HAL
open science

On the Effective Putinar's Positivstellensatz and Moment Approximation

Lorenzo Baldi, Bernard Mourrain

To cite this version:

Lorenzo Baldi, Bernard Mourrain. On the Effective Putinar's Positivstellensatz and Moment Approximation. Mathematical Programming, Series A, 2022, 10.1007/s10107-022-01877-6 . hal-03437328v4

HAL Id: hal-03437328

https://hal.science/hal-03437328v4

Submitted on 19 Jan 2023

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.

On the Effective Putinar's Positivstellensatz and Moment Approximation*

Lorenzo Baldi†, Bernard Mourrain
Inria at Université Côte d'Azur, Sophia Antipolis, France

January 19, 2023

Abstract

We analyse the representation of positive polynomials in terms of Sums of Squares. We provide a quantitative version of Putinar's Positivstellensatz over a compact basic semialgebraic set S, with a new polynomial bound on the degree of the positivity certificates. This bound involves a Łojasiewicz exponent associated to the description of S. We show that if the gradients of the active constraints are linearly independent on S (Constraint Qualification condition), this Łojasiewicz exponent is equal to 1. We deduce the first general polynomial bound on the convergence rate of the optima in Lasserre's Sum-of-Squares hierarchy to the global optimum of a polynomial function on S, and the first general bound on the Hausdorff distance between the cone of truncated (probability) measures supported on S and the cone of truncated pseudo-moment sequences, which are positive on the quadratic module of S.

1 Introduction

A fundamental question in Real Algebraic Geometry is how to describe effectively the set of polynomials which are positive[1] on a given domain.

Clearly, the set of positive polynomials on $\mathbb{R}^n$ contains the *Sums of Squares* of real polynomials (SoS). Let $\mathbb{R}[\mathbf{X}] = \mathbb{R}[X_1,\dots,X_n]$ be the $\mathbb{R}$-algebra of polynomials in the indeterminates $X_1,\dots,X_n$ with real coefficients. The convex cone of SoS polynomials

$$\Sigma^2 := \Sigma^2[\mathbf{X}] = \left\{ f \in \mathbb{R}[\mathbf{X}] \mid \exists r \in \mathbb{N},\ g_i \in \mathbb{R}[\mathbf{X}] \colon f = g_1^2 + \cdots + g_r^2 \right\}$$

is a subset of the convex cone of positive polynomials $\mathrm{Pos}(\mathbb{R}^n) := \{p \in \mathbb{R}[\mathbf{X}] \mid \forall x \in \mathbb{R}^n, p(x) \geq 0\}$. But it is known since Hilbert [Hil88], that these cones differ: not all positive polynomials are SoS. A famous counter-example is Motzkin polynomial [Mot67] of degree 6 in 2 variables, which is positive on $\mathbb{R}^2$ but not a SoS. Such polynomials exists for any dimension $n \geq 2$, but not in dimension 1 since univariate positive polynomials are SoS.

For a domain $S = \mathcal{S}(\mathbf{g}) = \mathcal{S}(g_1,\dots,g_r) = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0 \text{ for } i = 1,\dots,r\}$, defined by inequalities $g_i \geq 0$ with $g_i \in \mathbb{R}[\mathbf{X}]$, that is, a *basic closed semialgebraic set*, the set $\mathrm{Pos}(S)$ of positive polynomials on S contains the *quadratic module* generated by the tuple of polynomials $\mathbf{g} = (g_1,\dots,g_r)$, and defined by

$$Q = \mathcal{Q}(\mathbf{g}) := \Sigma^2 + \Sigma^2 \cdot g_1 + \cdots + \Sigma^2 \cdot g_r$$

and also the *preordering* $O = \mathcal{O}(g_1,\dots,g_r) := \mathcal{Q}(\prod_{j\in J} g_j \colon J \subset \{1,\dots,r\})$.

A complete description of positive polynomials on S is given by the Krivine–Stengle Positivstellensatz:

*This work has been supported by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions, grant agreement 813211 (POEMA)

†Corresponding Author: Lorenzo Baldi, Affiliation: Centre Inria Sophia Antipolis - Méditerranée, E-mail Address: lorenzo.baldi@inria.fr

[1] We follow the French tradition, and call a function f *positive* on a domain D if $f \geq 0$ on D and *strictly positive* on D if $f > 0$ on D.

Theorem 1.1 ([Kri64],[Ste74]). *Let* $\mathbf{g}$ *be a tuple of polynomials and* $S = S(\mathbf{g})$. *Then:*

$$\mathrm{Pos}(S) = \{p \in \mathbb{R}[\mathbf{X}] \mid \exists s \in \mathbb{N},\ q_1, q_2 \in \mathcal{O}(\mathbf{g}) \text{ s.t. } q_1 p = p^{2s} + q_2\}$$

This result is an extension of Artin's theorem [Art27], stating that globally positive polynomials are ratio of two SoS polynomials. But it induces a *denominator* in the representation of a positive polynomial.

Since for a general tuple $\mathbf{g}$ and $n > 1$, positive polynomials on $S(\mathbf{g})$ do not all belong to the quadratic module $Q(\mathbf{g})$ or even to the preordering $\mathcal{O}(\mathbf{g})$, it is natural to ask whether the convex cone $Q(\mathbf{g})$ (resp. $\mathcal{O}(\mathbf{g})$) is a good inner-approximation of $\mathrm{Pos}(S)$. A partial answer is given by two important results due to Schmüdgen and Putinar, and also known as denominator free Positivstellensatz. They require the following assumption:

Definition 1.2. Denote $\|\mathbf{X}\|_2^2 = X_1^2 + \cdots + X_n^2$. We say that a quadratic module Q is *Archimedean* if there exists $r \in \mathbb{R}$ such that $r^2 - \|\mathbf{X}\|_2^2 \in Q$.

We recall these two results, which are central in the paper:

Theorem 1.3 (Schmüdgen's Positivstellensatz [Sch91]). *Let* $\mathcal{S}(\mathbf{g})$ *be a compact basic semialgebraic set. Then* $f > 0$ *on* $\mathcal{S}(\mathbf{g})$ *implies* $f \in \mathcal{O}(\mathbf{g})$.

Theorem 1.4 (Putinar's Positivstellensatz [Put93]). *Let* $\mathcal{S}(\mathbf{g})$ *be a basic closed semialgebraic set. If* $\mathcal{Q}(\mathbf{g})$ *is Archimedean, then* $f > 0$ *on* $\mathcal{S}(\mathbf{g})$ *implies* $f \in \mathcal{Q}(\mathbf{g})$.

As a consequence of the first result, notice that $\mathcal{S}(\mathbf{g})$ compact implies that $\mathcal{O}(\mathbf{g})$ is Archimedean. On the other hand there are examples with $\mathcal{S}(\mathbf{g})$ compact but $\mathcal{Q}(\mathbf{g})$ not Archimedean (see e.g. [PD01, ex. 6.3.1]).

Since a positive polynomial $f \in \mathrm{Pos}(S)$ on a compact basic semialgebraic set S can be approximated uniformly on S by the polynomial $f + \varepsilon$, which is strictly positive on S for $\varepsilon > 0$, these results show that any $f = \lim_{\varepsilon \to 0} f + \varepsilon$ positive on S is the limit of polynomials in $Q(\mathbf{g})$ (resp. $\mathcal{O}(\mathbf{g})$). Unfortunately, the degree of the representation of $f + \varepsilon$ in $\mathcal{Q}(\mathbf{g})$ goes to infinity as $\varepsilon \to 0$, see [Ste96].

In this paper, we provide quantitative versions of Theorem 1.4. We give new bounds on the degree of the representation in $Q(\mathbf{g})$, which control the quality of approximation of positive polynomials by elements in $Q(\mathbf{g})$. For this problem, also known as *Effective Putinar Positivstellensatz*, our main result is Theorem 1.7, which provides the first polynomial bounds in the intrinsic parameters associated to $\mathbf{g}$ and f.

The proof of Theorem 1.7 is developed in Section 2 and Section 3. In the proof and in the bound of the theorem a special role is played by the Łojasiewicz exponent $Ł$, comparing the distance and the algebraic distance from S, see Definition 2.4. In Theorem 2.11 we prove that $Ł = 1$ in regular cases, i.e. when a regularity condition coming from Optimization is satisfied, see Definition 2.7. To our best knowledge this is the first analysis of the Łojasiewicz exponent under regularity assumptions of any kind. Corollaries to our main results in regular cases with $Ł = 1$ are described in Corollary 3.9, Corollary 4.4 and Corollary 5.9.

Polynomials whose representation in the quadratic module $Q(\mathbf{g})$ is of degree bounded by 2ℓ, $\ell \in \mathbb{N}$, are used to define a hierarchy of convex optimization problems, also known as Lasserre's hierarchy, whose optimum value converges to the *global optimum* of a polynomial $f \in \mathbb{R}[\mathbf{X}]$ on S under the Archimedean assumption [Las01]. We describe these hierarchies in Section 4. In Theorem 4.2 and Theorem 4.3, we deduce from Theorem 1.7 new polynomial bounds on the convergence rate of this hierarchy to the global optimum, in terms of the order ℓ of the hierarchy.

Considering the dual problem, we also analyse the quality of approximation of measures by truncated pseudo-moment sequences used in Lasserre moment hierarchy. In Theorem 1.8, we provide new bounds on the Hausdorff distance between the cone of truncated probability measures (supported on S) and the outer convex set of truncated positive pseudo-moment sequences of unit mass, and on the rate of convergence when the order ℓ goes to infinity. The proof of Theorem 1.8 is developed in Section 5. The bounds involve intrinsic parameters associated to $\mathbf{g}$ and the degree t of truncation. As an intermediate step, in Theorem 5.7 we also bound the Hausdorff distance between truncated positive pseudo-moment sequences and non-normalized measures.

1.1 Truncated quadratic modules and positive polynomials

To analyse the degree in these SoS representations, we introduce the *truncated* quadratic modules at degree (or *level*) $\ell \in \mathbb{N}$, i.e. the polynomials in $\mathcal{Q}(\mathbf{g})$ that are generated in degree $\leq \ell$:

$$\mathcal{Q}_\ell(\mathbf{g}) = \Big\{ s_0 + \sum_{i=1}^{r} s_i g_i \mid s_j \in \Sigma^2,\ \deg s_0 \leq \ell,\ \deg s_i g_i \leq \ell\ \forall i = 1, \dots r \Big\} \subset \mathcal{Q}(\mathbf{g}) \cap \mathbb{R}[\mathbf{X}]_\ell. \tag{1}$$

where $\mathbb{R}[\mathbf{X}]_\ell$ is the vector space of polynomials of degree $\leq \ell$.

Effective versions of Schmüdgen and Putinar's Positivstellensatz, that give degree bounds for the representation in truncaded preorderings and quadratic modules, have been proved by Schweighofer and Nie.

Theorem 1.5 ([Sch04]). *For all $\mathbf{g} = g_1, \dots g_r \subset \mathbb{R}[\mathbf{X}] = \mathbb{R}[X_1, \dots, X_n]$ defining $\emptyset \neq \mathcal{S}(\mathbf{g}) = S \subset (-1,1)^n$ there exists $0 < c \in \mathbb{R}$ (depending on $\mathbf{g}$ and n) such that, if $f \in \mathbb{R}[\mathbf{X}]_d$ is strictly positive on S with minimum $f^* = \min_{x \in S} f(x) > 0$, we have $f \in \mathcal{O}_\ell(\mathbf{g})$ if*

$$\ell \geq c d^2 \Big(1 + \Big(d^2 n^d \frac{\|f\|_{\mathbf{X}}}{f^*} \Big)^c \Big).$$

Theorem 1.6 ([NS07]). *For all $\mathbf{g} = g_1, \dots g_r \subset \mathbb{R}[\mathbf{X}] = \mathbb{R}[X_1, \dots, X_n]$ defining an Archimedean quadratic module $Q = \mathcal{Q}(\mathbf{g})$ and $\emptyset \neq \mathcal{S}(\mathbf{g}) = S \subset (-1,1)^n$, there exists $0 < c \in \mathbb{R}$ (depending on $\mathbf{g}$ and n) such that, if $f \in \mathbb{R}[\mathbf{X}]_d$ is strictly positive on S with minimum $f^* = \min_{x \in S} f(x) > 0$, we have $f \in \mathcal{Q}_\ell(\mathbf{g})$ if*

$$\ell \geq c \exp\Bigg(\Big(d^2 n^d \frac{\|f\|_{\mathbf{X}}}{f^*} \Big)^c \Bigg).$$

The norm $\|\cdot\|_{\mathbf{X}}$ used in [Sch04] and [NS07] is the max norm of the coefficients of the polynomial f w.r.t. the weighted monomial basis $\{\frac{|\alpha|!}{\alpha_1! \dots \alpha_n!} \mathbf{X}^\alpha : |\alpha| \leq d\}$, while the one we will use is the max norm on $[-1,1]^n$. We describe this norm and fix some notation.

Notation. Throughout the article:

- $f \in \mathbb{R}[\mathbf{X}]$ is a polynomial in n variables of degree $d = d(f)$;
- $S = \mathcal{S}(\mathbf{g}) = \mathcal{S}(g_1, \dots, g_r)$ is the basic closed semialgebraic set defined by $\mathbf{g} = g_1, \dots, g_r$;
- $d(\mathbf{g}) = \max_i \deg g_i$ is the maximum degree of the inequalities defining S;
- $f^* = \min_{x \in S} f(x)$ is the minimum of f on S, and unless otherwise stated $f^* > 0$;
- $\|\cdot\|$ denotes the max norm of a polynomial on $[-1,1]^n$, i.e. $\|h\| = \max_{x \in [-1,1]^n} |h(x)|$;
- $\varepsilon(f) = \frac{f^*}{\|f\|}$ is a measure of how close is f to have a zero on S.

For convenience we will prove our theorem in a normalized setting.

Normalisation assumptions. In the following, we assume that

$$\begin{aligned} &\bullet\ \ 1 - \|\mathbf{X}\|_2^2 \in \mathcal{Q}(\mathbf{g}), \\ &\bullet\ \ \|g_i\| \leq \tfrac{1}{2}\ \forall i \in \{1, \dots, r\}. \end{aligned} \tag{2}$$

We can always be in this setting by a change of variables if we start with an Archimedean quadratic module: if $r^2 - \|\mathbf{X}\|^2 \in \mathcal{Q}(\mathbf{g})$ then $1 - \|\mathbf{X}\|^2 \in \mathcal{Q}(\mathbf{g}(r\mathbf{X}))$ (i.e. the quadratic module generated by $g_i(rX_1, \dots, rX_n)$). By replacing g_i with $\frac{g_i}{2\|g_i\|}$, we can also assume without loss of generality that the second point is satisfied.

The main result of the paper is the following theorem. It is an effective, general version of Putinar's Positivstellensatz with polynomial bounds for fixed n.

Theorem 1.7. *Assume $n \geq 2$ and let $g_1, \dots, g_r \in \mathbb{R}[\mathbf{X}] = \mathbb{R}[X_1, \dots, X_n]$ satisfying the normalization assumptions* (2)*. Let $f \in \mathbb{R}[\mathbf{X}]$ such that $f^* = \min_{x \in S} f(x) > 0$. Let $\mathfrak{c}$, $Ł$ be the Łojasiewicz coefficient and exponent given by Definition 2.4. Then $f \in \mathcal{Q}_\ell(\mathbf{g})$ if*

$$\begin{aligned}\ell &\geq O(n^3 2^{5nŁ} r^n \mathfrak{c}^{2n} d(\mathbf{g})^n d(f)^{3.5nŁ} \varepsilon(f)^{-2.5nŁ}) \\ &= \gamma(n, \mathbf{g})\, d(f)^{3.5nŁ} \varepsilon(f)^{-2.5nŁ},\end{aligned}$$

where $\gamma(n, \mathbf{g}) \geq 1$ depends only on n and $\mathbf{g}$.

Notice that the only parameters in the bound that depend on f are $d(f)$ and $\varepsilon(f)$. We also remark that exponents in Theorem 1.7 have been simplified for the sake of readability and are not optimal: see Equation (21) for sharper bounds, especially for the case $n \gg 0$. Moreover we remark that the assumption $n \geq 2$, only used to do this simplification, is not a serious limitation since the univariate case is already well studied, see for instance [PR00].

