CONSTRAINT QUALIFICATIONS FOR NONLINEAR PROGRAMMING

RODRIGO G. EUSTÁQUIO[§], ELIZABETH W. KARAS[¶], AND ADEMIR A. RIBEIRO[¶]

Abstract. This paper deals with optimality conditions to solve nonlinear programming problems. The classical Karush-Kuhn-Tucker (KKT) optimality conditions are demonstrated through a cone approach, using the well known Farkas' Lemma. These conditions are valid at a minimizer of a nonlinear programming problem if a constraint qualification is satisfied. First we prove the KKT theorem supposing the equality between the polar of the tangent cone and the polar of the first order feasible variations cone. Although this condition is the weakest assumption, it is extremely difficult to be verified. Therefore, other constraints qualifications, which are easier to be verified, are studied, as: Slater's, linear independence of gradients, Mangasarian-Fromovitz's and quasiregularity. The relations among them are discussed.

Key words. Optimality conditions, Karush-Kuhn-Tucker, constraint qualifications.

1. Introduction. We shall study the nonlinear programming problem

$$(P) \qquad \begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \\ & g(x) \leq 0, \end{array}$$

where the functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. The feasible set is $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0, \quad g(x) \leq 0\}$.

Given $x^* \in \Omega$, the classical *Karush-Kuhn-Tucker* (KKT) conditions say that there exist *Lagrangian* multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that:

$$\begin{gathered} -\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*), \\ \mu_j^* \geq 0, \quad j = 1, \ldots, p, \\ \mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, p. \end{gathered}$$

In nonlinear programming we would like that KKT are necessary conditions for a given point to be a solution to the problem. When the problem is unconstrained ($\Omega = \mathbb{R}^n$), the KKT conditions reduce to $\nabla f(x^*) = 0$ which is a necessary optimality condition. However, this not always happens, as shown in the following example.

EXAMPLE 1. *Consider the problem* (P) *with* $f : \mathbb{R}^2 \to \mathbb{R}$ *and* $g : \mathbb{R}^2 \to \mathbb{R}^2$ *defined by* $f(x) = x_1$ *and* $g(x) = (x_2 - (1 - x_1)^3, -x_2)^T$*. Note that* $x^* = (1, 0)^T$ *is a minimizer of the problem but the KKT conditions do not hold.*

In this paper we shall discuss assumptions on the constraints in order to ensure that the KKT conditions hold at a minimizer. Such an assumption is called *constraint qualification (CQ)*. Formally, we say that the constraints $h(x) = 0$ and $g(x) \leq 0$ satisfy a constraint qualification at $x^* \in \Omega$ when, given any differentiable function f minimized at x^* with respect to Ω, the KKT conditions are valid.

Several authors have obtained different constraint qualifications. In this work, we will discuss many of them as well as some relations between them. A special interest is devoted to show the weakest such qualification. In this context, the concept of cones and their polars will be useful.

[§]Master Program in Numerical Methods in Engineering, Federal University of Paraná, Cx. Postal 19081, 81531-980, Curitiba, PR, Brazil; e-mail:`rodrigogarcia1@bol.com.br`.

[¶]Department of Mathematics, Federal University of Paraná, Cx. Postal 19081, 81531-980, Curitiba, PR, Brazil; e-mail:`karas@mat.ufpr.br`, `ademir@mat.ufpr.br`. Supported by PRONEX - Optimization.

Structure of the paper. Section 2 will be dedicated to define some important cones and to discuss their properties. In Section 3 we prove KKT theorem under the weakest constraint qualification and we discuss other ones easier to be verified as: Slater's, linear independence of gradients, Mangasarian-Fromovitz's and quasiregularity.

Notation. Given $\bar{x} \in \Omega$, consider the set $A(\bar{x})$ of the inequality active constraint indices, that is,

$$A(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}. \tag{1.1}$$

Given $\phi : \mathbb{R}^n \to \mathbb{R}^\ell$, we denote $\nabla\phi = (\nabla\phi_1 \dots \nabla\phi_\ell)$. So, the Jacobian matrix of ϕ is $\nabla\phi^T$.

2. Some important cones. In this section we present some useful cones based on the structure of the feasible set which play an important rule on the proof of the KKT theorem.

A subset $C \subset \mathbb{R}^n$ is a cone when $td \in C$, for all $t \geq 0$ and $d \in C$. Given a set $S \in \mathbb{R}^n$, the polar of S, is given by

$$P(S) = \{p \in \mathbb{R}^n \mid p^T x \leq 0 \quad \text{for all } x \in S\}.$$

Note that for any $S \subset \mathbb{R}^n$, $P(S)$ is a cone and $S \subset P(P(S))$. This holds with equality if S is a closed convex cone, as established by Farkas' Lemma below.

