

# OPTIMALITY CONDITIONS FOR MINIMIZERS AT INFINITY IN POLYNOMIAL PROGRAMMING

TIẾN-SON PHẠM

ABSTRACT. In this paper we study necessary optimality conditions for the optimization problem

$$\text{infimum } f_0(x) \quad \text{subject to } x \in S,$$

where  $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$  is a polynomial function and  $S \subset \mathbb{R}^n$  is a set defined by polynomial inequalities. Assume that the problem is bounded below and has the Mangasarian–Fromovitz property at infinity. We first show that if the problem does *not* have an optimal solution, then a version at infinity of the Fritz–John optimality conditions holds. From this we derive a version at infinity of the Karush–Kuhn–Tucker optimality conditions. As applications, we obtain a Frank–Wolfe type theorem which states that the optimal solution set of the problem is nonempty provided the objective function  $f_0$  is convenient. Finally, in the unconstrained case, we show that the optimal value of the problem is the smallest critical value of some polynomial. All the results are presented in terms of the Newton polyhedra of the polynomials defining the problem.

## 1. INTRODUCTION

Optimality conditions form the foundations of mathematical programming both theoretically and computationally. There is a large literature on all aspects of optimality conditions. We refer the reader to the classical papers [15, 16, 17, 21] and to the comprehensive monographs [6, 9, 19, 23] with the references therein.

In this paper, we are interested in necessary optimality conditions to polynomial optimization problems *whose solution sets are empty*. More precisely, let  $f_0, f_1, \dots, f_p: \mathbb{R}^n \rightarrow \mathbb{R}$  be polynomial functions and set

$$S := \{x \in \mathbb{R}^n : f_1(x) \leq 0, \dots, f_p(x) \leq 0\}.$$

---

*Date:* June 30, 2017.

1991 *Mathematics Subject Classification.* 90C46 · 90C26 · 90C30.

*Key words and phrases.* Existence of minimizers, Fermat theorem, Frank–Wolfe theorem, Fritz–John optimality conditions, Karush–Kuhn–Tucker optimality conditions, Mangasarian–Fromovitz constraint qualification, Newton polyhedron, polynomial programming.

The author was partially supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED), grant 101.04-2016.05.

Assume that  $S \neq \emptyset$  and the restriction of  $f_0$  on  $S$  is bounded from below. Consider the optimization problem

$$f^* := \inf f_0(x) \quad \text{such that} \quad x \in S. \quad (\text{P})$$

We first assume that  $f_0$  attains its infimum on  $S$ , i.e.,  $f^* = f_0(x^*)$  for some  $x^* \in S$ . Thanks to the Fritz-John necessary optimality conditions [15], there exist nonnegative real numbers  $\lambda_i^*, i = 0, 1, \dots, p$ , not all zero, such that

$$\begin{aligned} \lambda_0^* \nabla f_0(x^*) + \sum_{i=1}^p \lambda_i^* \nabla f_i(x^*) &= 0, \\ \lambda_i^* f_i(x^*) &= 0 \quad \text{for } i = 1, \dots, p. \end{aligned}$$

Here and in the following,  $\nabla f_i(x)$  denotes the gradient vector of  $f_i$  at  $x$ . Furthermore, if the Mangasarian–Fromovitz constraint qualification [21] holds at  $x^*$ : there exists a vector  $v \in \mathbb{R}^n$  such that

$$\langle \nabla f_i(x^*), v \rangle < 0 \quad \text{for all } i \geq 1 \quad \text{with} \quad f_i(x^*) = 0,$$

then we may obtain the more informative optimality conditions due to Karush [16], Kuhn and Tucker [17] where the real number  $\lambda_0^*$  can be taken to be 1. Note that the Mangasarian–Fromovitz constraint qualification is generally satisfied; see the very recent paper by Bolte, Hochart, and Pauwels [8].

We now assume that  $f_0$  does not attain its optimal value  $f^*$  on  $S$ . As far as we know, there are no results which are similar to the two necessary optimality conditions mentioned above. The purpose of this paper is to fill this gap for polynomial programs. Indeed, under the assumption that the considered problem has the Mangasarian–Fromovitz property at infinity (see Definition 2.2 in the next section), we will show that either  $f^* = 0$  or there exist a nonempty set  $J \subset \{1, \dots, n\}$ , a point  $x^* \in \mathbb{R}^n$ , and scalars  $\lambda_0^*, \lambda_1^*, \dots, \lambda_p^*$  satisfying the following conditions:

- (i)  $x_j^* = 0$  if, and only if,  $j \notin J$ ;
- (ii)  $f^* = f_{0, \Delta_0}(x^*)$ ;
- (iii)  $\lambda_0^* \nabla f_{0, \Delta_0}(x^*) + \sum_{i=1}^p \lambda_i^* \nabla f_{i, \Delta_i}(x^*) = 0$ ;
- (iv)  $f_{i, \Delta_i}(x^*) \leq 0$  and  $\lambda_i^* f_{i, \Delta_i}(x^*) = 0$  for  $i = 1, \dots, p$ ;
- (v) the numbers  $\lambda_i^*, i = 0, 1, \dots, p$ , are nonnegative and not all zero;

where  $f_{i, \Delta_i}, i = 0, \dots, p$ , are polynomials which correspond to some faces  $\Delta_i$  of the Newton polyhedra at infinity of  $f_i$ . Moreover, if the following constraint qualification holds: there exists a vector  $v \in \mathbb{R}^n$  such that

$$\langle \nabla f_{i, \Delta_i}(x^*), v \rangle < 0 \quad \text{for all } i \geq 1 \quad \text{with} \quad f_{i, \Delta_i}(x^*) = 0 \quad \text{and} \quad f_i|_{\mathbb{R}^J} \neq \text{constant}.$$

then we can take  $\lambda_0^* = 1$ .

In view of these results, we can say that  $x^*$  is a “minimizer at infinity” of the problem (P).

Next, we study the existence of optimal solutions to polynomial optimization problems. It is well-known that for linear programming, a bounded feasible problem always has an optimal solution. This property is remarkable, and fails to hold for general nonlinear programs. Frank and Wolfe [12] showed that when the objective function is quadratic and the feasible region is linear, the set of optimal solutions is nonempty provided the problem is bounded below. Many other authors generalized the Frank–Wolfe theorem to broader classes of functions. For example, Perold [27] extended the Frank–Wolfe theorem to a class of non-quadratic objective functions and linear constraints. Andronov, Belousov, and Shironin [1] generalized the Frank–Wolfe theorem to the case of a cubic polynomial objective function under linear constraints. Luo and Zhang [20] (see also [29]) extended the Frank–Wolfe theorem to various classes of general convex or non-convex quadratic constraint systems. Belousov and Klatte in [4] (see also [2, 3, 7, 26]) generalized the result on attainability to convex polynomial programs. Very recently, Đinh, Hà and the author [10] extended the Frank–Wolfe theorem for non-degenerate polynomial programs.

As a corollary of our theorem 3.1, we readily establish the attainability of the infimum (assumed to be finite) of the problem (P) when the objective function  $f_0$  is convenient and the considered problem has the Mangasarian–Fromovitz property at infinity; this improves the main result in the paper [10].

Finally, in the unconstrained case (i.e.,  $S = \mathbb{R}^n$ ), we show that the optimal value of the problem (P) is the smallest critical value of some polynomial. This property is useful because, using Theorem 9 in the paper of Nie, Demmel, and Sturmfels [25], it allows us to construct a sequence of semidefinite programming (SDP) relaxations whose optimal values converge monotonically, increasing to the optimal value of the original problem.

All the obtained results are presented in terms of the Newton polyhedra of  $f_0$  and the polynomials defining  $S$ . These results, together with those in [10, 11, 13, 14, 28]), show that many interesting properties in polynomial programming can be obtained from the geometry of Newton polyhedra.

The paper is structured as follows. In Section 2, we recall some notations, definitions and preliminary facts which are used throughout this paper. Optimality conditions and the Frank–Wolfe theorem for polynomial programs are given in Section 3. Further properties for the unconstrained case are given in Section 4.

## 2. PRELIMINARIES

Throughout this paper,  $\mathbb{R}^n$  denotes the Euclidean space of dimension  $n$ . The corresponding inner product (resp., norm) in  $\mathbb{R}^n$  is defined by  $\langle x, y \rangle$  for any  $x, y \in \mathbb{R}^n$  (resp.,

$\|x\| := \sqrt{\langle x, x \rangle}$  for any  $x \in \mathbb{R}^n$ ). Given a nonempty set  $J \subset \{1, \dots, n\}$ , we define

$$\mathbb{R}^J := \{x \in \mathbb{R}^n : x_j = 0, \text{ for all } j \notin J\}.$$

We denote by  $\mathbb{Z}_+$  the set of non-negative integer numbers. If  $\kappa = (\kappa_1, \dots, \kappa_n) \in \mathbb{Z}_+^n$ , we denote by  $x^\kappa$  the monomial  $x_1^{\kappa_1} \cdots x_n^{\kappa_n}$ .

