (or references to papers with proofs).

Section 3.1. The Moment-SOS hierarchy was first proposed in Lasserre (2000, 2001).

Sections 3.3–3.4. Theorems 3.5 and 3.7 are due to Nie (2014). Further, Nie (2013) shows that the flatness condition (3.8) at an optimal solution of the Moment-relaxation (3.5) also holds generically. A refinement of these results is provided in Baldi and Mourrain (2022). Those results are important as they guarantee that the Moment-SOS hierarchy has finite convergence, generically (in the sense of Theorem 3.7), and that one may extract global minimizers from an optimal solution of the semidefinite relaxation (3.5).

Sections 3.5–3.6. These are mainly taken from Lasserre (2015, Chapter 13). See also Lasserre (2009*a*) and de Klerk and Laurent (2011).

Section 3.7. This is essentially based on Slot (2022).

Section 3.8. Correlative sparsity was first proposed as a heuristic in Kojima, Kim and Maramatsu (2005) and Waki, Kim, Kojima and Maramatsu (2006), while its proof of convergence was provided in Lasserre (2006). Term sparsity was first described in Wang, Li and Xia (2019) and further exploited in the TSSOS hierarchy of Wang, Magron and Lasserre (2021). Finally, the combination of correlative and term sparsity which yields the CS-TSSOS hierarchy is described in Wang *et al.* (2022). The interested reader is referred to the book by Magron and Wang (2023), which, among other things, describes various forms of sparsity and associated Moment-relaxations based on appropriate cones of positive polynomials.

Finally let us mention that a lower bound $\tau_n \leq f^*$, obtained by the Moment-SOS hierarchy even at low degree n, can also be useful to 'gauge' how far from f^* is the value $f(\hat{\mathbf{x}})$ of a feasible solution $\hat{\mathbf{x}} \in S$ obtained by some numerical (local) optimization algorithm. Indeed, if $f(\hat{\mathbf{x}}) - \tau_n$ is not too large, then in some way τ_n certifies that the local optimization algorithm has produced a good feasible solution.

4. The Moment-LP hierarchy

As seen in Section 3, the Moment-SOS hierarchy is based on the use of Putinar's certificate of positivity (2.9) and its convergence relies on Theorem 2.2. We next provide a hierarchy of *LP relaxations* whose associated sequence of optimal values also converges to the global optimum from below. Similarly as for the Moment-SOS hierarchy, the Moment-LP hierarchy is also based on a positivity certificate, namely that in (2.13), and its convergence relies on Theorem 2.5.

Assumption 4.1. With $S \subset \mathbb{R}^d$ as in (2.7), assume that S is compact, $0 \leq g_j \leq 1$ on S, for every $j = 1, \ldots, m$, and the polynomials $\{\mathbf{1}, g_1, \ldots, g_m\}$ generate $\mathbb{R}[\mathbf{x}]$.

As S is compact, one can always rescale the g_j (and possibly add redundant constraints) to make the new definition of S satisfy Assumption 4.1. For more details the interested reader is referred to Lasserre (2009*b*).

Next, with the same notation $\mathbf{g}$ and $\mathbf{1} - \mathbf{g}$ as in (2.13), and $n \in \mathbb{N}$, introduce the following linear program (LP):

$$\rho_n = \min_{\boldsymbol{\phi}} \{\phi(f) \colon \phi(\mathbf{1}) = 1;\ \phi(\mathbf{g}^{\alpha}\,(\mathbf{1} - \mathbf{g})^{\beta}) \geq 0,\ (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{N}^{2m}_n\}, \tag{4.1}$$

where $\boldsymbol{\phi} = (\phi_{\gamma})_{\boldsymbol{\gamma} \in \mathbb{N}^d_{s_n}}$ with $s_n := \max_{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathbb{N}^{2m}_n} \deg(\mathbf{g}^{\alpha}(\mathbf{1} - \mathbf{g})^{\boldsymbol{\beta}})$.

By its very nature (4.1) is a linear program, and it is a relaxation of $\mathbf{P}$ because the constraints in (4.1) are only *necessary* conditions on $\boldsymbol{\phi}$ to be moments of a probability measure supported on S; see Theorem 2.5. The dual of (4.1) is the linear program

$$\rho^*_n = \max_{c_{\boldsymbol{\alpha\beta}} \geq 0, \lambda} \left\{ \lambda \colon f - \lambda = \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathbb{N}^{2m}_n} c_{\boldsymbol{\alpha\beta}}\, \mathbf{g}^{\alpha}\,(\mathbf{1} - \mathbf{g})^{\boldsymbol{\beta}} \right\}. \tag{4.2}$$

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

In a similar manner, just as (3.3) was an SOS-strengthening of **P** in (1.3), the LP (4.2) is now an LP-strengthening of **P** in (1.3). Of course, from duality for linear programs, $\rho_n = \rho_n^*$ for all n.

Example 4.2. To better visualize the LP (4.1), consider the toy example where $S = [0, 1] = \{x : x \geq 0;\ (1 - x) \geq 0\} \subset \mathbb{R}$. Then, for $n = 2$, $\boldsymbol{\phi} = (\phi_j)_{0\leq j\leq 2}$, and

$$\phi(1) = \phi_0, \quad \phi(x) = \phi_1, \quad \phi(1-x) = \phi_0 - \phi_1, \quad \phi(x^2) = \phi_2,$$
$$\phi(x(1-x)) = \phi_1 - \phi_2, \quad \phi((1-x)^2) = \phi_0 - 2\phi_1 + \phi_2,$$

so that with $f \in \mathbb{R}[x]_2$, $x \mapsto f(x) = \sum_{k=0}^{2} f_k\, x^k$,

$$\rho_2 = \min_{\boldsymbol{\phi}} \left\{ \sum_{k=0}^{2} \phi_k\, f_k : \phi_0 = 1;\ \phi_1 \geq 0;\ \phi_0 - \phi_1 \geq 0; \right.$$
$$\left. \phi_1 - \phi_2 \geq 0;\ \phi_0 - 2\phi_1 + \phi_2 \geq 0 \right\}.$$

Similarly,

$$\rho_2^* = \max_{\mathbf{c}\geq 0,\lambda} \{\lambda :\ f(x) - \lambda = c_{00} + c_{10}\, x + c_{01}(1-x)$$
$$+ c_{20}\, x^2 + c_{11} x(1-x) + c_{02}(1-x)^2,\ \forall x \in \mathbb{R}\},$$

or equivalently

$$\rho_2^* = \max_{\mathbf{c}\geq 0,\lambda} \{\lambda :\ f_0 - \lambda = c_{00} + c_{01};$$
$$f_1 = c_{10} - c_{01} + c_{11} + c_{02};\ f_2 = c_{20} - c_{11} + c_{02}\}.$$

Equality constraints are treated as for the Moment-SOS hierarchy. For instance, in (4.1), a Boolean constraint $x_i^2 = x_i$ of **P** translates into the moment equality constraints $\phi(x_i^k) = \phi(x_i)$ for all $k \leq n$.

Theorem 4.3 (Lasserre 2009*b*). With S as in (2.7), let Assumption 4.1 hold. Then, as n increases, the sequences $(\rho_n)_{n\in\mathbb{N}}$ and $(\rho_n^*)_{n\in\mathbb{N}}$ are monotone non-decreasing and converge to the global minimum f^* of **P**.

4.1. The case of a convex polytope

We now assume that $S \subset \mathbb{R}^d$ is a convex polytope (with non-empty interior), that is, for each $j = 1, \ldots, m$, $g_j \in \mathbb{R}[\mathbf{x}]_1$ (g_j is a polynomial of degree 1). In this case Theorem 2.5 takes the following specific form.

Theorem 4.4 (Handelman 1988). Let $S \subset \mathbb{R}^d$ be as in (2.7) with non-empty interior and with all g_j of degree 1, and assume that S is compact (hence S is a convex polytope). If $f \in \mathbb{R}[\mathbf{x}]$ is positive on S, then there exists $n \in \mathbb{N}$ and a

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

non-negative vector $\mathbf{c} = (c_\alpha)_{\alpha\in\mathbb{N}^m_n}$, such that

$$f = \sum_{\alpha\in\mathbb{N}^m_n} c_\alpha\,\mathbf{g}^\alpha. \tag{4.3}$$

So the obvious analogue of (4.1) for a convex polytope now reads

$$\rho_n = \min_{\boldsymbol{\phi}} \{\phi(f)\colon \phi(\mathbf{1}) = 1;\ \phi(\mathbf{g}^\alpha) \geq 0,\ \alpha \in \mathbb{N}^m_n\} \tag{4.4}$$

where $\boldsymbol{\phi} = (\phi_\gamma)_{\gamma\in\mathbb{N}^d_{s_n}}$ with $s_n := \max_{\alpha\in\mathbb{N}^m_n} \deg(\mathbf{g}^\alpha)$. The dual of (4.4) reads

$$\rho^*_n = \max_{c_\alpha\geq 0,\lambda} \biggl\{\lambda\colon f - \lambda = \sum_{\alpha\in\mathbb{N}^m_n} c_\alpha\,\mathbf{g}^\alpha\biggr\}. \tag{4.5}$$

So an analogue of Theorem 4.3 reads as follows.

Theorem 4.5. Let $S \subset \mathbb{R}^d$ be as in (2.7) with non-empty interior and with all g_j of degree 1, and assume that S is compact (hence S is a convex polytope). Let ρ_n (resp. ρ^*_n) be as in (4.4) (resp. (4.5)). Then, as n increases, both sequences $(\rho_n)_{n\in\mathbb{N}}$ and $(\rho^*_n)_{n\in\mathbb{N}}$ are monotone non-decreasing and converge to the global minimum f^* of **P**.

On $0/1$ *discrete problems and RLT.* Consider discrete optimization problems **P** $\min\{f(\mathbf{x})\colon \mathbf{x}\in S\}$ for which the set of feasible solutions is of the form

$$\begin{aligned} S &= \{\mathbf{x}\in\{0,1\}^d\colon g_j(\mathbf{x})\geq 0,\ j=1,\ldots,m\}\\ &= \{\mathbf{x}\in\mathbb{R}^d\colon g_j(\mathbf{x})\geq 0,\ j=1,\ldots,m;\ x_i^2 - x_i = 0,\ i=1,\ldots,d\}. \end{aligned}$$

This formulation includes many combinatorial optimization problems, including the celebrated Max-Cut problem and its variants on $\{-1,1\}^d$ (after a simple linear transformation). The so-called reformation–linearization technique (RLT) (Sherali and Adams 1990, 1999) solves **P** when f and the g_j are all linear. In the RLT we 'lift' **P** to a space of higher dimension. Namely, with $t \leq d$ fixed, we proceed as follows.

- We define *order-t bound-factor* constraints

  $$A_{J_1,J_2}(\mathbf{x}) := \prod_{i\in J_1} x_i \prod_{j\in J_2} (1-x_j) \geq 0 \quad \text{for all } (J_1,J_2),$$

  where $J_1, J_2 \subseteq \{1,\ldots,d\}$, $J_1\cap J_2 = \emptyset$ and $|J_1\cup J_2| = t$.
- For every $j = 1,\ldots,m$, we use the additional *constraint factor-based restrictions*

  $$A_{J_1,J_2}\, g_j(\mathbf{x}) \geq 0 \quad \text{for all } (J_1,J_2).$$
- Then, in each such constraint, we replace every occurrence of the power x_i^k with x_i, and 'linearize' the resulting polynomial constraint, that is, every occurrence of the non-linear monomial $\prod_{i\in J} x_i$ is replaced by a variable y_J constrained to be non-negative.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

We end up with a linear program in a space of higher dimension $s(t+1)$ which is an LP relaxation of **P**. In the RLT construction of this LP relaxation (which ignores constraints of the form $\mathbf{g}^\alpha \geq 0$), we recognize a particular case of the Moment-LP relaxation (4.1) of **P** in the presence of Boolean constraints $x_i^2 = x_i$, for all $i = 1, \ldots, d$. So it is fair to say that RLT, the first systematic construction of a 'hierarchy' of LP relaxations for $0/1$ programs, was implicitly based on positivity certificates of the flavour (2.13), i.e. *à la* Krivine–Vasilescu.

4.2. Contrasting the Moment-SOS hierarchy with the Moment-LP hierarchy

At first glance we are tempted to favour the Moment-LP hierarchy because state-of-the-art LP solvers are very efficient and can solve potentially very large-scale (even huge) LPs, whereas in its canonical form (3.3)–(3.5) the Moment-SOS hierarchy is limited to POPs of modest dimension and small degree of relaxation n, unless some sparsity or symmetries can be exploited. Unfortunately the Moment-LP hierarchy has some serious drawbacks that also limit its application to problems of modest dimension. Indeed, except for discrete and linear POPs, finite convergence is impossible in general, even for convex problems!

However, for discrete problems with $0/1$ variables, the Moment-LP hierarchy can be combined with *ad hoc* heuristics. For instance, we may try to solve such $0/1$ problems with Branch and Bound methods where, at each node of the search tree, a lower bound associated with the node is computed by solving an appropriate Moment-LP relaxation of the discrete subproblem associated with the node in the Branch and Bound strategy.

Finite convergence is not possible in general. For clarity and simplicity of exposition, we illustrate this claim in the case where S is a convex polytope.

Proposition 4.6. Let $S \subset \mathbb{R}^d$ be a convex polytope and consider the LP-strengthening (4.5) of **P**. If **P** has finitely many global minimizers and (4.5) is exact for some degree n, then necessarily every global minimizer $\mathbf{x}^*$ is a vertex of S.

Proof. Let $0 \leq \mathbf{c}^* \neq 0$ be an optimal solution of the degree-n LP relaxation (4.5), and assume that (4.5) is exact, i.e. $\rho_n^* = f^*$. Then

$$f(\mathbf{x}) - f^* = \sum_{\alpha \in \mathbb{N}_n^m} c_\alpha^* \, g_1(\mathbf{x})^{\alpha_1} \cdots g_m(\mathbf{x})^{\alpha_m} \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \tag{4.6}$$

In particular, at every global minimizer $\mathbf{x}^* \in S$,

$$0 = f(\mathbf{x}^*) - f^* = \sum_{\alpha \in \mathbb{N}_n^m} c_\alpha^* \, g_1(\mathbf{x}^*)^{\alpha_1} \cdots g_m(\mathbf{x}^*)^{\alpha_m} \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \tag{4.7}$$

So assume that there exists a global minimizer $\mathbf{x}^* \in S$ which is not a vertex, and let $I(\mathbf{x}^*) := \{j \in \{1, \ldots, m\} \colon g_j(\mathbf{x}^*) = 0\}$ be the set of active constraints at $\mathbf{x}^* \in S$. Observe that $I(\mathbf{x}^*) \neq \emptyset$ because otherwise we would have $g_j(\mathbf{x}^*) > 0$ for all $j = 1, \ldots, m$, which in turn by (4.7) implies $\mathbf{c}^* = 0$, in contradiction to $\mathbf{c}^* \neq 0$. So

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

(4.7) already rules out the possibility of having a global minimizer in int(S). Next, from (4.7) we may infer

$$c^*_\alpha > 0 \Rightarrow \alpha_j > 0 \quad \text{for some } j \in I(\mathbf{x}^*). \tag{4.8}$$

As $\mathbf{x}^*$ is not a vertex and $\mathbf{P}$ has only finitely many global minimizers, the set $\{\mathbf{y} \in S\colon I(\mathbf{y}) = I(\mathbf{x}^*)\}$ contains a point $\hat{\mathbf{y}} \in S$ which is not a global minimizer, i.e. $f(\hat{\mathbf{y}}) > f^*$. But then (4.6) combined with (4.8) yields the contradiction

$$0 = \sum_{\alpha \in \mathbb{N}^m_n} c^*_\alpha \, g_1(\hat{\mathbf{y}})^{\alpha_1} \cdots g_m(\hat{\mathbf{y}})^{\alpha_m} = f(\mathbf{y}) - f^* > 0. \qquad \square$$

So Proposition 4.6 implies that for all problems $\mathbf{P}$ (on convex polytopes) with finitely many minimizers, a necessary condition for some degree-n LP-strengthening (4.5) to be exact is that every global minimizer is a vertex of S. In particular, this condition rules out most convex POPs on a polytope, with a non-linear criterion f as in general a local (hence global) minimizer is *not* a vertex of S. (However, if f is linear, i.e. if $\mathbf{P}$ is a linear program, then (4.5) is exact with $n = 1$.)

A similar conclusion is also valid for POPs on even more general basic semi-algebraic sets S and the LP-strengthening (4.2). Indeed, as in Proposition 4.6 and for the same reasons, if such a POP has finitely many minimizers, then a necessary condition for (4.2) to be exact at some degree n, is that for every global minimizer $\mathbf{x}^* \in S$, the set $\{\mathbf{y} \in S\colon I(\mathbf{y}) = I(\mathbf{x}^*)\}$ contains only global minimizers of $\mathbf{P}$. Such a condition is very restrictive and rules out most problems $\mathbf{P}$, in particular convex problems!

4.3. Notes and sources

Section 4 is mainly taken from Lasserre (2002*b*, 2015, Chapter 9). In Laurent (2003) the Moment-SOS hierarchy for 0/1 variables is described with specific notation proper to graph theory and is embedded in the family of *lift-and-project* hierarchies, which include the Lovász–Schrijver and Sherali–Adams hierarchies. In particular, Laurent (2003) shows that the Moment-SOS dominates the other lift-and-project hierarchies.

If LP-hierarchies are not efficient when used alone to solve optimization problems, they can still be useful when associated with other techniques of discrete optimization, for instance as in Aloise and Hansen (2011) when used in conjunction with Branch and Bound.

