



HAL
open science

The Moment-SOS hierarchy: Applications and related topics

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. The Moment-SOS hierarchy: Applications and related topics. *Acta Numerica*, 2024, 33, pp.841-908. 10.1017/S0962492923000053 . hal-04687963

HAL Id: hal-04687963

<https://laas.hal.science/hal-04687963v1>

Submitted on 4 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THE MOMENT-SOS HIERARCHY: APPLICATIONS AND RELATED TOPICS

JEAN B. LASSERRE

ABSTRACT. The Moment-SOS hierarchy initially 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 be also applied to solve problems with positivity constraints “ $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in S$ ” and/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 in two other applications as well for illustration purpose. Finally we also introduce the Christoffel function and reveal its links with the Moment-SOS hierarchy and positive polynomials.

1. INTRODUCTION

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

$$(1) \quad \mathbf{P} : \quad f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in S \},$$

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

$$(2) \quad S := \{ \mathbf{x} \in \mathbb{R}^d : g_j(\mathbf{x}) \geq 0, j = 1, \dots, m \},$$

for some polynomials $g_j, j = 1, \dots, m$. Importantly, the description of \mathbf{P} is entirely *algebraic* via its polynomial data $f, g_j, j = 1, \dots, m$, which is a crucial feature. (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)? The answer depends on the meaning of f^* in (1). If one is interested in a *local* minimum only then the whole arsenal of efficient methods of NLP can be used for solving \mathbf{P} and its algebraic features are not really exploited.

On the other hand, if f^* in (1) is understood as the *global* minimum of \mathbf{P} then the situation is totally different. Why? First, to eliminate any ambiguity on the meaning of f^* , rewrite (1) as:

$$(3) \quad \mathbf{P} : \quad f^* = \sup \{ \lambda : f(\mathbf{x}) - \lambda \geq 0, \forall \mathbf{x} \in S \},$$

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

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 (3) one needs 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:

- (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 SDP (which can be done in time polynomial in the input size of the polynomial, up to arbitrary fixed precision¹).

After the pioneers works of [123] and [92], [54, 55] and [106, 107] have been the first to provide a systematic use of these two key ingredients in Optimization and Control, with convergence guarantees. It is also worth mentioning another closely related pioneer work, namely the celebrated SDP-relaxation of [32] which provides a 0.878 approximation guarantee for MAXCUT, 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. Lasserre hierarchy) when applied to the MAXCUT problem. This 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-relaxations 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 Conjecture² (UGC) in Theoretical Computer Science (TCS).

Another (and equivalent) “definition” of the global optimum f^* of \mathbf{P} reads:

$$(4) \quad f^* = \inf_{\phi \in \mathcal{M}(S)_+} \left\{ \int_S f d\phi : \phi(S) = 1 \right\},$$

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

$$\int_S f d\phi \geq \int_S f^* d\phi = f^*,$$

¹In fact see [102] for an update and more details

²For this conjecture and its theoretical and practical implications, S. Khot was awarded the prestigious Nevanlinna prize at the ICM 2014 in Seoul [41, 42].

so the infimum in (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 $\int_S f d\phi$ with ϕ being the Dirac probability measure $\delta_{\{\mathbf{x}\}}$ at $\mathbf{x} \in S$, and so the infimum in (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 (4).

In fact (3) is the LP dual of (4), with λ being the dual variable associated with the constraint $\phi(S) = 1$ in (4). In other words, standard LP duality between the two conic programs (4) and (3) 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 later).

Moreover, Problem (4) is a very particular instance (and even the simplest instance) of the more general *Generalized Moment Problem* (GMP) defined by:

$$(5) \quad \inf_{\phi_j \in \mathcal{M}(S_j)_+} \left\{ \sum_{j=1}^p \int_{S_j} f_j d\phi_j : \sum_{j=1}^p f_{ij} d\phi_j \geq b_i, i = 1, \dots, s \right\},$$

for some given functions $f_j, f_{ij} : \mathbb{R}^{d_j} \rightarrow \mathbb{R}$, $i = 1, \dots, s$, and sets $S_j \subset \mathbb{R}^{d_j}$, $j = 1, \dots, p$. The GMP (5) is an infinite-dimensional LP with dual:

$$(6) \quad \sup_{\lambda_1, \dots, \lambda_s \geq 0} \left\{ \sum_{i=1}^s \lambda_i b_i : f_j - \sum_{i=1}^s \lambda_i f_{ij} \geq 0 \text{ on } S_j, j = 1, \dots, p \right\}.$$

Therefore it should be of 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 crossroad of several area of mathematics [52] and the list of important applications of the GMP is almost endless; see e.g. [52, 60], references therein, and see also Section 6 where for illustration purpose we describe two particular applications.

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

- (i): its basic application in many (and diverse) areas after modeling the problem as an instance of the GMP. For illustration purpose two examples are described in Section 6; see also [37, 47] and references therein.
- (ii): its adaptation and extension to other domains, e.g. Operations Research [104], and entanglement, violation of Bell inequalities in quantum information where its non-commutative version (the Pironio-Navascues-Acin (NPA) hierarchy described in [91]) is also attracting a lot of attention; see also [16] and [105] and references therein.
- (iii): its detailed analysis by the TCS research community for hardness of approximation in combinatorial optimization (e.g. in relation with issues around the Unique Game Conjecture). See e.g. [11], [114] and [9].
- (iv): analysis of its rate of convergence with very interesting recent results on specific sets; see e.g. [126, 125, 6] and references therein.
- (v): development of algorithmic improvements to improve scalability of the standard Moment-SOS hierarchy. One direction is to take into account several

types of sparsity and/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. [1, 144, 79].

Structure of the paper. For ease of exposition and clarity, the proof of most results in the form of theorems and lemmas is not provided. However, at the end of each section one has included a *Notes & Sources* subsection 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 one describes the Moment-SOS hierarchy of *lower bounds* which converges to the global minimum in polynomial optimization. In Section 4 one provides a brief description of an alternative, the Moment-LP hierarchy. Section 5 describes the (less known) Moment-SOS hierarchy of *upper bounds* which also converges to the global minimum. Section 6 is devoted to other applications of the Moment-SOS hierarchy. For illustration purpose one describes how to apply the Moment-SOS hierarchy in two such applications and provides a (non exhaustive) list of references to other ones 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 links and connections with optimization and the Moment-SOS hierarchy.

2. NOTATION, DEFINITIONS AND SOME PRELIMINARIES

2.1. Notation, 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 $\mathbb{N}_n^d := \{\alpha \in \mathbb{N}^d : |\alpha| (= \sum_i \alpha_i) \leq n\}$, and let $\mathbf{v}_n(\mathbf{x}) = (\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}_n^d}$, be the vector of monomials of the canonical basis (\mathbf{x}^α) 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 nonnegative 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 $\langle \mathbf{B}, \mathbf{C} \rangle := \text{trace}(\mathbf{B}\mathbf{C})$ while the notation $\mathbf{B} \geq 0$ stands for \mathbf{B} is positive semidefinite (psd) whereas $\mathbf{B} > 0$ stands for \mathbf{B} is positive definite (pd). Denote by \mathcal{S}^n the space of real $n \times n$ symmetric matrices and \mathcal{S}_+^n its subset of positive semidefinite matrices.

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

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 $\phi : \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R}$ (not in bold) defined

by:

$$(7) \quad f (= \sum_{\alpha} f_{\alpha} \mathbf{x}^{\alpha}) \quad \mapsto \quad \phi(f) = \sum_{\alpha \in \mathbb{N}^d} f_{\alpha} \phi_{\alpha} = \langle \mathbf{f}, \boldsymbol{\phi} \rangle,$$

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,

$$\phi_{\alpha} = \int_{\mathbb{R}^d} \mathbf{x}^{\alpha} d\phi, \quad \forall \alpha \in \mathbb{N}^d.$$

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 $g \cdot \boldsymbol{\phi}$ defined by

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

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

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

In particular, if $\boldsymbol{\phi}$ has a representing measure ϕ and g is nonnegative, then the Riesz linear functional $g \cdot \phi$ is a representing measure, i.e.,

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

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 denoted $\mathbf{M}_n(\boldsymbol{\phi})$ (or $\mathbf{M}_n(\phi)$) with rows and columns indexed by \mathbb{N}_n^d , and whose entry (α, β) is just $\phi_{\alpha+\beta}$, for every $\alpha, \beta \in \mathbb{N}_n^d$. So $\mathbf{M}_n(\boldsymbol{\phi})$ depends only on moments ϕ_{α} of degree at most $2n$. Alternatively, if one introduces the real symmetric matrices $(\mathbf{B}_{\alpha}^1) \subset \mathcal{S}^{s(n)}$ defined by

$$(8) \quad \mathbf{v}_n(\mathbf{x}) \mathbf{v}_n(\mathbf{x})^T = \sum_{\alpha \in \mathbb{N}_{2n}^d} \mathbf{B}_{\alpha}^1 \mathbf{x}^{\alpha}, \quad \forall \mathbf{x} \in \mathbb{R}^d,$$

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

A measure whose all moments are finite, is *moment determinate* if there is no other measure with 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}_n^d , and whose entry (α, β) is just $(g \cdot \boldsymbol{\phi})_{\alpha+\beta}$, that is, $\mathbf{M}_n(g \cdot \boldsymbol{\phi})(\alpha, \beta) = \sum_{\gamma} g_{\gamma} \phi_{\alpha+\beta+\gamma}$, for every $\alpha, \beta \in \mathbb{N}_n^d$.

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

$$(9) \quad g(\mathbf{x}) \mathbf{v}_n(\mathbf{x}) \mathbf{v}_n(\mathbf{x})^T = \sum_{\alpha \in \mathbb{N}_{2(n+d_g)}^d} \mathbf{B}_{\alpha}^g \mathbf{x}^{\alpha}, \quad \forall \mathbf{x} \in \mathbb{R}^d,$$

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

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

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

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 $f_1, \dots, f_s \in \mathbb{R}[\mathbf{x}]$, such that $f(\mathbf{x}) = \sum_{k=1}^s f_k(\mathbf{x})^2$, for all $\mathbf{x} \in \mathbb{R}^d$. Denote by $\Sigma[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x}]_n$) the set of SOS polynomials (resp. SOS polynomials of degree at most $2n$). Of course every SOS polynomial is nonnegative. However the converse is not true.

Membership in $\Sigma[\mathbf{x}]_n$. Checking whether a given polynomial f is nonnegative 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 let $f \in \mathbb{R}[\mathbf{x}]_{2n}$ (for f to be SOS its degree must be even), $\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_{2n}^d} f_\alpha \mathbf{x}^\alpha$. Then $f \in \mathbb{R}[\mathbf{x}]_{2n}$ is SOS if and only if there exists a real symmetric matrix $\mathbf{X}^T = \mathbf{X}$ of size $s(n) = \binom{d+n}{d}$, such that:

$$(10) \quad \mathbf{X} \geq 0; \quad f_\alpha = \langle \mathbf{X}, \mathbf{B}_\alpha^1 \rangle, \quad \forall \alpha \in \mathbb{N}_{2n}^d,$$

where the matrices \mathbf{B}_α^1 have been introduced in (8). It turns out that (10) defines the feasible set of what is called a *semidefinite program*³ (in short, SDP).

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

$$(11) \quad Q(g) := \left\{ \sum_{j=0}^m \sigma_j g_j : \sigma_j \in \Sigma[\mathbf{x}], j = 0, \dots, m \right\},$$

and its degree- $2n$ truncated version

$$(12) \quad Q_n(g) := \left\{ \sum_{j=0}^m \sigma_j g_j : \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j}, j = 0, \dots, m \right\},$$

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 (12), $\deg(\sigma_j g_j) \leq 2n$, 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

$$(13) \quad S := \{ \mathbf{x} \in \mathbb{R}^d : g_j(\mathbf{x}) \geq 0, \quad j = 1, \dots, m \}$$

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.

³A semidefinite program (SDP) 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. [3] and also [102].

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

$$(14) \quad Q_n^*(g) = \{ \phi \in \mathbb{R}^{s(2n)} : \mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0, j = 0, \dots, m \},$$

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

For more details on the above notions of moment and localizing matrix, quadratic module, as well as their use in potential applications, the interested reader is referred to [76] and [60]. 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 which are important because they provide a theoretical justification (or rationale) behind convergence of the so-called SDP- and LP-relaxations for global optimization. In particular, the one below in (15) is at the core of the Moment-SOS hierarchy of lower bounds.

Theorem 2.2 ([113]). *Let $S \subset \mathbb{R}^d$ be as in (13) 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,*

$$(15) \quad f = \sum_{j=0}^m \sigma_j g_j,$$

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

(ii) *A sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^d} \subset \mathbb{R}$ 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 $j = 0, \dots, m$. Equivalently, if and only if $\mathbf{M}_n(g_j \cdot \phi) \geq 0$ for all $j = 0, \dots, m$, and all $n \in \mathbb{N}$.*

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

Theorem 2.3 ([116]). *Let the basic semi-algebraic set $S \subset \mathbb{R}^d$ in (13) be compact.*

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

$$(16) \quad f = \sum_{\alpha \in \{0,1\}^m} \sigma_\alpha g_1^{\alpha_1} \cdots g_m^{\alpha_m},$$

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

(ii) *A sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^d} \subset \mathbb{R}$ 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 $\alpha \in \{0, 1\}^m$. Equivalently, if and only if $\mathbf{M}_n(g_1^{\alpha_1} \cdots g_m^{\alpha_m} \cdot \phi) \geq 0$ for all $\alpha \in \{0, 1\}^m$, and all $n \in \mathbb{N}$.*

Observe that (15) is of the same flavor as (16) 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 (16)). On the other hand, in Theorem 2.3 the only condition is that the set S is compact whereas in Theorem 2.2 one also requires that the quadratic module $Q(g)$ is Archimedean (an additional condition on the representation of S).

The reader may have noticed that Theorem 2.2 and Theorem 2.3 have two facets (i) and (ii): The former is the algebraic facet (certificate of positivity) while the latter with a real analysis flavor 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 nonnegativity) of the author [61] 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 $\text{supp}(\phi) = S$, and with moment sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^d}$. If $f \in \mathbb{R}[\mathbf{x}]$ then*

$$(17) \quad f \geq 0 \text{ on } S \quad \Leftrightarrow \quad \mathbf{M}_n(f \cdot \phi) \geq 0, \quad \forall n \in \mathbb{N}.$$

Theorem 2.4 states that to decide whether f is nonnegative on S , one must check whether countably many Linear Matrix Inequalities (LMI) $\mathbf{M}_n(f \cdot \phi) \geq 0$, $n \in \mathbb{N}$, hold. Each constraint $\mathbf{M}_n(f \cdot \phi) \geq 0$ is indeed an LMI on the coefficients of the polynomial f because each entry of $\mathbf{M}_n(f \cdot \phi)$ 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

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

is a spectrahedron that contains the convex cone $\mathcal{P}_k(S)$ of polynomials of degree at most k that are nonnegative 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 Theorem 2.2 and 2.3, Theorem 2.4 is valid for arbitrary compact sets $S \subset \mathbb{R}^d$ and nonnegative (as opposed to positive) polynomials on S . On the other hand, its practical use to over-approximate the convex cone $\mathcal{P}_k(S)$ by Ω_n , requires knowledge of moments $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^d}$ of a measure ϕ with $\text{supp}(\phi) = S$. 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 $g_1, \dots, g_m \in \mathbb{R}[\mathbf{x}]$, introduce the notation $\mathbf{g}^\alpha \in \mathbb{R}[\mathbf{x}]$, and $(1 - \mathbf{g}^\alpha) \in \mathbb{R}[\mathbf{x}]$, with:

$$\begin{aligned} \mathbf{x} \mapsto \mathbf{g}^\alpha(\mathbf{x}) &:= g_1(\mathbf{x})^{\alpha_1} \cdots g_m(\mathbf{x})^{\alpha_m}, \quad \forall \mathbf{x} \in \mathbb{R}^d, \\ \mathbf{x} \mapsto (1 - \mathbf{g}^\alpha)(\mathbf{x}) &:= (1 - g_1(\mathbf{x}))^{\alpha_1} \cdots (1 - g_m(\mathbf{x}))^{\alpha_m}, \quad \forall \mathbf{x} \in \mathbb{R}^d, \end{aligned}$$

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

$$(18) \quad \mathcal{L}_n(g) := \left\{ \sum_{(\alpha, \beta) \in \mathbb{N}_n^{2m}} c_{\alpha\beta} \mathbf{g}^\alpha (1 - \mathbf{g})^\beta : \mathbf{c} = (c_{\alpha\beta}) \geq 0 \right\}.$$

Theorem 2.5 ([48, 49, 133]). *Let $S \subset \mathbb{R}^d$ as in (13) be compact and such that (possibly after scaling) $0 \leq g_j(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in S$, $j = 1, \dots, m$. Assume also that $[1, g_1, \dots, g_m]$ 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:

$$(19) \quad f = \sum_{(\alpha, \beta) \in \mathbb{N}^{2m}} c_{\alpha\beta} \mathbf{g}^\alpha (\mathbf{1} - \mathbf{g})^\beta,$$

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

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

Remark 2.6. Interestingly, as for Theorems 2.2-2.3, Theorem 2.5 has also 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, Theorem 2.2(i) and (2.5)(i) have another distinguishing feature. Both have a *practical implementation* that allows to perform interesting computations. Indeed:

- Testing membership in $Q_n(g)$ is just solving a single SDP, whereas
- testing membership in $\mathcal{L}_n(g)$ is just solving a single Linear Program (LP).

