# OPTIMIZATION ON THE EUCLIDEAN UNIT SPHERE

Jean-Bernard Lasserre

# OPTIMIZATION ON THE EUCLIDEAN UNIT SPHERE

JEAN B. LASSERRE

Abstract. We consider the problem of minimizing a continuously differentiable function $f$ of $m$ linear forms in $n$ variables on the Euclidean unit sphere. We show that this problem is equivalent to minimizing the same function of related $m$ linear forms (but now in $m$ variables) on the Euclidean *unit ball*. When the linear forms are known, this results in a drastic reduction in problem size whenever $m \ll n$ and allows to solve potentially large scale non-convex such problems. We also provide a test to detect when a polynomial is a polynomial in a fixed number of forms. Finally, we identify two classes of functions with no spurious local minima on the sphere: (i) quasi-convex polynomials of odd degree and (ii) nonnegative and homogeneous functions. Finally, odd degree-$d$ forms have only nonpositive local minima and at most $(d-1)^m$ are strictly negative.

## 1. Introduction

Optimization on the unit sphere is a fascinating topic. Indeed and despite it simple formulation (a single quadratic equality constraint), it it has important applications, in particular when searching for the global minimum. For instance, and as discussed in e.g. [6]:

- 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 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].

Moreover, with the celebrated MAXCUT problem in combinatorial optimization (see .e.g. [5]), it is one of the simplest NP-hard problems (a single quadratic equality constraint) and therefore serves as an important case of study to understand the efficiency of LP- and SDP-relaxations for solving non-convex polynomial optimization problems. In particular, recent

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

progress has been reported for convergence rates of the Moment-SOS hierarchy of lower bounds and the (different) SOS-hierarchy of upper bounds introduced in [10] and [12] respectively. Namely, Fang and Fawzi [8] (see also Doherty and Wehner [7]) for the lower bound hierarchy and de Klerk et al. and Laurent [6] for the upper bound hierarchy, have proved an interesting (and perhaps surprising) $O(1/r^2)$ rate of convergence if $r$ denotes step-$r$ of the hierarchy. The interested reader is referred to [8, 6] as well as Slot and Laurent [17] and the references therein for a detailed discussion on this topic. Recently, in the homogeneous case, the author has provided in [9] a complete characterization of first- and second-order necessary optimality conditions solely in terms of the first two smallest eigenvalues of the Hessian of the form to minimize.

In this paper we consider optimization on the unit sphere for the special class of functions of $m$ linear forms in $n$ variables, i.e., with $\mathbb{S}^{n-1}$ denoting the Euclidean unit sphere of $\mathbb{R}^n$,

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

$$= \min_{\mathbf{x}} \{f((\boldsymbol{\ell}_1 \cdot \mathbf{x}), \ldots, (\boldsymbol{\ell}_m \cdot \mathbf{x})) : \mathbf{x} \in \mathbb{S}^{n-1}\}, \tag{1.2}$$

for some continuously differentiable function $f : \mathbb{R}^m \to \mathbb{R}$ and some linear forms $\boldsymbol{\ell}_j : \mathbb{R}^n \to \mathbb{R}$, $j \in \{1, \ldots, m\}$ (where "$\boldsymbol{\ell}_j \cdot \mathbf{x}$" denotes the usual scalar product of vectors $\boldsymbol{\ell}_j$ and $\mathbf{x}$). As every optimization problem on $\mathbb{S}^{n-1}$ can be put in the form (1.2) with $m = n$, then clearly only the case $m < n$ is interesting. When $m \ll n$, Problem (1.2) can be thought of satisfying a sort of "sparsity", as the objective function is expressed as a function of *only a few* linear forms. In cite [15] a function of a few linear forms is called a *low-rank* function. This sparsity is quite different from the structured- or term-sparsity exploited in e.g. [11, 19, 20] (and references therein) in a more general context.

**Contribution.** In this paper we show that this sparsity can be exploited as indeed Problem (1.2) is completely equivalent to an optimization problem of the same form, but in $m$ variables only. More precisely, let $\mathscr{E}_m$ be the Euclidean ball $\{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x}\| \leq 1\}$. We show that Problem (1.2) reduces to solving the polynomial problem:

$$f^* = \min_{\mathbf{y}} \{f((\mathbf{L}_1 \cdot \mathbf{y}), \ldots, (\mathbf{L}_m \cdot \mathbf{y})) : \mathbf{y} \in \mathscr{E}_m\}, \tag{1.3}$$

where $\mathbf{L}_j$ is the $j$th-row of the matrix $\mathcal{L}^{1/2} \in \mathbb{R}^{m \times m}$, with $\mathcal{L} = (\boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_j)_{i,j \in [m]}$.

Problem (1.3) is equivalent to (1.2) in the sense that :

• To any critical point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ of (1.2) with $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$ (in particular, any local minimizer) is associated a critical point $\mathbf{y}^* \in \mathscr{E}_m$ of (1.3) with same value.

• The converse also holds. That is, from a critical point $\mathbf{y}^* \in \mathscr{E}_m$ of (1.3) one may easily obtain a critical point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ with same value.

• Moreover if $\mathbf{x}^*$ satisfies the second-order necessary optimality conditions **(SONC)** (resp. the second-order sufficient optimality conditions) then necessarily $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$, and so does the associated critical $\mathbf{y}^* \in \mathscr{E}_m$ for Problem (1.3). The converse also holds for all critical points $\mathbf{y}^* \in \mathscr{E}_m$.

• As a consequence (apparently unnoticed, at least to the best of our knowledge), if $m = 1$ then solving Problem (1.2) is straightforward as it reduces to minimizing the univariate polynomial $y \mapsto f(y\,\|\boldsymbol{\ell}\|)$ on the interval $[-1, 1]$, which can be done efficiently. Interestingly, this restrictive case $m = 1$ applies if $h$ in (1.2) is an odd-degree quasi-convex polynomial[1]. So if $h$ is quasi-convex of odd-degree then it has no spurious local minima on the sphere and $f^* = \min[f(-|\boldsymbol{\ell}\|)\,,\, f(\|\boldsymbol{\ell}\|)]$.

Hence $\mathbb{R}\mathbf{x}]_d$ is the union $\bigcup_{j=1}^n \mathscr{P}_j$ of disjoint subsets $(\mathscr{P}_j)_{j\leq n}$, where each member of $\mathscr{P}_j$ can be written as a degree-$d$ polynomial of exactly $j$ linear forms, and not less. For the smallest subset $\mathscr{P}_1$, minimization on $\mathbb{S}^{n-1}$ is quite easy, and when $d$ is odd $\mathscr{P}_1$ contains in particular all quasi-convex degree-$d$ polynomials. As $j$ increases the resulting equivalent problem (1.3) on $\mathscr{E}_j \subset \mathbb{R}^j$ is still easier than (1.1), while it brings nothing for $\mathscr{P}_n$.

• When in Problem (1.2) $f$ is *hidden*, that is when Problem (1.2) is only defined through $h$, one provides a simple numerical procedure based on sufficiently many evaluations of $h$ at arbitrary points (not necessarily on $\mathbb{S}^{n-1}$), which allows to reveal a formulation (1.2) for some $f_h$.

• It is rather straightforward to see that the only interesting local minima of (1.2) are negative. Indeed any $\mathbf{x} \in \mathbb{S}^{n-1}$ with $\boldsymbol{\ell}_j \cdot \mathbf{x} = 0$ for all $j$ (guaranteed to exist whenever $m < n$) satisfies $h(\mathbf{x}) = f(0) = 0$ and therefore as 0 is attained trivially, for minimization purposes we are mainly interested in *negative* local minima.

• If $h$ is nonnegative and the formulation (1.2) is known, then in view of the preceding remark, $f^* = 0$ is the global minimum. But when only $h$ is known then the hidden formulation (1.2) identifies a class of functions with no spurious local minima on the sphere. Any optimization algorithm converging to a local minimum of $h$ on $\mathbb{S}^{n-1}$ then converges to the global minimum $f^* = 0$.

• At last we consider the important case alluded to in the introduction, where $h$ in (1.1) is a positively homogeneous function (and a polynomial in particular). For homogeneous polynomials of the form (1.2), the recent characterization of **(SONC)** points by the author [9] in terms of the first- and second smallest eigenvalues of the Hessian, directly translates into a similar but more specific characterization for Problem (1.2) and Problem (1.3). Also when $f$ is an odd degree-$d$ form, all local minima of (1.3) are

[1]The class of quasi-convex polynomials include pseudo-convex and convex polynomials. An odd-degree polynomial $h$ is quasi-convex if and only if $h(\mathbf{x}) = f(\boldsymbol{\ell} \cdot \mathbf{x})$ for some $\boldsymbol{\ell} \in \mathbb{R}^n$ and some *monotonic* univariate polynomial $f$; see for instance Ahmadi et al. [1] and [13, Theorem 13.11]

necessarily nonpositive and there are at most $(d-1)^m$ negative local minima, independently of the dimension $n$.

When $m \ll n$, Problem (1.3) is easier to handle as it is an optimization problem (i) in a space of much smaller dimension and (ii), on the Euclidean unit ball $\mathscr{E}_m$ (a nice convex set) instead of $\mathbb{S}^{n-1}$. In particular, if $f$ is a polynomial or a rational function and $m$ is relatively small, then one may apply the Moment-SOS hierarchy described in e.g. [13, 14] with a good chance of obtaining the global minimum in a few steps, which could be impossible when treating the initial problem (1.2) in $\mathbb{R}^n$.

