SIAM J. OPTIM.
Vol. 34, No. 2, pp. 1515–1539


# ON ENHANCED KKT OPTIMALITY CONDITIONS FOR SMOOTH NONLINEAR OPTIMIZATION*

ROBERTO ANDREANI[†], MARÍA L. SCHUVERDT[‡], AND LEONARDO D. SECCHIN[§]

**Abstract.** The Fritz John (FJ) and Karush–Kuhn–Tucker (KKT) conditions are fundamental tools for characterizing minimizers and form the basis of almost all methods for constrained optimization. Since the seminal works of Fritz John, Karush, Kuhn, and Tucker, FJ/KKT conditions have been enhanced by adding extra necessary conditions. Such an extension was initially proposed by Hestenes in the 1970s and later extensively studied by Bertsekas and collaborators. In this work, we revisit enhanced KKT stationarity for standard (smooth) nonlinear programming. We argue that every KKT point satisfies the usual enhanced versions found in the literature. Therefore, enhanced KKT stationarity only concerns the Lagrange multipliers. We then analyze some properties of the corresponding multipliers under the quasi-normality constraint qualification (QNCQ), showing in particular that the set of so-called quasinormal multipliers is compact under QNCQ. Also, we report some consequences of introducing an extra abstract constraint to the problem. Given that enhanced FJ/KKT concepts are obtained by aggregating sequential conditions to FJ/KKT, we discuss the relevance of our findings with respect to the well-known sequential optimality conditions, which have been crucial in generalizing the global convergence of a well-established safeguarded augmented Lagrangian method. Finally, we apply our theory to mathematical programs with complementarity constraints and multiobjective problems, improving and elucidating previous results in the literature.

**Key words.** enhanced Fritz John, enhanced KKT, quasinormal multipliers, enhanced multipliers, quasi-normality, augmented Lagrangian method

**MSC codes.** 90C30, 90C46, 65K05

**DOI.** 10.1137/22M1539678

## 1. Introduction.

We consider the general constrained optimization problem

$$\text{(P)} \qquad \min_x f(x) \quad \text{subject to} \quad h(x)=0, \quad g(x)\leq 0,$$

where $f:\mathbb{R}^n\to\mathbb{R}$, $h:\mathbb{R}^n\to\mathbb{R}^m$, and $g:\mathbb{R}^n\to\mathbb{R}^p$ are continuously differentiable functions.

It is well known that it is not possible to define a practical algorithm that always reaches global minimizers of (P) for general nonlinear constraints. Even local minimizers are impracticable to guarantee, at least when convexity is not assumed. Practical algorithms aim to find a reasonable stationary point, that is, a computable point that exhibits key properties of minimizers. In this sense, the most important tool to characterize minimizers of (P) is the Karush–Kuhn–Tucker (KKT) conditions.

*Received by the editors December 6, 2022; accepted for publication (in revised form) January 15, 2024; published electronically April 17, 2024.
https://doi.org/10.1137/22M1539678

**Funding:** This work was partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2018/24293-0 and 2017/18308-2), CNPq (grants 306988/2021-6, 309136/2021-0, and 407147/2023-3), PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016), and ANPCyT (grants PICT 2016-0921 and PICT 2019-02172, Argentina). The third author was also partially supported by the University of La Plata during his visit to the university.

[†]Department of Applied Mathematics, UNICAMP (Universidade Estadual de Campinas), Campinas, SP, Brazil (andreani@ime.unicamp.br).
[‡]CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina (schuverd@mate.unlp.edu.ar).
[§]Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, ES, Brazil (leonardo.secchin@ufes.br).

They are used to assert theoretical convergence of practically every method in constrained optimization and have been specialized/adapted to different specific contexts, such as multiobjective optimization and nonsmooth optimization among others. Furthermore, KKT conditions inspire practical stopping criteria for several algorithms [5, 7, 12, 19, 20, 23, 25, 40].

Contrary to the KKT conditions, the well-known Fritz John (FJ) [30] conditions do not require the fulfillment of any constraint qualification (CQ) to hold at minimizers in general. They gained much attention with the works of Hestenes [28], Bertsekas, Ozdaglar, and collaborators [13, 14, 15, 16], Ye and Zhang [41, 42], and others, who dealt with extensions of the original FJ conditions, the so-called *enhanced FJ conditions.* Hestenes [28] was the first to propose such an extension, showing the existence of special (Lagrange) multipliers satisfying a property associated with infeasible sequences (see Theorem 2.1, item 3(a)). This enhanced FJ version is intimately linked to the pure external penalty method [13] and puts the quasi-normality CQ (QNCQ; see Definition 4.1) in evidence. Later on, other properties were incorporated, resulting in different enhanced FJ conditions (see Theorem 2.1, items 3(b) and (c)). In turn, enhanced KKT conditions were naturally derived from the enhanced FJ stationarity by imposing a nonnull multiplier associated with the objective function gradient (compare item 1 of Theorem 2.1 and item 1 of Definition 2.2).