5. A Moment-SOS hierarchy of upper bounds

In this section we consider another (less known) Moment-SOS hierarchy which provides a monotone non-increasing sequence $(\kappa_n)_{n\in\mathbb{N}}$ of *upper bounds* on the global minimum f^* of $\mathbf{P}$ defined in (1.3). For each 'degree' n, the upper bound $\kappa_n \geq f^*$ is now computed by solving a very specific semidefinite program as it has only a single variable. In fact its dual reduces to computing the smallest generalized eigenvalue of a pair of moment matrices.

5.1. A first multivariate formulation

Consider problem **P** in (1.3) with feasible set $S \subset \mathbb{R}^d$ as in (2.7), and let μ be a Borel (reference) measure on S with $\text{supp}(\mu) = S$.

Assumption 5.1.

(i) The set $S \subset \mathbb{R}^d$ is compact with non-empty interior.

(ii) The vector of moments $\boldsymbol{\mu} = (\mu_\alpha)_{\alpha\in\mathbb{N}^d}$ is available in closed form, or can be computed efficiently.

Of course, in view of Assumption 5.1(ii), the set S has to be rather specific, and indeed typical sets of this type are the unit box $[-1, 1]^d$, the Euclidean unit ball $\mathbf{B}(0, 1) = \{\mathbf{x}: \|\mathbf{x}\| \le 1\}$, the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x}: \|\mathbf{x}\| = 1\}$, the canonical simplex $\Delta^d = \{\mathbf{x} \in \mathbb{R}^d_+: \mathbf{e}^\top\mathbf{x} \le 1\}$, the discrete hypercube $\{-1, 1\}^d$, and their image by an affine transformation. Even though such sets S are rather specific, the associated problems **P** cover many interesting NP-hard optimization problems.

Theorem 5.2. Let Assumption 5.1 hold, and with $n \in \mathbb{N}$ fixed, consider the semidefinite problems

$$\kappa_n = \inf_{\sigma\in\Sigma[\mathbf{x}]_n} \left\{ \int f\,\sigma\,\mathrm{d}\mu : \int \sigma\,\mathrm{d}\mu = 1 \right\}, \quad n \in \mathbb{N}, \tag{5.1}$$

$$\kappa_n^* = \sup_\lambda \{\lambda : \lambda\,\mathbf{M}_n(\mu) \preceq \mathbf{M}_n(f\cdot\mu)\}, \quad n \in \mathbb{N}. \tag{5.2}$$

Then

$$\kappa_n = \kappa_n^*, \quad f^* \le \kappa_{n+1} \le \kappa_n \quad \text{for all } n \in \mathbb{N}, \quad \lim_{n\to\infty} \kappa_n = f^*. \tag{5.3}$$

Crucial to the proof of convergence, which can be found in Lasserre (2011, Theorem 4.2), is the Nichtnegativstellensatz Theorem 2.4. It turns out that (5.2) is just computing the smallest generalized eigenvalue associated with the pair of symmetric matrices $(\mathbf{M}_n(f\cdot\mu), \mathbf{M}_n(\mu))$, for which specialized software exists. If both matrices are expressed in a polynomial basis of $\mathbb{R}[\mathbf{x}]_n$ formed by polynomials that are orthonormal with respect to μ, then $\mathbf{M}_n(\mu)$ becomes the identity matrix $\mathbf{I}$, and τ_n^* is just the smallest eigenvalue of $\mathbf{M}_n(f\cdot\mu)$.

Computational consideration. It is straightforward to fill up entries of both matrices $\mathbf{M}_n(\mu)$ and $\mathbf{M}_n(f\cdot\mu)$, so the main effort is in computing the generalized eigenvalue of the pair of (symmetric) matrices $(\mathbf{M}_n(f\cdot\mu), \mathbf{M}_n(\mu))$, which can be done via standard software for eigenvalue computation. However, and even if they are symmetric, computing κ_n is quite challenging because of the size $O(n^d)$ of the matrices as n increases.

Remark 5.3. A refinement of (5.1) is to consider polynomial densities σ non-negative on S (instead of SOS). That is, we replace (5.1) with

$$\hat{\kappa}_n = \inf_{\sigma\in Q_n(g)} \left\{ \int f\,\sigma\,\mathrm{d}\mu : \int \sigma\,\mathrm{d}\mu = 1 \right\}, \quad n \in \mathbb{N}, \tag{5.4}$$

where $Q_n(g)$ is the degree-$2n$ truncated quadratic module associated with the generators g that define S; see (2.6). For instance, if S is the unit box,

$$Q_n(g) = \left\{ \sigma_0 + \sum_{j=1}^{d} \sigma_j\,(1 - x_j^2);\ \sigma_0 \in \Sigma[\mathbf{x}]_n;\ \sigma_j \in \Sigma[\mathbf{x}]_{n-1},\ j = 1, \ldots, m \right\}.$$

Of course, $f^* \le \hat{\kappa}_n \le \kappa_n$ for all n and therefore, in view of Theorem 5.2, $\hat{\kappa}_n \downarrow f^*$ as n increases.

5.2. *An alternative univariate formulation*

Let the (univariate) Borel measure $\#\mu$ be the pushforward of μ on $\mathbb{R}$ by the mapping f, that is,

$$\#\mu(B) = \mu(f^{-1}(B)) \quad \text{for all } B \in \mathcal{B}(\mathbb{R}), \tag{5.5}$$

where $\mathcal{B}(\mathbb{R})$ is the usual Borel σ-algebra generated by the open sets of $\mathbb{R}$. By construction of $\#\mu$, $\mathrm{supp}(\#\mu) = f(S)$, and therefore

$$f^* = \min_z \{z : z \in \mathrm{supp}(\#\mu)\}, \tag{5.6}$$

Moreover, the moments $(\#\mu_j)_{j\in\mathbb{N}}$ of $\#\mu$ satisfy

$$\#\mu_j = \int_{f(S)} z^j\,\mathrm{d}\#\mu(z) = \int_S f(\mathbf{x})^j\,\mathrm{d}\mu(\mathbf{x}) \quad \text{for all } j \in \mathbb{N}. \tag{5.7}$$

As all moments μ_α of μ are available, the $\#\mu_j$ can be obtained exactly, for instance, by expanding the polynomial f^j in the canonical basis $(\mathbf{x}^\alpha)$,

$$\mathbf{x} \mapsto f(\mathbf{x})^j = \sum_{\alpha\in\mathbb{N}^d} \theta_\alpha^{(j)}\,\mathbf{x}^\alpha, \quad \#\mu_j = \sum_\alpha \theta_\alpha^{(j)}\mu_\alpha.$$

However, notice that even though the above expansion is always possible, it can become very tedious if j is large, even for modest dimension d.

Theorem 5.4. Let $\#\mu$ be the measure on $\mathbb{R}$ in (5.5) (the pushforward of μ by f), and let

$$\rho_n := \inf_{\sigma\in\Sigma[z]_n} \left\{ \int z\,\sigma\,\mathrm{d}\#\mu : \int \sigma\,\mathrm{d}\#\mu = 1 \right\}, \quad n \in \mathbb{N}, \tag{5.8}$$

$$\rho_n^* := \sup_\lambda \{\lambda : \lambda\,\mathbf{M}_n(\#\mu) \preceq \mathbf{M}_n(z\cdot\#\mu)\}, \quad n \in \mathbb{N}. \tag{5.9}$$

Then $(\rho_n)_{n\in\mathbb{N}}$ is a monotone non-increasing sequence such that $\rho_n \downarrow f^*$ as n increases. In addition, and letting $d_f = \deg(f)$, we obtain $\rho_n \ge \kappa_{nd_f}$ for every

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

$n \in \mathbb{N}$ (where κ_n is defined in (5.1)), because if $\sigma \in \Sigma[z]_n$ is a feasible solution of (5.8) then $\sigma \circ f \in \Sigma[\mathbf{x}]_{nd_f}$ is a feasible solution of (5.1) with the same value.

A detailed proof can be found in Lasserre (2021). There is a striking difference between the hierarchies of upper bounds $(\kappa_n)_{n\in\mathbb{N}}$ in (5.1) and $(\rho_n)_{n\in\mathbb{N}}$ in (5.8). The latter involves *univariate* moment and localizing matrices. Both matrices are Hankel matrices of size $O(n)$ (in contrast to multivariate Hankel-type matrices of size $O(n^d)$ for computing κ_n). Therefore, computing the generalized eigenvalue ρ_n is much easier than computing κ_n.

On the other hand, filling up all entries of the Hankel moment matrix $\mathbf{M}_n(\#\mu)$ is in principle easy but tedious. Indeed, if f^j is expanded in the monomial basis then its integration (5.7) with respect to μ is straightforward. However, as already noted, such an expansion can be quite costly if j is not small (even for modest dimension d).

Rates of convergence. It is worth noting that, just as for the hierarchy of lower bounds, $O(1/n^2)$ rates of convergence have also been obtained for the hierarchy of upper bounds (5.1)–(5.2) on the sets $\mathbb{S}^{d-1}$, $\mathbf{B}(0,1)$, $[-1,1]^d$ and Δ^d; see Slot (2022, Table 2, p. 2615).

5.3. *Notes and sources*

Section 5 is essentially based on Lasserre (2011, 2013); the univariate formulation is from Lasserre (2021). In a series of papers, de Klerk, Laurent and collaborators (e.g. de Klerk and Laurent 2011, Slot and Laurent 2021, Laurent and Slot 2023) have obtained rates of convergence $\kappa_n \downarrow f^*$ (multivariate) and $\rho_n \downarrow f^*$ (univariate) as n grows, by playing with various reference measures μ on S and a clever choice of appropriate families of densities. The approach is also interesting in its own right as it is a mix of several sophisticated techniques, including polynomial kernels and asymptotics for roots of some distinguished orthogonal polynomials. Moreover, it turns out that such techniques have also been useful in obtaining rates of convergence for the Moment-SOS hierarchy of lower bounds on specific sets S; for more details the interested reader is referred to Slot (2022) and references therein.

6. Some applications of the Moment-SOS hierarchy

In this section we briefly describe how the Moment-SOS hierarchy can be used to help solve several problems in various fields of science and engineering. In brief, problems where the Moment-SOS hierarchy is a relevant tool are those which have an equivalent formulation as an instance of the so-called *generalized moment problem* (GMP) whose description is only through polynomials and semi-algebraic sets (i.e. GMPs with algebraic data). Indeed, the list of potential applications of the GMP is almost endless, with polynomial optimization being its simplest instance.

As is to be expected from what we have seen for optimization, the GMP can be formulated in a primal form (via moments) or a dual form (via polynomials). Its primal form is an infinite-dimensional and linear (hence convex) optimization problem on measure spaces, which reads

$$\text{GMP:}\quad \inf_{\phi_1,\ldots,\phi_s}\left\{\sum_{j=1}^{s}\int f_j\,\mathrm{d}\phi_j:\ \sum_{j=1}^{s}\int h_{kj}\,\mathrm{d}\phi_j \geq b_k,\ k\in\Gamma;\right.\\ \left.\operatorname{supp}(\phi_j)\subset S_j,\ j=1,\ldots,s\right\},\tag{6.1}$$

where all functions f_j, h_{kj} are polynomials, and the ϕ_j are Borel measures whose supports $S_j \subset \mathbb{R}^{r_j}$, $j = 1,\ldots,s$, are all basic semi-algebraic sets.

The reader will note that all constraints are *linear* constraints linking moments of the involved measures ϕ_j, $j = 1,\ldots,s$ (whence the name of generalized moment problem). So the GMP is an infinite-dimensional LP on spaces of measures.

Of course, the GMP in (6.1) can be extended to more general functions and sets, but for a practical application of the Moment-SOS hierarchy, we need *algebraic data* (polynomials and basic semi-algebraic sets). Notice also that formulation (1.4) of a polynomial optimization problem is the simplest instance of the GMP in which there is only one unknown measure ϕ and only one (equality) moment constraint $\phi(S) = \int 1\mathrm{d}\phi = 1$.

The dual GMP* of (6.1) is also an infinite-dimensional LP, and when $p := \#\Gamma < \infty$, it reads

$$\text{GMP}^*:\quad \sup_{\lambda\in\mathbb{R}^p_+}\left\{\sum_{k=1}^{p}\lambda_k\,b_k:\right.\\ \left. f_j(\mathbf{x})-\sum_{k=1}^{p}\lambda_k\,h_{kj}(\mathbf{x})\geq 0,\ \forall\,\mathbf{x}\in S_j,\ j=1,\ldots,s\right\}.\tag{6.2}$$

Moment equality constraints are also tolerated in (6.1), in which case the associated dual variable λ_k in (6.2) is not constrained to be non-negative. As is the case in some important applications, the set Γ is also allowed to be (countably) infinite. Finally, the objective function of (6.1) can also be a convex function of finitely many moments of measures ϕ_j (e.g. $-\log\det(\mathbf{M}_n(\phi_j))$ of the moment matrix $\mathbf{M}_n(\phi_j)$), though we do not discuss this further.

As we see next in two examples, in some applications the problem to solve is already in the form of a GMP (or GMP*), whereas in other applications, some equivalent formulation of the problem is an instance of the GMP.

Strategy of the Moment-SOS hierarchy. Roughly speaking, to apply the Moment-SOS hierarchy to the GMP (6.1), we proceed as follows.

- We replace the measures $(\phi_j)_{j=1,\ldots,s}$ with degree-$2n$ truncated pseudo-moment vectors $\boldsymbol{\phi}_j = (\phi_{j,\alpha})$, $\boldsymbol{\alpha}\in\mathbb{N}^{r_j}_{2n}$, $j = 1,\ldots,s$.

- We impose semidefinite constraints on the moment and localizing matrices associated with each $\boldsymbol{\phi}_j$ and each set S_j, which by Theorem 2.2(ii) are necessary conditions for $\boldsymbol{\phi}_j$ to be moments of a measure on S_j.

Then, as the moment constraints and the criterion are just linear on the pseudo-moment vectors $\boldsymbol{\phi}_j$, for each fixed n we end up with a finite-dimensional semidefinite relaxation that provides a lower bound on the optimal value of the GMP. Similarly, to apply the Moment-SOS hierarchy to GMP* in (6.2), we proceed as follows.

- For each $j = 1, \ldots, s$, we replace the positivity constraint

$$f_j - \sum_{k=1}^{p} \lambda_k \, h_{kj} \geq 0 \quad \text{on } S_j$$

with a Putinar certificate of positivity of degree $2n$; see Theorem 2.2(i). For instance, if $S_j = [-1, 1]^d$ then the above positivity constraint reads

$$f_j - \sum_{k=1}^{p} \lambda_k \, h_{kj} = \sum_{j=0}^{d} \sigma_j \, (1 - x_j^2), \quad \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j}, \; j = 0, \ldots, d.$$

Then, as the criterion $\sum_k \lambda_k b_k$ is linear, we end up with a finite-dimensional semidefinite program, which is the dual of the program on pseudo-moment vectors. In fact, a primal–dual semidefinite solver will solve both of them at the same time.

For completeness, below are two illustrative applications of the above strategy. In the first example, in computational geometry and probability, the problem itself is described as a GMP, while in the second example, in optimal control, an alternative and so-called 'weak formulation' of the optimal control problem is an instance of the GMP.

6.1. *Illustration in probability and computational geometry*

Let $S \subset \mathbb{R}^d$ be a compact set and suppose that $S \subset \mathbf{B} := [-1, 1]^d$ (possibly after scaling). The goal is to approximate the Lebesgue volume $\mathrm{vol}(S)$ of S, as closely as desired. This is known to be a very hard problem. In fact, even if S is convex, approximating its volume is quite hard; see e.g. Dyer and Frieze (1988), the discussion in Henrion, Lasserre and Savorgnan (2009) and references therein.

Let λ be the Lebesgue measure on $[-1, 1]^d$, so that its (infinite) vector of moments $\boldsymbol{\lambda} = (\lambda_\alpha)_{\alpha \in \mathbb{N}^d}$ is available in closed form. Let $\mathbf{1} \in \mathbb{R}[\mathbf{x}]$ be the constant polynomial (equal to 1 for all $\mathbf{x}$), and for two measures μ, ν on $\mathbb{R}^d$, the notation $\nu \leq \mu$ stands for $\nu(B) \leq \mu(B)$ for all $B \in \mathcal{B}(\mathbb{R}^d)$.

Proposition 6.1. We have

$$\mathrm{vol}(S) = \max_{\phi \in \mathscr{M}(S)_+, \nu \in \mathscr{M}(\mathbf{B})_+} \{\phi(\mathbf{1}) \colon \phi + \nu = \lambda\},$$

and $\phi^* := 1_S \, \lambda$ is the unique optimal solution.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Proof. As $\phi+\nu=\lambda$, $\phi\le\lambda$ and since $\mathrm{supp}(\phi)\subseteq S$, we have $\phi(\mathbf{1})=\phi(S)\le\lambda(S)=\mathrm{vol}(S)$. Next, with $\phi^*:=1_S\,\lambda\in\mathscr{M}(S)_+$, and $\nu^*:=\mu-\phi^*\in\mathscr{M}(\mathbf{B})_+$, we obtain $\phi^*(\mathbf{1})=\lambda(S)=\mathrm{vol}(S)$. □

The above formulation of vol(S) as an optimization problem is not yet in the form of the GMP (6.1). But notice that since $\mathbf{B}$ is compact,

$$\phi+\nu=\lambda \;\Leftrightarrow\; \phi_\alpha+\nu_\alpha=\lambda_\alpha \quad \text{for all } \boldsymbol{\alpha}\in\mathbb{N}^d,$$

and therefore

$$\mathrm{vol}(S)=\max_{\phi\in\mathscr{M}(S)_+,\,\nu\in\mathscr{M}(\mathbf{B})_+}\{\phi(\mathbf{1})\colon\ \phi_\alpha+\nu_\alpha=\lambda_\alpha,\ \forall\,\boldsymbol{\alpha}\in\mathbb{N}^d\},\tag{6.3}$$

which is an instance of the GMP (6.1) with $\Gamma=\mathbb{N}^d$ (a countable set). We next see how to implement the moment-SOS hierarchy. Let $g_0=\mathbf{1}$,

$$S=\{\mathbf{x}\in\mathbb{R}^d\colon\ g_j(\mathbf{x})\ge 0,\ j=1,\dots,m\},$$

and recall that $d_j=\lceil\deg(g_j)/2\rceil$ for all $j=0,\dots,m$. For each $n\in\mathbb{N}$, consider the optimization problem

$$\begin{aligned}\tau_n=\max_{\boldsymbol{\phi},\boldsymbol{\nu}}\{\phi_0\colon\ &\phi_\alpha+\nu_\alpha=\lambda_\alpha,\ \forall\,\boldsymbol{\alpha}\in\mathbb{N}^d_{2n};\\ &\mathbf{M}_{n-d_j}(g_j\cdot\boldsymbol{\phi})\succeq 0,\ j=0,\dots,m;\ \mathbf{M}_n(\boldsymbol{\nu})\succeq 0\}.\end{aligned}\tag{6.4}$$

For each fixed n, (6.4) is a semidefinite program and an obvious relaxation of (6.3) so that $\tau_n\ge\mathrm{vol}(S)$ for all n.