In the definition of $\varepsilon(f)$ we use the max norm $\|\cdot\|$ on $[-1,1]^n$ instead of $\|\cdot\|_{\mathbf{X}}$ used in [NS07], because it does not depend on the choice of a basis and on the representation of the polynomials. However, for polynomials of bounded degree, the two norms are equivalent. Using [NS07, lem. 7] to express our bound with $\|\cdot\|_{\mathbf{X}}$ would result in an extra factor $2^{2.5nŁ} n^{2.5d(f)nŁ} d(f)^{2.5nŁ}$, while keeping the exponent of $\varepsilon(f)$.

In Section 3 we develop in detail the proof of Theorem 1.7. The ingredients for the proof are introduced in Section 2. The main differences with the one of [NS07] is the use of an effective Schmüdgen's Positivstellensatz on the unit box [LS21], and an effective approximation of regular functions on the unit interval, see Theorem 2.12. Moreover in Section 2.2 we prove that the main exponent of the bound, i.e. the Łojasiewicz exponent $Ł$, is equal to 1 for regular polynomial optimization problems, see Definition 2.7 and Theorem 2.11. The corollary of Theorem 1.7 in these regular cases is Corollary 3.9.

As a corollary of Theorem 1.7 we analyse the convergence of Lasserre hierarchies used in polynomial optimization. In Section 4 we focus on the optimum of the hierarchy, proving in Theorem 4.2 and Theorem 4.3 a new, general polynomial convergence as corollary of our main result (see Corollary 4.4 for regular Polynomial Optimization Problems). On the other hand in Section 5 we focus on the convergence of the feasible truncated pseudo-moment sequences of the moment hierarchy to truncated moment sequences of measures supported on S: we prove in Theorem 1.8 that we can bound their Hausdorff distance using Theorem 1.7.

1.2 Truncated pseudo-moment sequences and measures

Dualizing our point of view, we consider the convex cone $\mathcal{M}(S)$ of Borel measures supported on S, which is dual to $\mathrm{Pos}(S)$. We denote by $\mathcal{M}^{(1)}(S)$ the set of Borel *probability* measures supported on S.

The dual of polynomials is described as follows (see [Mou18] for more details). For $L \in (\mathbb{R}[\mathbf{X}])^* = \hom_{\mathbb{R}}(\mathbb{R}[\mathbf{X}], \mathbb{R})$, we denote $\langle L | f \rangle = L(f)$ the application of L to $f \in \mathbb{R}[\mathbf{X}]$, to emphasize the dulity pairing between $\mathbb{R}[\mathbf{X}]$ and $(\mathbb{R}[\mathbf{X}])^*$. Recall that $(\mathbb{R}[\mathbf{X}])^* \cong \mathbb{R}[[\mathbf{Y}]] := \mathbb{R}[[Y_1, \dots, Y_n]]$, with the isomorphism given by:

$$(\mathbb{R}[\mathbf{X}])^* \ni L \mapsto \sum_{\alpha \in \mathbb{N}^n} \langle L | \mathbf{X}^\alpha \rangle \frac{\mathbf{Y}^\alpha}{\alpha!} \in \mathbb{R}[[\mathbf{Y}]],$$

where $\{\frac{\mathbf{Y}^\alpha}{\alpha!}\}$ is the dual basis[2] to $\{\mathbf{X}^\alpha\}$, i.e. $\left\langle \mathbf{Y}^\alpha \middle| \mathbf{X}^\beta \right\rangle = \alpha!\, \delta_{\alpha,\beta}$. With this basis we can also identify $L \in (\mathbb{R}[\mathbf{X}])^*$ with its sequence of coefficients (*pseudo-moments* of L) $(L_\alpha)_\alpha$, where $L_\alpha = \langle L | \mathbf{X}^\alpha \rangle$.

Among all the linear functionals of special importance are the ones *coming from a measure*, i.e. $L \in \mathbb{R}[\mathbf{X}]^*$ such that there exists a Borel measure $\mu \in \mathcal{M}(S)$ with $\langle L | f \rangle = \int f \, \mathrm{d}\mu$ for all $f \in \mathbb{R}[\mathbf{X}]$. In this case the sequence $(\mu_\alpha)_\alpha$ associated with μ is the sequence of *moments*: $\mu_\alpha = \int X^\alpha \, \mathrm{d}\mu$. We are interested in the case when S is compact. In such a case the moment problem is *determinate*, i.e. the sequence of moments $(\mu_\alpha)_\alpha$ determines uniquely $\mu \in \mathcal{M}(S)$, see for instance [Sch17,

[2] To be more precise, *basis* means here a Schauder basis of $\mathbb{R}[[\mathbf{Y}]]$ equipped with the $(Y_1, \dots Y_n)$-adic topology.

ch. 14]. Therefore we will identify μ with its associated linear functional acting on polynomials (or equivalently with its sequence of moments), so that $\mathcal{M}(S) \subset (\mathbb{R}[\mathbf{X}])^*$.

We will work in the *truncated* setting, i.e. when we restict our linear functionals to a fixed, finite dimensional subspace of $\mathbb{R}[\mathbf{X}]$. In particular we denote $(\cdot)^{[t]}$ the restriction of a linear functional (or of a family of linear functionals) to $\mathbb{R}[\mathbf{X}]_t$, i.e. to polynomials of degree at most t. In coordinates, if $L = (L_\alpha)_{|\alpha|\leq d} \in \mathbb{R}[\mathbf{X}]_d^*$ then $L^{[t]} = (L_\alpha)_{|\alpha|\leq t} \in \mathbb{R}[\mathbf{X}]_t^*$, i.e. $L^{[t]}$ is the truncation of the pseudo-moment sequence to degree t.

We are interested in the dual algebraic objects to truncated quadratic modules: the *truncated positive linear functionals*

$$\mathcal{L}_\ell(\mathbf{g}) = \{ L \in (\mathbb{R}[\mathbf{X}]_\ell)^* \mid \forall q \in \mathcal{Q}_\ell(g) \ \langle L|q\rangle \geq 0 \} = \mathcal{Q}_\ell(\mathbf{g})^\vee,$$

i.e. $\mathcal{L}_\ell(\mathbf{g})$ is the dual convex cone to $\mathcal{Q}_\ell(\mathbf{g})$. See [Roc97] for more about convex cones and convex duality, and [GPR12] for their use in Optimization and Convex Algebraic Geometry. We define the affine section

$$\mathcal{L}_\ell^{(1)}(\mathbf{g}) = \Big\{ L \in \mathcal{L}_\ell(\mathbf{g}) \mid \langle L|1\rangle = 1 \Big\}.$$

Let $t = \lfloor \frac{\ell}{2} \rfloor$. We verify that for $L \in \mathcal{L}_\ell(\mathbf{g})$, $\langle L|1\rangle = 0$ implies $L^{[t]} = 0$, in order to prove that $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}$ is a generating section of $\mathcal{L}_\ell(\mathbf{g})^{[t]}$. Assume that $\langle L|1\rangle = 0$. For all $h \in \mathbb{R}[\mathbf{X}]_t$ and $x \in \mathbb{R}$, we have $0 \leq \Big\langle L \Big| (1+xh)^2 \Big\rangle = \langle L|1\rangle + 2x\langle L|h\rangle + x^2 \Big\langle L \Big| h^2 \Big\rangle$. If $\langle L|1\rangle = 0$, then the polynomial $x \mapsto 2x\langle L|h\rangle + x^2 \Big\langle L \Big| h^2 \Big\rangle$ is positive on $\mathbb{R}$ and has a zero at $x = 0$. Thus $x = 0$ is a double root and $\langle L|h\rangle = 0$. This implies that L restricted to polynomials of degree $\leq t$ is zero, i.e. $L^{[t]} = 0$. Therefore if $L^{[t]} \neq 0$ then $\langle L|1\rangle > 0$ and $L^{[t]} = \langle L|1\rangle \frac{L^{[t]}}{\langle L|1\rangle}$, with $\frac{L^{[t]}}{\langle L|1\rangle} \in \mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}$. This shows that $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}$ is a generating section of $\mathcal{L}_\ell(\mathbf{g})^{[t]}$.

Truncated positive linear functionals are an outer approximation of measures supported on S. They are used in Polynomial Optimization Problems (POP) to compute lower approximations of the minimum of a polynomial function f on S, see Section 4. Under the Archimedean assumption, convergence to measures of the linear functionals realizing the lower approximations have been proved in [Sch05, th. 3.4] for POP.

However nothing is said about the rate of convergence. We use Theorem 1.7, quantifying how good is the inner approximation of positive polynomials by truncated quadratic modules, to answer the question we are interested in: how good is the outer approximation of (probability) measures by truncated positive linear functionals (of total mass one)? Theorem 1.8 gives the answer. In the theorem we bound the Hausdorff distance $\mathrm{d_H}(\cdot,\cdot)$ between the outer approximation and the measures supported on S, where $\mathrm{d_H}(A,B) = \max\Big\{ \sup_{a\in A} \mathrm{d}(a,B), \ \sup_{b\in B} \mathrm{d}(A,b) \Big\}$.

Theorem 1.8. *Assume $n \geq 2$ and that the normalisation assumptions* (2) *are satisfied, and in particular that $1 - \|\mathbf{X}\|_2^2 = q \in \mathcal{Q}_{\ell_0}(\mathbf{g})$. Let $0 < \varepsilon \leq \frac{1}{2}$, $t \in \mathbb{N}_+$ and $\ell \in \mathbb{N}$ such that $\ell \geq \gamma(n,\mathbf{g})\, 6^{2.5n\text{Ł}}\, t^{6n\text{Ł}} \binom{n+t}{t}^{2.5n\text{Ł}} \varepsilon^{-2.5n\text{Ł}}$ and $\ell \geq 2t + \ell_0$, with $\gamma(n,\mathbf{g})$ given by Theorem 1.7. Then*

$$\mathrm{d_H}(\mathcal{M}^{(1)}(S)^{[2t]}, \mathcal{L}_\ell^{(1)}(\mathbf{g})^{[2t]}) \leq \varepsilon.$$

The proof of Theorem 1.8 is developed in Section 5. The corollary of Theorem 2.11 in regular cases with $\text{Ł} = 1$ is Corollary 5.9.

1.3 Related works

Complexity analysis in Real Algebraic Geometry is an active area of research, where obtaining good upper bounds is challenging. See for instance [LPR20] for elementary recursive degree bounds in Kivrine-Stengle Positivstellensatz, and [SEDYZ18] for computation complexity of real radicals. Among all the Stellensätzen, we consider Putinar's Positivstellensatz, which allows a denominator free representation of strictly positive polynomials and has well-know applications in Polynomial Optimization. The representation of strictly positive polynomials has a long history. For instance Pólya [Pó28] gave a representation of homogeneous polynomials f strictly positive on the simplex

Δ as ratio of a polynomial with positive coefficients and $\|\mathbf{X}\|_2^k$, for some k. It is interesting to notice that, although no explicit degree bounds were presented, the degree of the representation depends on the sup norm of f on Δ and on its minimum $f^* > 0$, i.e. on $\varepsilon(f)$, in analogy with Theorem 1.7. Another representation for homogeneous polynomials has been proved by Reznick [Rez95], where it is shown that an homogeneous polynomial f strictly positive on $\mathbb{R}^n \setminus \{0\}$ (a *positive definite form*) can be written as ratio of even powers of linear forms and $\|\mathbf{X}\|_2^k$, for some k. Degree bounds for the representation are provided, and again we find a dependence on $\varepsilon(f)$ (defined restricting f to the $n-1$ hypersphere) with exponent equal to -1.

A general, effective version of Putinar's Positivstellensatz have been proved in [NS07] (see also [Sch04] for a general effective Schmüdgen's Positivstellensatz). This result is used in [MSED21] to give bounds on the degree of rational SoS positivity certificates, which are exponential in the bit-size of the input polynomials $f, \mathbf{g}$. Compared to [NS07], Theorem 1.7 gives degree bounds, which are polynomial and not exponential in $\varepsilon(f)$. This implies a polynomial rate convergence of Lasserre hierarchies, see Theorem 4.2, and not logarithmic as in [NS07]. For special semialgebraic sets, the bounds on the convergence rate can be improved: see for instance [LS21] for convergence on the unit box and [FF20] for the unit sphere. The convergence rate of the upper bounds of Lasserre SoS density hierarchy over the sphere has been studied in [DKL19].

The proof of Theorem 1.7 is based on the construction of a perturbation polynomial $q \in \mathcal{Q}(\mathbf{g})$ and the reduction to a simpler semialgebraic set. This construction of the perturbation polynomial q using univariate SoS, has already been used in [Sch04], [Sch05], [NS07], [Ave13], [KS15]. In [MM22] Mai and Magron investigate with a similar technique the representation of strictly positive polynomials on arbitrary semialgebraic sets as ratio of polynomials in the quadratic module and $(1+\|\mathbf{X}\|_2^2)^k$ for some k, giving degree bounds for the representation. These bounds are polynomial on f^* (and thus on $\varepsilon(f)$), but the exponent and the constant are not explicit in the general case. Moreover they remark that they were not able to derive a polynomial Effective Putinar's Positvstellensatz using their perturbation polynomials, defined recursively.

Our main improvements in the proof are the generalisation from univariate SoS or recursively defined perturbation polynomials to a positive polynomial echelon function, see Section 2.3, and the use of an Effective Schmüdgen's Positivstellensatz on the unit box from [LS21]. Moreover in Section 2.2 we analyse regular cases that result in very simple exponents, see Corollary 3.9. Corollary 3.9 can be applied in particular in the case of a single ball constraint, that was analysed in [MM22] for the Putinar-Vasilescu's Positivstellensatz, that introduces a denominator: the exponent in this case is equal to -65, while Corollary 3.9 gives $-2.5n$. We conjecture that it is possible to remove the dependence on n in the exponent of the Effective Putinar's Positivstellensatz.

This approach with a perturbation polynomial q has also been used in [KS15] to prove a Weierstrass Approximation theorem on compact sets for positive polynomials, where the approximation is done with polynomials in the quadratic module $\mathcal{Q}(\mathbf{g})$. We obtain an equivalent result with our polynomial echelon functions in Theorem 4.1 with bounds on the degree of q.

Convergence of pseudo-moments sequences to measures in Lasserre's hierarchies has been studied in [Sch05] for Polynomial Optimization Problems and more generally in [Tac21] for Generalized Moment Problems (GMP). The convergence rates of moment hierarchies in GMP over the simplex and the sphere have been studied in [KK21]. To our best knowledge there is no analysis of the convergence rate for general compact basic semialgebraic sets in the literature. In Theorem 1.8 we prove such a rate of convergence for the pseudo-moment sequences used in Polynomial Optimization, deducing this speed from Theorem 1.7.

2 The ingredients of the proof

To prove the polynomial bound for the Effective Putinar's Positivstellensatz (Theorem 1.7), we proceeds as follows, refining the approach in [Sch05], [NS07], [Ave13]:

- We perturb f into a polynomial p such that p is strictly positive on the box $[-1,1]^n$ and $f-p$ is in the quadratic module $\mathcal{Q}(\mathbf{g})$;

- We compute an SOS representation of p in $\mathcal{Q}(1-\|\mathbf{X}\|_2^2)$ to deduce the representation of $f \in \mathcal{Q}(\mathbf{g})$.

Notice that if $f > 0$ on $[-1,1]^n$ then we can directly apply Theorem 3.8 and Lemma 3.4 to conclude the proof. Therefore in the following we always assume that there exists $x \in [-1,1]^n \setminus S$ such that $f(x) \leq 0$.

To compute the perturbed polynomial p, we use a univariate echelon-like polynomial, which shape and degree depends on the distance between a level-set of f and S and on a lower bound of the algebraic distance to S. We detail these ingredients hereafter.

2.1 Distance between level sets of f and S

We define the complementary in $[-1,1]^n$ of a neighbourhood of S, where f is strictly smaller than f^*:

$$A = \{x \in [-1,1]^n \mid f(x) \leq \frac{3f^*}{4}\}$$

that is a sublevel set of the function f.

We are going to bound the distance from A to S in terms of $\varepsilon(f)$. We recall first a Markov inequality theorem, bounding the norm of the gradient of a polynomial function on a convex body, in the special case of the box $[-1,1]^n$.