All enhanced FJ/KKT conditions consist of the usual FJ/KKT ones together with extra conditions. Thus, one might expect that enhanced stationarity characterizes minima better than the usual FJ/KKT conditions, possibly, of course, under some CQ. Surprisingly, this is not true from the primal point of view. Specifically, we argue that not only local minimizers of (P) satisfy the enhanced KKT conditions (possibly under a CQ), as done in previous works, but also *all* KKT points. That is, from the primal point of view, enhanced KKT and usual KKT are equivalent. Thus, the difference only concerns the multipliers. We then investigate new properties of such multipliers, in particular, those called *quasinormal* in the literature. In [41] it was proved that the set of quasinormal multipliers $M_Q(x^*)$ associated with a KKT point $x^*$ is bounded if QNCQ holds at $x^*$. Here, we prove that $M_Q(x^*)$ is always closed, thus compact under QNCQ. In this sense, QNCQ has a similar status to the Mangasarian–Fromovitz CQ for the usual multiplier set $M(x^*)$ [24]. Contrary to $M(x^*)$, we show that $M_Q(x^*)$ is not convex in general. The relationship between quasinormal and *informative* multipliers [14], which are associated with the most stringent enhanced FJ version, is discussed. We introduce and investigate a novel intermediate concept of multiplier that we call *enhanced,* which provides sensitivity information regarding constraints in a way similar to that of informative multipliers. Furthermore, we apply our theory to mathematical programs with complementarity constraints and multiobjective optimization, improving and elucidating previous results in the literature.

Another relevant issue is the impact of enhanced FJ/KKT conditions and QNCQ on constrained nonlinear optimization methods. For many years these concepts remained restricted to the pure external method, so their applicability to practical algorithms was limited. Only recently, we proved [4] that the safeguarded augmented Lagrangian method known as ALGENCAN [3] converges to KKT points under QNCQ. ALGENCAN has an "official," mature, general-purpose implementation that has been successfully used in many applications [18]. The theory developed in [4] is supported by the sequential optimality condition *positive approximate KKT* (PAKKT), which captures the connection between the signs of the multipliers generated by the method and the infeasibility of its primal sequence. PAKKT is based on Hestenes' enhanced

FJ conditions [28]. One interesting consequence of PAKKT is that the sequences of multipliers generated by ALGENCAN are bounded if QNCQ holds at the limit point. This result is surprising, as QNCQ does not imply the boundedness of the set of usual multipliers. The validity of enhanced stationarity at all KKT points shown in this paper is an important step in properly studying practical methods that potentially generate quasinormal multipliers, at least under QNCQ. We provide some insights in this direction concerning ALGENCAN, paving the way to improving the previous results about the boundedness of the penalty parameter [3, 17], which requires, for example, uniqueness of the usual multiplier. This is an important issue regarding the stability of the method.

This paper is organized as follows. In section 2 we present the enhanced FJ/KKT stationarities. In section 3 we show that enhanced and classical KKT conditions are equivalent. Section 4 is devoted to discussing/proving properties of the set of quasinormal/enhanced multipliers, such as compactness and convexity. We also consider the inclusion of abstract constraints. Section 5 is dedicated to discussing the consequences of enhanced stationarity for ALGENCAN. In section 6 we apply our theory to the widely studied classes of problems cited above, namely, mathematical programs with complementarity constraints and multiobjective optimization. Finally, section 7 discusses our conclusions and future work.

*Notation.* $\operatorname{sgn} a$ will denote the sign function, that is, $\operatorname{sgn} a = 1$ if $a > 0$ and $\operatorname{sgn} a = -1$ if $a < 0$. We define $I_g(z) = \{j \mid g_j(z) = 0\}$. In addition, $\|\cdot\|$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ are an arbitrary norm, the Euclidean norm, and the sup-norm, respectively. Given $z \in \mathbb{R}^q$, we define $z_+ = (\max\{0, z_1\}, \dots, \max\{0, z_q\}) \in \mathbb{R}^q_+$. We use the "$o$" and "$O$" notation: given two sequences of real numbers $\{a_k\}$ and $\{b_k\}$, we write $b_k = o(a_k)$ (respectively, $b_k = O(a_k)$) to indicate that there is a sequence $\{m_k\}$, $m_k > 0$, converging to zero (respectively, a constant $M > 0$) such that $|b_k| \leq m_k|a_k|$ (respectively, $|b_k| \leq M|a_k|$) for all $k$, where $|\cdot|$ denotes the absolute value. Given a differentiable function $c : \mathbb{R}^n \to \mathbb{R}^q$, we denote by $\nabla c(x)$ the $n \times q$ matrix whose columns are the gradients $\nabla c_i(x)$, $i = 1, \dots, q$, i.e., $\nabla c(x)$ is the Jacobian transpose of $c$ at $x$. For convenience, we can interpret a vector $v \in \mathbb{R}^q$ as a matrix $q \times 1$; in this case, $v^t$ denotes its transpose.

