



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. 2020. hal-02966390v1

HAL Id: hal-02966390

<https://laas.hal.science/hal-02966390v1>

Preprint submitted on 14 Oct 2020 (v1), last revised 4 May 2021 (v2)

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.

HOMOGENEOUS POLYNOMIALS AND SPURIOUS LOCAL MINIMA ON THE UNIT SPHERE

JEAN B. LASSERRE

ABSTRACT. We consider degree- d forms on the Euclidean unit sphere. We specialize to our setting a genericity result by Nie obtained in a more general framework. We exhibit an homogeneous polynomial Res in the coefficients of f , such that if $\text{Res}(f) \neq 0$ then *all* points that satisfy first- and second-order necessary optimality conditions are in fact local minima of f on \mathbb{S}^{n-1} . Then we obtain a simple and compact characterization of all local minima of generic degree- d forms, 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 continuous differentiable functions that are positively homogeneous. Finally we obtain a characterization of generic degree- d forms with no spurious local minimum 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) denotes 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.)

In this paper we provide a characterization of forms which have *no spurious local minimum* on \mathbb{S}^{n-1} , i.e., every local minimum on \mathbb{S}^{n-1} is a global minimum as for convex polynomials. This issue is important as then any algorithm that searches for a local minimum (e.g. first- and/or second-order minimization algorithms) will eventually find the global minimum. To put it differently, the initial (in general hard) non-convex problem (1.1) becomes easy to solve.

Along the way we obtain a characterization of 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. Hence for *generic* forms (in a sense defined later) all local minima are characterized by some property of the spectrum off the Hessian; namely how its first two smallest eigenvalues

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.

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.

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 minimum exist (as e.g. in Ge et al. [9]). See also the discussions in [4, 9] and references therein.

In this paper, by restricting to optimization of forms on the unit Euclidean sphere, we characterize in relatively simple mathematical terms the “no spurious local minimum” situation. This is because for *generic* forms we can characterize all local minima as we show that they coincide with *all* points that satisfy first- and second-order necessary optimality conditions; to do so we specialize to our specific context (1.1) a result by Nie [13] in a much more general context.

We hope that this novel characterization may help in bringing some insights into less restrictive settings.

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} .
- 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.

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

three-fold:

- We first provide the following 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 minimum 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 $\nabla^2 f(\mathbf{x}^*)$.

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} .

- Next, we specialize to our simple setting (1.1) a result of Nie [13] obtained in a more general context of polynomial optimization, where (i) the criterion and constraints are all arbitrary but with bounded degree (fixed), and (ii) the number of constraints is also fixed. In contrast, (1.1) has a single constraint (the sphere constraint) with known associated polynomial $\mathbf{x} \mapsto \|\mathbf{x}\|^2 - 1$, and the criterion is also very specific as it is a degree- d form.

In this setting we obtain a single homogeneous polynomial $\text{Res} \in \mathbb{R}[\mathbf{f}]$ (a certain resultant) where the variables $\mathbf{f} \in \mathbb{R}^{s_0}$ (with $s_0 = \binom{n-1+d}{d}$) are coefficients of forms of degree d (fixed). If $\text{Res}(\mathbf{f}) \neq 0$ then in fact **(FONC)**-**(SONC)** (equivalently, (1.2)) characterize *all* local minima of f on \mathbb{S}^{n-1} . Hence such “nice” forms are “generic” as they are located outside the (real) zero set of a polynomial in the input space of coefficients, i.e., they belong to a Zariski open set.

Then for a generic form, checking whether a point \mathbf{x} is a local minimum 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 the 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 our characterization (1.2) of local minima for generic degree- d forms, and invoking a certain decomposition of gradient ideals already nicely exploited by Nie et al. [12] for unconstrained optimization, we provide a characterization of generic forms with no spurious local minimum on \mathbb{S}^{n-1} .