It is worth noticing that polynomials of linear forms are also interesting for *integration* of polynomials $f((\boldsymbol{\ell}_1\cdot\mathbf{x}),\ldots,(\boldsymbol{\ell}_m\cdot\mathbf{x}))$ of linear forms $(\boldsymbol{\ell}_j)$, on the simplex. Indeed Baldoni et al. [2] have shown that the computational complexity essentially depends on (i) the number of forms, (ii) the degree of $f$, and *not* on the dimension $n$, like in our setting. Finally notice that every degree-$d$ homogeneous polynomial $h$ admits a decomposition into a (signed) sum of $d$-powers of linear forms, i.e.,

$$\mathbf{x}\mapsto h(\mathbf{x}) \,=\, \sum_{j=1}^{s} \varepsilon_j\,(\boldsymbol{\ell}_j\cdot\mathbf{x})^d\,,\quad \forall \mathbf{x}\in\mathbb{R}^n\,,$$

for some $(\boldsymbol{\ell}_j)\subset\mathbb{R}^n$ and $\epsilon_j\in\{-1,1\}$. That is $f$ in (1.2) is even more specific as it reads $f(\mathbf{X}) = \sum_{j=1}^s \varepsilon_j\,\mathbf{X}_j^d$, and therefore all results of this paper are valid for the class of homogeneous polynomials. However if $h$ is not directly defined as a sum of powers of linear forms (or a polynomial of forms as in (1.2)), then getting the $\boldsymbol{\ell}_j$'s in the above decomposition is an NP-hard problem.

## 2. Main

2.1. **Notation, definitions and preliminaries.** Let $[m]$ denote the set of integers $\{1,2,\ldots,m\}$, $\mathbb{R}[\mathbf{x}]$ the ring of real polynomials in the variables $\mathbf{x}=(x_1,\ldots,x_n)$, and $\mathbb{R}[\mathbf{x}]_d$ its subset of polynomials of total degree at most $d$. For ease of notation the usual scalar product between two vectors $\mathbf{x},\mathbf{y}$ is denoted $\mathbf{x}\cdot\mathbf{y}$. For a real symmetric matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$, denote by $\lambda_1(\mathbf{A})\leq\lambda_2(\mathbf{A}),\ldots\leq\lambda_n(\mathbf{A})$, its eigenvalues arranged in increasing order.

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

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

As a constant term has no influence on the minimization of $h$ in (1.1) we may and will assume that $h(0)=0$ (and so $f(0)=0$ as well).

**Assumption 2.1.** *The family* $(\boldsymbol{\ell}_i)_{i\in[m]}\subset\mathbb{R}^n$ *in* (1.2) *is linearly independent.*

Let $\boldsymbol{\ell}\in\mathbb{R}^{m\times n}$ be the matrix with $i$-th row $\boldsymbol{\ell}_j$, $j\in[m]$. Given $\mathbf{x}\in\mathbb{R}^n$, let $\mathbf{X}:=\boldsymbol{\ell}\,\mathbf{x}$, and introduce the real symmetric psd matrix $\mathcal{L}=\mathcal{L}^T\in\mathbb{R}^{m\times m}$

defined by $\mathcal{L} = \boldsymbol{\ell}\boldsymbol{\ell}^T$, i.e.,

$$\mathcal{L}_{ij} := \boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_j\,, \quad i,j \in [m]\,. \tag{2.1}$$

By Assumption 2.1, $\mathcal{L}$ is positive definite (denoted $\mathcal{L} \succ 0$). Therefore we may and will introduce the well-defined symmetric matrix[2] $\mathbf{L} := \mathcal{L}^{1/2} \succ 0$, and let $\mathbf{L}_i$ denote its $i$-th row.

2.2. **First-order optimality conditions.** Given $\mathbf{y} \in \mathbb{R}^m$ let $\mathbf{Y} := \mathbf{L}\,\mathbf{y}$ and introduce the function $\tilde{f} : \mathbb{R}^m \to \mathbb{R}$, defined by:

$$\mathbf{y} \mapsto \tilde{f}(\mathbf{y}) \,:=\, f(\mathbf{Y}) \,=\, f((\mathbf{L}_1 \cdot \mathbf{y}), \ldots, (\mathbf{L}_m \cdot \mathbf{y}))\,. \tag{2.2}$$

So $\tilde{f}$ inherits basic properties of $f$, that is, $\tilde{f}$ is also continuously differentiable and if $f$ is a degree-$d$ polynomial then so does $\tilde{f}$. Similarly, if $f$ is positively homogeneous of degree $d$, (i.e., $f(\lambda\mathbf{X}) = \lambda^d f(\mathbf{X})$ for all $\lambda > 0$ and all $\mathbf{X} \in \mathbb{R}^m$) then so does $\tilde{f}$.

**Lemma 2.2.** *Let Assumption 2.1 hold. Let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ and $\mathbf{X}^* = \boldsymbol{\ell}\,\mathbf{x} \in \mathbb{R}^m$. The first-order necessary optimality conditions* **(FONC)** *at for problem* (1.2) *reads :*

$$\nabla f(\mathbf{X}^*) \;=\; 0 \quad \textit{or} \tag{2.3}$$

$$\boldsymbol{\ell}^T\,\nabla f(\mathbf{X}^*) \;=\; 2\lambda^*\,\mathbf{x}^*\,, \quad \textit{with } 0 \neq \lambda^* = \mathbf{X}^* \cdot \nabla f(\mathbf{X}^*)/2. \tag{2.4}$$

*Moreover if $f$ is positively homogeneous of degree $d$ then* (2.3)-(2.4) *reads:*

$$\boldsymbol{\ell}^T\,\nabla f(\mathbf{X}^*), = \; d\,f(\mathbf{X}^*)\,\mathbf{x}^*\,. \tag{2.5}$$

*Proof.* Notice that the standard constraint qualification of linear independent of active constraints at $\mathbf{x}^*$ trivially holds. Therefore **(FONC)** reads

$$\sum_{j=1}^m \frac{\partial f(\mathbf{X}^*)}{\partial X_j}\,\boldsymbol{\ell}_j \;=\; 2\lambda^*\,\mathbf{x}^*\,,$$

for some $\lambda^*$. If $\lambda^* = 0$ then in view of Assumption 2.1 we conclude that $\nabla f(\mathbf{X}^*) = 0$, i.e., (2.3) holds. If $\lambda^* \neq 0$ then the first statement of (2.4) holds while the second statement in (2.3) follows by multiplying both sides of the equality by $\mathbf{x}^*$.

Finally, when $f$ is positively homogeneous of degree $d$ we then exploit Euler's identity $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) = d\,f(\mathbf{X}^*)$. If (2.3) holds then $f(\mathbf{X}^*) = 0$ and therefore (2.3) implies (2.5). If (2.4) holds then again $2\lambda^* = df(\mathbf{X}^*)$, i.e., (2.5) holds. Conversely suppose that (2.5) holds. If $f(\mathbf{X}^*) \neq 0$ then (2.4) holds with $2\lambda^* = df(\mathbf{X}^*)$. If $f(\mathbf{X}^*) = 0$ then necessarily under Assumption 2.1, $\nabla f(\mathbf{X}^*) = 0$, i.e., (2.3) holds. □

[2]Writing the singular value decomposition $\mathcal{L} = \mathbf{P}^T\Lambda\mathbf{P}$ with $\Lambda = \mathrm{diag}((\lambda_i)_{i\in[m]})$, $\mathcal{L}^{1/2}$ is obtained as $\mathbf{P}^T\Lambda^{1/2}\mathbf{P}$ with $\Lambda^{1/2} := \mathrm{diag}((\sqrt{\lambda_i})_{i\in[m]})$.

**Remark 2.3.** *(i) Let* $0 \neq \mathbf{x} \in \mathrm{Ker}(\boldsymbol{\ell})$ *so that* $\mathbf{x}^* := \mathbf{x}/\|\mathbf{x}\| \in \mathbb{S}^{n-1}$ *with* $f(\mathbf{X}^*) = f(0) = h(0) = 0$. *If* $\mathbf{x}^*$ *satisfies* **(FONC)** *then necessarily* $\lambda^* = 0$ *and* $\nabla f(\mathbf{X}^*) = \nabla f(0) = 0$ *as well. Indeed multiplying both sides of* (2.4) *by* $\boldsymbol{\ell}$ *yields* $\boldsymbol{\ell}\boldsymbol{\ell}^T\nabla f(0) = 2\lambda^*\boldsymbol{\ell}\,\mathbf{x}^* = 0$, *and as* $\mathcal{L}$ *is non singular,* $\nabla f(\mathbf{X}^*) = \nabla f(0) = 0$. *But then again by* (2.4), $\lambda^* = 0$.

*(ii) The only interesting local minima are strictly negative because by (i) any* $\mathbf{x} \in \mathrm{Ker}(\boldsymbol{\ell})$ *provides a local minimizer* $\mathbf{x}^* := \mathbf{x}/\|\mathbf{x}\| \in \mathbb{S}^{n-1}$ *with value* $h(\mathbf{x}^*) = f(\mathbf{X}^*) = f(0) = 0$.

Next, consider the new optimization problem defined in (1.3).