## 2. Enhanced stationarity for smooth problems.

Enhanced FJ/KKT versions have been proposed in the literature since at least the 1970s [28] and gained prominence with the work of Bertsekas and collaborators [13, 14, 15, 16]. The most general enhanced FJ conditions for (P) are presented in the next theorem.

THEOREM 2.1 (see [15, Proposition 2.1]). *Let $x^*$ be a local minimizer of the problem* (P). *Then there are $\sigma \in \mathbb{R}_+$, $\lambda \in \mathbb{R}^m$, and $\mu \in \mathbb{R}^p_+$ such that*

1. $\sigma \nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$;
2. $(\sigma, \lambda, \mu) \neq 0$;
3. *if $I_{\neq} \cup J_+ \neq \emptyset$, where $I_{\neq} = \{i \mid \lambda_i \neq 0\}$ and $J_+ = \{j \mid \mu_j > 0\}$, then there is a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to $x^*$ such that, for all $k$, the following conditions are valid:*
   - (a) $\lambda_i h_i(x^k) > 0 \ \ \forall i \in I_{\neq}$, *and* $\mu_j g_j(x^k) > 0 \ \ \forall j \in J_+$;
   - (b) $f(x^k) < f(x^*)$;
   - (c) $|h_i(x^k)| = o(w(x^k)) \ \ \forall i \notin I_{\neq}$, *and* $g_j(x^k)_+ = o(w(x^k)) \ \ \forall j \notin J_+$, *where* $w(x) = \min\left\{\min_{i \in I_{\neq}} |h_i(x)|,\ \min_{j \in J_+} g_j(x)_+\right\}$.

Item 3(a) implies the complementarity of the classical KKT conditions, that is, $\mu_j = 0$ for all $j \notin I_g(x^*)$. Item 3(b) says that when we are at the boundary of the feasible set with nonnull multipliers ($I_{\neq} \cup J_+ \neq \emptyset$), it is possible to reach $x^*$ "from outside" the feasible set, so that $f(x^k)$ is smaller than the optimal value; this is a typical behavior of external penalty approaches. Finally, item 3(c) carries sensitivity properties of the constraints associated with null multipliers, which include inactive inequality constraints.

Enhanced KKT conditions are obtained by setting $\sigma = 1$ in Theorem 2.1 and imposing one or more conditions from item 3. Such conditions have been considered in the literature since the late 1990s. In [13, 41], only items 1 and 3(a) are considered; in [15, 14], items 3(b) and (c) are aggregated. These works state such conditions only for local minimizers, like Theorem 2.1, and their proofs were carried out using the pure external penalty method. Here, instead, we aim to establish the link between enhanced KKT points not only with qualified local minima, but also with usual KKT points. For this purpose, first we define an intermediate enhanced KKT stationary concept by assuming items 1 and 3(a) and (b) and changing 3(c), putting "$O(w(x^k))$" instead of "$o(w(x^k))$." The "$O$" notation is weaker than "$o$," so it results in a less stringent sensitivity measure of the constraints associated with null multipliers.

Definition 2.2. *We say that a feasible $x^*$ for* (P) *is an* enhanced KKT *(E-KKT) point if there are $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ such that*

1. $\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$;
2. *if $I_{\neq} \cup J_+ \neq \emptyset$, where*

$$I_{\neq} = \{i \mid \lambda_i \neq 0\} \quad \text{and} \quad J_+ = \{j \mid \mu_j > 0\}, \tag{2.1}$$

*then there is a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to $x^*$ such that, for all $k$, the following conditions are valid:*
   (a) $\lambda_i h_i(x^k) > 0 \;\; \forall i \in I_{\neq}$, *and* $\mu_j g_j(x^k) > 0 \;\; \forall j \in J_+$;
   (b) $f(x^k) < f(x^*)$;
   (c) $|h_i(x^k)| = O(w(x^k)) \;\; \forall i \notin I_{\neq}$, *and* $g_j(x^k)_+ = O(w(x^k)) \;\; \forall j \notin J_+$, *where*

$$w(x) = \min\left\{ \min_{i \in I_{\neq}} |h_i(x)|, \; \min_{j \in J_+} g_j(x)_+ \right\}.$$

Given a feasible point $x^*$ for (P), we define the set of usual associated Lagrange multipliers and the set of *quasinormal multipliers* [41] by

$$M(x^*) = \left\{ (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \;\middle|\; \begin{array}{l} \nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0, \\ \mu_j = 0 \quad \forall j \notin I_g(x^*) \end{array} \right\}$$

and

$$M_Q(x^*) = \left\{ (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \;\middle|\; \begin{array}{l} \nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0, \quad \exists \{x^k\} \to x^* \\ \text{such that item 2(a) of Definition 2.2 holds} \end{array} \right\},$$

respectively. Of course, as with KKT, some additional hypothesis (constraint qualification) is needed to ensure that a local minimizer satisfies enhanced KKT stationarity.