Theorem 2.1 ([KR99, th. 3])**.** *Let $p \in \mathbb{R}[\mathbf{X}]_d$ be a polynomial of degree $\leq d$. Then:*

$$\left\| \|\nabla p\|_2 \right\| \leq (2d^2 - d)\|p\|.$$

Recall that the Lipschitz constant L_f of f is the smallest real number such that $\left|f(x) - f(y)\right| \leq L_f \left\|x-y\right\|_2$ for all x, y in the domain of f. Using Theorem 2.1 to bound the Lipschitz constant of f on $[-1,1]^n$, we can lower bound the distance between A and S.

Proposition 2.2. *Let A and S be as above. Then* $\mathrm{d_H}(A,S) \geq \frac{\varepsilon(f)}{8d^2}$.

Proof. We first relate the Lipschitz constant L_f of f on $[-1,1]^n$ with $\|f\|$.

From the mean value theorem we deduce that for all $x, y \in [-1,1]^n$ we have $\left|f(x) - f(y)\right| \leq \left\| \|\nabla f\|_2 \right\| \left\|x-y\right\|_2$. Then from the definition of Lipschitz constant and Theorem 2.1:

$$L_f \leq \left\| \|\nabla f\|_2 \right\| \leq (2d^2 - d)\|f\| \Rightarrow \frac{1}{L_f} \geq \frac{1}{(2d^2-d)\|f\|}. \tag{3}$$

Now let $x \in A$ and $y \in S$. By definition of L_f we have $\left|f(x) - f(y)\right| \leq L_f \left\|x-y\right\|_2$. Since $x \in A$ we have $f(x) \leq \frac{3f^*}{4}$; since $y \in S$ we have $f(y) \geq f^*$: thus $\left|f(x) - f(y)\right| \geq \frac{f^*}{4}$ and $\left\|x-y\right\|_2 \geq \frac{f^*}{4L_f}$. As the inequality hold for all $x \in A$ and $y \in S$ we can use Equation (3) to conclude:

$$\mathrm{d_H}(A,S) \geq \frac{f^*}{4L_f} \geq \frac{f^*}{4(2d^2-d)\|f\|} = \frac{\varepsilon(f)}{4(2d^2-d)} \geq \frac{\varepsilon(f)}{8d^2}.$$

□

2.2 Bounds on the algebraic distance to S

The algebraic distance to the set S is the continuous semialgebraic function defined by

$$G(x) = |\min\{g_1(x), \ldots, g_r(x), 0\}|.$$

Clearly, $G(x) = 0$ if and only if $x \in S$, and $G(x) > 0$ if $x \notin S$. We are going to bound from below the function G on A, that is find $\delta \in \mathbb{R}_{>0}$ such that

$$\forall x \in A,\ G(x) \geq \delta \tag{4}$$

(such a δ exists since A is compact and $G(x) > 0$ on A).

To express such a δ in terms of $\varepsilon(f)$, we use Łojasiewicz inequalities, introduced by Łojasiewicz in [Łoj59], following and expanding the approach in [NS07, lem. 13].

Theorem 2.3 ([Łoj59], [BCR98, cor. 2.6.7])**.** *Let B be a closed and bounded semialgebraic set and let f, g be two continuous semialgebraic functions from B to R such that $f^{-1}(0) \subset g^{-1}(0)$. Then there exists $c, Ł \in \mathbb{R}_{>0}$ such that $\forall x \in B$:*

$$|g(x)|^{Ł} \leq c|f(x)|.$$

We use now Theorem 2.3 and Proposition 2.2 to bound δ in terms of $\varepsilon(f)$.

Definition 2.4. Let $\mathfrak{c}, Ł$ be the constant and exponent of Łojasiewicz inequalities (Theorem 2.3) for the functions $G : x \in [-1,1]^n \mapsto G(x) = |\min\{g_1, \ldots, g_r(x), 0\}|$ and $D : x \in [-1,1]^n \mapsto D(x) = \mathrm{d}(x, S)$, that is, for $x \in [-1,1]^n$

$$D(x)^{Ł} \leq \mathfrak{c}\, G(x). \tag{5}$$

These constant and exponent are well-defined by Theorem 2.3, since the functions D and G are continuous semialgebraic and $S = D^{-1}(0) = G^{-1}(0)$.

Lemma 2.5. *We can choose $\delta = \frac{1}{\mathfrak{c}}\left(\frac{\varepsilon(f)}{8d^2}\right)^{Ł}$ in Equation* (4)*, where $\mathfrak{c}, Ł$ are defined in Definition 2.4.*

Proof. By Proposition 2.2, we have $D(x) = \mathrm{d}(x, S) \geq \frac{\varepsilon(f)}{8d^2}$ for $x \in A$. Then from Equation (5), we deduce that for $x \in A$,

$$\left(\frac{\varepsilon(f)}{8d^2}\right)^{Ł} \leq \mathfrak{c}G(x) \Rightarrow G(x) \geq \frac{1}{\mathfrak{c}}\left(\frac{\varepsilon(f)}{8d^2}\right)^{Ł}.$$

Therefore, we can choose $\delta = \frac{1}{\mathfrak{c}}\left(\frac{\varepsilon(f)}{8d^2}\right)^{Ł}$. □

The exponent $Ł$ in Definition 2.4 will play an important role in the bounds of the Effective Putinar's Positivstellensatz. We show now that, under generic assumptions, we can choose $Ł = 1$, as suggested in the following example.

Example 2.6. Consider the univariate polynomial $g(X) = \varepsilon^2 - X^2$ and let $S = S(g) = [-\varepsilon, \varepsilon] \subset [-1,1]$. Now let $x \in [-1,1]$ and D, G be as in Definition 2.4. It is easy to show that:

$$D(x) \leq \frac{1}{2\varepsilon} G(x).$$

Indeed, if for example $\varepsilon \leq x \leq 1$, we have $D(x) = x - \varepsilon$ and $G(x) = x^2 - \varepsilon^2 = (x+\varepsilon)(x-\varepsilon)$ and $D(x) = \frac{1}{x+\varepsilon} G(x) \leq \frac{1}{2\varepsilon} G(x)$. This shows that we can choose $Ł = 1$ for all $\varepsilon > 0$.

On the other hand if $\varepsilon = 0$, i.e. $g(X) = -X^2$ and $S = \{0\}$, we have a singular equation. A simple computation shows that it is not possible to choose $Ł = 1$ in this case. The minimum $Ł$ satisfying the inequality is $Ł = 2$.

We introduce a regularity condition needed to prove $Ł = 1$, generalizing Example 2.6. This is a standard condition in optimization (see [Ber99, sec. 3.3.1]), which implies the so-called Karush–Kuhn–Tucker (KKT) conditions [Ber99, prop. 3.3.1].

Definition 2.7. Let $x \in \mathcal{S}(\mathbf{g})$. The active constraints at x are the constraints $g_{i_1}, \ldots, g_{i_m}$ such that $g_{i_j}(x) = 0$. We say that the *Constraint Qualification condition* (CQC) holds at x if for the active constraints $g_{i_1}, \ldots, g_{i_l}$ at x, the gradients $\nabla g_{i_1}(x), \ldots, \nabla g_{i_m}(x)$ are linearly independent.

Lemma 2.8. *Let $y \in \mathbb{R}^n \setminus \mathcal{S}(\mathbf{g})$, and let z be a point in $S = \mathcal{S}(\mathbf{g})$ minimizing the distance of y to S, that is $\mathrm{d}(y, S) = \left\|y - z\right\|_2$. If $\{g_i : i \in I\}$ are the active constraints at z and the CQC holds, then there exist $\lambda_i \in \mathbb{R}_{\geq 0}$ such that:*

$$y - z = \sum_{i \in I} \lambda_i \nabla(-g_i)(z).$$

Proof. Fix $y \in \mathbb{R}^n$. Notice that $y - x = -\frac{\nabla\|y-x\|_2^2}{2}$, where the gradient is take w.r.t. x. Moreover $z \in S$ such that $\mathrm{d}(y, S) = \left\|y - z\right\|_2$ is a minimizer of the following Polynomial Optimization Problem:

$$\min_x \frac{\left\|y - x\right\|_2^2}{2} \colon g_i(x) \geq 0 \ \forall i \in \{1, \ldots, r\}.$$

Since the CQC holds at z, we deduce from [Ber99, prop. 3.3.1] that the KKT conditions hold. In particular:

$$\frac{\nabla\left\|y - z\right\|_2^2}{2} = \sum_{i \in I} \lambda_i \nabla g_i(z)$$

For some $\lambda_i \in \mathbb{R}_{\geq 0}$. Therefore $y - z = -\frac{\nabla \mathrm{d}(y,z)^2}{2} = \sum_{i \in I} \lambda_i \nabla(-g_i)(z)$. □

We first fix a point $z \in S$ and consider the points y such that the closest point to y on S is z. We prove that $Ł = 1$ in the sector of these y where all the active constraints at z are strictly negative at y.

Lemma 2.9. *Let D, G as in Definition 2.4 and let $z \in \mathcal{S}(\mathbf{g})$ with active constraints $g_i \colon i \in I$. Then there exists $\varepsilon' = \varepsilon'(z) > 0$ and constant $\mathfrak{c}' = \mathfrak{c}'(z) > 0$ such that for all y with:*

- $D(y) = \mathrm{d}(y, S) = \left\|y - z\right\|_2$;
- $D(y) \leq \varepsilon'$;
- $g_i(y) < 0$ *for all* $i \in I$,

we have $D(y) \leq \mathfrak{c}' G(y)$.

Proof. Let $z \in S$ and $y \in \mathbb{R}^n$ be such that $D(y) = \left\|y - z\right\|_2$. Consider the Taylor expansion of g_i at z for the active constraints $\{g_i \colon i \in I\}$: there exists $\mathbf{h}_i(y) = (h_{i,1}(y), \ldots, h_{i,n}(y))$, where $\lim_{y \to z} h_{i,j}(y) = 0$, such that:

$$g_i(y) = \nabla g_i(z) \cdot (y - z) + \sum_{j=1}^{n} h_{i,j}(y)(y_i - z_i) = \nabla g_i(z) \cdot (y - z) + \mathbf{h}_i(y) \cdot (y - z). \tag{6}$$

In other words, the $\mathbf{h}_i(y) \cdot (y - z)$ is the remainder of the first order Taylor approximation. Since the CQC is satisfied at z we can apply Lemma 2.8: there exists $\lambda = (\lambda_i \colon i \in I)$ with $y - z = \sum_{i \in I} \lambda_i \nabla(-g_i)(z)$. Substituting we obtain $\forall i \in I$:

$$g_i(y) = -\sum_{j \in I} \lambda_j (\nabla g_i(z) \cdot \nabla g_j(z)) + \mathbf{h}_i(y) \cdot (y - z).$$

We denote:

- $\mathbf{g}_I(y) = (g_i(y) \colon i \in I)$;
- $\mathbf{J}_I(z) = \mathrm{Jac}(\mathbf{g}_I)(z) = (\nabla g_i(z))_{i \in I}$ the Jacobian matrix;
- $\mathbf{N}_I(z) = (\nabla g_i(z) \cdot \nabla g_j(z))_{i,j} = \mathbf{J}_I(z)^t \mathbf{J}_I(z)$ the Gram matrix of $\nabla g_i(z)$; and
- $\mathbf{h}(y) = (h_{i,j}(y))_{i,j}$.

With this notation we get:

$$\mathbf{g}_I(y) = -\mathbf{N}_I(z)\lambda + \mathbf{h}(y)(y - z).$$

Since CQC hold at z, the $\nabla g_i(z)$ are linearly independent and thus $\mathbf{N}_I(z)$ is invertible. Indeed, if $\mathbf{N}_I(z)$ is not invertible then there exists $0 \neq v \in \mathbb{R}^{|I|}$ such that $\mathbf{N}_I(z)v = 0$. Therefore $0 = v^t \mathbf{N}_I(z) v =$

$(\mathbf{J}_I(z)v)^t\mathbf{J}_I(z)v = \|\mathbf{J}_I(z)v\|_2^2$. Hence $\mathbf{J}_I(z)v = 0$, contradicting the linear independence of $\nabla g_i(z)\colon i \in I$. Thus we can solve for λ:

$$\lambda = -\mathbf{N}_I(z)^{-1}\mathbf{g}_I(y) + \mathbf{N}_I(z)^{-1}\mathbf{h}(y)(y-z). \tag{7}$$

Recall from Lemma 2.8 that we have:

$$y - z = \sum_{i\in I} \lambda_i \nabla(-g_i)(z) = -\mathbf{J}_I(z)\lambda.$$

Substituting λ from Equation (7) we obtain:

$$y - z = \mathbf{J}_I(z)\mathbf{N}_I(z)^{-1}\mathbf{g}_I(y) - \mathbf{J}_I(z)\mathbf{N}_I(z)^{-1}\mathbf{h}(y)(y-z).$$

Taking the norm we deduce that:

$$\begin{aligned}\left\|y-z\right\|_2 &= \left\|\mathbf{J}_I(z)\mathbf{N}_I(z)^{-1}\mathbf{g}_I(y) - \mathbf{J}_I(z)\mathbf{N}_I(z)^{-1}\mathbf{h}(y)(y-z)\right\|_2 \\ &\le \|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}\left\|\mathbf{g}_I(y)\right\|_2 + \|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}\left\|\mathbf{h}(y)\right\|_2\left\|y-z\right\|_2\end{aligned}$$

where $\|\cdot\|_2$ denotes the 2-norm (resp. operator norm) of vectors (resp. matrices). Therefore:

$$\left(1 - \|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}\left\|\mathbf{h}(y)\right\|_2\right)\left\|y-z\right\|_2 \le \|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}\left\|\mathbf{g}_I(y)\right\|_2 \tag{8}$$

Notice that:

- $\left\|\mathbf{g}_I(y)\right\|_2 = \sqrt{\sum_{i\in I} g_i(y)^2} \le \sqrt{|I|}\max_{i\in I}\left|g_i(y)\right| \le \sqrt{r}\,G(y)$, since

  $$G(y) = \left|\min\{0, g_i(y)\colon i \in \{1,\dots,r\}\}\right| = \max\{0, \left|g_i(y)\right|\colon g_i(y) < 0\}$$

  and $g_i(y) < 0$ for all $i \in I$ by hypothesis;
- $1 - \|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}\left\|\mathbf{h}(y)\right\|_2 \ge \frac{1}{2}$ if y is close enough to S. Indeed $h_{i,j} \to 0$ when $y \to z$, i.e. when $\left\|y-z\right\|_2 = \mathrm{d}(y,S) = D(y)$ is going to zero. Thus we can choose ε' such that $D(y) \le \varepsilon'$ implies $1 - \|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}\left\|\mathbf{h}(y)\right\|_2 \ge \frac{1}{2}$.

Then we deduce from Equation (8):

$$D(y) = \left\|y-z\right\|_2 \le 2\sqrt{r}\|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}G(y)$$

when $D(y) \le \varepsilon'$, that proves the lemma with $\mathfrak{c}' = 2\sqrt{r}\|\mathbf{J}_I(z)\|_2\|\mathbf{N}_I(z)\|_2^{-1}$. □

We generalize the previous lemma, removing the assumption that all the active constraints are negative at y.

Lemma 2.10. *Let D, G as in Definition 2.4 and assume that the CQC hold at $z \in S = \mathcal{S}(\mathbf{g})$. Then there exists $\varepsilon'' = \varepsilon''(z) > 0$ and constant $\mathfrak{c}'' = \mathfrak{c}''(z) > 0$ such that for all y with:*

- $D(y) = \mathrm{d}(y,S) = \left\|y-z\right\|_2$;
- $D(y) \le \varepsilon''$;

we have $D(y) \le \mathfrak{c}''G(y)$.

Proof. Let y and z be as in the hypothesis and let $g_i\colon i \in I$ be the active constraints at z. Notice that if $y = z \in S$ then $D(y) = G(y) = 0$ and there is nothing to prove. So we assume that $y \notin S$: there exists $i \in \{1,\dots,r\}$ s.t. $g_i(y) < 0$. Moreover, from Lemma 2.9 we only need to consider the case where there exists $i \in I$ such that $g_i(y) \ge 0$.