**Lemma 2.4.** *Let Assumption 2.1 hold. Let* $\mathbf{y}^* \in \mathscr{E}_m$ *and* $\mathbf{Y}^* := \mathbf{L}\,\mathbf{y}^*$. *The first-order necessary optimality conditions* **(FONC)** *at* $\mathbf{y}^*$ *for problem* (1.3) *reads :*

$$\begin{aligned} &(2.6) & \nabla f(\mathbf{Y}^*) &= 0 \quad \textit{or} \\ &(2.7) & \mathbf{L}\,\nabla f(\mathbf{Y}^*) &= -2\theta^*\,\mathbf{y}^*, \quad \textit{with} \left\{ \begin{array}{l} 0 < \theta^* = -\mathbf{Y}^*\cdot\nabla f(\mathbf{Y}^*)/2\,, \\ \textit{and } \mathbf{y}^* \in \mathbb{S}^{m-1}\,. \end{array}\right. \end{aligned}$$

*Moreover if* $f$ *is positively homogeneous of degree* $d$ *then* (2.6)-(2.7) *reads:*

$$\mathbf{L}\,\nabla f(\mathbf{Y}^*) \,=\, -d\,f(\mathbf{Y}^*)\,\mathbf{y}^*\,. \tag{2.8}$$

*Proof.* The Fritz-John necessary optimality conditions [4] at $\mathbf{y}^* \in \mathscr{E}_m$ read:

$$\theta_0^*\,\mathbf{L}\,\nabla f(\mathbf{Y}^*) + 2\,\theta^*\,\mathbf{y}^* \,=\, 0\,; \quad \lambda_1^*\,(1 - \|\mathbf{y}^*\|^2) \,=\, 0\,,$$

for some nonnegative couple $0 \neq (\theta_0^*, \theta^*) \in \mathbb{R}^2_+$. The case $\theta_0^* = 0$ is not possible since as $\theta^* \neq 0$, it implies $\mathbf{y}^* = 0$ and $\|\mathbf{y}^*\|^2 = 1$. Therefore one obtains the first-order Karush-Kuhn-Tucker (KKT) optimality conditions **(FONC)**

$$\mathbf{L}\,\nabla f(\mathbf{Y}^*) + 2\theta^*\,\mathbf{y}^* \,=\, 0\,; \quad \theta^*(1 - \|\mathbf{y}^*\|^2) \,= 0\,,$$

for some $\theta^* \geq 0$. Suppose that $\theta^* = 0$. In view of Assumption 2.1 we have seen that the matrix $\mathcal{L} \succeq 0$ in (2.1) is non singular and so is $\mathbf{L} = \mathcal{L}^{1/2}$. Therefore $\nabla f(\mathbf{Y}^*) = 0$, which yields (2.6).

Next suppose that $\theta^* > 0$ so that necessarily $\|\mathbf{y}^*\|^2 = 1$ and then $\nabla f(\mathbf{Y}^*) \neq 0$. Multiplying by $\mathbf{y}^*$ yields

$$0 \,<\, 2\,\theta^* \,=\, -\mathbf{Y}^*\cdot\nabla f(\mathbf{Y}^*) \quad \text{and} \quad \mathbf{y}^* = -\mathbf{L}\nabla f(\mathbf{Y}^*)/2\theta^*\,,$$

which yields (2.7)

Finally, when $f$ is positively homogeneous of degree $d$ we then again exploit Euler's identity $\mathbf{Y}^*\cdot\nabla f(\mathbf{Y}^*) = d\,f(\mathbf{Y}^*)$ and (2.8) covers the two cases. □

2.3. **Main result.** We now can state our first main result. Recall that $\mathcal{L} = \boldsymbol{\ell}\,\boldsymbol{\ell}^T$.

**Theorem 2.5.** *Let Assumption 2.1 hold and let* $\mathbf{x}^* \in \mathbb{S}^{n-1}$ *satisfy* **(FONC)** *for Problem* (1.2)*.*

*(i) If* $\mathbf{x}^* \in \mathbb{S}^{n-1}$ *satisfies* **(FONC)** *for Problem* (1.2) *with* $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$ *then there exists* $\mathbf{y}^* \in \mathscr{E}_m$ *which satisfies* **(FONC)** *for problem* (1.3) *with same value* $f(\mathbf{Y}^*) = f(\mathbf{X}^*)$*, in addition,* $\mathbf{Y}^* = \mathbf{X}^*$*.*

*(ii) Conversely if* $\mathbf{y}^* \in \mathscr{E}_m$ *satisfies* **(FONC)** *for Problem* (1.3) *then there exists* $\mathbf{x}^* \in \mathbb{S}^{n-1}$ *which satisfies* **(FONC)** *for problem* (1.2)*, with same value* $f(\mathbf{X}^*) = f(\mathbf{Y}^*)$ *and with* $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$*.*

A detailed proof is postponed to §5.1.

**Remark 2.6.** *(a) If* $h$ *is positively homogeneous of degree* $d$ *then by Euler's identity,* $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) > 0$ *reads* $f(\mathbf{X}^*) > 0$*. But if* $m < n$ *the interesting local minima are strictly negative because* $f(\mathbf{X}) = 0$ *can always be attained at* $\mathbf{X} = \boldsymbol{\ell}\,\mathbf{x}/\|\mathbf{x}\|$ *for any* $0 \neq \mathbf{x} \in \mathrm{Ker}(\boldsymbol{\ell})$ *(and* $\mathbf{x}/\|\mathbf{x}\| \in \mathbb{S}^{n-1}$*). Therefore the interesting critical points of* (1.2) *are those* $\mathbf{x}^*$ *for which* $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$*, i.e., exactly those of Problem* (1.3)*.*

*(b) If* $h$ *is convex (or even pseudo-convex) then necessarily* $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \geq 0 \Rightarrow h(\mathbf{x}^*) \geq 0$*. Indeed by convexity of* $h$*,* $-\mathbf{x}^* \cdot \nabla h(\mathbf{x}^*) \geq 0 \Rightarrow 0\,(= h(0)) \geq h(\mathbf{x}^*)$*, i.e.,* $-\boldsymbol{\ell}\,\mathbf{x}^* \cdot \nabla f(\mathbf{X}^*) \geq 0 \Rightarrow 0\,(= h(0)) \geq f(\mathbf{X}^*)$*. Therefore again as in (a), the interesting critical points of* (1.2) *are those* $\mathbf{x}^*$ *for which* $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$*, i.e., exactly those of Problem* (1.3)*.*

As we next see, the second-order optimality conditions **(SONC)** show that problem (1.2) is indeed equivalent to Problem (1.3).

## 2.4. **Second-order optimality conditions.**

We now consider second-order optimality conditions **(SONC)** that complement **(FONC)** . Given $\mathbf{x} \in \mathbb{R}^n$, let $\mathbf{x}^\perp := \{\mathbf{u} \in \mathbb{S}^{n-1} : \mathbf{x} \cdot \mathbf{u} = 0\}$, and let $\boldsymbol{\ell} \in \mathbb{R}^{m \times n}$ be the matrix $(\boldsymbol{\ell}_{ij})_{i \leq m, j \leq n}$. Recall that $\mathbf{L} = (\boldsymbol{\ell}\,\boldsymbol{\ell}^T)^{1/2}$ is non singular and symmetric. Next, observe that with $h(\mathbf{x}) := f((\boldsymbol{\ell}_1 \cdot \mathbf{x}), \ldots, (\boldsymbol{\ell}_m \cdot \mathbf{x}))$,

$$\nabla^2 h(\mathbf{x}) \,=\, \boldsymbol{\ell}^T\, \nabla^2 f(\mathbf{X})\, \boldsymbol{\ell}\,.$$

Therefore, at a point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ that satisfies **(FONC)** for Problem (1.2), the second-order necessary optimality condition **(SONC)** reads:

$$\mathbf{u} \cdot \nabla^2 h(\mathbf{x}^*)\, \mathbf{u} \,\geq\, 2\lambda^*\,, \quad \forall \mathbf{u} \in (\mathbf{x}^*)^\perp\,,$$

where

$$(\mathbf{x}^*)^\perp \,=\, \{\, \mathbf{u} \in \mathbb{S}^{n-1} : \, \boldsymbol{\ell}\,\mathbf{u} \cdot \nabla f(\mathbf{X}^*) \,=\, 0 \,\}\,. \tag{2.9}$$

Equivalently, by Lemma 2.2:

$$\boldsymbol{\ell}\,\mathbf{u} \cdot \nabla^2 f(\mathbf{X}^*)\, \boldsymbol{\ell}\,\mathbf{u} \,\geq\, 2\,\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*)\,, \quad \forall \mathbf{u} \in \, (\mathbf{x}^*)^\perp\,, \tag{2.10}$$

and (2.10) covers the two cases $\nabla f(\mathbf{X}^*) = 0$ and $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \neq 0$. Similarly, the second-order sufficient optimality condition **(SOSC)** reads:

$$\boldsymbol{\ell}\,\mathbf{u} \cdot \nabla^2 f(\mathbf{X}^*)\, \boldsymbol{\ell}\,\mathbf{u} \,>\, -2\,\theta^* 2\,\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*)\,, \quad \forall \mathbf{u} \in \, (\mathbf{x}^*)^\perp\,.$$

**Remark 2.7.** *If $m < n$ notice that with $0 \neq \mathbf{u}/\|\mathbf{u}\| \in \mathrm{Ker}(\boldsymbol{\ell}) \in (\mathbf{x}^*)^\perp$ in* (2.10)*, one immediately sees that necessarily* **(SONC)** *implies* $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$*. In particular if $f$ is positively homogeneous then* **(SONC)** *implies* $f(\mathbf{X}^*) \leq 0$*.*

Similarly, at a **(FONC)** point $\mathbf{y}^* \in \mathscr{E}_m$ for (1.3), **(SONC)** reads

$$\mathbf{u} \cdot \mathbf{L}\, \nabla^2 f(\mathbf{Y}^*)\, \mathbf{L}\, \mathbf{u} \,\geq\, -2\theta^*\,, \quad \forall \mathbf{u} \in (\mathbf{y}^*)^\perp\,,$$

and **(SOSC)** reads

$$\mathbf{u} \cdot \mathbf{L}\, \nabla^2 f(\mathbf{Y}^*)\, \mathbf{L}\, \mathbf{u} \,>\, -2\theta^*\,, \quad \forall \mathbf{u} \in (\mathbf{y}^*)^\perp\,,$$

where $\theta^* \geq 0$ is as in (2.7), and

$$(\mathbf{y}^*)^\perp \,=\, \{\, \mathbf{u} \in \mathbb{S}^{m-1} :\, \mathbf{u} \cdot \mathbf{L}\, \nabla f(\mathbf{Y}^*) \,=\, 0\,\}\,. \tag{2.11}$$

We next show an analogue of Theorem 2.5 for second-order necessary optimality conditions **(SONC)** .