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

$$(20) \quad f_\alpha = \sum_{j=0}^m \langle \mathbf{X}_j, \mathbf{B}_\alpha^{g_j} \rangle, \quad \forall \alpha \in \mathbb{N}_{2n}^d; \quad \mathbf{X}_j \in \mathcal{S}^{n-d_j};$$

$$(21) \quad \mathbf{X}_j \geq 0, \quad j = 0, \dots, m.$$

where the real symmetric matrices $\mathbf{B}_\alpha^{g_j} \in \mathcal{S}^{s(n-d_j)}$ are defined in (9) (here with $n-d_j$ instead of n). Each real symmetric matrix \mathbf{X}_j is a Gram matrix of a polynomial σ_j , $j = 0, \dots, m$. Next, (20) are *linear* equality constraints on the unknown entries of \mathbf{X}_j , while (21) is a positive semidefinite constraint on $(\mathbf{X}_j)_{j=0}^m$ to ensure that every σ_j is an SOS. (In (20), $f_\alpha = 0$ whenever $k < |\alpha| \leq 2n$ because $f \in \mathbb{R}[\mathbf{x}]_k$.) Observe that multiplying (20) by \mathbf{x}^α and summing up yield

$$\begin{aligned} f(\mathbf{x}) &= \sum_{\alpha \in \mathbb{N}_{2n}^d} f_\alpha \mathbf{x}^\alpha = \sum_{j=0}^m \langle \mathbf{X}_j, \sum_{\alpha \in \mathbb{B}_{2n}^d} \mathbf{B}_\alpha^{g_j} \mathbf{x}^\alpha \rangle \\ &= \sum_{j=0}^m \langle \mathbf{X}_j, \mathbf{v}_{n-d_j}(\mathbf{x}) \mathbf{v}_{n-d_j}(\mathbf{x})^T \rangle g_j(\mathbf{x}) \quad [\text{see (9)}] \\ &= \sum_{j=0}^m \underbrace{\mathbf{v}_{n-d_j}(\mathbf{x})^T \mathbf{X}_j \mathbf{v}_{n-d_j}(\mathbf{x})}_{\sigma_j(\mathbf{x})} g_j(\mathbf{x}). \end{aligned}$$

Hence checking whether (20)-(21) has a solution reduces to solving an SDP.

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 a linear programming problem (LP). Indeed, with n such that $s := (\max_j \deg(g_j))^n \geq k$, it amounts to find a nonnegative vector $\mathbf{c} = (c_{\alpha,\beta})$, $(\alpha, \beta) \in \mathbb{N}_n^{2m}$, such that

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

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

2.5. Notes and sources. Most of the material is from [60, 63]. A good reference for exhaustive results on positive polynomials and moment problems is [76, 85, 111, 117, 98, 44]; see also [15] for related material on convex algebraic geometry.

Section 2.3. Theorem 2.4 is from [61, 62]. Interestingly, it is also valid for some non-compact sets like e.g. 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 $d\phi = \exp(-\sum_i x_i) d\mathbf{x}$ while for the latter one may choose the gaussian measure $d\phi = \exp(-\|\mathbf{x}\|^2/2) d\mathbf{x}$. In both cases one obtains a monotone sequence $(\Omega_n)_{n \in \mathbb{N}}$ of outer-approximations which converges to \mathbb{R}_+^d and \mathbb{R}^d respectively.

Section 2.4. It is worth mentioning that other certificates of positivity (via convex cones of positive polynomials) have been also defined to overcome (or at least mitigate) the computational burden associated with testing membership in $Q_n(g)$ in (20)-(21) (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 [1]; see also [81]. 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 Moment-SOS hierarchy which can handle non-convex POPs with more than a thousand variables; see also [78].

3. THE MOMENT-SOS HIERARCHY IN OPTIMIZATION

Consider the polynomial optimization problem (POP) \mathbf{P} in (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.

When viewing \mathbf{P} in its equivalent formulation (3) (real algebraic glasses).

step 1:: One replaces the hard constraint “ $f - \lambda \geq 0$ on S ” with the equivalent constraint $f \in Q(g)$ (with $Q(g)$ being the quadratic module defined in (11)).

Indeed

$$(22) \quad \begin{aligned} f^* &= \sup_{\lambda} \{ \lambda : f - \lambda \geq 0 \text{ on } S \} = \sup_{\lambda} \{ \lambda : f - \lambda > 0 \text{ on } S \} \\ &= \sup_{\lambda} \{ \lambda : f - \lambda \in Q(g) \}, \end{aligned}$$

where the second equality follows from Theorem 2.2(i) if the quadratic module $Q(g)$ is Archimedean. However, (22) 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 $n \geq n_0$, one replaces (22) with the more restrictive constraint

$$(23) \quad \tau_n^* = \sup_{\lambda} \{ \lambda : f - \lambda \in Q_n(g) \} \quad (n \geq n_0)$$

$$(24) \quad = \sup_{\lambda, \sigma_j} \{ \lambda : f - \lambda = \sum_{j=0}^m \sigma_j g_j ; \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j}, \forall j \},$$

so that $f^* \geq \tau_n^*$ for all $n \geq n_0$. A crucial feature of (24) is to be a finite-dimensional convex optimization problem, and more precisely a *semidefinite program*. Therefore (24) can be solved (up to arbitrary fixed precision) in time polynomial in its input size.

So in solving (23) for increasing values of $k \in \mathbb{N}$, one obtains a monotone non decreasing sequence $(\tau_n^*)_{n \geq n_0}$ of lower bounds on the global minimum of f^* of \mathbf{P} . When viewing \mathbf{P} in its equivalent formulation (4) (real analysis glasses). First observe that since $f \in \mathbb{R}[\mathbf{x}]$,

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

Therefore

$$\begin{aligned} f^* &= \inf_{\phi \in \mathcal{M}(S)_+} \left\{ \int f d\phi : \phi(S) = 1 \right\} \\ &= \inf_{\phi = (\phi_{\alpha})_{\alpha \in \mathbb{N}^d}} \left\{ \langle \mathbf{f}, \phi \rangle : \phi_0 = 1 ; \right. \\ &\quad \left. \phi \text{ has a representing measure supported on } S \right\}. \end{aligned}$$

Again one proceeds in two steps:

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

$$(25) \quad f^* = \inf_{\phi = (\phi_{\alpha})} \{ \langle \mathbf{f}, \phi \rangle : \phi_0 = 1 ; \mathbf{M}_n(g_j \cdot \phi) \geq 0, j = 0, \dots, m, n \in \mathbb{N} \}$$

is still infinite-dimensional.

Step 2:: Next, one then replaces (25) with its truncated versions

$$(26) \quad \tau_n = \inf_{\phi = (\phi_{\alpha})} \{ \langle \mathbf{f}, \phi \rangle : \phi_0 = 1 ; \mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0, j = 0, \dots, m \},$$

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 , (26) is a finite-dimensional semidefinite program.

Clearly, $\tau_n \leq \tau_{n+1} \leq f^*$ for all $n \geq n_0$, and therefore in solving (26) for increasing values of $n \in \mathbb{N}$, one obtains 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 (24) is the dual of the semidefinite program (26), and by *weak duality* in convex optimization,

$$(27) \quad \tau_n^* \leq \tau_n \leq f^*, \quad \forall n \geq n_0.$$

As it is clear from its formulation, (26) is a semidefinite *relaxation* of \mathbf{P} as its constraints are only necessary conditions for ϕ to have a representing measure ϕ on S . We call (26) a *Moment-relaxation* of \mathbf{P} .

On the other hand, its dual (24) is a *reinforcement* (or *strengthening*) of \mathbf{P} (viewed as the maximization problem (3)) as one has replaced “ $f - \lambda \geq 0$ on S ” with the sufficient condition “ $f - \lambda \in Q_n(g)$ ”. We call (24) an *SOS-reinforcement* (or *SOS-strengthening*) of \mathbf{P} , whence the name “Moment-SOS hierarchy” for (26)-(24). In both cases one obtains a lower bound τ_n (resp. τ_n^*) on f^* .

When $\tau_n = f^*$ (resp. $\tau_n^* = f^*$) for some n , one says that the degree- n Moment-relaxation (26) (resp. the degree- n SOS-reinforcement (24)) of \mathbf{P} is *exact*. In addition, if $\tau_n = f^*$ and an optimal solution ϕ^* of (26) satisfies $\text{rank}(\mathbf{M}_n(\phi^*)) = 1$ then ϕ^* is nothing less than 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 (26) is to obtain at some step n , an optimal solution $\phi^* \in \mathbb{R}^{s(2n)}$ 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 \mathbf{P} has a unique global minimizer $\mathbf{x}^* \in S$, this happens generically; in case of finely many global minimizers it will happen generically that at some step n , ϕ^* is be 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 just explained above, it is more appropriate to call (24) an *SOS-reinforcement of \mathbf{P}* instead of an *SOS-relaxation of \mathbf{P}* as is sometimes written in the literature. Of course, it is also the dual of the *Moment-relaxation* (26) of \mathbf{P} . (Relaxing in the primal is equivalent to reinforcing in the dual.)

(b) We call (3) (resp. (26)) a *primal formulation* (resp. a *primal semidefinite relaxation*) of \mathbf{P} because in solving \mathbf{P} one is mainly interested in an optimal solution $\mathbf{x}^* \in S$ (a global minimizer of \mathbf{P}), and 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 \mathbf{P} then $\phi^* = ((\mathbf{x}^*)^\alpha)_{\alpha \in \mathbb{N}^d}$ is an optimal solution of (25).

Then (23) is the dual of the semidefinite relaxation (26), and one will see that indeed when $\tau_n^* = f^*$ for some n and (23) has an optimal solution $\sigma_j^* \in \Sigma[\mathbf{x}]_{k-d_j}$, $j = 0, \dots, m$, then $\sigma_j^*(\mathbf{x}^*) = \lambda_j^* \geq 0$ for all $j = 1, \dots, m$, where the λ_j^* are optimal KKT-Lagrange multipliers associated with a global minimizer $\mathbf{x}^* \in S$.

Computational considerations. The Moment-relaxation (26) 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 (24) 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 d . Therefore in principle they can be solved efficiently (up to arbitrary fixed precision) in time polynomial in their input size⁴. However, in view of their non-modest size and the current status of semidefinite solvers, such semidefinite programs can be solved only for POPs of modest dimension d and small degree- n relaxation. So in its canonical form (26)-(24), the Moment-SOS hierarchy is limited to POPs of modest dimension. However and fortunately:

- Practice reveals that finite convergence often occurs at low degree n .
- As is often the case for many POP of large dimension d , some *sparsity pattern* and/or *symmetries* are present. It turns out that they can be exploited to define appropriate Moment-relaxations (resp. SOS-strengthenings) of \mathbf{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 (24) and (26) by first-order methods; see e.g. [144], [80].

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 always recommended to indeed add the redundant constraint $g_1(\mathbf{x}) := M^2 - \|\mathbf{x}\|^2 \geq 0$ in the definition (13) of S . Moreover, to avoid possible numerical ill-conditioning if M is large, it is even recommended to scale problem \mathbf{P} in such a manner that $S \subset \mathbf{B}_1 := \{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$ so that $g_1(\mathbf{x}) = 1 - \|\mathbf{x}\|^2$. Hence:

Assumption 3.2. $S \subset \mathbb{R}^d$ defined in (13) is compact with $\mathbf{x} \mapsto g_1(\mathbf{x}) = 1 - \|\mathbf{x}\|^2$ (so that $S \subset \mathbf{B}_1$).

The reason to do that is because under Assumption 3.2, the quadratic module $Q(g)$ in (11) is guaranteed to be Archimedean, a crucial property for convergence

⁴More precisions can be found in e.g. [102]

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 (13), let Assumption 3.2 hold and let τ_n (resp. τ_n^*) be as in (26) (resp. (24)) for all $n \geq n_0$. Then:*

(a) $\tau_n^* = \tau_n$ for all $n \geq n_0$, and for every $n \geq n_0$, the semidefinite relaxation (26) has an optimal solution $\phi^n = (\phi_\alpha^n)_{\alpha \in \mathbb{N}_{2n}^d}$. Moreover, both sequences $(\tau_n^*)_{n \geq n_0}$ and $(\tau_n)_{n \geq n_0}$ are monotone non decreasing, and

$$(28) \quad \lim_{n \rightarrow \infty} \tau_n^* = \lim_{n \rightarrow \infty} \tau_n = f^* .$$

(b) With ϕ^n an optimal solution of (26), let $v := \max_{j=1, \dots, m} [\deg(g_j)/2]$. If

$$(29) \quad \text{rank } \mathbf{M}_t(\phi^n) = \text{rank } \mathbf{M}_{t-v}(\phi^n) (=: s) ,$$

for some $v \leq t \leq n$, then $\tau_n^* = \tau_n = f^*$ and from the vector ϕ^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 \mathbf{P} , that is, $f(\mathbf{x}^*(\ell)) = f^*$, $\ell = 1, \dots, s$.

(c) If $\text{int}(S) \neq \emptyset$ then for every $n \geq n_0$, the SOS-strengthening (24) of \mathbf{P} has an optimal solution $(\tau_n^*, \sigma_0^*, \dots, \sigma_m^*)$.

Convergence of minimizers. Theorem 3.3 states that the sequence $(\tau_n)_{n \geq n_0}$ of optimal values converges to the global minimum f^* of \mathbf{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 (29) holds. But what about the sequence of minimizers $(\phi^n)_{n \geq n_0}$ in case when the convergence is only asymptotic (as opposed to finite)?

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

$$(30) \quad \lim_{n \rightarrow \infty} \phi_\alpha^n (= \lim_{n \rightarrow \infty} \phi^n(\mathbf{x}^\alpha)) = (\mathbf{x}^*)^\alpha, \quad \forall \alpha \in \mathbb{N}^d .$$

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

So in case when \mathbf{P} has a unique global minimizer and convergence of the Moment-relaxation (26) 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 (30) is also interesting even if finite convergence takes place at some n because one may already obtain a good approximation of $\mathbf{x}^* \in S$ from the degree-1 moments of ϕ^t for $t < n$.

Equality constraints. Of course in (13) one may tolerate equality constraints $g_j(\mathbf{x}) \geq 0$ and $g_{j+1}(\mathbf{x}) \geq 0$ 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 $g_{j+1} \geq 0$). The resulting modifications are as follows:

- In (24) the unknown SOS weight $\sigma_j \in \Sigma[\mathbf{x}]_{k-d_j}$ is now a polynomial in $\mathbb{R}[\mathbf{x}]_{2(n-d_j)}$ and not an SOS anymore.
- In (26) the psd constraint $\mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0$ now reads as the equality constraints $\mathbf{M}_{n-d_j}(g_j \cdot \phi) = 0$ on the variables (ϕ_α) .

- Theorem 3.3(a)-(b) remains valid, whereas Theorem 3.3(c) needs some adjustment since $\text{int}(S) = \emptyset$. For instance, if the ideal $\langle g_j \rangle_{j \in J} \subset \mathbb{R}[\mathbf{x}]$ generated by the polynomials g_j 's associated with the equality constraints, is real radical, then for n sufficiently large, the SOS-strengthening (24) of \mathbf{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

$$(31) \quad S = \{ \mathbf{x} \in \mathbb{R}^d : x_j^2 - 1 = 0, j \in J; g_j(\mathbf{x}) \geq 0, j \notin J \},$$

of which the celebrated Maxcut problem is a particular case (no inequality constraint). Then the ideal $\langle x_j^2 - 1 \rangle_{j \in J} = \langle x_1^2 - 1, \dots, x_d^2 - 1 \rangle$ 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 (26) is not interesting as it contains 2^d variables ϕ_α . But the interest of Theorem 3.3 is that (29) may take place for $n \ll d$. For instance in most random instances of Maxcut problems with $d = 11$, one observes $f^* = \tau_2$ (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, one recovers the global optimum f^* , and moreover by Theorem 3.3(b), *finite* convergence takes place whenever the so-called *flatness condition* (29) holds at some degree n . In the latter case we can say more. Indeed and remarkably, one is able to provide a global optimality-condition for non-convex POPs, of the same flavor as the celebrated KKT-optimality conditions for convex optimization, and under the same second-order sufficiency condition.

One first recalls 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 (13), 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 $\lambda^* \in \mathbb{R}_+^m$ such that

$$(32) \quad \nabla f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*) = 0; \quad \lambda_j^* g_j(\mathbf{x}^*) = 0, \quad j = 1, \dots, m.$$

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

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

is convex, nonnegative, and satisfies $\nabla L(\mathbf{x}^*) = 0$. Hence \mathbf{x}^* is also a global 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) (32) and strict complementarity hold at $(\mathbf{x}^*, \lambda^*)$, and (ii) in addition,

$$(33) \quad \langle \mathbf{u}, \nabla^2 f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla^2 g_j(\mathbf{x}^*) \rangle > 0, \quad \forall \mathbf{u} \in (\mathbf{x}^*)^\perp.$$

If (32), strict complementarity and (33) hold at a global minimizer $\mathbf{x}^* \in S$ of \mathbf{P} , then one obtains a remarkable certificate of global optimality.