There are others sets of Lagrange multipliers in the literature. In [15], those satisfying items 1 and 2(a) and (b) of Definition 2.2 were called *strong multipliers.* Also in [15], the multipliers satisfying item 1 with $\sigma = 1$ and items 3(a)–(c) of Theorem 2.1 were called *informative multipliers*;; they will be considered in subsection 4.3. Here, we define an intermediate concept, which we will refer to as *enhanced multipliers*:

$$M_E(x^*) = \left\{ (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \;\middle|\; \begin{array}{l} \nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0, \quad \exists \{x^k\} \to x^* \\ \text{such that items 2(a)–(c) of Definition 2.2 hold} \end{array} \right\}.$$

Clearly, $M_E(x^*) \subset M_Q(x^*) \subset M(x^*)$ and every informative multiplier is an enhanced multiplier, which in turn is strong. Examples 2 and 5 below show that $M(x^*) \not\subset M_Q(x^*)$ in general. The next example shows that $M_Q(x^*) \not\subset M_E(x^*)$ in general.

*Example* 1. Consider the bidimensional problem

$$\min x_1 \quad \text{subject to} \quad x_1 = 0, \quad x_2 \leq 0, \quad x_2^3 = 0$$

and the unique feasible point $x^* = (0,0)$. Defining $x^k = (-1/k, 1/k)$, $k \geq 1$, we conclude that any $\omega_a = (-1,0,a)$, $a \geq 0$, is in $M_Q(x^*)$ (note that $x_1^k < 0$ for all $k$ and thus $\omega_a$ is also a strong multiplier). On the other hand, for $a > 0$ we have $w(x) = \min\{|x_1|, |x_2^3|\} \leq |x_2^3|$; thus $|x_2| = O(w(x))$ is impossible for any sequence $\{x^k\}$ converging to $x^*$ with $x_2^k > 0$ for all $k$. Therefore, $\omega_a \notin M_E(x^*)$ for all $a > 0$. Actually, it is easy to see that $\omega_0$ is the unique element of $M_E(x^*)$.

The elements of $M_Q(x^*)$ are called quasinormal multipliers because of their close connection with QNCQ (see Definition 4.1). Next, we establish new connections between E-KKT and KKT, as well as between $M_Q(x^*)$ and/or $M_E(x^*)$ with QNCQ.

**3. Equivalence between enhanced and classical KKT.** The main purpose of this section is to highlight the equivalence between KKT and E-KKT, that is, that every KKT point $x^*$ is E-KKT and vice versa. We start by recalling the primal and dual versions of MFCQ for (P).

DEFINITION 3.1. *We say that a feasible $x^*$ for* (P) *satisfies the* Mangasarian–Fromovitz CQ *(MFCQ) if $\nabla h_i(x^*)$, $i = 1,\ldots,m$, are linearly independent and there exists $d \in \mathbb{R}^n$ such that*

$$\nabla h_i(x^*)^t d = 0, \; i = 1,\ldots,m, \qquad \nabla g_j(x^*)^t d < 0 \;\; \forall j \in I_g(x^*).$$

THEOREM 3.2 (see [28]). *A feasible $x^*$ for* (P) *satisfies MFCQ if and only if the system*

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in I_g(x^*)} \mu_j \nabla g_j(x^*) = 0, \quad \mu \geq 0,$$

*admits only the trivial solution $(\lambda, \mu) = 0$.*

The next lemma contains an interesting geometric property of smooth constraints. Roughly speaking, it states that for active constraints with linearly independent gradients at $x^*$, it is possible to approach $x^*$ from "either side of each of these constraints." This result constitutes the core idea behind the proof of [4, Lemma 2.6]. Although the arguments are similar, we provide a complete proof for the sake of clarification.

LEMMA 3.3. *Let $c : \mathbb{R}^n \to \mathbb{R}^q$ be a differentiable function at $x^*$ such that $c(x^*) = 0$. Suppose that $\nabla c_i(x^*)$, $i = 1,\ldots,q$, are linearly independent. Then for any fixed $s \in \{-1,+1\}^q$, there exist $\gamma > 0$ and a sequence $\{x^k\} \subset \mathbb{R}^n \backslash \{x^*\}$ converging to $x^*$ such that $s_i c_i(x^k) \geq \gamma \|x^k - x^*\|_2$ for all $k$ and $i = 1,\ldots,q$.*

*Proof.* Let $\mathcal{I}_+ = \{i \mid s_i = 1\}$, $\mathcal{I}_- = \{i \mid s_i = -1\}$, and

$$D = \{x \in \mathbb{R}^n \mid c_j(x) \geq 0, \; j \in \mathcal{I}_+, \quad c_l(x) \leq 0, \; l \in \mathcal{I}_-\}.$$