LEMMA 2.1. *Let $C \subset \mathbb{R}^n$ be a closed convex cone. Then $P(P(C)) = C$.*

Proof. Consider $z \in P(P(C))$ and $\hat{z} = proj_C(z) \in C$. We will show that $z = \hat{z}$. By definition of $\hat{z}$, for all $x \in C$, we have

$$(z - \hat{z})^T(x - \hat{z}) \leq 0. \tag{2.1}$$

Since C is a cone, $x = 0$ and $x = 2\hat{z}$ belongs to C. So,

$$-\hat{z}^T(z - \hat{z}) \leq 0 \quad \text{and} \quad \hat{z}^T(z - \hat{z}) \leq 0,$$

which imply

$$\hat{z}^T(z - \hat{z}) = 0. \tag{2.2}$$

Substituting this on (2.1), we conclude that $(z - \hat{z})^T x \leq 0$ for all $x \in C$, which by definition of polar cone, yields $(z - \hat{z}) \in P(C)$. Since $z \in P(P(C))$, we have that

$$(z - \hat{z})^T z \leq 0. \tag{2.3}$$

But

$$\|z - \hat{z}\|^2 = (z - \hat{z})^T z - (z - \hat{z})^T \hat{z}.$$

Using (2.2) and (2.3), we conclude that $\|z - \hat{z}\| \leq 0$. So, $z = \hat{z}$. ☐

Given $\bar{x} \in \Omega$, we say that $d \in \mathbb{R}^n$ is a *feasible direction* at $\bar{x}$, with respect to set Ω, when there exists $\delta > 0$ such that,

$$\bar{x} + td \in \Omega, \quad \forall t \in [0, \delta].$$

We denote by $V(\bar{x})$ the cone of feasible directions at $\bar{x}$.

A direction $d \in \mathbb{R}^n$ is a *descent direction* of the function f at $\bar{x}$, if there exists $\delta > 0$ such that

$$f(\bar{x} + td) < f(\bar{x}), \quad \forall t \in (0, \delta].$$

The set of descent directions of f at $\bar{x}$ is denoted by $F(\bar{x})$.

The next result characterizes the descent directions and its proof follows from the derivative definition.

LEMMA 2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function at a point $\bar{x} \in \mathbb{R}^n$. Then*
(i) $\nabla f(\bar{x})^T d \leq 0$, for all $d \in F(\bar{x})$.
(ii) If $d \in \mathbb{R}^n$ satisfies $\nabla f(\bar{x})^T d < 0$, then $d \in F(\bar{x})$.

We denote

$$F_0(\bar{x}) = \{d \in \mathbb{R}^n \mid \nabla f(\bar{x})^T d < 0\}. \tag{2.4}$$

We define the set of *first order feasible variations* at a point $\bar{x} \in \Omega$ by

$$D(\bar{x}) = \left\{d \in \mathbb{R}^n \mid \nabla h_i(\bar{x})^T d = 0, \forall i = 1, \ldots, m, \quad \text{and} \quad \nabla g_j(\bar{x})^T d \leq 0, \forall j \in A(\bar{x})\right\}, \tag{2.5}$$

where the active set $A(\bar{x})$ is defined by (1.1).

It is easy to see that $D(\bar{x})$ is a closed convex nonempty cone. We can say that this cone is a linear approximation of the feasible set.

Given $\bar{x} \in \Omega$, we define the cone $G(\bar{x})$, that will be useful ahead, by

$$G(\bar{x}) = \left\{\sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{j \in A(\bar{x})} \mu_j \nabla g_j(\bar{x}) \mid \mu_j \geq 0, \forall j \in A(\bar{x})\right\}. \tag{2.6}$$

Let us see some properties of this cone. For this, we shall need the classical result, named Carathéodory's Lemma.

LEMMA 2.3. *Let $u_1, \ldots, u_r$ be nonzero vectors in $\mathbb{R}^n$, $m < r$ and $x \in \mathbb{R}^n$ such that*

$$x = \sum_{i=1}^{r} \gamma_i u_i, \tag{2.7}$$

with $\gamma_i \geq 0$ for all $i > m$. Then, there exist indices subsets $I \subset \{1, \ldots, m\}$, $J \subset \{m+1, \ldots, r\}$ and scalars γ_i', $i \in I \cup J$, with $\gamma_i' \geq 0$, for $i \in J$, such that

$$x = \sum_{i \in I \cup J} \gamma_i' u_i$$

and the vectors u_i, $i \in I \cup J$, are linearly independent.

Proof. If the vectors $u_1, \ldots, u_r$ are linearly independent, there is nothing to prove. Suppose then that they are linearly dependent. So, there exist scalars α_i with $i = 1, \ldots, r$, not all $\alpha_i = 0$, such that

$$\sum_{i=1}^{r} \alpha_i u_i = 0.$$

Therefore, for all $t \in \mathbb{R}$,

$$x = \sum_{i=1}^{r} (\gamma_i - t\alpha_i) u_i.$$

Define $\bar{t}$ as t of minimum absolute value that vanish one of the coefficients $\gamma_i - t\alpha_i$. Then

$$x = \sum_{i=1}^{r} (\gamma_i - \bar{t}\alpha_i) u_i$$

with $\gamma_i - \bar{t}\alpha_i \geq 0$, for all $i > m$. Therefore x is written as a linear combination using no more than $r-1$ vectors. We can repeat this process until that all vectors of the linear combination are linearly independent. $\square$