Theorem 6.2 (Henrion *et al.* 2009). The sequence of optimal values $(\tau_n)_{n\in\mathbb{N}}$ is monotone non-increasing, bounded below, and $\lim_{n\to\infty}\tau_n=\mathrm{vol}(S)$.

So (6.4), indexed by $n\in\mathbb{N}$, provides a hierarchy of semidefinite relaxations of (6.3) such that $(\tau_n)_{n\in\mathbb{N}}$ converges (from above) to the desired value vol(S) as n increases. However, in its basic form (6.4), its convergence is quite slow. To see why, consider the dual of (6.4), which is the semidefinite program

$$\begin{aligned}\tau_n^*=\min_{p\in\mathbb{R}[\mathbf{x}]_{2n}}\biggl\{\int_{\mathbf{B}}p\,\mathrm{d}\lambda\colon\ &p-1=\sum_{j=0}^m\sigma_j\,g_j\\ &p\in\Sigma[\mathbf{x}]_n;\ \sigma_j\in\Sigma[\mathbf{x}]_{n-d_j};\ j=0,\dots,m\biggr\}.\end{aligned}\tag{6.5}$$

It turns out that if S has non-empty interior then $\tau_n=\tau_n^*$ for all n. Next observe that

$$p\in\Sigma[\mathbf{x}]_n \quad\text{and}\quad p-1=\sum_{j=0}^m\sigma_j\,g_j\ \Rightarrow\ p\ge 1_S \quad \text{for all } \mathbf{x}\in\mathbf{B},$$

and since $\int p\,\mathrm{d}\lambda\downarrow 1_S\,\mathrm{d}\lambda$ as n grows, in the dual (6.5) we search for a degree-$2n$ SOS polynomial p that tries to approximate from above the indicator function of

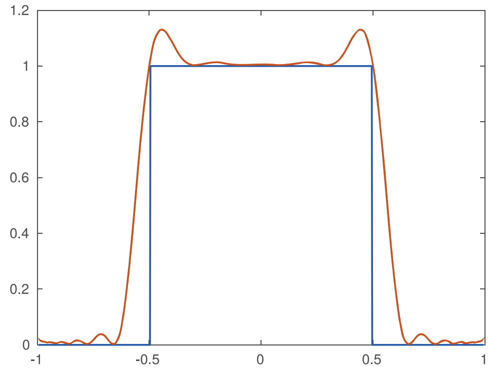

Figure 6.1. $S = [-0.5, 0.5] \subset [0, 1]$; polynomial approximation (in red) of 1_S on $[0, 1]$ with Gibbs phenomenon.

S for all $\mathbf{x} \in \mathbf{B}$. As is well known, for such polynomial approximations a typical Gibbs phenomenon (oscillations) occurs at points of discontinuities, which makes the convergence quite slow; see Figure 6.1.

Fortunately one can significantly attenuate (or even remedy) this problem. Indeed, as we know in advance the (unique) optimal solution $\phi^* = 1_S\lambda$ of (6.3), every available additional information on ϕ^* in terms of linear constraints on its moments can be added to (6.3) without changing its optimal value and solution. While such additional redundant constraints do not change (6.3), they have a dramatic impact on the relaxations (6.4) and yield a fairly significant acceleration of their convergence. This is indeed the case if we add additional moment constraints (satisfied by ϕ^*) coming from Stokes' theorem.

Stokes constraints. Let us see how it works for the case where $S = \{\mathbf{x} \colon g(\mathbf{x}) \geq 0\}$ for some polynomial $g \in \mathbb{R}[\mathbf{x}]$ with compact sublevel set S. As g vanishes on ∂S, by Stokes' theorem,

$$\int_S \mathrm{Div}(\mathbf{x}\, g(\mathbf{x})\, \mathbf{x}^\alpha)\, \mathrm{d}\mathbf{x} = \int_{\partial S} \langle \vec{n}_\mathbf{x}, \mathbf{x}\rangle g(\mathbf{x})\mathbf{x}^\alpha\, \mathrm{d}\sigma(\mathbf{x}) = 0 \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^d,$$

where $\vec{n}_\mathbf{x}$ is the outward pointing normal at $\mathbf{x} \in \partial S$. Hence each $\boldsymbol{\alpha} \in \mathbb{N}^d$ provides us with the moment constraint

$$\phi^*(s_\alpha) := \phi^*(\mathrm{Div}(\mathbf{x}\, g(\mathbf{x})\mathbf{x}^\alpha) = 0 \quad \text{on } \phi^*,$$

because $\mathbf{x} \mapsto s_\alpha(\mathbf{x}) := \mathrm{Div}(\mathbf{x}\, g(\mathbf{x})\, \mathbf{x}^\alpha)$ is a polynomial (of degree $\deg(g) + |\alpha| + 1$). Hence, for every $n \in \mathbb{N}$, the additional moment constraints

$$\phi(s_\alpha) = 0 \quad \text{for all } \boldsymbol{\alpha} \colon |\alpha| \leq 2n - 1 - \deg(g) \tag{6.6}$$

can be included in the semidefinite relaxation (6.4). The effect on the dual (6.5) is to change the initial constraint $p - 1 = \sigma_0 + \sigma_1\, g$ to

$$p + q - 1 = \sigma_0 + \sigma_1\, g,$$

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

where $q := \sum_{|\alpha|\leq 2n-\deg(g)-1} \theta_\alpha\, s_\alpha$ for the dual variables (θ_α) associated with (6.6). Hence, p is *no longer* required to approximate 1_S from above! For more details on volume computation via the moment-SOS hierarchy, the interested reader is referred to Henrion *et al.* (2009), Tacchi, Weisser, Lasserre and Henrion (2021) and Tacchi, Lasserre and Henrion (2023). In particular, this technique has also been implemented in Tacchi *et al.* (2021) to approximate the volume of certain non-convex sets $S \subset \mathbb{R}^{100}$ where the description of S exhibits some structured sparsity, as described in Section 3.8.

Remark 6.3. One can consider (6.3) with a measure λ that is not Lebesgue measure on $\mathbf{B}$. For example, if λ is the Gaussian measure $\exp(-\|\mathbf{x}\|^2)\,\mathrm{d}\mathbf{x}$ on $\mathbb{R}^d$, or the exponential measure $\exp(-\sum_j x_j)\,\mathrm{d}\mathbf{x}$ on $\mathbb{R}^d_+$, one can approximate the value $\lambda(S)$ for non-compact semi-algebraic sets as closely as desired; see e.g. Lasserre (2017).

Application in probability. Suppose that X is a $\mathbb{R}^d$-valued random vector whose distribution is only partially known through a few of its moments $\boldsymbol{m} = (m_\alpha)_{\alpha\in\Gamma}$, where $\Gamma \subset \mathbb{N}^d$ is a finite set (typically the index set of moments up to order 3, 4). Next, let $S \subset \mathbb{R}^d$ be a given compact basic semi-algebraic set with non-empty interior. The goal is to provide the best upper bound on $\mathrm{Prob}(X \in S)$, under the partial knowledge of $\boldsymbol{m} = (m_\alpha)_{\alpha\in\Gamma}$, that is, compute

$$\rho = \max_{\mu\in\mathscr{M}(\mathbb{R}^d)_+} \left\{\mu(S) \colon \int \mathbf{x}^\alpha\,\mathrm{d}\mu = m_\alpha,\ \forall\,\boldsymbol{\alpha}\in\Gamma\right\}. \tag{6.7}$$

Observe that (6.7) is an instance of the GMP (6.1) but with non-polynomial data because with $\mu \in \mathscr{M}(\mathbb{R}^d)_+$, $\mu(S) = \mu(1_S)$ and 1_S is not a polynomial. We need to consider that $\mu = \phi + \nu$ with $\phi \in \mathscr{M}(S)_+$ (in which case $\phi(S) = \phi(\mathbf{1}) = \phi_0$), and hence for every $n \in \mathbb{N}$, consider the optimization problem

$$\begin{aligned}\tau_n = \max_{\boldsymbol{\phi},\boldsymbol{\nu}} \{\phi_0 \colon \phi_\alpha + \nu_\alpha = m_\alpha,\ \forall\,\boldsymbol{\alpha}\in\Gamma;\\ \mathbf{M}_{n-d_j}(g_j\cdot\boldsymbol{\phi}) \succeq 0,\ j = 0,\ldots,m;\ \mathbf{M}_n(\boldsymbol{\nu}) \succeq 0\},\end{aligned} \tag{6.8}$$

which is a relaxation of (6.7) and a variant of (6.4) (in fact even easier because now Γ is a finite set instead of the countable set $\mathbb{N}^d$). For instance, with $\Gamma = \mathbb{N}^d_{2t}$ for some fixed t, the dual of (6.8) reads

$$\begin{aligned}\tau_n^* = \min_p \left\{\sum_{\alpha\in\Gamma} p_\alpha\, m_\alpha \colon p - 1 = \sum_{j=0}^m \sigma_j\, g_j\right.\\ \left. p\in\Sigma[\mathbf{x}]_t;\ \sigma_j\in\Sigma[\mathbf{x}]_{n-d_j};\ j=0,\ldots,m\right\}.\end{aligned} \tag{6.9}$$

The difference from (6.5) is that now, even when n changes, we still search for a degree-$2t$ polynomial $p \geq 1_S$ (where the degree $2t$ is fixed by the number of moments in Γ). In this case we do not have a Gibbs phenomenon because the degree of p is fixed.

Next, let

$$\Phi := \left\{\mu \in \mathscr{M}(\mathbb{R}^d)_+ : \int \mathbf{x}^\alpha \, \mathrm{d}\mu = m_\alpha, \ \forall \boldsymbol{\alpha} \in \Gamma\right\}.$$

If there is some $\mu \in \Phi$ with a strictly positive density with respect to Lebesgue measure, and with all moments finite, then each relaxation (6.9) is solvable and there is no duality gap, i.e. $\tau_n = \tau_n^*$, and there is an optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_t$ for all n. Moreover, $\tau_n \downarrow \tau \geq \rho$ as $n \to \infty$. Finally, the bound τ may be sharp, i.e. $\tau = \rho$, if at some step n, the moment matrices of an optimal solution $(\boldsymbol{\phi}, \boldsymbol{\nu})$ of (6.8) satisfy some 'flatness' property; for more details see e.g. Lasserre (2002*a*) and Bertsimas and Popescu (2005).

6.2. *Illustration in optimal control of dynamical systems*

In this section we briefly describe how to apply the Moment-SOS hierarchy to help solve optimal control problems (OCPs) with algebraic data, i.e. those problems whose description is through polynomials and basic semi-algebraic sets. Consider the optimal control problem

$$\begin{aligned} \text{OCP:} \quad J(\mathbf{x}_0, 0) := \min_{\mathbf{u}} \Bigg\{ \int_0^1 & h(\mathbf{x}(t), \mathbf{u}(t)) \, \mathrm{d}t + H(\mathbf{x}(1)) : \\ & \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)), \ \forall\, t \in (0,1), \\ & \mathbf{x}(t) \in X, \ \mathbf{u}(t) \in U, \ \forall\, t \in (0,1), \\ & \mathbf{x}(0) = \mathbf{x}_0 \Bigg\}, \end{aligned} \tag{6.10}$$

where h, H, f are polynomials and $X \subset \mathbb{R}^d$ and $U \subset \mathbb{R}^m$ are basic semi-algebraic set, and $\mathbf{x}_0 \in X$ is the initial condition.

Equation (6.10) describes a dynamical system whose evolution in the time interval $[0, 1]$ of its state $\mathbf{x}(t) \in \mathbb{R}^d$, $t \in [0, 1]$, is governed by a controlled ODE with vector field $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$, and control $\mathbf{u}(t) \in U$ for all $t \in (0, 1)$. The goal is to approximate an optimal (or close to optimal) control trajectory $t \mapsto \mathbf{u}^*(t) \in \mathbb{R}^m$, $t \in [0, 1]$, which minimizes the functional $\int_0^1 h(\mathbf{x}(t), \mathbf{u}(t)) \, \mathrm{d}t + H(\mathbf{x}(1))$. Here we will not discuss the appropriate function spaces in which we have to search the state and control trajectories $\mathbf{x}(t), \mathbf{u}(t)$. Instead we show:

(i) how to model (6.10) as a particular instance of the GMP (6.1), and

(ii) how to define an appropriate Moment-SOS hierarchy for solving (6.10).

Note that in contrast to POPs, where we search for a point $\mathbf{x}^* \in S \subset \mathbb{R}^d$, we now search for maps $(\mathbf{x}^*, \mathbf{u}^*) \colon [0, 1] \to \mathbb{R}^d \times \mathbb{R}^m$, a much more difficult problem which is already infinite-dimensional in its description.

Strategy. As for (static) optimization problems (1.1), where in the Moment-SOS hierarchy we search for a probability measure (the Dirac measure $\delta_{\{\mathbf{x}^*\}}$ at a global

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

minimizer $\mathbf{x}^* \in S$), here we will also search for a measure μ, now supported on an optimal state-control trajectory $\{(\mathbf{x}^*(t), \mathbf{u}^*(t))\}$ from time $t = 0$ up to time $t = 1$.

By integrating polynomial test functions along feasible trajectories, the ordinary differential equation that governs the dynamical system will provide *linear* constraints on moments of μ and moments of a terminal measure ν on X at time $t = 1$ (because the vector field f in (6.10) is a polynomial). The state and control constraints in (6.10) translate into support constraints on μ and ν.

Modelling (6.10) *as a GMP via occupation measures.* The idea is to look at (6.10) via its impact on the evaluation of test functions all along feasible trajectories. Let $\{(\mathbf{x}(t), \mathbf{u}(t)) \colon t \in [0, 1]\}$ be an admissible trajectory, and let $(\mathbf{x}, t) \mapsto w(\mathbf{x}, t)$ be an arbitrary *test* function in $\mathscr{C}^1(X \times [0, 1])$. Then observe that

$$\begin{aligned} w(\mathbf{x}(1), 1) - w(\mathbf{x}(0), 0) &= \int_0^1 \mathrm{d}w(\mathbf{x}(t), t) \\ &= \int_0^1 \frac{\partial w(\mathbf{x}(t), t)}{\partial t} + \langle \nabla_{\mathbf{x}} w(\mathbf{x}(t), t), f(\mathbf{x}(t), \mathbf{u}(t))\rangle \,\mathrm{d}t. \end{aligned} \tag{6.11}$$

Introduce the measures μ on $X \times U \times [0, 1]$, and ν_0, ν on $X \times [0, 1]$:

$$\mu(A \times B \times C) = \int_{C\cap[0,1]} 1_{A\cap X}(\mathbf{x}(t))\, 1_{B\cap U}(\mathbf{u}(t))\,\mathrm{d}t, \tag{6.12}$$

$$\nu(A \times C) = 1_{(A\cap X)\times(C\cap\{1\})}(\mathbf{x}, t), \tag{6.13}$$

$$\nu_0(A \times C) = 1_{(A\cap X)\times(C\cap\{0\})}(\mathbf{x}, t), \tag{6.14}$$

for all Borel sets $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(U)$ and $C \in \mathcal{B}([0, 1])$. The measure μ is called the *occupation measure* up to time 1, while ν_0 (resp. ν) is called the occupation measure at time $t = 0$ (resp. at time $t = 1$), all associated with the trajectory $\{(\mathbf{x}(t), \mathbf{u}(t)) \colon t \in [0, 1]\}$. Another equivalent characterization of μ is via its *disintegration*

$$\mathrm{d}\mu(\mathbf{x}, \mathbf{u}, t) = \delta_{\{(\mathbf{x}(t),\mathbf{u}(t))\}}(d(\mathbf{x}, \mathbf{u}))\, 1_{[0,1]}(t)\,\mathrm{d}t, \tag{6.15}$$

into:

- its *marginal* $1_{[0,1]}(t)\,\mathrm{d}t$ on $[0, 1]$, and
- its *conditional probability* $\delta_{\{(\mathbf{x}(t),\mathbf{u}(t))\}}(d(\mathbf{x}, \mathbf{u}))$ on $X \times U$, given $t \in [0, 1]$ (which is the Dirac measure at the point $(\mathbf{x}(t), \mathbf{u}(t)) \in X \times U$).

It is important to observe that the support of μ is the *graph* $\{(t, \mathbf{x}(t), \mathbf{u}(t)) \colon t \in [0, 1]\}$ of state-control trajectories $(\mathbf{x}(t), \mathbf{u}(t))$. Ideally we search for the measure

$$\mathrm{d}\mu^*(\mathbf{x}, \mathbf{u}, t) = \delta_{\{(\mathbf{x}^*(t),\mathbf{u}^*(t))\}}(d(\mathbf{x}, \mathbf{u}))\, 1_{[0,1]}(t)\,\mathrm{d}t,$$

whose support is exactly the graph of optimal state-control trajectories $(\mathbf{x}^*(t), \mathbf{u}^*(t))$, $t \in [0, 1]$, when the latter exist.