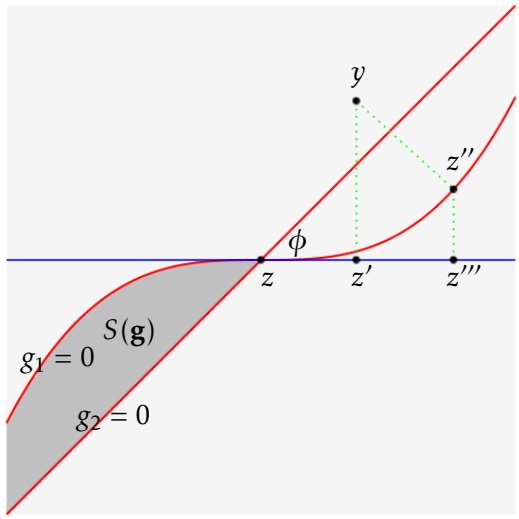

Figure 1: Proof of Lemma 2.10

So let $I_+ = I_+(y) = \{i \in I \colon g_i(y) \geq 0\}$ and $I_- = I_-(y) = \{i \in I \colon g_i(y) < 0\}$. Notice that I_- and I_+ depend on y, but to obtain a result independent from I_- and I_+ it is enough to take the minimum ε'' and the maximum $\mathfrak{c}''$ as I_- and I_+ vary.

If we consider the Taylor expansion of g_i at z, we obtain:

$$i \in I_- \Rightarrow 0 > g_i(y) = \nabla g_i(z) \cdot (y - z) + \mathbf{h}_i(y) \cdot (y - z),$$

with the same notation as in Equation (6). This implies that there exists $\delta > 0$ such that $\nabla(-g_i)(z) \cdot (y - z) \geq \delta$, for all $i \in I_-$, when y is close enough to z.

We want to reduce to the case of only negative inequalities. We define:

- $G_-(y) = \left|\min\{0, g_i(y) \colon i \in I_-\}\right|$;
- $S_- = \mathcal{S}(g_i \colon i \in I_-)$;
- $D_-(y) = \mathrm{d}(y, S_-)$;
- $T_z S_-$ the (affine) tangent space of S_- at z.

Notice that, since the gradients are linearly independent, $T_z S_-$ is the affine subspace passing through z and orthogonal to $\nabla(-g_i)(z)$ for $i \in I_-$. In particular, since $\nabla(-g_i)(z) \cdot (y - z) \geq \delta$, $y - z \notin T_z S_-$ the angle between $y - z$ and $T_z S_-$ is lower bounded by a strictly positive angle $\phi > 0$ for all y close enough to z.

For a geometric intuition of the following discussion, see Figure 1. Let z' be the projection of y on $T_z S_-$. By definition of ϕ we have $\left\|y - z\right\|_2 \leq \frac{\|y - z'\|_2}{\sin \phi}$. Now let z'' be the projection of y on S_-. Since y is close to z, z'' is close to z', i.e. the projection of y on S_- is close to the projection of y on $T_z S_-$. Thus there exists a constant c such that $\left\|y - z\right\|_2 \leq c \frac{\|y - z''\|_2}{\sin \phi}$. More precisely, let z''' be the projection of z'' on $T_z S_-$. Thus $z'' - z = (z'' - z''') + (z''' - z)$, and since we project z'' on $T_z S_-$ we have:

- $z'' - z''' = \sum_{i \in I_-} \gamma_i \nabla g_i(z) = \mathbf{J}_{I_-}(z)\gamma$ for some $\gamma = (\gamma_i \colon i \in I_-)$;
- $z''' - z$ is orthogonal to $\nabla g_i(z)$ for $i \in I_-$.

By definition of z' we have $\left\|y - z'\right\|_2 \leq \left\|y - z'''\right\|_2 \leq \left\|y - z''\right\|_2 + \|z'' - z'''\|_2$. We show now that $\|z'' - z'''\|_2$ is small compared to $\left\|y - z\right\|_2$. Expanding at z for $i \in I_-$ we obtain:

$$0 = g_i(z'') = \nabla g_i(z) \cdot (z'' - z) + \mathbf{h}_i(z'') \cdot (z'' - z) = \nabla g_i(z) \cdot (z'' - z''') + \mathbf{h}_i(z'') \cdot (z'' - z).$$

Proceeding as in Equation (7), we have $\gamma = \mathbf{N}_{I_-}(z)^{-1}\mathbf{h}(z'')(z''-z)$. Now, since z'' is the projection of y on S_- and $z \in S_-$ we have $\|z''-z\|_2 \leq 2\left\|y-z\right\|_2$. Thus we deduce:

$$\begin{aligned}\left\|y-z'\right\|_2 &\leq \left\|y-z''\right\|_2 + \left\|z''-z'''\right\|_2 \\ &\leq \left\|y-z''\right\|_2 + \left\|\mathbf{J}_{I_-}(z)\right\|_2\left\|\gamma\right\|_2 \\ &\leq \left\|y-z''\right\|_2 + \left\|\mathbf{J}_{I_-}(z)\right\|_2\left\|\mathbf{N}_{I_-}(z)^{-1}\mathbf{h}(z'')(z''-z)\right\|_2 \\ &\leq \left\|y-z''\right\|_2 + 2\left\|\mathbf{J}_{I_-}(z)\right\|_2\left\|\mathbf{N}_{I_-}(z)^{-1}\right\|_2\left\|\mathbf{h}(z'')\right\|_2\left\|y-z\right\|_2.\end{aligned}$$

Therefore

$$\left\|y-z\right\|_2 \leq \frac{\left\|y-z'\right\|_2}{\sin\phi} \leq \frac{\left\|y-z''\right\|_2}{\sin\phi} + \frac{2\left\|\mathbf{J}_{I_-}(z)\right\|_2\left\|\mathbf{N}_{I_-}(z)^{-1}\right\|_2\|\mathbf{h}(z'')\|_2\left\|y-z\right\|_2}{\sin\phi},$$

and finally

$$\Big(1 - \frac{2\left\|\mathbf{J}_{I_-}(z)\right\|_2\left\|\mathbf{N}_{I_-}(z)^{-1}\right\|_2\|\mathbf{h}(z'')\|_2}{\sin\phi}\Big)\left\|y-z\right\|_2 \leq \frac{\left\|y-z''\right\|_2}{\sin\phi}.$$

As $z'' \to z$ if $y \to z$, $\|\mathbf{h}(z'')\|_2 \to 0$ for $y \to z$. Then there exists $\varepsilon'' > 0$ such that $D(y) \leq \varepsilon''$ implies $1 - \frac{2\|\mathbf{J}_{I_-}(z)\|_2\|\mathbf{N}_{I_-}(z)^{-1}\|_2\|\mathbf{h}(z'')\|_2}{\sin\phi} \geq \frac{1}{2}$ and thus

$$\left\|y-z\right\|_2 \leq 2\frac{\left\|y-z''\right\|_2}{\sin\phi}. \tag{9}$$

In other words, we just proved in Equation (9) that

$$D(y) \leq \varepsilon'' \Rightarrow D(y) \leq \frac{2}{\sin\phi}D_-(y).$$

Since D_- is the distance function to S_-, that is defined by inequalities negative at y, we can apply Lemma 2.9: there exists $\mathfrak{c}'$ such that if ε'' is small enough, $D_-(y) \leq \varepsilon''$ implies $D_-(y) \leq \mathfrak{c}'G_-(y)$ (notice that this is possible because $D(y) \to 0$ implies $D_-(y) \to 0$). Moreover observe that $G(y) = G_-(y)$ since only the g_i that are negative at y contribute to $G(y)$. Then, if we set $\mathfrak{c}'' = \frac{2\mathfrak{c}'}{\sin\phi}$ we can conclude:

$$D(y) \leq \varepsilon'' \Rightarrow D(y) \leq \mathfrak{c}''G(y).$$

□

We can now show that if the CQC hold at every point of the semialgebraic set the Łojasiewicz exponent is equal to 1.

Theorem 2.11. *Let D, G as in Definition 2.4 and assume that the CQC holds at every point of $S = \mathcal{S}(\mathbf{g}) \subset [-1,1]^n$. Then there exists a constant $\mathfrak{c} \in \mathbb{R}_{>0}$ such that:*

$$D(y) \leq \mathfrak{c}G(y)$$

for all $y \in [-1,1]^n$.

Proof. Let $\varepsilon = \min_{z\in S}\varepsilon''(z)$ and $\mathfrak{c}' = \max_{z\in S}\mathfrak{c}''(z)$, with $\varepsilon''(z)$ and $\mathfrak{c}''(z)$ as in Lemma 2.10. Notice that $\varepsilon > 0$ and $\mathfrak{c}' < \infty$ exist because S is compact and $0 < \varepsilon''(z)$, $\mathfrak{c}''(z) < \infty$ are respectively lower and upper semicontinuous functions of z. Let $U = \{y \in [-1,1]^n \mid \mathrm{d}(y,S) < \varepsilon\} \subset [-1,1]^n$ (an open set containing S): by definition of ε and $\mathfrak{c}'$ we have $D(y) \leq \mathfrak{c}'G(y)$ for all $y \in U$ from Lemma 2.10.

Now consider the compact set $C = [-1,1]^n \setminus U$ and let $G^* > 0$ be the minimum of G on C. Moreover since $S \subset [-1,1]^n$ we have $D(y) \leq 2\sqrt{n}$ for $y \in [-1,1]^n$. Then:

$$D(y) \leq 2\sqrt{n} = \frac{2\sqrt{n}}{G^*}G^* \leq \frac{2\sqrt{n}}{G^*}G(y)$$

for all $y \in C$.

Finally, taking $\mathfrak{c} = \max(\mathfrak{c}', \frac{2\sqrt{n}}{G^*})$ we obtain:

$$D(y) \leq \mathfrak{c} G(y)$$

for all $y \in [-1,1]^n$. □

Remark. In Theorem 2.11 we prove that in regular cases the Łojasiewicz exponent is 1. On the other hand we don't give a precise estimate for the constant $\mathfrak{c}$, even if we can revisit the proof of Lemma 2.9, Lemma 2.10 and Theorem 2.11 to bound it in terms of the following parameters:

- the max norm of the Jacobian of the $\mathbf{g}$: we could bound this parameter bounding the norm of $\mathbf{g}$;
- the min norm of the Gram matrix of the $\nabla\mathbf{g}$: this measures how close are the gradients to be linearly dependend;
- the minimum of $G(y)$ on the complementary in $[-1,1]^n$ of a small neighbourhood of S: this measures how close are the $\mathbf{g}$ to have a common zero outside of S;
- the convergence rate to 0 of the Taylor remainder $\mathbf{h}(z)$.

A detailed analysis of these parameters would also give an upper bound for $\mathfrak{c}$, but we don't develop it for the sake of simplicity.

Remark. On the contrary when the problem is not regular the bounds on the exponent $Ł$ can be large. We have:

$$Ł \leq d(\mathbf{g})(6d(\mathbf{g})-3)^{n+r-1}$$

see [KS15, sec. 3.1] and [KSS16].

2.3 Construction of a polynomial echelon function

In this section, we describe the polynomial echelon function $h_{k,m}$ used to perturb f. This echelon polynomial depends on a parameter $\delta \in \mathbb{R}_{>0}$ controlling the width of the step (and defined in Section 2.2) and on a parameter $k \in \mathbb{R}_{>0}$ controlling the minimum of the function. To show that the degree and the norm of the perturbation polynomial depend polynomially on $\varepsilon(f)$ (in Section 3.1), we are going to bound the degree of the echelon polynomials in terms of δ and k.

Consider the following function:

$$H(t) = \begin{cases} 1 & t \in [-1,-\delta] \\ -\frac{9(k-1)}{2d^3k}t^3 - \frac{27(k-1)}{2d^2k}t^2 - \frac{27(k-1)}{2dk}t - \frac{7k-9}{2k} & t \in [-\delta, -\delta+\frac{\delta}{3}] \\ \frac{9(k-1)}{d^3k}t^3 + \frac{27(k-1)}{2d^2k}t^2 + \frac{9(k-1)}{2dk}t + \frac{k+1}{2k} & t \in [-\delta+\frac{\delta}{3}, -\delta+\frac{2\delta}{3}] \\ -\frac{9(k-1)}{2d^3k}t^3 + \frac{1}{k} & t \in [-\delta+\frac{2\delta}{3}, 0] \\ \frac{1}{k} & t \in [0,1] \end{cases} \tag{10}$$

The piecewise polynomial function $H(t)$ is a C^2 cubic spline on $[-1,1]$. Indeed an explicit computation shows that the functions $H, H^{(1)}, H^{(2)}$ are absolutely continuous, and moreover the piecewise constant function $H^{(3)}$ is of total variation $V = \frac{216(k-1)}{\delta^3 k}$. Finally notice that H is non-increasing on $[-1,1]$.

We approximate this function by a polynomial $\in \mathbb{R}[T]$, using Chebyshev approximation (see Figure 2):

Theorem 2.12 (Chebyshev approximation on $[-1,1]$ [Tre13]). *For an integer u, let $h : [-1,1] \to \mathbb{R}$ be a function such that its derivatives through $h^{(u-1)}$ be absolutely continuous on $[-1,1]$ and its u-th derivative $h^{(u)}$ is of bounded variation V. Then its Chebyshev approximation p_m of degree m satisfies:*

$$\|h - p_m\| \leq \frac{4V}{\pi u (m-u)^u}.$$

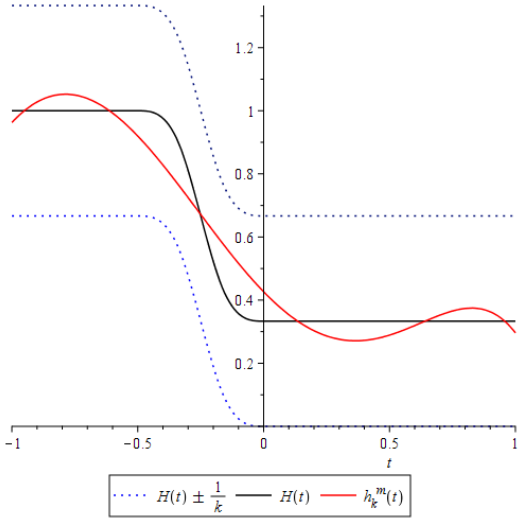

Figure 2: Polynomial Echelon Function

Proposition 2.13. *There exists a univariate polynomial $h_{k,m} \in \mathbb{R}[T]$ such that:*

- $\deg h_{k,m} = m$ *with* $m = \left\lceil \frac{6}{\delta}\sqrt[3]{\frac{4(k-1)}{3\pi}} + 3 \right\rceil$;
- *for* $t \in [-1, -\delta]$ *we have* $1 - \frac{1}{k} \leq h_{k,m}(t) \leq 1 + \frac{1}{k}$;
- *for* $t \in [0, 1]$ *we have* $h_{k,m}(t) \leq \frac{2}{k}$;
- *for* $t \in [-1, 1]$ *we have* $0 \leq h_{k,m}(t) \leq 1 + \frac{1}{k}$.

Proof. We construct a degree m Chebyshev approximation $h_{k,m} \in \mathbb{R}[T]$ of H such that

$$\left\| H - h_{k,m} \right\| \leq \frac{1}{k}, \tag{11}$$

so that the last three points of the proposition are satisfied. As H, $H^{(1)}$ and $H^{(2)}$ are absolutely continuous and $H^{(3)}$ has total variation $V = \frac{216(k-1)}{\delta^3 k}$, by Theorem 2.12, it suffices to take m such $\frac{4V}{3\pi(m-3)^3} \leq \frac{1}{k}$, i.e.

$$m \geq \sqrt[3]{\frac{4Vk}{3\pi}} + 3 = \frac{6}{\delta}\sqrt[3]{\frac{4(k-1)}{3\pi}} + 3,$$

which proves the first point.

The other points follow from Equation (11) and the definition of H in (10). □

3 Effective Putinar's Positivstellensatz

This section is devoted to the proof of Theorem 1.7.

3.1 From S to $[-1, 1]^n$

Let $h_{k,m}$ be as in Proposition 2.13. We want to show that, for a suitable choice of k, m and $s \in \mathbb{R}_{>0}$, the polynomial:

$$p = f - s \sum_{i=1}^{r} h_{k,m}(g_i) g_i \tag{12}$$

is such that $p \geq \frac{f^*}{2}$ on $[-1, 1]^n$.