LEMMA 2.4. *For any $\bar{x} \in \Omega$, $G(\bar{x})$ is a closed convex cone.*

Proof. Consider without loss of generality that $A(\bar{x}) = \{1, \dots, q\}$. First we prove that $G(\bar{x})$ is convex. Consider s_1, $s_2 \in G(\bar{x})$ and $t \in [0,1]$. Then there exist $\lambda, \alpha \in \mathbb{R}^m$, $\mu, \beta \in \mathbb{R}^q_+$ such that

$$s_1 = \sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{j=1}^{q} \mu_j \nabla g_j(\bar{x}) \qquad \text{and} \qquad s_2 = \sum_{i=1}^{m} \alpha_i \nabla h_i(\bar{x}) + \sum_{j=1}^{q} \beta_j \nabla g_j(\bar{x}).$$

Therefore,

$$ts_1 + (1-t)s_2 = \sum_{i=1}^{m} (t\lambda_i + (1-t)\alpha_i) \nabla h_i(\bar{x}) + \sum_{j=1}^{q} (t\mu_j + (1-t)\beta_j) \nabla g_j(\bar{x}).$$

Since $t\mu_j + (1-t)\beta_j \geq 0$, we conclude that $ts_1 + (1-t)s_2 \in G(\bar{x})$.

Now we will show that $G(\bar{x})$ is closed. For that, consider a sequence $(s^k) \subset G(\bar{x})$ satisfying $s^k \to s^* \in \mathbb{R}^n$. We need to prove that $s^* \in G(\bar{x})$. For suitable matrices B and C, we have

$$G(\bar{x}) = \{B\lambda + C\zeta \mid \zeta \geq 0\}.$$

By the Caratheodory's Lemma 2.3, we can assume that $D = (B\ C)$ have linearly independent columns, so that $D^T D$ is nonsingular. Since $(s^k) \subset G(\bar{x})$, there exists $\gamma^k = \begin{pmatrix} \lambda^k \\ \zeta^k \end{pmatrix}$ with $\zeta^k \geq 0$ such that

$$s^k = D\gamma^k. \tag{2.8}$$

Since $D^T D$ is nonsingular, $\gamma^k = (D^T D)^{-1} D^T s^k$. Taking limit, we obtain

$$\begin{pmatrix} \lambda^* \\ \zeta^* \end{pmatrix} = \gamma^* = \lim_{k \to \infty} \gamma^k = (D^T D)^{-1} D^T s^*$$

with $\zeta^* \geq 0$. Taking limit in (2.8), we have

$$s^* = D\gamma^* \in G(\bar{x}),$$

completing the proof. $\square$

LEMMA 2.5. *For any $\bar{x} \in \Omega$, $G(\bar{x}) = P(D(\bar{x}))$.*

Proof. From Lemmas 2.1 and 2.4 it is enough to prove that $D(\bar{x}) = P(G(\bar{x}))$. Consider $d \in D(\bar{x})$. Given $s \in G(\bar{x})$, we have

$$d^T s = \sum_{i=1}^{m} \lambda_i d^T \nabla h_i(\bar{x}) + \sum_{j \in A(\bar{x})} \mu_j d^T \nabla g_j(\bar{x}). \tag{2.9}$$

By definition of $D(\bar{x})$ and since $\mu_j \geq 0$, it follows that $d^T s \leq 0$. So, $d \in P(G(\bar{x}))$.

Conversely, consider $d \in P(G(\bar{x}))$, that is,

$$d^T s \leq 0, \quad \forall s \in G(\bar{x}).$$

In particular, since $\nabla h_i(\bar{x})$ and $-\nabla h_i(\bar{x})$ belongs to $G(\bar{x})$, for all $i = 1, \ldots, m$, we have $d^T \nabla h_i(\bar{x}) = 0$. Furthermore, since $\nabla g_j(\bar{x}) \in G(\bar{x})$, for all $j \in A(\bar{x})$, we have $d^T \nabla g_j(\bar{x}) \leq 0$, completing the proof. □

The tangent cone. Let us discuss now another linear approximation of the feasible set, defined by means of tangent directions.

A vector $d \in \mathbb{R}^n$ is called *tangent direction* to $\Omega \subset \mathbb{R}^n$ from $\bar{x} \in \Omega$ when either $d = 0$ or there exists a sequence of feasible points $(x^k) \subset \Omega$ such that $x^k \to \bar{x}$ and

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \frac{d}{\|d\|}. \tag{2.10}$$

Clearly, the set $T(\bar{x})$ of the tangent directions to Ω from $\bar{x}$ is a cone. This set will be called *tangent cone.*

The next lemma states that this cone is closed. However, $T(\bar{x})$ is not necessarily convex.