**2.1. Newton polyhedra and non-degeneracy conditions.** Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be a polynomial function. Suppose that  $f$  is written as  $f = \sum_{\kappa} a_{\kappa} x^{\kappa}$ . Then the support of  $f$ , denoted by  $\text{supp}(f)$ , is defined as the set of those  $\kappa \in \mathbb{Z}_+^n$  such that  $a_{\kappa} \neq 0$ . The *Newton polyhedron (at infinity)* of  $f$ , denoted by  $\Gamma(f)$ , is defined as the convex hull in  $\mathbb{R}^n$  of the set  $\text{supp}(f)$ .<sup>1</sup> The polynomial  $f$  is said to be *convenient* if  $\Gamma(f)$  intersects each coordinate axis in a point different from the origin 0 in  $\mathbb{R}^n$ , that is, if for any  $j \in \{1, \dots, n\}$  there exists some  $\kappa_j > 0$  such that  $\kappa_j e^j \in \Gamma(f)$ , where  $\{e^1, \dots, e^n\}$  denotes the canonical basis in  $\mathbb{R}^n$ . For each (closed) face  $\Delta$  of  $\Gamma(f)$ , we will denote by  $f_{\Delta}$  the polynomial  $\sum_{\kappa \in \Delta} a_{\kappa} x^{\kappa}$ ; if  $\Delta \cap \text{supp}(f) = \emptyset$  we let  $f_{\Delta} := 0$ .

Given a nonzero vector  $q \in \mathbb{R}^n$ , we define

$$d(q, \Gamma(f)) := \min\{\langle q, \kappa \rangle : \kappa \in \Gamma(f)\},$$

$$\Delta(q, \Gamma(f)) := \{\kappa \in \Gamma(f) : \langle q, \kappa \rangle = d(q, \Gamma(f))\}.$$

By definition, for each nonzero vector  $q \in \mathbb{R}^n$ ,  $\Delta(q, \Gamma(f))$  is a closed face of  $\Gamma(f)$ . Conversely, if  $\Delta$  is a closed face of  $\Gamma(f)$  then there exists a nonzero vector<sup>2</sup>  $q \in \mathbb{R}^n$  such that  $\Delta = \Delta(q, \Gamma(f))$ . The *Newton boundary (at infinity)* of  $f$ , denoted by  $\Gamma_{\infty}(f)$ , is defined as the union of all closed faces  $\Delta(q, \Gamma(f))$  for some  $q \in \mathbb{R}^n$  with  $d(q, \Gamma(f)) < 0$  (and so  $\min_{j=1, \dots, n} q_j < 0$ ).

Following Némethi and Zaharia [24] we say that a closed face  $\Delta$  of  $\Gamma(f)$  is *bad* if the following two conditions hold:

- (i) the affine subvariety of dimension =  $\dim(\Delta)$  spanned by  $\Delta$  contains the origin, and
- (ii) there exists a hyperplane  $H \subset \mathbb{R}^n$  with equation  $q_1 \kappa_1 + q_2 \kappa_2 + \cdots + q_n \kappa_n = 0$ , where  $\kappa_1, \kappa_2, \dots, \kappa_n$  are the coordinates in  $\mathbb{R}^n$ , such that:
  - (ii<sub>a</sub>) there exist  $i$  and  $j$  with  $q_i \cdot q_j < 0$ , and
  - (ii<sub>b</sub>)  $H \cap \Gamma(f) = \Delta$ .

More geometrically, the condition (ii<sub>a</sub>) says that the hyperplane  $H$  intersects the interior of the positive octant of  $\mathbb{R}^n$ .

<sup>1</sup>Note that we do not include the origin in the definition of the Newton polyhedron at infinity  $\Gamma(f)$ .

<sup>2</sup>Since  $\Gamma(f)$  is an integer polyhedron, we can assume that all the coordinates of  $q$  are rational numbers.

**Remark 2.1.** The following statements follow immediately from definitions:

(i) For each nonempty subset  $J$  of  $\{1, \dots, n\}$ , if the restriction of  $f$  on  $\mathbb{R}^J$  is not constant, then  $\Gamma(f) \cap \mathbb{R}^J = \Gamma(f|_{\mathbb{R}^J})$ .

(ii) If  $f$  is convenient, then there is no bad face in  $\Gamma(f)$  and, moreover,  $\Gamma(f) \cap \mathbb{R}^J = \Gamma(f|_{\mathbb{R}^J})$  for all nonempty subset  $J$  of  $\{1, \dots, n\}$ .

(iii) Let  $\Delta := \Delta(q, \Gamma(f))$  for some nonzero vector  $q := (q_1, \dots, q_n) \in \mathbb{R}^n$ . By definition,  $f_\Delta = \sum_{\kappa \in \Delta} a_\kappa x^\kappa$  is a weighted homogeneous polynomial of type  $(q, d := d(q, \Gamma(f)))$ , i.e., we have for all  $t > 0$  and all  $x \in \mathbb{R}^n$ ,

$$f_\Delta(t^{q_1} x_1, \dots, t^{q_n} x_n) = t^d f_\Delta(x_1, \dots, x_n).$$

This implies the Euler relation

$$\sum_{j=1}^n q_j x_j \frac{\partial f_\Delta}{\partial x_j}(x) = d f_\Delta(x). \quad (1)$$

In particular, if  $d \neq 0$  and  $\nabla f_\Delta(x) = 0$ , then  $f_\Delta(x) = 0$ .

The following notion (see [18]) will play an important role in Section 4.

**Definition 2.1.** We say that  $f$  is (*Kouchnirenko*) *nondegenerate at infinity* if, and only if, for all faces  $\Delta \in \Gamma_\infty(f)$ , the system of equations

$$\frac{\partial f_\Delta}{\partial x_1}(x) = \dots = \frac{\partial f_\Delta}{\partial x_n}(x) = 0$$

has no solution in  $(\mathbb{R} \setminus \{0\})^n$ .

**Remark 2.2.** It is worth emphasizing that the condition of non-degeneracy at infinity is a generic property in the sense that it holds in an open and dense semialgebraic set of the entire space of input data (see, for example, [14]).

The following definition is inspired from the work of Mangasarian and Fromovitz [21].

**Definition 2.2.** Let  $f_0, f_1, \dots, f_p: \mathbb{R}^n \rightarrow \mathbb{R}$  be polynomial functions and set

$$S := \{x \in \mathbb{R}^n : f_1(x) \leq 0, \dots, f_p(x) \leq 0\}.$$

We say that the *Mangasarian–Fromovitz property at infinity* ( $(MF)_\infty$  for short) holds for the problem  $\inf_{x \in S} f_0(x)$  if, and only if, for every nonempty set  $J \subset \{1, \dots, n\}$ , for every vector  $q \in \mathbb{R}^n$ , and for every  $x \in \mathbb{R}^n$  satisfying the conditions

- (i)  $d(q, \Gamma(f_0)) < 0$ ,
- (ii)  $\Delta_i := \Delta(q, \Gamma(f_i)) \subset \mathbb{R}^J$  for all  $i \in I := \{i \in \{0, 1, \dots, p\} : f_i|_{\mathbb{R}^J} \neq \text{constant}\}$ ,
- (iii)  $x_j = 0 \iff j \notin J$ ,
- (iv)  $f_{i, \Delta_i}(x) \leq 0$  for  $i = 1, \dots, p$ ,

there exists a nonzero vector  $v \in \mathbb{R}^n$  such that

$$\langle \nabla f_{i,\Delta_i}(x), v \rangle < 0 \quad \text{for all } i \in I \text{ with } f_{i,\Delta_i}(x) = 0.$$

Note that the  $(\text{MF})_\infty$  property in the above definition is not a constraint qualification since it involves the objective function  $f_0$ .

As shown in the next lemma, Definitions 2.1 and 2.2 are equivalent in the unconstrained case.

**Lemma 2.1.** *Let  $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$  be a polynomial function. Then  $f_0$  is nondegenerate at infinity if, and only if, the problem  $\inf_{x \in \mathbb{R}^n} f_0(x)$  has the  $(\text{MF})_\infty$  property.*

*Proof.*  $\Rightarrow$ . Take arbitrary a nonempty set  $J \subset \{1, \dots, n\}$ , a nonzero vector  $q \in \mathbb{R}^n$ , and a point  $x^0 \in \mathbb{R}^n$  such that the following conditions hold:

- (a)  $d_0 := d(q, \Gamma(f_0)) < 0$ ,
- (b)  $\Delta_0 := \Delta(q, \Gamma(f_0)) \subset \mathbb{R}^J$ ,  $f_0|_{\mathbb{R}^J} \neq \text{constant}$ ,
- (c)  $x_j^0 = 0 \iff j \notin J$ ,
- (d)  $f_{0,\Delta_0}(x^0) = 0$ .

By definition,  $\Delta_0 \in \Gamma_\infty(f_0)$  and  $f_{0,\Delta_0}$  does not depend on  $x_j$  for  $j \notin J$ . The nondegeneracy assumption implies that there exists  $j^* \in J$  such that  $\frac{\partial f_{0,\Delta_0}}{\partial x_{j^*}}(x^0) \neq 0$ . Let  $v := (v_1, \dots, v_n) \in \mathbb{R}^n$  with

$$v_j := \begin{cases} -\left(\frac{\partial f_{0,\Delta_0}}{\partial x_{j^*}}(x^0)\right)^{-1} & \text{if } j = j^*, \\ 0 & \text{otherwise.} \end{cases}$$

We have  $\langle \nabla f_{0,\Delta_0}(x^0), v \rangle = -1 < 0$ , and so the problem  $\inf_{x \in \mathbb{R}^n} f_0(x)$  has the  $(\text{MF})_\infty$  property.