Notice that the previous genericity result is important as otherwise (1.2) would characterize all local minima only for some small set of “exotic” degree- d forms, then our characterization of those with no spurious local minimum would be even more exotic, and so with little value.

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 the larger set of degree- d forms with no spurious *negative* local minimum 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:

$$(2.1) \quad \text{Res}(f_1, f_2, \dots, f_n) = 0 \Leftrightarrow \exists \mathbf{u} (\neq 0) \in \mathbb{C}^n; f_1(\mathbf{u}) = \dots = f_n(\mathbf{u}) = 0.$$

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:*

$$(2.2) \quad \nabla f(\mathbf{x}^*) + 2\lambda^* \mathbf{x}^* = 0.$$

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

$$(2.3) \quad \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u} + 2\lambda^* \geq 0, \quad \forall \mathbf{u} \in (\mathbf{x}^*)^\perp.$$

(iii) *Conversely, if $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfies (2.2) and the **Second-Order Sufficiency Optimality-Condition (SOSC)***

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

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 \mathbf{x}^* ($\neq 0$) and therefore is linearly independent, i.e., a basic constraint qualification holds true. Therefore (2.2)-(2.3) 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 (2.2), $2\lambda^* = -d f(\mathbf{x}^*)$. In addition, (2.2) holds if and only if*

$$(2.5) \quad \|\nabla f(\mathbf{x}^*)\|^2 = d^2 f(\mathbf{x}^*)^2,$$

and (SONC) reads:

$$(2.6) \quad \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u} \geq d f(\mathbf{x}^*), \quad \forall \mathbf{u} \in (\mathbf{x}^*)^\perp.$$

Proof. In (2.2) we obtain

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

and therefore $\|\nabla f(\mathbf{x}^*)\|^2 = (2\lambda^*)^2 \|\mathbf{x}^*\|^2 = d^2 f(\mathbf{x}^*)^2$. Then (2.6) follows from (2.3). Conversely, assume that (2.5) holds at $\mathbf{x}^* \in \mathbb{S}^{n-1}$. Then

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

that is,

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

and so (2.2) holds with $\lambda^* = -d f(\mathbf{x}^*)/2$, and again (2.6) follows from (2.3). \square

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:*

$$(2.7) \quad \tau(\mathbf{x}^*) := \min_{\mathbf{u} \in (\mathbf{x}^*)^\perp} \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u}.$$

Then

$$(2.8) \quad \lambda_1(\nabla^2 f(\mathbf{x}^*)) = \min [d(d-1) f(\mathbf{x}^*), \tau(\mathbf{x}^*)].$$

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

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

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 (2.8). 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 (2.8), $\lambda_1(\nabla^2 f(\mathbf{x}^*)) = d(d-1) f(\mathbf{x}^*)$. \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:*

$$(2.9) \quad \lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*) \quad \text{if } f(\mathbf{x}^*) \geq 0$$

$$(2.10) \quad \lambda_2(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*) \quad \text{if } f(\mathbf{x}^*) < 0.$$

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 $\lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*)$ 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 (2.9) 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)**. (2.9) \Rightarrow **(SONC)**. Again by Lemma 2.3, $\tau(\mathbf{x}^*) \geq \lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*)$, **(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 $\lambda_k = 1$ 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 .

Remark 2.5. *It is worth noticing that the characterization of **(FONC)** in (2.5) 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 (2.2) (or equivalently (2.5)) 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 on

\mathcal{E}_n are non positive and are local minima on \mathbb{S}^{n-1} ; hence no local maximum on \mathbb{S}^{n-1} can be negative.