Theorem 3.5 (Certificate of global optimality). *With S as in (13), let $\mathbf{x}^* \in S$ be a global minimizer of \mathbf{P} , and assume that:*

(i) *The gradients $\nabla g_j(\mathbf{x}^*)$, $j \in I(\mathbf{x}^*)$, are linearly independent (so that (32) holds for some $\lambda^* \in \mathbb{R}_+^m$) and strict complementarity holds at $(\mathbf{x}^*, \lambda^*)$.*

(ii) *Second-order sufficient condition (33) holds at $(\mathbf{x}^*, \lambda^*)$.*

Then there exists $n \in \mathbb{N}$ such that the SOS-strengthening (24) is exact, i.e.:

$$(34) \quad f(\mathbf{x}) - f^* = \sum_{j=0}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^d;$$

$$(35) \quad \sigma_j^*(\mathbf{x}^*) g_j(\mathbf{x}^*) = 0, \quad j = 0, \dots, m,$$

for some SOS polynomials $\sigma_j^* \in \Sigma[\mathbf{x}]_{n-d_j}$. Moreover, let $\hat{\lambda} = (\hat{\lambda}_j) \in \mathbb{R}_+^m$ with $\hat{\lambda}_j := \sigma_j^*(\mathbf{x}^*)$, $j = 1, \dots, m$. Then the couple $(\mathbf{x}^*, \hat{\lambda}) \in S \times \mathbb{R}_+^m$ satisfies (32) and (33).

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

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 (32) implies*

$$(36) \quad \mathbf{x} \mapsto L(\mathbf{x}) = f(\mathbf{x}) - f^* - \sum_{j=1}^m \lambda_j^* g_j(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^d;$$

$$(37) \quad L(\mathbf{x}^*) = 0.$$

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 .*

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,*

$$(38) \quad \hat{L}(\mathbf{x}) = f(\mathbf{x}) - f^* - \sum_{j=1}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x}) = \sigma_0^*(\mathbf{x}) (\geq 0), \quad \forall \mathbf{x} \in \mathbb{R}^d;$$

$$(39) \quad \hat{L}(\mathbf{x}^*) = 0.$$

So in (38) the extended Lagrangian \hat{L} looks like the standard Lagrangian L in (36) except that the scalar weight λ_j^* is now replaced with 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 (36)).

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 one removes that constraint in the definition (13) 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 (34) 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, one may wonder how "generic" are the results of Theorem 3.5. It turns out that Theorem 3.5 holds generically in a rigorous sense. More precisely:

Let $m \in \mathbb{N}$ and $r_j \in \mathbb{N}$, $j = 0, \dots, m$, be fixed, and consider the family of POPs whose (possibly empty) feasible set $S \subset \mathbb{R}^d$ is as in (13) 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 $\theta \in \mathbb{R}^a$ with $a := \sum_{j=0}^m s(r_j)$ completely specifies an instance $\mathbf{P}(\theta)$ of such a problem \mathbf{P} . The next result is due to [96].

Theorem 3.7. *There exists an integer L and finitely many real polynomials $\varphi_1, \dots, \varphi_L \in \mathbb{R}[\theta]$ in the coefficients θ of the polynomials f, g_1, \dots, g_m , such that if $\varphi_\ell(\theta) \neq 0$ for all $\ell = 1, \dots, L$, then (32), strict complementarity and second-order sufficient KKT-optimality condition (33) hold at any global minimizer of problem $\mathbf{P}(\theta)$.*

As a result, there exists $n \in \mathbb{N}$ such that (34) and (35) hold at every global minimizer $\mathbf{x}^ \in S$ of $\mathbf{P}(\theta)$, i.e., finite convergence of the Moment-SOS hierarchy is generic.*

3.5. The convex case. In this section we consider the particular case when \mathbf{P} is a convex program⁵, that is when the polynomials f and $-g_j$ are all convex, and so the set S in (13) 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 \mathbf{P} is a global optimum, then \mathbf{P} can be solved by several powerful local optimization algorithms.

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

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

⁵The set S in (13) may be convex even if the $-g_j$'s 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.

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, $\nabla^2 f = LL^T$ 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.

One has 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 f and $-g_j$ being 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.

Theorem 3.10. Let S be as in (13) 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 $n' := \max[\deg(f)/2, \max_j[\deg(g_j)/2]]$

$$(40) \quad f - f^* = \sigma_0^* + \sum_{j=1}^m \lambda_j^* g_j$$

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

$$(41) \quad f^* = \min_{\phi} \{ \phi(f) : \phi(\mathbf{1}) = 1; \mathbf{M}_{n'}(\phi) \geq 0; \phi(g_j) \geq 0, j = 1, \dots, m \}.$$

Moreover, $\phi^*(x_i) = x_i^*$, for all $i = 1, \dots, d$, where ϕ^* is an optimal solution of (41), and $\mathbf{x}^* \in S$ is a local (hence global) minimizer of \mathbf{P} .

It is also rather straightforward to check that in (40), $\lambda^* = (\lambda_j^*)_{1 \leq j \leq m}$ are Lagrange-KKT multipliers at an optimal solution $\mathbf{x}^* \in S$ of \mathbf{P} .

Next, observe that the semidefinite Moment-relaxation (41) is a particular case of (26) where the constraints $\mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0$ are replaced with the simpler $\mathbf{M}_0(g_j \cdot \phi) \geq 0$ (i.e. the scalar linear inequality constraint $\phi(g_j) \geq 0$).

However, if one does not know that f and $-g_j$ are SOS-convex and one solves (26) as for a general POP, then one still obtains $f^* = \tau_{n_0}$, that is, the first Moment-relaxation of the hierarchy is exact. In other words, the Moment-SOS hierarchy has recognized that \mathbf{P} was a convex (easy) problem.

The reason why the Moment-relaxation (26) can be replaced with the simpler (41) is because linear functionals $\phi \in \mathbb{R}[\mathbf{x}]_{2n}^*$ such that $\mathbf{M}_n(\phi) \geq 0$ 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 $\mathbf{M}_n(\phi) \geq 0$, $\phi(\mathbf{1}) = 1$, and let $\mathbf{x}^* := (\phi(x_1), \dots, \phi(x_d)) \in \mathbb{R}^d$. Then*

$$(42) \quad \phi(f) \geq f(\mathbf{x}^*), \quad \text{for every SOS-convex polynomial } f \in \mathbb{R}[\mathbf{x}]_{2n}.$$

So let ϕ^* be an optimal solution of

$$\tau'_n = \min_{\phi} \{ \phi(f) : \phi(\mathbf{1}) = 1; \mathbf{M}_n(\phi) \geq 0; \phi(g_j) \geq 0, j = 1, \dots, m \}.$$

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 POP one also obtains 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 (13), assume that $Q(g)$ is Archimedean, Slater's condition holds, and f and $-g_j$ are convex, $j = 1, \dots, m$. If $\nabla^2 f(\mathbf{x}^*) > 0$ at every global minimizer $\mathbf{x}^* \in S$ (assumed to be finitely many) then finite convergence takes place, that is, the Moment-relaxation (26) of \mathbf{P} is exact at some degree n . Moreover, the SOS-strengthening (24) of \mathbf{P} is also exact, and both (26) and (24) have an optimal solution ϕ^n and $(\lambda^*, \sigma_0^*, \dots, \sigma_m^*)$, respectively.*

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

3.6. General versus ad-hoc. One would like to emphasize that Theorem 3.5 is a quite 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 usual algorithmic practice of optimization suggests to use *ad-hoc* algorithms, i.e., algorithms tailored to the type of problem one has to solve. 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 a (only local) optimum. One typically uses branch & bound (or branch & 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 QCQP), Theorem 3.9 ensures that finite convergence takes places at first step of the hierarchy, without the need of specifying that the

POP is SOS-convex. Similarly, if f and $-g_j$ are all convex, and $\nabla^2 f(\mathbf{x}) > 0$ 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 to solve such convex problems, and indeed other efficient algorithms exist. But this remark is to simply emphasize that somehow the Moment-SOS hierarchy recognizes easy problems (as one usually considers convex programs to be 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 like Maxcut and its variants, which are notoriously difficult NP-hard problems. It is now considered as an important tool for proving/disproving the celebrated Khot's Unique Games conjecture.

3.7. Rates of convergence. In this section one provides 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 $\Delta^d = \{\mathbf{x} \in \mathbb{R}_+^d : 1 - \mathbf{e}^T \mathbf{x} \geq 0\}$.

Table 3.7 below is taken from [125]. The first column is related to the set $S \subset \mathbb{R}^d$ and "Archimedean" means that the quadratic module $Q(g)$ in (11) associated with S , is Archimedean (an algebraic certificate of its compactness used in Putinar's Positivstellensatz). The third column "Certificate" specifies if in the Moment-SOS hierarchy one uses Putinar's certificate (Theorem 2.2) or the more costly Schmüdgen's certificate (Theorem 2.3) for the semidefinite relaxation (24). In the notation $O(1/n^c)$, n is the order (or degree) of the semidefinite relaxations (24)-(26), 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 (26). Finally, for optimization of trigonometric polynomials (and so for POPs on the box

TABLE 3.1. Rates of convergence for the hierarchy of lower bounds [125]

S	Error $f^* - \tau_n$	Certificate	Reference
Archimedean	$O(1/n^c)$	Putinar	[10]
Compact	$O(1/n^c)$	Schmüdgen	[118]
\mathbb{S}^{d-1}	$O(1/n^2)$	Putinar	[28]
\mathbf{B}^d	$O(1/n^2)$	Putinar	[125]
$[-1, 1]^d$	$O(1/n^2)$	Schmüdgen	[77]
Δ^d	$O(1/n^2)$	Schmüdgen	[125]

$[0, 1]^d$ as well) and under some condition on global minimizers (isolated and with positive definite Hessian), an exponential rate of convergence has been provided in [6]. This shows that beyond general results like those in Table 3.7, there is hope for even faster rates convergence, at the price of some additional conditions on the minimizers and/or the set S .

3.8. Handling sparsity. As already mentioned, in its canonical form (24)-(26), the Moment-SOS hierarchy is limited to problems \mathbf{P} of modest dimension, even though for fixed dimension d the size-parameters of each moment-relaxation (26) are polynomial in the degree n . This is because (26) being a semidefinite program, efficient algorithms based on interior-point methods are still very time-consuming. Fortunately, practice reveals that moment-relaxations (26) of low degree n already provide tight lower bounds on f^* , and are sometimes *exact*. In addition, large-scale problems \mathbf{P} usually exhibit some *sparsity* pattern and/or symmetries. For instance, as is typically for applications in large dimension d , (i) each constraint $g_j \geq 0$ in (13) 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 \mathbf{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 (13) of \mathbf{P} contains 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 \mathbf{P} . These two types of sparsity can even be combined for further efficiency; see e.g. [138] and [78].

For sake of completeness, one next briefly describes how correlative sparsity allows to define an appropriate sparsity-adapted Moment-SOS hierarchy that can handle large-scale POP.

Assumption 3.13. *With \mathbf{P} as in (1) with $S \subset \mathbb{R}^d$ as in (13):*

- $I_0 := \{1, \dots, d\} = \cup_{k=1}^p I_k$ (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 $k \in \{1, \dots, p\}$, and so let $J_k := \{j : g_j \in \mathbb{R}[\mathbf{x}; I_k]\}$, $k = 1, \dots, p$.

Of course, as one next sees, 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

$$\begin{aligned} \mathbb{N}^{(k)} &:= \{ \alpha \in \mathbb{N}^d : \alpha_i = 0, \forall i \notin I_k \}, \quad k = 1, \dots, p; \\ \mathbb{N}_n^{(k)} &:= \{ \alpha \in \mathbb{N}^{(k)} : \sum_i \alpha_i \leq n \} \quad k = 1, \dots, p. \end{aligned}$$

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

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

$$(43) \quad \tau_n^{\text{sparse}} = \inf_{\phi=(\phi_\alpha)} \left\{ \sum_{k=1}^p \phi(f_k) : \phi_0 = 1 ; \right. \\ \left. \mathbf{M}_n(\phi; I_k) \geq 0, \quad k = 1, \dots, p ; \right. \\ \left. \mathbf{M}_{n-d_j}(g_j \cdot \phi; I_k) \geq 0, \quad \forall j \in J_k ; k = 1, \dots, p \right\}.$$

The reason why (43) is appealing is because the size of the matrix $\mathbf{M}_n(\phi; I_k)$ is $s_k(n) := \binom{\#I_k+n}{n}$ while that of $\mathbf{M}_{n-d_j}(g_j \cdot \phi)$ 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 even possibly a lot) of small size matrices rather than a single large size matrix constrained to be psd.

It is quite straightforward to obtain that $\tau_n^{\text{sparse}} \leq f^*$ for all $n \geq n_0$. Indeed $\tau_n^{\text{sparse}} \leq \tau_n \leq f^*$ (with τ_n being the optimal value of the degree- n moment relaxation (26)). Moreover the sequence $(\tau_n^{\text{sparse}})_{n \geq k_0}$ is monotone non decreasing and being bounded above, converges to some $\gamma \leq f^*$.

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

$$(44) \quad 1 - \sum_{i \in I_k} x_i^2 \geq 0, \quad k = 1, \dots, p,$$

in the definition of S .

Theorem 3.14 ([58]). *Let $S \subset \mathbb{R}^d$ be as in (13) with constraints (44) in its definition (13), and consider the hierarchy of semidefinite relaxations (43) with optimal value τ_n^{sparse} . Then $\tau_n^{\text{sparse}} \uparrow \gamma \leq f^*$, as n increases. Moreover, if for every $k = 2, \dots, p$,*

$$(45) \quad I_k \cap \left(\bigcup_{j=1}^{k-1} I_j \right) \subseteq I_\ell,$$

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

The condition (45) 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, i.e., such that for all pairs (i, j) with $I_i \cap I_j \neq \emptyset$,

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

If (45) holds then there exists a probability measure ϕ on \mathbb{R}^d such that

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

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

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

Theorem 3.15 (Sparse Positivstellensatz [58]). *Let $S \subset \mathbb{R}^d$ be as in (13) with constraints (44) in its definition (13) and let $f = \sum_{k=1}^p f_k$ with $f_k \in \mathbb{R}[\mathbf{x}; I_k]$, $k = 1, \dots, p$. If (45) holds and $f > 0$ on S then*

$$(46) \quad f(\mathbf{x}) = \sum_{k=1}^p \left(\sigma_{0,k}(\mathbf{x}) + \sum_{j \in J_k} \sigma_{j,k}(\mathbf{x}) g_j(\mathbf{x}) \right), \quad \forall \mathbf{x} \in \mathbb{R}^d,$$

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

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 management of large-scale electricity networks [89, 88], in geometric perception [141], Robotics [115], sensor network localization [94], as well as in [99].

3.9. Notes and sources. Section 3 is mainly based on [60, 63] where the reader can find all proofs (or references to papers with proofs).

Section 3.1. The Moment-SOS hierarchy was first proposed in [54, 55].

Section 3.3-Section 3.4 Theorem 3.5 and Theorem 3.7 are due to [96]. In [95] the author also shows that the Flatness condition (29) at an optimal solution of the Moment-relaxation (26), also holds generically. A refinement of these results is provided in [10]. Those results are important as they guarantee that the Moment-SOS hierarchy as 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 (26).

Section 3.5 and 3.6 are mainly taken from [63, Chapter 13]. See also [59, 43].

Section 3.7 is essentially based on [125].

Section 3.8. Correlative-sparsity was first proposed as a heuristic in [45, 135] while its proof of convergence was provided in [58]. Term-sparsity was initially described in [136] and further exploited in the TSSOS hierarchy in [137]. Finally, the combination of correlative- and term-sparsity which yields the CS-TSSOS hierarchy is described in [138]. The interested reader is referred to the book [78] 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 somehow 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 (15) and its convergence relies on Theorem 2.2. One next provides 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 (19), and its convergence relies on Theorem 2.5.

Assumption 4.1. *With $S \subset \mathbb{R}^d$ as in (13), assume that S is compact, $0 \leq g_j \leq 1$ on S , for every $j = 1, \dots, m$, and the polynomials $\{\mathbf{1}, g_1, \dots, g_m\}$ generate $\mathbb{R}[\mathbf{x}]$.*

As S is compact one may always re-scale the g_j 's (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 [60].

Next, with same notation \mathbf{g} and $\mathbf{1} - \mathbf{g}$ as in (19), and $n \in \mathbb{N}$, introduce the following linear program (LP):

$$(47) \quad \rho_n = \min_{\phi} \{ \phi(f) : \phi(\mathbf{1}) = 1; \quad \phi(\mathbf{g}^\alpha (\mathbf{1} - \mathbf{g})^\beta) \geq 0, \quad (\alpha, \beta) \in \mathbb{N}_n^{2m} \},$$

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

By its very nature, (47) is a linear program and is a relaxation of \mathbf{P} because the constraints in (47) are only *necessary* conditions on ϕ to be moments of a probability measure supported on S ; see Theorem 2.5. The dual of (47) is the linear program

$$(48) \quad \rho_n^* = \max_{c_{\alpha\beta} \geq 0, \lambda} \{ \lambda : f - \lambda = \sum_{(\alpha, \beta) \in \mathbb{N}_n^{2m}} c_{\alpha\beta} \mathbf{g}^\alpha (\mathbf{1} - \mathbf{g})^\beta \}.$$

In similar manner as (24) was a SOS-strengthening of \mathbf{P} in (3), the LP (48) is now an LP-strengthening of \mathbf{P} in (3). Of course, from duality for linear programs, $\rho_n = \rho_n^*$ for all n .

Example 4.2. *To better visualize how the LP (47) looks like, consider the toy example where $S = [0, 1] = \{x : x \geq 0; (1-x) \geq 0\} \subset \mathbb{R}$. Then for $n = 2$, $\phi = (\phi_j)_{0 \leq j \leq 2}$, and*

$$\begin{aligned} \phi(1) &= \phi_0; \quad \phi(x) = \phi_1; \quad \phi(1-x) = \phi_0 - \phi_1; \quad \phi(x^2) = \phi_2 \\ \phi(x(1-x)) &= \phi_1 - \phi_2; \quad \phi((1-x)^2) = \phi_0 - 2\phi_1 + \phi_2, \end{aligned}$$

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

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

Similarly,

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

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 (47), a boolean constraint $x_i^2 = x_i$ of \mathbf{P} , translates into the moment equality constraints $\phi(x_i^k) = \phi(x_i)$ for all $k \leq n$.