LEMMA 2.6. *For any $\bar{x} \in \Omega$, $T(\bar{x})$ is closed.*

Proof. Consider $(d^k) \subset T(\bar{x})$, with $d^k \to d$. We need to show that $d \in T(\bar{x})$. This is immediate if $d = 0$. Consider then $d \neq 0$ and suppose without loss of generality that $d^k \neq 0$ for all $k \in \mathbb{N}$. Fixed $k \in \mathbb{N}$, since $d^k \in T(\bar{x})$, there exists $(x^{k,j})_{j \in \mathbb{N}} \subset \Omega$ such that

$$x^{k,j} \xrightarrow{j} \bar{x} \quad \text{and} \quad q^{k,j} = \frac{x^{k,j} - \bar{x}}{\|x^{k,j} - \bar{x}\|} \xrightarrow{j} \frac{d^k}{\|d^k\|}.$$

So, there exists $j_k \in \mathbb{N}$ such that

$$\|x^k - \bar{x}\| < \frac{1}{k} \quad \text{and} \quad \left| q^k - \frac{d^k}{\|d^k\|} \right| < \frac{1}{k},$$

where $x^k = x^{k,j_k}$ and $q^k = q^{k,j_k}$. Taking limits, we obtain $x^k \to \bar{x}$ and

$$\left| q^k - \frac{d}{\|d\|} \right| \leq \left| q^k - \frac{d^k}{\|d^k\|} \right| + \left| \frac{d^k}{\|d^k\|} - \frac{d}{\|d\|} \right| \to 0.$$

Thus, $\dfrac{x^k - \bar{x}}{\|x^k - \bar{x}\|} = q^k \to \dfrac{d}{\|d\|}$, which implies $d \in T(\bar{x})$, completing the proof. □

We have presented two different linear approximations of the feasible set at a point $\bar{x}$: the first order feasible variations cone $D(\bar{x})$ and the tangent cone $T(\bar{x})$. The next result shows that $T(\bar{x})$ is a subset of $D(\bar{x})$.

LEMMA 2.7. *For any $\bar{x} \in \Omega$, $T(\bar{x}) \subset D(\bar{x})$.*

Proof. Consider $d \in T(\bar{x})$, $d \neq 0$. Then there exists a sequence $(x^k) \subset \Omega$ with , such that $x^k \to \bar{x}$ and $\dfrac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \dfrac{d}{\|d\|}$. From the smoothness of g and h it follows that

$$h(x^k) = h(\bar{x}) + \nabla h(\bar{x})^T (x^k - \bar{x}) + o(\|x^k - \bar{x}\|)$$

and

$$g(x^k) = g(\bar{x}) + \nabla g(\bar{x})^T (x^k - \bar{x}) + o(\|x^k - \bar{x}\|).$$

Since $x^k, \bar{x} \in \Omega$, we have for $i \in A(\bar{x})$,

$$\nabla h(\bar{x})^T \frac{(x^k - \bar{x})}{\|x^k - \bar{x}\|} + \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} = 0 \qquad \text{and} \qquad \nabla g_i(\bar{x})^T \frac{(x^k - \bar{x})}{\|x^k - \bar{x}\|} + \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} \leq 0.$$

Taking limits, we obtain

$$\nabla h(\bar{x})^T \frac{d}{\|d\|} = 0 \qquad \text{and} \qquad \nabla g_i(\bar{x})^T \frac{d}{\|d\|} \leq 0,$$

for all $i \in A(\bar{x})$. Thus, $d \in D(\bar{x})$, completing the proof. $\square$

The converse of the above lemma is not true, as we can see in the following example.

EXAMPLE 2. *Consider the functions $h : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ defined by*

$$h(x) = x_1 x_2 \qquad \textit{and} \qquad g(x) = -x_1 - x_2$$

and the point $\bar{x} = (0\ 0)^T$. Thus,

$$T(\bar{x}) = \{d \in \mathbb{R}^2 \mid d_1 \geq 0, d_2 \geq 0, d_1 d_2 = 0\},$$

$$D(\bar{x}) = \{d \in \mathbb{R}^2 \mid -d_1 - d_2 \leq 0\}$$

and $T(\bar{x}) \neq D(\bar{x})$.

In the next section we shall prove that $T(\bar{x}) = D(\bar{x})$ is a constraint qualification known as "quasiregularity".

3. Optimality Conditions and Constraint Qualifications. In this section we prove KKT theorem assuming the weakest qualification condition and discuss other ones easier to be verified.

Next lemma roughly says that at a minimizer, the objective function increases along tangent directions.

LEMMA 3.1. *If $x^* \in \Omega$ is a local minimizer of the problem (P), then $\nabla f(x^*)^T d \geq 0$ for all $d \in T(x^*)$.*

Proof. This follows directly from the relation

$$0 \leq f(x^k) - f(x^*) = \nabla f(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|),$$

which is valid for $(x^k) \subset \Omega$. □

Now we state the classical Karush-Kuhn-Tucker theorem.

