



HAL
open science

Homogeneous polynomials and spurious local minima on the unit sphere

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. Homogeneous polynomials and spurious local minima on the unit sphere. *Optimization Letters*, 2022, 46 (4), pp.1105-1118. <10.1007/s11590-021-01811-3>. <hal-02966390v2>

HAL Id: hal-02966390

<https://laas.hal.science/hal-02966390v2>

Submitted on 4 May 2021

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.



HAL Authorization

HOMOGENEOUS POLYNOMIALS AND SPURIOUS LOCAL MINIMA ON THE UNIT SPHERE

JEAN B. LASSERRE

ABSTRACT. We consider forms on the Euclidean unit sphere. We obtain a simple and complete characterization of all points that satisfy the standard second-order necessary condition of optimality. It is stated solely in terms of the value of (i) f , (ii) the norm of its gradient, and (iii) the first two smallest eigenvalues of its Hessian, all evaluated at the point. In fact this property also holds for twice continuously differentiable functions that are positively homogeneous. We also characterize a class of degree- d forms with no spurious local minima on \mathbb{S}^{n-1} by using a property of gradient ideals in algebraic geometry.

1. INTRODUCTION

Let \mathbb{S}^{n-1} (resp. \mathcal{E}_n) denote the unit sphere (resp. Euclidean unit ball) in \mathbb{R}^n , and consider the optimization problem

$$(1.1) \quad f^* = \min_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbb{S}^{n-1} \},$$

where f is a degree- d form and f^* is understood as the global minimum. (For linear f or degree-2 forms, (1.1) can be solved efficiently.)

Background. In large-scale optimization problems (as is typical in machine learning applications), so far only first-order methods (e.g. stochastic gradient and its variants) can be implemented. Therefore in the quest of the global minimum it is important to be able to escape spurious local minima (see e.g. works by Jin et al. [4]) or identify and characterize cases where no spurious local minima exist (as e.g. in Ge et al. [9]). See also the discussions in [4, 9] and references therein.

Even though minimizing forms on the unit sphere is a quite specific problem, it has important applications. For instance:

- Finding the maximal cardinality of $\alpha(G)$ of a stable set in a graph G reduces to minimizing a cubic form on the unit sphere.
- Deciding convexity of an n -variate form reduces to minimizing a form on \mathbb{S}^{2n-1} .
- Deciding nonnegativity of an even degree form reduces to minimizing this form on \mathbb{S}^{n-1} .

Work partly funded by the AI Interdisciplinary Institute ANITI through the French “Investing for the Future PISA” program under the Grant agreement ANR-19-PI3A-0004.

- Deciding copositivity of a symmetric matrix reduces to check whether some associated quartic form is nonnegative on \mathbb{R}^n (equivalently on \mathbb{S}^{n-1}).
- In quantum information, the *Best Separable State problem* also relates to homogeneous polynomial optimization; see e.g. [8].

Crucial in the above problems is the search for the *global* optimum and if possible rates of convergence of specialized algorithms like e.g., the Moment-SOS-hierarchy [7, 8] for converging sequences of *lower bounds* and another (different) Moment-SOS-hierarchy for converging sequences of *upper bounds* described in Lasserre [10] with rates provided in de Klerk and Laurent [6]. For more details on applications of homogeneous optimization on the sphere, the interested reader is referred to the discussion in Fang and Fawzi [8], de Klerk and Laurent [6] and the references therein.

In this paper, by restricting to optimization of *forms* on the unit sphere, we provide a complete and rather simple characterization of all points which satisfy first- and second-order optimality conditions, solely in terms of the norm of the gradient of f and the first two smallest eigenvalues of its Hessian, which to the best of our knowledge seems to be new. Therefore all such points (and local minimizers in particular) are characterized by some property of the spectrum of the Hessian; namely how its first two smallest eigenvalues relate to the value of f , an algebraic property of the form. Indeed in the context (1.1), convexity plays little if no role for the absence of spurious local minima. For instance, an arbitrary *quadratic* form $\mathbf{x} \mapsto f(\mathbf{x}) := \mathbf{x}^T \mathbf{Q} \mathbf{x}$ has always a unique local (hence global) minimum (the smallest eigenvalue of \mathbf{Q}) no matter if f is convex or not.

This simple characterization could help to understand the “no spurious local minima” situation. Then combining this characterization with a decomposition property of gradient ideals, one also obtains a sufficient condition that identifies a class of forms with no spurious local minima.

Moreover this characterization is also particularly useful for algorithmic purposes. Indeed it provides an easy practical test in first- and/or second-order minimization algorithms, to check whether a current iterate can be a candidate local minimum.

Contribution. We restrict (1.1) to degree- d forms, with $d > 2$ since for $d \leq 2$ the problem has an easy solution in closed form. Our contribution is two-fold:

- We first provide the following simple and complete characterization of standard first-order and second-order necessary optimality conditions (respectively denoted by **(FONC)** and **(SONC)**).

If $\mathbf{x}^* \in \mathbb{S}^{n-1}$ is a local minimizer then **(FONC)**-**(SONC)** reads:
(1.2)

$$\|\nabla f(\mathbf{x}^*)\| = d|f(\mathbf{x}^*)| \quad \text{and} \quad \begin{cases} \lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq d f(\mathbf{x}^*), & \text{if } f(\mathbf{x}^*) \geq 0, \\ \lambda_2(\nabla^2 f(\mathbf{x}^*)) \geq d f(\mathbf{x}^*), & \text{if } f(\mathbf{x}^*) < 0, \end{cases}$$

where $\nabla^2 f(\mathbf{x}^*)$ is the Hessian of f at \mathbf{x}^* and $\lambda_1(\nabla^2 f(\mathbf{x}^*))$ (resp. $\lambda_2(\nabla^2 f(\mathbf{x}^*))$) denotes the smallest (resp. second smallest) eigenvalue of $\nabla^2 f(\mathbf{x}^*)$.