$\Leftarrow$ . By contradiction, assume that there exist  $\Delta_0 \in \Gamma_\infty(f_0)$  and  $x^0 \in (\mathbb{R} \setminus \{0\})^n$  such that  $\nabla f_{0,\Delta_0}(x^0) = 0$ . By definition, there exists a vector  $q \in \mathbb{R}^n$  such that  $\Delta_0 = \Delta(q, \Gamma(f_0))$  and  $d_0 := d(q, \Gamma(f_0)) < 0$ . Remark 2.1(iii) now leads to  $f_{0,\Delta_0}(x^0) = 0$ . Let  $J$  be the smallest subset of  $\{1, \dots, n\}$  such that the space  $\mathbb{R}^J$  contains  $\Delta_0$ . We have for all  $\kappa \in \Delta_0 \subset \mathbb{R}^J$ ,

$$\sum_{j \in J} q_j \kappa_j = \sum_{j=1}^n q_j \kappa_j = d_0 < 0.$$

It turns out that  $\Gamma(f_0) \cap \mathbb{R}^J$  is nonempty and different from  $\{0\}$ . Consequently,  $f_0|_{\mathbb{R}^J}$  is not constant. Let  $y^0 := (y_1^0, \dots, y_n^0) \in \mathbb{R}^n$ , where

$$y_j^0 := \begin{cases} x_j^0 & \text{if } j \in J, \\ 0 & \text{otherwise.} \end{cases}$$

Since  $f_{0,\Delta_0}$  does not depend on  $x_j$  for  $j \notin J$ , we get

$$f_{0,\Delta_0}(y^0) = f_{0,\Delta_0}(x^0) = 0 \quad \text{and} \quad \nabla f_{0,\Delta_0}(y^0) = \nabla f_{0,\Delta_0}(x^0) = 0.$$

Combining these facts with the  $(\text{MF})_\infty$  property, we obtain the absurd conclusion:

$$0 = \langle \nabla f_{0,\Delta_0}(y^0), v \rangle < 0$$

for some vector  $v \in \mathbb{R}^n$ . □

**2.2. Semi-algebraic geometry.** This subsection contains some background material on semi-algebraic geometry and preliminary results which will be used later. We give only concise definitions and results that will be needed in the paper. For more detailed information on the subject, see, for example, [5] and [14, Chapter 1].

**Definition 2.3.** (i) A subset of  $\mathbb{R}^n$  is called *semi-algebraic* if it is a finite union of sets of the form

$$\{x \in \mathbb{R}^n \mid f_i(x) = 0, i = 1, \dots, k; f_i(x) > 0, i = k + 1, \dots, p\}$$

where all  $f_i$  are polynomials.

(ii) Let  $A \subset \mathbb{R}^n$  and  $B \subset \mathbb{R}^m$  be semi-algebraic sets. A map  $F: A \rightarrow B$  is said to be *semi-algebraic* if its graph

$$\{(x, y) \in A \times B \mid y = F(x)\}$$

is a semi-algebraic subset in  $\mathbb{R}^n \times \mathbb{R}^m$ .

The class of semi-algebraic sets is closed under taking finite intersections, finite unions and complements; a Cartesian product of semi-algebraic sets is a semi-algebraic set. Moreover, a major fact concerning the class of semi-algebraic sets is its stability under linear projections; in particular, the closure and the interior of a semi-algebraic set are semi-algebraic sets.

In the sequel, we will need the following useful results (see, for example, [14, Chapter 1]).

**Lemma 2.2** (Curve Selection Lemma at infinity). *Let  $A \subset \mathbb{R}^n$  be a semi-algebraic set, and let  $f := (f_1, \dots, f_p): \mathbb{R}^n \rightarrow \mathbb{R}^p$  be a semi-algebraic map. Assume that there exists a sequence  $\{x^\ell\}$  such that  $x^\ell \in A$ ,  $\lim_{\ell \rightarrow \infty} \|x^\ell\| = \infty$  and  $\lim_{\ell \rightarrow \infty} f(x^\ell) = y \in (\overline{\mathbb{R}})^p$ , where  $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ . Then there exists a smooth semi-algebraic curve  $\varphi: (0, \epsilon) \rightarrow \mathbb{R}^n$  such that  $\varphi(t) \in A$  for all  $t \in (0, \epsilon)$ ,  $\lim_{t \rightarrow 0} \|\varphi(t)\| = \infty$ , and  $\lim_{t \rightarrow 0} f(\varphi(t)) = y$ .*

**Lemma 2.3** (Growth Dichotomy Lemma). *Let  $f: (0, \epsilon) \rightarrow \mathbb{R}$  be a semi-algebraic function with  $f(t) \neq 0$  for all  $t \in (0, \epsilon)$ . Then there exist constants  $c \neq 0$  and  $q \in \mathbb{Q}$  such that  $f(t) = ct^q + o(t^q)$  as  $t \rightarrow 0^+$ .*

**Lemma 2.4** (Monotonicity Lemma). *Let  $a < b$  in  $\mathbb{R}$ . If  $f: [a, b] \rightarrow \mathbb{R}$  is a semi-algebraic function, then there is a partition  $a =: t_1 < \dots < t_N := b$  of  $[a, b]$  such that  $f|_{(t_l, t_{l+1})}$  is  $C^1$ , and either constant or strictly monotone, for  $l \in \{1, \dots, N-1\}$ .*

### 3. THE CONSTRAINED CASE

From now on we let  $f_0, f_1, \dots, f_p: \mathbb{R}^n \rightarrow \mathbb{R}$  be nonconstant polynomial functions and set

$$S := \{x \in \mathbb{R}^n : f_1(x) \leq 0, \dots, f_p(x) \leq 0\}.$$

We will assume that  $S \neq \emptyset$  and  $f_0$  is bounded from below on  $S$ . Consider the problem (P) formulated in the introduction section:

$$f^* := \inf f_0(x) \quad \text{such that} \quad x \in S. \quad (\text{P})$$

The main result of this paper is the following theorem, which is a version at infinity of the Fritz-John and Karush–Kuhn–Tucker optimality conditions.

**Theorem 3.1** (Optimality conditions for minimizers at infinity). *Assume that the  $(\text{MF})_\infty$  property holds for the problem (P). If  $f_0$  does not attain its infimum  $f^*$  on  $S$ , then either  $f^* = 0$  or there exist a nonempty set  $J \subset \{1, \dots, n\}$ , a vector  $q \in \mathbb{R}^n$  with  $\min_{j \in J} q_j < 0$ , a point  $x^* \in \mathbb{R}^n$ , and scalars  $\lambda_0^*, \lambda_1^*, \dots, \lambda_p^*$  such that the following conditions hold:*

- (i)  $x_j^* = 0$  if, and only if,  $j \notin J$ ;
- (ii)  $f^* = f_{0, \Delta_0}(x^*)$ ,  $\Delta_0 := \Delta(q, \Gamma(f_0)) \subset \mathbb{R}^J$ , and  $d(q, \Gamma(f_0)) = 0$ ;
- (iii)  $\lambda_0^* \nabla f_{0, \Delta_0}(x^*) + \sum_{i=1}^p \lambda_i^* \nabla f_{i, \Delta_i}(x^*) = 0$ ;
- (iv)  $f_{i, \Delta_i}(x^*) \leq 0$  and  $\lambda_i^* f_{i, \Delta_i}(x^*) = 0$  for  $i = 1, \dots, p$ ;
- (v) the numbers  $\lambda_i^*, i = 0, 1, \dots, p$ , are nonnegative and not all zero;

where  $\Delta_i := \Delta(q, \Gamma(f_i))$ . Moreover, we can take  $\lambda_0^* = 1$  provided the following constraint qualification holds: there exists a nonzero vector  $v \in \mathbb{R}^n$  such that

$$\langle \nabla f_{i, \Delta_i}(x^*), v \rangle < 0 \quad \text{for all } i \geq 1 \quad \text{with } f_{i, \Delta_i}(x^*) = 0 \quad \text{and } f_i|_{\mathbb{R}^J} \neq \text{constant}.$$

*Proof.* Since  $f_0$  does not attain its infimum  $f^*$  on  $S$ , there exists a sequence  $\{a^\ell\}_{\ell \geq 1} \subset S$  such that

$$\lim_{\ell \rightarrow \infty} \|a^\ell\| = +\infty \quad \text{and} \quad \lim_{\ell \rightarrow \infty} f_0(a^\ell) = f^*.$$

For each  $\ell \geq 1$ , we consider the problem

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, p, \quad \|x\|^2 = \|a^\ell\|^2. \end{aligned}$$



Since the objective function  $f_0$  is continuous and the constraint set is compact, by the Weierstrass theorem, an optimal solution  $x^\ell \in S$  of the problem exists. We have

$$\|x^\ell\|^2 = \|a^\ell\|^2 \quad \text{and} \quad f^* \leq f_0(x^\ell) \leq f_0(a^\ell).$$

Hence,

$$\lim_{\ell \rightarrow \infty} \|x^\ell\| = +\infty \quad \text{and} \quad \lim_{\ell \rightarrow \infty} f_0(x^\ell) = f^*.$$

Since the number of all subsets of the set  $\{1, \dots, n\}$  is finite, by passing to a subsequence if necessary, we may assume that  $\{j \in \{1, \dots, n\} : x_j^\ell \neq 0\} = J$  for all  $\ell$  and for some nonempty set  $J \subset \{1, \dots, n\}$ .