Theorem 4.3 ([60]). *With S as in (13), 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 \mathbf{P} .*

4.1. The case of a convex polytope. We now assume that $S \subset \mathbb{R}^d$ is a convex polytope (with nonempty interior), i.e., for each $j = 1, \dots, m$, $g_j \in \mathbb{R}[\mathbf{x}]_1$ (g_j is a polynomial of degree 1). In this case Theorem 2.5 specializes.

Theorem 4.4 ([34]). *Let $S \subset \mathbb{R}^d$ be as in (13) with nonempty 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 nonnegative vector $\mathbf{c} = (c_\alpha)_{\alpha \in \mathbb{N}_n^m}$, such that:*

$$(49) \quad f = \sum_{\alpha \in \mathbb{N}_n^m} c_\alpha \mathbf{g}^\alpha.$$

So the obvious analogue of (47) for a convex polytope now reads:

$$(50) \quad \rho_n = \min_{\phi} \{ \phi(f) : \phi(\mathbf{1}) = 1; \quad \phi(\mathbf{g}^\alpha) \geq 0, \quad \alpha \in \mathbb{N}_n^m \}$$

where $\phi = (\phi_\gamma)_{\gamma \in \mathbb{N}_{s_n}^d}$ with $s_n := \max_{\alpha \in \mathbb{N}_n^m} \deg(\mathbf{g}^\alpha)$. The dual of (50) reads:

$$(51) \quad \rho_n^* = \max_{\mathbf{c} \geq 0, \lambda} \{ \lambda : f - \lambda = \sum_{\alpha \in \mathbb{N}_n^m} c_\alpha \mathbf{g}^\alpha \}.$$

So an analogue of Theorem 4.3 reads:

Theorem 4.5. *Let $S \subset \mathbb{R}^d$ be as in (13) with nonempty 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 (50) (resp. (51)). 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 \mathbf{P} .*

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

$$S = \{ \mathbf{x} \in \{0, 1\}^d : g_j(\mathbf{x}) \geq 0, \quad j = 1, \dots, m \} \\ = \{ \mathbf{x} \in \mathbb{R}^d : g_j(\mathbf{x}) \geq 0, \quad j = 1, \dots, m; x_i^2 - x_i = 0, \quad i = 1, \dots, d \}.$$

This formulation includes many combinatorial optimization problems, including the celebrated MAXCUT problems and its variants on $\{-1, 1\}^d$ (after a simple linear transformation). The so-called RLT [121, 122] is a reformation-linearization technique (already proposed in the nineties) to solve \mathbf{P} when f and the g_j 's are all

linear. In RLT one “lifts” \mathbf{P} in a space of higher dimension: Namely, with $t \leq d$ fixed:

- One defines *order- t bound-factors* constraints

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

where $J_1, J_2 \subseteq \{1, \dots, d\}$, $J_1 \cap J_2 = \emptyset$ and $|J_1 \cup J_2| = t$, and

- For every $j = 1, \dots, m$, the additional *constraint factor-based restrictions*

$$A_{J_1, J_2} g_j(\mathbf{x}) \geq 0, \quad \forall (J_1, J_2).$$

- Then in each such constraints, one replaces every occurrence of the power x_i^k with x_i , and one “linearizes” the resulting polynomial constraint, i.e., every occurrence of the non-linear monomial $\prod_{i \in J} x_i$ is replaced with a variable y_J constrained to be nonnegative.

One ends up with a linear program in a space of higher dimension $s(t+1)$ which is an LP-relaxation of \mathbf{P} . In the RLT construction of this LP-relaxation (which ignores constraints of the form $\mathbf{g}^\alpha \geq 0$), one recognizes a particular case of the Moment-LP relaxation (47) of \mathbf{P} in the presence of boolean constraints $x_i^2 = x_i$, for all $i = 1, \dots, d$. 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 flavor (19), i.e., *à la* Krivine-Vasilescu.

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

At first glance one is tempted to favor the Moment-LP hierarchy because state-of-the-art LP solvers are very efficient and can solve potentially very large (even huge) scale LPs, whereas in its canonical form (26)-(24) the Moment-SOS hierarchy is limited to POPs of modest dimension and small degree of relaxation n , unless some sparsity and/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 one may try to solve such 0/1 problems with Branch & 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 & 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 (51) of \mathbf{P} . If \mathbf{P} has finitely many global minimizers and (51) 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 (51), and assume that (51) is exact, i.e., $\rho_n^* = f^*$. Then

$$(52) \quad f(\mathbf{x}) - f^* = \sum_{\alpha \in \mathbb{N}_n^m} c_\alpha^* g_1(\mathbf{x})^{\alpha_1} \cdots g_m(\mathbf{x})^{\alpha_m}, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

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

$$(53) \quad 0 = f(\mathbf{x}^*) - f^* = \sum_{\alpha \in \mathbb{N}_n^m} c_\alpha^* g_1(\mathbf{x}^*)^{\alpha_1} \cdots g_m(\mathbf{x}^*)^{\alpha_m}, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

So assume that there exists a global minimizer $\mathbf{x}^* \in S$ which is not a vertex, and let $I(\mathbf{x}^*) := \{j \in \{1, \dots, m\} : g_j(\mathbf{x}^*) = 0\}$ 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 $j = 1, \dots, m$, which in turn by (53) implies $\mathbf{c}^* = 0$, in contradiction with $\mathbf{c}^* \neq 0$. So (53) already rules out the possibility of having a global minimizer in $\text{int}(S)$. Next, from (53) one may infer

$$(54) \quad c_\alpha^* > 0 \quad \Rightarrow \quad \alpha_j > 0 \text{ for some } j \in I(\mathbf{x}^*).$$

As \mathbf{x}^* is not a vertex and \mathbf{P} has only finitely many global minimizers, the set $\{\mathbf{y} \in S : I(\mathbf{y}) = I(\mathbf{x}^*)\}$ contains a point $\hat{\mathbf{y}} \in S$ which is not a global minimizer, i.e., $f(\hat{\mathbf{y}}) > f^*$. But then (52) combined with (54) yields the contradiction

$$0 = \sum_{\alpha \in \mathbb{N}_n^m} c_\alpha^* g_1(\hat{\mathbf{y}})^{\alpha_1} \cdots g_m(\hat{\mathbf{y}})^{\alpha_m} = f(\hat{\mathbf{y}}) - f^* > 0.$$

□

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 (51) 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 nonlinear 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 (51) 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 (48). Indeed, as in Proposition 4.6 and for same reasons, if such a POP has finitely many minimizers, then a necessary condition for (48) to be exact at some degree n , is that for every global minimizer $\mathbf{x}^* \in S$, the set $\{\mathbf{y} \in S : I(\mathbf{y}) = I(\mathbf{x}^*)\}$ 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 [57, 63, Chapter 9]. In [75] 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, with among them the Lovasz-Schrijver and Sherali-Adams hierarchies. In particular the author 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 still can be useful when associated with other techniques of discrete

optimization; for instance as in [2] when used in conjunction with Branch-and-Bound.

5. A MOMENT-SOS HIERARCHY OF UPPER BOUNDS

In this section one considers 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 (3). For each “degree” n , the upper bound $\kappa_n \geq f^*$ is now computed by solving a very specific semidefinite program as its 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 \mathbf{P} in (3) with feasible set $S \subset \mathbb{R}^d$ as in (13), 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 nonempty 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 such sets are the unit box $[-1, 1]^d$, the Euclidean unit ball $\mathbf{B}(0, 1) = \{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$, the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} : \|\mathbf{x}\| = 1\}$, the canonical simplex $\Delta^d = \{\mathbf{x} \in \mathbb{R}_+^d : \mathbf{e}^T \mathbf{x} \leq 1\}$, the discrete hypercube $\{-1, 1\}^d$, and their image by an affine transformation. Even though such sets S are rather specific, the associated problems \mathbf{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 semi-definite problems:*

$$(55) \quad \kappa_n = \inf_{\sigma \in \Sigma[\mathbf{x}]_n} \left\{ \int f \sigma d\mu : \int \sigma d\mu = 1 \right\}, \quad n \in \mathbb{N};$$

$$(56) \quad \kappa_n^* = \sup_{\lambda} \{ \lambda : \lambda \mathbf{M}_n(\mu) \leq \mathbf{M}_n(f \cdot \mu) \}, \quad n \in \mathbb{N}.$$

Then :

$$(57) \quad \kappa_n = \kappa_n^*; \quad f^* \leq \kappa_{n+1} \leq \kappa_n, \quad \forall n \in \mathbb{N}; \quad \lim_{n \rightarrow \infty} \kappa_n = f^*.$$

Crucial in the proof of convergence, which can be found in [61, Theorem 4.2], is the Nichtnegativstellensatz Theorem 2.4. It turns out that (56) 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 softwares exist. If both matrices are expressed in a polynomial basis of $\mathbb{R}[\mathbf{x}]_n$ formed by polynomials that are orthonormal with respect to (w.r.t.) 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. Filling up entries of both matrices $\mathbf{M}_n(\mu)$ and $\mathbf{M}_n(f \cdot \mu)$ is straightforward. So the main effort is about computing the generalized eigenvalue of the couple of (symmetric) matrices $(\mathbf{M}_n(f \cdot \mu), \mathbf{M}_n(\mu))$ which can be done via standard softwares 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 involved matrices as n increases.

Remark 5.3. A refinement of (55) is to consider polynomial densities σ nonnegative on S (instead of SOS). That is, replace (55) with

$$(58) \quad \hat{\kappa}_n = \inf_{\sigma \in Q_n(g)} \left\{ \int f \sigma d\mu : \int \sigma d\mu = 1 \right\}, \quad n \in \mathbb{N},$$

where $Q_n(g)$ is the degree- $2n$ truncated quadratic module associated with the generators g that define S ; see (12). For instance with S being the unit box,

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

Of course $f^* \leq \hat{\kappa}_n \leq \kappa_n$ 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:

$$(59) \quad \#\mu(B) = \mu(f^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

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

$$(60) \quad f^* = \min_z \{ z : z \in \text{supp}(\#\mu) \}.$$

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

$$(61) \quad \#\mu_j = \int_{f(S)} z^j d\#\mu(z) = \int_S f(\mathbf{x})^j d\mu(\mathbf{x}), \quad \forall j \in \mathbb{N}.$$

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

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

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 (59) (the pushforward of μ by f), and let

$$(62) \quad \rho_n := \inf_{\sigma \in \Sigma[z]_n} \left\{ \int z \sigma d\#\mu : \int \sigma d\#\mu = 1 \right\}, \quad n \in \mathbb{N};$$

$$(63) \quad \rho_n^* := \sup_{\lambda} \{ \lambda : \lambda \mathbf{M}_n(\#\mu) \leq \mathbf{M}_n(z \cdot \#\mu) \}, \quad n \in \mathbb{N}.$$

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 $d_f = \deg(f)$, one obtains $\rho_n \geq \kappa_{nd_f}$ for every $n \in \mathbb{N}$ (where κ_n is defined in (55)) because if $\sigma \in \Sigma[z]_n$ is a feasible solution of (62) then $\sigma \circ f \in \Sigma[\mathbf{x}]_{nd_f}$ is a feasible solution of (55) with same value.

A detailed proof can be found in [65]. There is a striking difference between the hierarchies of upper bounds $(\kappa_n)_{n \in \mathbb{N}}$ in (55) and $(\rho_n)_{n \in \mathbb{N}}$ in (62). 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 (61) w.r.t. 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 noticing that similarly as for the hierarchy of lower bounds, $O(1/n^2)$ rates of convergence have also been obtained for the hierarchy of upper bounds (55)-(56) on the sets \mathbb{S}^{-1} , $\mathbf{B}(0, 1)$, $[-1, 1]^d$, and Δ^d ; see [125, Table 2, p. 2615].

5.3. Notes and Sources. Section 5 is essentially based on [61] and [62]; the univariate formulation is from [65]. In a series of papers, de Klerk, Laurent and collaborators 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 various and sophisticated techniques, including polynomial kernels and asymptotics for roots of some distinguished orthogonal polynomials. Moreover it turns out that such techniques have been also useful to obtain rates of convergence for the Moment-SOS hierarchy of lower bounds on specific sets S ; for more details the interested reader is referred to [125] and references therein.

6. SOME APPLICATIONS OF THE MOMENT-SOS HIERARCHY

In this section one briefly describes how the Moment-SOS hierarchy can be applied 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 problems which have an equivalent formulation as an instance of the the so-called *Generalized Moment Problem* (GMP in short) whose description is only through polynomials and semi-algebraic sets (i.e., GMP 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 one has 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:

$$(64) \quad \mathbf{GMP} : \inf_{\phi_1, \dots, \phi_s} \left\{ \sum_{j=1}^s \int f_j d\phi_j \text{ s.t. } \sum_{j=1}^s \int h_{kj} d\phi_j \geq b_k, \quad k \in \Gamma; \right. \\ \left. \text{supp}(\phi_j) \subset S_j, \quad j = 1, \dots, s \right\},$$

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

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

Of course GMP in (64) can be extended to more general functions and sets, but for a practical application of the Moment-SOS hierarchy, one needs *algebraic data* (polynomials and basic semi-algebraic sets). Notice also that formulation (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 $\phi(S) = \int 1 d\phi = 1$.

The dual GMP* of (64) is also an infinite-dimensional LP, and when $p := \#\Gamma < \infty$, it reads

$$(65) \quad \begin{aligned} \mathbf{GMP}^* : \sup_{\lambda \in \mathbb{R}_+^p} \{ & \sum_{k=1}^p \lambda_k b_k : \\ \text{s.t. } & f_j(\mathbf{x}) - \sum_{k=1}^p \lambda_k h_{kj}(\mathbf{x}) \geq 0, \forall \mathbf{x} \in S_j, \quad j = 1, \dots, s \}. \end{aligned}$$

Moment equality-constraints are also tolerated in (64) in which case the associated dual variable λ_k in (65) is not constraint to be nonnegative. As is the case in some important applications, the set Γ is also tolerated to be (countably) infinite. Finally, and not touched upon here, one may also tolerate as objective function of (64), a convex function of finitely many moments of measures ϕ_j (e.g. $-\log \det(\mathbf{M}_n(\phi_j))$) of the moment matrix $\mathbf{M}_n(\phi_j)$.

As one next sees in two examples, in some applications the problem to solve is already under the format of an instance of a GMP (or GMP*), whereas in other applications, some equivalent formulation 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 (64):

- One replaces the measures $(\phi_j)_{j=1, \dots, s}$ by degree- $2n$ truncated pseudo-moment vectors $\phi_j = (\phi_{j, \alpha}), \alpha \in \mathbb{N}_{2n}^j$, $j = 1, \dots, s$.
- One imposes semidefinite constraints on the moment and localizing matrices associated with each ϕ_j and each set S_j , which by Theorem 2.2(ii) are necessary conditions for ϕ_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 ϕ_j 's, for each fixed n one ends up with a finite-dimensional semi-definite relaxation which provides a lower bound on the optimal value of the GMP. Similarly, to apply the Moment-SOS hierarchy to GMP* in (65):

- For each $j = 1, \dots, s$, one replaces 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:

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

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

For sake of 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 (OCP), an alternative and so-called ‘‘weak formulation’’ of the OCP is an instance of the GMP.

6.1. Illustration in Probability and Computational Geometry. Let $S \subset \mathbb{R}^d$ be compact set and suppose that $S \subset \mathbf{B} := [-1, 1]^d$ (possibly after scaling). The goal is approximate the Lebesgue volume $\text{vol}(S)$ 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. [24], the discussion in [39], and references therein.

Let λ be the Lebesgue measure on $[-1, 1]^d$, so that its (infinite) vector of moments $\lambda = (\lambda_\alpha)_{\alpha \in \mathbb{N}^d}$ 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. $\text{vol}(S) = \max_{\phi \in \mathcal{M}(S)_+, \nu \in \mathcal{M}(\mathbf{B})_+} \{ \phi(\mathbf{1}) : \phi + \nu = \lambda \}$, and $\phi^* := 1_S \lambda$ is the unique optimal solution.

Proof. As $\phi + \nu = \lambda$, $\phi \leq \lambda$ and since $\text{supp}(\phi) \subseteq S$, $\phi(\mathbf{1}) = \phi(S) \leq \lambda(S) = \text{vol}(S)$. Next, with $\phi^* := 1_S \lambda \in \mathcal{M}(S)_+$, and $\nu^* := \lambda - \phi^* \in \mathcal{M}(\mathbf{B})_+$, one obtains $\phi^*(\mathbf{1}) = \lambda(S) = \text{vol}(S)$. \square

The above formulation of $\text{vol}(S)$ as an optimization problem is not yet in the format of an instance of the GMP (64). But notice that since \mathbf{B} is compact,

$$\phi + \nu = \lambda \quad \Leftrightarrow \quad \phi_\alpha + \nu_\alpha = \lambda_\alpha, \quad \forall \alpha \in \mathbb{N}^d,$$

and therefore

$$(66) \quad \text{vol}(S) = \max_{\phi \in \mathcal{M}(S)_+, \nu \in \mathcal{M}(\mathbf{B})_+} \{ \phi(\mathbf{1}) : \phi_\alpha + \nu_\alpha = \lambda_\alpha, \quad \forall \alpha \in \mathbb{N}^d \},$$

which is an instance of the GMP (64) with $\Gamma = \mathbb{N}^d$ (a countable set). We next see how to implement the moment-SOS hierarchy. Let $g_0 = \mathbf{1}$,

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

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

$$(67) \quad \tau_n = \max_{\phi, \nu} \{ \phi_0 : \phi_\alpha + \nu_\alpha = \lambda_\alpha, \quad \forall \alpha \in \mathbb{N}_{2n}^d; \\ \mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0, \quad j = 0, \dots, m; \mathbf{M}_n(\nu) \geq 0 \}.$$

For each fixed n , (67) is a semidefinite program and an obvious relaxation of (66) so that $\tau_n \geq \text{vol}(S)$ for all n .