**Theorem 2.8.** *(i) Let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfy* **(FONC)** *and* **(SONC)** *for Problem* (1.2)*. Then $\mathbf{y}^* \in \mathbb{S}^{m-1}$ in Theorem 2.5(i) also satisfies* **(FONC)** *as well as* **(SONC)** *for Problem* (1.3)*, and with same value $f(\mathbf{Y}^*) = f(\mathbf{X}^*)$.*

*(ii) Conversely, let $\mathbf{y}^* \in \mathbb{S}^{m-1}$ satisfy* **(FONC)** *and* **(SONC)** *for Problem* (1.3) *Then $\mathbf{x}^* \in \mathbb{S}^{n-1}$ in Theorem 2.5(ii) also satisfies* **(FONC)** *as well as* **(SONC)** *for Problem* (1.2)*, and with same value $f(\mathbf{X}^*) = f(\mathbf{Y}^*)$.*

*In addition, (i) and (ii) also hold true if we replace* **(SONC)** *by* **(SOSC)** .

A detailed proof is postponed to §5.2.

Theorem 2.8 nicely complements Theorem 2.5. Indeed now, to *every* point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ that satisfies **(SONC)** for (1.2), corresponds a point $\mathbf{y}^* \in \mathbb{S}^{m-1}$ (with same value) that satisfies **(SONC)** for (1.3), and the converse is also true. We have eliminated the ambiguity in **(FONC)** for (1.2) where $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*)$ can be positive or negative as indeed a local maximum also satisfies the same **(FONC)** condition. This is because with an equality constraint, the Lagrange multiplier $\lambda^*$ is unsigned. This is not true for (1.3) where with the *inequality* constraint, the associated Lagrange multiplier is nonnegative (whence $\mathbf{Y}^* \cdot f(\mathbf{Y}^*) \leq 0$). However in **(SONC)** the ambiguity "local minimum versus local maximum" disappears.

An important consequence of Theorem 2.8 is that if the form (1.2) of Problem (1.1) is known (i.e., the $\boldsymbol{\ell}_j$'s are available) then instead of solving Problem (1.2) in $\mathbb{R}^n$, one may rather consider solving Problem (1.3) in $\mathbb{R}^m$, especially when $m \ll n$. In particular, it is much easier to optimize on $\mathscr{E}_m$ (a nice convex set) than optimize on the non convex set $\mathbb{S}^{n-1}$. Moreover if $f$ is a polynomial, then one may apply the Moment-SOS hierarchy described in e.g. [13]. Generically it has finite convergence and fast in practice (as observed at least for reasonable dimension $m$).

2.5. **The case $m = 1$.** Interestingly, the case $m = 1$ which seems rather restrictive in fact contains the class of odd-degree *quasi-convex polynomials.* Indeed if $h \in \mathbb{R}[\mathbf{x}]$ is quasi-convex of odd-degree then $h(\mathbf{x}) = f(\boldsymbol{\ell} \cdot \mathbf{x})$ for some $\boldsymbol{\ell} \in \mathbb{R}^m$ and some monotonic univariate polynomial $f$. See for instance Ahmadi et al [1] or [13, Theorem 13.11].

**Corollary 2.9.** *Let $m = 1$ and let $f$ be the univariate polynomial*

$$t \mapsto f(t) = \sum_{k=1}^{d} f_k\, t^k, \quad t \in \mathbb{R},$$

*so that* (1.2) *reads* $\min_{\mathbf{x}} \{ f(\boldsymbol{\ell} \cdot \mathbf{x}) : \mathbf{x} \in \mathbb{S}^{n-1} \}$ *for some* $\boldsymbol{\ell} \in \mathbb{R}^n$*. Let* $\mathbf{T} := \{ t : f'(t) = 0 ; |t| \leq \|\boldsymbol{\ell}\| \}$ *(so that* $\#\mathbf{T} \leq d-1$*). Then:*

$$f^* = \min \left[ f(\|\boldsymbol{\ell}\|) , f(-\|\boldsymbol{\ell}\|) , \min[ f(t) : t \in \mathbf{T} ] \right] , \tag{2.12}$$

*that is, it suffices to compare the values of $f$ at $d+1$ points of* $[-1, 1]$*.*

*In particular, if $h$ in* (1.2) *is quasi-convex of odd degree then*

$$f^* = \min \left[ f(\|\boldsymbol{\ell}\|) , f(-\|\boldsymbol{\ell}\|) \right] ,$$

*and $h$ has no spurious minimum on the sphere.*

*Proof.* By Theorem 2.8, the global minimum $f^*$ of (1.2) is the same as that of Problem (1.3), i.e., the global minimum of a univariate polynomial on the interval $\mathscr{E}_1 = [-1, 1]$. That is, (1.3) reads:

$$\min \{ \tilde{f}(y) : y^2 \leq 1 \} = \min \{ f(\|\boldsymbol{\ell}\|\, y) : y^2 \leq 1 \} .$$

trivial to solve. The critical points are the end points $y = \pm 1$ and the points $t \in [-1, 1]$ where $t\,\|\boldsymbol{\ell}\|$ is a zero of the derivative $f'$, which yields (2.12). Finally if $h$ in (1.2) is quasi-convex of odd-degree then as $f$ is monotonic, its global minimum is attained at one of the end-points $y = \pm 1$, and it is the unique local (hence global) minimum on $[-1, 1]$. ☐

Corollary 2.9 states that Problem (1.2) with $m = 1$, has at most $d$ spurious local minima, and in fact at most roughly $d/2$ since only half of the points in $\mathbf{T}$ can be local minima, and among them only those with negative value since $f(0) = 0$.

2.6. **Detecting the formulation** (1.2)**.** An optimization problem (1.1) may have a *hidden* formulation (1.2), i.e., (1.2) is valid but only $h$ is known. Therefore if the form (1.2) exists and is to be exploited, an important issue is how to reveal a formulation (1.2) from the sole knowledge of $h$ in the $n$ variables $\mathbf{x} = (x_1, \ldots, x_n)$.

Let $\mathscr{E}_n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|^2 \leq 1\}$. and let $\mu$ be the Lebesgue measure on $\mathscr{E}_n$ normalized to a probability measure.

**Theorem 2.10.** *Let $h \in \mathbb{R}[\mathbf{x}]_d$ in* (1.1) *be such that the hidden formulation* (1.2) *exists for some unknown $f \in \mathbb{R}[\mathbf{X}]_d$ and some unknown linear forms $\boldsymbol{\ell}_j$, $j \in [m]$, that are linearly independent.*

*Let $(\mathbf{x}(i))_{i\in[m]} \subset \mathscr{E}_n$ be a sample of $m$ points drawn from $\mu$, and assume that the vectors $\mathbf{s}_i := \nabla h(\mathbf{x}(i))$, $i \in [m]$, are linearly independent.*

*Let $\mathbf{s} \in \mathbb{R}^{m\times n}$ be the matrix with ith row vector $\mathbf{s}_i := \nabla h(\mathbf{x}(i))$, $i \in [m]$, so that $\mathcal{L} := \mathbf{s}\,\mathbf{s}^T \in \mathbb{R}^{m\times m}$ is non singular. Let $\mathbf{L} := \mathcal{L}^{-1/2}$ and introduce the polynomial $f_h \in \mathbb{R}[\mathbf{y}]_d$:*

$$\mathbf{y} \mapsto f_h(\mathbf{y}) \,:=\, h(\mathbf{s}\,\mathbf{L}^{-1}\mathbf{y})\,,\quad \mathbf{y} \in \mathbb{R}^m\,.$$

*Then every critical point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ of* (1.1) *with $\nabla h(\mathbf{x}^*) \neq 0$, corresponds a critical point $\mathbf{y}^* \in \mathbb{S}^{m-1}$ of*

$$\min_{\mathbf{y}} \{\, f_h((\mathbf{L}_1\cdot\mathbf{y}),\ldots,(\mathbf{L}_m\cdot\mathbf{y})) :\, \mathbf{y} \in \mathbb{S}^{m-1}\,\}\,. \tag{2.13}$$

*with same value $h(\mathbf{x}^*) = f_h((\mathbf{L}_1\cdot\mathbf{y}^*),\ldots,(\mathbf{L}_m\cdot\mathbf{y}^*))$.*

*Proof.* As $h \in \mathbb{R}[\mathbf{x}]_d$ satisfies (1.2) for some hidden $f$, then by construction and under our assumption on the $\boldsymbol{\ell}_j$'s,

$$\nabla h(\mathbf{x}) \,=\, \boldsymbol{\ell}^T\,\nabla f(\mathbf{X}) \,\in\, \mathrm{Span}(\boldsymbol{\ell}_1,\ldots,\boldsymbol{\ell}_m) \,=:\, V\,.$$

Therefore $H := \mathrm{Span}\{\nabla h(\mathbf{x}) : \mathbf{x} \in \mathscr{E}_n\} \subset V$ is an $r$-dimensional subspace of $V$ with $r \leq m$, and in particular the $\mathbf{s}_j$'s form a basis of $H$. In fact as the vectors $\{\nabla h(\mathbf{x}(i))\}_{i\in[m]}$ are linearly independent, $r = m$ and the $\mathbf{s}_j$'s form a basis of $V$.