THEOREM 3.2. *Let $x^* \in \Omega$ be a local minimizer of the problem (P). If $P(T(x^*)) = P(D(x^*))$, then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that*

$$-\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_i(x^*),$$

$$\mu_j^* \geq 0, \quad j = 1, \dots, p,$$

$$\mu_j^* g_j(x^*) = 0, \quad j = 1, \dots, p.$$

Proof. Consider x^* a local minimizer of (P). By Lemma 3.1, we have $-\nabla f(x^*)^T d \leq 0$ for all $d \in T(x^*)$. Thus, using the hypothesis and Lemma 2.5, we obtain

$$-\nabla f(x^*) \in P(T(x^*)) = P(D(x^*)) = G(x^*).$$

This means that there exist $\lambda \in \mathbb{R}^m$ and $\mu_j \geq 0$, $j \in A(x^*)$, such that

$$-\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*).$$

Defining $\lambda^* = \lambda$ and $\mu_j^* = \begin{cases} \mu_j, & \text{for } j \in A(x^*) \\ 0, & \text{otherwise} \end{cases}$, we complete the proof. □

The hypothesis $P(T(x^*)) = P(D(x^*))$ used to prove last theorem was introduced by Monique Guignard [8] for infinite dimensional space and reformulated for finite case by Gould and Tolle [7]. We next discuss other conditions which imply $P(T(x^*)) = P(D(x^*))$.

Constraint Qualifications. Now we present some other constraint qualification conditions found in the literature (see [1, 3, 4, 5, 6, 10, 11]) and discuss the relationship between them.

We start by a very weak condition.

Quasiregularity constraint qualification. We say that the quasiregularity constraint qualification is satisfied at $\bar{x}$ when $T(\bar{x}) = D(\bar{x})$. Note that this condition implies trivially $P(T(x^*)) = P(D(x^*))$.

Next example shows that these conditions are not equivalent.

EXAMPLE 3. *Consider the functions $h : \mathbb{R}^2 \to \mathbb{R}$, $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $h(x) = x_1 x_2$, $g(x) = (-x_1, -x_2)^T$ and the feasible point $\bar{x} = (0,0)^T$.* It is easy to see that

$$D(\bar{x}) = \{(d_1, d_2) \mid d_1 \geq 0, d_2 \geq 0\},$$

$$T(\bar{x}) = \{(d_1, d_2) \mid d_1 \geq 0, d_2 \geq 0, d_1 d_2 = 0\}$$

and

$$P(T(\bar{x})) = P(D(\bar{x})) = \{(d_1, d_2) \mid d_1 \leq 0, d_2 \leq 0\}.$$

See Fig. 3.1.

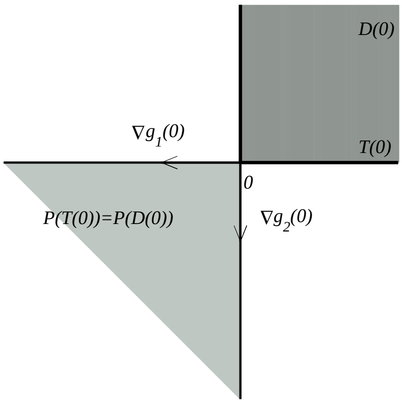

FIG. 3.1. *$P(T(\bar{x})) = P(D(\bar{x}))$ does not imply $T(\bar{x}) = D(\bar{x})$.*

Problems with Linear Constraints. Consider the problem (P) with linear constraints,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Mx = c \\ & Ax \le b \end{array} \tag{3.1}$$

where $A \in \mathbb{R}^{p\times n}$, $M \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^p$ and $c \in \mathbb{R}^m$. In this case it is easy to prove that the constraints are qualified, as shown in the next theorem.

THEOREM 3.3. *Any local minimizer $x^* \in \Omega$ of the problem* (3.1) *satisfies the KKT conditions.*

Proof. Using Lemma 2.7 and Theorem 3.2, it is enough to prove that $D(x^*) \subset T(x^*)$. Given $d \in D(x^*)$, we have $Md = 0$ and $Ad \le 0$. If $d = 0$, there is nothing to prove. So, suppose $d \neq 0$ and define $x^k = x^* + \dfrac{1}{k}d$. Thus, $Mx^k = c$, $Ax^k \le b$, $x^k \to x^*$ and $\dfrac{x^k - x^*}{\|x^k - x^*\|} = \dfrac{d}{\|d\|}$. Therefore, $d \in T(x^*)$, completing the proof. □

Slater constraint qualification. Regarding the problem (P), we say that the Slater constraint qualification holds if h is linear, g is convex and there exists $\widetilde{x} \in \Omega$ such that

$$h(\widetilde{x}) = 0 \quad \text{and} \quad g(\widetilde{x}) < 0.$$

The next theorem shows that Slater condition is, in fact, a constraint qualification.