One reason for introducing occupation measures is that the *time* integral (6.11) reads as the *spatial* integral

$$\int w\,\mathrm{d}\nu - \int w\,\mathrm{d}\nu_0 = \int \frac{\partial w}{\partial t}(\mathbf{x},t) + \langle \nabla_{\mathbf{x}} w(\mathbf{x},t), f(\mathbf{x},\mathbf{u})\rangle\,\mathrm{d}\mu(\mathbf{x},\mathbf{u},t), \tag{6.16}$$

where the variables $(\mathbf{x},\mathbf{u},t)$ are now treated as *independent* variables. The respective dependence of $(\mathbf{x},\mathbf{u})$ on t is implicit through the support of μ.

Next, introduce the operator $\mathcal{L}\colon \mathscr{C}^1(X\times[0,1]) \to \mathscr{C}(X\times U\times[0,1])$:

$$w \mapsto \mathcal{L}w := \frac{\partial w}{\partial t} + \langle \nabla_{\mathbf{x}} w, f\rangle,$$

and its adjoint $\mathcal{L}^*\colon \mathscr{C}(X\times U\times[0,1])^* \to \mathscr{C}^1(X\times[0,1])^*$ by

$$\mu \mapsto \mathcal{L}^*\mu := -\frac{\partial \mu}{\partial t} - \sum_{i=1}^{d} \frac{\partial(f_i\mu)}{\partial x_i} = -\frac{\partial\mu}{\partial t} - \operatorname{div}(f\mu),$$

where derivatives of measures are understood in a weak sense via their actions on smooth test functions (and the change of signs comes from integration by parts). Then (6.16) reads

$$\langle w,\nu\rangle - \langle w,\nu_0\rangle = \langle \mathcal{L}w,\mu\rangle = \langle w,\mathcal{L}^*\mu\rangle,$$

and as it must be valid for all test functions w in a dense subset $\mathcal{D}\subset\mathscr{C}^1(X\times[0,1])$, we obtain the equation $\mathcal{L}^*\mu = \nu - \nu_0$, that is,

$$\frac{\partial\mu}{\partial t} + \operatorname{div}(f\,\mu) + \nu = \nu_0. \tag{6.17}$$

Equation (6.17) is a linear transport equation (transporting ν_0 to ν) which is classical in fluid mechanics, statistical physics and PDEs. It is known under several names, including the *equation of conservation of mass*, the *advection equation* and *Liouville's equation*.

This observation allows us to define the so-called *measure-valued weak formulation* of the OCP:

$$\begin{aligned}\rho = \inf_{\mu,\nu}\bigg\{ &\int h\,\mathrm{d}\mu + \int H\,\mathrm{d}\nu\colon\\ &\int \frac{\partial w}{\partial t} + \langle \nabla_{\mathbf{x}} w, f\rangle\,\mathrm{d}\mu = \int w\,\mathrm{d}\nu - \int w\,\mathrm{d}\nu_0,\ \forall\, w\in\mathcal{D};\\ &\mu\in\mathscr{M}(X\times U\times[0,1])_+,\ \nu\in\mathscr{M}(X\times\{1\})_+\bigg\},\end{aligned} \tag{6.18}$$

introduced by Vinter (1993).

Observe that if $\mathcal{D}$ is a countable set of polynomials, then (6.18) is an instance of the GMP in (6.1) and of course a relaxation of (6.10) so that $\rho \le J(\mathbf{x}_0, 0)$. The

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

dual of (6.18) reads

$$\rho^* = \sup_{w\in\mathscr{C}^1(X\times[0,1])} \left\{ \int w\,\mathrm{d}\nu_0 \ (= w(\mathbf{x}_0,0)) : \right.$$
$$h + \mathcal{L}w \geq 0,\ \forall\,(\mathbf{x},\mathbf{u},t) \in X \times U \times [0,1];$$
$$\left. w(\mathbf{x},1) \leq H(\mathbf{x}),\ \forall\,\mathbf{x} \in X \right\}. \tag{6.19}$$

It turns out that under some convexity assumptions, $\rho = J(\mathbf{x}_0, 0)$, that is, the measure-valued weak formulation (6.18) is equivalent to the strong formulation (6.10). In the dual (6.19) we approximate the optimal value function $J\colon X\times[0,1] \to \mathbb{R}$ all along an optimal trajectory $\{(\mathbf{x}^*(t),t)\colon t \in [0,1]; \mathbf{x}^*(0) = \mathbf{x}_0\}$ (but *not* for all $(\mathbf{x},t) \in X \times [0,1]$). For more details see e.g. Lasserre, Henrion, Prieur and Trélat (2008), Korda *et al.* (2022) and references therein.

Next, to implement the moment SOS hierarchy we first select a countable set of test functions $\mathcal{D}$, namely the set of monomials

$$\mathcal{D} := \{(\mathbf{x}^\alpha\,\mathbf{u}^\beta\,t^k)\colon \boldsymbol{\alpha} \in \mathbb{N}^d, \boldsymbol{\beta} \in \mathbb{N}^m, k \in \mathbb{N}\},$$

which is dense in $\mathscr{C}^1(X \times [0,1])$. Then, for every $n \in \mathbb{N}$, let $\mathcal{D}_n := \{(\mathbf{x}^\alpha\,\mathbf{u}^\beta\,t^k) \in \mathcal{D}\colon |\boldsymbol{\alpha}+\boldsymbol{\beta}| + k \leq 2n\}$, and consider the optimization problem

$$\rho_n = \min_{\boldsymbol{\mu},\boldsymbol{\nu}} \left\{ \mu(h) + \nu(H) : \right.$$
$$\mu\left(\frac{\partial w}{\partial t} + \langle \nabla_{\mathbf{x}} w, f\rangle\right) = \nu(w) - \nu_0(w),\ \forall\,w \in \mathcal{D}_n;$$
$$\mathbf{M}_n(\boldsymbol{\mu}),\ \mathbf{M}_n(\boldsymbol{\nu}) \succeq 0;\ \mathbf{M}_n((1-t)\boldsymbol{\nu}) = 0;$$
$$\mathbf{M}_{n-d_g}(g\cdot\boldsymbol{\mu}),\ \mathbf{M}_{n-d_g}(g\cdot\boldsymbol{\nu}) \succeq 0,\ \forall\,g \in G;$$
$$\left. \mathbf{M}_{n-1}(t(1-t)\cdot\boldsymbol{\mu}) \succeq 0,\ \mathbf{M}_{n-d_\theta}(\theta\cdot\boldsymbol{\mu}) \succeq 0,\ \forall\,\theta \in \Theta \right\}. \tag{6.20}$$

where $X = \{\mathbf{x}\colon g(\mathbf{x}) \geq 0,\ g \in G\}$, $U = \{\mathbf{u}\colon \theta(\mathbf{u}) \geq 0,\ \theta \in \Theta\}$, $d_g = \lceil \deg(g)/2 \rceil$, $g \in G$, and $d_\theta = \lceil \deg(\theta)/2 \rceil$, $\theta \in \Theta$.

So the sequence of optimal values $(\rho_n)_{n\in\mathbb{N}}$ is monotone non-decreasing, and under the convexity assumptions alluded to above, $\rho_n \uparrow J(\mathbf{x}_0, 0)$ as $n \to \infty$.

Reconstruction of optimal trajectories from moments. So far, by solving the semidefinite relaxations (6.20) we obtain a sequence $(\rho_n)_{n\in\mathbb{N}}$ of lower bounds on the optimal value $J(\mathbf{x}_0, 0)$ of the initial OCP (6.10). But from the vector of pseudo-moments $(\boldsymbol{\mu}^n, \boldsymbol{\nu}^n)$ optimal solution of (6.20) for some degree n, can we retrieve or approximate optimal trajectories $(\mathbf{x}^*(t), \mathbf{u}^*(t))$ when they exist, or provide ε-suboptimal trajectories otherwise?

Again, and ideally, when n is sufficiently large, we expect $(\boldsymbol{\mu}^n, \boldsymbol{\nu}^n)$ to provide quite good approximations of the moments of the measures

$$\mathrm{d}\mu(\mathbf{x}, \mathbf{u}, t) = \delta_{\{\mathbf{x}^*(t), \mathbf{u}^*(t))\}}(d(\mathbf{x}, \mathbf{u}))\, 1_{[0,1]}(t)\, \mathrm{d}t$$

and ν, supported respectively on the trajectories $(\mathbf{x}^*(t), \mathbf{u}^*(t))$ and on the point $(\mathbf{x}^*(1), 1) \in X \times \{1\}$. In Section 7.3 we describe an efficient strategy via the Christoffel function, a tool from approximation theory and orthogonal polynomials, particularly well suited to identifying the support of a measure solely from knowledge of its moments.

6.3. *Other applications*

Here we provide the reader with some references to other applications of the Moment-SOS hierarchy. The purpose of this list, which is not exhaustive, is to convince the reader that the Moment-SOS hierarchy is indeed a versatile tool, widely applicable when the problem data are algebraic, and provided that some sparsity or symmetries can be exploited when the problem size demands.

- Control and stochastic control: Pauwels, Henrion and Lasserre (2017), Henrion and Garulli (2005), Parrilo (2000), Parrilo and Lall (2003), Aylward, Parrilo and Slotine (2008), Parrilo (2003), Korda, Henrion and Jones (2014), Fantuzzi, Goluskin, Huang and Chernyshenko (2016) and Goluskin and Fantuzzi (2019).
- For convex computation of the region of attraction for dynamical systems, see e.g. Henrion *et al.* (2021, Chapter 10), and for analysis and control of some types on non-linear PDEs, see e.g. Marx, Weisser, Henrion and Lasserre (2020), Henrion *et al.* (2021, Chapter 11) and Korda *et al.* (2022).
- Tensor computation: Nie (2023), Nie and Yang (2020), Fan, Nie and Zhou (2018) and Nie (2017).
- Algorithmic game theory: Stein, Ozdaglar and Parrilo (2008), Laraki and Lasserre (2012) and Nie and Tang (2023).
- Management of energy networks, and in particular for solving the optimal power flow problem for (large) electricity networks. The Moment-SOS hierarchy has been able to handle problems with thousands of variables by exploiting some inherent sparsity, in the spirit of Section 3.8; see e.g. Molzhan and Hiskens (2015), Tian, Wei and Tan (2015), Molzahn and Josz (2018), Cicconet and Almeida (2019) and Haussmann, Liers, Stingl and Vera (2018).
- Computer science, e.g. for coding and packing problems: Bachoc and Vallentin (2008), Bachoc, Passuello and Vallentin (2013), Dostert and Vallentin (2020) and de Laat *et al.* (2022).

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

- Computer vision, geometric perception and pattern recognition: Yang and Carlone (2020), Yang, Shi and Carlone (2020), Yang and Pavone (2023) and Probst, Paudel, Chhatkuli and Van Gool (2019).
- Mathematical finance, for portfolio optimization and option pricing: Lasserre, Priéto-Rumeau and Zervos (2006) and Gepp, Harris and Vanstone (2020). When the evolution in time is modelled by Itô's stochastic differential equations, a weak formulation of the problem via occupation measures is almost identical to that of OCPs, the only difference being that a second-order differential operator appears in the infinitesimal generator. It is also the case for computing exit-time distribution (of a given set) in stochastic models, as described in Lasserre and Priéto-Rumeau (2004), for example.
- The Internet of Things (IoT): Sedighi, Mishra, Shankar and Ottersten (2021).
- Computer graphics and geometry processing: Marschner, Palmer, Zhang and Solomon (2020, 2021).
- Signal processing: Marmin, Castella, Pesquet and Duval (2021) and de Castro, Gamboa, Henrion and Lasserre (2017).
- Optimal design in statistics: de Castro *et al.* (2019).
- Physics, for bounding ground-state energy of interacting particle systems: de Laat (2020).
- Chemistry, for deriving bounds on stochastic chemical kinetic systems: Dowdy and Barton (2018).
- Traffic networks, for bounding travel time: Xiangfeng Ji, Xuegang (Jeff) Ban, Jian Zhang and Bin Ran (2019).
- Engineering: Courtier *et al.* (2022).
- Machine learning, for certification of robustness for neural networks: Latorre, Rolland and Cevher (2020) and Chen Tong, Lasserre, Magron and Pauwels (2021).
- Quantum information (e.g. for several problems in entanglement theory): Eisert, Hyllus, Gühne and Curty (2004), Bacari, Gogolin, Wittek and Acín (2020), Parekh and Thompson (2021), Doherty, Parrilo and Spedalieri (2005) and Selby *et al.* (2023).
- Data analysis of citation networks: e.g. Wittek, Darányi and Nelhans (2017).
- Radar and wireless communications: Jie Pan and Fu Jiang (2020).
- Medical applications of cancer treatment: Moussa, Fiacchini and Alamir (2020).
- Truss topology design: Tyburec, Zeman, Kruzik and Henrion (2021).

6.4. *Notes and sources*

Section 6.1 is based on Lasserre (2002*a*) and Henrion *et al.* (2009), while Section 6.2 is based on Lasserre *et al.* (2008). Infinite-dimensional LP formulations of optimal control problems can be traced back to works of L. C. Young, A. F. Filippov, R. V. Gamkrelidze and J. Warga; see Fattorini (1999) for a historical survey. The novelty is the observation that such problems (with algebraic data) can be approximated numerically by semidefinite relaxations. In particular, it should be noted that while state constraints '$\mathbf{x}(t) \in X$ for all t' are usually considered a source of additional difficulties for classical numerical methods, they pose no problem for the Moment-SOS hierarchy as they simply appear in the support of the occupation measure.

The Moment-SOS hierarchy approach to analysis and control of some non-linear PDEs in Korda *et al.* (2022) follows the same principles as for solving OCPs. Namely, we consider a measure-valued weak formulation of the problem, similar to that in (6.18) for OCPs, using test functions and occupation measures. Appropriate conditions are required for the weak formulation to be equivalent to the initial (strong) formulation. For instance, for the Burgers equation, additional entropy constraints (due to Kruzkhov) on the occupation measures are needed; see Marx *et al.* (2020) and Henrion *et al.* (2021, Chapter 11).

The reconstruction technique of state and control trajectories based on the Christoffel function is detailed in Marx *et al.* (2021). In particular, this technique has been used with success in Marx *et al.* (2020) to recover solutions to Burgers' PDE from moments of the measure supported on their graph. A (remarkably accurate) approximation of such moments has been obtained by solving semidefinite relaxations of the Moment-SOS hierarchy applied to the weak measure-valued formulation of the Burgers equation (and in a spirit similar to (6.20) for optimal control problems). The role of the Christoffel function is treated in more detail in Section 7, including its remarkable ability to recover a function solely from knowledge of the moments of the (degenerate) measure supported on its graph.

7. Positive polynomials and the Christoffel function

In this section we introduce the Christoffel–Darboux (CD) kernel and the Christoffel function (CF), which are classical tools from the fields of orthogonal polynomials and approximation theory. In addition to being interesting in their own right, they have proved to be useful in understanding and interpreting the Moment-SOS hierarchy of lower bounds. Further, the CF also appears in a certain distinguished representation of polynomials that are positive on a semi-algebraic set $S \subset \mathbb{R}^d$, as in (2.7), extensively used in the Moment-SOS hierarchy. In particular, every SOS polynomial p in the interior of the convex cone $\Sigma[\mathbf{x}]_n$ of degree-$2n$ SOS polynomials is the reciprocal of the CF of some linear functional ϕ in $\mathbb{R}[\mathbf{x}]_{2n}^*$. If $n = 2$, then ϕ has a clear interpretation in terms of a Gaussian measure, but in the general case, the link between p and ϕ is only partially understood and remains to be interpreted.

7.1. Christoffel–Darboux kernel and Christoffel function

The CF is usually defined for a measure μ with moments $\boldsymbol{\mu} = (\mu_\alpha)_{\alpha\in\mathbb{N}^d}$, whose support $S \subset \mathbb{R}^d$ is compact and such that its moment matrix $\mathbf{M}_n(\boldsymbol{\mu})$ (or equivalently, $\mathbf{M}_n(\mu)$) is positive definite for every degree $n \in \mathbb{N}$. However, it can also be defined for a Riesz linear functional $\phi \in \mathbb{R}[\mathbf{x}]^*$ (with $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^d}$) such that $\mathbf{M}_n(\boldsymbol{\phi}) \succ 0$ for every $n \in \mathbb{N}$, not necessarily coming from a measure μ.

So after fixing some ordering of monomials in $\mathbb{N}^d$, and since $\mathbf{M}_n(\boldsymbol{\phi}) \succ 0$ for every n, let $(P_\alpha)_{\alpha\in\mathbb{N}^d}$ be a family of polynomials that are *orthonormal* with respect to ϕ, that is, such that

$$\phi(P_\alpha\, P_\beta) = \delta_{\alpha=\beta} \quad \text{for all } \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^d, \tag{7.1}$$

where $\delta_{\cdot=\cdot}$ is the usual Kronecker symbol (with value 1 if $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and 0 otherwise).