Moreover, if $f(\mathbf{x}^*) < 0$ then $f(\mathbf{x}^*) = \lambda_1(\nabla^2 f(\mathbf{x}^*)) / d(d-1)$, and the second condition which also reads $\lambda_2(\nabla^2 f(\mathbf{x}^*)) \geq \lambda_1(\nabla^2 f(\mathbf{x}^*)) / (d-1)$, states that the second smallest eigenvalue should be sufficiently separated from the smallest one.

Notice that (1.2) is stated solely in terms of (i) the value of f , (ii) the norm of its gradient, and (iii) the first two smallest eigenvalues of its Hessian, evaluated at the point \mathbf{x}^* . To the best of our knowledge this characterization appears to be new. It is also worth noticing that this characterization remains valid for functions that are positively homogeneous (of degree d) and twice continuously differentiable, i.e., such that $f(\lambda \mathbf{x}) = \lambda^d f(\mathbf{x})$ for all $\lambda > 0$ and all \mathbf{x} .

Then checking whether a point \mathbf{x} satisfies **SONC** is remarkably simple. It reduces to check (1.2), i.e., check whether $\|\nabla f(\mathbf{x})\| = d|f(\mathbf{x})|$ and then compare the value $f(\mathbf{x})$ with the two smallest eigenvalues of the Hessian. This is very useful for any local optimization algorithm since one can easily check whether a current iterate satisfies (1.2).

- Finally, with any degree- d form f we associate a polynomial g of degree d such that (i) g coincide with f on \mathbb{S}^{n-1} , and (ii) all points $\mathbf{x} \in \mathbb{S}^{n-1}$ that satisfy **(FONC)** are critical points of g (i.e. $\nabla g(\mathbf{x}) = 0$) and the converse is also true. Then by using the characterization (1.2) and invoking a certain decomposition of gradient ideals already nicely exploited by Nie et al. [12] for unconstrained optimization, we provide a characterization of a class of degree- d forms with no spurious local minima on \mathbb{S}^{n-1} .

At last but not least, we also remark that if a form f can take negative values then minimizing f on the (convex) Euclidean unit ball \mathcal{E}_n is easier than on \mathbb{S}^{n-1} and yields same (negative) minima and minimizers. In this case one may adapt the previous result and characterize a larger class of degree- d forms with no spurious *negative* local minima on \mathbb{S}^{n-1} .

2. HOMOGENEOUS OPTIMIZATION ON THE SPHERE

2.1. Notation and preliminary results. Let $\mathbb{R}[\mathbf{x}]$ denote the ring of polynomials in the variables $\mathbf{x} = (x_1, \dots, x_n)$ and let $\Sigma[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$ be the space of sums-of-squares polynomials (SOS). Denote by $\mathbb{R}[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]$ the space of polynomials of degree at most d . Let $\nabla f(\mathbf{x})$ (resp. $\nabla^2 f(\mathbf{x})$) denote the gradient (resp. Hessian) of f at \mathbf{x} . Recall that given polynomials

$g_1, \dots, g_s \in \mathbb{R}[\mathbf{x}]$, the notation $I = \langle g_1, g_2, \dots, g_s \rangle$ stands for the ideal

$$\left\{ \sum_{j=1}^s h_j g_j : h_j \in \kappa[\mathbf{x}] \right\}, \quad (\kappa = \mathbb{R} \text{ or } \mathbb{C}),$$

of $\kappa[\mathbf{x}]$ generated by the polynomials g_1, \dots, g_m .

A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is homogeneous of degree d (and called a form) if $f(\lambda \mathbf{x}) = \lambda^d f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and all $\lambda \in \mathbb{R}$. Then the important Euler's identity states that $\langle \nabla f(\mathbf{x}), \mathbf{x} \rangle = d f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Similarly, $\mathbf{x} \mapsto \nabla f(\mathbf{x})$ is homogeneous of degree $d-1$ and so $\nabla^2 f(\mathbf{x}^*) \mathbf{x} = (d-1) \nabla f(\mathbf{x})$.

Given a polynomial $p \in \mathbb{R}[\mathbf{x}]_d$, its homogenization $\tilde{p} \in \mathbb{R}[x_0, \mathbf{x}]_d$ is defined by

$$(x_0, \mathbf{x}) \mapsto \tilde{p}(x_0, \mathbf{x}) := x_0^d p(\mathbf{x}/x_0), \quad (x_0, \mathbf{x}) \in \mathbb{R}^{n+1}.$$

Given n forms $f_1, \dots, f_n \in \mathbb{R}[\mathbf{x}]$ with respective coefficient vectors $\mathbf{f}_1, \dots, \mathbf{f}_n$, and given the system of polynomial equations

$$f_1(\mathbf{x}) = \dots = f_n(\mathbf{x}) = 0,$$

the *resultant* $\text{Res}(f_1, f_2, \dots, f_n) \in \mathbb{R}[\mathbf{f}_1, \dots, \mathbf{f}_n]$ is a homogeneous polynomial in $(\mathbf{f}_1, \dots, \mathbf{f}_n)$ with the property:

$$(2.1) \quad \text{Res}(f_1, f_2, \dots, f_n) = 0 \quad \Leftrightarrow$$

$$(2.2) \quad \exists \mathbf{u} (\neq 0) \in \mathbb{C}^n : \quad f_1(\mathbf{u}) = \dots = f_n(\mathbf{u}) = 0.$$

See e.g. [5, 14, 12].

For a real symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, denote by $\lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}), \dots \leq \lambda_n(\mathbf{A})$, its eigenvalues arranged in increasing order.

Optimization on the Euclidean sphere. A point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ is said to be a local minimizer (and $f(\mathbf{x}^*)$ a local minimum) if there exists $\varepsilon > 0$ and a ball $\mathbf{B}(\mathbf{x}^*, \varepsilon) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon\}$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{S}^{n-1} \cap \mathbf{B}(\mathbf{x}^*, \varepsilon)$.

Below we recall some standard results in optimization, concerned with necessary and/or sufficient for optimality, in the context of the optimization problem (1.1); for a detailed account see e.g. Bertsekas [2].