So if f can take negative values then it is definitely better and easier (as \mathcal{E}_n is a convex set) to minimize on \mathcal{E}_n . 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 $f^* \leq 0$, and it is attained at some $\mathbf{x}^* \in \mathbb{S}^{n-1}$; therefore f^* is also a local minimum 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 $\mathbf{x}^* \in \mathcal{E}_n$ is a local minimizer with $\|\mathbf{x}^*\| < 1$. Then $\mathbf{z}^* := \lambda \mathbf{x}^* \in \mathbb{S}^{n-1}$ for some $\lambda > 1$, and $f(\mathbf{z}^*) = \lambda^d f(\mathbf{x}^*) \leq f(\mathbf{x}^*) = f^* \leq 0$ and therefore, necessarily $f^* = 0$ and $\mathbf{z} \in \mathbb{S}^{n-1}$ is also a (local) minimizer. If $f^* < 0$ then necessarily $\mathbf{x}^* \in \mathbb{S}^{n-1}$.

(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 Lemma 2.6 states that 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 . In addition, notice that there may be no spurious negative local minimum on \mathbb{S}^{n-1} (hence no spurious local minimum 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 converging to a Karush-Kuhn-Tucker point (i.e. a point that satisfies **(FONC)**) will find a 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 MINIMUM ON \mathbb{S}^{n-1}

In this section we provide a necessary and sufficient condition for a generic degree- d form to have *no* spurious local minimum on \mathbb{S}^{n-1} . As a preliminary result we show that for a generic degree- d form, every point that satisfies **(FONC)** and **(SONC)**, also satisfies **(SOSC)**. Therefore **(FONC)**-**(SONC)** completely characterizes *all* local minimizers of (1.1).

3.1. Genericity of (SOSC). Let $d \in \mathbb{N}$ be fixed and consider polynomial optimization problems (POPs) of the form

$$(3.1) \quad \mathbf{P} : \quad \inf_{\mathbf{x}} \{ p_0(\mathbf{x}) : p_1(\mathbf{x}) = 0 \},$$

where $p_0 \in \mathbb{R}[\mathbf{x}]_d$ and $p_1 \in \mathbb{R}[\mathbf{x}]_2$ with respective coefficient vector $\mathbf{p}_0 \in \mathbb{R}^{s_0}$ and $\mathbf{p}_1 \in \mathbb{R}^{s_1}$, with $s_0 = \binom{n+d}{n}$ and $s_1 = \binom{n+2}{n}$. Then we may and will identify the input data $(\mathbf{p}_0, \mathbf{p}_1)$ of \mathbf{P} with a point in $\mathbb{R}^{s_0+s_1}$. The following result is from Nie [13], adapted to problem (3.1).

Proposition 3.1. (Nie [13]) *There exist an integer m and polynomials $\varphi_1, \dots, \varphi_m \in \mathbb{R}[\mathbf{p}_0, \mathbf{p}_1]$ such that if $\varphi_j(\mathbf{p}_0, \mathbf{p}_1) \neq 0$ for all $j = 1, \dots, m$, then (SOSC) holds at every point \mathbf{x}^* of \mathbf{P} that satisfies (FONC)-(SONC).*

Said differently, “generically” every local minimizer of (3.1) satisfies (SOSC) as the “good” input data of couples of polynomials (p_0, p_1) in Proposition 3.1, lie in a Zariski open set of $\mathbb{R}^{s_0+s_1}$.

Remark 3.2. *Proposition 3.1 is very nice as it provides a purely algebraic characterization of the locus of (potentially) “bad polynomials”, and moreover this locus is “negligible” in terms of its Lebesgue volume in the input space.*

However Proposition 3.1 is mainly a result of theoretical nature. Indeed even though the (φ_j) ’s can be obtained in principle, their computation requires computing discriminant and resultants, and so far is limited to very modest dimensions. However some recent progress described in Bender [1] suggest that in some cases where structure can be exploited, problems of larger dimension can be addressed.