By (2.4) in Lemma 2.2, a critical point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ of (1.2) with $\nabla h(\mathbf{x}^*) \neq 0$ is of the form $\mathbf{x}^* = \nabla h(\mathbf{x}^*)/2\lambda^*$ and so necessarily $\mathbf{x}^* \in V$. Therefore, as for optimization purpose we are only interested in such critical points, write $\mathbf{x} = \mathbf{s}^T\mathbf{y}$, with $\mathbf{y} \in \mathbb{R}^m$, so that $h(\mathbf{x}) = h(\mathbf{s}^T\,\mathbf{y}) \in \mathbb{R}[\mathbf{y}]_d$ for all such points. Moreover, the constraint $\mathbf{x}^* \in \mathbb{S}^{n-1}$ reads $\mathbf{y}\cdot\mathbf{s}\,\mathbf{s}^T\mathbf{y} = \mathbf{y}^T\mathcal{L}\mathbf{y} = 1$. Therefore to every critical point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ of (1.1) with $\nabla h(\mathbf{x}^*) \neq 0$, corresponds a critical point $\mathbf{y}^* \in \mathbb{S}^{m-1}$ of the minimization problem:

$$\min_{\mathbf{y}} \{\, f_h((\mathbf{L}_1\cdot\mathbf{y}),\ldots,(\mathbf{L}_m\cdot\mathbf{y})) :\, \mathbf{y} \in \mathbb{S}^{m-1}\,\}\,,$$

where $\mathbf{L} = \mathcal{L}^{-1/2}$ and $\mathbf{y} \mapsto f_h(\mathbf{y}) := h(\mathbf{s}^T\mathbf{L}^{-1}\mathbf{y})$, which yields (2.13). □

As the points $\{\mathbf{x}(i)\}_{i\in[m]} \subset \mathscr{E}_n$ are drawn from the uniform distribution $\mu$ on $\mathscr{E}_n$, then under reasonable hypotheses on the unknown $f$ in (1.3), with probability 1 the vectors $\{\nabla h(\mathbf{x}(i))\}_{i\in[m]}$ are linearly independent. Indeed $\nabla h(\mathbf{x}) = \boldsymbol{\ell}^T\nabla f(\boldsymbol{\ell}\,\mathbf{x})$ and therefore it enough that with probability 1, the vectors $\{\nabla f(\boldsymbol{\ell}\,\mathbf{x}(i))\}_{i\in[m]}$ are linearly independent. So consider the polynomial $\theta \in \mathbb{R}[\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_m]$, defined by:

$$(\mathbf{U}_1,\ldots,\mathbf{U}_m) \mapsto \theta(\mathbf{U}_1,\ldots,\mathbf{U}_m) := \det[\nabla f(\boldsymbol{\ell}\,\mathbf{U}_1),\ldots,\nabla f(\boldsymbol{\ell}\,\mathbf{U}_m)]\,.$$

If $\theta$ is not identically zero, then with $\mu^{\otimes m}$ being the product measure $\underbrace{\mu \otimes \mu \otimes \cdots \otimes \mu}_{m \text{ times}}$ on $\mathscr{E}_n^m$,

$$\mu^{\otimes m}\left(\{(\mathbf{U}_1, \ldots, \mathbf{U}_m) : \theta(\mathbf{U}_1, \ldots, \mathbf{U}_m) = 0\}\right) = 0$$

because level sets of polynomials have zero-Lebesgue measure. Therefore if $\theta$ is not identically null then with probability 1, the sample $\{\nabla h(\boldsymbol{\ell}\,\mathbf{x}(i))\}_{i\in[m]}$ of Theorem 2.10 is a linearly independent family.

**Remark 2.11.** *Theorem 2.10 also holds if $h$ in* (1.1) *and the hidden function $f$ in* (1.2) *are continuously differentiable functions. Indeed by construction, $\nabla h(\mathbf{x}) \in V$ for all $\mathbf{x}$, where $V$ is the $m$-dimensional vector space spanned by the unknown $\{\boldsymbol{\ell}_j\}_{j\in[m]}$. Again by Lemma 2.2, all critical points $\mathbf{x}^* \in \mathbb{S}^{n-1}$ with $\nabla f(\mathbf{X}^*) \neq 0$, are in $V$. However, some more detailed analysis is needed to check whether the family $\{\nabla h(\mathbf{x}(i))\}_{i\in[m]}$ spans $V$, or a larger sample may be needed. But the same conclusion with* (2.13) *is valid.*

## 3. The homogeneous case

In this section we assume that $f$ is a homogeneous polynomial of degree $d > 2$, that is $f(\lambda\,\mathbf{X}) = \lambda^d f(\mathbf{X})$ for all $\lambda > 0$ and all $\mathbf{x} \in \mathbb{R}^n$. Then of course $\mathbf{x} \mapsto h(\mathbf{x}) := f(\boldsymbol{\ell}\,\mathbf{x})$ is also positively homogeneous of degree $d$.

Any local minimum $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfies **(SONC)** and so if $m < n$, then by Remark 2.7, necessarily $f(\mathbf{X}^*) \leq 0$. Moreover for any $\mathbf{x} \in \mathrm{Ker}(\boldsymbol{\ell}) \cap \mathbb{S}^{n-1}$, $f(\mathbf{X}) = f(0) = 0$; see Remark 2.6. As a matter of fact, we even obtain the following consequence:

**Corollary 3.1.** *Let $m < n$ and let $\mathbf{x} \mapsto h(\mathbf{x}) := f(\boldsymbol{\ell}\,\mathbf{x})$ be a twice continuously differentiable and positively homogeneous function. Then necessarily $f(\mathbf{X}^*) \leq 0$ at any point $\mathbf{x}^* \in \mathbb{S}^{n-1}$ that satisfies* **(SONC)** *(and so at any local minimizer).*

*In particular, if $h$ is nonnegative then $f^* = 0$ is the unique local (hence global) minimum of $h$ on $\mathbb{S}^{n-1}$, and is attained at any point $\mathbf{x} \in \mathrm{Ker}(\boldsymbol{\ell}) \cap \mathbb{S}^{n-1}$. Moreover $f^* = 0$ is also the unique local (hence global) minimum of Problem* (1.3) *and is attained at some point $\mathbf{y}^* \in \mathscr{E}_m$.*

*Proof.* Let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfy **(SONC)** so that necessarily $f(\mathbf{X}^*) \leq 0$; see Remark 2.7. If $\nabla f(\mathbf{X}^*) = 0$ then by homogeneity $h(\mathbf{x}^*) = f(\mathbf{X}^*) = 0$. Next, assume that $h$ is nonnegative. As $m < n$, any point $\mathbf{x}^* \in \mathrm{Ker}(\boldsymbol{\ell}) \cap \mathbb{S}^{n-1}$ (guaranteed to exist, see Remark 2.6) is a global minimizer with $h(\mathbf{x}^*) = f(0) = 0$. As $\mathbf{x}^*$ necessarily satisfies **(SONC)** , let $\mathbf{y}^* \in \mathscr{E}_m$ be obtained from $\mathbf{x}^*$ as in Theorem 2.8(i). Then as $\mathbf{X}^* = \mathbf{Y}^*$, $f(\mathbf{Y}^*) = f^*$. Suppose that $\mathbf{y}^*$ is not global minimizer so that there exists a global minimizer $\mathbf{z} \in \mathscr{E}_m$ with $f(\mathbf{Z}) < f^*$. As necessarily $\mathbf{z}$ satisfies **(SONC)** , then by Theorem 2.8(ii) there exists $\mathbf{x}' \in \mathbb{S}^{n-1}$ with same value $f(\mathbf{X}') = f(\mathbf{Z}) < f^*$, in contradiction with $\mathbf{x}^*$ being a global minimizer. ∎

Corollary 3.1 characterizes a class of positively homogeneous functions that have no spurious local minima on $\mathbb{S}^{n-1}$. Of course if a nonnegative $h$ is given in the form (1.2) then it suffices to find a point in $\mathbf{x}^* \in \mathrm{Ker}(\boldsymbol{\ell}) \cap \mathbb{S}^{n-1}$. But if only $h$ is given (i.e., without knowing that $h$ is of the form (1.2) for some matrix $\boldsymbol{\ell} \in \mathbb{R}^{m\times n}$), then any local minimization algorithm converging to a point that satisfies **(FONC)** would converges to the global optimum $f^* = 0$.

Therefore for homogeneous problems in the form (1.2) we are especially interested with negative local minima $h(\mathbf{x}^*)\,(= f(\mathbf{X}^*)) < 0$ whenever they exist (e.g. when $d > 2$ is odd). Given a real symmetric matrix $\mathbf{A} \in \mathbb{R}^{t\times t}$ denote by

$$\lambda_1(\mathbf{A}) \,\leq\, \lambda_2(\mathbf{A}) \,\leq\, \cdots \,\leq\, \lambda_t(\mathbf{A})$$

its eigenvalues, counting multiplicities and ordered by increasing values. In the following we assume that $h$ is a homogeneous polynomial of degree $d > 2$ (and then so is $f$). In this case the author has recently provided in [9] a complete characterization of first- and second-order necessary optimality conditions solely in terms of the first two smallest eigenvalues of the Hessian of $h$. We first need the following result whose proof is quite straightforward.