Theorem 6.2 ([39]). *The sequence of optimal values $(\tau_n)_{n \in \mathbb{N}}$ is monotone non increasing, bounded below, and $\lim_{n \rightarrow \infty} \tau_n = \text{vol}(S)$ as n grows.*

So (67) indexed by $n \in \mathbb{N}$, provides a hierarchy of semidefinite relaxations of (66) such that $(\tau_n)_{n \in \mathbb{N}}$ converges (from above) to the desired value $\text{vol}(S)$ as n increases. However in its basic form (67) its convergence is quite slow. To see why, consider the dual of (67), which is the semidefinite program

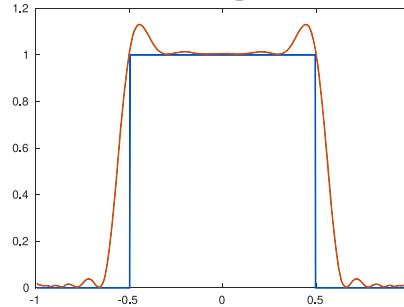
$$(68) \quad \tau_n^* = \min_{p \in \mathbb{R}[\mathbf{x}]_{2n}} \left\{ \int_{\mathbf{B}} p \, d\lambda : p - 1 = \sum_{j=0}^m \sigma_j g_j \right. \\ \left. p \in \Sigma[\mathbf{x}]_n; \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j}; j = 0, \dots, m \right\}.$$

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

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

and since $\int p \, d\lambda \downarrow \int 1_S \, d\lambda$ as n grows, in the dual (68) one searches of a degree- $2n$ SOS polynomial p that tries to approximate from above, the indicator function of 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.

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



Fortunately one may significantly attenuate (or even remedy to) this problem. Indeed as one knows in advance the (unique) optimal solution $\phi^* = 1_S \lambda$ of (66), every available additional information on ϕ^* in terms of linear constraints on its moments can be added to (66) without changing its optimal value and solution. While such additional redundant constraints do not change (66), they have a dramatic impact on the relaxations (67) and yield a quite significant acceleration of their convergence. And it is indeed the case when one adds additional moment constraints (satisfied by ϕ^*) coming from Stokes' theorem.

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

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

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

$$\phi^*(s_\alpha) := \phi^*(\operatorname{Div}(\mathbf{x} g(\mathbf{x}) \mathbf{x}^\alpha)) = 0,$$

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

$$(69) \quad \phi(s_\alpha) = 0, \quad \forall \alpha : |\alpha| \leq 2n - 1 - \deg(g),$$

can be included in the semidefinite relaxation (67). The effect on the dual (68) is to change the initial constraint $p - 1 = \sigma_0 + \sigma_1 g$, to now

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

where $q := \sum_{|\alpha| \leq 2n - \deg(g) - 1} \theta_\alpha s_\alpha$ for the dual variables (θ_α) associated with (69). Hence now p is *not* required anymore to approximate 1_S from above! For more details on volume computation via the moment-SOS hierarchy, the interested reader is referred to [39], [129], and [128]. In particular, this technique has also been implemented in [129] for approximating 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 (66) with a measure λ different from Lebesgue measure on \mathbf{B} . For example with λ being the Gaussian measure $\exp(-\|\mathbf{x}\|^2) d\mathbf{x}$ on \mathbb{R}^d , or the exponential measure $\exp(-\sum_j x_j) d\mathbf{x}$ on \mathbb{R}_+^d , one may approximate the value $\lambda(S)$ for non-compact semi-algebraic sets, as closely as desired; see e.g. [64].*

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 $\mathbf{m} = (m_\alpha)_{\alpha \in \Gamma}$, 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 nonempty interior. The goal is to provide the best upper bound on $\operatorname{Prob}(X \in S)$, under the partial knowledge of $\mathbf{m} = (m_\alpha)_{\alpha \in \Gamma}$, that is, compute:

$$(70) \quad \rho = \max_{\mu \in \mathcal{M}(\mathbb{R}^d)_+} \left\{ \mu(S) : \int \mathbf{x}^\alpha d\mu = m_\alpha, \forall \alpha \in \Gamma \right\}.$$

Observe that (70) is an instance of the GMP (64) but with non-polynomial data because with $\mu \in \mathcal{M}(\mathbb{R}^d)_+$, $\mu(S) = \mu(1_S)$ and 1_S is not a polynomial. We need consider that $\mu = \phi + \nu$ with $\phi \in \mathcal{M}(S)_+$ (in which case $\phi(S) = \phi(\mathbf{1}) = \phi_0$), and so for every $n \in \mathbb{N}$, consider the optimization problem,

$$(71) \quad \tau_n = \max_{\phi, \nu} \left\{ \phi_0 : \phi_\alpha + \nu_\alpha = m_\alpha, \quad \forall \alpha \in \Gamma; \right. \\ \left. \mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0, \quad j = 0, \dots, m; \mathbf{M}_n(\nu) \geq 0 \right\},$$

which is a relaxation of (70) and a variant of (67) (in fact even easier because now Γ is a finite set instead of the countable set \mathbb{N}^d). For instance with $\Gamma = \mathbb{N}_{2t}^d$ for some fixed t , the dual of (71) reads:

$$(72) \quad \tau_n^* = \min_p \left\{ \sum_{\alpha \in \Gamma} p_\alpha m_\alpha : p - 1 = \sum_{j=0}^m \sigma_j g_j \right. \\ \left. p \in \Sigma[\mathbf{x}]_t; \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j}; j = 0, \dots, m \right\},$$

The difference with (68) is that now, even when n changes one still searches 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 $\Phi := \{\mu \in \mathcal{M}(\mathbb{R}^d)_+ : \int \mathbf{x}^\alpha d\mu = m_\alpha, \forall \alpha \in \Gamma\}$. If there is some $\mu \in \Phi$ with a strictly positive density w.r.t. Lebesgue measure, and with all moments finite, then each relaxation (72) 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 $n \rightarrow \infty$. Finally, the bound τ may be sharp, i.e., $\tau = \rho$ if at some step n , the moment matrices of an optimal solution (ϕ, ν) of (71) satisfy some ‘‘flatness’’ property; for more details see e.g. [56] and [14].

6.2. Illustration in optimal control of dynamical systems. In this section one briefly describes how to apply the Moment-SOS hierarchy to help solve optimal control problems (OCP) with algebraic data, i.e., whose description is through polynomials and basic semi-algebraic sets. Consider the optimal control problem:

$$(73) \quad \text{OCP : } J(\mathbf{x}_0, 0) := \min_{\mathbf{u}} \left\{ \int_0^1 h(\mathbf{x}(t), \mathbf{u}(t)) dt + H(\mathbf{x}(1)) \right. \\ \text{s.t. } \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)), \quad \forall t \in (0, 1) \\ \mathbf{x}(t) \in X, \mathbf{u}(t) \in U, \quad \forall t \in (0, 1) \\ \left. \mathbf{x}(0) = \mathbf{x}_0 \right\},$$

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.

Eq. (73) 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 ordinary differential equation (o.d.e.) with vector field $f : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$, 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 $\int_0^1 h(\mathbf{x}(t), \mathbf{u}(t)) dt + H(\mathbf{x}(1))$. We here do not discuss the appropriate function spaces in which one has to search the state and control trajectories $\mathbf{x}(t), \mathbf{u}(t)$, but rather to show

- (i) how to model (73) as a particular instance of the GMP (64), and
- (ii) how to define an appropriate Moment-SOS hierarchy for solving (73).

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

Strategy. As for (static) optimization problems (1) where in the Moment-SOS hierarchy one searches for a probability measure (the Dirac measure $\delta_{\{\mathbf{x}^*\}}$ at a global minimizer $\mathbf{x}^* \in S$), here one 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 (73) is a polynomial). The state and control constraints in (73) translate into support constraints on μ and ν .

Modeling (73) as a GMP via occupation measures. The idea is to look at (73) via its impact on the evaluation of test functions all along feasible trajectories. Let $\{(\mathbf{x}(t), \mathbf{u}(t)) : t \in [0, 1]\}$ be an admissible trajectory, and let $(\mathbf{x}, t) \mapsto w(\mathbf{x}, t)$ be an arbitrary *test* function in $\mathcal{C}^1(X \times [0, 1])$: Then observe that

$$\begin{aligned} (74) \quad w(\mathbf{x}(1), 1) - w(\mathbf{x}(0), 0) &= \int_0^1 dw(\mathbf{x}(t), t) \\ &= \int_0^1 \frac{\partial w(\mathbf{x}(t), t)}{\partial t} + \langle \nabla_{\mathbf{x}} w(\mathbf{x}(t), t), f(\mathbf{x}(t), \mathbf{u}(t)) \rangle dt. \end{aligned}$$

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

$$(75) \quad \mu(A \times B \times C) = \int_{C \cap [0, 1]} 1_{A \cap X}(\mathbf{x}(t)) 1_{B \cap U}(\mathbf{u}(t)) dt$$

$$(76) \quad \nu(A \times C) = 1_{(A \cap X) \times (C \cap \{1\})}(\mathbf{x}, t)$$

$$(77) \quad \nu_0(A \times C) = 1_{(A \cap X) \times (C \cap \{0\})}(\mathbf{x}, t),$$

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 $\{(\mathbf{x}(t), \mathbf{u}(t)) : t \in [0, 1]\}$. Another equivalent characterization of μ is via its *disintegration*

$$(78) \quad d\mu(\mathbf{x}, \mathbf{u}, t) = \delta_{\{(\mathbf{x}(t), \mathbf{u}(t))\}}(d(\mathbf{x}, \mathbf{u})) 1_{[0, 1]}(t) dt,$$

into:

- its *marginal* $1_{[0, 1]}(t) dt$ 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 $(\mathbf{x}(t), \mathbf{u}(t)) \in X \times U$).

Importantly, observe that the support of μ is the *graph* $\{(t, \mathbf{x}(t), \mathbf{u}(t)) : t \in [0, 1]\}$ of state-control trajectories $(\mathbf{x}(t), \mathbf{u}(t))$. Ideally one searches for the measure

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

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

One reason why one introduces occupation measures is that the *time* integral (74) reads as the *spatial* integral

$$(79) \quad \int w \, d\nu - \int w \, d\nu_0 = \int \frac{\partial w}{\partial t}(\mathbf{x}, t) + \langle \nabla_{\mathbf{x}} w(\mathbf{x}, t), f(\mathbf{x}, \mathbf{u}) \rangle \, d\mu(\mathbf{x}, \mathbf{u}, t),$$

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 $\mathcal{L} : \mathcal{C}^1(X \times [0, 1]) \rightarrow \mathcal{C}(X \times U \times [0, 1])$:

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

and its adjoint $\mathcal{L}^* : \mathcal{C}(X \times U \times [0, 1])^* \rightarrow \mathcal{C}^1(X \times [0, 1])^*$ 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 (79) 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 $\mathcal{D} \subset \mathcal{C}^1(X \times [0, 1])$, one obtains the equation $\mathcal{L}^* \mu = \nu - \nu_0$, i.e.,

$$(80) \quad \frac{\partial \mu}{\partial t} + \operatorname{div}(f \mu) + \nu = \nu_0.$$

Eq (80) 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, as *equation of conservation of mass*, *advection equation* or *Liouville's equation*.

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

$$(81) \quad \begin{aligned} \rho &= \inf_{\mu, \nu} \int h \, d\mu + \int H \, d\nu \\ \text{s.t.} \quad &\int \frac{\partial w}{\partial t} + \langle \nabla_{\mathbf{x}} w, f \rangle \, d\mu = \int w \, d\nu - \int w \, d\nu_0, \quad \forall w \in \mathcal{D}; \\ &\mu \in \mathcal{M}(X \times U \times [0, 1])_+, \quad \nu \in \mathcal{M}(X)_+, \end{aligned}$$

introduced by [134].

Observe that if \mathcal{D} is a countable set of polynomials then (81) is an instance of the GMP in (64) and of course a relaxation of (73) so that $\rho \leq J$. The dual of (81) reads

$$(82) \quad \begin{aligned} \rho^* &= \sup_{w \in \mathcal{C}^1(X \times [0, 1])} \int w \, d\nu_0 (= w(\mathbf{x}_0, 0)) : \\ \text{s.t.} \quad &h + \mathcal{L}w \geq 0, \quad \forall (\mathbf{x}, \mathbf{u}, t) \in X \times U \times [0, 1]; \\ &w(\mathbf{x}, 1) \leq H(\mathbf{x}), \quad \forall \mathbf{x} \in X. \end{aligned}$$

It turns out that under some convexity assumptions, $\rho = J(\mathbf{x}_0, 0)$, i.e., the measure-valued weak formulation (81) is equivalent to the strong formulation (73). In the dual (82) one approximates the optimal value function $J : X \times [0, 1] \rightarrow \mathbb{R}$ all

along an optimal trajectory $\{(\mathbf{x}^*(t), t) : t \in [0, 1]; \mathbf{x}^*(0) = \mathbf{x}_0\}$ (but *not* for all $(\mathbf{x}, t) \in X \times [0, 1]$). For more details see e.g. [68], [47], and references therein.

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

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

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

$$(83) \quad \begin{aligned} \rho_n = \min_{\mu, \nu} \quad & \mu(h) + \nu(H) \\ \text{s.t.} \quad & \mu\left(\frac{\partial w}{\partial t} + \langle \nabla_{\mathbf{x}} w, f \rangle\right) = \nu(w) - \nu_0(w), \quad \forall w \in \mathcal{D}_n; \\ & \mathbf{M}_n(\mu), \mathbf{M}_n(\nu) \geq 0; \\ & \mathbf{M}_{n-d_g}(g \cdot \mu), \mathbf{M}_{n-d_g}(g \cdot \nu) \geq 0, \quad \forall g \in G; \\ & \mathbf{M}_{n-1}(t(1-t) \cdot \mu) \geq 0, \mathbf{M}_{n-d_\theta}(\theta \cdot \mu) \geq 0, \quad \forall \theta \in \Theta. \end{aligned}$$

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

Reconstruction of optimal trajectories from moments. So far, by solving the semi-definite relaxations (83) one obtains a sequence $(\rho_n)_{n \in \mathbb{N}}$ of lower bounds on the optimal value J of the initial OCP (73). But from the vector of pseudo-moments (μ^n, ν^n) optimal solution of (83) for some degree n , can one 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, one expects that (μ^n, ν^n) approximates quite well moments of the measures

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

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 one describes an efficient strategy via the Christoffel function, a tool from approximation theory and orthogonal polynomials, particularly well suited to identify the support of a measure from the sole knowledge of its moments.

6.3. Other applications. We here 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 indeed the Moment-SOS hierarchy is a versatile tool, widely applicable as soon as the problem data are algebraic, and provided that some sparsity and/or symmetries can be exploited when the problem size demands.

- In control and stochastic control: [110], [36], [106, 109, 4, 108], [46], [29], [33].
- For convex computation of region of attraction for dynamical systems, see e.g. [37, Chapter 10], and for analysis and control of some types on non-linear PDEs, see e.g. [87], [37, Chapter 11] and [47].

- In tensor computation: [98], [101, 27, 97].
- In algorithmic game theory: [127], [53], [100].
- In management of energy networks, and in particular the for solving the Optimal Power Flow Problem (OPF) problem for (large) electricity networks. The Moment-SOS hierarchy has been able to handle problems with thousands variables by exploiting some inherent sparsity, in the spirit of Section 3.8; see e.g. [89, 130, 88, 19, 35].
- In Computer Science, e.g., for coding, packing problems: [8], [7], [22],[51].
- In Computer Vision, Geometric perception and Pattern Recognition : [140, 143, 142, 112].
- In mathematical finance for portfolio optimization and option pricing see [72], [31]. When the evolution in time is modeled by Ito's stochastic differential equations, a weak formulation of the problem via occupation measures, is almost identical to that of OCPs with the only difference 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 e.g. [71].
- In Internet of Things (IoT): [119].
- In computer graphics and geometry processing. [83, 84].
- In signal processing. [82, 18].
- For optimal design in Statistics: [17].
- In physics for bounding ground-state energy of interacting particle systems: [50].
- In Chemistry for deriving bounds on stochastic chemical kinetic systems: [23].
- In traffic networks for bounding travel time: [40].
- in Engineering: [20].
- In Machine Learning for certification of robustness for Neural Networks: [74, 131].
- In Quantum Information (e.g., for several problems in entanglement theory): see [26, 5, 105, 21, 120].
- In data analysis of citation networks: see e.g. [139].
- In radar and wireless communications: see [103].
- In medical applications of cancer treatment: see [90].
- In truss topology design: see [132].

6.4. Notes and Sources. Section 6.1 is based on [56] and [39] while Section 6.2 is based on [68]. Infinite-dimensional LP formulations of optimal control problems can be traced back to works of L.C. Young, A.F. Filippov R.V. Gamkrelidze, J. Warga; see [30] 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 [47] follows same principles as for solving OCPs. Namely one considers a measure-valued weak formulation of the problem, similar as the one in (81) for OCP's, 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 Kruzhkov) on the occupation measures are needed; see [87] and [37, Chapter 11].