For every $n \in \mathbb{N}$, the Christoffel–Darboux (CD) kernel $K_n^\phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is then defined by

$$(\mathbf{x}, \mathbf{y}) \mapsto K_n^\phi(\mathbf{x}, \mathbf{y}) := \sum_{\alpha\in\mathbb{N}_n^d} P_\alpha(\mathbf{x})\, P_\alpha(\mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,\ n \in \mathbb{N}, \tag{7.2}$$

and the Christoffel function (CF) $\Lambda_n^\phi : \mathbb{R}^d \to \mathbb{R}_+$ is defined by

$$\mathbf{x} \mapsto \Lambda_n^\phi(\mathbf{x}) := K_n^\phi(\mathbf{x}, \mathbf{x})^{-1} \quad \text{for all } \mathbf{x} \in \mathbb{R}^d,\ n \in \mathbb{N}, \tag{7.3}$$

that is, the CF is the reciprocal of the 'diagonal' of the CD kernel. Hence, by construction, $1/\Lambda_n^\phi$ is an SOS polynomial of degree $2n$.

A reproducing property. Let $p \in \mathbb{R}[\mathbf{x}]_n$, and since $(P_\alpha)_{\alpha\in\mathbb{N}_n^d}$ form a basis of $\mathbb{R}[\mathbf{x}]_n$, write

$$\mathbf{x} \mapsto p(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}_n^d} p_\alpha\, P_\alpha(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d,$$

for some vector of coefficients $\mathbf{p} = (p_\alpha)_{\alpha\in\mathbb{N}_n^d}$ in $\mathbb{R}^{s(n)}$ (with $s(n) = \binom{n+d}{d}$).

With $\mathbf{x} \in \mathbb{R}^d$ fixed, $\mathbf{y} \mapsto K_n^\phi(\mathbf{x}, \mathbf{y}) \in \mathbb{R}[\mathbf{y}]_n$, and we have

$$\begin{aligned} \phi(K_n^\phi(\mathbf{x}, \cdot)\, p) &= \phi\Bigg(\Bigg(\sum_{\alpha\in\mathbb{N}_n^d} P_\alpha(\mathbf{x})\, P_\alpha(\mathbf{y})\Bigg) \cdot \Bigg(\sum_{\beta\in\mathbb{N}_n^d} p_\beta\, P_\beta(\mathbf{y})\Bigg)\Bigg) \\ &= \sum_{\alpha\in\mathbb{N}_n^d} p_\alpha\, P_\alpha(\mathbf{x}) = p(\mathbf{x}) \quad \text{for all } p \in \mathbb{R}[\mathbf{x}]_n, \end{aligned} \tag{7.4}$$

where we have used that

$$\phi(p_\beta P_\beta(\mathbf{y})\, P_\alpha(\mathbf{x}) P_\alpha(\mathbf{y})) = p_\beta\, P_\alpha(\mathbf{x})\, \phi(P_\beta\, P_\alpha) = p_\beta\, P_\alpha(\mathbf{x})\, \delta_{\beta=\alpha}.$$

For this reason, if μ is a measure on $S \subset \mathbb{R}^d$, and $L^2(\mu)$ is the Hilbert space of square-integrable functions with respect to μ, with scalar product

$$\langle f, g \rangle = \int_S f\, g \,\mathrm{d}\mu \quad \text{for all } f, g \in L^2(\mu),$$

then $(\mathbb{R}[\mathbf{x}]_n, \langle\cdot,\cdot\rangle) \subset L^2(\mu)$ is called a reproducing kernel Hilbert space (RKHS) with kernel K_n^μ, because

$$\int_S K_n^\mu(\mathbf{x},\mathbf{y})\, p(\mathbf{y})\,\mathrm{d}\mu(\mathbf{y}) = p(\mathbf{x}) \quad \text{for all } p \in \mathbb{R}[\mathbf{x}]_n.$$

Alternative formulations of the CF. The CF can be defined by

$$\Lambda_n^\phi(\boldsymbol{\xi})^{-1} = \mathbf{v}_n(\boldsymbol{\xi})^\top \mathbf{M}_n(\boldsymbol{\phi})^{-1}\,\mathbf{v}_n(\boldsymbol{\xi}) \quad \text{for all } \boldsymbol{\xi} \in \mathbb{R}^d \tag{7.5}$$

(the ABC theorem in Simon (2008)), and it also has the variational formulation

$$\Lambda_n^\phi(\boldsymbol{\xi}) = \min_{p\in\mathbb{R}[\mathbf{x}]_n} \{\phi(p^2)\colon p(\boldsymbol{\xi}) = 1\} \quad \text{for all } \boldsymbol{\xi} \in \mathbb{R}^d. \tag{7.6}$$

In particular, observe that (7.6) can be rewritten as

$$\Lambda_n^\phi(\boldsymbol{\xi}) = \min_{\mathbf{p}\in\mathbb{R}^{s(n)}} \{\mathbf{p}^\top \mathbf{M}_n(\boldsymbol{\phi})\mathbf{p}\colon \langle \mathbf{p}, \mathbf{v}_n(\boldsymbol{\xi})\rangle = 1\} \quad \text{for all } \boldsymbol{\xi} \in \mathbb{R}^d,$$

which is a convex quadratic optimization problem that can be solved efficiently even for large dimension d. After some algebra, the unique optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_n$ of (7.6) reads

$$\mathbf{x} \mapsto p^*(\mathbf{x}) = \frac{K_n^\phi(\boldsymbol{\xi},\mathbf{x})}{K_n^\phi(\boldsymbol{\xi},\boldsymbol{\xi})} \quad \text{for all } \mathbf{x} \in \mathbb{R}^d.$$

7.2. *Some useful properties of the CF*

A crucial property of the CFs $(\Lambda_n^\mu)_{n\in\mathbb{N}}$ associated with a measure μ on a compact set $S \subset \mathbb{R}^d$ is to identify the support of μ. Indeed, its decay with the degree n exhibits the following interesting dichotomy:

- for all $\boldsymbol{\xi} \in \operatorname{supp}(\mu)$, $\Lambda_n^\mu(\boldsymbol{\xi})^{-1}$ grows at most as a polynomial in n,
- for all $\boldsymbol{\xi} \notin \operatorname{supp}(\mu)$, $\Lambda_n^\mu(\boldsymbol{\xi})^{-1}$ grows at least as an exponential in n.

This property has been exploited in data analysis to provide a simple and easy-to-use tool (with no tuning of parameters), e.g. to detect outliers, with performance similar to (and sometimes better than) state-of-the-art techniques; see Lasserre, Pauwels and Putinar (2022) and Lasserre and Pauwels (2016).

Next, let μ have a density f with respect to Lebesgue measure on S. Under some additional regularity properties of μ and its support S,

$$\lim_{n\to\infty} s(n)\,\Lambda_n^\mu(\boldsymbol{\xi}) = f(\boldsymbol{\xi})/\mu_E(\boldsymbol{\xi}), \tag{7.7}$$

uniformly on compact subsets of $\operatorname{int}(S)$, where μ_E is the density of a so-called *equilibrium measure* of S.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Equilibrium measure. A Borel measure μ supported on a compact set $S \subset \mathbb{R}^d$ satisfies the Bernstein–Markov property if there exists a sequence of positive numbers $(M_n)_{n\in\mathbb{N}}$ such that for all n and $p \in \mathbb{R}[\mathbf{x}]_n$,

$$\sup_{\mathbf{x}\in S} |p(\mathbf{x})| \leq M_n \cdot \left(\int_S p^2 \,\mathrm{d}\mu\right)^{1/2} \quad \text{and} \quad \lim_{n\to\infty} \log(M_n)/n = 0 \tag{7.8}$$

(see e.g. Lasserre *et al.* (2022, Section 4.3.3)). The Bernstein–Markov property allows a qualitative description of asymptotics of the Christoffel function as n grows.

The notion of *equilibrium measure*, associated with a given set, originates from logarithmic potential theory (working in $\mathbb{C}$ in the univariate case) to minimize some energy functional. For instance, the (Chebyshev) measure $\mathrm{d}\mu := \mathrm{d}x/\pi\sqrt{1-x^2}$ is the equilibrium measure of the interval $[-1,1]$. Some generalizations have been obtained in the multivariate case via pluripotential theory in $\mathbb{C}^d$. In particular, if $S \subset \mathbb{R}^d \subset \mathbb{C}^d$ is compact then its equilibrium measure (let us denote it by λ_S) is equivalent to Lebesgue measure on compact subsets of $\mathrm{int}(S)$. It has an even explicit expression if S is convex and symmetric about the origin; see e.g. Baran (1995) and Bedford and Taylor (1986, Theorems 1.1, 1.2). Moreover, if μ is a Borel measure on S and (S,μ) has the Bernstein–Markov property (7.8), then the sequence of measures

$$\nu_n = \frac{\mu}{s(n)\Lambda_n^\mu(\mathbf{x})},$$

$n \in \mathbb{N}$, converges to λ_S for the weak-$\star$ topology, and therefore in particular

$$\lim_{n\to\infty}\int_S \mathbf{x}^\alpha \,\mathrm{d}\nu_n = \lim_{n\to\infty}\int_S \frac{\mathbf{x}^\alpha \,\mathrm{d}\mu(\mathbf{x})}{s(n)\Lambda_n^\mu(\mathbf{x})} = \int_S \mathbf{x}^\alpha \,\mathrm{d}\lambda_S \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{N}^d. \tag{7.9}$$

(See e.g. Lasserre *et al.* (2022, Theorem 4.4.4).) In addition, if the compact $S \subset \mathbb{R}^d$ is regular, then (S, λ_S) has the Bernstein–Markov property. For a brief account of equilibrium measures, see Baran (1995), Bedford and Taylor (1986) and the discussion in Lasserre *et al.* (2022, Sections 4–5, pp. 56–60), while for more detailed expositions see some of the references therein.

7.3. The CF for interpolation and approximation

In this section we briefly address the following issue, which is interesting in its own right and also central to the recovery of an optimal (or ε-optimal) trajectory $\{\mathbf{x}(t) \colon t \in [0,1]\}$ in optimal control problems, solely from knowledge of the moments of the occupation measure supported on the graph $\{(t,\mathbf{x}(t)) \colon t \in [0,1]\}$; see Section 6.2.

So with $X \subset \mathbb{R}$, let μ be a measure on $[0,1] \times X$, defined by

$$\mathrm{d}\mu(t,x) = \delta_{\{f(t)\}}(\mathrm{d}x)\, 1_{[0,1]}(t)\,\mathrm{d}t,$$

for some unknown measurable function $f \colon [0,1] \to X$, that is, μ is supported

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

on the graph $\{(t, f(t)) \colon t \in [0,1]\}$ of f. The goal is to recover f solely from knowledge of the moments $\boldsymbol{\mu} = (\mu_{ij})_{(i,j)\in\mathbb{N}^2}$, where

$$\mu_{ij} = \int t^i x^j \, \mathrm{d}\mu(t,x) = \int_0^1 t^i f(t)^j \, \mathrm{d}t, \quad (i,j) \in \mathbb{N}^2.$$

We propose using the Christoffel function Λ_n^μ to recover f from $\boldsymbol{\mu}$ because, as seen earlier, Λ_n^μ is a good tool for identifying the support of μ, and in this case the support is precisely the graph of the unknown function f to recover. Here, observe that μ is a degenerate measure on $[0,1] \times X$, that is, its support has Lebesgue measure zero on $[0,1] \times X$. Therefore its moment matrix $\mathbf{M}_n(\mu)$ can be ill-conditioned and even singular if f is a polynomial (because then the vector of coefficients of $f \in \mathbb{R}[t]$ is in the kernel of $\mathbf{M}_n(\mu)$ when n is sufficiently large). So we first 'perturbate' (or regularize) $\mathbf{M}_n(\mu)$ to $\mathbf{M}_n(\mu) + \varepsilon\mathbf{I}$ with $\mathbf{I}$ the identity matrix and some small regularization parameter $\varepsilon > 0$, and we define a new perturbated Christoffel function $\hat{\Lambda}_n^\mu$ by

$$(t,x) \mapsto \hat{\Lambda}_n^\mu(t,x)^{-1} \coloneqq \mathbf{v}_n(t,x)^\top (\mathbf{M}_n(\mu) + \varepsilon\mathbf{I})^{-1} \mathbf{v}_n(t,x) \quad \text{for all } (t,x) \in \mathbb{R}^2. \tag{7.10}$$

We then define the following n-approximant $f_n \colon [0,1] \to X$ of f by

$$t \mapsto f_n(t) \coloneqq \arg\min_{x \in X} \hat{\Lambda}_n^\mu(t,x)^{-1}, \quad t \in [0,1]. \tag{7.11}$$

(If there are several minimizers in (7.11) then just take the smallest one as a tie-breaker rule.) For every fixed $t \in [0,1]$, $f_n(t)$ can be computed efficiently as $x \mapsto \hat{\Lambda}_n^\mu(t,x)^{-1}$ is a *univariate* SOS polynomial in x.

Next, as n increases, pointwise convergence (except at points of discontinuity) and L^1-norm convergence to f are proved in Marx *et al.* (2021). Observe that the f_n approximant (7.11) is *not* a polynomial, and since it is semi-algebraic, it is able to approximate some discontinuous functions quite well with no Gibbs phenomenon.

For instance, in Figure 7.1(a) we may observe a typical Gibbs phenomenon (oscillations) when approximating the (discontinuous) step function $t \mapsto f(t) = 0$ if $t \in [0,1/2]$ and $f(t) = 1$ if $t \in (1/2,1]$ (in red) by a polynomial $p^* \in \mathbb{R}[t]_n$ (in black) that minimizes the integral of the mean squared error, that is,

$$p^* = \arg \min_{p \in \mathbb{R}[t]_n} \int_0^1 (p - f)^2 \, \mathrm{d}t$$

(even with degree $n = 12$). This L^2-norm approximation of f is a standard application of the CD kernel K_n^ν associated with the univariate measure $\nu = \mathrm{d}t$ on $[0,1]$. On the other hand, with $\varepsilon > 0$ very small and f_n as in (7.11), the step function is recovered almost exactly (in black) with no Gibbs phenomenon and with small degree $n = 4$. This is what we may call a non-standard application of the CD kernel, as we consider the degenerate bivariate measure μ on $[0,1] \times X$ instead of the univariate measure $f(t)\,\mathrm{d}t$ on $[0,1]$.

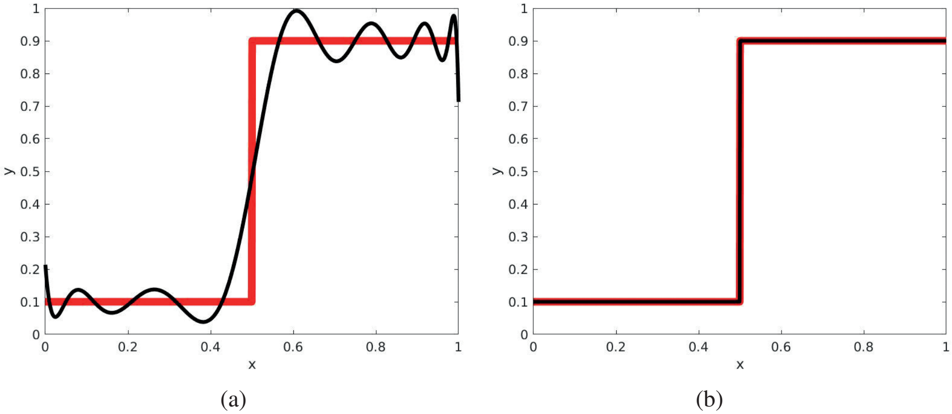

Figure 7.1. (a) Degree-12 polynomial approximation of step function with Gibbs phenomenon. (b) Step function approximated by f_4 in (7.11). © Springer Nature, reproduced with permission from Marx *et al.* (2021).

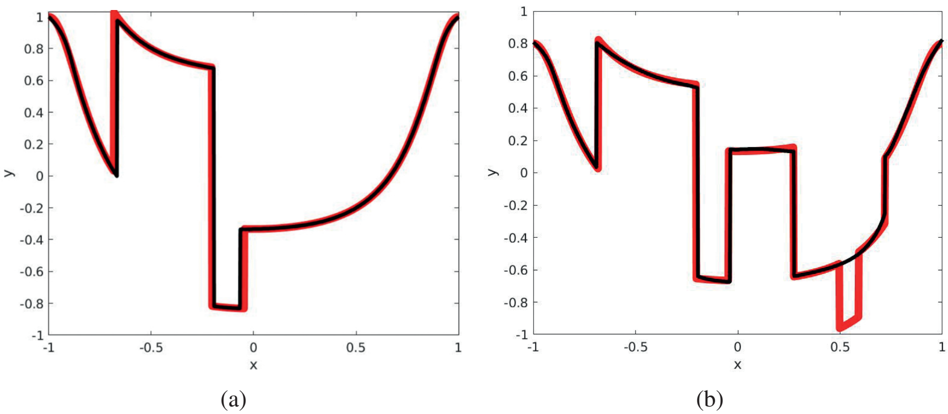

Figure 7.2. Two Eckhoff functions (Eckhoff 1993) approximated by f_{10} in (7.11). © Springer Nature, reproduced with permission from Marx *et al.* (2021).

Similarly, in Figure 7.2, two discontinuous Eckhoff functions from Eckhoff (1993) (in red) are also recovered (in black) with very good precision via f_n in (7.11) with $n = 10$, and again with no Gibbs phenomenon; for more details the reader is referred to Marx *et al.* (2021).

Application to optimal control. As already mentioned in Section 6.2, such an approximation technique can be used to recover the graph of functions supported on trajectories $\{(\mathbf{x}^*(t), \mathbf{u}^*(t)) : t \in [0, 1]\}$, i.e. the optimal solutions of optimal control problems (6.10) described in Section 6.2. Indeed, when applying the Moment-SOS hierarchy to solve (6.10), at an optimal solution of (6.20), we obtain

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

approximate moments up to degree $2n$, of the measure

$$d\mu^*(\mathbf{x},\mathbf{u},t) = \delta_{\{(\mathbf{x}^*(t),\mathbf{u}^*(t))\}}\, 1_{[0,1]}(t)\, dt,$$

supported on the graph of the map $(\mathbf{x}^*,\mathbf{u}^*)\colon [0,1] \to \mathbb{R}^d \times \mathbb{R}^m$.