This set is nonempty since $c(x^*) = 0$. As $\nabla c_i(x^*)$, $i = 1,\ldots,q$, are linearly independent, the gradients of all constraints in $D$ evaluated at $x^*$ are also linearly independent. Thus, MFCQ holds at $x^*$, and so there is an unitary $d \in \mathbb{R}^n$ such that

$$\nabla c_j(x^*)^t d > 0, \quad \nabla c_l(x^*)^t d < 0, \quad j \in \mathcal{I}_+,\ l \in \mathcal{I}_-. \tag{3.1}$$

Defining $x^k = x^* + d/k$ for all $k \geq 1$, we have $\|x^k - x^*\|_2 = 1/k \to 0$ and $(x^k - x^*)/\|x^k - x^*\|_2 = d$ for all $k$. By the differentiability of $c_i$ at $x^*$, $i = 1, \ldots, q$, we have

$$c_i(x^k) = c_i(x^* + d/k) = c_i(x^*) + 1/k \nabla c_i(x^*)^t d + o(d/k). \tag{3.2}$$

As $c(x^*) = 0$, dividing the above expression by $\|x^k - x^*\|_2 = 1/k$ and using (3.1), we obtain

$$c_j(x^k) \geq 1/2 \nabla c_j(x^*)^t d \cdot \|x^k - x^*\|_2 \ \ \forall j \in \mathcal{I}_+,$$
$$c_l(x^k) \leq 1/2 \nabla c_l(x^*)^t d \cdot \|x^k - x^*\|_2 \ \ \forall l \in \mathcal{I}_-$$

for all $k$ large enough. Therefore, $s_i c_i(x^k) \geq \gamma \|x^k - x^*\|_2$ for all $i = 1, \ldots, q$ and $k$ large enough by taking $\gamma = 1/2 \cdot \min\{\nabla c_j(x^*)^t d,\ -\nabla c_l(x^*)^t d \mid j \in \mathcal{I}_+,\ l \in \mathcal{I}_-\} > 0$. □

The next useful result is a consequence of the well-known Carathéodory lemma. It is a simplified version of [8, Lemma 1].

LEMMA 3.4. *Let $\mathcal{Z}$ be a finite set of indices, and suppose that $z = \sum_{i \in \mathcal{Z}} \alpha_i v_i \neq 0$ is a linear combination of vectors $v_i \in \mathbb{R}^n$. If $\alpha_i \neq 0$ for all $i \in \mathcal{Z}$, then there are $\widehat{\mathcal{Z}} \subset \mathcal{Z}$ and $\widehat{\alpha}_i$, $i \in \widehat{\mathcal{Z}}$, such that $z = \sum_{i \in \widehat{\mathcal{Z}}} \widehat{\alpha}_i v_i$, $\alpha_i \cdot \widehat{\alpha}_i > 0$ for all $i \in \widehat{\mathcal{Z}}$, and $\{v_i\}_{i \in \widehat{\mathcal{Z}}}$ is linearly independent.*

In the following, we present one of the main results of this paper. We show that the stationarity concept of Definition 2.2 is equivalent to the classical KKT conditions. This is an interesting issue since

- classical KKT conditions are at the root of nonlinear programming. They have been used for decades, serving as the theoretical basis for practically all existing methods and have been adapted to nonstandard contexts such as multiobjective and nonsmooth optimization;
- enhanced FJ/KKT conditions have been extensively studied since at least the 1990s (see, for example, [13]);
- and finally, the result says that, from the primal point of view, the classical KKT conditions characterize minimizers of (P) as well as their enhanced versions. This becomes particularly interesting when we confront item 2(a) of Definition 2.2 with interior point strategies.

THEOREM 3.5. *Every KKT point $x^*$ is E-KKT and vice versa.*

*Proof.* Trivially, every E-KKT point is KKT. Let us prove the converse. Let $x^*$ be a KKT point, $\lambda$ and $\mu$ be any associated multipliers, and $I_{\neq}$ and $J_+$ be the induced sets as in (2.1). Clearly, it is sufficient to consider only the nontrivial case $\nabla f(x^*) \neq 0$, which implies $I_{\neq} \cup J_+ \neq \emptyset$. Lemma 3.4 asserts that there are sets $\mathcal{I} \subset I_{\neq}$ and $\mathcal{J} \subset J_+$, not both empty, and vectors $\widehat{\lambda}_{\mathcal{I}}$, $\widehat{\mu}_{\mathcal{J}}$ such that

$$\nabla f(x^*) + \sum_{i \in \mathcal{I}} \widehat{\lambda}_i \nabla h_i(x^*) + \sum_{j \in \mathcal{J}} \widehat{\mu}_j \nabla g_j(x^*) = 0, \tag{3.3}$$

$\widehat{\lambda}_i \neq 0$ for all $i \in \mathcal{I}$, $\widehat{\mu}_j > 0$ for all $j \in \mathcal{J}$, and the gradients of constraints in (3.3) are linearly independent. Applying Lemma 3.3 on such gradients, we obtain a sequence $\{x^k\}$ converging to $x^*$ and $\gamma > 0$ satisfying

$$\operatorname{sgn}(\widehat{\lambda}_i) \cdot h_i(x^k) \geq \gamma \|x^k - x^*\|_2 \ \ \forall i \in \mathcal{I} \quad \text{and} \quad g_j(x^k) \geq \gamma \|x^k - x^*\|_2 \ \ \forall j \in \mathcal{J} \tag{3.4}$$

for all $k$. Defining $\widehat{\lambda}_i = 0$, $i \notin \mathcal{I}$, and $\widehat{\mu}_j = 0$, $j \notin \mathcal{J}$, items 1 and 2(a) of Definition 2.2 are valid with the sequence $\{x^k\}$ and multipliers $\widehat{\lambda}$, $\widehat{\mu}$.