Problem (1.1) is an instance of a POP (3.1) with $p_0^* = f$ and $p_1^* := 1 - \|\mathbf{x}\|^2$, and with respective vector of coefficients \mathbf{p}_0^* and \mathbf{p}_1^* . Therefore if $\varphi_j(\mathbf{p}_0^*, \mathbf{p}_1^*) \neq 0$ for all $j = 1, \dots, m$, then (SOSC) holds at every local minimizer of f on \mathbb{S}^{n-1} . However, observe that

- p_1^* is a specific degree-2 polynomial, and
- f is a degree- d form and so also a very specific polynomial as all its coefficients associated with monomials of degree strictly less than d , vanish.

Therefore it is not clear whether $\varphi_j(\mathbf{p}_0^*, \mathbf{p}_1^*) \neq 0$ for all $j = 1, \dots, m$, for a generic degree- d form $p_0 \in \mathbb{R}[\mathbf{x}]$. However in the framework of Problem (1.1) one can provide further insights. Namely it suffices to prove that for a generic form of degree d , (2.4) holds whenever \mathbf{x}^* satisfies (FONC) and (SONC) in Corollary 2.2.

Lemma 3.3. *Let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfy (FONC) and (SONC), and let*

$$(3.2) \quad \mathbf{x} \mapsto \mathcal{D}(\mathbf{x}) := \det \begin{bmatrix} \nabla^2 f(\mathbf{x}) - d f(\mathbf{x}) \mathbf{I} & \mathbf{x} \\ \mathbf{x}^T & 0 \end{bmatrix} \in \mathbb{R}[\mathbf{x}].$$

Then (SOSC) $\Leftrightarrow \mathcal{D}(\mathbf{x}^) \neq 0$.*

Proof. Observe that $\mathcal{D}(\mathbf{x}^*) = 0$ if and only if the matrix in (3.2) has at least one zero eigenvalue. Equivalently, if and only if there exists a vector $\mathbf{u}^* \in (\mathbf{x}^*)^\perp$ and scalars γ, θ, λ with $\gamma^2 + \theta^2 + \lambda^2 = 1$, such that

$$\begin{bmatrix} \nabla^2 f(\mathbf{x}^*) - df(\mathbf{x}^*)\mathbf{I} & \mathbf{x}^* \\ (\mathbf{x}^*)^T & 0 \end{bmatrix} \begin{bmatrix} \gamma \mathbf{u}^* + \theta \mathbf{x}^* \\ \lambda \end{bmatrix} = 0,$$

that is,

$$\begin{aligned} (\nabla^2 f(\mathbf{x}^*) - df(\mathbf{x}^*)\mathbf{I})(\gamma \mathbf{u}^* + \theta \mathbf{x}^*) + \lambda \mathbf{x}^* &= 0 \\ \theta \|\mathbf{x}^*\|^2 &= 0, \end{aligned}$$

so that $\theta = 0$. Next, multiplying the first equation by $\gamma \mathbf{u}^*$ yields

$$\gamma^2 (\langle \mathbf{u}^*, \nabla^2 f(\mathbf{x}^*) \mathbf{u}^* \rangle - df(\mathbf{x}^*)) = 0.$$

The case $\gamma = 0$ (combined with $\theta = 0$) yields $\lambda^2 = 1$ and $(\mathbf{x}^*, 0) = (0, 0)$, which is not possible. Therefore $\gamma \neq 0$ and $\langle \mathbf{u}^*, \nabla^2 f(\mathbf{x}^*) \mathbf{u}^* \rangle = df(\mathbf{x}^*)$, which shows that $\mathcal{D}(\mathbf{x}^*) = 0$ if and only if **(SOSC)** does not hold. \square

Next, consider the system of $(n+1)$ polynomial equations:

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial x_i} - df(\mathbf{x})x_i &= 0, \quad i = 1, \dots, n \\ \mathcal{D}(\mathbf{x}) &= 0. \end{aligned}$$