THEOREM 3.4. *If Slater condition holds, then $T(\bar{x}) = D(\bar{x})$ for all $\bar{x} \in \Omega$.*

Proof. Using Lemma 2.7, it is enough to prove that $D(\bar{x}) \subset T(\bar{x})$. Consider an arbitrary direction $d \in D(\bar{x})$ and $\widetilde{x} \in \Omega$ given by Slater condition. Define $\overline{d} = \widetilde{x} - \bar{x}$. By the convexity of g_i, we have

$$0 > g_i(\widetilde{x}) \ge g_i(\bar{x}) + \nabla g_i(\bar{x})^T \overline{d}.$$

Thus, for $i \in A(\bar{x})$, $\nabla g_i(\bar{x})^T \overline{d} < 0$. Given $\lambda \in (0, 1]$, define

$$\hat{d} = (1 - \lambda)d + \lambda \overline{d}.$$

We will prove that $\hat{d} \in T(\bar{x})$ for all $\lambda \in (0,1)$.

For $i \in A(\bar{x})$, we have $\nabla g_i(\bar{x})^T d \leq 0$ and $\nabla g_i(\bar{x})^T \overline{d} < 0$. Consequently $\nabla g_i(\bar{x})^T \hat{d} < 0$. Therefore, there exists $\hat{x} = \bar{x} + t\hat{d}$, with $t > 0$ such that $g_i(\hat{x}) < g_i(\bar{x}) = 0$ Taking a sequence (t_k), with $t_k > 0$ and $t_k \to 0$, define $x^k = (1-t_k)\bar{x} + t_k\hat{x} = \bar{x} + t_k t\hat{d}$. Thus,

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} = \frac{t_k t\hat{d}}{\|t_k t\hat{d}\|} = \frac{\hat{d}}{\|\hat{d}\|},$$

For $i \notin A(\bar{x})$, $g(\bar{x}) < 0$. By continuity of g, $g(x^k) \leq 0$ for all k sufficiently large. To conclude that $\hat{d} \in T(\bar{x})$ it is enough to show that $h(x^k) = 0$ for all k sufficiently large.

Since $d \in D(\bar{x})$, $Md = \nabla h(x)^T d = 0$. Furthermore, $M\overline{d} = M(\widetilde{x} - \bar{x}) = 0$. Consequently $M\hat{d} = 0$. Thus, $h(x^k) = Mx^k - c = M\bar{x} - c + t_k t M\hat{d} = 0$, since $\bar{x} \in \Omega$. So, $\hat{d} \in T(\bar{x})$, which implies $d \in T(\bar{x})$, since $T(\bar{x})$ is a closed set. □

Linear Independence constraint qualification - LICQ. This is the most known constraint qualification and states that the equality constraint gradients $\nabla h_i(\bar{x})$, $i = 1, \ldots, m$ and the active inequality constraint gradients $\nabla g_i(\bar{x})$, $i \in A(\bar{x})$ are linearly independent. Although easy to check, this condition is a very strong assumption. Many problems satisfy KKT without LICQ, as we can see in the following example with $x^* = 0$.

$$\begin{array}{ll} \text{minimize} & f(x) = x_2 \\ \text{subject to} & g_1(x) = -x_1^2 + x_2 \leq 0 \\ & g_2(x) = -x_2 \leq 0 \qquad . \end{array}$$

Mangasarian-Fromovitz constraint qualification - MFCQ. Another well known condition which ensures KKT is due to Mangasarian and Fromovitz [13]. We say that MFCQ holds at $\bar{x}$ when the equality constraint gradients are linearly independent and there exists a vector $d \in \mathbb{R}^n$ such that

$$\nabla h(\bar{x})^T d = 0 \quad \text{and} \quad \nabla g_j(\bar{x})^T d < 0, \text{ for all } j \in A(\bar{x}).$$

Relation between LICQ and MFCQ.

THEOREM 3.5. *If $\bar{x} \in \Omega$ satisfies LICQ, then $\bar{x}$ satisfies MFCQ.*

Proof. Suppose without loss of generality that $A(\bar{x}) = \{1, \ldots, q\}$. Consider the matrix

$$M = (\nabla h_1(\bar{x}) \ \cdots \nabla h_m(\bar{x}) \ \nabla g_1(\bar{x}) \cdots \nabla g_q(\bar{x}))^T$$

and $b \in \mathbb{R}^{m+q}$ given by $b_i = 0$, for all $i = 1, \ldots, m$ and $b_j = -1$, for all $j \in \{m+1, \ldots, m+q\}$. Since the rows of M are linearly independent, the system $Md = b$ has a solution. Let $\overline{d}$ be a solution. Then

$$\nabla h(\bar{x})^T \overline{d} = 0, \qquad \text{and} \qquad \nabla g_i(\bar{x})^T \overline{d} = -1 < 0, \quad \text{for all } i \in A(\bar{x}),$$

completing the proof. □

The next example shows that MFCQ does not imply LICQ. Consider the functions $g_j : \mathbb{R}^2 \to \mathbb{R}$ $(j = 1, 2, 3)$ defined by

$$\begin{aligned} g_1(x) &= (x_1 - 1)^2 + (x_2 - 1)^2 - 2 \\ g_2(x) &= (x_1 - 1)^2 + (x_2 + 1)^2 - 2 \\ g_3(x) &= -x_1, \end{aligned}$$

and the feasible point $\bar{x} = (0,0)^T$. Note that $\{\nabla g_j(\bar{x}), j = 1,2,3\}$ is linearly dependent. On the other hand, taking $d = (1,0)^T$, we have $\nabla g_j(\bar{x})^T d < 0$ for $j = 1,2,3$, which means that MFCQ holds.