For instance, to recover the particular trajectory $\{x_i^*(t)\colon t \in [0,1]\}$ for some coordinate $i \in \{1,\dots,d\}$, we proceed as follows.

- We extract the sub-matrix $\mathbf{M}_n^{(x_i,t)}$ of $\mathbf{M}_n(\mu)$ obtained by restricting to rows and columns indexed by monomials $(x_i^k\, t^j)$, $(k,j) \in \mathbb{N}_n^2$ (i.e. $\mathbf{M}_n^{(x_i,t)}$ is the degree-n moment matrix of the marginal μ_i of μ on (x_i,t)).
- We compute the perturbed Christoffel function in (7.10) associated with μ_i, that is,

  $$\hat{\Lambda}_n^{\mu_i}(x_i,t)^{-1} = \mathbf{v}_n(x_i,t)^\top (\mathbf{M}_n^{(x_i,t)} + \varepsilon\,\mathbf{I})^{-1} \mathbf{v}_n(x_i,t),$$

  and then the f_n approximant of the function $x_i(t)$ is obtained via (7.11).

The same procedure is repeated for all coordinates $x_i^*(t)$, $i \in \{1,\dots,d\}$, of $\mathbf{x}^*(t)$, and all coordinates $u_j^*(t)$, $j \in \{1,\dots,m\}$, of $\mathbf{u}^*(t)$, independently.

7.4. *Christoffel function and positive polynomials*

First notice that by construction, the reciprocal of a Christoffel function is an SOS polynomial. Next, with $S \subset \mathbb{R}^d$ as in (2.7), recall the convex cone

$$Q_n(g) := \left\{ \sum_{j=0}^{m} \sigma_j\, g_j \colon \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j},\ j = 0,\dots,m \right\}, \tag{7.12}$$

which is a degree-$2n$ *truncated* version of the quadratic module $Q(g)$ (with $d_j = \lceil \deg(g_j)/2 \rceil$ and $g_0 = \mathbf{1}$). For every polynomial $p = \sum_j \sigma_j g_j \in Q_n(g)$, the SOS weights σ_j provide p with an algebraic certificate of its positivity on S.

Recall that the dual of $Q_n(g)$ is the convex cone

$$Q_n^*(g) = \{\boldsymbol{\phi} \in \mathbb{R}^{s(2n)} \colon \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \succeq 0,\ j = 0,\dots,m\}, \tag{7.13}$$

where $\mathbf{M}_n(g_j \cdot \boldsymbol{\phi})$ is the localizing matrix associated with the polynomial g_j and the sequence $\boldsymbol{\phi}$ (or equivalently the moment matrix associated with the sequence $g_j \cdot \boldsymbol{\phi}$), defined in Section 2.1. We saw in Section 3 that $Q_n(g)$ and its dual $Q_n^*(g)$ are crucial to the construction of the Moment-SOS hierarchy of lower bounds. It turns out that there is a nice one-to-one correspondence between the respective interiors of $Q_n(g)$ and $Q_n^*(g)$, stated in terms of Christoffel functions.

Theorem 7.1. If $p \in \mathrm{int}(Q_n(g))$, then there exists $\boldsymbol{\phi} \in \mathrm{int}(Q_n^*(g))$ such that

$$p(\mathbf{x}) = \sum_{j=0}^{m} \Lambda_{n-d_j}^{g_j\cdot\boldsymbol{\phi}}(\mathbf{x})^{-1}\, g_j(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{7.14}$$

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

or, equivalently,

$$\operatorname{int}(Q_n(g)) = \left\{ \sum_{j=0}^{m} \left(\Lambda_{n-d_j}^{g_j \cdot \boldsymbol{\phi}}\right)^{-1} g_j : \boldsymbol{\phi} \in \operatorname{int}(Q_n^*(g)) \right\}. \tag{7.15}$$

Theorem 7.1 is an interpretation from Lasserre (2022) of a duality result of Nesterov (2000). Remarkably, it states that every p in the interior of $Q_n(g)$ has a distinguished certificate of its positivity on S, with very specific SOS weights $\sigma_j = (\Lambda_{n-d_j}^{g_j \cdot \boldsymbol{\phi}})^{-1}$ in its Putinar's representation (2.9). Indeed, those weights all come from a *unique* element $\boldsymbol{\phi} \in \operatorname{int}(Q_n^*(g))$ and its Christoffel functions associated with the Riesz linear functionals $g_j \cdot \boldsymbol{\phi}$, $j = 0, \ldots, m$. It also turns out that those weights have an extremal property: consider the optimization problem

$$\rho_n = \inf_{\boldsymbol{\phi} \in \mathbb{R}^{s(2n)}} \left\{ -\sum_{j=0}^{m} \log\det(\mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi})) : \right.$$
$$\left. \phi(p) = 1,\ \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \succeq 0,\ \forall\, j = 0, \ldots, m \right\}. \tag{7.16}$$

It is a convex optimization problem which has an explicit dual, namely

$$\rho_n^* = \sup_{\mathbf{Q}_j} \left\{ \sum_{j=0}^{m} \log\det(\mathbf{Q}_j) : \mathbf{Q}_j \succeq 0,\ \forall\, j = 0, \ldots, m \right.$$
$$\left. p(\mathbf{x}) \sum_{j=0}^{m} s(n-d_j) = \sum_{j=0}^{m} g_j(\mathbf{x}) \cdot \mathbf{v}_{n-d_j}(\mathbf{x})^\top \mathbf{Q}_j \mathbf{v}_{n-d_j}(\mathbf{x}),\ \forall\, \mathbf{x} \in \mathbb{R}^d \right\}, \tag{7.17}$$

where the supremum is taken over real symmetric matrices $\mathbf{Q}_j$ of respective size $s(n-d_j)$, $j = 0, \ldots, m$. The criterion to maximize in (7.17) is minus the log-barrier of the convex cone $Q_n(g)$.

Theorem 7.2. With $n \in \mathbb{N}$ fixed, problems (7.16) and (7.17) have the same finite optimal value $\rho_n = \rho_n^*$ if and only if $p \in \operatorname{int}(Q_n(g))$. Moreover, both have a unique optimal solution $\boldsymbol{\phi}_{2n}^* \in \mathbb{R}^{s(2n)}$ and $(\mathbf{Q}_j^*)_{j=0,\ldots,m}$ respectively, which satisfy $\mathbf{Q}_j^* = \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}_{2n}^*)^{-1}$ for all $j = 0, \ldots, m$. As a consequence,

$$\begin{aligned} p(\mathbf{x}) &= \frac{1}{\sum_{j=0}^{m} s(n-d_j)} \sum_{j=0}^{m} g_j(\mathbf{x})\, \mathbf{v}_{n-d_j}(\mathbf{x})^\top \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}_{2n}^*)^{-1} \mathbf{v}_{n-d_j}(\mathbf{x}) \\ &= \frac{1}{\sum_{j=0}^{m} s(n-d_j)} \sum_{j=0}^{m} g_j(\mathbf{x})\, \Lambda_{n-d_j}^{g_j \cdot \boldsymbol{\phi}_{2n}^*}(\mathbf{x})^{-1} \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \end{aligned} \tag{7.18}$$

Notice that $\boldsymbol{\phi}$ in (7.14) is just $(\sum_{j=0}^{m} s(n-d_j))\, \boldsymbol{\phi}_{2n}^*$ with $\boldsymbol{\phi}_{2n}^*$ as in (7.18).

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Of course, Theorem 7.1 immediately raises the following question: Given $p \in \mathrm{int}(Q_n(g))$, what is this linear functional $\phi \in \mathbb{R}[\mathbf{x}]_{2n}^*$ with associated moment sequence $\boldsymbol{\phi} \in \mathbb{R}^{s(2n)}$ in Theorem 7.1? It turns out that there is a simple and remarkable answer for special sets S and the constant polynomial $p = \mathbf{1}$.

Relating the constant polynomial and the equilibrium measure. Let $S \subset \mathbb{R}^d$ in (2.7) be a compact set with non-empty interior, generated by a finite set $\tilde{G} = \{g_1, \ldots, g_m\} \subset \mathbb{R}[\mathbf{x}]$ of polynomials. Let $G \subset \mathbb{R}[\mathbf{x}]$ be a certain finite set of polynomials formed with some products of polynomials in $\tilde{G}$. We provide three examples.

- If $S \subset \mathbb{R}^d$ is the Euclidean unit ball, then $\tilde{G} = \{g\}$, $G = \{\mathbf{1}, g\}$, with $\mathbf{x} \mapsto g(\mathbf{x}) = 1 - \|\mathbf{x}\|^2$. Then the equilibrium measure μ is proportional to

$$\mathrm{d}\mathbf{x}\big(1 - \|\mathbf{x}\|^2\big)^{-1/2}.$$

- If S is the unit box $[-1,1]^d$, then $\tilde{G} = \{g_1, \ldots, g_d\}$ with $g_j(\mathbf{x}) = 1 - x_j^2$, $j = 1, \ldots, d$, and $G = \{g_{\boldsymbol{\varepsilon}} : \boldsymbol{\varepsilon} \in \{0,1\}^d\}$, where

$$\mathbf{x} \mapsto g_{\varepsilon}(\mathbf{x}) := \prod_{j=1}^{d} g_j(\mathbf{x})^{\varepsilon_j} \quad \text{for all } \mathbf{x} \in \mathbb{R}^d.$$

  The equilibrium measure μ of S is proportional to

$$\mathrm{d}\mathbf{x} \prod_{j=1}^{d} \big(1 - x_j^2\big)^{-1/2}.$$

- If $S \subset \mathbb{R}^d$ is the canonical simplex, then $\tilde{G} = \{g_1, \ldots, g_{d+1}\}$ with $g_j(\mathbf{x}) = x_j$, $j = 1, \ldots, d$, $g_{d+1}(\mathbf{x}) = 1 - \sum_j x_j$ and $G = \{g_{\boldsymbol{\varepsilon}} : \boldsymbol{\varepsilon} \in \{0,1\}^{d+1};\ |\boldsymbol{\varepsilon}| \in 2\mathbb{N}\}$, where

$$\mathbf{x} \mapsto g_{\varepsilon}(\mathbf{x}) := \prod_{j=1}^{d+1} g_j(\mathbf{x})^{\varepsilon_j} \quad \text{for all } \mathbf{x} \in \mathbb{R}^d.$$

  The equilibrium measure μ of S is proportional to

$$\mathrm{d}\mathbf{x}\left(\left(1 - \sum_j x_j\right) \prod_{j=1}^{d} x_j\right)^{-1/2}.$$

For every $g \in G$ let $t_g := \lceil \deg(g)/2 \rceil$. In addition, given $n \in \mathbb{N}$, let

$$G_n := \{g \in G : \deg(g) \leq 2n\}$$

so that $G_n = G$ if $n \geq \lceil d/2 \rceil$.

Theorem 7.3 (Lasserre 2023, Lasserre and Xu 2023). Let $S \subset \mathbb{R}^d$ be the Euclidean unit ball, the unit box or the simplex, and let μ be its equilibrium measure.

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Then, for all integer n,

$$1=\frac{1}{\sum_{g\in G_n} s(t-t_g)}\sum_{g\in G_n} g(\mathbf{x})\,\Lambda^{g\cdot\mu}_{n-t_g}(\mathbf{x})^{-1}\quad\text{for all } \mathbf{x}\in\mathbb{R}^d. \tag{7.19}$$

So, remarkably, the constant polynomial $p=\mathbf{1}\in\operatorname{int}(Q_n(g))$, for all n, is strongly related to the equilibrium measure μ of S. Its corresponding element $\boldsymbol{\phi}\in\operatorname{int}(Q^*_n(g))$, in Theorem 7.1, is the moment vector $\boldsymbol{\mu}\in\mathbb{R}^{s(2n)}$ of μ.

In addition, for every n, the polynomials $(g/\Lambda^{g\cdot\mu}_{n-t_g})_{g\in G_n}$ (all non-negative on S) provide S with a polynomial *partition of unity*. We have called (7.19) a *generalized polynomial Pell's equation* solved by the Christoffel functions $(\Lambda^{g\cdot\mu}_{n-t_g})_{g\in G_n}$ (and the polynomials $g\in G_n$), because (7.19) is an exact multivariate generalization of the polynomial Pell's equation[6]

$$1=T_n(x)^2+(1-x^2)\,U_{n-1}(x)^2\quad\text{for all } x\in\mathbb{R}, \tag{7.20}$$

satisfied by the univariate Chebyshev polynomials of the first kind $(T_n)_{n\in\mathbb{N}}$ and Chebyshev polynomials of the second kind $(U_n)_{n\in\mathbb{N}}$, orthogonal with respect to the measures $\mathrm{d}x/\sqrt{1-x^2}$ and $\sqrt{1-x^2}\,\mathrm{d}x$, respectively. Indeed, after normalization to orthonormal polynomials, and summing up (7.20) over n, we obtain

$$1=\frac{1}{s(n)+s(n-1)}\bigl(\Lambda^{\mu}_n(x)^{-1}+g(x)\,\Lambda^{g\cdot\mu}_{n-1}(x)^{-1}\bigr)\quad\text{for all } x\in\mathbb{R},\ n\in\mathbb{N}, \tag{7.21}$$

where $g(x)=1-x^2$, and $\mathrm{d}\mu(x)=\mathrm{d}x/\pi\sqrt{1-x^2}$ is the equilibrium measure of the interval $S=[-1,1]$. The term 'generalized' is justified because in (7.19) we have a sum of squares in $\mathbb{R}[\mathbf{x}]$ and several generators $g\in G_n$, instead of two single squares in $\mathbb{Z}[x]$ and a single generator g in (7.20). But formally, (7.19) is of exactly the same flavour as (7.21).

Remark 7.4. When $S=\mathbb{R}^d$, there is still a nice well-known and somewhat related fact. Let $p\in\mathbb{R}[\mathbf{x}]_2$ be a quadratic polynomial which is strictly positive on $\mathbb{R}^d$. With $\mathbf{v}_1(\mathbf{x})=(1,x_1,\ldots,x_d)$, p is written as

$$\mathbf{x}\mapsto p(\mathbf{x}):=\mathbf{v}_1(\mathbf{x})^\top\mathbf{Q}\,\mathbf{v}_1(\mathbf{x})\quad\text{for all } \mathbf{x}\in\mathbb{R}^d,$$

for some (unique) Gram matrix $\mathbf{Q}\succ 0$. It is well known that $\mathbf{Q}^{-1}$ is the moment matrix $\mathbf{M}_1(\mu)$ of a Gaussian measure μ on $\mathbb{R}^d$, and therefore

$$p(\mathbf{x})=\Lambda^{\mu}_1(\mathbf{x})^{-1}\quad\text{for all } \mathbf{x}\in\mathbb{R}^d.$$

This is another particular case (but in a non-compact setting) where one can identify the linear functional $\boldsymbol{\phi}$ in Theorem 7.1 (now with $Q_1(g)=\Sigma[\mathbf{x}]_1$ and $\Sigma_1[\mathbf{x}]^*_1=\{\boldsymbol{\phi}\in\mathbb{R}^{2d+1}:\mathbf{M}_1(\boldsymbol{\phi})\succeq 0\}$). For instance, the scaled Hermite polynomials of degree at most 1,

$$\widehat{H}_0(\mathbf{x})=(2\pi)^{-d/4},\quad \widehat{H}_j(\mathbf{x})=(2\pi)^{-d/4}x_j,\quad j=1,\ldots,d,$$

[6] A triple (F,g,H) of polynomials in $\mathbb{Z}[x]$ satisfies (polynomial) Pell's equation if $F^2+g\,H^2=1$.

are orthonormal with respect to the Gaussian (product measure)

$$\mathrm{d}\mu(\mathbf{x}) = \exp\biggl(-\sum_j x_j^2/2\biggr)\,\mathrm{d}\mathbf{x},$$

and

$$\Lambda_1^{\mu}(\mathbf{x})^{-1} = \sum_{j=0}^{d} \widehat{H}_j(\mathbf{x})^2 = (2\pi)^{-d/2}(1+\|\mathbf{x}\|^2).$$

7.5. Comparing Moment-SOS hierarchies of upper and lower bounds

To compare the two Moment-SOS hierarchies of upper and lower bounds for solving the POP

$$\mathbf{P}\colon\ f^* = \min\,\{f(\mathbf{x})\colon \mathbf{x}\in S\},$$

we express them in the same language of *polynomial densities* with respect to a reference finite Borel probability measure μ whose support is exactly the set $S\subset\mathbb{R}^d$ (assumed to be compact with non-empty interior). Let $\mathscr{P}(S)$ be the space of probability measures on S, and let $(P_\alpha)_{\alpha\in\mathbb{N}^d}$ be a family of polynomials that are orthonormal with respect to μ.

Moment-SOS hierarchy of lower bounds. With $\phi\in\mathbb{R}[\mathbf{y}]_{2n}^*$ arbitrary, and from the reproducing property (7.4) of K_{2n}^{μ}, observe that

$$\begin{aligned}\phi(f) &= \phi\biggl(\int_S \sum_{\alpha\in\mathbb{N}_{2n}^d} P_\alpha(\mathbf{y})P_\alpha(\mathbf{x})\,f(\mathbf{x})\,\mathrm{d}\mu(\mathbf{x})\biggr)\\ &= \sum_{\alpha\in\mathbb{N}_{2n}^d}\phi(P_\alpha)\int_S P_\alpha(\mathbf{x})\,f(\mathbf{x})\,\mathrm{d}\mu(\mathbf{x})\\ &= \int_S f(\mathbf{x})\biggl(\sum_{\alpha\in\mathbb{N}_{2n}^d}\phi(P_\alpha)\,P_\alpha(\mathbf{x})\biggr)\,\mathrm{d}\mu(\mathbf{x})\\ &= \int_S f(\mathbf{x})\,\sigma_\phi(\mathbf{x})\,\mathrm{d}\mu(\mathbf{x}),\end{aligned}$$

where the degree-$2n$ polynomial

$$\mathbf{x}\mapsto\sigma_{\boldsymbol{\phi}}(\mathbf{x}) := \sum_{\alpha\in\mathbb{N}_{2n}^d}\phi(P_\alpha)\,P_\alpha(\mathbf{x}) \tag{7.22}$$

is a *signed density* with respect to μ.