Remark. Our construction of the perturbed polynomial p is similar to the one in [Sch05], [NS07], or [Ave13] where the polynomial h is a univariate (sum of) squares. That choice is simpler, but it results in worst bounds for the degree and the norm of $s\sum_{i=1}^{r} h_{k,m}(g_i)g_i$, than the one we obtain using the polynomial echelon function $h_{k,m}$.

These univariate SoS coefficients have also been used in [KS15], to prove that one can uniformly approximate positive polynomials on compact sets, using the proper subcone of the quadratic module $\mathcal{Q}(\mathbf{g})$ where the SoS coefficient of g_i is of the form $\sum_j (h_j(g_i))^2$, for h_j univariate. They derive a Putinar's Positivstellensatz and apply it to Polynomial Optimization problems. We describe the equivalent of the uniform approximation result in Theorem 4.1, with our coefficients $h_{k,m} \in \mathcal{Q}(1+T, 1-T)$.

Proposition 3.1. *Assume that the normalisation assumptions* (2) *are satisfied. If*

$$s > \frac{6\|f\|}{\delta}; \tag{13}$$

$$k > \frac{2r-2}{\delta} + 1; \tag{14}$$

$$k > \frac{4rs}{f^*}; \tag{15}$$

then $p = f - s\sum_{i=1}^{r} h_{k,m}(g_i)g_i \geq \frac{f^*}{2}$ *on* $[-1,1]^n$.

Proof. Let $x \in A$ so that $G(x) \geq \delta$, i.e. $\min\{g_1(x), \dots, g_r(x), 0\} \leq -\delta$ (see Section 2), and WLOG assume $g_1(x) \leq -\delta$. Notice that from Proposition 2.13 we have $h_{k,m}(g_1(x)) \geq 1 - \frac{1}{k}$ and, if $g_i(x) \geq 0$, $h_{k,m}(g_1(x)) \leq \frac{2}{k}$. Moreover recall that $\|g_i\| \leq \frac{1}{2}$ from the normalisation assumptions (2). Then:

$$\begin{aligned}
p(x) &= f(x) - s\sum_{i=1}^{r} h_{k,m}(g_i(x))g_i(x) \\
&\geq f(x) + s\delta(1-\frac{1}{k}) - s\sum_{i=2}^{r} h_{k,m}(g_i(x))g_i(x) \\
&\geq f(x) + s\delta(1-\frac{1}{k}) - s\frac{r-1}{k} = f(x) + s\frac{\delta}{2}(1-\frac{1}{k}) + s(\frac{\delta}{2}(1-\frac{1}{k}) - \frac{r-1}{k}).
\end{aligned}$$

From Equation (13) and Equation (14), we have respectively $f(x) + s\frac{\delta}{2}(1-\frac{1}{k}) > \frac{\|f\|}{2} \geq \frac{f^*}{2}$ and $\frac{\delta}{2}(1-\frac{1}{k}) - \frac{r-1}{k} > 0$, so that $p(x) > \frac{f^*}{2}$ for $x \in A$.

By Equation (15), $\frac{3f^*}{4} - \frac{sr}{k} > \frac{f^*}{2}$. By the normalization assumptions (2) and as $h_{k,m}$ is upper bounded by $\frac{2}{k}$ on $[0,1]$ (see Proposition 2.13), we therefore deduce that for $x \in [-1,1]^n \setminus A$

$$p(x) = f(x) - s\sum_{i=1}^{r} h_{k,m}(g_i(x))g_i(x) \geq \frac{3f^*}{4} - sr\frac{2}{k}\frac{1}{2} = \frac{3f^*}{4} - \frac{sr}{k} > \frac{f^*}{2}.$$

This shows that $p(x) > \frac{f^*}{2}$ for $x \in [-1,1]^n = A \cup ([-1,1]^n \setminus A)$. □

Proposition 3.2. *Let* p *be as in* (12), *with* (13), (14), (15) *and the normalization assumptions* (2) *satisfied. Let* $d(\mathbf{g}) = \max_i \deg g_i$. *Then*

$$\|p\| = O(\|f\| 2^{3L} r \mathfrak{c} d(f)^{2L} \varepsilon(f)^{-L}), \tag{16}$$

$$\deg p = O(2^{4L} r^{\frac{1}{3}} \mathfrak{c}^{\frac{4}{3}} d(\mathbf{g}) d(f)^{\frac{8L}{3}} \varepsilon(f)^{-\frac{4L+1}{3}}). \tag{17}$$

Proof. Let $d = d(f) = \deg f$. We start bounding m in terms of $\varepsilon(f)$.

We can choose $m = \left\lceil \frac{6}{\delta}\sqrt[3]{\frac{4(k-1)}{3\pi}} + 3\right\rceil$ from Proposition 2.13, thus it is enough to bound k and δ.

From Lemma 2.5 we can choose $\delta = \frac{1}{\mathfrak{c}}(\frac{\varepsilon(f)}{8d^2})^{Ł} = \mathfrak{c}^{-1}\varepsilon(f)^{Ł}2^{-3Ł}d^{-2Ł}$. From Equation (13) we deduce that:

$$s = O(\frac{\|f\|}{\delta}) = O(\|f\|\mathfrak{c}2^{3Ł}d^{2Ł}\varepsilon(f)^{-Ł}). \tag{18}$$

From Equation (14) we deduce that $k = O(\frac{r}{\delta})$, while from Equation (15) (together with Equation (13)) we deduce that $k = O(\frac{r}{\varepsilon(f)\delta})$: the latter has an higher order in terms of $\varepsilon(f)$, and finally:

$$k = O(\mathfrak{c}2^{3Ł}rd^{2Ł}\varepsilon(f)^{-(Ł+1)}). \tag{19}$$

Now we plug Equation (19) in $m = \left\lceil \frac{6}{\delta}\sqrt[3]{\frac{4(k-1)}{3\pi}} + 3\right\rceil$ and obtain:

$$m = O(\frac{k^{\frac{1}{3}}}{\delta}) = O((\mathfrak{c}^{\frac{1}{3}}r^{\frac{1}{3}}2^{Ł}d^{\frac{2Ł}{3}}\varepsilon(f)^{-\frac{Ł+1}{3}})(\mathfrak{c}2^{3Ł}d^{2Ł}\varepsilon(f)^{-Ł})) = O(\mathfrak{c}^{\frac{4}{3}}r^{\frac{1}{3}}2^{4Ł}d^{\frac{8Ł}{3}}\varepsilon(f)^{-\frac{4Ł+1}{3}}). \tag{20}$$

By the normalization assumptions (2), the properties of $h_{k,m}$ (Proposition 2.13) and Equation (18) we obtain:

$$\begin{aligned}\|p\| &\leq \|f\| + s\sum_{i=1}^{r}\left\|h_{k,m}(g_i)g_i\right\| \leq \|f\| + sr(1+\frac{1}{k})\frac{1}{2}\\ &\leq \|f\| + sr = O(\|f\| + \|f\|\mathfrak{c}r2^{3Ł}d^{2Ł}\varepsilon(f)^{-Ł})\\ &= O(\|f\|\mathfrak{c}r2^{3Ł}d^{2Ł}\varepsilon(f)^{-Ł}).\end{aligned}$$

Similarly, using Equation (20) we have:

$$\deg(f-p) \leq \max_i\{\deg(h_{k,m}(g_i)g_i), i = 1,\dots,r\} = O(d(\mathbf{g})m + d(\mathbf{g})) = O(2^{4Ł}r^{\frac{1}{3}}\mathfrak{c}^{\frac{4}{3}}d(\mathbf{g})d^{\frac{8Ł}{3}}\varepsilon(f)^{-\frac{4Ł+1}{3}}),$$

where $d(\mathbf{g}) = \max_i \deg g_i$. □

We now show that $f - p = s\sum_{i=1}^{r} h_{k,m}(g_i(x))g_i(x)$ is in $Q_\ell(\mathbf{g})$, giving degree bounds for the degree ℓ that is necessary to represent $f - p$ (see Proposition 3.7).

Theorem 3.3 (Fekete - Lukács , [PR00])**.** *Let $f \in \mathbb{R}[T]_d$ be a univariate polynomial of degree d. If $f \geq 0$ on $[-1,1]$ then there exists $s_0, s_1, s_2 \in \Sigma^2$ such that $f = s_0 + s_1(1-T) + s_2(1+T)$, where the degree of every addendum is $\leq d+1$. In other words,* $\mathrm{Pos}([-1,1])_d \subset \mathcal{Q}_{d+1}(1-T, 1+T)$.

Proof. From [PR00] (see also [PS76, part VI, 46–47]) there exists polynomials h_i such that $f = h_0^2 + h_1^2(1-T) + h_2^2(1+T) + h_3^2(1-T^2)$, where the degree of every addendum is $\leq d$. Now notice that $1 - T^2 = \frac{1}{2}\Big((1+T)^2(1-T) + (1-T)^2(1+T)\Big)$ to conclude. □

Our assumption is that $\mathcal{Q}(1 - \|\mathbf{X}\|_2^2) \subset \mathcal{Q}(\mathbf{g})$, while we are trying to reduce to the case of $[-1,1]^n$. We show that we can move from the latter to the former with a constant degree shift in Lemma 3.4.

Lemma 3.4. *The preordering associated with the box $[-1,1]^n$ is included in the quadratic module of the unit ball. In particular $\mathcal{O}_d(1 \pm X_i \colon i \in \{1,\dots,n\}) \subset \mathcal{Q}_{d+n}(1 - \|\mathbf{X}\|_2^2)$.*

Proof. Notice that:

$$1 \pm X_i = \frac{1}{2}((1 - X_i^2 + (1 \pm X_i)^2)) = \frac{1}{2}((1 - \|\mathbf{X}\|_2^2 + \sum_{j\neq i} X_j^2 + (1 \pm X_i)^2)).$$

This implies that $\mathcal{Q}_d(1 \pm X_i \colon i \in \{1,\dots,n\}) \subset \mathcal{Q}_{d+1}(1 - \|\mathbf{X}\|^2)$. Since $\mathcal{Q}(1 - \|\mathbf{X}\|^2)$ is a preordering (i.e. it is closed under multiplication) we also have $\mathcal{O}_d(1 \pm X_i \colon i \in \{1,\dots,n\}) \subset \mathcal{Q}_{d+n}(1 - \|\mathbf{X}\|_2^2)$. □

Lemma 3.4 implies that we have a Putinar-like representation of polynomials strictly positive on the box as elements of the quadratic module of the ball.

Lemma 3.5. *Let $\mathcal{Q}(\mathbf{g})$ be a quadratic module such that $1-\|\mathbf{X}\|_2^2 \in \mathcal{Q}(\mathbf{g})$, and let f be a polynomial such that $f>0$ on $[-1,1]^n$. Then $f \in \mathcal{Q}(\mathbf{g})$.*

Proof. Since $f>0$ on $[-1,1]^n$, then $f \in \mathcal{Q}(1-\|\mathbf{X}\|^2)$ by Schmüdgen's Positivstellensatz and Lemma 3.4. Now by hypothesis $\mathcal{Q}(1-\|\mathbf{X}\|^2) \subset \mathcal{Q}(\mathbf{g})$ and thus $f \in \mathcal{Q}(\mathbf{g})$. □

Lemma 3.5 shows that we can use a Schmüdgen theorem on $[-1,1]^n$, for instance Theorem 3.8, to prove that $f \in \mathcal{Q}(\mathbf{g})$, without having proved a general Putinar's Positivstellensatz for $\mathcal{Q}(\mathbf{g})$ yet. Another alternative to prove the result would have been to notice that $f>0$ on $[-1,1]^n$ implies $f>0$ on the unit ball, and then apply a Schmüdgen/Putinar theorem for $\mathcal{Q}(1-\|\mathbf{X}\|^2)$.

We are ready to show that the addenda $h(g_i)g_i$ belong to $\mathcal{Q}(\mathbf{g})$, with degree bounds for the representation.

Lemma 3.6. *Let $h \in \mathrm{Pos}([-1,1])_m$ be a univariate polynomial of degree m. If the normalization assumptions (2) are satisfied and $d(\mathbf{g}) = \max_i \deg g_i$, then $h(g_i)g_i \in \mathcal{Q}_{d(\mathbf{g})m+\ell_0+2}(\mathbf{g})$, where $\ell_0 = \min\{k\colon 1-g_i \in \mathcal{Q}_k(\mathbf{g})\ \forall i=1,\dots,r\}$.*

Proof. By Theorem 3.3, $h \in \mathcal{Q}_{m+1}(1+T,1-T)$, i.e. $h = s_0 + s_1(1+T) + s_2(1-T)$, where s_i is a SoS where $\deg s_0$, $\deg s_1 + 1$ and $\deg s_2 + 1$ are $\le m+1$. Let $d_i = \deg g_i$. Notice that:

- $s_0(g_i)g_i \in \mathcal{Q}_{d_i(m+1)+d_i}(\mathbf{g}) = \mathcal{Q}_{d_i(m+2)}(\mathbf{g})$ since s_0 is a SoS of degree $\le m+1$;
- $s_1(g_i)(1+g_i)g_i = s_1(g_i)g_i + s_1(g_i)g_i^2 \in \mathcal{Q}_{d_i m+2d_i}(\mathbf{g})$ since s_1 is a SoS of degree $\le m$;
- $s_2(g_i)(1-g_i)g_i = s_2(g_i)(g_i-g_i^2) \in \mathcal{Q}(\mathbf{g})$. Indeed $g_i - g_i^2 = (1-g_i)^2 g_i + g_i^2(1-g_i)$, and since $\|g_i\| \le \frac{1}{2}$ we have $(1-g_i) \in \mathcal{Q}(\mathbf{g})$ by Lemma 3.5. In particular let ℓ_0 be minimal such that for all i we have $1-g_i \in \mathcal{Q}_{\ell_0}(\mathbf{g})$. Then $g_i - g_i^2 \in Q_{\ell_0+2}(\mathbf{g})$ and finally $s_2(g_i)(g_i-g_i^2) \in \mathcal{Q}_{d_i m+\ell_0+2}(\mathbf{g})$.

This shows that $h(g_i)g_i = s_0(g_i)g_i + s_1(g_i)(1+g_i)g_i + s_2(g_i)(1-g_i)g_i \in \mathcal{Q}_{d(\mathbf{g})m+\ell_0+2}(\mathbf{g})$, where $d(\mathbf{g}) = \max_i d_i$. □

We now apply Lemma 3.6 to p to determine the degree of the representation of $f-p \in \mathcal{Q}(\mathbf{g})$.

Proposition 3.7. *Let $s\sum_{i=1}^r h_{k,m}(g_i)g_i = f-p$ be as in (12). If the normalization assumptions (2) are satisfied, then $f-p \in \mathcal{Q}_\ell(\mathbf{g})$ when $\ell = O(2^{4Ł} r^{\frac{1}{3}} \mathfrak{c}^{\frac{4}{3}} d(\mathbf{g}) d(f)^{\frac{8Ł}{3}} \varepsilon(f)^{-\frac{4Ł+1}{3}})$, where $\mathfrak{c}, Ł$ are given by Lemma 2.5.*

Proof. It is enough to prove that for all i we have $h_{k,m}(g_i)g_i \in \mathcal{Q}_\ell(\mathbf{g})$. Notice that $h_{k,m}(g_i)g_i \in \mathcal{Q}_{d(\mathbf{g})m+\ell_0+2}(\mathbf{g})$ for all i, see Lemma 3.6. From Equation (20) we can choose $m = O(\mathfrak{c}^{\frac{4}{3}} r^{\frac{1}{3}} 2^{4Ł} d^{\frac{8Ł}{3}} \varepsilon(f)^{-\frac{4Ł+1}{3}})$ and thus if $\ell = O(2^{4Ł} r^{\frac{1}{3}} \mathfrak{c}^{\frac{4}{3}} d(\mathbf{g}) d(f)^{\frac{8Ł}{3}} \varepsilon(f)^{-\frac{4Ł+1}{3}})$ we have $s\sum_{i=1}^r h_{k,m}(g_i)g_i = f-p \in \mathcal{Q}_\ell(\mathbf{g})$. □

3.2 The Polynomial Effective Positivstellensatz

We will use an effective version of Schmüdgen's Positivstellensatz for the box $[-1,1]^n$.