As in the proof of Lemma 3.3, we have $x^k = x^* + d/k$ where $d$ is unitary satisfying

$$\operatorname{sgn}(\widehat{\lambda}_i) \cdot \nabla h_i(x^*)^t d > 0, \quad \nabla g_j(x^*)^t d > 0, \quad i \in \mathcal{I},\ j \in \mathcal{J}.$$

Multiplying (3.3) by $d$ and using the inequalities above, we conclude that $\nabla f(x^*)^t d < 0$. Thus, by the first-order Taylor expansion of $f$ around $x^*$ with increment $d/k$ (see (3.2)), we obtain item 2(b) of Definition 2.2 using $x^k$ for all $k$ sufficiently large.

Finally, we claim that the same sequence $\{x^k\}$ satisfies item 2(c) of Definition 2.2 for all $k$ large enough. In fact, as $g$ and $h$ are differentiable at $x^*$, taking

$$L > \max\{|\nabla h_i(x^*)^t d|, |\nabla g_j(x^*)^t d| \mid i \notin \mathcal{I},\ j \notin \mathcal{J}\},$$

we have

$$|h_i(x^k)| = |h_i(x^k) - h_i(x^*)| \leq L\|x^k - x^*\|_2, \tag{3.5}$$

$$g_j(x^k)_+ \leq \left[g_j(x^k) - g_j(x^*)\right]_+ \leq L\|x^k - x^*\|_2 \tag{3.6}$$

for all $k$ sufficiently large, $i \notin \mathcal{I}$ and $j \notin \mathcal{J}$. Considering the multipliers $\widehat{\lambda}$ and $\widehat{\mu}$ in (3.3), $w(x^k)$ takes the form

$$w(x^k) = \min\left\{\min_{i\in\mathcal{I}} |h_i(x^k)|,\ \min_{j\in\mathcal{J}} g_j(x^k)_+\right\}.$$

By (3.4), (3.5), and (3.6) we have, for all $k$ large enough,

$$w(x^k) \geq \gamma\|x^k - x^*\|_2 = \left(\frac{\gamma}{L}\right) L\|x^k - x^*\|_2 \geq \frac{\gamma}{L} \max\left\{\max_{i\notin\mathcal{I}} |h_i(x^k)|,\ \max_{j\notin\mathcal{J}} g_j(x^k)_+\right\}.$$

Thus $\{x^k\}$ fulfills item 2(c) of Definition 2.2 as we wanted, concluding the proof. ∎

*Remark* 3.6. In [26], the equivalence between KKT points and those satisfying items 1 and 2(a) of Definition 2.2 was proved in the context of multiobjective optimization, which clearly includes the case of a single objective function. On the other hand, Theorem 3.5 includes items 2(b) and (c) of Definition 2.2, which, due to the strength of the enhanced multipliers over the quasinormal ones, results in a stronger result.

It is easy to exhibit examples where a multiplier vector fulfills the classical KKT conditions, but not the enhanced KKT (see Example 2). From the proof of Theorem 3.5, given an arbitrary multiplier vector $(\lambda, \mu)$, the vector $(\widehat{\lambda}, \widehat{\mu})$ obtained by applying Lemma 3.4 on $(\lambda, \mu)$ is suitable for both KKT and E-KKT stationarity. This lemma also ensures that these multipliers preserve signs, that is, $\lambda_i\widehat{\lambda}_i \geq 0$ $\forall i$ and $\mu_j\widehat{\mu}_j \geq 0$ $\forall j$. To put such multipliers in perspective, we say that a multiplier vector $(\lambda, \mu)$ has a *linearly independent support* if the gradients of equality and inequality constraints associated with their nonnull entries are linearly independent. Note that a possible result of an application of Lemma 3.4 on $(\lambda, \mu)$ with linearly independent support is $(\lambda, \mu)$ itself. Also, it is clear that every KKT point admits such multipliers.