Proposition 2.1. *Let $f \in \mathbb{R}[\mathbf{x}]$ and for every $\mathbf{x} \in \mathbb{S}^{n-1}$, let $\mathbf{x}^\perp := \{\mathbf{u} \in \mathbb{S}^{n-1} : \mathbf{u}^T \mathbf{x} = 0\}$. If $\mathbf{x}^* \in \mathbb{S}^{n-1}$ is local minimizer of (1.1) then there exists $\lambda^* \in \mathbb{R}$ such that:*

(i) *The **First-Order Necessary Optimality-Condition (FONC)** holds:*

$$(2.3) \quad \nabla f(\mathbf{x}^*) + 2\lambda^* \mathbf{x}^* = 0.$$

(ii) *The **Second-Order Necessary Optimality-Condition (SONC)** holds:*

$$(2.4) \quad \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u} + 2\lambda^* \geq 0, \quad \forall \mathbf{u} \in (\mathbf{x}^*)^\perp.$$

(iii) *Conversely, if $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfies (2.3) and the **Second-Order Sufficiency Optimality-Condition (SOSC)***

$$(2.5) \quad \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u} + 2\lambda^* > 0, \quad \forall \mathbf{u} \in (\mathbf{x}^*)^\perp,$$

then \mathbf{x}^* is a local minimizer of (1.1).

Proof. At $\mathbf{x}^* \in \mathbb{S}^{n-1}$ the gradient of the constraint $\|\mathbf{x}\|^2 = 1$ at \mathbf{x}^* is simply $2\mathbf{x}^*$ ($\neq 0$) and therefore is linearly independent, i.e., a basic constraint qualification holds true. Therefore (2.3)-(2.4) and (iii) follow from standard results in non-linear programming [2]. \square

The following result is an easy consequence of Proposition 2.1 but useful for our purpose.

Corollary 2.2. *Let f be a degree- d form and $\mathbf{x}^* \in \mathbb{S}^{n-1}$ be a local minimizer. Then in (2.3), $2\lambda^* = -df(\mathbf{x}^*)$. In addition, (2.3) holds if and only if*

$$(2.6) \quad \|\nabla f(\mathbf{x}^*)\|^2 = d^2 f(\mathbf{x}^*)^2,$$

and **(SONC)** reads:

$$(2.7) \quad \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u} \geq df(\mathbf{x}^*), \quad \forall \mathbf{u} \in (\mathbf{x}^*)^\perp.$$

Proof. In (2.3) we obtain

$$df(\mathbf{x}^*) = \langle \nabla f(\mathbf{x}^*), \mathbf{x}^* \rangle = -2\lambda^* \|\mathbf{x}^*\|^2 = -2\lambda^*,$$

and therefore $\|\nabla f(\mathbf{x}^*)\|^2 = (2\lambda^*)^2 \|\mathbf{x}^*\|^2 = d^2 f(\mathbf{x}^*)^2$. Then (2.7) follows from (2.4). Conversely, assume that (2.6) holds at $\mathbf{x}^* \in \mathbb{S}^{n-1}$. Then

$$\|\nabla f(\mathbf{x}^*) - df(\mathbf{x}^*) \mathbf{x}^*\|^2 = \|\nabla f(\mathbf{x}^*)\|^2 - \underbrace{2df(\mathbf{x}^*) \langle \nabla f(\mathbf{x}^*), \mathbf{x}^* \rangle}_{=-2d^2 f(\mathbf{x}^*)^2} + d^2 f(\mathbf{x}^*)^2 \|\mathbf{x}^*\|^2,$$

that is,

$$\|\nabla f(\mathbf{x}^*) - df(\mathbf{x}^*) \mathbf{x}^*\|^2 = \|\nabla f(\mathbf{x}^*)\|^2 - d^2 f(\mathbf{x}^*)^2 = 0,$$

and so (2.3) holds with $\lambda^* = -df(\mathbf{x}^*)/2$, and again (2.7) follows from (2.4). \square

Note in passing that all **FONC** points are solutions of

$$\nabla f(\mathbf{x}) = df(\mathbf{x}) \mathbf{x},$$

a system of n polynomial equations in n variables (the dual variable λ^* in (2.4) has been identified, thanks to Euler' identity). Then generically, by Bezout's theorem it has at most $(d+1)^n$ solutions.

2.2. A distinguished representation. In this section we obtain a more specific characterization of points that satisfies **(FONC)**-**(SONC)** solely in terms of $f(\mathbf{x}^*)$, $\lambda_1(\nabla^2 f(\mathbf{x}^*))$ and $\lambda_2(\nabla^2 f(\mathbf{x}^*))$.

When $d \leq 2$, Problem (1.1) is easy and completely solved analytically so we only consider the case $d > 2$.

Lemma 2.3. *Let $f \in \mathbb{R}[\mathbf{x}]$ be a form of degree $d > 2$, and let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfy **(FONC)**. Define:*

$$(2.8) \quad \tau(\mathbf{x}^*) := \min_{\mathbf{u} \in (\mathbf{x}^*)^\perp} \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u}.$$

Then

$$(2.9) \quad \lambda_1(\nabla^2 f(\mathbf{x}^*)) = \min [d(d-1)f(\mathbf{x}^*), \tau(\mathbf{x}^*)],$$

and if $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$ then $\tau(\mathbf{x}^*) = \lambda_2(\nabla^2 f(\mathbf{x}^*))$.

Hence if $\tau(\mathbf{x}^*) \neq \lambda_1(\nabla^2 f(\mathbf{x}^*))$ then $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$ and $\tau(\mathbf{x}^*) = \lambda_2(\nabla^2 f(\mathbf{x}^*))$.