Its homogenization is a system of $(n+1)$ forms $q_i \in \mathbb{R}[x_0, \mathbf{x}]$, which reads:

$$(3.3) \quad q_i(x_0, \mathbf{x}) = 0, \quad i = 1, \dots, n$$

$$(3.4) \quad q_0(x_0, \mathbf{x}) (:= \tilde{\mathcal{D}}(x_0, \mathbf{x})) = 0,$$

with associated resultant $\text{Res}(q_0, q_1, \dots, q_n)$. By (2.1) :

$$\text{Res}(q_0, q_1, \dots, q_n) = 0 \Leftrightarrow \exists \mathbf{u} \in \mathbb{C}^{n+1} (\mathbf{u} \neq 0); q_0(\mathbf{u}) = \dots = q_n(\mathbf{u}) = 0.$$

Moreover, $\text{Res}(q_0, q_1, \dots, q_n)$ is an homogeneous polynomial in the coefficients of f , that is, letting $\mathbf{f} \in \mathbb{R}^{s_0}$ be the coefficient vector of $f \in \mathbb{R}[\mathbf{x}]_d$,

$$(3.5) \quad \text{Res}(q_0, q_1, \dots, q_n) =: Q(f) \in \mathbb{R}[\mathbf{f}].$$

Theorem 3.4. *Consider problem (1.1) and let $Q(f)$ be as in (3.5). If a degree- d form satisfies $Q(f) \neq 0$ then every point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ which satisfies **(FONC)** and **(SONC)** also satisfies **(SOSC)** and is a local minimizer of f on \mathbb{S}^{n-1} .*

Hence this last property is true for a generic degree- d form and so $\mathbf{x}^ \in \mathbb{S}^{n-1}$ is a local minimum of a generic degree- d form f if and only if $\nabla f(\mathbf{x}^*) = df(\mathbf{x}^*)\mathbf{x}^*$ and*

$$(3.6) \quad \begin{cases} \lambda_1(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*) & \text{if } f(\mathbf{x}^*) \geq 0 \\ \lambda_2(\nabla^2 f(\mathbf{x}^*)) \geq df(\mathbf{x}^*) & \text{if } f(\mathbf{x}^*) < 0. \end{cases}$$

Proof. The vector of coefficients of f is an element $\mathbf{f} \in \mathbb{R}^{s_0}$ (recall that $s_0 = \binom{n-1+d}{d}$). Let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfy **(FONC)** and **(SONC)**. Then (3.3) holds at $0 \neq \mathbf{u} := (1, \mathbf{x}^*) \in \mathbb{R}^{n+1}$. But then (3.4) cannot hold otherwise $Q(f) = 0$. Hence, $\mathcal{D}(\mathbf{x}^*) \neq 0$ and by Lemma 3.3, **(SOSC)** holds at \mathbf{x}^* .

Finally, the set $\{\mathbf{f} \in \mathbb{R}^{s_0} : Q(\mathbf{f})\} = 0$ is a set of zero Lebesgue measure in the input space of coefficients, and therefore the property is true for a generic degree- d form f . Finally (3.6) follows from Corollary 2.4 since all local minima of a generic degree- d form satisfy **(FONC)**-**(SONC)**. \square

3.2. Forms with no spurious local minimum on \mathbb{S}^{n-1} . We are now in position to characterize all generic degree- d forms that have no spurious local minimum on \mathbb{S}^{n-1} . The following result is a direct consequence of Theorem 3.4

Corollary 3.5. *Let f be a degree- d ($d > 2$) generic form, i.e. such that $Q(f) \neq 0$ with Q as in (3.5), and let*

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

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 minimum on \mathbb{S}^{n-1} if and only if f is constant on the set*

$$\{\mathbf{x} \in \Theta : \lambda_1(\nabla^2 f(\mathbf{x})) \geq \|\nabla f(\mathbf{x})\|\}.$$

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

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

Proof. By Corollary 2.2, $\|\nabla f(\mathbf{x})\| = d|f(\mathbf{x})|$ on Θ . Then we just apply Theorem 3.4 when there is only one local (hence global) minimum. \square