Let  $I_1$  be the (possibly empty) set of all indices  $i \in \{1, \dots, p\}$  such that the restriction of  $f_i$  on  $\mathbb{R}^J = \{x \in \mathbb{R}^n : x_j = 0, j \notin J\}$  is not constant. We have for all  $i \in \{1, \dots, p\} \setminus I_1$ ,

$$f_i(x) = f_i(x^\ell) \leq 0 \quad \text{for all } x \in \mathbb{R}^J \text{ and all } \ell \geq 1. \quad (2)$$

Consequently,  $x^\ell$  is also an optimal solution of the problem

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, i \in I_1, \quad h_k(x) = 0, k \notin J, \quad \|x\|^2 = \|a^\ell\|^2, \end{aligned}$$

here and in the following we let  $h_k(x) := x_k$ . The Fritz-John optimality conditions (see, for example, [6]) imply that there exist some real numbers  $\lambda_i^\ell$  for  $i \in I_1 \cup \{0\}$ ,  $\nu_k^\ell$  for  $k \notin J$ , and  $\mu^\ell$  such that the following relations hold:

$$\begin{aligned} & \sum_{i \in I_1 \cup \{0\}} \lambda_i^\ell \nabla f_i(x^\ell) + \sum_{k \notin J} \nu_k^\ell \nabla h_k(x^\ell) + \mu^\ell x^\ell = 0, \\ & \lambda_0^\ell \geq 0, \lambda_i^\ell \geq 0 \quad \text{and} \quad \lambda_i^\ell f_i(x^\ell) = 0 \quad \text{for } i \in I_1, \\ & \|(\lambda_i^\ell, \nu_k^\ell, \mu^\ell)_{i \in I_1 \cup \{0\}, k \notin J}\| = 1. \end{aligned}$$

For simplicity, we write  $\lambda := (\lambda_i)_{i \in I_1 \cup \{0\}} \in \mathbb{R}^{\#I_1+1}$  and  $\nu := (\nu_k)_{k \notin J} \in \mathbb{R}^{n-\#J}$ . Let

$$\begin{aligned} \mathcal{A} := \{ & (x, \lambda, \nu, \mu) \in \mathbb{R}^n \times \mathbb{R}^{\#I_1+1} \times \mathbb{R}^{n-\#J} \times \mathbb{R} : f_i(x) \leq 0 \text{ for } i \in I_1, \\ & x_j \neq 0 \text{ for } j \in J, \quad h_k(x) = 0 \text{ for } k \notin J, \\ & \sum_{i \in I_1 \cup \{0\}} \lambda_i \nabla f_i(x) + \sum_{k \notin J} \nu_k \nabla h_k(x) + \mu x = 0, \\ & \lambda_i f_i(x) = 0 \text{ for } i \in I_1, \quad \lambda_i \geq 0 \text{ for } i \in I_1 \cup \{0\}, \quad \|(\lambda, \nu, \mu)\| = 1\}. \end{aligned}$$

Then  $\mathcal{A}$  is a semi-algebraic set and the sequence  $(x^\ell, \lambda^\ell, \nu^\ell, \mu^\ell) \in \mathcal{A}$  tends to infinity as  $\ell \rightarrow \infty$ . Applying Lemma 2.2 to the semi-algebraic function  $\mathcal{A} \rightarrow \mathbb{R}, (x, \lambda, \nu, \mu) \mapsto f_0(x)$ , we get a smooth semi-algebraic curve

$$(\varphi, \lambda, \nu, \mu): (0, \epsilon) \rightarrow \mathbb{R}^n \times \mathbb{R}^{\#I_1+1} \times \mathbb{R}^{n-\#J} \times \mathbb{R}, \quad t \mapsto (\varphi(t), \lambda(t), \nu(t), \mu(t)),$$

satisfying the following conditions

- (c1)  $\lim_{t \rightarrow 0^+} \|\varphi(t)\| = +\infty$ ;
- (c2)  $\lim_{t \rightarrow 0^+} f_0(\varphi(t)) = f^*$ ;
- (c3)  $f_i(\varphi(t)) \leq 0$  for  $i \in I_1$ ;
- (c4)  $J = \{j \in \{1, \dots, n\} : \varphi_j(t) \not\equiv 0\}$ ;
- (c5)  $h_k(\varphi(t)) \equiv 0$  for  $k \notin J$ ;
- (c6)  $\sum_{i \in I_1 \cup \{0\}} \lambda_i(t) \nabla f_i(\varphi(t)) + \sum_{k \notin J} \nu_k(t) \nabla h_k(\varphi(t)) + \mu(t) \varphi(t) \equiv 0$ ;
- (c7)  $\lambda_i(t) f_i(\varphi(t)) \equiv 0$ , for  $i \in I_1$ ;
- (c8)  $\lambda_i(t) \geq 0$  for  $i \in I_1 \cup \{0\}$ ;
- (c9)  $\|(\lambda(t), \nu(t), \mu(t))\| \equiv 1$ .

Since the (smooth) functions  $\lambda_i$  and  $f_i \circ \varphi$  are semi-algebraic, by shrinking  $\epsilon$  if necessary, we can assume, that these functions are either constant or strictly monotone (see Lemma 2.4). Then, by Condition (c7), one can see for all  $i \in I_1$  that either  $\lambda_i(t) \equiv 0$  or  $f_i \circ \varphi(t) \equiv 0$ . Consequently, we obtain

$$\lambda_i(t) \frac{d}{dt} (f_i \circ \varphi)(t) \equiv 0 \quad \text{for } i \in I_1.$$

Let  $I_2 := \{i \in I_1 : f_i \circ \varphi(t) \equiv 0\}$ . We have  $\lambda_i(t) \equiv 0$  for  $i \in I_1 \setminus I_2$  because of Condition (c7). It follows from Condition (c6) that

$$\begin{aligned} \frac{\mu(t)}{2} \frac{d\|\varphi(t)\|^2}{dt} &= \mu(t) \left\langle \varphi(t), \frac{d\varphi(t)}{dt} \right\rangle \\ &= - \sum_{i \in I_2 \cup \{0\}} \lambda_i(t) \left\langle \nabla f_i(\varphi(t)), \frac{d\varphi(t)}{dt} \right\rangle - \sum_{k \notin J} \nu_k(t) \left\langle \nabla h_k(\varphi(t)), \frac{d\varphi(t)}{dt} \right\rangle \\ &= - \sum_{i \in I_2 \cup \{0\}} \lambda_i(t) \frac{d}{dt} (f_i \circ \varphi)(t) - \sum_{k \notin J} \nu_k(t) \frac{d}{dt} (h_k \circ \varphi)(t) \\ &= -\lambda_0(t) \frac{d}{dt} (f_0 \circ \varphi)(t) - \sum_{k \notin J} \nu_k(t) \frac{d}{dt} (h_k \circ \varphi)(t). \end{aligned}$$

From Condition (c5) one has  $\frac{d}{dt} (h_k \circ \varphi)(t) = 0$  for all  $k \notin J$  and hence

$$\frac{\mu(t)}{2} \frac{d\|\varphi(t)\|^2}{dt} = -\lambda_0(t) \frac{d}{dt} (f_0 \circ \varphi)(t). \quad (3)$$

This, together with Condition (c1), implies that if  $\|(\lambda_i(t))_{i \in I_1 \cup \{0\}}\| = 0$  then  $\mu(t) = \lambda_0(t) = 0$ , and hence, by Condition (c6),

$$\sum_{k \notin J} \nu_k(t) \nabla h_k(\varphi(t)) = 0.$$

Combining this with the definition of  $h_k$  we see that  $\nu_k(t) = 0$  for all  $k \notin J$ , which contradicts Condition (c9). Thus,  $\|(\lambda_i(t))_{i \in I_1 \cup \{0\}}\| \neq 0$ , and so, after a scaling, we can assume that

$$\|(\lambda_i(t))_{i \in I_1 \cup \{0\}}\| = 1 \quad \text{for all } t \in (0, \epsilon). \quad (4)$$

From Condition (c4) we have  $\varphi_j \not\equiv 0$  for all  $j \in J$ . By Lemma 2.3, for each  $j \in J$ , we can expand the coordinate  $\varphi_j$  as follows

$$\varphi_j(t) = x_j^* t^{q_j} + \text{higher order terms in } t,$$

where  $x_j^* \neq 0$  and  $q_j \in \mathbb{Q}$ . From Condition (c1), we get  $\min_{j \in J} q_j < 0$ .