Proof. Observe that $\mathbb{R}^n = \theta \mathbf{x}^* \oplus \gamma (\mathbf{x}^*)^\perp$ where θ, γ runs over \mathbb{R} . Then writing $\mathbf{v} \in \mathbb{S}^{n-1}$ as $\theta \mathbf{x}^* + \gamma \mathbf{u}$ with $\mathbf{u} \in (\mathbf{x}^*)^\perp$, one obtains $\|\mathbf{v}\|^2 = \theta^2 + \gamma^2$. Next,

$$\mathbf{v}^T \nabla^2 f(\mathbf{x}^*) \mathbf{v} = \theta^2 \langle \mathbf{x}^*, \nabla^2 f(\mathbf{x}^*) \mathbf{x}^* \rangle + 2\gamma\theta \langle \mathbf{u}, \nabla^2 f(\mathbf{x}^*) \mathbf{x}^* \rangle + \gamma^2 \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u}.$$

Using homogeneity of f (hence of $\nabla f(\mathbf{x})$ as well), yields

$$\langle \mathbf{x}^*, \nabla^2 f(\mathbf{x}^*) \mathbf{x}^* \rangle = (d-1) \langle \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle = d(d-1)f(\mathbf{x}^*),$$

and

$$\langle \mathbf{u}, \nabla^2 f(\mathbf{x}^*) \mathbf{x}^* \rangle = (d-1) \langle \mathbf{u}, \nabla f(\mathbf{x}^*) \rangle = d(d-1)f(\mathbf{x}^*) \mathbf{u}^T \mathbf{x}^* = 0,$$

so that

$$\mathbf{v}^T \nabla^2 f(\mathbf{x}^*) \mathbf{v} = \theta^2 d(d-1)f(\mathbf{x}^*) + \gamma^2 \langle \mathbf{u}, \nabla^2 f(\mathbf{x}^*) \mathbf{u} \rangle.$$

This yields

$$\lambda_1(\nabla^2 f(\mathbf{x}^*)) = \min_{\|\mathbf{v}\|=1} \mathbf{v}^T \nabla^2 f(\mathbf{x}^*) \mathbf{v} = \min [d(d-1)f(\mathbf{x}^*), \tau(\mathbf{x}^*)],$$

which is the desired result (2.9). Next, if $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$ (hence with associated eigenvector \mathbf{x}^*), then

$$\lambda_2(\nabla^2 f(\mathbf{x}^*)) = \min_{\mathbf{v} \perp \mathbf{x}^*; \|\mathbf{v}\|=1} \mathbf{v}^T \nabla^2 f(\mathbf{x}^*) \mathbf{v} = \tau(\mathbf{x}^*).$$

Conversely, if $\tau(\mathbf{x}^*) = \lambda_2(\nabla^2 f(\mathbf{x}^*)) > \lambda_1(\nabla^2 f(\mathbf{x}^*))$ then by (2.9), $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$. \square

We are now in position to characterize in a simple compact form, all points of \mathbb{S}^{n-1} that satisfy **(SONC)** when f is a degree- d form.

Corollary 2.4. *Let f be a degree- d form with $d > 2$, and let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfy **(FONC)**. Then \mathbf{x}^* satisfies **(SONC)** if and only if:*

$$(2.10) \quad \lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq d f(\mathbf{x}^*) \quad \text{if } f(\mathbf{x}^*) \geq 0$$

$$(2.11) \quad \lambda_2(\nabla^2 f(\mathbf{x}^*)) \geq d f(\mathbf{x}^*) \quad \text{if } f(\mathbf{x}^*) < 0.$$

Moreover, if $f(\mathbf{x}^*) < 0$ then $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$.

If $d = 2$ then \mathbf{x}^* satisfies **(SONC)** if and only if $\lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq d f(\mathbf{x}^*)$ and there is only one local (hence global) minimum.

Proof. i) $d > 2$. First consider the case $f(\mathbf{x}^*) < 0$. By **(SONC)**, $\tau(\mathbf{x}^*) \geq df(\mathbf{x}^*) > d(d-1)f(\mathbf{x}^*)$, and therefore by Lemma 2.3, $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$ and $\lambda_2(\nabla^2 f(\mathbf{x}^*)) = \tau(\mathbf{x}^*) \geq df(\mathbf{x}^*)$.

Conversely, suppose that $\lambda_2(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*)$. Then $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$ because $d(d-1)f(\mathbf{x}^*) < df(\mathbf{x}^*) \leq \lambda_2(\nabla^2 f(\mathbf{x}^*))$ and $d(d-1)f(\mathbf{x}^*)$ is an eigenvalue. Hence by Lemma 2.3, $\lambda_2(\nabla^2 f(\mathbf{x}^*)) = \tau(\mathbf{x}^*) \geq df(\mathbf{x}^*)$, i.e., **(SONC)** holds.

Next, consider the case $f(\mathbf{x}^*) \geq 0$. Then **(SONC)** \Rightarrow (2.10) follows from Lemma 2.3. Indeed if $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1)f(\mathbf{x}^*)$ then $\lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*)$, and if $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = \tau(\mathbf{x}^*)$ then $\lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*)$ by **(SONC)**. (2.10) \Rightarrow **(SONC)**. Again by Lemma 2.3, $\tau(\mathbf{x}^*) \geq \lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*)$, and therefore **(SONC)** holds.

ii) $d = 2$. Then $d(d-1) = d$ and $f(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x}$ for some real matrix \mathbf{Q} . Then each point \mathbf{x}^* that satisfies **(FONC)** is an eigenvector of \mathbf{Q} with associated eigenvalue $f(\mathbf{x}^*) \in \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and $\nabla^2 f(\mathbf{x}) = 2\mathbf{Q}$ for all \mathbf{x} . So let \mathbf{x}^* satisfies **(FONC)**.