**Proposition 3.2.** *Let $h \in \mathbb{R}[\mathbf{x}]_d$ in* (1.1) *be homogeneous of odd degree $d > 1$ and let $\mathbf{y}^*$ be* **(SONC)** *point of* (1.3) *with negative value $f(\mathbf{Y}^*) < 0$. Then $\mathbf{y}^*$ is a* **(SONC)** *point of*

$$\min\{\, f(\mathbf{L}\,\mathbf{y}) \,:\, \mathbf{y} \in \mathbb{S}^{m-1}\,\}. \tag{3.1}$$

*Proof.* As $\mathbf{Y}^*\cdot\nabla f(\mathbf{Y}^*) = d\,f(\mathbf{Y}^*) < 0$ then by Lemma 2.7, $\mathbf{y}^* \in \mathbb{S}^{m-1}$. Next, by Lemma 2.2 applied to (3.1), $\mathbf{y}^*$ is a **(FONC)** point of (3.1). Finally as $\mathbf{y}^* \in \mathbb{S}^{m-1}$ satisfies **(SONC)** for (1.3) it also satisfies **(SONC)** for (3.1) since there is no restriction of sign for $\theta^*$. □

**Lemma 3.3.** *Let $m < n$ and let $h$ in* (1.1) *be a degree-$d$ homogeneous polynomial with $d > 2$. Then:*

*(i) $\mathbf{x}^* \in \mathbb{S}^{n-1}$ with $f(\mathbf{X}^*) < 0$ satisfies* **(SONC)** *for* (1.2) *if and only if*

$$\lambda_2(\boldsymbol{\ell}^T\nabla^2 f(\mathbf{X}^*)\,\boldsymbol{\ell}) \,\geq\, d\,f(\mathbf{X}^*)\,. \tag{3.2}$$

*In particular $\mathbf{X}^*$ is an eigenvector of $\mathcal{L}\nabla^2 f(\mathbf{X}^*)$ with eigenvalue $d(d-1)f(\mathbf{X}^*)$, i.e.,*

$$\mathcal{L}\,\nabla^2 f(\mathbf{X}^*)\,\mathbf{X}^* \,=\, d(d-1)\,f(\mathbf{X}^*)\,\mathbf{X}^*\,. \tag{3.3}$$

*(ii) Similarly, $\mathbf{y}^* \in \mathscr{E}_m$ with $f(\mathbf{Y}^*) < 0$ satisfies* **(SONC)** *for* (1.3) *if and only if*

$$\lambda_2(\mathbf{L}\,\nabla^2 f(\mathbf{Y}^*)\mathbf{L}) \,\geq\, d\,f(\mathbf{Y}^*)\,. \tag{3.4}$$

*In particular $\mathbf{Y}^*$ is an eigenvector of $\mathcal{L}\nabla^2 f(\mathbf{Y}^*)$ associated with the eigenvalue $d(d-1)f(\mathbf{Y}^*)$, i.e.,*

$$\mathcal{L}\,\nabla^2 f(\mathbf{Y}^*)\,\mathbf{Y}^* \,=\, d(d-1)\,f(\mathbf{Y}^*)\,\mathbf{Y}^*\,. \tag{3.5}$$

*Proof.* (i) (3.2) is a direct consequence of [9, Corollary 2.4], in which it is also proved that $\mathbf{x}^* \in \mathbb{S}^{n-1}$ is en eigenvector of $\nabla^2 h(\mathbf{x}^*)$ associated with its smallest eigenvalue $\lambda_1(\nabla^2 h(\mathbf{x}^*)) = d\,(d-1)\,h(\mathbf{x}^*)$. Therefore

$$\begin{aligned} d\,(d-1)\,h(\mathbf{x}^*)\,\mathbf{x}^* &= \nabla^2 h(\mathbf{x}^*)\,\mathbf{x}^* \\ &= \boldsymbol{\ell}^T\,\nabla^2 f(\mathbf{X}^*)\,\boldsymbol{\ell}\mathbf{x}^* \,=\, \boldsymbol{\ell}^T \nabla^2 f(\mathbf{X}^*)\,\mathbf{X}^*\,, \end{aligned}$$

and multiplying on the left by $\boldsymbol{\ell}$ yields (3.3).

(ii) Let $\mathbf{y}^*$ be a **(SONC)** point of (1.3) with negative value $f(\mathbf{Y}^*) < 0$. By Lemma 2.7, $\mathbf{y}^* \in \mathbb{S}^{m-1}$ and $\mathbf{y}^*$ is a **(SONC)** point of (3.1) (a homogeneous problem on $\mathbb{S}^{m-1}$). Then again as a direct consequence of [9, Corollary 2.4] applied to (3.1),

$$\lambda_2(\mathbf{L}\,\nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}) \,\geq\, d\,f(\mathbf{Y}^*)$$

which is (3.4), and $\mathbf{y}^*$ is an eigenvector of $\mathbf{L}\,\nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}$ associated with the smallest eigenvalue $\lambda_1(\mathbf{L}\,\nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}) = d\,(d-1) f(\mathbf{Y}^*)$. That is,

$$\mathbf{L}\,\nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}\,\mathbf{y}^* \,=\, d\,(d-1)\,f(\mathbf{Y}^*)\,\mathbf{y}^*\,,$$

and multiplying on the left by $\mathbf{L}$ yields

$$\begin{aligned} \mathcal{L}\,\nabla^2 f(\mathbf{Y}^*)\,\mathbf{Y}^* &= \mathbf{L}\,\mathbf{L}\,\nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}\,\mathbf{y}^* \\ &= d\,(d-1)\,f(\mathbf{Y}^*)\,\mathbf{L}\,\mathbf{y}^* \,=\, d\,(d-1)\,f(\mathbf{Y}^*)\,, \end{aligned}$$

which is (3.5). □

3.1. **Odd-degree forms of $m$ linear forms.** We mentioned in the introduction that interesting NP-hard combinatorial problems reduce to minimizing a cubic form on the Euclidean sphere. In this section we consider the case $(m, d) = (m, 2p + 1)$ with $p \geq 1$, that is, $f$ is a $m$-variate form of odd degree (whose cubic forms are a particular case).

So Problem (1.2) of minimizing an odd-degree form of $m$ linear forms on $\mathbb{S}^{n-1}$, reduces to (1.3), i.e., minimizing an $m$-variate odd-degree form on $\mathscr{E}_m$. As the degree is odd, we have seen that all local minima are nonpositive and are attained on $\mathbb{S}^{m-1}$. Therefore with $g(\mathbf{y}) = f(\mathbf{L}_1\mathbf{y}, \ldots, \mathbf{L}_m\mathbf{y})$ consider the generic problem

$$\min\,\{\, g(\mathbf{y}) :\, \|\mathbf{y}\|^2 \,=\, 1\,\},$$

where $g$ is a cubic form. Then **(FONC)** condition in Lemma 2.4 yields that that any critical point $\mathbf{y}^* \in \mathbb{S}^{m-1}$ should solve

$$\nabla g(\mathbf{y}) + \lambda\,\mathbf{y} \,=\, 0\,;\quad \|\mathbf{y}\|^2 \,=\, 1\,,$$

for some $\lambda$, and by Bézout theorem, the above system has at most $2\,(d-1)^m = 2.(2p)^m$ solutions.

Moreover, $-(\mathbf{y}^*, \lambda)$ is solution whenever $(\mathbf{y}^*, \lambda)$ is solution. But only one of the two provides a negative value for $g(\mathbf{y}^*)$. So there are at most $(2p)^m$ local minimizers for Problem (1.3), all with negative associated value $f(\mathbf{Y}^*)$. These are also the only "interesting" local minima for problem (1.2) since all others (if any) have $f^* = 0$ associated value.

**Corollary 3.4.** *If $f$ is an odd degree-$d$ form, then independently of the dimension $n$, Problem* (1.2) *has at most $(2\lfloor d/2\rfloor)^m$ local minima with negative value. In particular, cubic forms have at most $2^m$ local minima with negative value.*

## 4. Conclusion

We have presented some results on optimization of functions of $m\,(<n)$ linear forms on the Euclidean sphere $\mathbb{S}^{n-1}$. The key result is to reduce the problem to one of the same type but now on the Euclidean unit ball $\mathscr{E}_m\subset\mathbb{R}^m$. Indeed, to every point that satisfies the standard second-order necessary optimality conditions is associated a point that satisfies the standard second-order necessary optimality conditions of the latter, and the converse is also true. It thus results in a drastic reduction of the computational effort if $m\ll n$ and the forms are known. If they are not known then there is a practical algorithm to reveal this hidden formulation. As a by-product we also obtain that quasi-convex polynomials of odd-degree have no spurious local minima on $\mathbb{S}^{n-1}$ and the global minimum has an analytical easy expression. The general homogeneous case has also some interesting properties.

## 5. Appendix

### 5.1. **Proof of Theorem 2.5.**

**Cases $\nabla f(\mathbf{X}^*)\neq 0$ and $\nabla f(\mathbf{Y}^*)\neq 0$.**

*Proof.* (i) If $\mathbf{x}^*\in\mathbb{S}^{n-1}$ satisfies **(FONC)** for problem (1.2) then by Lemma 2.2:

$$\mathbf{x}^*\,=\,\tau^{-1}\sum_{j=1}^s\frac{\partial f(\mathbf{X}^*)}{\partial X_j}\,\boldsymbol{\ell}_j\,=:\,\sum_{j=1}^s z_j^*\,\boldsymbol{\ell}_j$$

with $\tau\,\mathbf{z}^*=\nabla f(\mathbf{X}^*)$. Next, as $\mathbf{x}^*\in\mathbb{S}^{n-1}$, and letting $\mathbf{y}^*:=\mathbf{L}\mathbf{z}^*$,

$$1\,=\,\|\mathbf{x}^*\|^2\,=\,\|\sum_{j=1}^s z_j^*\,\ell_j\|^2\,=\,\mathbf{z}^*\cdot\mathcal{L}\,\mathbf{z}^*\,=\,\|\mathbf{L}\,\mathbf{z}^*\|^2\,=\,\|\mathbf{y}^*\|^2\,.$$

It remains to prove that $f(\mathbf{L}\,\mathbf{y}^*)=f(\mathbf{X}^*)$. But this follows from

$$\begin{aligned}\mathbf{X}_j^*\,=\,\boldsymbol{\ell}_j\cdot\mathbf{x}^*\;&=\;\sum_{i=1}^s(\boldsymbol{\ell}_j\cdot\boldsymbol{\ell}_i)\,z_i^*\,=\,(\mathcal{L}\,\mathbf{z}^*)_j\\&=\;(\mathbf{L}\,\mathbf{L}\,\mathbf{z}^*)_j\,=\,(\mathbf{L}\,\mathbf{y}^*)_j\,=\,\mathbf{L}_j\cdot\mathbf{y}^*\,=\,\mathbf{Y}_j^*,\quad j\in[s]\,,\end{aligned}$$

and therefore

$$\begin{aligned}f(\mathbf{X}^*)\;&=\;f((\boldsymbol{\ell}_1\cdot\mathbf{x}^*),\ldots,(\boldsymbol{\ell}_m\cdot\mathbf{x}^*))\\&=\;f((\mathbf{L}_1\cdot\mathbf{y}^*),\ldots,(\mathbf{L}_m\cdot\mathbf{y}^*))\,=\,f(\mathbf{Y}^*)\,,\end{aligned}$$

which completes the proof of (i).