Let  $q_j := M$  for  $j \notin J$  with

$$M \gg \max_{i=0,1,\dots,p} \left\{ \sum_{j \in J} q_j \kappa_j : \kappa \in \Gamma(f_i) \right\}.$$

For each  $i \in \{0, 1, \dots, p\}$ , let  $d_i$  be the minimal value of the linear function  $\sum_{j=1}^n q_j \kappa_j$  on  $\Gamma(f_i)$  and let  $\Delta_i$  be the maximal face of  $\Gamma(f_i)$  (maximal with respect to the inclusion of faces) where the linear function takes this value, i.e.,

$$d_i := d(q, \Gamma(f_i)) \quad \text{and} \quad \Delta_i := \Delta(q, \Gamma(f_i)).$$

Recall that  $\mathbb{R}^J := \{\kappa := (\kappa_1, \dots, \kappa_n) \in \mathbb{R}^n : \kappa_j = 0 \text{ for } j \notin J\}$ . Take any  $i \in I_1 \cup \{0\}$ . Then the restriction of  $f_i$  on  $\mathbb{R}^J$  is not constant, and so  $\Gamma(f_i) \cap \mathbb{R}^J = \Gamma(f_i|_{\mathbb{R}^J})$  is nonempty and different from  $\{0\}$ . Furthermore, by definition of the vector  $q$ , one has

$$d_i = d(q, \Gamma(f_i|_{\mathbb{R}^J})) \quad \text{and} \quad \Delta_i = \Delta(q, \Gamma(f_i|_{\mathbb{R}^J})) \subset \mathbb{R}^J. \quad (5)$$

Consequently, for each  $j \notin J$ , the polynomial  $f_{i, \Delta_i}$  does not depend on the variable  $x_j$ . Now suppose that  $f_i$  is written as  $f_i(x) = \sum_{\kappa} a_{i, \kappa} x^{\kappa}$ . Then

$$\begin{aligned} f_i(\varphi(t)) &= \sum_{\kappa \in \Gamma(f_i) \cap \mathbb{R}^J} a_{i, \kappa} (\varphi(t))^{\kappa} \\ &= \sum_{\kappa \in \Gamma(f_i) \cap \mathbb{R}^J} a_{i, \kappa} (\varphi_1(t))^{\kappa_1} \dots (\varphi_n(t))^{\kappa_n} \\ &= \sum_{\kappa \in \Gamma(f_i) \cap \mathbb{R}^J} (a_{i, \kappa} (x_1^* t^{q_1})^{\kappa_1} \dots (x_n^* t^{q_n})^{\kappa_n} + \text{higher order terms in } t) \\ &= \sum_{\kappa \in \Gamma(f_i) \cap \mathbb{R}^J} (a_{i, \kappa} (x^*)^{\kappa} t^{\sum_{j \in J} q_j \kappa_j} + \text{higher order terms in } t) \\ &= \sum_{\kappa \in \Delta_i} a_{i, \kappa} (x^*)^{\kappa} t^{d_i} + \text{higher order terms in } t, \end{aligned}$$

where  $x^* := (x_1^*, \dots, x_n^*)$  with  $x_j^* := 0$  for  $j \notin J$ . By definition,  $f_{i, \Delta_i}(x) = \sum_{\kappa \in \Delta_i} a_{i, \kappa} x^{\kappa}$ . Hence

$$f_i(\varphi(t)) = f_{i, \Delta_i}(x^*) t^{d_i} + \text{higher order terms in } t. \quad (6)$$

If  $d_0 > 0$ , then it follows from Condition (c2) that  $f^* = 0$  and the theorem is proved. So, in the rest of the proof, we assume that  $d_0 \leq 0$ . Observe that if  $d_0 < 0$  then  $f_{0, \Delta_0}(x^*) = 0$ .

Therefore,

$$d_0 \leq 0 \quad \text{and} \quad d_0 f_{0,\Delta_0}(x^*) = 0. \quad (7)$$

Furthermore, it follows from (c3), (6) and the definition of the sets  $I_1, I_2$  that

$$f_{i,\Delta_i}(x^*) \leq 0 \quad \text{for all} \quad i \in I_1 \quad \text{and} \quad f_{i,\Delta_i}(x^*) = 0 \quad \text{for all} \quad i \in I_2. \quad (8)$$

Let  $I_3 := \{i \in I_2 \cup \{0\} : \lambda_i \neq 0\}$ . Since  $\lambda_i \equiv 0$  for all  $i \in I_1 \setminus I_2$ , we obtain from (4) that  $I_3 \neq \emptyset$ . For  $i \in I_3$ , expand the coordinate  $\lambda_i$  in terms of the parameter (cf. Lemma 2.3) as follows

$$\lambda_i(t) = \lambda_i^0 t^{\theta_i} + \text{higher order terms in } t,$$

where  $\lambda_i^0 \neq 0$  and  $\theta_i \in \mathbb{Q}$ . By Condition (c8), then  $\lambda_i^0 > 0$ . Furthermore, from (4) one has  $\theta_i \geq 0$  for all  $i \in I_3$  with the equality occurring for some  $i \in I_3$ .

For  $i \in I_3$  and  $j \in J$ , by some similar calculations as with  $f_i(\varphi(t))$ , we have

$$\frac{\partial f_i}{\partial x_j}(\varphi(t)) = \frac{\partial f_{i,\Delta_i}}{\partial x_j}(x^*) t^{d_i - q_j} + \text{higher order terms in } t.$$

Since  $h_k(x) = x_k$  for all  $k \notin J$ , it holds that

$$\frac{\partial h_k}{\partial x_j}(\varphi(t)) \equiv 0 \quad \text{for all} \quad j \in J.$$

Consequently, we have for all  $j \in J$ ,

$$\begin{aligned} \sum_{i \in I_3} \lambda_i(t) \frac{\partial f_i}{\partial x_j}(\varphi(t)) + \sum_{k \notin J} \nu_k(t) \frac{\partial h_k}{\partial x_j}(\varphi(t)) &= \sum_{i \in I_3} \lambda_i(t) \frac{\partial f_i}{\partial x_j}(\varphi(t)) \\ &= \sum_{i \in I_3} \left( \lambda_i^0 \frac{\partial f_{i,\Delta_i}}{\partial x_j}(x^*) t^{d_i + \theta_i - q_j} + \text{higher order terms in } t \right) \\ &= \left( \sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i,\Delta_i}}{\partial x_j}(x^*) \right) t^{\ell - q_j} + \text{higher order terms in } t, \end{aligned} \quad (9)$$

where  $\ell := \min_{i \in I_3} (d_i + \theta_i)$  and  $I_4 := \{i \in I_3 : d_i + \theta_i = \ell\} \neq \emptyset$ .

**Claim.** *We have*

$$\sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i,\Delta_i}}{\partial x_j}(x^*) = 0 \quad \text{for all} \quad j = 1, \dots, n. \quad (10)$$

*Proof.* Indeed, for each  $j \notin J$ , the polynomial  $f_{i,\Delta_i}$  does not depend on the variable  $x_j$ , so  $\frac{\partial f_{i,\Delta_i}}{\partial x_j} \equiv 0$ . Consequently,

$$\sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i,\Delta_i}}{\partial x_j}(x^*) = 0 \quad \text{for all} \quad j \notin J.$$

If  $\mu(t) \equiv 0$  then Condition (c6) and (9) give

$$\sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i, \Delta_i}}{\partial x_j}(x^*) = 0 \quad \text{for all } j \in J,$$

and there is nothing to prove. So assume that  $\mu(t) \not\equiv 0$ . By Lemma 2.3, we may write

$$\mu(t) = \mu^0 t^\delta + \text{higher order terms in } t,$$

where  $\mu^0 \neq 0$  and  $\delta \in \mathbb{Q}$ . Let  $J' := \{j \in J : \ell - q_j = \delta + q_j\}$ . Assume that  $J' \neq \emptyset$ . We have from (c6) and (9) that  $\ell - q_j \leq \delta + q_j$  for all  $j \in J$  and that

$$\sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i, \Delta_i}}{\partial x_j}(x^*) = \begin{cases} -\mu^0 x_j^* & \text{if } j \in J', \\ 0 & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} \sum_{j=1}^n \left( \sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i, \Delta_i}}{\partial x_j}(x^*) \right) q_j x_j^* &= \sum_{j \in J'} \left( \sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i, \Delta_i}}{\partial x_j}(x^*) \right) q_j x_j^* \\ &= \sum_{j \in J'} -q_j \mu^0 (x_j^*)^2 \\ &= -\frac{\ell - \delta}{2} \mu^0 \sum_{j \in J'} (x_j^*)^2. \end{aligned}$$

On the other hand,  $f_{i, \Delta_i}$  is a weighted homogeneous polynomial of type  $(q, d_i)$ . Thus, from the Euler relation (1) we obtain for all  $i \in I_4$ ,

$$\sum_{j=1}^n q_j x_j^* \frac{\partial f_{i, \Delta_i}}{\partial x_j}(x^*) = d_i f_{i, \Delta_i}(x^*) = 0,$$

where the last equality follows from (7) and (8). But  $\ell - \delta \neq 0$  because  $\ell - q_j \leq \delta + q_j$  for all  $j \in J$  and  $\min_{j \in J} q_j < 0$ . Hence, we obtain the absurd equality

$$\begin{aligned} 0 &= \sum_{i \in I_4} \left( \sum_{j=1}^n q_j x_j^* \frac{\partial f_{i, \Delta_i}}{\partial x_j}(x^*) \right) \lambda_i^0 = \sum_{j=1}^n \left( \sum_{i \in I_4} \lambda_i^0 \frac{\partial f_{i, \Delta_i}}{\partial x_j}(x^*) \right) q_j x_j^* \\ &= -\frac{\ell - \delta}{2} \mu^0 \sum_{j \in J'} (x_j^*)^2 \neq 0. \end{aligned}$$

Therefore,  $J' = \emptyset$ . The claim is proved.  $\square$

*Proof of Theorem 3.1 (continued).* Let  $\lambda^* := (\lambda_0^*, \lambda_1^*, \dots, \lambda_p^*) \in \mathbb{R}^{p+1}$ , where

$$\lambda_i^* := \begin{cases} \lambda_i^0 & \text{if } i \in I_4, \\ 0 & \text{otherwise.} \end{cases}$$

Then the real numbers  $\lambda_i^*, i = 0, 1, \dots, p$ , are nonnegative and not all zero. This proves the statements (i), (iii) and (v) when combined with (10).