If $f(\mathbf{x}^*) = \lambda_j$ with $j > 1$, then necessarily $\tau(\mathbf{x}^*) = d\lambda_1 \leq df(\mathbf{x}^*)$ with equality only if $\lambda_k = \lambda_1$ for all $2 \leq k \leq j$. Hence **(SONC)** holds only if $f(\mathbf{x}^*) = \lambda_1$ and therefore $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d\lambda_1 \geq df(\mathbf{x}^*)$. Conversely let $\lambda_1(\nabla^2 f(\mathbf{x}^*)) (= d\lambda_1) \geq df(\mathbf{x}^*)$ then necessarily $f(\mathbf{x}^*) = \lambda_1$ and **(SONC)** holds because $\tau(\mathbf{x}^*) = d\lambda_2 \geq d\lambda_1 = df(\mathbf{x}^*)$. \square

So Corollary 2.4 states that in homogeneous optimization on the Euclidean sphere, first- and second-order necessary optimality conditions can be easily checked by inspection of the gradient and the first two smallest eigenvalues of the Hessian of f . In particular, if $\mathbf{x}^* \in \mathbb{S}^{n-1}$ is a local minimizer with $f(\mathbf{x}^*) < 0$ then $f(\mathbf{x}^*) = \lambda_1(\nabla^2 f(\mathbf{x}^*)) / d(d-1)$ with \mathbf{x}^* being the corresponding eigenvector of $\nabla^2 f(\mathbf{x}^*)$.

Remark 2.5. *It is worth noticing that the characterization of **(FONC)** in (2.6) and **(SONC)** in Corollary 2.4 remains valid for twice continuously differentiable and positively homogeneous functions of degree d , that is, functions f that satisfy $f(\lambda \mathbf{x}) = \lambda^d f(\mathbf{x})$ for all $\lambda > 0$ and all $\mathbf{x} \in \mathbb{R}^n$. Indeed nowhere in the proof we have used the fact that f is a polynomial.*

2.3. Minimizing on \mathcal{E}_n rather than on \mathbb{S}^{n-1} . Notice that (2.3) (or equivalently (2.6)) also holds at a local maximum.

In this section we remark that if f is a form, all non positive local minima of f in (1.1) are also local minima on \mathcal{E}_n . Conversely, *all* local minima f^* on \mathcal{E}_n are non positive (i.e., necessarily $f^* \leq 0$) and are local minima on \mathbb{S}^{n-1} (except if $f^* = 0$ is attained only at $\mathbf{x}^* = 0$); hence in particular, no local maximum on \mathcal{E}_n can be negative.

So if f can take negative values then it is definitely better and easier to minimize on \mathcal{E}_n because \mathcal{E}_n is a convex set. In doing so one obtains a negative local minimum and avoid any positive local minimum on \mathbb{S}^{n-1} .

Lemma 2.6. *Let $f \in \mathbb{R}[\mathbf{x}]$ be a form of degree d . Then:*

(i) Every local minimum f^* on \mathcal{E}_n satisfies $f^* \leq 0$. If $f^* < 0$ then it is attained at some $\mathbf{x}^* \in \mathbb{S}^{n-1}$ and so f^* is also a local minimum on \mathbb{S}^{n-1} . If $f^* = 0$ then either f^* is attained only at $\mathbf{x}^* = 0$ or f^* is also a local minimum (also attained) on \mathbb{S}^{n-1} .

(ii) Every local minimum $f^* \leq 0$ on \mathbb{S}^{n-1} is also a local minimum on \mathcal{E}_n .

Proof. (i) Assume that $f^* > 0$ is a local minimum on \mathcal{E}_n hence for some local minimizer $0 \neq \mathbf{x}^* \in \mathcal{E}_n$. Then $\lambda \mathbf{x}^* \in \mathcal{E}_n$ for every $\lambda \in (0, 1)$, and by homogeneity of f one obtains $f(\lambda \mathbf{x}^*) = \lambda^d f(\mathbf{x}^*) = \lambda^d f^* < f^*$, in contradiction with the hypothesis. Next, assume that $f^* < 0$ and $\mathbf{x}^* \in \mathcal{E}_n$ is a local minimizer (hence $\mathbf{x}^* \neq 0$) with $\|\mathbf{x}^*\| < 1$. Then $\mathbf{z}^* := \lambda \mathbf{x}^* \in \mathbb{S}^{n-1}$ for some $\lambda > 1$, and $f(\mathbf{z}^*) = \lambda^d f(\mathbf{x}^*) < f^*$, a contradiction and so necessarily $\mathbf{x}^* \in \mathbb{S}^{n-1}$. If $f^* = 0$ and f^* is attained at $\mathbf{x}^* \neq 0$, then $f^* = 0$ is also attained at $\mathbf{z}^* = \mathbf{x}^*/\|\mathbf{x}^*\| \in \mathbb{S}^{n-1}$ and so is also a local minimum on \mathbb{S}^{n-1} .

(ii) We proceed by contradiction. Assume $\mathbf{x}^* \in \mathbb{S}^{n-1}$ is a local minimizer of f on \mathbb{S}^{n-1} with $f^* \leq 0$ and not a local minimizer on \mathcal{E}_n . Let $\mathbf{B}_j(\mathbf{x}^*) := \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}^*\|^2 < 1/j\}$. Then for every integer $j > n_0$, there exists $\mathbf{y}_j \in \mathbf{B}_j(\mathbf{x}^*) \cap \mathcal{E}_n$ with $f(\mathbf{y}_j) < f(\mathbf{x}^*) \leq 0$. Letting $\mathbf{z}_j := \mathbf{y}_j/\|\mathbf{y}_j\| \in \mathbb{S}^{n-1}$, one obtains $f(\mathbf{z}_j) = \|\mathbf{y}_j\|^{-d} f(\mathbf{y}_j) \leq f(\mathbf{y}_j) < f(\mathbf{x}^*)$. By letting j increase one has exhibited a sequence $(\mathbf{z}_j)_{j \in \mathbb{N}} \subset \mathbb{S}^{n-1}$ converging to \mathbf{x}^* and with cost $f(\mathbf{z}_j) < f(\mathbf{x}^*)$ for all j , in contradiction with our hypothesis. \square

So if f is homogeneous and not nonnegative on \mathbb{R}^n , then its global minimum f^* on \mathbb{S}^{n-1} is strictly negative. Then by Lemma 2.6, searching for the global minimum f^* is equivalent to searching for the global minimum of f on the larger (but convex) set \mathcal{E}_n . Therefore consider the case where f has no spurious negative local minima on \mathbb{S}^{n-1} (hence no spurious negative local minima on \mathcal{E}_n) while spurious positive local minima on \mathbb{S}^{n-1} may exist. In such a case, any local minimization algorithm on \mathcal{E}_n (starting at $\mathbf{x}_0 \in \mathcal{E}_n$ with $f(\mathbf{x}_0) < 0$) converging to a Karush-Kuhn-Tucker point (i.e. a point that satisfies **(FONC)**) will find the global minimum on \mathcal{E}_n (and hence on \mathbb{S}^{n-1}), and optimizing over \mathcal{E}_n is certainly easier than on \mathbb{S}^{n-1} .