Relation between MFCQ and quasiregularity.

LEMMA 3.6. *Let $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ be a differentiable curve such that $h(\gamma(t)) = 0$, for all $t \in (-\varepsilon, \varepsilon)$. If $\gamma(0) = \bar{x}$ and $\gamma'(0) = d \neq 0$, then there exists a sequence (x^k) with $h(x^k) = 0$, $x^k \to \bar{x}$ and*

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \frac{d}{\|d\|}.$$

Proof. We have

$$\lim_{t\to 0} \frac{\gamma(t) - \bar{x}}{t} = \lim_{t\to 0} \frac{\gamma(t) - \gamma(0)}{t} = \gamma'(0) = d \neq 0,$$

which means that $\gamma(t) \neq \bar{x}$, for all $t \neq 0$ sufficiently small. Taking a sequence (t_k), with $t_k > 0$ and $t_k \to 0$, define $x^k = \gamma(t_k)$. Thus,

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} = \frac{x^k - \bar{x}}{t_k} \frac{t_k}{\|x^k - \bar{x}\|} \to \frac{d}{\|d\|},$$

completing the proof. □

THEOREM 3.7. *If $\bar{x} \in \Omega$ satisfies MFCQ, then $T(\bar{x}) = D(\bar{x})$.*

Proof. Consider an arbitrary direction $d \in D(\bar{x})$ and $\overline{d}$ given by MFCQ. Given $\lambda \in (0,1]$, define

$$\hat{d} = (1-\lambda)d + \lambda\overline{d}.$$

We will prove that $\hat{d} \in T(\bar{x})$ for all $\lambda \in (0,1)$.

Denote $M = \nabla h(\bar{x})^T$. By MFCQ, $\operatorname{rank}(M) = m$. Consider the matrix $Z = (v^1 \cdots v^{n-m}) \in \mathbb{R}^{n\times(n-m)}$, whose columns are a basis of $\mathcal{N}(M)$. Since $\{\nabla h_1(\bar{x}), \ldots, \nabla h_m(\bar{x})\}$ is a basis of $\operatorname{Im}(M^T)$, the matrix $\begin{pmatrix} M \\ Z^T \end{pmatrix}$ is nonsingular. Define $\varphi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ by

$$\varphi\begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} h(x) \\ Z^T(x - \bar{x} - t\hat{d}) \end{pmatrix}.$$

Since $\nabla_x \varphi^T = \begin{pmatrix} M \\ Z^T \end{pmatrix}$ is nonsingular, by the Implicit Function Theorem, there exists a differentiable curve $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that $\varphi\begin{pmatrix} \gamma(t) \\ t \end{pmatrix} = 0$, for all $t \in (-\varepsilon, \varepsilon)$. Thus

$$h(\gamma(t)) = 0 \qquad \text{and} \qquad Z^T(\gamma(t) - \bar{x} - t\hat{d}) = 0. \tag{3.2}$$

Since $\varphi\begin{pmatrix} \bar{x} \\ 0 \end{pmatrix} = 0$, by unicity of γ, we have $\gamma(0) = \bar{x}$. Taking the derivative at $t = 0$ on the first equation in (3.2), we obtain

$$M\gamma'(0) = 0. \tag{3.3}$$

Using again (3.2), for $t \neq 0$,

$$Z^T \left(\frac{\gamma(t) - \bar{x}}{t} - \hat{d} \right) = 0.$$

Taking the limit, we have

$$Z^T \gamma'(0) = Z^T \hat{d}. \tag{3.4}$$

As $d, \overline{d} \in D(\bar{x})$, $M\hat{d} = 0$. Using this, (3.3) and (3.4), we obtain

$$\begin{pmatrix} M \\ Z^T \end{pmatrix} \gamma'(0) = \begin{pmatrix} M \\ Z^T \end{pmatrix} \hat{d},$$

which implies $\hat{d} = \gamma'(0)$. By Lemma 3.6 there exists a sequence (x^k) with $h(x^k) = 0$, $x^k \to \bar{x}$ and

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \frac{\hat{d}}{\|\hat{d}\|}.$$

To conclude that $\hat{d} \in T(\bar{x})$ it is enough to show that $g(x^k) \leq 0$ for all k sufficiently large.

For $i \notin A(\bar{x})$, $g(\bar{x}) < 0$. By continuity of g, $g(x^k) \leq 0$ for all k sufficiently large.