Take any  $i \in \{1, \dots, p\} \setminus I_1$ . The restriction of  $f_i$  on  $\mathbb{R}^J$  is constant. Combining this with (2), we get

$$f_{i,\Delta_i}(x^*) = f_i(x^*) = f_i(x^\ell) \leq 0.$$

Hence, the statement (iv) follows from (8).

We next show that  $d_0 = 0$ . If it is not the case, then the  $(\text{MF})_\infty$  property shows the existence of a vector  $v \in \mathbb{R}^n$  satisfying

$$\langle \nabla f_{i,\Delta_i}(x^*), v \rangle < 0 \quad \text{for all } i \in I_1 \cup \{0\} \quad \text{with } f_{i,\Delta_i}(x^*) = 0.$$

(Note that  $I_1 \cup \{0\}$  is the set of indices  $i \geq 0$  for which the restriction of  $f_i$  on  $\mathbb{R}^J$  is not constant.) These inequalities, together with the proved statements (iii)-(v), give a contradiction:

$$\begin{aligned} 0 &= \lambda_0^* \langle \nabla f_{0,\Delta_0}(x^*), v \rangle + \sum_{i=1}^p \lambda_i^* \langle \nabla f_{i,\Delta_i}(x^*), v \rangle \\ &= \sum_{i \in I_4} \lambda_i^* \langle \nabla f_{i,\Delta_i}(x^*), v \rangle < 0. \end{aligned}$$

Therefore,  $d(q, \Gamma(f_0)) = d_0 = 0$ . Furthermore, combining (6) with Condition (c2), we deduce that  $f^* = f_{0,\Delta_0}(x^*)$ , and the statement (ii) follows.

Finally, let  $v \in \mathbb{R}^n$  be a vector such that

$$\langle \nabla f_{i,\Delta_i}(x^*), v \rangle < 0 \quad \text{for all } i \in I_1 \quad \text{with } f_{i,\Delta_i}(x^*) = 0.$$

Since the multipliers  $\lambda_i^*$  can be normalized by multiplication with a positive scalar, it is sufficient to show that  $\lambda_0^* > 0$ .

To the contrary, assume that  $\lambda_0^* = 0$ , so that

$$\sum_{i=1}^p \lambda_i^* \nabla f_{i,\Delta_i}(x^*) = 0.$$

This leads to an absurd situation

$$0 = \sum_{i=1}^p \lambda_i^* \langle \nabla f_{i,\Delta_i}(x^*), v \rangle = \sum_{i \in I_4, i \geq 1} \lambda_i^* \langle \nabla f_{i,\Delta_i}(x^*), v \rangle < 0.$$

Therefore, we must have  $\lambda_0^* > 0$ , which completes the proof.  $\square$

**Remark 3.1.** (i) We do not know whether  $x^*$  is a minimizer or not of the polynomial optimization problem

$$\inf f_{0,\Delta_0}(x) \quad \text{subject to } f_{i,\Delta_i}(x) \leq 0 \quad \text{for } i = 1, \dots, p.$$

(ii) Since the restriction of  $f_0$  on  $S$  does not attain its infimum  $f^*$ , Condition (c2) shows that the function  $(0, \epsilon) \rightarrow \mathbb{R}, t \mapsto (f_0 \circ \varphi)(t)$ , is strictly decreasing (after perhaps shrinking

$\epsilon$ ). Now, by Lemma 2.3, we may write

$$f_0(\varphi(t)) = f^* + ct^\alpha + \text{higher order terms in } t,$$

where  $c > 0$  and  $\alpha > 0$ .

On the other hand, we deduce from (3) and Condition (c6) that

$$\left| \lambda_0(t) \frac{d}{dt}(f_0 \circ \varphi)(t) \right| = \frac{\left\| \sum_{i \in I_1 \cup \{0\}} \lambda_i(t) \nabla f_i(\varphi(t)) + \sum_{k \notin J} \nu_k(t) \nabla h_k(\varphi(t)) \right\|}{2\|\varphi(t)\|} \left| \frac{d\|\varphi(t)\|^2}{dt} \right|.$$

Note by (c9) that  $|\lambda_0(t)| \leq 1$  for all  $t \in (0, \epsilon)$ . Then a simple calculation shows that

$$\|\varphi(t)\| \left\| \sum_{i \in I_1 \cup \{0\}} \lambda_i(t) \nabla f_i(\varphi(t)) + \sum_{k \notin J} \nu_k(t) \nabla h_k(\varphi(t)) \right\| = c't^\alpha + \text{higher order terms in } t,$$

for some constant  $c' \geq 0$ . Since  $\alpha > 0$ , we obtain

$$\lim_{t \rightarrow 0^+} \|\varphi(t)\| \left\| \sum_{i \in I_1 \cup \{0\}} \lambda_i(t) \nabla f_i(\varphi(t)) + \sum_{k \notin J} \nu_k(t) \nabla h_k(\varphi(t)) \right\| = 0.$$

Since the curve  $\varphi(t)$  lies in the constraint set  $S$ , it follows from (c1), (c2) and the above equation that the restriction of  $f_0$  on  $S$  does not satisfy the so-called (*weak*) *Palais–Smale condition*<sup>3</sup> at the optimal value  $f^*$ . We refer the reader to the survey of Mawhin and Willem [22] for more details about the history and genesis of the Palais–Smale condition.

**Remark 3.2.** It is worth mentioning that, very recently, relying on results from real algebraic geometry, Lasserre [19] derived *global optimality conditions* for polynomial optimization which generalize the local optimality conditions due to Fritz–John and Karush–Kuhn–Tucker for nonlinear optimization.

We now study the existence of optimal solutions to the optimization problem (P). Very recently, it was proved in [10] that (P) has an optimal solution provided the following conditions hold:

- (i) all the polynomial functions  $f_0, f_1, \dots, f_p$  are convenient, and
- (ii) the polynomial map  $(f_0, f_1, \dots, f_p): \mathbb{R}^n \rightarrow \mathbb{R}^{p+1}$  satisfies the so-called condition of non-degeneracy at infinity. (See also Definition 2.1.)

We would like to mention that Theorem 3.1 (and hence Theorem 3.2 below) still holds if the  $(MF)_\infty$  property is replaced by the condition of non-degeneracy at infinity; we leave

---

<sup>3</sup>Given a differentiable function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  and a value  $y \in \mathbb{R}$ , we say that  $f$  satisfies the *weak Palais–Smale condition* at  $y$  if any sequence  $\{x^\ell\} \subset \mathbb{R}^n$  such that  $f(x^\ell) \rightarrow y$  and  $\|x^\ell\| \|\nabla f(x^\ell)\| \rightarrow 0$  as  $\ell \rightarrow \infty$  contains a convergent subsequence (whose limit is then a critical point with critical value  $y$ ).

it the reader to verify these facts. Moreover, as a first application of Theorem 3.1, we obtain the following result which improves Theorem 1.1 in [10].

**Theorem 3.2** (A Frank–Wolfe type theorem). *Let the  $(\text{MF})_\infty$  property hold for the problem  $\inf_{x \in S} f_0(x)$ . In addition, if the polynomial  $f_0$  is convenient, then the problem has at least an optimal solution.*

*Proof.* Suppose, the assertion of the theorem is false. Keeping the notations as in the proof of Theorem 3.1. We have  $\min_{j \in J} q_j < 0$  and  $d_0 := d(q, \Gamma(f_0)) \geq 0$ . Let  $j^* \in J$  be such that  $q_{j^*} = \min_{j \in J} q_j$ . Since  $f_0$  is convenient, there exists some  $\alpha > 0$  such that  $(0, \dots, 0, \hat{\alpha}, 0, \dots, 0) \in \Gamma(f_0)$ . Therefore,

$$0 \leq d(q, \Gamma(f_0)) = \min \left\{ \sum_{j=1}^n q_j \kappa_j : \kappa \in \Gamma(f_0) \right\} \leq q_{j^*} \alpha < 0,$$

which is impossible. The theorem is proved.  $\square$

**Example 3.1.** Let  $f_0(x, y, z) := x^2 + y^2 + z$  and  $S := \{(x, y, z) \in \mathbb{R}^3 : f_1(x, y, z) := -z \leq 0\}$ . It is easy to check that the problem  $\inf_{x \in S} f_0(x)$  is bounded below and has the  $(\text{MF})_\infty$  property. Since  $f_0$  is convenient, it follows from Theorem 3.2 that the polynomial  $f_0$  attains its infimum on  $S$ ; namely, we can see that  $(0, 0, 0) \in S$  and  $f^* = f(0, 0, 0)$ . Notice that [10, Theorem 1.1] cannot be applied for this example because  $f_1$  is not convenient.