COROLLARY 3.7. *Let $x^*$ be a KKT point. Then any multiplier vector with linearly independent support satisfies all the conditions of Definition* 2.2. *In other words, these multipliers are enhanced (and therefore, quasinormal).*

*Remark* 3.8. Bertsekas, Nedic, and Ozdaglar [14] proved Theorem 2.1 for the case where $f$, $h$, and $g$ are continuously differentiable. Ye and Zhang [41] provided a nonsmooth version of this result, without items 2(b) and (c), where all data functions are only Lipschitz continuous around $x^*$. It is straightforward to verify that Theorem 3.5 remains valid if we suppose that $f$, $h$, and $g$ are differentiable, not necessarily with continuous derivatives (Lemma 3.3 is valid with differentiability only). As a consequence, Corollary 3.7 is still valid in this case.

In a reasoning similar to Corollary 3.7, we may ask which known CQs ensure that *any* Lagrange multiplier is enhanced/quasinormal. Certainly the *linear independence of the gradients of the active constraints* (LICQ) is one of them, since in this case the unique multiplier is one of those from Corollary 3.7. The same does not occur when constraints are linear and, consequently, with any implied CQ.

*Example* 2. The point $x^* = 0$ is KKT for the problem of minimizing $x$ subject to $g_1(x) = x \le 0$ and $g_2(x) = -x \le 0$ with, for instance, multiplier vector $\mu = (1,2)$. However, for any sequence $\{x^k\}$ converging to $x^*$, we cannot have $\mu_1 g_1(x^k) > 0$ and $\mu_2 g_2(x^k) > 0$ simultaneously.

Next, we prove that MFCQ, like LICQ, ensures that any multiplier vector works for E-KKT.

THEOREM 3.9. *Let $x^*$ be a KKT point. If MFCQ holds at $x^*$ then $M(x^*) = M_Q(x^*) = M_E(x^*)$.*

*Proof.* The proof is an adaptation of that of Theorem 3.5. Let $(\lambda,\mu) \in M(x^*)$, and let $I_{\neq}$, $J_+$ be defined as in (2.1). We have

$$\nabla f(x^*) + \sum_{i \in I_{\neq}} \lambda_i \nabla h_i(x^*) + \sum_{j \in J_+} \mu_j \nabla g_j(x^*) = 0.$$

If $\nabla f(x^*) = 0$, then $M_E(x^*) = M_Q(x^*) = M(x^*) = \{0\}$ by Theorem 3.2. Suppose that $\nabla f(x^*) \neq 0$. Let us define

$$D' = \{x \in \mathbb{R}^n \mid h_i(x) \ge 0,\ h_l(x) \le 0,\ g_j(x) \ge 0, \quad i \in I_+,\ l \in I_-,\ j \in J_+\},$$

where $I_+ = \{i \mid \lambda_i > 0\}$ and $I_- = \{i \mid \lambda_i < 0\}$. Since $x^*$ conforms to MFCQ, we claim that MFCQ holds at $x^*$ considering $x \in D'$. In fact, if there were $0 \le (\bar\lambda^+, \bar\lambda^-, \bar\mu) \neq 0$ such that

$$-\sum_{i \in I_+} \bar\lambda_i^+ \nabla h_i(x^*) + \sum_{l \in I_-} \bar\lambda_i^- \nabla h_l(x^*) - \sum_{j \in J_+} \bar\mu_j \nabla g_j(x^*) = 0,$$

we could define $\bar\lambda$ by setting $\bar\lambda_i = \bar\lambda_i^+$ for all $i \in I_+$, $\bar\lambda_l = -\bar\lambda_l^-$ for all $l \in I_-$, and $\bar\lambda_j = 0$ otherwise, which, together with $\bar\mu$, violate MFCQ at $x^*$ with respect to the original constraints by Theorem 3.2. Thus, there is a unitary $d \in \mathbb{R}^n$ such that

$$\nabla h_i(x^*)^t d > 0, \quad \nabla h_l(x^*)^t d < 0, \quad \nabla g_j(x^*)^t d > 0, \quad i \in I_+,\ l \in I_-,\ j \in J_+.$$

Analogously to the proof of Lemma 3.3, defining $x^k = x^* + d/k$ for all $k \ge 1$ we can conclude that $x^k$ verifies (3.4) with $\widehat\lambda_i = \lambda_i$, $i \in I_{\neq}$, and $\widehat\mu_i = \mu_i$, $j \in J_+$, if we take

$$\gamma = \min\left\{\nabla h_i(x^*)^t d,\ -\nabla h_l(x^*)^t d,\ \nabla g_j(x^*)^t d \mid i \in I_+,\ l \in I_-,\ j \in J_+\right\} > 0.$$

Items 2(a) and (b) of Definition 2.2 hold using $x^k$ for all $k$ sufficiently large. Taking $L > \max\{|\nabla h_i(x^*)^t d|, |\nabla g_j(x^*)^t d| \mid i \notin I_{\neq},\ j \notin J\}$ we obtain (3.5) and (3.6) for all $i \notin I_{\neq}$ and $j \notin J_+$. The last inequalities of the proof of Theorem 3.5 hold with $\mathcal{I} = I_{\neq}$ and $\mathcal{J} = J_+$, and then item 2(c) of Definition 2.2 is valid. This concludes the proof. ∎

Recently [35], it was proved that some CQs like the *(relaxed) constant positive linear dependence* ((R)CPLD) [8], the *(relaxed) constant rank CQ* ((R)CRCQ) [36], and the *constant rank of the subspace component condition* (CRSC) [9], can be reduced in some sense to MFCQ by locally rewriting the feasible set of the problem. This is the case in Example 2: we can rewrite the constraints simply as $x = 0$. However, we do not know a priori how such a reformulation can be done. It is worth mentioning that Example 2 satisfies all the aforementioned CQs, QNCQ (see Definition 4.1), and the pseudo-normality CQ defined in [15]. Thus, Theorem 3.9 cannot be improved using any other known CQ from the literature (see [6, Figure 1]).