Theorem 3.8 ([LS21])**.** *Let $f \in \mathbb{R}[\mathbf{X}]$, $\deg f = d$ and $f>0$ on $[-1,1]^n$. Let $f_{\min} = \min_{x\in[-1,1]^n} f(x)$ and $f_{\max} = \max_{x\in[-1,1]^n} f(x)$. Then there exists a constant $C(n,d)$ (depending only on n and d) such that $f \in \mathcal{O}_{nr}(1 \pm X_i\colon i \in \{1,\dots,n\})$, where:*

$$r \ge \max\left\{\pi d\sqrt{2n}, \sqrt{\frac{C(n,d)(f_{\max}-f_{\min})}{f_{\min}}}\right\}.$$

Moreover the constant $C(n,d)$ is a polynomial in d for fixed n:

$$C(n,d) \le 2\pi^2 d^2 (d+1)^n n^3 = O(d^{n+2} n^3)$$

Our assumption is that $\mathcal{Q}(1-\|\mathbf{X}\|_2^2)\subset\mathcal{Q}(\mathbf{g})$, while Theorem 3.8 involves $\mathcal{O}(1\pm X_i\colon i\in\{1,\dots,n\})$. But we have already shown in Lemma 3.4 that we can move from the latter to the former with a constant degree shift.

We are now ready to prove the main theorem.

Proof of Theorem 1.7. Let $p=f-s\sum_{i=1}^r h_{k,m}(g_i)g_i$ be as in Equation (12), with s,k,m satisfying *Equation* (13), *Equation* (14), *Equation* (15) and $h_{k,m}$ as in Proposition 2.13. In particular:

- $p\geq\frac{f^*}{2}$ on $[-1,1]^n$ from Proposition 3.1;
- $\|p\|=O(2^{3Ł}r\mathfrak{c}\,d(f)^{2Ł}\|f\|\varepsilon(f)^{-Ł})$ from Equation (16);
- $\deg p=O(2^{4Ł}r^{\frac{1}{3}}\mathfrak{c}^{\frac{4}{3}}d(\mathbf{g})\,d(f)^{\frac{8Ł}{3}}\varepsilon(f)^{-\frac{4Ł+1}{3}})$ from Equation (17).

We apply Theorem 3.8 to p: $p\in\mathcal{O}_{n\ell_0}(1\pm X_i\colon i\in\{1,\dots,n\})$, if $\ell_0\geq\sqrt{\frac{C(n,\deg p)(p_{\max}-p_{\min})}{p_{\min}}}$. Recall also from Theorem 3.8 that $C(n,m)=O(n^3m^{n+2})$. We now deduce the asymptotic order of ℓ_0:

$$\begin{aligned}\sqrt{\frac{C(n,\deg p)(p_{\max}-p_{\min})}{p_{\min}}}&=O\Big(\sqrt{n^3(\deg p)^{n+2}(\frac{2\|p\|}{f^*}+1)}\Big)\\&=O\Big(\sqrt{n^3(2^{4Ł}r^{\frac{1}{3}}\mathfrak{c}^{\frac{4}{3}}d(\mathbf{g})\,d(f)^{\frac{8Ł}{3}}\varepsilon(f)^{-\frac{4Ł+1}{3}})^{n+2}\frac{\|f\|2^{3Ł}r\mathfrak{c}d(f)^{2Ł}\varepsilon(f)^{-Ł}}{f^*}}\Big)\\&=O\Big((n^3 2^{(4n+11)Ł}r^{\frac{n+5}{3}}\mathfrak{c}^{\frac{4n+11}{3}}d(\mathbf{g})^{n+2}d(f)^{\frac{2(4n+11)Ł}{3}}\varepsilon(f)^{-\frac{(4Ł+1)n+11Ł+5}{3}})^{\frac{1}{2}}\Big)\\&=O\Big(n^{\frac{3}{2}}2^{\frac{(4n+11)Ł}{2}}r^{\frac{n+5}{6}}\mathfrak{c}^{\frac{4n+11}{6}}d(\mathbf{g})^{\frac{n+2}{2}}d(f)^{\frac{(4n+11)Ł}{3}}\varepsilon(f)^{-\frac{(4Ł+1)n+11Ł+5}{6}}\Big),\end{aligned}$$

so we can choose $\ell_0=O(n^{\frac{3}{2}}2^{\frac{(4n+11)Ł}{2}}r^{\frac{n+5}{6}}\mathfrak{c}^{\frac{4n+11}{6}}d(\mathbf{g})^{\frac{n+2}{2}}d(f)^{\frac{(4n+11)Ł}{3}}\varepsilon(f)^{-\frac{(4Ł+1)n+11Ł+5}{6}})$ and $p\in\mathcal{O}_{n\ell_0}(1\pm X_i\colon i\in\{1,\dots,n\})$. Now, from Lemma 3.4 we have $\mathcal{O}_{n\ell_0}(1\pm X_i\colon i\in\{1,\dots,n\}\subset\mathcal{Q}_{n\ell_0+n}(1-\|\mathbf{X}\|_2^2)$. Moreover from Equation (2) we have that $1-\|\mathbf{X}\|_2^2\in\mathcal{Q}(\mathbf{g})$. In particular if $1-\|\mathbf{X}\|_2^2\in\mathcal{Q}_{\ell_1}(\mathbf{g})$ and thus $\mathcal{Q}_{n\ell_0+n}(1-\|\mathbf{X}\|_2^2)\subset\mathcal{Q}_{n\ell_0+n+\ell_1}(\mathbf{g})$, i.e. choosing $\ell=nO(\ell_0)=O(n^{\frac{5}{2}}2^{\frac{(4n+11)Ł}{2}}r^{\frac{n+5}{6}}\mathfrak{c}^{\frac{4n+11}{6}}d(\mathbf{g})^{\frac{n+2}{2}}d(f)^{\frac{(4n+11)Ł}{3}}\varepsilon(f)^{-\frac{(4Ł+1)n+11Ł+5}{6}})$ we have $p\in\mathcal{Q}_\ell(\mathbf{g})$. Finally notice that $f=(f-p)+p$ and

- $p\in\mathcal{Q}_\ell(\mathbf{g})$ from the discussion above;
- $f-p\in\mathcal{Q}_\ell(\mathbf{g})$ from Proposition 3.7, since the degree of the truncated quadratic module in Proposition 3.7 is smaller than ℓ.

Then $f\in\mathcal{Q}_\ell(\mathbf{g})$ with

$$\ell=O(n^{\frac{5}{2}}2^{\frac{(4n+11)Ł}{2}}r^{\frac{n+5}{6}}\mathfrak{c}^{\frac{4n+11}{6}}d(\mathbf{g})^{\frac{n+2}{2}}d(f)^{\frac{(4n+11)Ł}{3}}\varepsilon(f)^{-\frac{(4Ł+1)n+11Ł+5}{6}}).\tag{21}$$

We simplify the exponents for readibility. Recall that $Ł\geq1$ and $\mathfrak{c}\geq1$, and assume $n\geq2$. Under these assumptions the inequalities $(4n+11)Ł\leq10nŁ$, $n+5\leq6n$, $4n+11\leq10n$, $n+2\leq2n$ and $(4Ł+1)n+11Ł+5\leq14nŁ$ hold. Therefore we deduce that $f\in\mathcal{Q}_\ell(\mathbf{g})$ if

$$\begin{aligned}\ell&\geq O(n^3 2^{5nŁ}r^n\mathfrak{c}^{2n}d(\mathbf{g})^n d(f)^{3.5nŁ}\varepsilon(f)^{-2.5nŁ})\\&=\gamma(n,\mathbf{g})d(f)^{3.5nŁ}\varepsilon(f)^{-2.5nŁ},\end{aligned}$$

where $\gamma(n,\mathbf{g})=O(n^3 2^{5nŁ}r^n\mathfrak{c}^{2n}d(\mathbf{g})^n)\geq1$. □

Remark. From Equation (21), we have $\ell=O(n^{\frac{5}{2}}2^{\frac{(4n+11)Ł}{2}}r^{\frac{n+5}{6}}\mathfrak{c}^{\frac{4n+11}{6}}d(\mathbf{g})^{\frac{n+2}{2}}d(f)^{\frac{(4n+11)Ł}{3}}\varepsilon(f)^{-\frac{(4Ł+1)n+11Ł+5}{6}})$, where $\mathfrak{c},Ł$ are defined in Definition 2.4. The exponents in Theorem 1.7 have been simplified for the sake of readability and are not optimal.

If the inequalities defining S satisfy a regularity condition we can simplify the bound, since $Ł = 1$ in this case (see Section 2.2).

Corollary 3.9. *Assume $n \geq 2$ and let $g_1, \ldots, g_r \in \mathbb{R}[\mathbf{X}] = \mathbb{R}[X_1, \ldots, X_n]$ satisfying the normalization assumptions* (2) *and such that the CQC (Definition 2.7) hold at every point of $\mathcal{S}(\mathbf{g})$. Let $f \in \mathbb{R}[\mathbf{X}]$ such that $f^* = \min_{x \in S} f(x) > 0$. Then $f \in \mathcal{Q}_\ell(\mathbf{g})$ if*

$$\ell = O(n^3 2^{5n} r^n \mathfrak{c}^{2n} d(\mathbf{g})^n d(f)^{3.5n} \varepsilon(f)^{-2.5n}),$$

where $\mathfrak{c}$ is given by Theorem 2.11.

Proof. Apply Theorem 1.7 and Theorem 2.11. □

4 Convergence of Lasserre's relaxations optimum

We begin with a short description of Polynomial Optimization Problems (POP) and of the Lasserre hierarchies to approximately solve them, and refer to [Las01], [Las15] for more details.

Let $f, g_1, \ldots, g_s \in \mathbb{R}[\mathbf{X}]$. The goal of Polynomial Optimization is to find:

$$f^* := \inf \left\{ f(x) \in \mathbb{R} \mid x \in \mathbb{R}^n,\ g_i(x) \geq 0 \ \text{ for } i = 1, \ldots, s \right\} = \inf_x f(x) \colon g_i(x) \geq 0 \ \forall i \in \{1, \ldots, r\}, \tag{22}$$

that is the infimum f^* of the *objective function* f on the *basic closed semialgebraic set* $S = \mathcal{S}(\mathbf{g})$. It is a general problem, which appears in many contexts and with many applications, see for instance [Las10].

We define the *SoS relaxation of order* ℓ of problem (22) as $\mathcal{Q}_{2\ell}(\mathbf{g})$ and the supremum:

$$f^*_{\mathrm{SoS},\ell} := \sup \left\{ \lambda \in \mathbb{R} \mid f - \lambda \in \mathcal{Q}_{2\ell}(\mathbf{g}) \right\}. \tag{23}$$

Now we want to define the dual approximation of the polynomial optimization problem. We are interested in an affine hyperplane section of the cone $\mathcal{L}_\ell(\mathbf{g}) = \mathcal{Q}_\ell(\mathbf{g})^\vee$:

$$\mathcal{L}^{(1)}_\ell(\mathbf{g}) = \left\{ L \in \mathcal{L}_\ell(\mathbf{g}) \mid \langle L | 1 \rangle = 1 \right\}.$$

With this notation we define the *MoM relaxation of order* ℓ of problem (22) as $\mathcal{L}_{2\ell}(\mathbf{g})$ and the infimum:

$$f^*_{\mathrm{MoM},\ell} := \inf \left\{ \langle L | f \rangle \in \mathbb{R} \mid L \in \mathcal{L}^{(1)}_{2\ell}(\mathbf{g}) \right\}. \tag{24}$$

It is easy to show that the relaxations (23) and (24) are lower approximations of f^*. Their convergenge to f^* as the order ℓ goes to infinity is deduced from Putinar's Positivstellensatz. In particular the rate of convergence can be deduced from the Effective Putinar's Positivstellensatz: see Theorem 4.3. The proof of this result is the purpose of Section 4.

Remark. We have that $f^*_{\mathrm{SoS},\ell} \leq f^*_{\mathrm{MoM},\ell} \leq f^*$ for all ℓ. Thus the results of this section, stated for the SoS relaxations $f^*_{\mathrm{SoS},\ell}$, are also valid for the MoM relaxations $f^*_{\mathrm{MoM},\ell}$.

A first step for the proof of Theorem 4.3 is to recognise Theorem 1.7 as a quantitative result of approximation of polynomials with polynomials in the truncated quadratic module.

Theorem 4.1. *Assume $n \geq 2$ and let $\mathbf{g}$ satisfy the normalization conditions* (2)*. Let Ł be the Łojasiewicz exponent defined in Definition 2.4 and let $f \geq 0$ on $\mathcal{S}(\mathbf{g})$. Then for $0 < \varepsilon \leq \|f\|$, we have $f - f^* + \varepsilon = q \in \mathcal{Q}_\ell(\mathbf{g})$ for*

$$\ell \geq \gamma'(n, \mathbf{g})\, d(f)^{3.5nŁ} \|f\|^{2.5nŁ} \varepsilon^{-2.5nŁ} \tag{25}$$

where $\gamma'(n, \mathbf{g}) = 3^{2.5nŁ} \gamma(n, \mathbf{g}) \geq 1$ depends only on n and $\mathbf{g}$ and $\gamma(n, \mathbf{g})$ is given by Theorem 1.7.

Proof. Notice that $f - f^* + \varepsilon > 0$ on $\mathcal{S}(\mathbf{g})$ and

$$\varepsilon(f - f^* + \varepsilon) = \frac{\varepsilon}{\|f - f^* + \varepsilon\|} \geq \frac{\varepsilon}{\|f\| + |f^*| + \varepsilon} \geq \frac{\varepsilon}{3\|f\|}$$

for $\varepsilon \leq \|f\|$. Moreover $\deg f - f^* + \varepsilon = \deg f = d(f)$. By Theorem 1.7, we have $f - f^* + \varepsilon = q \in Q_\ell(\mathbf{g})$ if

$$\begin{aligned}\ell &\geq O(n^3 2^{5n\text{Ł}} r^n \mathfrak{c}^{2n} d(\mathbf{g})^n d(f)^{3.5n\text{Ł}} (\frac{\varepsilon}{3\|f\|})^{-2.5n\text{Ł}}) \\ &= \gamma'(n,\mathbf{g})\, d(f)^{3.5n\text{Ł}} \|f\|^{2.5n\text{Ł}} \varepsilon^{-2.5n\text{Ł}}\end{aligned}$$

where $\gamma'(n,\mathbf{g}) = 3^{2.5n\text{Ł}}\gamma(n,\mathbf{g}) = O(n^3 2^{5n\text{Ł}} 3^{2.5n\text{Ł}} r^n \mathfrak{c}^{2n} d(\mathbf{g})^n) \geq 1$ depends only on n and $\mathbf{g}$, and not on f, and $\gamma(n,\mathbf{g})$ is given by Theorem 1.7. □

Remark. From Equation (21), we have $\gamma(n,\mathbf{g}) = O(n^{\frac{3}{2}} 2^{\frac{4\text{Ł}n+11\text{Ł}}{2}} r^{\frac{n+5}{6}} \mathfrak{c}^{\frac{4n+11}{6}} d(\mathbf{g})^{\frac{n+2}{2}})$, where $\mathfrak{c}, \text{Ł}$ are defined in Definition 2.4. The exponents of $\gamma'(n,\mathbf{g}) = 3^{2.5n\text{Ł}}\gamma(n,\mathbf{g})$ in the proof have been simplified for the sake of readability and are not optimal.

Remark. Theorem 4.1 is a quantitive version of Weierstrass approximation theorem for positive polynomials on S, showing that a polynomial $f \in \mathrm{Pos}(S(\mathbf{g}))$ can be approximated uniformly on $[-1,1]^n$ (within distance ε) by an element $f^* + q \in Q_\ell(\mathbf{g})$ for $\ell \geq \gamma'(n,\mathbf{g})\, d(f)^{3.5n\text{Ł}} \|f\|^{2.5n\text{Ł}} \varepsilon^{-2.5n\text{Ł}}$.

We are now ready to prove the rate of convergence for Lasserre hierarchies.