We next show that the characterization in Corollary 3.5 is also related to property of gradient ideals of $\mathbb{R}[\mathbf{x}]$. We introduce a polynomial 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 points that satisfy **(FONC)** for 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

$$(3.7) \quad \mathbf{x} \mapsto g(\mathbf{x}) := f(\mathbf{x}) \left(1 - \frac{d}{d+2} \|\mathbf{x}\|^2\right), \quad \mathbf{x} \in \mathbb{R}^n.$$

Proposition 3.6. *Let f be a degree- d form and let $g \in \mathbb{R}[\mathbf{x}]$ be as in (3.7). Then on \mathbb{S}^{n-1} :*

$$(3.8) \quad \nabla g(\mathbf{x}) = 0 \Leftrightarrow \nabla f(\mathbf{x}) = d f(\mathbf{x}) \cdot \mathbf{x} \Leftrightarrow \|\nabla f(\mathbf{x})\|^2 = (d f(\mathbf{x}))^2,$$

That is, all critical points $\mathbf{x}^* \in \mathbb{S}^{n-1}$ of g satisfy **(FONC)** for f , and conversely, all points $\mathbf{x}^* \in \mathbb{S}^{n-1}$ 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 (2.2) holds with $2\lambda^* = -d f(\mathbf{x}^*)$, which yields the desired result (3.8). \square

Moreover, on \mathbb{S}^{n-1} minimizing f is strictly equivalent to minimizing g since on \mathbb{S}^{n-1} , $g(\mathbf{x}) = 2f(\mathbf{x})/(d+2)$. Next, with g as in (3.7), 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

$$(3.9) \quad V_{\text{grad}}(g) = W_0 \cup \tilde{W}_1 \dots \cup \tilde{W}_r,$$

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 degree- d forms f with no spurious local minimum on \mathbb{S}^{n-1} .

Theorem 3.7. *Consider problem (1.1) where f is a degree- d form ($d > 2$) that satisfies $Q(f) \neq 0$ (with Q as in (3.5)). Then f has no spurious local minimum on \mathbb{S}^{n-1} if and only if there is only one index j^* in (3.9) such that*

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

where

$$(3.10) \quad \begin{aligned} \Omega &:= \{ \mathbf{x} \in \mathbb{S}^{n-1} : f(\mathbf{x}) < 0; \lambda_2(\nabla^2 f(\mathbf{x})) \geq -\|\nabla f(\mathbf{x})\| \} \\ &\cup \{ \mathbf{x} \in \mathbb{S}^{n-1} : f(\mathbf{x}) \geq 0; \lambda_1(\nabla^2 f(\mathbf{x})) \geq \|\nabla f(\mathbf{x})\| \}. \end{aligned}$$

Proof. Under the assumption of the theorem and (3.8), f has no spurious local minimum on \mathbb{S}^{n-1} if and only if f is constant (hence g is constant) on $V_{\text{grad}}(g) \cap \Omega$ with Ω as in (3.10). In view of the decomposition (3.9) and its properties, one obtains the statement of the theorem. \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 those degree- d forms with the less restrictive condition of no spurious *negative* local minimum on \mathcal{E}_n (hence on \mathbb{S}^{n-1}).

Corollary 3.8. *Consider problem (1.1) where f is a degree- d form ($d > 2$) that satisfies $Q(f) \neq 0$ (with Q as in (3.5)). Then f has no spurious negative local minimum on \mathbb{S}^{n-1} if and only if there is only one index j^* in (3.9) such that*

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

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

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. This characterization is valid for a generic degree- d form f where “genericity” is fully characterized in terms of the resultant of a single system of polynomial equations. Then for those generic forms f , the absence of spurious local minimum can be characterized, 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 positively homogeneous functions, characterizing the absence of spurious minimum is challenging as we cannot invoke any more algebraic properties of f and the genericity property needs to be explicitated. Also, the second characterization of no spurious minimum via the 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