For $i \in A(\bar{x})$, we have $\nabla g_i(\bar{x})^T d \leq 0$ and $\nabla g_i(\bar{x})^T \overline{d} < 0$. Consequently $\nabla g_i(\bar{x})^T \hat{d} < 0$. From the smoothness of g_i it follows that

$$g_i(x^k) = g_i(\bar{x}) + \nabla g_i(\bar{x})^T (x^k - \bar{x}) + o(\|x^k - \bar{x}\|).$$

Thus,

$$\frac{g_i(x^k)}{\|x^k - \bar{x}\|} = \nabla g_i(\bar{x})^T \frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} + \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} \to \nabla g_i(\bar{x})^T \frac{\hat{d}}{\|\hat{d}\|} < 0.$$

Therefore, $g_i(x^k) < 0$, for all k sufficiently large. So, $\hat{d} \in T(\bar{x})$, which implies $d \in T(\bar{x})$, since $T(\bar{x})$ is a closed set. □

The example below shows that quasiregularity does not imply MFCQ. Consider the functions $g_j : \mathbb{R}^2 \to \mathbb{R}$ $(j = 1, 2)$ defined by

$$g_1(x) = -x_1^2 + x_2 \qquad \text{and} \qquad g_2(x) = -x_1^2 - x_2$$

and the feasible point $\bar{x} = (0, 0)^T$. In this case, $D(\bar{x}) = \{(d_1, 0) \mid d_1 \in \mathbb{R}\}$. For obtaining $T(\bar{x})$, consider the sequence $x^k = (t_k, 0)$ with $t_k \to 0$ and $t_k > 0$. Thus $x^k \to \bar{x}$ and

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} = \frac{(t_k, 0)}{t_k} = (1, 0).$$

So $d = (1, 0)$ is a tangent direction. In the same way, for $t_k < 0$, we see that $d = (-1, 0)$ also is a tangent direction. Since $T(\bar{x})$ is a cone, we have $(d_1, 0) \in T(\bar{x})$ for all $d_1 \in \mathbb{R}$. So, $D(\bar{x}) \subset T(\bar{x})$. Using Lemma 2.7 we conclude that $T(\bar{x}) = D(\bar{x})$. Note that there is not $d \in \mathbb{R}^2$ such that $\nabla g_i(\bar{x})^T d < 0$, for $i = 1, 2$. Furthermore, $\{\nabla g_1(\bar{x}), \nabla g_2(\bar{x})\}$ is linearly dependent. Then, $\bar{x}$ does not satisfy neither MFCQ nor LICQ.

4. Conclusions. We have proved the classical KKT optimality condition for nonlinear programming problem assuming the equality of the polar of the tangent cone and the polar of the first order feasible variations cone. Since this condition is a somewhat abstract property, it is useful to have more readily verifiable conditions for the admittance of Lagrange multipliers. Such conditions called constraint qualifications have been investigated extensively in the literature. We discussed some of them such as Slater, linear independence of gradients (LICQ), Mangasarian-Fromovitz (MFCQ) and quasiregularity condition.

However, there are other constraint qualifications not discussed in this paper. Hestenes [9] introduced the *quasinormality* condition and showed that it implies quasiregularity. Recently, Andreani, Martínez and Schuverdt [2, 15] proved that the *Constant Positive Linear Dependence* (CPLD) introduced by Qi and Wei [14] is weaker than MFCQ and implies quasinormality. A drawback of the MFCQ, as well as LICQ, is that they are no longer valid if we substitute an equality $h_i(x) = 0$ by two equivalent inequalities $h_i(x) \leq 0$ and $-h_i(x) \leq 0$. The *Constant Rank Contraint Qualification* (CRCQ) introduced by Janin [12] is not affected by this change and implies CPLD.

REFERENCES

[1] J. Abadie. On the Kuhn-Tucker Theorem. In J. Abadie, editor, *Nonlinear Programming*, pages 19–36. North Holland, Amsterdam, 1967.

[2] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125(2):473–485, 2005.

[3] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming Theory and Algorithms.* John Wiley, New York, 2nd edition, 1993.

[4] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts, 1995.

[5] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar. *Convex Analysis and Optimization.* Athena Scientific, Belmont, Massachusetts, 2003.

[6] M. L. Flegel. *Constraint Qualification and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints.* PhD thesis, Institute of Applied Mathematics and Statistics, University of Würzburg, Alemanha, 2005.

[7] F. J. Gould and J. W. Tolle. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20:164–172, 1971.

[8] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7:232–241, 1969.

[9] M. R. Hestenes. *Optimization Theory: The finite dimensional case.* John Wiley, New York, 1975.

[10] T. Hoheisel and C. Kanzow. On the Abadie and Guignard constraint qualification for mathematical programs with constraints. Preprint 272, Institute of Mathematics, University of Würzburg, Würzburg, September 2006.

[11] A. Izmailov and M. Solodov. *Condições de otimalidade, elementos de análise convexa e de dualidade*, volume 1. IMPA, 2005.

[12] R. Janin. Direction derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, pages 127–138, 1984.

[13] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. and Appl*, 17:37–47, 1967.

[14] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to sqp methods. *SIAM J. on Optimization*, 10(4):963–981, 2000.

[15] M. L. Schuverdt. *Métodos de Langrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante.* PhD thesis, Instituto de Matemática Estatística e Computação Científica, Universidade Estadual de Campinas, Brasil, 2006.