Theorem 4.2. *With the same hypothesis of Theorem 4.1, let $f^*_{\mathrm{SoS},\ell}$ be the Lasserre SoS (lower) approximation. Then $f^* - f^*_{\mathrm{SoS},\ell} \leq \varepsilon$ for*

$$\ell \geq \gamma'(n,\mathbf{g})\, d(f)^{3.5n\text{Ł}} \|f\|^{2.5n\text{Ł}} \varepsilon^{-2.5n\text{Ł}}. \tag{26}$$

Proof. Notice that

$$f^*_{\mathrm{SoS},\ell} = \sup\{\lambda \in \mathbb{R} \mid f - \lambda \in \mathcal{Q}_{2\ell}(\mathbf{g})\} = \inf\{\varepsilon \in \mathbb{R}_{\geq 0} \mid f - f^* + \varepsilon \in \mathcal{Q}_{2\ell}(\mathbf{g})\}.$$

By Theorem 4.1, for $\ell \geq \gamma'(n,\mathbf{g})\, d(f)^{3.5n\text{Ł}} \|f\|^{2.5n\text{Ł}} \varepsilon^{-2.5n\text{Ł}}$, $f - f^* + \varepsilon \in \mathcal{Q}_\ell(\mathbf{g})$. This implies that $f^* - f^*_{\mathrm{SoS},\ell} \leq \varepsilon$ and concludes the proof. □

Theorem 4.3. *With the same hypothesis of Theorem 4.2 and $\gamma''(n,\mathbf{g}) = \gamma'(n,\mathbf{g})^{\frac{1}{2.5n\text{Ł}}}$, we have*

$$0 \leq f^* - f^*_{\mathrm{SoS},\ell} \leq \gamma''(n,\mathbf{g}) \|f\| d(f)^{\frac{7}{5}} \ell^{-\frac{1}{2.5n\text{Ł}}}.$$

Proof. We apply Theorem 4.2 with $\varepsilon \leq \|f\|$ such that $\ell = \lceil \gamma'(n,\mathbf{g}) d(f)^{3.5n\text{Ł}} \|f\|^{2.5n\text{Ł}} \varepsilon^{-2.5n\text{Ł}} \rceil$ and $\gamma''(n,\mathbf{g}) = \gamma'(n,\mathbf{g})^{\frac{1}{2.5n\text{Ł}}}$. □

In conclusion Theorem 1.7 allows to prove Theorem 4.3, a polynomial convergence of the Lasserre's lower approximations to f^*. In comparison with [NS07, th. 8], where the convergence is logarithmic in level ℓ of the hierarchy, Theorem 4.3 gives a polynomial convergence to f^*.

In regular POP we can simplify the bound, since $\text{Ł} = 1$ in this case (see Section 2.2).

Corollary 4.4. *With the same hypothesis of Theorem 4.2 and $\gamma''(n,\mathbf{g}) = \gamma'(n,\mathbf{g})^{\frac{1}{2.5n}}$, we have*

$$0 \leq f^* - f^*_{\mathrm{SoS},\ell} \leq \gamma''(n,\mathbf{g}) \|f\| d(f)^{\frac{7}{5}} \ell^{-\frac{1}{2.5n}}$$

if the CQC (Definition 2.7) hold at every point of $\mathcal{S}(\mathbf{g})$.

Proof. Apply Theorem 4.3 and Theorem 2.11. □

5 Convergence of pseudo-moment sequences to measures

We are interested in the study of the truncated positive linear functionals $\mathcal{L}_\ell(\mathbf{g}) = \mathcal{Q}_\ell(\mathbf{g})^\vee$, i.e. the dual convex cone of the truncated quadratic modules, and in particular of its section $\mathcal{L}_d^{(1)}(\mathbf{g})$. This cone is used to define the Lasserre MoM relaxations (24). In the following we often restrict the linear functionals to polynomials of degree $\leq t$, that is we consider the cones $\mathcal{L}_\ell(\mathbf{g})^{[t]}$.

Notice in particular that, if $\mu \in \mathcal{M}(S)^{[t]}$ and $q \in \mathcal{Q}_\ell(\mathbf{g}) \cap \mathbb{R}[\mathbf{X}]_t$ then $\langle \mu | q \rangle = \int q \, \mathrm{d}\mu \geq 0$, since $q \geq 0$ on S. In other words: $\mathcal{M}(S)^{[t]} \subset \mathcal{L}_\ell(\mathbf{g})^{[t]}$ for all ℓ, i.e. our dual cone is an outer approximation of the cone of measures supported on S. To compare quantitatively these cones we consider their affine sections $\mathcal{M}^{(1)}(S)^{[t]}$ and $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}$. Recall that $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}$ is a generating section of $\mathcal{L}_\ell(\mathbf{g})^{[t]}$ when $t \leq \frac{\ell}{2}$, see Section 1.2. In this section, we prove Theorem 1.8, which shows the convergence of the outer approximation as ℓ goes to infinity, and deduce the speed rate from Theorem 1.7. To measure this convergence we use the Hausdorff distance of sets $\mathrm{d_H}(\cdot,\cdot)$.

Before the proof of the main theorem, recall that in the finite dimensional vector space $\mathbb{R}[\mathbf{X}]_t$, all the norms are equivalent: we specify in Lemma 5.1 a constant that we will need in the proof of Theorem 5.7, for the following norms. For $f = \sum_{|\alpha| \leq t} a_\alpha \mathbf{X}^\alpha \in \mathbb{R}[\mathbf{X}]_t$, as usual $\|f\| = \max_{\mathbf{x} \in [-1,1]^n} |f(\mathbf{x})|$, and $\|f\|_2 = \sqrt{\sum_{|\alpha| \leq t} a_\alpha^2}$.

Lemma 5.1. *For $f \in \mathbb{R}[\mathbf{X}]_t$, we have $\|f\| \leq \sqrt{\binom{n+t}{t}} \|f\|_2$.*

Proof. Let $x \in [-1,1]^n$ such that $|f(x)| = \|f\|$. Denote $\bar{\mathbf{x}} = (x^\alpha)_{|\alpha| \leq t}$ and $\bar{a} = (a_\alpha)_{|\alpha| \leq t}$. Then:

$$\|f\| = |f(x)| = |\bar{a} \cdot \bar{x}| \leq \|\bar{a}\|_2 \|\bar{x}\|_2 = \|f\|_2 \|\bar{x}\|_2$$

using the Cauchy-Schwarz inequality. Finally notice that $|x^\alpha| \leq 1$ for all α since $x \in [-1,1]^n$, and thus $\|\bar{x}\|_2 \leq \sqrt{\dim \mathbb{R}[\mathbf{X}]_t} = \sqrt{\binom{n+t}{t}}$, which implies $\|f\| \leq \sqrt{\binom{n+t}{t}} \|f\|_2$. □

We recall a version of Haviland's theorem that characterize linear functionals that are represented by measures supported on a compact set.

Theorem 5.2 ([Sch17, th.17.3])**.** *Let $S \subset \mathbb{R}^n$ be compact and let $\mathrm{Pos}(S)_t = \{f \in \mathbb{R}[\mathbf{X}] \mid \deg f \leq t,\ f(x) \geq 0\ \forall x \in S\}$. Then for a linear functional $L \in \mathbb{R}[\mathbf{X}]_t^*$, $L \in \mathcal{M}(S)^{[t]}$ if and only if $\langle L | f \rangle \geq 0$ for all $f \in \mathrm{Pos}(S)_t$.*

We slightly modify Theorem 5.2 in order to consider only polynomials of unit norm.

Corollary 5.3. *Let $P = \{f \in \mathrm{Pos}(S)_t \mid \|f\|_2 = 1\}$ and let $L \in \mathbb{R}[\mathbf{X}]_t^*$. Then $L \in \mathcal{M}(S)^{[t]} \subset \mathbb{R}[\mathbf{X}]_t^*$ if and only if $\langle L | f \rangle \geq 0$ for all $f \in P$.*

Proof. Notice that $\langle L | f \rangle \geq 0 \iff \left\langle L \middle| \frac{f}{\|f\|_2} \right\rangle \geq 0$. Then apply Theorem 5.2. □

We interpret Corollary 5.3 in terms of convex geometry. The convex set

$$\mathcal{M}(S)^{[t]} = \{L \in \mathbb{R}[\mathbf{X}]_t^* \mid \forall f \in P, \langle L | f \rangle \geq 0\}$$

is the convex cone dual to P. Any $f \in P$ is defining an hyperplane $\langle L | f \rangle = 0$ in $\mathbb{R}[\mathbf{X}]_t^*$, and an associated halfspace $H_f = \{L \in \mathbb{R}[\mathbf{X}]_t^* \mid \langle L | f \rangle \geq 0\}$ such that $\mathcal{M}(S)^{[t]} \subset H_f$. Corollary 5.3 means that $\mathcal{M}(S)^{[t]} = \bigcap_{f \in P} H_f$.

We consider a relaxation of the positivity condition to prove our convergence.

Definition 5.4. For $\varepsilon \geq 0$ and P as in Corollary 5.3, we define $C(\varepsilon) = \{L \in \mathbb{R}[\mathbf{X}]_t^* \mid \forall f \in P, \langle L | f \rangle \geq -\varepsilon\}$.

Notice that by definition and Corollary 5.3 we have $C(0) = \mathcal{M}(S)^{[t]}$.

We show now that $C(\varepsilon)$ contains the truncated positive linear functionals of total mass one for a large enough order of the hierarchy.

Lemma 5.5. *Let $\ell \geq \gamma'(n,\mathbf{g})\, t^{3.5n\text{Ł}} \binom{n+t}{t}^{\frac{5n\text{Ł}}{4}} \varepsilon^{-2.5n\text{Ł}}$, where $\mathbf{g}$ satisfy assumption* (2) *and $\gamma'(n,\mathbf{g})$ is given by Equation* (25)*. Then $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]} \subset C(\varepsilon)$.*

Proof. By Lemma 5.1, for all $f \in P$ we have $\|f\| \leq \binom{n+t}{t}^{\frac{1}{2}}$. From Theorem 4.1, we deduce that for $\ell \geq \gamma'(n,\mathbf{g})\, t^{3.5n\text{Ł}} \binom{n+t}{t}^{\frac{5n\text{Ł}}{4}} \varepsilon^{-2.5n\text{Ł}}$, we have $f - f^* + \varepsilon = q \in \mathcal{Q}_\ell(\mathbf{g})$. Thus for $L \in \mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}$ we obtain $\langle L|f+\varepsilon\rangle = \langle L|q+f^*\rangle \geq 0$. Therefore $\langle L|f\rangle \geq -\varepsilon$: this shows that $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]} \subset C(\varepsilon)$. □

The convex set $C(\varepsilon)$ can be seen as a *tubular* neighborhood of $\mathcal{M}(S)^{[t]}$. We are going to bound its Hausdorff distance to the measures. We state and prove the result in the general setting of convex geometry, and finally use it to prove Theorem 5.7.

Lemma 5.6. *Let $C = \bigcap_{H\in\mathcal{H}} H$ be a closed convex set described as intersection of half spaces $H = \{\mathbf{x} \in \mathbb{R}^N \mid c_H \cdot \mathbf{x} + b_H \geq 0\}$, where*

- *$\|c_H\|_2 = 1$ for all $H \in \mathcal{H}$;*
- *$\mathcal{H}$ is the set of all the half-spaces containing C (of unit normal).*

If $H(\varepsilon) = \{\mathbf{x} \in \mathbb{R}^N \mid c_H \cdot \mathbf{x} + b_H \geq -\varepsilon\}$ and $C(\varepsilon) = \bigcap_{H\in\mathcal{H}} H(\varepsilon)$, then $\mathrm{d_H}(C, C(\varepsilon)) \leq \varepsilon$.

Proof. By definition $C \subset C(\varepsilon)$. Assume that this inclusion is proper, otherwise there is nothing to prove, and let $\xi \in C(\varepsilon) \setminus C$. Consider the closest point η in C of ξ on C, and the half space $H = \{\mathbf{x} \in \mathbb{R}^N \mid \frac{\eta-\xi}{\|\eta-\xi\|_2} \cdot \mathbf{x} + b \geq 0\} \in \mathcal{H}$ defined by the affine supporting hyperplane orthogonal to $\eta - \xi$ passing through η (and thus $\frac{\eta-\xi}{\|\eta-\xi\|_2} \cdot \eta = -b$). Notice that $H \in \mathcal{H}$ since H is defined by a normalized supporting hyperplane of C.

Finally notice that $\left\|\eta - \xi\right\|_2 = \frac{(\eta-\xi)\cdot(\eta-\xi)}{\|\eta-\xi\|_2} = -\frac{\eta-\xi}{\|\eta-\xi\|_2} \cdot \xi + \frac{\eta-\xi}{\|\eta-\xi\|_2} \cdot \eta = -(\frac{\eta-\xi}{\|\eta-\xi\|_2} \cdot \xi + b)$. Since $\xi \in C(\varepsilon)$ and $H \in \mathcal{H}$, we have $(\frac{\eta-\xi}{\|\eta-\xi\|_2} \cdot \xi + b) \geq -\varepsilon$, and thus $0 < \left\|\eta - \xi\right\|_2 \leq \varepsilon$. Then the distance between any $\xi \in C(\varepsilon) \setminus C$ and its closest point $\eta \in C$ is $\leq \varepsilon$, which implies $\mathrm{d_H}(C, C(\varepsilon)) \leq \varepsilon$. □

Theorem 5.7. *Let $\mathcal{Q}(\mathbf{g})$ be a quadratic module where $\mathbf{g}$ satisfy assumption* (2) *and let*

$$\ell \geq \gamma'(n,\mathbf{g})\, t^{3.5n\text{Ł}} \binom{n+t}{t}^{\frac{5n\text{Ł}}{4}} \varepsilon^{-2.5n\text{Ł}}$$

with $\gamma'(n,\mathbf{g})$ given by Equation (25)*. Then $\mathrm{d_H}(\mathcal{M}(S)^{[t]}, \mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}) \leq \varepsilon$.*

Proof. By Corollary 5.3 we have:

$$\mathcal{M}(S)^{[t]} = \{L \in \mathbb{R}[\mathbf{X}]_t^* \mid \ \forall f \in P, \langle L|f\rangle \geq 0\} = \cap_{f\in P} H_f,$$

where $H_f = \{L \in \mathbb{R}[\mathbf{X}]_t^* \mid \langle L|f\rangle \geq 0\}$ with $\|f\|_2 = 1$ and $f \in \mathrm{Pos}(S)_t$. We check that the hyperplanes H_f with $f \in P$ defining $\mathcal{M}(S)^{[t]}$ satisfy the hypothesis of Lemma 5.6:

- The half-space H_f has a unit normal since $\|f\|_2 = 1$;
- Any supporting hyperplane of $\mathcal{M}(S)^{[t]}$ defines an half-space $H_f = \{L \in \mathbb{R}[\mathbf{X}]_t^* \mid \langle L|f\rangle \geq 0$ with $f \in P$. Indeed if f defines a supporting hyperplane of $\mathcal{M}(S)^{[t]}$, then $\langle \mu \big| f\rangle = \int f \,\mathrm{d}\mu \geq 0$ for all $\mu \in \mathcal{M}(S)^{[t]}$. In particular for all $x \in S$ we have $f(x) = \int f \,\mathrm{d}\delta_x \geq 0$ (where δ_x denotes the dirac measure concentred at x). This proves that $f \in \mathrm{Pos}(S)_t$ and, normalizing it, we can assume $f \in P$.

Then from Lemma 5.6 we have $\mathrm{d_H}(\mathcal{M}(S)^{[t]}, C(\varepsilon)) \leq \varepsilon$.

Finally by Lemma 5.5 we deduce that $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]} \subset C(\varepsilon)$ and conclude that

$$\mathrm{d_H}(\mathcal{M}(S)^{[t]}, \mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}) \leq \mathrm{d_H}(\mathcal{M}(S)^{[t]}, C(\varepsilon)) \leq \varepsilon.$$

□

Notice that in Theorem 5.7 we are bounding the distance between normalized linear functionals and measures that may be *not* normalized (i.e. not a probability measure). In the following we solve this problem.

We recall and adapt to our context [JH16, lem. 3] to obtain a bound on the norm of pseudo-moment sequences. In particular we do not assume that the ball constraint is an explicit inequality, but only that the quadratic module is Archimedean.