(ii) Conversely let $\mathbf{y}^* \in \mathscr{E}_m$ satisfy **(FONC)** for Problem (1.3) and let $\mathbf{Y}^* := (\mathbf{L}_j \cdot \mathbf{y}^*)_{j\in[m]}$. That is, by Lemma 2.4:

$$\sum_{j=1}^m \frac{\partial f(\mathbf{Y}^*)}{\partial X_j}\,\mathbf{L}_j \;=\; -2\theta^*\,\mathbf{y}^* \quad\text{and}\quad -2\theta^* = \mathbf{Y}^*\cdot\nabla f(\mathbf{Y}^*) \neq 0.$$

Hence

$$\mathbf{y}^* \;=\; (-2\theta^*)^{-1}\sum_{j=1}^m \frac{\partial f(\mathbf{Y}^*)}{\partial X_j}\,\mathbf{L}_j$$

and with $\mathbf{z}^* := \mathbf{L}^{-1}\,\mathbf{y}^*$, let $\mathbf{x}^* := \sum_j z_j^*\,\boldsymbol{\ell}_j$. Then

$$\|\mathbf{x}^*\|^2 = \mathbf{z}^*\mathcal{L}\,\mathbf{z}^* = \mathbf{z}^*\mathbf{L}\,\mathbf{L}\mathbf{z}^* = \|\mathbf{y}^*\|^2 = 1\,,$$

so that $\mathbf{x}^* \in \mathbb{S}^{n-1}$. Next,

$$\mathbf{X}^* = \boldsymbol{\ell}\,\mathbf{x}^* \;=\; \boldsymbol{\ell}\,\boldsymbol{\ell}^T\mathbf{z}^* \;=\; \mathbf{L}\,\mathbf{L}\,\mathbf{z}^* \;=\; \mathbf{L}\,\mathbf{y}^* \;=\; \mathbf{Y}^*\,,$$

and in view of the definition of $\tilde{f}$ in (2.2), $f(\mathbf{X}^*) = f(\mathbf{Y}^*) = \tilde{f}(\mathbf{y}^*)$. Next,

$$\begin{aligned}
\mathbf{z}^* \;=\; \mathbf{L}^{-1}\mathbf{y}^* \;&=\; (-2\theta^*)^{-1}\sum_{j=1}^m \frac{\partial f(\mathbf{Y}^*)}{\partial X_j}\,\mathbf{L}^{-1}\mathbf{L}_j\\
&=\; (-2\theta^*)^{-1}\sum_{j=1}^m \frac{\partial f(\mathbf{Y}^*)}{\partial X_j}\,\mathbf{e}_j\,,
\end{aligned}$$

and thus $z_j^* = (-2\theta^*)^{-1}\partial f(\mathbf{Y}^*)/\partial X_j$, for all $j \in [m]$. Next, recalling that $\mathbf{y}^* = \mathbf{L}\,\mathbf{z}^*$, one obtains $\mathbf{Y}^* = \mathbf{L}\,\mathbf{L}\,\mathbf{z}^*$ and therefore,

$$\mathbf{x}^* \;=\; \sum_{j=1}^s z_j^*\,\boldsymbol{\ell}_j \;=\; (-2\theta^*)^{-1}\sum_{j=1}^m \frac{\partial f(\mathbf{Y}^*)}{\partial X_j}\,\boldsymbol{\ell}_j\,.$$

$$\sum_{j=1}^s \frac{\partial f(\mathbf{X}^*)}{\partial X_j}\,\boldsymbol{\ell}_j \;=\; \sum_{j=1}^s \frac{\partial f(\mathbf{Y}^*)}{\partial X_j}\,\boldsymbol{\ell}_j \;=\; -2\,\theta^*\,\mathbf{x}^*\,,$$

that is, $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfies **(FONC)** for problem (1.2) with

$$-2\,\theta^* \;=\; \mathbf{Y}^*\cdot\nabla f(\mathbf{Y}^*) \;=\; \mathbf{X}^*\cdot\nabla f(\mathbf{X}^*)\,.$$

□

**Cases $\nabla f(\mathbf{X}^*) = 0$ and $\nabla f(\mathbf{Y}^*) = 0$.**

*Proof.* (i) Let $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfy **(FONC)** for (1.2) with $\nabla f(\mathbf{X}^*) = 0$. Write $\mathbf{x}^* = \boldsymbol{\ell}^T\,\mathbf{u} + \mathbf{v}$ with $\mathbf{v} \in \mathrm{Ker}(\boldsymbol{\ell})$, so that $\boldsymbol{\ell}\,\mathbf{v} = 0$ and $\boldsymbol{\ell}^T\mathbf{u} \perp \mathbf{v}$. Then by orthogonality:

$$1 \;=\; \|\mathbf{x}^*\|^2 \;=\; \|\boldsymbol{\ell}^T\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \;\Rightarrow\quad \|\boldsymbol{\ell}^T\mathbf{u}\| \leq 1\,.$$

Let $\mathbf{y}^* := \mathbf{L}\,\mathbf{u}$ so that $\|\mathbf{y}^*\|^2 \;=\; \mathbf{u}\cdot\mathbf{L}\,\mathbf{L}\,\mathbf{u} \;=\; \mathbf{u}\cdot\boldsymbol{\ell}\,\boldsymbol{\ell}^T\mathbf{u} \;=\; \|\boldsymbol{\ell}^T\mathbf{u}\|^2 \;\leq\; 1$, and so $\mathbf{y}^* \in \mathscr{E}_m$. In addition,

$$\mathbf{X}^* \;=\; \boldsymbol{\ell}\,\mathbf{x}^* \;=\; \boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{u} = \mathbf{L}\,\mathbf{L}\,\mathbf{u} \;=\; \mathbf{L}\,\mathbf{y}^* \;=\; \mathbf{Y}^*\,,$$

and therefore $f(\mathbf{X}^*) = f(\mathbf{Y}^*)$.

Finally, $0 = \nabla f(\mathbf{X}^*) = \nabla f(\mathbf{Y}^*)$ so that $\mathbf{y}^*$ satisfies **(FONC)** with $\theta^* = 0$.

(ii) Let $\mathbf{y}^* \in \mathscr{E}_m$ satisfy **(FONC)** for (1.3) with $\nabla f(\mathbf{Y}^*) = 0$. Define $\mathbf{x}^* := \boldsymbol{\ell}^T \mathbf{L}^{-1}\mathbf{y}^* + \mathbf{v}$ where $\mathbf{v} \in \mathrm{Ker}(\boldsymbol{\ell})$ is arbitrary with $\|\mathbf{v}\|^2 = 1 - \|\mathbf{y}^*\|^2$. Then by orthogonality, $\|\mathbf{x}^*\|^2 = \|\boldsymbol{\ell}^T \mathbf{L}^{-1}\mathbf{y}^*\|^2 + \|\mathbf{v}\|^2$. Moreover,

$$\|\boldsymbol{\ell}^T \mathbf{L}^{-1}\mathbf{y}^*\|^2 \,=\, \mathbf{y}^* \cdot \mathbf{L}^{-1}\boldsymbol{\ell}\,\boldsymbol{\ell}^T\mathbf{L}^{-1}\,\mathbf{y}^* \,=\, \|\mathbf{y}^*\|^2 \,\leq\, 1\,,$$

and therefore $\|\mathbf{x}^*\|^2 = 1$, i.e., $\mathbf{x}^* \in \mathbb{S}^{n-1}$. Moreover,

$$\nabla f(\mathbf{X}^*) \,=\, \nabla f(\boldsymbol{\ell}\,\boldsymbol{\ell}^T \mathbf{L}^{-1}\,\mathbf{y}^*) \,=\, \nabla f(\mathbf{L}\,\mathbf{y}^*) \,=\, \nabla f(\mathbf{Y}^*) = 0,$$

so that $\mathbf{x}^* \in \mathbb{S}^{n-1}$ satisfies **(FONC)** for (1.2) with $\lambda^* = 0$. □

### 5.2. **Proof of Theorem 2.8.**

*Proof.* (i) Recall that $\mathbf{L} = (\boldsymbol{\ell}\,\boldsymbol{\ell}^T)^{1/2}$, and recall that $\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*) \leq 0$ whenever $\mathbf{x}^*$ satisfies **(SONC)** . Therefore, let $\mathbf{y}^* \in \mathbb{S}^{m-1}$ be as in Theorem 2.5(i) so that $\mathbf{Y}^* = \mathbf{X}^*$. Next, the second-order necessary optimality condition **(SONC)** for problem (1.3) reads:

$$\mathbf{L}\,\mathbf{v} \cdot \nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}\,\mathbf{v} \,\geq\, 2\,\mathbf{Y}^* \cdot \nabla f(\mathbf{Y}^*)\,, \quad \forall\, \mathbf{v} \in (\mathbf{y}^*)^{\perp}\,. \tag{5.1}$$

where

$$(\mathbf{y}^*)^{\perp} \,=\, \{\,\mathbf{v} \in \mathbb{S}^{m-1} :\, \mathbf{L}\,\mathbf{v} \cdot \nabla f(\mathbf{Y}^*) \,=\, 0\,\}\,.$$