The reconstruction technique of state and control trajectories based on the Christoffel function is detailed in [86]. In particular, this technique has been used with success in [87] to recover solutions to Burgers PDE from moments of the measure supported on their graph. A (remarkably accurate) approximation of such moments have been obtained by solving semidefinite relaxations of the Moment-SOS hierarchy applied to the weak measure-valued formulation of Burgers equation (and in spirit similar to (83) for optimal control problems). The role and remarkable properties of the Christoffel function to recover a function from the sole knowledge of moments of the (degenerate) measure supported on its graph, is also treated in more details in Section 7.

7. POSITIVE POLYNOMIALS AND CHRISTOFFEL FUNCTION

In this section one introduces 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 of being interesting in their own right,

- they prove to be useful to understand and interpret the Moment-SOS hierarchy of lower bounds,
- 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 (13), 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 be also defined for a Riesz linear functional $\phi \in \mathbb{R}[\mathbf{x}]^*$ (with $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha \in \mathbb{N}^d}$) such that $\mathbf{M}_n(\boldsymbol{\phi}) > 0$ 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 $\mathbf{M}_n(\boldsymbol{\phi}) > 0$ for every n , let $(P_\alpha)_{\alpha \in \mathbb{N}^d}$ be a family of polynomials that are *orthonormal* w.r.t. ϕ , that is, such that

$$(84) \quad \phi(P_\alpha P_\beta) = \delta_{\alpha=\beta}, \quad \forall \alpha, \beta \in \mathbb{N}^d,$$

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

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

$$(85) \quad (\mathbf{x}, \mathbf{y}) \mapsto K_n^\phi(\mathbf{x}, \mathbf{y}) := \sum_{\alpha \in \mathbb{N}_n^d} P_\alpha(\mathbf{x}) P_\alpha(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \quad n \in \mathbb{N},$$

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

$$(86) \quad \mathbf{x} \mapsto \Lambda_n^\phi(\mathbf{x}) := K_n^\phi(\mathbf{x}, \mathbf{x})^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^d, \quad n \in \mathbb{N},$$

i.e., 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 as $(P_\alpha)_{\alpha \in \mathbb{N}_n^d}$ form a basis of $\mathbb{R}[\mathbf{x}]_n$, write

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

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

$$(87) \quad \begin{aligned} \phi(K_n^\phi(\mathbf{x}, \cdot) p) &= \phi \left(\left(\sum_{\alpha \in \mathbb{N}_n^d} P_\alpha(\mathbf{x}) P_\alpha(\mathbf{y}) \right) \cdot \left(\sum_{\beta \in \mathbb{N}_n^d} p_\beta P_\beta(\mathbf{y}) \right) \right) \\ &= \sum_{\alpha \in \mathbb{N}_n^d} p_\alpha P_\alpha(\mathbf{x}) = p(\mathbf{x}), \quad \forall p \in \mathbb{R}[\mathbf{x}]_n, \end{aligned}$$

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 w.r.t. μ , with scalar product

$$\langle f, g \rangle = \int_S f g d\mu, \quad \forall f, g \in L^2(\mu),$$

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

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

Alternative formulations of the CF. Alternatively, the CF can be also be defined by:

$$(88) \quad \Lambda_n^\phi(\xi)^{-1} = \mathbf{v}_n(\xi)^T \mathbf{M}_n(\phi)^{-1} \mathbf{v}_n(\xi), \quad \forall \xi \in \mathbb{R}^d,$$

(the ABC theorem in [124]) and it also has the variational formulation:

$$(89) \quad \Lambda_n^\phi(\xi) = \min_{p \in \mathbb{R}[\mathbf{x}]_n} \{ \phi(p^2) : p(\xi) = 1 \}, \quad \forall \xi \in \mathbb{R}^d.$$

In particular, observe that (89) can be rewritten

$$\Lambda_n^\phi(\xi) = \min_{\mathbf{p} \in \mathbb{R}^{s(n)}} \{ \mathbf{p}^T \mathbf{M}_n(\phi) \mathbf{p} : \langle \mathbf{p}, \mathbf{v}_n(\xi) \rangle = 1 \}, \quad \forall \xi \in \mathbb{R}^d,$$

which is a convex quadratic optimization problem which can be solved efficiently even for large dimension d . After some algebra, the unique optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_n$ of (89) reads

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

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:

- $\forall \xi \in \text{supp}(\mu)$, $\Lambda_n^\mu(\xi)^{-1}$ grows at most as a polynomial in n .
- $\forall \xi \notin \text{supp}(\mu)$, $\Lambda_n^\mu(\xi)^{-1}$ 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 similar (and sometimes better) performance as state-of-the-art techniques; see [70, 69].

Next let μ have a density f w.r.t. Lebesgue measure on S . Under some additional regularity properties of μ and its support S ,

$$(90) \quad \lim_{n \rightarrow \infty} s(n) \Lambda_n^\mu(\xi) = f(\xi) / \mu_E(\xi),$$

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

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 $p \in \mathbb{R}[\mathbf{x}]_n$,

$$(91) \quad \sup_{\mathbf{x} \in S} |p(\mathbf{x})| \leq M_n \cdot \left(\int_S p^2 d\mu \right)^{1/2}, \quad \text{and} \quad \lim_{n \rightarrow \infty} \log(M_n)/n = 0$$

(see e.g. [70, Section 4.3.3]). The Bernstein-Markov property allows qualitative description for asymptotics of the Christoffel function as n grows.

The notion of *equilibrium measure* associated to 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 $d\mu := dx/\pi\sqrt{1-x^2}$ 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 $\text{int}(S)$. It has an even explicit expression if S is convex and symmetric about the origin; see e.g. [12] [13, Theorem 1.1 and Theorem 1.2]. Moreover if μ is a Borel measure on S and (S, μ) has the Bernstein-Markov property (91) 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:

$$(92) \quad \lim_{n \rightarrow \infty} \int_S \mathbf{x}^\alpha d\nu_n = \lim_{n \rightarrow \infty} \int_S \frac{\mathbf{x}^\alpha d\mu(\mathbf{x})}{s(n)\Lambda_n^\mu(\mathbf{x})} = \int_S \mathbf{x}^\alpha d\lambda_S, \quad \forall \alpha \in \mathbb{N}^d.$$

(See e.g. [70, 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 on equilibrium measures see [12, 13], the discussion in [70, Section 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 one briefly addresses the following issue which is interesting in its own right and is also central in the recovery of an optimal (or ε -optimal) trajectory $\{\mathbf{x}(t) : t \in [0, 1]\}$ in optimal control problems, from the sole knowledge of moments of the occupation measure supported on the graph $\{(t, \mathbf{x}(t)) : t \in [0, 1]\}$; see Section 6.2.

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

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

for some unknown measurable function $f : [0, 1] \rightarrow X$, i.e., μ is supported on the graph $\{(t, f(t)) : t \in [0, 1]\}$ of f . The goal is to recover f from the sole knowledge of moments $\boldsymbol{\mu} = (\mu_{ij})_{(i,j) \in \mathbb{N}^2}$ where :

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

We propose to use the Christoffel function Λ_n^μ to recover f from $\boldsymbol{\mu}$ because as seen before, Λ_n^μ is a good tool to identify 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 $[0, 1] \times X$, i.e. 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 one first ‘‘perturbates’’ (or regularizes) $\mathbf{M}_n(\mu)$ to $\mathbf{M}_n(\mu) + \varepsilon \mathbf{I}$ with \mathbf{I} the identity matrix and some small regularization parameter $\varepsilon > 0$, and one defines a new perturbed Christoffel function $\hat{\Lambda}_n^\mu$ by:

$$(93) \quad (t, x) \mapsto \hat{\Lambda}_n^\mu(t, x)^{-1} := \mathbf{v}_n(t, x)^T (\mathbf{M}_n(\mu) + \varepsilon \mathbf{I})^{-1} \mathbf{v}_n(t, x), \quad \forall (t, x) \in \mathbb{R}^2.$$

Then define the following n -approximant $f_n : [0, 1] \rightarrow X$ of f by:

$$(94) \quad t \mapsto f_n(t) := \arg \min_{x \in X} \hat{\Lambda}_n^\mu(t, x)^{-1}, \quad t \in [0, 1].$$

(In case of several minimizers in (94) then as a tie-breaker rule just take the smallest one.) For every fixed $t \in [0, 1]$, computing $f_n(t)$ can be done 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 [86]. Observe that the f_n approximant (94) is *not* a polynomial and being semi-algebraic, it is able approximate quite well some discontinuous functions with no Gibbs phenomenon.

For instance in Figure 7.3 (left) one 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, i.e.,

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

(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 $\nu = f(t)dt$ on $[0, 1]$. On the other hand, with $\varepsilon > 0$ very small and f_n as in (94), the step function is recovered almost exactly (in black) with no Gibbs phenomenon and with small degree $n = 4$. This is what one may call a non-standard application of the CD kernel as one considers the degenerate bivariate measure μ on $X \times [0, 1]$ instead of the univariate measure $\nu = f(t)dt$ on $[0, 1]$.

Similarly in Figure 7.3, two discontinuous Eckhoff functions from [25] in red are also recovered (in black) with very good precision via f_n in (94) with $n = 10$, and again with no Gibbs phenomenon; for more details the reader is referred to [86].

FIGURE 7.1. Left: Degree-12 polynomial approximation of step function with Gibbs phenomenon and right: step function approximated by f_4 in (94). © Reprinted from [86]

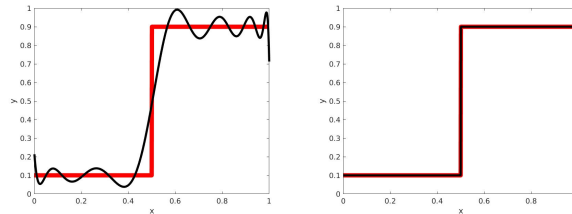
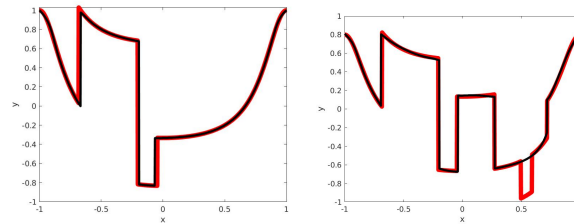


FIGURE 7.2. Two Eckhoff functions [25] approximated by f_{10} in (94). © Reprinted from [70]



Application to optimal control. As was already mentioned in Section 6.2, such a technique of approximation can be used to recover the graph of functions supported on trajectories $\{(\mathbf{x}^*(t), \mathbf{u}^*(t)) : t \in [0, 1]\}$, optimal solutions of optimal control problems (73) described in Section 6.2. Indeed when applying the Moment-SOS

hierarchy to solve (73), at an optimal solution of (83) one obtains 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 $(\mathbf{x}^*, \mathbf{u}^*) : [0, 1] \rightarrow \mathbb{R}^d \times \mathbb{R}^m$.

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

- 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)), and
- compute the perturbed Christoffel function in (93) associated with μ_i , i.e.,

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

and then the f_n approximant of the function $x_i(t)$ is obtained via (94).

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. One has first noticed that by construction, the reciprocal of a Christoffel function is an SOS polynomial. Next, with $S \subset \mathbb{R}^d$ as in (13), recall the convex cone

$$(95) \quad Q_n(g) := \left\{ \sum_{j=0}^m \sigma_j g_j : \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j}, j = 0, \dots, m \right\},$$

which is 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

$$(96) \quad Q_n^*(g) = \{ \boldsymbol{\phi} \in \mathbb{R}^{s(2n)} : \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \geq 0, j = 0, \dots, m \},$$

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. One has seen that $Q_n(d)$ and its dual $Q_n^*(g)$ are crucial in the construction of the Moment-SOS hierarchy of lower bounds in Section 3. 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 $p \in \text{int}(Q_n(g))$ then there exists $\boldsymbol{\phi} \in \text{int}(Q_n^*(g))$ such that*

$$(97) \quad p(\mathbf{x}) = \sum_{j=0}^m \Lambda_{n-d_j}^{g_j \cdot \boldsymbol{\phi}}(\mathbf{x})^{-1} g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^d,$$

or, equivalently:

$$(98) \quad \text{int}(Q_n(g)) = \left\{ \sum_{j=0}^m (\Lambda_{n-d_j}^{g_j \cdot \boldsymbol{\phi}})^{-1} g_j : \boldsymbol{\phi} \in \text{int}(Q_n^*(g)) \right\}.$$

Theorem 7.1 is an interpretation in [66] of a duality result of [92]. 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 $\sigma_j = (\Lambda_{n-d_j}^{g_j \cdot \phi})^{-1}$ in its Putinar's representation (15). Indeed those weights are all coming from a *unique* element $\phi \in \text{int}(Q_n^*(g))$ and its Christoffel functions associated with the Riesz linear functionals $g_j \cdot \phi$, $j = 0, \dots, m$. It also turns out that those weights have an extremal property: Consider the optimization problem:

$$(99) \quad \rho_n = \inf_{\phi \in \mathbb{R}^{s(2n)}} \left\{ - \sum_{j=0}^m \log \det(\mathbf{M}_{n-d_j}(g_j \cdot \phi)) : \right. \\ \left. \phi(p) = 1, \mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0, \forall j = 0, \dots, m \right\},$$

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

$$(100) \quad \rho_n^* = \sup_{\mathbf{Q}_j} \left\{ \sum_{j=0}^m \log \det(\mathbf{Q}_j) : \mathbf{Q}_j \geq 0, \forall j = 0, \dots, m \right. \\ \left. p(\mathbf{x}) \sum_{j=0}^m s(n-d_j) = \sum_{j=0}^m g_j(\mathbf{x}) \cdot \mathbf{v}_{n-d_j}(\mathbf{x})^T \mathbf{Q}_j \mathbf{v}_{n-d_j}(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^d \right\},$$

where the supremum is taken over real symmetric matrices \mathbf{Q}_j of respective size $s(n-d_j)$, $j = 0, \dots, m$. The criterion to maximize in (100) is minus the log-barrier of the convex cone $Q_n(g)$.

Theorem 7.2. *With $n \in \mathbb{N}$ fixed, Problems (99) and (100) have the same finite optimal value $\rho_n = \rho_n^*$ if and only if $p \in \text{int}(Q_n(g))$. Moreover, both have a unique optimal solution $\phi_{2n}^* \in \mathbb{R}^{s(2n)}$ and $(\mathbf{Q}_j^*)_{j=0, \dots, m}$ respectively, which satisfy $\mathbf{Q}_j^* = \mathbf{M}_{n-d_j}(g_j \cdot \phi_{2n}^*)^{-1}$ for all $j = 0, \dots, m$. And, as a consequence,*

$$(101) \quad p(\mathbf{x}) = \frac{1}{\sum_{j=0}^m s(n-d_j)} \sum_{j=0}^m g_j(\mathbf{x}) \mathbf{v}_{n-d_j}(\mathbf{x})^T \mathbf{M}_{n-d_j}(g_j \cdot \phi_{2n}^*)^{-1} \mathbf{v}_{n-d_j}(\mathbf{x}) \\ = \frac{1}{\sum_{j=0}^m s(n-d_j)} \sum_{j=0}^m g_j(\mathbf{x}) \Lambda_{n-d_j}^{g_j \cdot \phi_{2n}^*}(\mathbf{x})^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

Notice that ϕ in (97) is just $(\sum_{j=0}^m s(n-d_j)) \phi_{2n}^*$ with ϕ_{2n}^* as in (101).

Of course, Theorem 7.1 immediately raises the following question: *given $p \in \text{int}(Q_n(g))$ what is this linear functional $\phi \in \mathbb{R}[\mathbf{x}]_{2n}^*$ with associated moment sequence $\phi \in \mathbb{R}^{s(2n)}$ 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 (13) be a compact set with nonempty interior, generated by a finite set $\tilde{G} = \{g_1, \dots, g_m\} \subset \mathbb{R}[\mathbf{x}]$ 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} . For instance

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

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

The equilibrium measure μ of S is proportional to $d\mathbf{x}/\prod_{j=1}^d \sqrt{1 - x_j^2}$.

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

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

The equilibrium measure μ of S is proportional to $d\mathbf{x}/\sqrt{(1 - \sum_j x_j) \prod_{j=1}^d x_j}$.

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 $G_n = G$ as soon as $n \geq \lceil d/2 \rceil$.

Theorem 7.3 ([67, 73]). *Let $S \subset \mathbb{R}^d$ be the Euclidean unit ball, the unit box, or the simplex, and let μ be its equilibrium measure. Then for all integer n :*

$$(102) \quad 1 = \frac{1}{\sum_{g \in G_n} s(t - t_g)} \sum_{g \in G_n} g(\mathbf{x}) \Lambda_{n-t_g}^{g \cdot \mu}(\mathbf{x})^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

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

In addition, for every n , the polynomials $(g/\Lambda_{n-t_g}^{g \cdot \mu})_{g \in G_n}$ (all nonnegative on S) provide S with a polynomial *partition of unity*. We have called (102) a *generalized polynomial Pell's equation* solved by the Christoffel functions $(\Lambda_{n-t_g}^{g \cdot \mu})_{g \in G_n}$ (and the polynomials $g \in G_n$) because (102) is an exact multivariate generalization of the polynomial Pell's equation⁶

$$(103) \quad 1 = T_n(x)^2 + (1 - x^2)U_{n-1}(x)^2, \quad \forall x \in \mathbb{R},$$

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

satisfied by the univariate Chebyshev polynomials of first kind $(T_n)_{n \in \mathbb{N}}$ and Chebyshev polynomials of second kind $(U_n)_{n \in \mathbb{N}}$. Indeed after normalization to orthonormal polynomials, and summing up (103) over n , one obtains

$$(104) \quad 1 = \frac{1}{s(n) + s(n-1)} \left(\Lambda_n^\mu(x)^{-1} + g(x) \Lambda_{n-1}^{g \cdot \mu}(x)^{-1} \right), \quad \forall x \in \mathbb{R}, n \in \mathbb{N},$$

where $g(x) = 1 - x^2$, and $d\mu(x) = dx/\pi\sqrt{1-x^2}$ is the equilibrium measure of the interval $S = [-1, 1]$. The term “generalized” is justified because in (102) one has 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 (103). But formally, (102) is exactly of the same flavor as (104).