Next, recall that in the semidefinite relaxation (3.5) of the Moment-SOS hierarchy of lower bounds on f^*, we search for a linear functional $\phi\in\mathbb{R}[\mathbf{x}]_{2n}^*$ that satisfies

$$\phi(\mathbf{1}) = 1,\quad \mathbf{M}_{n-d_j}(g_j\cdot\boldsymbol{\phi})\succeq 0\quad \text{for all } j=0,\ldots,m,$$

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

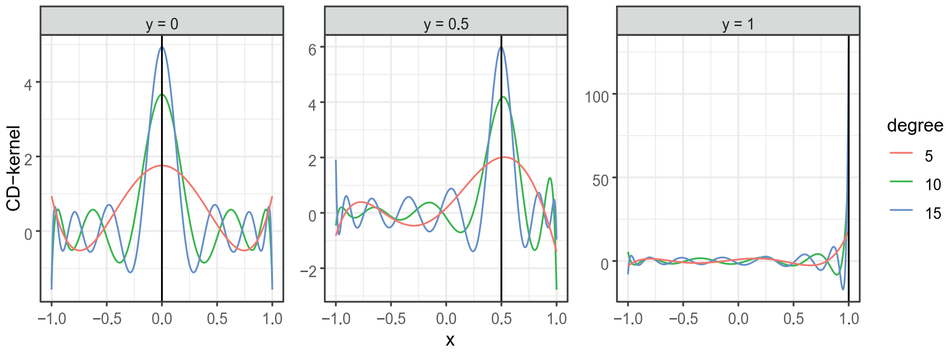

Figure 7.3. $S = [-1, 1]$, $\mu = \mathrm{d}x/2$; the signed density $x \mapsto \sigma_{\phi^*}(x) = K^{\mu}_{2n}(x, y)$ with (a) $y = 0$, (b) $y = 0.5$ and (c) $y = 1$. © Cambridge University Press, reproduced with permission from Lasserre *et al.* (2022).

and that minimizes $\langle \mathbf{f}, \boldsymbol{\phi} \rangle = \phi(f) = \int_S f\, \sigma_{\boldsymbol{\phi}}\, \mathrm{d}\mu$. Moreover, observe that

$$1 = \phi(\mathbf{1}) = \int_S \sigma_\phi \, \mathrm{d}\mu,$$

which means that σ_ϕ is a signed *probability* density. Therefore we have proved the following.

Theorem 7.5. Let μ be a finite Borel measure whose support is S in (2.7), and consider the Moment-SOS hierarchy of semidefinite relaxations (3.5) for solving $\mathbf{P}$. Then, with n fixed, (3.5) reads

$$\min_{\boldsymbol{\phi} \in \mathbb{R}^{s(2n)}} \left\{ \int_S f\, \sigma_\phi \, \mathrm{d}\mu : \phi(\mathbf{1}) = 1;\ \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \succeq 0,\ j = 0, \ldots, m \right\}, \tag{7.23}$$

where σ_ϕ is the signed probability density with respect to μ in (7.22).

So, again, solving the semidefinite relaxation (3.5) in the Moment-SOS hierarchy is searching for a polynomial signed probability density $\sigma_\phi \in \mathbb{R}[\mathbf{x}]_{2n}$ of the form (7.22), and as already mentioned, when the relaxation (3.5) is exact, $\phi^* = \delta_{\{\mathbf{y}\}}$, where $\mathbf{y} \in S$ is a global minimizer of f. Then the associated polynomial signed probability density $\sigma_{\phi^*} \in \mathbb{R}[\mathbf{x}]_{2n}$ reads

$$\mathbf{x} \mapsto \sigma_{\phi^*}(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^d_{2n}} P_\alpha(\mathbf{y})\, P_\alpha(\mathbf{x}) = K^{\mu}_{2n}(\mathbf{y}, \mathbf{x}).$$

It is interesting to see σ_{ϕ^*} in Figure 7.3 for the toy example where $S = [-1, 1]$ and $\mu = \mathrm{d}x/2$. Indeed, σ_{ϕ^*} has a peak at $x = y$, and thus mimics the Dirac measure at y (as far as moments up to degree $2n$ are concerned).

https://doi.org/10.1017/S0962492923000053 Published online by Cambridge University Press

Table 7.1. Hierarchies of upper and lower bounds interpreted as searching for respective positive probability density $\sum_\alpha \sigma_\alpha P_\alpha$ and signed density $\sum_\alpha \phi(P_\alpha)P_\alpha$, with respect to μ.

Lower bounds	Upper bounds
Primal	Primal
$\tau_n = \inf_{\boldsymbol{\phi}} \int_S \big(\sum_{\alpha\in\mathbb{N}^d_{2n}} \phi(P_\alpha)\, P_\alpha\big)\, f\, \mathrm{d}\mu$	$\hat{\kappa}_n = \inf_{\sigma,\psi_j} \int_S \big(\sum_{\alpha\in\mathbb{N}^d_{2n}} \sigma_\alpha\, P_\alpha\big)\, f\, \mathrm{d}\mu$
s.t. $\phi_0\,(=\phi(\mathbf{1})) = 1$;	s.t. $\sigma_0 = 1$;
$\mathbf{M}_n(g_j\cdot\boldsymbol{\phi}) \succeq 0,\ 0\le j\le m.$	$\sum_{\alpha\in\mathbb{N}^d_{2n}} \sigma_\alpha\, P_\alpha = \sum_{j=0}^m \psi_j\, g_j.$
	$\psi_j \in \Sigma[\mathbf{x}]_{n-d_j}, 0\le j\le m$
Dual	Dual
$\tau^*_n = \sup_{\lambda,\psi_j}\ \lambda$	$\hat{\kappa}^*_n = \sup_{\lambda,\boldsymbol{\phi}}\ \lambda$
s.t. $f-\lambda = \sum_{j=0}^m \psi_j\, g_j$	s.t. $f-\lambda = \sum_{\alpha\in\mathbb{N}^d_{2n}} \phi(P_\alpha)\, P_\alpha$
$\psi_j \in \Sigma[\mathbf{x}]_{n-d_j},\ 0\le j\le m$	$\mathbf{M}_n(g_j\cdot\boldsymbol{\phi}) \succeq 0,\ 0\le j\le m$

Comparing with the Moment-SOS hierarchy of upper bounds. Let μ be the same (reference) measure on S as in Theorem 7.5. By construction, the (refined) hierarchy of upper bounds $(\hat{\kappa}_n)_{n\in\mathbb{N}}$ in (5.4) is searching for a *positive* probability density $\sigma\in Q_n(g)$. Hence $\mathbf{x}\mapsto\sigma(\mathbf{x}) = \sum_\alpha \sigma_\alpha\, P_\alpha(\mathbf{x})$, with

$$1 = \int_S \sigma\,\mathrm{d}\mu = \sum_\alpha \sigma_\alpha \int_S P_\alpha\,\mathrm{d}\mu = \sigma_0,$$

as $P_0 = \mathbf{1}$ (because μ is a probability measure), and

$$\sigma\in Q_n(g) \;\Rightarrow\; \sum_{\alpha\in\mathbb{N}^d_{2n}} \sigma_\alpha\, P_\alpha = \sum_{j=0}^m \psi_j\, g_j, \quad \psi_j\in\Sigma[\mathbf{x}]_{n-d_j},\ j=0,\ldots,m.$$

As we can see, Table 7.1 exhibits a complete symmetry between the primal and dual formulations of the respective Moment-SOS hierarchies of lower bounds and upper bounds, when the involved polynomials are expressed in the orthonormal basis $(P_\alpha)_{\alpha\in\mathbb{N}^d}$.

7.6. Notes and sources

Sections 7.1–7.2. For more details and historical background on the CD kernel and the Christoffel function, the interested reader is referred to Simon (2008), Nevai and Freud (1986), Lasserre *et al.* (2022) and references therein.

Section 7.3. This is based on Marx *et al.* (2021) and Henrion and Lasserre (2022).

Section 7.4. This is essentially from Lasserre (2022, 2023) and Lasserre and Xu (2023). Remarkably, the generalized Pell's equation establishes links between seemingly unrelated fields such as optimization, positivity certificates, conic duality on one hand, and orthogonal polynomials and equilibrium measures on the other hand. It is likely that the generalized Pell's equation is valid only for sets with specific geometries and with an appropriate set of generators. Indeed, from the proof in Lasserre and Xu (2023), a property of Gegenbauer polynomials (in particular a summation property) is crucial. However, for more general basic semi-algebraic sets S, there is an even weaker result that links the constant polynomial $\mathbf{1} \in Q_n(g)$ and moments $\boldsymbol{\phi} \in Q_n^*(g)$ of the equilibrium measure μ of S; see Lasserre and Xu (2023).

Section 7.5. Here we interpret both Moment-SOS hierarchies of upper and lower bounds in the common language of densities with respect to a reference measure μ whose support is S. In contrast to the hierarchy of upper bounds, finite convergence for the hierarchy of lower bounds is possible (and in fact takes place generically) because a *signed* density with respect to μ may have all its moments up to order $2n$ equal to those of the Dirac measure at a global minimizer, which is not possible for a *positive* density with respect to μ on S (with non-empty interior) as in the hierarchy of upper bounds.

8. Conclusion

We have described the Moment-SOS hierarchy methodology for polynomial optimization (hierarchies of lower and upper bounds). We have also used it to solve the generalized moment problem (GMP) with algebraic data, whose list of applications in many areas of science and engineering is almost endless. The basic principle behind the Moment-SOS hierarchy is quite simple, and for illustration we have described its application to two problems (viewed as instances of the GMP) in computational geometry and optimal control.

It is a powerful methodology but the computational cost of its basic formulation can be quite heavy, even for problems of modest dimension. Fortunately, large-scale problems often exhibit sparsity or symmetries in their formulation, and we have also described how such properties can be exploited to define a sparsity-adapted Moment-SOS hierarchy whose associated computational burden can be drastically reduced.

Much remains to be done in several research directions, some of which have been briefly mentioned. As it is at the intersection of several disciplines, it is very likely that we will see even more contributions in the coming years.

Acknowledgements
This work is supported by the ANR-NuSCAP 20-CE-48-0014 programme and also the AI Interdisciplinary Institute ANITI funding through the French programme 'Investing for the Future PI3A' under grant agreement ANR-19-PI3A-0004. This research is also part of the DesCartes programme, and is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

A. A. Ahmadi and A. Majumdar (2019), DSOS and SDSOS optimization: More tractable alternatives to sum of squares and semidefinite optimization, *SIAM J. Appl. Algebra Geometry* **3**, 193–230.

D. Aloise and P. Hansen (2011), Evaluating a branch-and-bound RLT-based algorithm for minimum sum-of-squares clustering, *J. Global Optim.* **49**, 449–465.

M. Anjos and J. B. Lasserre, eds (2012), *Handbook on Semidefinite, Conic and Polynomial Optimization*, Vol. 166 of International Series in Operations Research and Management Science, Springer.

E. M. Aylward, P. A. Parrilo and J.-J. E. Slotine (2008), Stability and robustness analysis of nonlinear systems via contraction metrics and SOS programming, *Automatica* **44**, 2163–2170.

F. Bacari, C. Gogolin, P. Wittek and A. Acín (2020), Verifying the output of quantum optimizers with ground-state energy lower bounds, *Phys. Rev. Res.* **2**, art. 043163.

F. Bach and A. Rudi (2023), Exponential convergence of Sum-of-Squares hierarchies for trigonometric polynomials, *SIAM J. Optim.* **33**, 2137–2159.

C. Bachoc and F. Vallentin (2008), New upper bounds for kissing numbers from semidefinite programming, *J. Amer. Math. Soc.* **21**, 909–924.

C. Bachoc, A. Passuello and F. Vallentin (2013), Bounds for projective codes from semidefinite programming, *Adv. Math. Commun.* **7**, 127–145.

M. Bafna, B. Barak, P. K. Kothari, T. Schramm and D. Steurer (2021), Playing unique games on certified small-set expanders, in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, Association for Computing Machinery (ACM), pp. 1629–1642.

L. Baldi and B. Mourrain (2022), Exact moment representation in Polynomial Optimization. Available at arXiv:2012:14652.

B. Barak and D. Steurer (2014), Sum-of-squares proofs and the quest towards optimal algorithms, in *Proceedings of the International Congress of Mathematicians (ICM 2014)* (S. Y. Sang *et al.*, eds), Kyung Moon Sa, pp. 509–533.

M. Baran (1995), Complex equilibrium measure and Bernstein type theorems for compact sets in $\mathbb{R}^n$, *Proc. Amer. Math. Soc.* **123**, 485–494.

E. Bedford and B. A. Taylor (1986), The complex equilibrium measure of a symmetric convex set in $\mathbb{R}^n$, *Trans. Amer. Math. Soc.* **294**, 705–717.

D. Bertsimas and I. Popescu (2005), Optimal inequalities in probability theory: A convex optimization approach, *SIAM J. Optim.* **15**, 780–804.

G. Blekherman, P. Parrilo and R. Thomas, eds (2012), *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series on Optimization, SIAM.

S. Burgdorf, I. Klep and J. Povh (2016), *Optimization of Polynomials in Non-Commuting Variables*, SpringerBriefs in Mathematics, Springer.

Chen Tong, J. B. Lasserre, V. Magron and E. Pauwels (2021), Semialgebraic representation of monotone deep equilibrium models and application to certification, in *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)* (M. Ranzato *et al.*, eds), Curran Associates, pp. 27146–27159.

F. Cicconet and K. C. Almeida (2019), Moment-SOS relaxation of the medium term hydrothermal dispatch problem, *Int. J. Elec. Power Energy Systems* **104**, 124–133.

N. E. Courtier, R. Drummond, P. Ascensio, L. D. Couto and D. A. Howey (2022), Discretisation-free battery fast-charging optimisation using the measure-moment approach, in *2022 European Control Conference (ECC)*, IEEE, pp. 628–634.

Y. de Castro, F. Gamboa, D. Henrion and J. B. Lasserre (2017), Exact solutions to super resolution on semi-algebraic domains in higher dimensions, *IEEE Trans. Inform. Theory* **63**, 621–630.

Y. de Castro, F. Gamboa, D. Henrion, R. Hess and J. B. Lasserre (2019), Approximate optimal designs for multivariate polynomial regression, *Ann. Statist.* **47**, 125–157.

E. de Klerk and M. Laurent (2011), On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems, *SIAM J. Optim.* **21**, 824–832.

D. de Laat (2020), Moment methods in energy minimization: New bounds for Riesz minimal energy problems, *Trans. Amer. Math. Soc.* **373**, 1407–1453.

D. de Laat, F. C. Machado, F. M. de Oliveira Filho and F. Vallentin (2022), k-point semidefinite programming bounds for equiangular lines, *Math. Program.* **194**, 533–567.

A. C. Doherty, P. A. Parrilo and F. Spedalieri (2005), Detecting multipartite entanglement, *Phys. Rev. A* **71**, art. 032333.

M. Dostert and F. Vallentin (2020), New dense superball packings in three dimensions, *Adv. Geom.* **20**, 473–482.

G. R. Dowdy and P. I. Barton (2018), Dynamics bounds on stochastic chemical kinetic systems using semidefinite programming, *J. Chem. Phys.* **149**, art. 74103.

M. E. Dyer and A. M. Frieze (1988), The complexity of computing the volume of a polyhedron, *SIAM J. Comput.* **17**, 967–974.

K. S. Eckhoff (1993), Accurate and efficient reconstruction of discontinuous functions from truncated series expansions, *Math. Comp.* **61**(204), 745–763.

J. Eisert, P. Hyllus, O. Gühne and M. Curty (2004), Complete hierarchies of efficient approximations to problems in entanglement theory, *Phys. Rev. A* **70**, art. 062317.

J. Fan, J. Nie and A. Zhou (2018), Tensor eigenvalue complementarity problems, *Math. Program. Ser. A* **170**, 507–539.

K. Fang and H. Fawzi (2021), The sum-of-squares hierarchy on the sphere and applications in quantum information theory, *Math. Program.* **190**, 331–360.

G. Fantuzzi, D. Goluskin, D. Huang and S. I. Chernyshenko (2016), Bounds for deterministic and stochastic dynamical systems using sum-of-squares optimization, *SIAM J. Appl. Dyn. Systems* **15**, 1962–1988.

H. O. Fattorini (1999), *Infinite Dimensional Optimization*, Cambridge University Press.

A. Gepp, G. Harris and B. Vanstone (2020), Financial applications of semidefinite programming: A review and call for interdisciplinary research, *Account. Finance* **60**, 3527–3555.

M. X. Goemans and D. P. Williamson (1995), Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. Assoc. Comput. Mach.* **42**, 1115–1145.

D. Goluskin and G. Fantuzzi (2019), Bounds on mean energy in the Kuramoto–Sivashinsky equation computed using semidefinite programming, *Nonlinearity* **32**, 1705–1730.

D. Handelman (1988), Representing polynomials by positive linear functions on compact convex polyhedra, *Pacific J. Math.* **132**, 35–62.

D. Haussmann, F. Liers, M. Stingl and J. C. Vera (2018), Deciding robust feasibility and infeasibility using a set containment approach: An application to stationary passive gas network operations, *SIAM J. Optim.* **28**, 2489–2517.

D. Henrion and A. Garulli, eds (2005), *Positive Polynomials in Control*, Vol. 312 of Lecture Notes on Control and Information Sciences, Springer.