Hence suppose that $\mathbf{v} \in (\mathbf{y}^*)^{\perp}$ violates **(SONC)** for Problem (1.3), i.e.,

$$\mathbf{L}\,\mathbf{v} \cdot \nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}\,\mathbf{v} \,<\, 2\,\mathbf{Y}^* \cdot \nabla f(\mathbf{Y}^*)$$

or, equivalently, since $\mathbf{Y}^* = \mathbf{X}^*$,

$$\mathbf{L}\,\mathbf{v} \cdot \nabla^2 f(\mathbf{X}^*)\,\mathbf{L}\,\mathbf{v} \,<\, 2\,\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*)\,. \tag{5.2}$$

Next, since $\mathbf{L} = (\boldsymbol{\ell}\,\boldsymbol{\ell}^T)^{1/2}$ is symmetric and nonsingular, $\mathbf{v} = \mathbf{L}\mathbf{w}$ for some for some $\mathbf{w} \in \mathbb{R}^m$. In addition,

$$\|\boldsymbol{\ell}^T\mathbf{w}\,\|^2 = \|\boldsymbol{\ell}^T(\boldsymbol{\ell}\,\boldsymbol{\ell}^T)^{-1/2}\mathbf{v}\,\|^2 \,=\, \|\mathbf{v}\|^2 \,=\, 1\,.$$

So letting $\mathbf{u} := \boldsymbol{\ell}^T\mathbf{w}$, observe that

$$\begin{aligned}
\boldsymbol{\ell}\,\mathbf{u} \cdot \nabla f(\mathbf{X}^*) \;&=\; \mathbf{w}^T\boldsymbol{\ell}\boldsymbol{\ell}^T\,\nabla f(\mathbf{X}^*) \,=\, \mathbf{w}^T\mathbf{L}\,\mathbf{L}\,\nabla f(\mathbf{X}^*)\\
&=\; \mathbf{v} \cdot \mathbf{L}\,\nabla f(\mathbf{Y}^*) = \mathbf{L}\,\mathbf{v} \cdot \nabla f(\mathbf{Y}^*) \,=\, 0\,.
\end{aligned}$$

In addition, since $\mathbf{L}\,\mathbf{v} = \mathbf{L}\,\mathbf{L}\,\mathbf{w} = \boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{w} = \boldsymbol{\ell}\,\mathbf{u}$, (5.2) reads

$$\boldsymbol{\ell}\,\mathbf{u} \cdot \nabla^2 f(\mathbf{X}^*)\,\boldsymbol{\ell}\,\mathbf{u} \,<\, 2\,\mathbf{X}^* \cdot \nabla f(\mathbf{X}^*)\,,$$

and therefore we have exhibited $\mathbf{u} \in (\mathbf{x}^*)^{\perp}$ which violates (2.10).

If we now replace **(SONC)** by **(SOSC)** , the same conclusion indeed holds and with exactly same above arguments.

(ii) We show that (2.10) holds by contradiction. Let $\mathbf{y}^* \in \mathbb{S}^{m-1}$ be as in the statement of (ii). Then $\mathbf{x}^* \in \mathbb{S}^{n-1}$ as in Theorem 2.5(ii) satisfies **(FONC)** and $\mathbf{X}^* = \mathbf{Y}^*$. Observe that from (2.10)-(2.9)

$$(\mathbf{x}^*)^{\perp} \,=\, \{\,(\mathbf{a} + \boldsymbol{\ell}^T\mathbf{b}) :\, \mathbf{a} \in \mathrm{Ker}(\boldsymbol{\ell})\,; \quad \mathcal{L}\,\mathbf{b} \cdot \nabla f(\mathbf{X}^*) \,=\, 0\,;\, \|\mathbf{a} + \boldsymbol{\ell}^T\mathbf{b}\| = 1\,\}\,.$$

Suppose that there exists $(\mathbf{a}+\boldsymbol{\ell}^T\mathbf{b}) \in (\mathbf{x}^*)^\perp$ such that (2.10) is violated, i.e.,

$$\boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{b}\cdot\nabla^2 f(\mathbf{X}^*)\,\boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{b}\,<\,2\,\mathbf{X}^*\cdot\nabla f(\mathbf{X}^*)\,(\,\leq 0\,).$$

Next, by orthogonality $1 = \|\mathbf{a}+\boldsymbol{\ell}^T\mathbf{b}\|^2 = \|\mathbf{a}\|^2+\|\boldsymbol{\ell}^T\,\mathbf{b}\|^2$, so that $\|\boldsymbol{\ell}^T\mathbf{b}\| \leq 1$. So let $\mathbf{L}\,\mathbf{v} := \gamma\,\boldsymbol{\ell}\,\boldsymbol{\ell}^T\mathbf{b}$ with $\gamma = 1/\|\boldsymbol{\ell}^T\mathbf{b}\| \geq 1$. Hence (recalling $\boldsymbol{\ell}\,\boldsymbol{\ell}^T = \mathbf{L}^2$)

$$\|\mathbf{v}\|^2 \,=\, \gamma^2\mathbf{b}^T\boldsymbol{\ell}\,\boldsymbol{\ell}^T\mathbf{L}^{-1}\mathbf{L}^{-1}\boldsymbol{\ell}\,\boldsymbol{\ell}^T\mathbf{b} \,=\, \gamma^2\mathbf{b}^T\boldsymbol{\ell}\,\boldsymbol{\ell}^T\mathbf{b} \,=\, \gamma^2\|\boldsymbol{\ell}^T\mathbf{b}\|^2 \,=\, 1\,.$$

In addition

$$\begin{aligned}\mathbf{L}\,\mathbf{v}\cdot\nabla^2 f(\mathbf{Y}^*)\,\mathbf{L}\,\mathbf{v} &= \gamma^2\,\boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{b}\cdot\nabla^2 f(\mathbf{X}^*)\,\boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{b}\\ &\leq \boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{b}\cdot\nabla^2 f(\mathbf{X}^*)\,\boldsymbol{\ell}\,\boldsymbol{\ell}^T\,\mathbf{b}\quad[\text{as } \gamma^2 \geq 1]\\ &< 2\,\mathbf{X}^*\cdot\nabla f(\mathbf{X}^*) \,=\, 2\,\mathbf{Y}^*\cdot\nabla f(\mathbf{Y}^*)\,,\end{aligned}$$

and

$$\mathbf{L}\,\mathbf{v}\cdot\nabla f(\mathbf{Y}^*) \,=\, \gamma\,\mathcal{L}\,\mathbf{b}\cdot\nabla f(\mathbf{Y}^*) \,=\, \gamma\,\mathcal{L}\,\mathbf{b}\cdot\nabla f(\mathbf{X}^*) \,=\, 0,$$

so that $\mathbf{v} \in (\mathbf{y}^*)^\perp$. Therefore we have exhibited a contradiction with $\mathbf{y}^*$ satisfying **(SONC)** and so (2.10) holds.

The same arguments are valid when **(SONC)** is replaced with **(SOSC)** .

□

## References


[1] A.A Ahmadi, A. Olshevsky, P.A. Parrilo, J.N. Tsitsiklis. NP-hardness of deciding convexity of quartic polynomials and related problems. Math. Program. 137, pp. 453–476, 2013.

[2] Baldoni, V., Berline, N., De Loera, J.A., Kppe, M., Vergne, M.: How to integrate a polynomial over a simplex. Math. Comput. 80, 297Ð325 (2011)

[3] 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.

[4] F. John. Extremum problems with inequalities as side constraints, In *Studies and Essays, Courant Anniversary*, K.O. Friedrichs, O.E. Neugebauer and J. J. Stoker, eds.), pp. 187Ð204, Wiley (interscience), New York, 1948.

[5] M.X. Goemans, D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, Journal of the ACM (JACM), 42, 1115–1145, 1995.

[6] E. de Kerk, M. Laurent. Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. *Math. Program.*, 2020. doi.org/10.1007/s10107-019-01465-1

[7] A. C. Doherty, S. Wehner. Convergence of SDP hierarchies for polynomial optimization on the hypersphere. `arXiv:1210.5048v2, 2013.`

[8] K. Fang, H. Fawzi. The sum-of-squares hierarchy on the sphere, and applications in quantum information theory. *Math. Program.*, 2020. `arXiv: 1908.05155`, 2019.

[9] J.B. Lasserre. Homogeneous polynomials and spurious local minima on the unit sphere. `arXiv:2010.07066`, 2020. Submitted.

[10] J.B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM J. Optimization 11, pp 796–817, 2001.

[11] J.B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity, SIAM J. Optim. 17, pp. 822–843, 2006

[12] J.B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM J. Optim. 2(3)*, pp. 864–885, 2011.

[13] Jean B. Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge University Press, 2015.

[14] Lasserre J.B. (2019) The Moment-SOS Hierarchy, In *Proceedings of the International Congress of Mathematicians* (ICM 2018), vol 4, B. Sirakov, P. Ney de Souza and M. Viana (Eds.), World Scientific, 2019, pp. 3773–3794

[15] Trung Thanh Nguyen, K. Elbassioni. A PTAS for class of binary nonlinear programs with low-rank functions, *Oper. Res. Letters* 2021. DOI: 10.1016/j.orl.2021.06.017

[16] 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.

[17] L. Slot, M. Laurent. Near-optimal analysis of Lasserre's univariate measure-based bounds for multivariate polynomial optimization. *Math. Program.*, 2020

[18] B. Sturmfels. *Solving Systems of Polynomial; Equations*, CBMS Regional Conference Series in Mathematics, vol 97, AMS, Providence, 2002.

[19] H. Waki, S. Kim, M. Kojima, M. Muramatsu. Sums of Squares and Semidefinite Programming Relaxations for Polynomial Optimization Problems with Structured Sparsity, SIAM J. Optim. 17(1), pp. 218–242, 2006 ?

[20] J. Wang, V. Magron, J.B. Lasserre. Chordal-TSSOS: A Moment-SOS hierarchy that exploits term sparsity with chordal extension, SIAM J. Optim 31(1), pp. 114–141, 2021

LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4, France, email: lasserre@laas.fr