3. NO SPURIOUS LOCAL MINIMA ON \mathbb{S}^{n-1}

In this section we are concerned with the “no spurious local minima” situation, and characterize a class of degree- d forms that have no spurious local minima on \mathbb{S}^{n-1} .

Corollary 3.1. *Let f be a degree- d form and*

$$\Theta := \{\mathbf{x} \in \mathbb{S}^{n-1} : \|\nabla f(\mathbf{x})\| = d|f(\mathbf{x})|\},$$

*i.e., Θ is the set of all points of \mathbb{S}^{n-1} that satisfy **(FONC)**.*

i) If f is nonnegative then it has no spurious local minima on \mathbb{S}^{n-1} if f is constant on the set

$$\Delta^+ := \{\mathbf{x} \in \Theta : \lambda_1(\nabla^2 f(\mathbf{x})) \geq \|\nabla f(\mathbf{x})\|\},$$

*in which case all **SONC** points are global minimizers.*

ii) If f can take negative values then it has no spurious local minima on \mathbb{S}^{n-1} if

$$\{\mathbf{x} \in \Theta : f(\mathbf{x}) \geq 0; \lambda_1(\nabla^2 f(\mathbf{x})) \geq \|\nabla f(\mathbf{x})\|\} = \emptyset,$$

and f is constant on the set

$$\Delta^- := \{\mathbf{x} \in \Theta : f(\mathbf{x}) < 0; \lambda_2(\nabla^2 f(\mathbf{x})) \geq -\|\nabla f(\mathbf{x})\|\}.$$

Proof. (i) By Corollary 2.2, $\|\nabla f(\mathbf{x})\| = d|f(\mathbf{x})|$ on Θ . Next, if f is constant on Δ^+ and f is nonnegative, then by Corollary 2.4, all points $\mathbf{x}^* \in \mathbb{S}^{n-1}$ that satisfy **SONC** (in particular all local minimizers) belong to Δ^+ . So if f is constant on Δ^+ , all **SONC** points (and local minimizers in particular) have same nonnegative value, and therefore they all are global minimizers. A similar argument applies to prove (ii) \square

Notice that Corollary 3.1 also holds for twice continuously differentiable positively homogeneous functions.

We next show that the characterization in Corollary 3.1 is also related to a property of gradient ideals of $\mathbb{R}[\mathbf{x}]$. We introduce a polynomial (not a form) with the following nice property. On \mathbb{S}^{n-1} :

- (i) it coincides with f (up to a multiplicative constant), and
- (ii) all its critical points coincide with **FONC** points of f .

We then invoke a property of gradient ideals nicely exploited in Nie et al. [12].

Given a degree- d form f , let $g \in \mathbb{R}[\mathbf{x}]_d$ be the polynomial

$$(3.1) \quad \mathbf{x} \mapsto g(\mathbf{x}) := f(\mathbf{x}) \left(1 - \frac{d}{d+2} \|\mathbf{x}\|^2\right), \quad \mathbf{x} \in \mathbb{R}^n.$$

Proposition 3.2. *Let f be a degree- d form and let $g \in \mathbb{R}[\mathbf{x}]$ be as in (3.1). Then on \mathbb{S}^{n-1} :*

$$(3.2) \quad \nabla g(\mathbf{x}) = 0 \Leftrightarrow \nabla f(\mathbf{x}) = d f(\mathbf{x}) \cdot \mathbf{x} \Leftrightarrow \|\nabla f(\mathbf{x})\|^2 = (d f(\mathbf{x}))^2.$$

*That is, all critical points of g in \mathbb{S}^{n-1} satisfy **(FONC)** for f , and conversely, all points of \mathbb{S}^{n-1} that satisfy **(FONC)** for f are critical points of g .*

Proof. Observe that

$$\nabla g(\mathbf{x}) = \nabla f(\mathbf{x}) \left(1 - \frac{d}{d+2} \|\mathbf{x}\|^2\right) - \frac{2d}{d+2} f(\mathbf{x}) \mathbf{x}.$$

Therefore if $\mathbf{x} \in \mathbb{S}^{n-1}$ then $g(\mathbf{x}) = \frac{2}{d+2} f(\mathbf{x})$ and ,

$$\nabla g(\mathbf{x}) = \frac{2}{d+2} \nabla f(\mathbf{x}) - \frac{2d}{d+2} f(\mathbf{x}) \cdot \mathbf{x} = \frac{2}{d+2} (\nabla f(\mathbf{x}) - d f(\mathbf{x}) \mathbf{x}),$$

and so $\nabla g(\mathbf{x}^*) = 0$ if and only if (2.3) holds with $2\lambda^* = -d f(\mathbf{x}^*)$, which yields the desired result (3.2). \square

Moreover, on \mathbb{S}^{n-1} minimizing f is strictly equivalent to minimizing g since $g(\mathbf{x}) = 2f(\mathbf{x})/(d+2)$ on \mathbb{S}^{n-1} . Next, with g as in (3.1), define the *gradient ideal*:

$$\mathcal{I}_{\text{grad}}(g) := \left\langle \frac{\partial g(\mathbf{x})}{x_1}, \dots, \frac{\partial g(\mathbf{x})}{x_n} \right\rangle,$$

and its associated variety

$$V_{\text{grad}}(g) := V(\mathcal{I}_{\text{grad}}(g)) = \{\mathbf{z} \in \mathbb{C}^n : \nabla g(\mathbf{z}) = 0\}.$$