#### 4. THE UNCONSTRAINED CASE

In the rest of this paper we assume that  $S = \mathbb{R}^n$ . Then (P) is a unconstrained optimization problem:

$$f^* := \inf_{x \in \mathbb{R}^n} f_0(x).$$

**Theorem 4.1** (Fermat’s theorem). *Assume that the polynomial  $f_0$  is bounded from below and non-degenerate at infinity. If  $f_0$  does not attain its infimum  $f^*$ , then there exist a point  $x^* \in (\mathbb{R} \setminus \{0\})^n$  and a bad face  $\Delta_0$  of  $\Gamma(f_0)$  such that*

$$f^* = f_{0, \Delta_0}(x^*) \quad \text{and} \quad \nabla f_{0, \Delta_0}(x^*) = 0.$$

*Proof.* Since  $f_0$  is non-degenerate at infinity, it follows from Lemma 2.1 that the problem  $\inf_{x \in \mathbb{R}^n} f_0(x)$  has the  $(\text{MF})_\infty$  property. Keeping the notations as in the proof of Theorem 3.1. There exist a nonempty set  $J \subset \{1, \dots, n\}$ , a vector  $q \in \mathbb{R}^n$  with  $\min_{j \in J} q_j < 0$ , and a point  $x^* \in \mathbb{R}^n$  such that the conditions (c1)–(c9) hold.

Observe that, if  $d_0 := d(q, \Gamma(f_0)) > 0$  then  $f^* = 0$  and  $0 \notin \Gamma(f_0)$ ; hence  $f(0) = 0 = f^*$  which contradicts our assumption. By Theorem 3.1, therefore

- (a)  $x_j^* = 0$  if, and only if,  $j \notin J$ ;



- (b)  $\Delta_0 := \Delta(q, \Gamma(f_0)) \subset \mathbb{R}^J$  and  $d_0 = 0$ ;  
(c)  $f^* = f_{0, \Delta_0}(x^*)$  and  $\nabla f_{0, \Delta_0}(x^*) = 0$ .

(The assumption  $S = \mathbb{R}^n$  yields that  $I_1 = \emptyset$  and hence, by (4), that  $\lambda_0^* = 1$ ). Furthermore, Conditions (c6) reads

$$\nabla f_0(\varphi(t)) + \sum_{k \notin J} \nu_k(t) \nabla h_k(\varphi(t)) + \mu(t) \varphi(t) \equiv 0. \quad (11)$$

On the other hand, since  $\Delta_0 \subset \mathbb{R}^J$ , the polynomial  $f_{0, \Delta_0}$  does not depend on  $x_j$  for all  $j \notin J$ . Hence, by re-assigning  $x_j^* := 1$  for all  $j \notin J$ , we obtain the new point  $x^* \in (\mathbb{R} \setminus \{0\})^n$  for which the property (c) still holds.

We next prove that  $\max_{j \in J} q_j > 0$ . If it is not the case, then we have

$$\sum_{j=1}^n q_j \kappa_j = \sum_{j \in J} q_j \kappa_j \leq 0 \quad \text{for all } \kappa := (\kappa_1, \dots, \kappa_n) \in \Gamma(f_0) \cap \mathbb{R}^J.$$

Since  $d_0 = 0$  is the smallest value of the linear function  $\sum_{j=1}^n q_j \kappa_j$  on  $\Gamma(f_0) \cap \mathbb{R}^J$  (see the equation (5)), this follows that

$$\kappa_{j^*} = 0 \quad \text{for all } \kappa \in \Gamma(f_0) \cap \mathbb{R}^J,$$

where  $j^* \in J$  is such that  $q_{j^*} = \min_{j \in J} q_j < 0$ . Consequently, the restriction of  $f_0$  on  $\mathbb{R}^J$  does not depend on the variable  $x_{j^*}$ , and so  $\frac{\partial f_0}{\partial x_{j^*}}(\varphi(t)) \equiv 0$ , in contradiction to (11) because we know that  $\frac{\partial h_k}{\partial x_{j^*}}(\varphi(t)) \equiv 0$  for all  $k \notin J$  and  $\varphi_{j^*}(t) \not\equiv 0$ .

In summary, we have shown that  $d(q, \Gamma(f_0)) = 0$  and  $\min_{j \in J} q_j < 0 < \max_{j \in J} q_j$ . Let

$$H := \{x \in \mathbb{R}^n : \sum_{j=1}^n q_j \kappa_j = 0\}.$$

Then the equality  $H \cap \Gamma(f_0) = \Delta_0$  follows immediately from definitions. On the other hand, if the affine subvariety of dimension  $= \dim(\Delta_0)$  spanned by  $\Delta_0$  does not contain the origin, then  $\Delta_0 \in \Gamma_\infty(f_0)$ , which, together with Condition (c) above, leads to a contradiction with the nondegeneracy condition of  $f_0$ . Hence the conditions (i) and (ii) in the definition of a bad face are fulfilled. The theorem is proved.  $\square$

Let  $K_0(f_0)$  be the set of critical values of  $f_0$ , i.e.,

$$K_0(f_0) := \{f_0(x) : x \in \mathbb{R}^n \text{ and } \nabla f_0(x) = 0\};$$

we also put

$$\begin{aligned} \Sigma_\infty(f_0) &:= \bigcup_{\Delta} \{f_{0, \Delta}(x) : x \in \mathbb{R}^n \text{ and } \nabla f_{0, \Delta}(x) = 0\}, \\ \Sigma'_\infty(f_0) &:= \bigcup_{\Delta} \{f_{0, \Delta}(x) : x \in (\mathbb{R} \setminus \{0\})^n \text{ and } \nabla f_{0, \Delta}(x) = 0\}, \end{aligned}$$

where the unions are taken over all bad faces  $\Delta$  of  $\Gamma(f_0)$ . Clearly, we have  $\Sigma'_\infty(f_0) \subset \Sigma_\infty(f_0)$ . Furthermore, by a semi-algebraic version of Sard's theorem (see, for example, [14]), the sets  $K_0(f_0)$ ,  $\Sigma_\infty(f_0)$  and  $\Sigma'_\infty(f_0)$  are finite.

**Theorem 4.2.** *Assume that the polynomial  $f_0$  is bounded from below on  $\mathbb{R}^n$  and non-degenerate at infinity. We have*

$$\begin{aligned} f^* &= \min\{c : c \in K_0(f_0) \cup \Sigma_\infty(f_0)\} \\ &= \min\{c : c \in K_0(f_0) \cup \Sigma'_\infty(f_0)\}. \end{aligned}$$

*Proof.* We first show that

$$f^* \leq \min\{c : c \in K_0(f_0) \cup \Sigma_\infty(f_0)\}. \quad (12)$$

Indeed, it is clear that

$$f^* \leq \min\{c : c \in K_0(f_0)\}.$$

Let  $\Delta$  be a bad face of  $\Gamma(f_0)$  and let  $x^* \in \mathbb{R}^n$  be a critical point of  $f_{0,\Delta}$  such that

$$f_{0,\Delta}(x^*) = \min\{c : c \in \Sigma_\infty(f_0)\}.$$

By definition, there exist a nonzero vector  $q \in \mathbb{R}^n$  with  $\min_{j=1,\dots,n} q_j < 0$  such that  $\Delta = \Delta(q, \Gamma(f_0))$  and  $d(q, \Gamma(f_0)) = 0$ . Let  $J$  be the smallest subset of  $\{1, \dots, n\}$  such that  $\Delta \subset \mathbb{R}^J$ . Then  $J \neq \emptyset$  because  $\Delta$  is a bad face of  $\Gamma(f_0)$ . Moreover, the restriction of  $f_0$  on  $\mathbb{R}^J$  is not constant and the polynomial  $f_{0,\Delta}$  does not depend on the variables  $x_j$  for  $j \notin J$ .

Let  $J' := \{j \in J : x_j^* \neq 0\}$ . If  $J' = \emptyset$  or  $f_0|_{\mathbb{R}^{J'}}$  is constant, then it follows from definitions that

$$f^* \leq f_0(0) = f_{0,\Delta}(0) = f_{0,\Delta}(x^*)$$

and (12) holds. So assume that  $J' \neq \emptyset$  and  $f_0|_{\mathbb{R}^{J'}}$  is not constant. Let  $\Delta' := \Delta \cap \mathbb{R}^{J'}$ . Then  $\Delta'$  is a closed face of  $\Gamma(f_0|_{\mathbb{R}^{J'}})$ . Furthermore, by definition, we have

$$f_{0,\Delta'}(x^*) = f_{0,\Delta}(x^*).$$

Define the monomial curve  $\varphi: (0, 1) \rightarrow \mathbb{R}^n, t \mapsto (\varphi_1(t), \dots, \varphi_n(t))$ , by

$$\varphi_j(t) := \begin{cases} x_j^* t^{q_j} & \text{if } j \in J', \\ 0 & \text{otherwise.} \end{cases}$$

A simple calculation shows that

$$f_0(\varphi(t)) = f_{0,\Delta'}(x^*) + \text{higher order terms in } t.$$

Consequently, we obtain

$$f^* \leq \lim_{t \rightarrow 0^+} f_0(\varphi(t)) = f_{0,\Delta'}(x^*) = f_{0,\Delta}(x^*)$$

and (12) is proved.