Lemma 5.8. *Assume that $r^2 - \|\mathbf{X}\|_2^2 = q \in \mathcal{Q}_{\ell_0}(\mathbf{g})$. Then for all $t \in \mathbb{N}$ and $\ell \geq 2t - 2 + \ell_0$, if $L \in \mathcal{L}_\ell^{(1)}(\mathbf{g})$ we have $\left\|L^{[2t]}\right\|_2 \leq \sqrt{\binom{n+t}{t}} \sum_{k=0}^t r^{2k}$.*

Proof. For $L \in \mathcal{L}_\ell^{(1)}(\mathbf{g})$, let H_L^k be the Moment matrix of L in degree $\leq 2k$, which is semi-definite positive. Let $\left\|H_L^k\right\|_F$ be its Frobenius norm, i.e. $\left\|H_L^k\right\|_F = \sqrt{\sum_{|\alpha|,|\beta|\leq k} L_{\alpha+\beta}^2}$, and $\left\|H_L^k\right\|_2$ its ℓ^2 operator norm, i.e. the maximal eigenvalue of H_L^k. Notice that by definition we have $\left\|L^{[2k]}\right\|_2 \leq \left\|H_L^k\right\|_F$ and $\left\|H_L^k\right\|_2 \leq \sqrt{\operatorname{tr} H_L^k}$, Moreover recall $\left\|H_L^k\right\|_F \leq \sqrt{\operatorname{rank}(H_L^k)}\left\|H_L^k\right\|_2$. To obtain a bound on $\left\|L^{[2k]}\right\|_2$, we are going to use $\operatorname{tr} H_L^k = \sum_{|\alpha|\leq k} L_{2\alpha} = \left\langle L^{[2k]} \middle| \sum_{|\alpha|\leq k} \mathbf{X}^{2\alpha} \right\rangle$. As for $k \leq t$,

$$(r^2 - \|\mathbf{X}\|_2^2)(\sum_{|\alpha|\leq k-1} \mathbf{X}^{2\alpha}) \in \mathcal{Q}_{2t-2+\ell_0}(\mathbf{g}) \subset \mathcal{Q}_\ell(\mathbf{g}).$$

we have

$$\begin{aligned} 0 &\leq \left\langle L \middle| (r^2 - \|\mathbf{X}\|_2^2)(\sum_{|\alpha|\leq k-1} \mathbf{X}^{2\alpha}) \right\rangle = r^2 \left\langle L \middle| \sum_{|\alpha|\leq k-1} \mathbf{X}^{2\alpha} \right\rangle - \left\langle L \middle| \|\mathbf{X}\|_2^2 (\sum_{|\alpha|\leq k-1} \mathbf{X}^{2\alpha}) \right\rangle \\ &= r^2 \operatorname{tr} H_L^{k-1} - \left(\left\langle L \middle| \sum_{|\alpha|\leq k} \mathbf{X}^{2\alpha} \right\rangle - \langle L|1\rangle \right) = r^2 \operatorname{tr} H_L^{k-1} + 1 - \operatorname{tr} H_L^k, \end{aligned}$$

that is, $\operatorname{tr} H_L^k \leq r^2 \operatorname{tr} H_L^{k-1} + 1$. Since $\operatorname{tr} H_L^0 = L_0 = 1$, we deduce by induction on k that $\operatorname{tr} H_L^t \leq \sum_{k=0}^t r^{2k}$ and thus

$$\left\|L^{[2t]}\right\|_2 \leq \left\|H_L^t\right\|_F \leq \sqrt{\operatorname{rank}(H_L^t)}\left\|H_L^t\right\|_2 \leq \sqrt{\binom{n+t}{t}} \operatorname{tr} H_L^t \leq \sqrt{\binom{n+t}{t}} \sum_{k=0}^t r^{2k}.$$

□

Finally we are ready to prove Theorem 1.8, where we obtain the bound of the distance between normalized linear functionals and probability measures.

Proof of Theorem 1.8. Let $\varepsilon' = \frac{1}{2}\varepsilon t^{-1}\binom{n+t}{t}^{-\frac{1}{2}} \leq \frac{1}{4}$, $L \in \mathcal{L}_\ell^{(1)}(\mathbf{g})^{[2t]}$ and $\mu \in \mathcal{M}(S)^{[2t]}$ be the closest point to L. We first bound the norm of μ. As

$$\ell \geq \gamma(n, \mathbf{g})\, 6^{2.5n\text{Ł}}\, t^{6n\text{Ł}} \binom{n+t}{t}^{\frac{5n\text{Ł}}{2}} \varepsilon^{-2.5n\text{Ł}} = \gamma'(n, \mathbf{g})\, t^{3.5n\text{Ł}} \binom{n+t}{t}^{\frac{5n\text{Ł}}{4}} (\varepsilon')^{-2.5n\text{Ł}},$$

by Theorem 5.7 we have $\mathrm{d}(L, \mu) \leq \varepsilon'$.

Let $\mu_0 = \int 1 \, d\mu$. We want to bound the distance between L and $\frac{\mu}{\mu_0} \in \mathcal{M}^{(1)}(S)^{[2t]}$. Notice that

$$\mathrm{d}(L, \frac{\mu}{\mu_0}) \leq \mathrm{d}(L, \mu) + \mathrm{d}(\mu, \frac{\mu}{\mu_0}) \leq \varepsilon' + \left|\frac{1-\mu_0}{\mu_0}\right| \left\|\mu\right\|_2. \tag{27}$$

Since $L_0 = 1$, $\mathrm{d}(L, \mu) \leq \varepsilon'$ implies $1 - \varepsilon' \leq \mu_0 \leq 1 + \varepsilon'$, and therefore $\left|\frac{1-\mu_0}{\mu_0}\right| \leq \frac{\varepsilon'}{1-\varepsilon'}$. Moreover, using Lemma 5.8 we have

$$\left\|\mu\right\|_2 = \left\|\mu - L + L\right\| \leq \mathrm{d}(\mu, L) + \|L\|_2 \leq \varepsilon' + (t+1)\sqrt{\binom{n+t}{t}}.$$

Then from Equation (27) we conclude that

$$\mathrm{d}(L, \frac{\mu}{\mu_0}) \leq \varepsilon' + \frac{\varepsilon'}{1-\varepsilon'}(\varepsilon' + (t+1)\sqrt{\binom{n+t}{t}}) = \frac{\varepsilon'}{1-\varepsilon'} + \frac{\varepsilon'}{1-\varepsilon'}(t+1)\sqrt{\binom{n+t}{t}} \leq 2\varepsilon' t\sqrt{\binom{n+t}{t}} = \varepsilon,$$

since $\varepsilon' \leq \frac{1}{4}$, $n \geq 1$ and $t \geq 1$. □

Corollary 5.9. *With the hypothesis of Theorem 1.8 and the CQC (Definition 2.7) satisfied at every point of $\mathcal{S}(\mathbf{g})$, then*

$$\mathrm{d}_{\mathrm{H}}(\mathcal{M}^{(1)}(S)^{[2t]}, \mathcal{L}_\ell^{(1)}(\mathbf{g})^{[2t]}) \leq \varepsilon$$

if $\ell \geq \gamma(n, \mathbf{g})\, 6^{2.5n}\, t^{6n} \binom{n+t}{t}^{2.5nL} \varepsilon^{-2.5n}$.

Proof. Apply Theorem 1.8 and Theorem 2.11. □

In Theorem 1.8 we prove a bound for the convergence of Lasserre truncated pseudo-moments to moments of measures. The convergence, without bounds, can be deduced from [Sch05, th. 3.4] by taking as objective function a constant. On the other hand, we can deduce [Sch05, th. 3.4] from Theorem 1.8, by considering the sections of $\mathcal{L}_\ell^{(1)}(\mathbf{g})^{[t]}$ given by $\langle L|f\rangle = f^*_{\mathrm{MoM},k}$.

In the context of Generalized Moment Problems (GMP), general convergence to moments of measures has been studied in [Tac21]. The uniform bounded mass assumption in [Tac21] is trivially satisfied in the context of Polynomial Optimization, since $L_0 = \langle L|1\rangle = 1$: the convergence result of [Tac21] is thus more general than [Sch05, th. 3.4] and the one implied by Theorem 1.8. But we conjecture, and leave it for future exploration, that it is possible to extend the proof technique of Theorem 1.8 to the GMP and give bounds on the rate of convergence also in this extended context.

Aknowkedgments. The authors thank M. Laurent and L. Slot for the discussion about Schmüdgen's theorem on $[-1,1]^n$, A. Parusiński and K. Kurdyka for the useful suggestions on the Łojasiewicz and Markov inequalities and F. Kirschner for discussions on half space descriptions of convex bodies. The authors thank the anonymous referees for their suggestion, that helped improving the presentation and pointed out errors present in previous versions of the article.

References

[Art27] Emil Artin. "Über die Zerlegung definiter Funktionen in Quadrate". In: *Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg* 5.1 (1927), pp. 100–115.

[Ave13] Gennadiy Averkov. "Constructive Proofs of some Positivstellensätze for Compact Semialgebraic Subsets of R d". In: *Journal of Optimization Theory and Applications* 158.2 (2013), pp. 410–418.

[BCR98] Jacek Bochnak, Michel Coste, and Marie-Francoise Roy. *Real Algebraic Geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Berlin Heidelberg: Springer-Verlag, 1998. ISBN: 978-3-540-64663-1.

[Ber99] Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999. ISBN: 978-1-886529-00-7.

[DKL19] Etienne De Klerk and Monique Laurent. *Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere*. 2019.

[FF20] Kun Fang and Hamza Fawzi. "The sum-of-squares hierarchy on the sphere and applications in quantum information theory". In: *Mathematical Programming* (2020).

[GPR12] Blekherman Grigoriy, A. Parrilo Pablo, and R. Thomas Rekha, eds. *Semidefinite Optimization and Convex Algebraic Geometry*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2012. ISBN: 978-1-61197-228-3.

[Hil88] D. Hilbert. "Ueber die Darstellung definiter Formen als Summe von Formenquadraten". In: *Mathematische Annalen* 32 (1888), pp. 342–350.

[JH16] Cédric Josz and Didier Henrion. "Strong duality in Lasserre's hierarchy for polynomial optimization". In: *Optimization Letters* 10.1 (2016), pp. 3–10.

[KK21] Felix Kirschner and Etienne de Klerk. *Convergence rates of RLT and Lasserre-type hierarchies for the generalized moment problem over the simplex and the sphere*. 2021.

[KR99] András Kroó and Szilárd Révész. "On Bernstein and Markov-Type Inequalities for Multivariate Polynomials on Convex Bodies". In: *Journal of Approximation Theory* 99.1 (1999), pp. 134–152.

[Kri64] J. L. Krivine. "Anneaux préordonnés". In: *Journal d'Analyse Mathématique* 12.1 (1964), pp. 307–326.

[KS15] Krzysztof Kurdyka and StanisŁaw Spodzieja. "Convexifying Positive Polynomials and Sums of Squares Approximation". In: *SIAM Journal on Optimization* 25.4 (2015), pp. 2512–2536.

[KSS16] Krzysztof Kurdyka, Stanisław Spodzieja, and Anna Szlachcińska. "Metric Properties of Semialgebraic Mappings". In: *Discrete & Computational Geometry* 55.4 (2016), pp. 786–800.

[Las01] Jean B. Lasserre. "Global Optimization with Polynomials and the Problem of Moments". In: *SIAM Journal on Optimization* 11.3 (2001), pp. 796–817.

[Las10] Jean-Bernard Lasserre. *Moments, positive polynomials and their applications*. Imperial College Press optimization series v. 1. London : Signapore ; Hackensack, NJ: Imperial College Press ; Distributed by World Scientific Publishing Co, 2010. ISBN: 978-1-84816-445-1.

[Las15] Jean Bernard Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge: Cambridge University Press, 2015. ISBN: 978-1-107-44722-6.

[Łoj59] S. Łojasiewicz. "Sur le problème de la division". In: *Studia Math.* 18 (1959), pp. 87–136.

[LPR20] Henri Lombardi, Daniel Perrucci, and Marie-Françoise Roy. "An elementary recursive bound for effective Positivstellensatz and Hilbert 17-th problem". In: *Memoirs of the American Mathematical Society* 263.1277 (2020).

[LS21] Monique Laurent and Lucas Slot. "An effective version of Schmüdgen's Positivstellensatz for the hypercube". In: *arXiv:2109.09528 [math]* (2021).

[MM22] Ngoc Hoang Anh Mai and Victor Magron. "On the complexity of Putinar–Vasilescu's Positivstellensatz". In: *Journal of Complexity* (2022), p. 101663.

[Mot67] T. S. Motzkin. "The arithmetic-geometric inequality". In: *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)* (1967), pp. 205–224.

[Mou18] Bernard Mourrain. "Polynomial–Exponential Decomposition From Moments". In: *Foundations of Computational Mathematics* 18.6 (2018), pp. 1435–1492.

[MSED21] Victor Magron and Mohab Safey El Din. "On Exact Reznick, Hilbert-Artin and Putinar's Representations". In: *Journal of Symbolic Computation* 107 (2021), pp. 221–250.

[NS07] Jiawang Nie and Markus Schweighofer. "On the complexity of Putinar's Positivstellensatz". In: *Journal of Complexity* 23.1 (2007), pp. 135–150.

[PD01] Alexander Prestel and Charles Delzell. *Positive Polynomials: From Hilbert's 17th Problem to Real Algebra*. Springer Monographs in Mathematics. Berlin Heidelberg: Springer-Verlag, 2001. ISBN: 978-3-540-41215-1.

[PR00] Victoria Powers and Bruce Reznick. "Polynomials that are positive on an interval". In: *Transactions of the American Mathematical Society* 352.10 (2000), pp. 4677–4692.

[PS76] George Pólya and Gabor Szegö. *Problems and Theorems in Analysis II: Theory of Functions. Zeros. Polynomials. Determinants. Number Theory. Geometry*. 1st ed. Classics in Mathematics 216. Springer-Verlag Berlin Heidelberg, 1976. ISBN: 978-3-540-63686-1.

[Put93] Mihai Putinar. "Positive Polynomials on Compact Semi-algebraic Sets". In: *Indiana University Mathematics Journal* 42.3 (1993), pp. 969–984.

[Pó28] G. Pólya. "Über positive Darstellung von Polynomen." In: *Vierteljahrsschrift Zürich* 73 (1928), pp. 141–145.

[Rez95] Bruce Reznick. "Uniform denominators in Hilbert's seventeenth problem". In: *Mathematische Zeitschrift* 220 (1995), pp. 75–97.

[Roc97] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1997.

[Sch04] Markus Schweighofer. "On the complexity of Schmüdgen's Positivstellensatz". In: *Journal of Complexity* 20.4 (2004), pp. 529–543.

[Sch05] Markus Schweighofer. "Optimization of Polynomials on Compact Semialgebraic Sets". In: *SIAM Journal on Optimization* 15.3 (2005), pp. 805–825.

[Sch17] Konrad Schmüdgen. *The Moment Problem*. Graduate Texts in Mathematics. Springer International Publishing, 2017. ISBN: 978-3-319-64545-2.

[Sch91] Konrad Schmüdgen. "The K-moment problem for compact semi-algebraic sets". In: *Mathematische Annalen* 289.1 (1991), pp. 203–206.

[SEDYZ18] Mohab Safey El Din, Zhi-Hong Yang, and Lihong Zhi. "On the complexity of computing real radicals of polynomial systems". In: *ISSAC '18 - The 2018 ACM on International Symposium on Symbolic and Algebraic Computation*. New-York, United States: ACM, 2018, pp. 351–358.

[Ste74] Gilbert Stengle. "A nullstellensatz and a positivstellensatz in semialgebraic geometry". In: *Mathematische Annalen* 207.2 (1974), pp. 87–97.

[Ste96] Gilbert Stengle. "Complexity Estimates for the SchmüDgen Positivstellensatz". In: *J. Complex.* 12.2 (1996), 167–174.

[Tac21] Matteo Tacchi. "Convergence of Lasserre's hierarchy: the general case". In: *Optimization Letters* (2021).

[Tre13] Lloyd N. Trefethen. *Approximation Theory and Approximation Practice*. SIAM, 2013. ISBN: 978-1-61197-240-5.