Then $V_{\text{grad}}(g)$ is a finite union of irreducible subvarieties W_j 's, that is,

$$V_{\text{grad}}(g) = W_0 \cup W_1 \dots \cup W_s,$$

with $W_0 \cap \mathbb{R}^n = \emptyset$ and in addition, g is a real constant on each W_j , $j \geq 1$; see e.g. [3, §2] and [12, p. 592]. So we can regroup all components on which g takes the same value, and write

$$(3.3) \quad V_{\text{grad}}(g) = W_0 \cup \tilde{W}_1 \dots \cup \tilde{W}_r,$$

where $g(\mathbf{x}) = g_j$ on \tilde{W}_j and $g_j \neq g_i$ for all $1 \leq i, j$ with $i \neq j$.

We are now in position to provide a characterization of a class of degree- d forms f with no spurious local minima on \mathbb{S}^{n-1} .

Theorem 3.3. *Consider problem (1.1) where f is a degree- d form ($d > 2$). Then f has no spurious local minima on \mathbb{S}^{n-1} if there is only one index $j^* \geq 1$ in (3.3) such that*

$$\tilde{W}_{j^*} \cap \Omega \neq \emptyset,$$

where

$$(3.4) \quad \Omega := \{\mathbf{x} \in \mathbb{S}^{n-1} : f(\mathbf{x}) < 0; \lambda_2(\nabla^2 f(\mathbf{x})) \geq -\|\nabla f(\mathbf{x})\|\}$$

$$(3.5) \quad \cup \{\mathbf{x} \in \mathbb{S}^{n-1} : f(\mathbf{x}) \geq 0; \lambda_1(\nabla^2 f(\mathbf{x})) \geq \|\nabla f(\mathbf{x})\|\}.$$

If f has no spurious local minima on \mathbb{S}^{n-1} then all **SONC** points with value f^* (i.e., all local hence global minimizers) belong to a unique set $\tilde{W}_{j^*} \cap \Omega$. The other nonempty sets $\tilde{W}_j \cap \Omega \neq \emptyset$ contain **SONC** points which are not local minimizers.

Proof. By construction each nonempty set $\tilde{W}_j \cap \Omega$ identifies a subset of **SONC** points of f which share the same value, say f_j^* . Therefore $f_j^* \geq f^*$ for all j (where f^* is the global minimum). So if there only one such set $\tilde{W}_{j^*} \cap \Omega \neq \emptyset$ then necessarily $f_{j^*}^* = f^*$, otherwise a spurious local minimum $\tau > f^*$ would correspond to some **SONC** point in another nonempty set $\tilde{W}_j \cap \Omega$, with $j \neq j^*$, and $f_j^* = \tau$.

With similar arguments, if f has no spurious local minima on \mathbb{S}^{n-1} then necessarily all local (hence global) minimizers are **SONC** points and belong to the same set $\tilde{W}_{j^*} \cap \Omega$ for some index j^* . All other nonempty sets $\tilde{W}_j \cap \Omega \neq \emptyset$ contain **SONC** points which cannot be local minimizers as their associated value $f_j^* \neq f^*$ must be larger than f^* . \square

We have seen that if a degree- d form can take negative values then all its local minima on \mathcal{E}_n are negative local minima on \mathbb{S}^{n-1} , and the converse is true. Then for minimizing on \mathcal{E}_n , it is interesting to characterize a class of degree- d forms with the less restrictive condition of no spurious *negative* local minima on \mathcal{E}_n (hence on \mathbb{S}^{n-1}).

Corollary 3.4. *Consider problem (1.1) where f is a degree- d form ($d > 2$). Then f has no spurious negative local minima on \mathbb{S}^{n-1} if there is only one index $j^* \geq 1$ in (3.3) such that*

$$\tilde{W}_{j^*} \cap \Omega \neq \emptyset,$$

where $\Omega = \{\mathbf{x} \in \mathbb{S}^{n-1} : f(\mathbf{x}) < 0; \lambda_2(\nabla^2 f(\mathbf{x})) \geq -\|\nabla f(\mathbf{x})\|\}$.

The proof is similar to that of Theorem 3.3.

Example 1. *To illustrate Theorem 3.3 and Corollary 3.4, consider the following toy example with $n = 2$, $d = 3$ and $\mathbf{x} \mapsto f(\mathbf{x}) := x_1 x_2^2$. The polynomial g in (3.1) reads $\mathbf{x} \mapsto g(\mathbf{x}) = x_1 x_2^2 - 3 x_1^3 x_2^2 / 5 - 3 x_1 x_2^4 / 5$. Then:*

$$\nabla g(\mathbf{z}) = 0 \Leftrightarrow \begin{cases} z_2^2 (1 - 9 z_1^2 / 5 - 3 z_2^2 / 5) & = 0 \\ z_1 z_2 (2 - 6 z_1^2 / 5 - 12 z_2^2 / 5) & = 0 \end{cases}.$$

So with \tilde{W}_j as in (3.3), we find that g is constant on the four subvarieties

$$\tilde{W}_1 = \{(0, \pm\sqrt{5/3})\} : \quad \tilde{W}_2 = \{(x, 0) : x \in \mathbb{R}\},$$

and

$$\tilde{W}_3 = \frac{1}{\sqrt{3}}\{(1, \pm\sqrt{2})\}; \quad \tilde{W}_4 = \frac{1}{\sqrt{3}}\{(-1, \pm\sqrt{2})\}.$$

with different values on each one of them. Then

$$\mathbb{S}^{n-1} \cap \tilde{W}_1 = \emptyset; \quad \mathbb{S}^{n-1} \cap \tilde{W}_2 = \{(1, 0)\},$$

while

$$\mathbb{S}^{n-1} \cap \tilde{W}_3 = \{(\sqrt{1/3}, \pm\sqrt{2/3})\}; \quad \mathbb{S}^{n-1} \cap \tilde{W}_4 = \{(-\sqrt{1/3}, \pm\sqrt{2/3})\}.$$