We now assume that  $f_0$  attains its infimum  $f^*$  on  $\mathbb{R}^n$ . Then

$$\begin{aligned} f^* &= \min\{c : c \in K_0(f_0)\} \\ &\geq \min\{c : c \in K_0(f_0) \cup \Sigma'_\infty(f_0)\} \\ &\geq \min\{c : c \in K_0(f_0) \cup \Sigma_\infty(f_0)\}. \end{aligned}$$

These, together with (12), yield the desired relations.

Finally, we assume that  $f_0$  does not attain its infimum  $f^*$  on  $\mathbb{R}^n$ . By Theorem 4.1, we have

$$\begin{aligned} f^* &\geq \min\{c : c \in \Sigma'_\infty(f_0)\} \\ &\geq \min\{c : c \in K_0(f_0) \cup \Sigma'_\infty(f_0)\} \\ &\geq \min\{c : c \in K_0(f_0) \cup \Sigma_\infty(f_0)\}. \end{aligned}$$

Combining this with (12) again we get the desired relations.  $\square$

**Example 4.1.** Let  $f_0: \mathbb{R}^2 \rightarrow \mathbb{R}$  be the polynomial defined by  $f_0(x, y) := (xy - 1)^2 + x^2$ . A simple calculation shows that  $(0, 0)$  is the unique critical point of  $f_0$ , and so

$$K_0(f_0) = \{f_0(0, 0)\} = \{1\}.$$

On the other hand, by definition, the Newton polyhedron  $\Gamma(f_0)$  of  $f_0$  is the triangle with the vertices at  $O(0, 0)$ ,  $A(2, 2)$  and  $B(2, 0)$ . It follows that the edge  $\Delta := OA$  is the unique bad face of  $\Gamma(f_0)$  and that  $f_{0, \Delta} = (xy - 1)^2$ . Hence, by definition again, we obtain

$$\Sigma_\infty(f_0) := \{0, 1\} \quad \text{and} \quad \Sigma'_\infty(f_0) := \{0\}.$$

Note that, the polynomial  $f_0$  is bounded from below and non-degenerate at infinity. Therefore, due to Theorem 4.2,

$$f^* = \min\{c : c \in K_0(f_0) \cup \Sigma_\infty(f_0)\} = 0.$$

**Remark 4.1.** (i) In the paper [25], Nie, Demmel, and Sturmfels established sum of squares representations of positive polynomials modulo gradient ideals, i.e., the ideals generated by all the partial derivatives. Based on these representations, the authors constructed a sequence of SDP relaxations whose optimal values converge monotonically, increasing to the smallest critical value of a polynomial. Combining this fact with Theorem 4.2, we can find an appropriate sequence of computationally feasible SDP relaxations, whose optimal values converge to the infimum value  $f^* := \inf_{x \in \mathbb{R}^n} f_0(x)$ . These facts open up the possibility of solving previously intractable polynomial optimization problems.

(ii) We do not know whether Theorem 4.2 can be extended to the case of optimization problems with constraints. This question will be studied in the future work.

**Acknowledgments.** The author thanks to Jean Bernard Lasserre for useful discussions. The final version of this paper was completed while the author was visiting LAAS–CNRS in April 2017. He wishes to thank the institute and Jean Bernard Lasserre in particular for the hospitality and financial support from the European Research Council Advanced Grant for the TAMING project, No. 666981.

## REFERENCES

- [1] V. G. ANDRONOV, E. G. BELOUSOV, AND V. M. SHIRONIN, *On solvability of the problem of polynomial programming*, Izvestija Akadem. Nauk SSSR, Tekhnicheskaja Kibernetika, no. 4, 194–197 (1982) (in Russian). Translation Appeared in News of the Academy of Science of USSR, Dept. of Technical Sciences, Technical Cybernetics, no. 4, 194–197 (1982).
- [2] B. BANK AND R. MANDEL, *Parametric integer optimization*, Mathematical Research, Akademie-Verlag: Berlin, vol. **39**, (1988).
- [3] E. G. BELOUSOV, *Introduction to convex analysis and integer programming*, Moscow University Publ.: Moscow, (1977) (in Russian).
- [4] E. G. BELOUSOV AND D. KLATTE, *A Frank-Wolfe type theorem for convex polynomial programs*, Comput. Optim. Appl., **22** (1) (2002), 37–48.
- [5] J. BOCHNAK, M. COSTE, AND M.-F. ROY, *Real algebraic geometry*, vol. **36**, Springer, 1998.
- [6] D.P. BERTSEKAS, *Nonlinear programming*, Athena Scientific Publishers, Belmont, MA, 1999.
- [7] D. P. BERTSEKAS AND P. TSENG, *Set intersection theorems and existence of optimal solutions*, Math. Program. Series A, **110** (2) (2007) 287–314.
- [8] J. BOLTE, A. HOCHART, AND E. PAUWELS, *Qualification conditions in semi-algebraic programming*, Available at <https://arxiv.org/pdf/1705.08219.pdf>.
- [9] F. H. CLARKE, *Optimization and nonsmooth analysis*, Classics in Applied Mathematics, SIAM, Philadelphia, PA, 1990.
- [10] S. T. ĐINH, H. V. HÀ, AND T. S. PHẠM, *A Frank–Wolfe type theorem for nondegenerate polynomial programs*, Math. Program. Ser. A., **147** (1–2) (2014), 519–538.
- [11] D. V. ĐOẠT, H. V. HÀ, AND T. S. PHẠM, *Well-posedness in unconstrained polynomial optimization problems*, SIAM J. Optim., **26** 3 (2016), 1411–1428.
- [12] M. FRANK, AND P. WOLFE, *An algorithm for quadratic programming*, Naval Res. Logist. Quart., **3** (1956), 95–110.
- [13] H. V. HÀ, *Global Hölderian error bound for non-degenerate polynomials*, SIAM J. Optim., **23** (2) (2013) 917–933.
- [14] H. V. HÀ AND T. S. PHẠM, *Genericity in polynomial optimization*, vol. 3 of Series on Optimization and Its Applications, World Scientific, 2017.
- [15] F. JOHN, *Extremum problems with inequalities as side constraints*, In “Studies and Essays, Courant Anniversary Volume” (K.O. Friedrichs, O.E. Neugebauer and J. J. Stoker, eds), 187–204. Wiley (interscience), New York, 1948.
- [16] W. KARUSH *Minima of functions of several variables with inequalities as side constraints*, M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois, 1939.
- [17] H. W. KUHN AND A. W. TUCKER, *Nonlinear programming*, Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press. (1951), 481–492.

- [18] A. G. KOUCHNIRENKO, *Polyhedres de Newton et nombre de Milnor*, Invent. Math., **32** (1976), 1–31.
- [19] J.B. LASSERRE, *An introduction to polynomial and semi-algebraic optimization*, Cambridge University Press, Cambridge, 2015.
- [20] Z.-Q. LUO AND S. ZHANG, *On extensions of the Frank-Wolfe theorems*, Comput. Optim. Appl., **13** (1-3) (1999), 87–110.
- [21] O. L. MANGASARIAN AND S. FROMOVITZ, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl., **17** (1967), 37–47.
- [22] J. MAWHIN AND M. WILLEM, *Origin and evolution of the Palais–Smale condition in critical point theory*, J. Fixed Point Theory Appl., **7** (2010), 265–290.
- [23] B. S. MORDUKHOVICH, *Variational analysis and generalized differentiation, I: Basic Theory; II: Applications*, Springer, Berlin, 2006.
- [24] A. NÉMETHI AND A. ZAHARIA, *On the bifurcation set of a polynomial. function and Newton boundary*, Publ. RIMS. Kyoto Univ., **26** (1990), 681–689.
- [25] J. NIE, J. DEMMEL, AND B. STURMFELS, *Minimizing polynomials via sum of squares over the gradient ideal*, Math. Program. Ser. A, **106** (3) (2006), 587–606.
- [26] W. T. OBUCHOWSKA, *On generalizations of the Frank-Wolfe theorem to convex and quasi-convex programmes*, Comput. Optim. Appl., **33** (2-3) (2006), 349–364.
- [27] A. F. PEROLD, *Generalization of the Frank-Wolfe Theorem*, Math. Program., **18** (1980), 215–227.
- [28] B. TOMÁŠ AND O. STEIN, *Coercive polynomials and their Newton polytopes*, SIAM J. Optim., **25** (3) (2015), 1542–1570.
- [29] T. TERLAKY, *On  $l_p$  programming*, Euro. J. Oper. Res., **22** (1985), 70–100.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF DALAT, 1 PHU DONG THIEN VUONG, DALAT, VIETNAM

*E-mail address:* sonpt@dlu.edu.vn