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

$$\mathbf{x} \mapsto p(\mathbf{x}) := \mathbf{v}_1(\mathbf{x})^T \mathbf{Q} \mathbf{v}_1(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^d,$$

for some (unique) Gram matrix $\mathbf{Q} > 0$. 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

$$p(\mathbf{x}) = \Lambda_1^\mu(\mathbf{x})^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

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

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

are orthonormal w.r.t. the Gaussian (product measure) $d\mu(\mathbf{x}) = \exp(-\sum_j x_j^2/2) d\mathbf{x}$, 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

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

one expresses them in the same language of *polynomial densities* w.r.t. a reference finite Borel probability measure μ whose support is exactly the set $S \subset \mathbb{R}^d$ (assumed to be compact with nonempty interior). Let $\mathcal{P}(S)$ 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 w.r.t. μ .

Moment-SOS hierarchy of Lower bounds. With $\phi \in \mathbb{R}[\mathbf{y}]_{2n}^*$ arbitrary, and from the reproducing property (87) of K_{2n}^μ , observe that

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

where the degree- $2n$ polynomial

$$(105) \quad \mathbf{x} \mapsto \sigma_\phi(\mathbf{x}) := \sum_{\alpha \in \mathbb{N}_{2n}^d} \phi(P_\alpha) P_\alpha(\mathbf{x}),$$

is a *signed density* w.r.t. μ .

Next, recall that in the semidefinite relaxation (26) of the Moment-SOS hierarchy of lower bounds on f^* , one searches for a linear functional $\phi \in \mathbb{R}[\mathbf{x}]_{2n}^*$ which satisfies

$$\phi(\mathbf{1}) = 1; \quad \mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0, \quad \forall j = 0, \dots, m,$$

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

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

which means that σ_ϕ is a signed *probability density*. Therefore one has proved:

Theorem 7.5. *Let μ be a finite Borel measure whose support is S in (13) and consider the Moment-SOS hierarchy of semidefinite relaxations (26) for solving \mathbf{P} . Then with n fixed, (26) reads*

$$(106) \quad \min_{\phi \in \mathbb{R}^{s(2n)}} \left\{ \int_S f \sigma_\phi d\mu : \phi(\mathbf{1}) = 1; \mathbf{M}_{n-d_j}(g_j \cdot \phi) \geq 0, j = 0, \dots, m \right\},$$

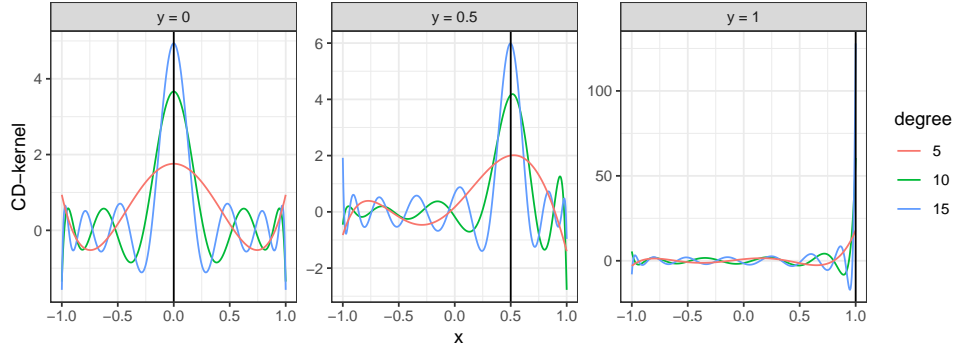
where σ_ϕ is the signed probability density w.r.t. μ in (105).

So, again, solving the semidefinite relaxation (26) in the Moment-SOS hierarchy is searching for a polynomial signed probability density $\sigma_\phi \in \mathbb{R}[\mathbf{x}]_{2n}$ of the form (105), and as already mentioned, when the relaxation (26) 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

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

It is interesting to visualize in Figure 7.5 how σ_{ϕ^*} looks like in the toy example where $S = [-1, 1]$ and $\mu = dx/2$. Indeed σ_{ϕ^*} has a peak at $x = y$, and so mimics the Dirac measure at y (as long as moments up to degree $2n$ are concerned).

FIGURE 7.3. $S = [-1, 1]$, $\mu = dx/2$; the signed density $x \mapsto \sigma_{\phi^*}(x) = K_{2n}^{\mu}(x, y)$ with $y = 0$ (left), $y = 0.5$ (middle) and $y = 1$ (right). © Reprinted from [70]



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 $(\hat{v}_n)_{n \in \mathbb{N}}$ in (58) 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

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

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

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

As one may see, Table 7.5 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 $(P_{\alpha})_{\alpha \in \mathbb{N}^d}$.

Lower bounds	Upper bounds
Primal	Primal
$\tau_n = \inf_{\phi} \int_S \left(\sum_{\alpha \in \mathbb{N}_{2n}^d} \phi(P_\alpha) P_\alpha \right) f d\mu$	$\hat{\kappa}_n = \inf_{\sigma, \psi_j} \int_S \left(\sum_{\alpha \in \mathbb{N}_{2n}^d} \sigma_\alpha P_\alpha \right) f d\mu$
<p>s.t. $\phi_0 (= \phi(\mathbf{1})) = 1;$</p> $\mathbf{M}_n(g_j \cdot \phi) \geq 0, 0 \leq j \leq m.$	<p>s.t. $\sigma_0 = 1;$</p> $\sum_{\alpha \in \mathbb{N}_{2n}^d} \sigma_\alpha P_\alpha = \sum_{j=0}^m \psi_j g_j.$ $\psi_j \in \Sigma[\mathbf{x}]_{n-d_j}, 0 \leq j \leq m$
Dual	Dual
$\tau_n^* = \sup_{\lambda, \psi_j} \lambda :$	$\hat{\kappa}_n^* = \sup_{\lambda, \phi} \lambda :$
<p>s.t. $f - \lambda = \sum_{j=0}^m \psi_j g_j$</p> $\psi_j \in \Sigma[\mathbf{x}]_{n-d_j}, 0 \leq j \leq m$	<p>s.t. $f - \lambda = \sum_{\alpha \in \mathbb{N}_{2n}^d} \phi(P_\alpha) P_\alpha$</p> <p style="text-align: center;">-</p> $\mathbf{M}_n(g_j \cdot \phi) \geq 0, 0 \leq j \leq m$

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 $\sum_{\alpha} \phi(P_{\alpha}) P_{\alpha}$, w.r.t. μ .

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 [124, 93, 70] and references therein.

Section 7.3 is based on [86] and [38].

Section 7.4 is essentially from [66, 67] and [73]. Remarkably, the generalized Pell's equation establishes links between seemingly unrelated fields like optimization, positivity certificates, conic duality on one side, and orthogonal polynomials and equilibrium measures on the other side. 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 [73], crucial is a property of Gegenbauer polynomials (in particular a summation property). However for more general basic

semi-algebraic sets S , there is still a weaker result that links the constant polynomial $\mathbf{1} \in Q_n(g)$ and moments $\phi \in Q_n^*(g)$ of the equilibrium measure μ of S ; see [73].

In Section 7.5 one interprets both Moment-SOS hierarchies of upper and lower bounds in the common language of densities w.r.t. 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 w.r.t. 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 w.r.t. μ on S (with nonempty interior) as in the hierarchy of upper bounds.

8. CONCLUSION

One has described the Moment-SOS hierarchy methodology for polynomial optimization (hierarchies of lower and upper bounds). It also applies solve the generalized Moment Problem (GMP) with algebraic data, whose list of applications in many area of Science and Engineering is almost endless. The basic principle behind the Moment-SOS hierarchy is quite simple and for illustration purpose we have described its application on 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, as large scale problems often exhibit sparsity and/or symmetries in their formulation, one has 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 have been briefly mentioned). As it is at the crossroad of several disciplines, it is very likely that we should see more and more contributions in the coming years.

REFERENCES

- [1] Ahmadi, A.A., Majumdar, A.: DSOS and SDSOS Optimization: More tractable alternatives to Sum of Squares and semidefinite optimization. *SIAM J. Appl. Algebra Geometry* **3**(2), 193–230 (2019)
- [2] Aloise, D., Hansen, P.: Evaluating a branch-and-bound RLT-based algorithm for minimum sum-of-squares clustering. *J. Global Optim.* **49**, 449–465 (2011)
- [3] Anjos, M., Lasserre, J.B. (eds.): *Handbook on Semidefinite, Conic and Polynomial Optimization*, *Internat. Ser. Oper. Res. Management Sci.*, vol. 166. Springer, New York (2012)
- [4] Aylward, E.M., Parrilo, P.A., Slotine, J.J.E.: Stability and robustness analysis of nonlinear systems via contraction metrics and sos programming. *Automatica* **44**(8), 2163–2170 (2008)
- [5] Bacari, F., Gogolin, C., Wittek, P., Acin, A.: Verifying the output of quantum optimizers ground-state energy lower bounds. *Physical Review Research* **2**, 043163 (2020)
- [6] Bach, F., Rudi, A.: Exponential convergence of Sum-of-Squares hierarchies for trigonometric polynomials. *SIAM J. Optim.* **33**(3), 2137–2159 (2023)
- [7] Bachoc, C., Passuello, A., Vallentin, F.: Bounds for projective codes from semidefinite programming. *Adv. Mathematics of Communications* **7**, 127–145 (2013)
- [8] Bachoc, C., Vallentin, F.: New upper bounds for kissing numbers from semidefinite programming. *J. Amer. Math. Soc.* **21**, 909–924 (2008)

- [9] Bafna, M., Barak, B., Kothari, P.K., Schramm, T., Steurer, D.: Playing unique games on certified small-set expanders. In: Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, pp. 1629–1642. Association for Computing Machinery (ACM), New York (2021)
- [10] Baldi, L., Mourrain, B.: Exact moment representation in Polynomial Optimization (2022). [arXiv:2012:14652](https://arxiv.org/abs/2012.14652)
- [11] Barak, B., Steurer, D.: Sum-of-squares proofs and the quest towards optimal algorithms. In: S.Y. Sang, Y.R. Kim, D.W. Lee, I. Tie (eds.) Proceedings of the International Congress of Mathematicians - Seoul 2014, pp. 509–533. Kyung Moon Sa, Seoul (2014)
- [12] Baran, M.: Complex equilibrium measure and Bernstein type theorems for compact sets in \mathbb{R}^n . Proc. Amer. Math. Soc. **123**(2), 485–494 (1995)
- [13] Bedford, E., Taylor, B.: The complex equilibrium measure of a symmetric convex set in \mathbb{R}^n . Trans. Amer. Math. Soc. **294**(2), 705–717 (1986)
- [14] Bertsimas, D., Popescu, I.: Optimal inequalities in probability theory. a convex optimization approach. SIAM J. Optim. **15**(3), 780–804 (2005)
- [15] Blekherman, G., Parrilo, P., Thomas, R. (eds.): Semidefinite Optimization and Convex Algebraic Geometry. MOS-SIAM Series on Optimization. SIAM, Philadelphia, PA (2012)
- [16] Burgdorf, S., Klep, I., Povh, J.: Optimization of Polynomials in Non-Commuting Variables. SpringerBriefs in Mathematics. Springer (2016)
- [17] de Castro, Y., Gamboa, F., Henrion, D., Hess, R., Lasserre, J.B.: Approximate optimal designs for multivariate polynomial regression. Ann. Statistics **47**, 125–157 (2019)
- [18] de Castro, Y., Gamboa, F., Henrion, D., Lasserre, J.B.: Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. IEEE Trans. Info. Theory **63**, 621–630 (2017)
- [19] Cicconet, F., Almeida, K.C.: Moment-SOS relaxation of the medium term hydrothermal dispatch problem. Int. J. Elec. Power Energy Systems **104**, 124–133 (2019)
- [20] Courtier, N., Drummond, R., Ascensio, P., Couto, L., Howey, D.: Discretisation-free battery fast-charging optimisation using the measure-moment approach. In: 2022 European Control Conference (ECC), July 2022, London, UK, pp. 628–634. IEEE (2022)
- [21] Doherty, A.C., Parrilo, P.A., Spedalieri, F.: Detecting multipartite entanglement. Phys. Rev. A **71**, 032333 (2005)
- [22] Dostert, M., Vallentin, F.: New dense superball packings in three dimensions. Adv. Geom. **20**, 473–482 (2020)
- [23] Dowdy, G.R., Barton, P.I.: Dynamics bounds on stochastic chemical kinetic systems using semidefinite programming. J. Chemical Physics **149**, 74103 (2018)
- [24] Dyer, M., Frieze, A.: The complexity of computing the volume of a polyhedron. SIAM J. Comp. **17**, 967–974 (1988)
- [25] Eckhoff, K.S.: Accurate and efficient reconstruction of discontinuous functions from truncated series expansions. Math. Comput. **61**(204), 745–763 (1993)
- [26] Eisert, J., Hyllus, P., Gühne, O., Curty, M.: Complete hierarchies of efficient approximations to problems in entanglement theory. Physical Review A **70**, 062317 (2004)
- [27] Fan, J., Nie, J., Zhou, A.: Tensor eigenvalue complementarity problems. Math. Program. Ser. A **170**, 507–539 (2018)
- [28] Fang, K., Fawzi, H.: The sum-of-squares hierarchy on the sphere and applications in quantum information theory. Math. Program. **190**, 331–360 (2021)
- [29] Fantuzzi, G., Goluskin, D., Huang, D., Chernyshenko, S.I.: Bounds for deterministic and stochastic dynamical systems using sum-of-squares optimization. SIAM J. Appl. Dyn. Systems **15**(4), 1962–1988 (2016)
- [30] Fattorini, H.O.: Infinite Dimensional Optimization. Cambridge University Press, Cambridge, UK (1999)
- [31] Gepp, A., Harris, G., Vanstone, B.: Financial applications of semidefinite programming: a review and call for interdisciplinary research. Accounting and Finance **60**(4), 3527–3555 (2020)

- [32] Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* **42**, 1115–1145 (1995)
- [33] Goluskin, D., Fantuzzi, G.: Bounds on mean energy in the Kuramoto-Sivashinsky equation computed using semidefinite programming. *Nonlinearity* **32**(5), 1705–1730 (2019)
- [34] Handelman, D.: Representing polynomials by positive linear functions on compact convex polyhedra. *Pac. J. Math.* **132**, 35–62 (1988)
- [35] Haussmann, D., Liers, F., Stingl, M., Vera, J.C.: Deciding robust feasibility and infeasibility using a set containment approach: An application to stationary passive gas network operations. *SIAM J. Optim.* **28**, 2489–2517 (2018)
- [36] Henrion, D., Garulli, A. (eds.): Positive Polynomials in Control, *Lecture Notes on Control and Information Sciences*, vol. 312. Springer Verlag, Berlin (2005)
- [37] Henrion, D., Korda, M., Lasserre, J.B.: The Moment-SOS Hierarchy: Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs. Optimization and Applications 4. World Scientific Co. Pte Ltd., Hackensack, NJ (2021)
- [38] Henrion, D., Lasserre, J.B.: Graph recovery from incomplete moment information. *Constr. Approx.* **56**, 165–187 (2022)
- [39] Henrion, D., Lasserre, J.B., Savorgnan, C.: Approximate volume and integration for basic semi-algebraic sets. *SIAM Review* **51**, 722–743 (2009)
- [40] Xiangfeng Ji, Xuegang (Jeff) Ban, Jian Zhang, Bin Ran: Moment-based travel time reliability assessment with Lasserre’s relaxation. *Transportmetrica B: Transport Dynamics* **7**(1), 401–422 (2019)
- [41] Khot, S.: Inapproximability of NP-complete problems, discrete Fourier analysis, and geometry. In: *Proceedings of the International Congress of Mathematicians*, vol. IV, pp. 2676–2697. Hindustan Book Agency, New Delhi, Hyderabad, India (2010)
- [42] Khot, S.: Hardness of approximation. In: *Proceedings of the International Congress of Mathematicians*, pp. 711–728. Kyung Moon Sa, Seoul, Korea (2014)
- [43] de Klerk, E., Laurent, M.: On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems. *SIAM J. Optim.* **21**(3), 824–832 (2011)
- [44] Kocvara, M., Murrain, B., Riemer, C. (eds.): *Polynomial Optimization, Moments, and Applications*. Springer Optimization and Its Applications. Springer Cham (2023)
- [45] Kojima, M., Kim, S., Maramatsu, M.: Sparsity in sums of squares of polynomials. *Math. Program.* **103**, 45–62 (2005)
- [46] Korda, M., Henrion, D., Jones, C.N.: Convex computation of the maximum controlled invariant set for polynomial control systems. *SIAM J. Control Optim.* **52**(5), 2944–2969 (2014)
- [47] Korda, M., Henrion, D., Lasserre, J.B.: Moments and convex optimization for analysis and control of nonlinear PDEs. In: E. Zuazua, E. Trélat (eds.) *Handbook of Numerical Analysis: Numerical Control: Part A*, vol. 23, pp. 339–366. Elsevier (2022)
- [48] Krivine, J.L.: Anneaux préordonnés. *J. Anal. Math.* **12**, 307–326 (1964)
- [49] Krivine, J.L.: Quelques propriétés des préordres dans les anneaux commutatifs unitaires. *C.R. Acad. Sci. Paris* **258**, 3417–3418 (1964)
- [50] de Laat, D.: Moment methods in energy minimization: New bounds for Riesz minimal energy problems. *Trans. Amer. Math. Soc.* **373**, 1407–1453 (2020)
- [51] de Laat, D., Machado, F.C., de Oliveira Filho, F.M., Vallentin, F.: k -point semidefinite programming bounds for equiangular lines. *Math. Program.* **194**, 533–56 (2022)
- [52] Landau, H.: Classical background of the moment problem. In: *Moments in Mathematics*, pp. 1–15. American Mathematical Society, Providence, RI (1987)
- [53] Laraki, R., Lasserre, J.B.: Semidefinite programming for min-max problems and games. *Math. Program. Ser. A* **131**, 305–332 (2012)
- [54] Lasserre, J.B.: Optimisation globale et théorie des moments. *C. R. Acad. Sci. Paris, Série I* **331**, 929–934 (2000)
- [55] Lasserre, J.B.: Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* **11**, 796–817 (2001)