Next from $f(\mathbf{x}) = x_1 x_2^2$,

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 0 & 2x_2 \\ 2x_2 & 2x_1 \end{bmatrix} \Rightarrow \lambda_1(\nabla^2 f(\mathbf{x})) = x_1 - \sqrt{x_1^2 + 4x_2^2}.$$

Recall the definition of Ω in (3.4). Then $\Omega \cap \tilde{W}_4 \neq \emptyset$. Indeed $\mathbf{x}^* = (-\sqrt{1/3}, \pm\sqrt{2/3}) \in \Omega$, with $f^* < 0$, because (2.11) holds; indeed:

$$\begin{aligned} \lambda_2(\nabla^2 f(\mathbf{x}^*)) &= x_1^* + \sqrt{(x_1^*)^2 + 4(x_2^*)^2} \\ &= -\sqrt{1/3} + 3\sqrt{1/3} = 2\sqrt{1/3} \geq d f(\mathbf{x}^*) = -2\sqrt{1/3}. \end{aligned}$$

As $j = 4$ is the only index for which $f(\mathbf{x}) < 0$ in \tilde{W}_j , one concludes that f has no spurious negative local minima on \mathbb{S}^{n-1} and $f^* = \frac{-2\sqrt{1/3}}{3}$ is the global minimum.

On the other hand, $\tilde{W}_2 \cap \Omega \neq \emptyset$. Indeed $\mathbf{x}^* = (1, 0) \in \Omega$ with nonnegative value $f^* = 0$ because (2.10) holds, since

$$\lambda_1(\nabla^2 f(\mathbf{x}^*)) = x_1^* - \sqrt{(x_1^*)^2 + 4(x_2^*)^2} = 1 - 1 = 0 \geq df(\mathbf{x}^*) = 0.$$

Finally $\tilde{W}_3 \cap \Omega = \emptyset$ because $\mathbf{x}^* = (\sqrt{1/3}, \pm\sqrt{2/3}) \in \mathbb{S}^{n-1} \cap \tilde{W}_3$ does not satisfy (2.10) (i.e., \mathbf{x}^* is not a **SONC** point), since

$$\lambda_1(\nabla^2 f(\mathbf{x}^*)) = \sqrt{1/3} - 3\sqrt{1/3} = -2\sqrt{1/3} \not\geq df(\mathbf{x}^*) = 2\sqrt{1/3}.$$

4. CONCLUSION

In this paper we have considered homogeneous polynomial optimization on the Euclidean sphere \mathbb{S}^{n-1} and completely characterize all points that satisfy first- and second-order necessary optimality conditions, solely in terms of f , its gradient and the two smallest eigenvalues of its Hessian. Then one may characterize a class of degree- d forms with no spurious local minima, in particular via some decomposition of a related gradient ideal.

While the characterization of all points that satisfy first- and second-order necessary optimality conditions is also valid for twice continuously differentiable positively homogeneous functions, characterizing a class of such functions with no spurious local minima is challenging as we cannot invoke algebraic properties of f any more; indeed the second characterization of no spurious local minima via a certain gradient variety is proper to forms. Another issue for further investigation is the case where f is an arbitrary degree- d polynomial and not a form any more.

Acknowledgement: The author gratefully acknowledges Professor Jiawang Nie (UCSD at San Diego) for fruitful discussions that helped improve the paper.

REFERENCES

- [1] M.R. Bender. Algorithms for sparse polynomial systems: Gröbner bases and resultants, PhD thesis, Sorbonne Université, September 2020.
- [2] D. Bertsekas. *Nonlinear Programming*, 2nd. edition, Athena Scientific, 1995.
- [3] J. Bochnak, M. Coste, M-F. Roy. *Real Algebraic Geometry*, Springer, 1998
- [4] Chi Jin, Rong Ge, P. Netrapalli, S. Kakade, M.I. Jordan. How to escape saddle points efficiently. In *Proceedings of the 34 th International Conference on Machine Learning*, Sydney, Australia, PMLR 70, pp. 1724–1732, 2017.
- [5] D. Cox, J. Little, D. O’Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 3rd. edition, Springer, New York, 1997.
- [6] E. de Kerk, M. Laurent. Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. *Math. Program.*, 2020.
- [7] A. C. Doherty, S. Wehner. Convergence of SDP hierarchies for polynomial optimization on the hypersphere. [arXiv:1210.5048v2](https://arxiv.org/abs/1210.5048v2), 2013.
- [8] K. Fang, H. Fawzi. The sum-of-squares hierarchy on the sphere, and applications in quantum information theory. *Math. Program. Ser. A*, 2020. [arXiv:1908.05155](https://arxiv.org/abs/1908.05155), 2019.

- [9] Rong Ge, J.D. Lee, T. Ma. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems*, pp. 2973–2981, 2016.
- [10] J.B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM J. Optim.* 2(3), pp. 864–885, 2011.
- [11] Jean B. Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge University Press, 2015.
- [12] J. Nie, J. Demmel, B. Sturmfels. Minimizing polynomials via sum of squares over the gradient ideal, *Math. Program. Ser. A* 106, pp. 587–606, 2005.
- [13] J. Nie. Optimality conditions and finite convergence of Lasserre’s hierarchy. *Math. Program. Ser. A* 146, pp. 97–121, 2014.
- [14] B. Sturmfels. *Solving Systems of Polynomial Equations*, CBMS Regional Conference Series in Mathematics, vol 97, AMS, Providence, 2002.

LAAS-CNRS AND INSTITUTE OF MATHEMATICS, UNIVERSITY OF TOULOUSE, LAAS,
7 AVENUE DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE, EMAIL: LASSERRE@LAAS.FR