D. Henrion and J. B. Lasserre (2022), Graph recovery from incomplete moment information, *Constr. Approx.* **56**, 165–187.

D. Henrion, M. Korda and J. B. Lasserre (2021), *The Moment-SOS Hierarchy: Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs*, Vol. 4 of Optimization and Applications, World Scientific.

D. Henrion, J. B. Lasserre and C. Savorgnan (2009), Approximate volume and integration for basic semi-algebraic sets, *SIAM Rev.* **51**, 722–743.

Jie Pan and Fu Jiang (2020), Low complexity beamspace super resolution for DOA estimation of linear array, *Sensors* **20**, art. 2222.

S. Khot (2010), Inapproximability of NP-complete problems, discrete Fourier analysis, and geometry, in *Proceedings of the 2010 International Congress of Mathematicians (ICM 2010)* (R. Bhatia *et al.*, eds), Vol. IV, Hindustan Book Agency, pp. 2676–2697.

S. Khot (2014), Hardness of approximation, in *Proceedings of the 2014 International Congress of Mathematicians (ICM 2014)* (S. Y. Sang *et al.*, eds), Kyung Moon Sa, pp. 711–728.

M. Kojima, S. Kim and M. Maramatsu (2005), Sparsity in sums of squares of squares of polynomials, *Math. Program.* **103**, 45–62.

M. Korda, D. Henrion and C. N. Jones (2014), Convex computation of the maximum controlled invariant set for polynomial control systems, *SIAM J. Control Optim.* **52**, 2944–2969.

M. Korda, D. Henrion and J. B. Lasserre (2022), Moments and convex optimization for analysis and control of nonlinear PDEs, in *Numerical Control: Part A* (E. Trélat and E. Zuazua, eds), Vol. 23 of Handbook of Numerical Analysis, Elsevier, pp. 339–366.

M. Kočvara, B. Mourrain and C. Riener, eds (2023), *Polynomial Optimization, Moments, and Applications*, Springer Optimization and its Applications, Springer.

J. L. Krivine (1964*a*), Anneaux préordonnés, *J. Anal. Math.* **12**, 307–326.

J. L. Krivine (1964*b*), Quelques propriétés des préordres dans les anneaux commutatifs unitaires, *C.R. Acad. Sci. Paris* **258**, 3417–3418.

H. Landau (1987), Classical background of the moment problem, in *Moments in Mathematics*, Vol. 37 of Proceedings of Symposia in Applied Mathematics, American Mathematical Society (AMS), pp. 1–15.

R. Laraki and J. B. Lasserre (2012), Semidefinite programming for min-max problems and games, *Math. Program. Ser. A* **131**, 305–332.

J. B. Lasserre (2000), Optimisation globale et théorie des moments, *C.R. Acad. Sci. Paris, Sér. I* **331**, 929–934.

J. B. Lasserre (2001), Global optimization with polynomials and the problem of moments, *SIAM J. Optim.* **11**, 796–817.

J. B. Lasserre (2002*a*), Bounds on measures satisfying moment conditions, *Ann. Appl. Prob.* **12**, 1114–1137.

J. B. Lasserre (2002*b*), Semidefinite programming vs. LP relaxations for polynomial programming, *Math. Oper. Res.* **27**, 347–360.

J. B. Lasserre (2006), Convergent SDP-relaxations in polynomial optimization with sparsity, *SIAM J. Optim.* **17**, 822–843.

J. B. Lasserre (2009*a*), Convexity in semi-algebraic geometry and polynomial optimization, *SIAM J. Optim.* **19**, 1995–2014.

J. B. Lasserre (2009*b*), *Moments, Positive Polynomials and their Applications*, Imperial College Press.

J. B. Lasserre (2011), A new look at nonnegativity on closed sets and polynomial optimization, *SIAM J. Optim.* **21**, 864–885.

J. B. Lasserre (2013), The K-moment problem for continuous functionals, *Trans. Amer. Math. Soc.* **365**, 2489–2504.

J. B. Lasserre (2015), *An Introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge University Press.

J. B. Lasserre (2017), Computing Gaussian & exponential measures of semi-algebraic sets, *Adv. Appl. Math.* **91**, 137–163.

J. B. Lasserre (2021), Connecting optimization with spectral analysis of tri-diagonal matrices, *Math. Program. Ser. A* **190**, 795–809.

J. B. Lasserre (2022), A disintegration of the Christoffel function, *Comptes Rendus Math.* **360**, 1071–1079.

J. B. Lasserre (2023), Pell's equation, sum-of-squares and equilibrium measures on a compact set, *Comptes Rendus Math.* **361**, 935–952.

J. B. Lasserre and E. Pauwels (2016), Sorting out typicality via the inverse moment matrix SOS polynomial, in *Advances in Neural Information Processing Systems 29 (NIPS2016)* (D. Lee *et al.*, eds), Curran Associates, pp. 190–198.

J. B. Lasserre and T. Priéto-Rumeau (2004), SDP vs. LP relaxations for the Moment approach in some performance evaluation problems, *Stoch. Models* **20**, 439–456.

J. B. Lasserre and Y. Xu (2023), A generalized Pell's equation for a class of multivariate orthogonal polynomials. Technical report, LAAS-CNRS, Toulouse, France. Available at hal-04163153v1. To appear in *Trans. Amer. Math. Soc.*

J. B. Lasserre, D. Henrion, C. Prieur and E. Trélat (2008), Nonlinear optimal control via occupation measures and LMI-relaxations, *SIAM J. Control Optim.* **47**, 1649–1666.

J. B. Lasserre, E. Pauwels and M. Putinar (2022), *The Christoffel Function for Data Analysis*, Cambridge University Press.

J. B. Lasserre, T. Priéto-Rumeau and M. Zervos (2006), Pricing a class of exotic options via moments and SDP relaxations, *Math. Finance* **16**, 469–494.

F. Latorre, P. Rolland and V. Cevher (2020), Lipschitz constant estimation for neural networks via sparse polynomial optimization, in *8th International Conference on Learning Representations (ICLR 2020)*, OpenReview.

M. Laurent (2003), A comparison of the Sherali–Adams, Lovász–Schrijver and Lasserre relaxations for 0-1 programming, *Math. Oper. Res.* **28**, 470–496.

M. Laurent (2008), Sums of squares, moment matrices and optimization over polynomials, in *Emerging Applications of Algebraic Geometry* (M. Putinar and S. Sullivant, eds), Vol. 149 of The IMA Volumes in Mathematics and its Applications, Springer, pp. 157–270.

M. Laurent and L. Slot (2023), An effective version of Schmüdgen's Positivstellensatz for the hypercube, *Optim. Lett.* **17**, 515–530.

V. Magron and J. Wang (2023), *Sparse Polynomial Optimization: Theory and Practice*, Series on Optimization and its Applications, World Scientific.

A. Majumdar, G. Hall and A. A. Ahmadi (2020), A survey of recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics, *Annu. Rev. Control Robot Auton. Syst.* **3**, 331–60.

A. Marmin, M. Castella, J.-C. Pesquet and L. Duval (2021), Sparse signal reconstruction for nonlinear models via piecewise rational optimization, *Signal Process.* **179**, art. 107835.

Z. Marschner, D. Palmer, P. Zhang and J. Solomon (2020), Hexahedral mesh repair via sum-of-squares relaxations, *Comput. Graph. Forum* **39**, 133–147.

Z. Marschner, D. Palmer, P. Zhang and J. Solomon (2021), Sum-of-squares geometry processing, *ACM Trans. Graphics* **40**, 1–13.

M. Marshall (2008), *Positive Polynomials and Sums of Squares*, Vol. 146 of AMS Math. Surveys and Monographs, American Mathematical Society (AMS).

S. Marx, E. Pauwels, T. Weisser, D. Henrion and J. B. Lasserre (2021), Semi-algebraic approximation using Christoffel–Darboux kernel, *Constr. Approx.* **54**, 391–429.

S. Marx, T. Weisser, D. Henrion and J. B. Lasserre (2020), A moment approach for entropy solutions to nonlinear hyperbolic PDEs, *Math. Control Related Fields* **10**, 113–140.

D. Molzahn and C. Josz (2018), Lasserre hierarchy for large-scale polynomial optimization in real and complex variabales, *SIAM J. Optim.* **28**, 1017–1048.

D. Molzhan and I. Hiskens (2015), Sparsity-exploiting moment-based relaxations of the optimal power flow problem, *IEEE Trans. Power Systems* **30**, 3168–3180.

K. Moussa, M. Fiacchini and M. Alamir (2020), Robust optimal scheduling of combined chemo- and immunotherapy: Considerations on chemotherapy detrimental effects, in *Proceedings of the 2020 American Control Conference (ACC)*, IEEE, pp. 4252–4257.

M. Navascués, S. Pironio and A. Acín (2012), SDP relaxations for non-commutative polynomial optimization, in *Handbook on Semidefinite, Conic and Polynomial Optimization* (M. Anjos and J. B. Lasserre, eds), Vol. 166 of International Series in Operations Research and Management Science, Springer, pp. 601–634.

Y. Nesterov (2000), Squared functional systems and optimization problems, in *High Performance Optimization* (H. Frenk *et al.*, eds), Kluwer, pp. 405–440.

P. Nevai and G. Freud (1986), Orthogonal polynomials and christoffel functions, *J. Approx. Theory* **48**, 3–167.

Ngoc Hoang Anh Mai, J. B. Lasserre and V. Magron (2023), A hierarchy of spectral relaxations for polynomial optimization, *Math. Program. Comput* **15**, 651–701.

Ngoc Hoang Anh Mai, J. B. Lasserre, V. Magron and J. Wang (2022), Exploiting constant trace property in large scale polynomial optimization, *ACM Trans. Math. Software* **48**, art. 40.

J. Nie (2009), Sum of squares method for sensor network localization, *Comput. Optim. Appl.* **43**, 151–179.

J. Nie (2013), Certifying convergence of Lasserre's hierarchy via flat truncation, *Math. Program. Ser. A* **142**, 485–510.
J. Nie (2014), Optimality conditions and finite convergence of Lasserre's hierarchy, *Math. Program. Ser. A* **146**, 97–121.
J. Nie (2017), Low rank symmetric tensor approximations, *SIAM J. Matrix Anal. Appl.* **38**, 1517–1540.
J. Nie (2023), *Moment and Polynomial Optimization*, Vol. 31 of MOS/SIAM Series in Optimization, SIAM.
J. Nie and J. Demmel (2008), Sparse SOS relaxations for minimizing functions that are summations of small polynomials, *SIAM J. Optim.* **19**, 1534–1558.
J. Nie and X. Tang (2023), Convex generalized nash equilibrium problems and polynomial optimization, *Math. Program.* **198**, 1485–1518.
J. Nie and Z. Yang (2020), Hermitian tensor decompositions, *SIAM J. Matrix Anal. Appl.* **41**, 1115–1144.
R. O'Donnell (2017), SoS is not fully automatizable, even approximately, in *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)* (C. H. Papadimitriou, ed.), Vol. 67 of Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, pp. 59:1–59:10.
O. Parekh (2023), Synergies between operations research and quantum information science, *INFORMS J. Comput.* **35**, 266–273.
O. Parekh and K. Thompson (2021), Application of the level-2 quantum Lasserre hierarchy in quantum approximation algorithms, in *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)* (N. Bansal *et al.*, eds), Vol. 198 of Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, pp. 102:1–102:20.
P. A. Parrilo (2000), Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. PhD thesis, California Institute of Technology, Pasadena, CA.
P. A. Parrilo (2003), Semidefinite programming relaxations for semialgebraic problems, *Math. Program.* **96**, 293–320.
P. A. Parrilo and S. Lall (2003), Semidefinite programming relaxations and algebraic optimization in control, *Eur. J. Control* **9**, 307–321.
E. Pauwels, D. Henrion and J. B. Lasserre (2017), Positivity certificates in optimal control, in *Geometric and Numerical Foundations of Movements* (J. P. Laumond *et al.*, eds), Vol. 117 of Springer Tracts in Advanced Robotics, Springer, pp. 113–132.
A. Prestel and C. N. Denzel (2001), *Positive Polynomials*, Springer Monographs in Mathematics, Springer.
T. Probst, D. D. Paudel, A. Chhatkuli and L. Van Gool (2019), Convex relaxations for consensus and non-minimal problems in 3D vision, in *Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV)*, IEEE, pp. 10233–10242.
M. Putinar (1993), Positive polynomials on compact semi-algebraic sets, *Indiana Univ. Math. J.* **42**, 969–984.
P. Raghavendra, T. Schramm and D. Steurer (2018), High dimensional estimation via sum-of-squares proofs, in *Proceedings of the International Congress of Mathematicians (ICM 2018)* (B. Sirakov *et al.*, eds), World Scientific, pp. 3389–3423.
D. M. Rosen, L. Carlone, A. S. Bandera and J. J. Leonard (2019), Se-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group, *Int. J. Robotics Research* **38**, 95–125.

K. Schmüdgen (1991), The K-moment problem for compact semi-algebraic sets, *Math. Ann.* **289**, 203–206.

K. Schmüdgen (2017), *The Moment Problem*, Graduate Texts in Mathematics, Springer.

M. Schweighofer (2005), On the complexity of Schmüdgen's Positivstellensatz, *J. Complexity* **20**, 529–543.

S. Sedighi, K. V. Mishra, B. Shankar and M. R. B. Ottersten (2021), Localization with 1-Bit passive radars in Narrow Band IoT-applications using multivariate polynomial optimization, *IEEE Trans. Signal Process.* **69**, 2525–2540.

J. H. Selby, A. B. Sainz, V. Magron, L. Czekaj and M. Horodecki (2023), Correlations constrained by composite measurements, *Quantum* **7**, art. 1080.

H. D. Sherali and W. P. Adams (1990), A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems, *SIAM J. Discrete Math.* **3**, 411–430.

H. D. Sherali and W. P. Adams (1999), *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*, Kluwer.

N. Z. Shor (1987), Quadratic optimization problems, *Tekhnicheskaya Kibernetika* **1**, 128–139.

B. Simon (2008), The Christoffel–Darboux kernel, in *Perspectives in Partial Differential Equations, Harmonic Analysis and Applications*, Vol. 79 of Proceedings of Symposia in Pure Mathematics, American Mathematical Society (AMS), pp. 295–335.

L. Slot (2022), Sum-of-Ssquares hierarchies for polynomial optimization and the Christoffel–Darboux kernel, *SIAM J. Optim.* **32**, 2612–2635.

L. Slot and M. Laurent (2021), Near optimal analysis of Lasserre's univariate measure-based bounds for multivariate polynomial optimization, *Math. Program.* **188**, 443–460.

N. Stein, A. Ozdaglar and P. A. Parrilo (2008), Separable and low-rank continuous games, *Int. J. Game Theory* **37**, 457–474.

M. Tacchi, J. B. Lasserre and D. Henrion (2023), Stokes, Gibbs and volume computation of semi-algebraic sets, *Discrete Comput. Geom.* **69**, 260–283.

M. Tacchi, T. Weisser, J. B. Lasserre and D. Henrion (2021), Exploiting sparsity in semi-algebraic set volume computation, *Found. Comput. Math.* **22**, 161–209.

J. Tian, H. Wei and J. Tan (2015), Global optimization for power dispatch problems based on theory of moments, *Int. J. Electr. Power Energy Syst.* **71**, 184–194.

M. Tyburec, J. Zeman, M. Kruzik and D. Henrion (2021), Global optimality in minimum compliance topology optimization of frames and shells by moment-sum-of-squares hierarchy, *Struct. Multidiscipl. Optim.* **64**, 1963–1981.

F.-H. Vasilescu (2003), Spectral measures and moment problems, in *Spectral Theory and its Applications*, Vol. 2 of Theta Series in Advanced Mathematics, Theta, pp. 173–215.

R. Vinter (1993), Convex duality and nonlinear optimal control, *SIAM J. Control Optim.* **31**, 518–538.

S. Waki, S. Kim, M. Kojima and M. Maramatsu (2006), Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity, *SIAM J. Optim.* **17**, 218–242.

J. Wang, H. Li and B. Xia (2019), A new sparse SOS decomposition algorithm based on term sparsity, in *Proceedings of the 2019 International Symposium on Symbolic and Algebraic Computation (ISSAC '19)*, Association for Computing Machinery (ACM), pp. 347–354.

J. Wang, V. Magron and J. B. Lasserre (2021), TSSOS: A moment-SOS hierarchy that exploits term sparsity, *SIAM J. Optim.* **31**, 30–58.

J. Wang, V. Magron, J. B. Lasserre and Ngoc Hoang Anh Mai (2022), CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization, *ACM Trans. Math. Software* **48**, 1–26.

P. Wittek, S. Darányi and G. Nelhans (2017), Ruling out static latent homophily in citation networks, *Scientometrics* **110**, 765–777.

Xiangfeng Ji, Xuegang (Jeff) Ban, Jian Zhang and Bin Ran (2019), Moment-based travel time reliability assessment with Lasserre's relaxation, *Transp. B* **7**, 401–422.

H. Yang and L. Carlone (2020), In perfect shape: Certifiably optimal 3D shape reconstruction from 2D landmarks, in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 618–627.

H. Yang and L. Carlone (2023), Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization, *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 2816–2834.

H. Yang and M. Pavone (2023), Object pose estimation with statistical guarantees: Conformal keypoint detection and geometric uncertainty propagation, in *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 8947–8958.

H. Yang, J. Shi and L. Carlone (2020), TEASER: Fast and certifiable point cloud registration, *IEEE Trans. Robotics* **37**, 314–333.

A. Yurtsever, J. A. Tropp, O. Fercoq, M. Udell and V. Cevher (2021), Scalable semidefinite programming, *SIAM J. Math. Data Sci.* **3**, 171–200.