- [56] Lasserre, J.B.: Bounds on measures satisfying moment conditions. *Annals Appl. Prob.* **12**, 1114–1137 (2002)
- [57] Lasserre, J.B.: Semidefinite programming vs. LP relaxations for polynomial programming. *Math. Oper. Res.* **27**, 347–360 (2002)
- [58] Lasserre, J.B.: Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM J. Optim.* **17**, 822–843 (2006)
- [59] Lasserre, J.B.: Convexity in semi-algebraic geometry and polynomial optimization. *SIAM J. Optim.* **19**(4), 1995–2014 (2009)
- [60] Lasserre, J.B.: *Moments, Positive Polynomials and Their Applications*. Imperial College Press, London (2009)
- [61] Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. *SIAM J. Optim.* **21**, 864–885 (2011)
- [62] Lasserre, J.B.: The K-moment problem for continuous functionals. *Trans. Amer. Math. Soc.* **365**, 2489–2504 (2013)
- [63] Lasserre, J.B.: *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge University Press, Cambridge, UK (2015)
- [64] Lasserre, J.B.: Computing gaussian & exponential measures of semi-algebraic sets. *Adv. Appl. Math.* **91**, 137–163 (2017)
- [65] Lasserre, J.B.: Connecting optimization with spectral analysis of tri-diagonal matrices. *Math. Program. Ser. A* **190**, 795–809 (2021)
- [66] Lasserre, J.B.: A disintegration of the Christoffel function. *Comptes Rendus Mathématique* **360**, 1071–1079 (2022)
- [67] Lasserre, J.B.: Pell’s equation, sum-of-squares and equilibrium measures on a compact set. *Comptes Rendus Mathématique* **361**, 935–952 (2023)
- [68] Lasserre, J.B., Henrion, D., Prieur, C., Trélat, E.: Nonlinear optimal control via occupation measures and LMI-relaxations. *SIAM J. Control Optim.* **47**, 1649–1666 (2008)
- [69] Lasserre, J.B., Pauwels, E.: Sorting out typicality via the inverse moment matrix sos polynomial. In: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (eds.) *Advances in Neural Information Processing Systems 29*, pp. 190–198. Curran Associates, Inc., NIP2016, Barcelona (2016)
- [70] Lasserre, J.B., Pauwels, E., Putinar, M.: *The Christoffel Function for Data Analysis*. Cambridge University Press, Cambridge, UK (2022)
- [71] Lasserre, J.B., Priéto-Rumeau, T.: SDP vs. LP relaxations for the Moment approach in some performance evaluation problems. *Stoch. Models* **20**, 439–456 (2004)
- [72] Lasserre, J.B., Priéto-Rumeau, T., Zervos, M.: Pricing a class of exotic options via moments and sdp relaxations. *Math. Finance* **16** (2006)
- [73] Lasserre, J.B., Xu, Y.: A generalized Pell’s equation for a class of multivariate orthogonal polynomials. Tech. rep., LAAS-CNRS, Toulouse, France (2023). Hal-04163153v1
- [74] Latorre, F., Rolland, P., Cevher, V.: Lipschitz constant estimator of neural networks via sparse polynomial optimization. In: *Proceedings of the ICLR, December 2020* (2020)
- [75] Laurent, M.: A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.* **28**, 470–496 (2003)
- [76] Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. In: M. Putinar, S. Sullivant (eds.) *Emerging Applications of Algebraic Geometry*, vol. 149, pp. 157–270. Springer, New York, NY (2008)
- [77] Laurent, M., Slot, L.: An effective version of Schmüdgen’s Positivstellensatz for the hypercube. *Optim. Letters* **17**, 515–530 (2023)
- [78] Magron, V., Wang, J.: *Sparse Polynomial Optimization: Theory and Practice*. Series on Optimization and Its Applications. World Scientific Press, London (2023)
- [79] Ngoc Hoang Anh Mai, Lasserre, J.B., Magron, V., Wang, J.: Exploiting constant trace property in large scale polynomial optimization. *ACM Trans. Math. Software* **48**(4), 1–39, art. 40 (2022)
- [80] Mai, N.H.A., Lasserre, J.B., Magron, V.: A hierarchy of spectral relaxations for polynomial optimization. *Math. Program. Comput.* (2023). To appear.

- [81] Majumdar, A., Hall, G., Ahmadi, A.A.: A survey of recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics. *Annual Rev. Control Robot Auton. Syst.* **3**, 331–60 (2020)
- [82] Marmin, A., Castella, M., Pesquet, J.C., Duval, L.: Sparse signal reconstruction for nonlinear models via piecewise rational optimization. *Signal Processing* **179** (2021)
- [83] Marschner, Z., Palmer, D., Zhang, P., Solomon, J.: Hexahedral mesh repair via sum-of-squares relaxations. In: *Proceedings of Eurographics Symposium on Geometry Processing*, vol. 39, pp. 133–147 (2020)
- [84] Marschner, Z., Palmer, D., Zhang, P., Solomon, J.: Sum-of-squares geometry processing. *ACM Transactions on Graphics* **40**(6), 1–13 (2021)
- [85] Marshall, M.: *Positive Polynomials and Sums of Squares, AMS Math. Surveys and Monographs*, vol. 146. American Mathematical Society, Providence, RI (2008)
- [86] Marx, S., Pauwels, E., Weisser, T., Henrion, D., Lasserre, J.B.: Semi-algebraic approximation using Christoffel-Darboux kernel. *Constr. Approx.* **54**, 391–429 (2021)
- [87] Marx, S., Weisser, T., Henrion, D., Lasserre, J.B.: A moment approach for entropy solutions to nonlinear hyperbolic PDEs. *Math. Control Related Fields* **10**(1), 113–140 (2020)
- [88] Molzahn, D., Jozs, C.: Lasserre hierarchy for large-scale polynomial optimization in real and complex variables. *SIAM J. Optim.* **28**, 1017–1048 (2018)
- [89] Molzahn, D., Hiskens, I.: Sparsity-exploiting moment-based relaxations of the optimal power flow problem. *IEEE Trans. Power Systems* **30**(6), 3168–3180 (2015)
- [90] Moussa, K., Fiacchini, M., Alamir, M.: Robust optimal scheduling of combined chemo- and immunotherapy: Considerations on chemotherapy detrimental effects. In: *Proceedings of the American Control Conference*, Denver, CO, pp. 4252–4257. IEEE (2020)
- [91] Navascués, M., Pironio, S., Acín, A.: SDP relaxations for non-commutative polynomial optimization. In: M. Anjos, J.B. Lasserre (eds.) *Handbook on Semidefinite, Conic and Polynomial Optimization, Internat. Ser. Oper. Res. Management Sci.*, vol. 166, pp. 601–634. Springer, New York (2012)
- [92] Nesterov, Y.: Squared functional systems and optimization problems. In: H. Frenk, K. Roos, T. Terlaky, S. Zhang (eds.) *High Performance Optimization*, pp. 405–440. Kluwer, Dordrecht (2000)
- [93] Nevai, P., Freud, G.: Orthogonal polynomials and christoffel functions. *J. Approx. theory* **48**, 3–167 (1986)
- [94] Nie, J.: Sum of Squares method for sensor network localization. *Comput. Optim. Appl.* **43**(2), 151–179 (2009)
- [95] Nie, J.: Certifying convergence of Lasserre’s hierarchy via flat truncation. *Math. Program. Ser. A* **142**(1-2), 485–510 (2013)
- [96] Nie, J.: Optimality conditions and finite convergence of Lasserre’s hierarchy. *Math. Program. Ser. A* **146**(2), 97–121 (2014)
- [97] Nie, J.: Low rank symmetric tensor approximations. *SIAM J. Matrix Anal. Appl.* **38**(4), 1517–1540 (2017)
- [98] Nie, J.: *Moment and Polynomial Optimization, MOS/SIAM series in Optimization*, vol. 31. SIAM, Philadelphia (2023)
- [99] Nie, J., Demmel, J.: Sparse SOS relaxations for minimizing functions that are summations of small polynomials. *SIAM J. Optim.* **19**(4), 1534–1558 (2008)
- [100] Nie, J., Tang, X.: Convex generalized nash equilibrium problems and polynomial optimization. *Math. Program.* **198**, 1485–1518 (2023)
- [101] Nie, J., Yang, Z.: Hermitian tensor decompositions. *SIAM J. Matrix Anal. Appl.* **41**(3), 1115–1144 (2020)
- [102] O’Donnell, R.: SoS is not fully automatisable, even approximately. In: H. Papadimitriou (ed.) *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, Leibniz Intern. Proc. Informatics (LIPIcs), vol. 67, pp. 59:1–59:10. Dagstuhl, Germany (2017)
- [103] Jie Pan, Fu Jiang: Low complexity beamspace super resolution for DOA estimation of linear array. *Sensors* **20**(8), 2222 (2020)

- [104] Parekh, O.: Synergies between operations research and quantum information science. *INFORMS J. Comput.* **35**(2), 266–273 (2023)
- [105] Parekh, O., Thompson, K.: Application of the level-2 quantum Lasserre hierarchy in quantum approximation algorithms. In: *LIPICs. Leibniz Int. Proc. Inform.*, vol. 198. Schloss Dagstuhl, Leibniz-Zentrum für Informatik, Wadern (2021). Art. No. 102, 20 pp.
- [106] Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology, Pasadena, CA (2000)
- [107] Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. *Math. Program.* **96**, 293–320 (2003)
- [108] Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. *Math. Program. Ser. B* **96**(2), 293–320 (2003)
- [109] Parrilo, P.A., Lall, S.: Semidefinite programming relaxations and algebraic optimization in control. *Eur. J. Control* **9**(2-3) (2003)
- [110] Pauwels, E., Henrion, D., Lasserre, J.B.: Positivity certificates in optimal control. In: J.P. Laumond, N. Mansard, J.B. Lasserre (eds.) *Geometric and Numerical Foundations of Movements, Springer Tracts in Advanced Robotics*, vol. 117, pp. 113–132. Springer (2017)
- [111] Prestel, A., Denzel, C.N.: *Positive Polynomials. Springer Monographs in Mathematics.* Springer, Berlin Heidelberg (2001)
- [112] Probst, T., Paudel, D.D., Chhatkuli, A., Van Gool, L.: Convex relaxations for consensus and non-minimal problems in 3D vision. In: *Proceedings of the IEEE Int. Conference on Computer Vision (ICCV)*, 2019, pp. 10233–10242. IEEE (2019)
- [113] Putinar, M.: Positive polynomials on compact semi-algebraic sets. *Ind. Univ. Math. J.* **42**, 969–984 (1993)
- [114] Raghavendra, P., Schramm, T., Steurer, D.: High dimensional estimation via sum-of-squares proofs. In: B. Sirakov, P.N. de Sousa, M. Viana (eds.) *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pp. 3389–3423. World Scientific Publishing Co. Pte Ltd., Hackensack, NJ (2018)
- [115] Rosen, D.M., Carlone, L., Bandera, A.S., Leonard, J.J.: Se-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. *The Int. J. Robotics Research* **38**(2-3), 95–125 (2019)
- [116] Schmüdgen, K.: The K-moment problem for compact semi-algebraic sets. *Math. Ann.* **289**, 203–206 (1991)
- [117] Schmüdgen, K.: *The Moment Problem. Graduate Texts in Mathematics.* Springer Cham (2017)
- [118] Schweighofer, M.: On the complexity of Schmüdgen’s positivstellensatz. *J. Complexity* **20**, 529–543 (2005)
- [119] Sedighi, S., Mishra, K., Shankar, B., Ottersten, M.R.B.: Localization with 1-Bit passive radars in Narrow Band IoT-applications using multivariate polynomial optimization. *IEEE Trans. Signal Processing* **69**, 2525–2540 (2021)
- [120] Selby, J.H., Sainz, A.B., Magron, V., Czekaj, L., Horodecki, M.: Correlations constrained by composite measurements. *Quantum* **7**, 1080 (2023)
- [121] Sherali, H.D., Adams, W.P.: A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discr. Math.* **3**, 411–430 (1990)
- [122] Sherali, H.D., Adams, W.P.: *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems.* Kluwer Academic Publishers, Dordrecht/Boston/London (1999)
- [123] Shor, N.Z.: Quadratic optimization problems. *Tekhnicheskaya Kibernetika* **1**, 128–139 (1987)
- [124] Simon, B.: The Christoffel-Darboux kernel. In: *Perspective in Partial Differential Equations, Harmonic Analysis and Applications: Proc. Symp. Pure Math.*, vol. 79, pp. 295–335. American Mathematical Society, Providence, RI (2008)
- [125] Slot, L.: Sum-of-Squares hierarchies for polynomial optimization and the Christoffel-Darboux kernel. *SIAM J. Optim* **32**(4), 2612–2635 (2022)

- [126] Slot, L., Laurent, M.: Near optimal analysis of Lasserre’s univariate measure-based bounds for multivariate polynomial optimization. *Math. Program.* **188**, 443–460 (2021)
- [127] Stein, N., Ozdaglar, A., Parrilo, P.A.: Separable and low-rank continuous games. *Int. J. Game Theory* **37**(4), 457–474 (2008)
- [128] Tacchi, M., Lasserre, J.B., Henrion, D.: Stokes, Gibbs and volume computation of semi-algebraic sets. *Discrete Comput. Geom.* **69**, 260–283 (2023)
- [129] Tacchi, M., Weisser, T., Lasserre, J.B., Henrion, D.: Exploiting sparsity in semi-algebraic set volume computation. *Found. Comput. Math.* **22**, 161–209 (2021)
- [130] Tian, J., Wei, H., Tan, J.: Global optimization for power dispatch problems based on theory of moments. *Electrical Power and Energy Systems* **71**, 184–194 (2015)
- [131] Chen Tong, Lasserre, J.B., Magron, V., Pauwels, E.: Semialgebraic representation of monotone deep equilibrium models and application to certification. In: *Proceedings of NEURIPS 2021* (2021)
- [132] Tyburec, M., Zeman, J., Kruzik, M., Henrion, D.: Global optimality in minimum compliance topology optimization of frames and shells by moment-sum-of-squares hierarchy. *Structural and Multidisciplinary Optimization* **64**, 1963–1981 (2021)
- [133] Vasilescu, F.H.: Spectral measures and moment problems. In: *Spectral Theory and Its Applications*, Theta Ser. Adv. Math. 2, pp. 173–215. Theta, Bucharest (2003)
- [134] Vinter, R.: Convex duality and nonlinear optimal control. *SIAM J. Control Optim.* **31**, 518–538 (1993)
- [135] Waki, S., Kim, S., Kojima, M., Marmatsu, M.: Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optim.* **17**, 218–242 (2006)
- [136] Wang, J., Li, H., Xia, B.: A new sparse SOS decomposition algorithm based on term sparsity. In: *Proceedings of the 2019 on International Symposium on Symbolic and Algebraic Computation*, pp. 347–354 (2019)
- [137] Wang, J., Magron, V., Lasserre, J.B.: TSSOS: A moment-SOS hierarchy that exploits term sparsity. *SIAM J. Optim.* **31**(1), 30–58 (2021)
- [138] Wang, J., Magron, V., Lasserre, J.B., Hong Mai, N.: CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization. *ACM Trans. Math. Software* **48**(4), 1–26 (2022)
- [139] Wittek, P., Darányi, S., Nelhans, G.: Ruling out static latent homophily in citation networks. *Scientometrics* **110**, 765–777 (2017)
- [140] Yang, H., Carlone, L.: In perfect shape: Certifiably optimal 3D shape reconstruction from 2D landmarks. In: *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, (CVPR)*, Seattle, WA, pp. 618–627. IEEE (2020)
- [141] Yang, H., Carlone, L.: Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence* **45**(3), 2816–2834 (2023)
- [142] Yang, H., Pavone, M.: Object pose estimation with statistical guarantees: Conformal keypoint detection and geometric uncertainty propagation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2023*. IEEE (2023)
- [143] Yang, H., Shi, J., Carlone, L.: TEASER: Fast and certifiable point cloud registration. *IEEE Trans. Robotics* **37**(2), 314–333 (2020)
- [144] Yurtsever, A., Tropp, J.A., Fercoq, O., Udell, M., Cevher, V.: Scalable Semidefinite Programming. *SIAM J. Math. Data Sci.* **3**(1), 171–200 (2021)

LAAS-CNRS AND INSTITUTE OF MATHEMATICS, UNIVERSITY OF TOULOUSE,, 7 AVENUE DU COLONEL ROCHE, BP54200, 31031 TOULOUSE CÉDEX 4, FRANCE
Email address: lasserre@laas.fr