## 4. Properties of quasinormal and enhanced multipliers.

### 4.1. Compactness of $M_Q(x^*)$ under QNCQ.

It is well known that MFCQ is necessary and sufficient for the compactness of $M(x^*)$ [24]. Due to the peculiar nature of quasinormal multipliers, a question arises: which condition on $x^*$ is equivalent to the compactness of $M_Q(x^*)$? In this section, we give a partial answer to this issue.

The connection between enhanced stationarity and the quadratic penalty method brings QNCQ to the discussion. In fact, QNCQ was designed to eliminate possible "wrong" multipliers generated by this method. It was introduced by Hestenes in [28] and generalized in [15] to the case where additional abstract constraints are present.

DEFINITION 4.1. *We say that a feasible $x^*$ for* (P) *satisfies the quasi-normality* CQ *(QNCQ) if there are no $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^{|I_g(x^*)|}$, $\mu \geq 0$, such that*

1. $\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in I_g(x^*)} \mu_j \nabla g_j(x^*) = 0$;
2. $(\lambda, \mu) \neq 0$;
3. *there is a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to $x^*$ such that, for each $k$, $\lambda_i h_i(x^k) > 0$ for all $i \in I_{\neq}$ and $\mu_j g_j(x^k) > 0$ for all $j \in J_+$, where $I_{\neq}$ and $J_+$ are defined as in Definition* 2.2.

In [41], it was proved that the set of quasinormal multipliers $M_Q(x^*)$ is bounded when $x^*$ satisfies QNCQ. Next we prove that $M_Q(x^*)$ is always closed, and thus compact under QNCQ.

THEOREM 4.2. *Let $x^*$ be a KKT point. Then $M_Q(x^*)$ is nonempty and closed. Also, if QNCQ holds at $x^*$, then $M_Q(x^*)$ is compact.*

*Proof.* By Theorem 3.5, $M_Q(x^*) \neq \emptyset$. Let us show that $M_Q(x^*)$ is closed. Take a convergent sequence $M_Q(x^*) \supset \{(\lambda^k, \mu^k)\} \to (\lambda, \mu)$ and define $I^k_{\neq} = \{i \mid \lambda^k_i \neq 0\}$, $J^k_+ = \{j \mid \mu^k_j > 0\}$ for each $k$. If $(\lambda, \mu) = 0$, then it is in $M_Q(x^*)$ trivially. Suppose that $(\lambda, \mu) \neq 0$, so the sets $I_{\neq} = \{i \mid \lambda_i \neq 0\}$ and $J_+ = \{j \mid \mu_j > 0\}$ are not both empty. As there are only finitely many distinct $I^k_{\neq}$, $J^k_+$, there exist sets $I^*_{\neq}$, $J^*_+$ and a subsequence $\{(\lambda^k, \mu^k)\}_{k \in K}$ such that $I^k_{\neq} = I^*_{\neq}$ and $J^k_+ = J^*_+$ for all $k \in K$. For each $k \in K$, as $(\lambda^k, \mu^k) \in M_Q(x^*)$ there is a sequence $\{x^{k,p}\}_{p \in \mathbb{N}} \to x^*$ satisfying item 2(a) of Definition 2.2 with respect to $I^*_{\neq}$ and $J^*_+$. Note that $I_{\neq} \subset I^*_{\neq}$ and $J_+ \subset J^*_+$, and then we can obtain a sequence $\{x^k\}_{k \in K}$ converging to $x^*$ and satisfying item 2(a) in the following way: for the first index $\ell_1 \in K$, take $p_1 \in \mathbb{N}$ such that $\|x^{\ell_1, p_1} - x^*\| \leq 1$; for the second index $\ell_2 \in K$, take $p_2 \in \mathbb{N}$ such that $\|x^{\ell_2, p_2} - x^*\| \leq 1/2$; in general, for the $k$th index $\ell_k \in K$, take $p_k \in \mathbb{N}$ such that $\|x^{\ell_k, p_k} - x^*\| \leq 1/k$. So, just put

$x^k = x^{\ell_k, p_k}$ for all $k \in K$ so that item 2(a) holds with respect to $I_{\neq}$ and $J_+$. Thus $(\lambda, \mu) \in M_Q(x^*)$, from which we conclude the closedness of $M_Q(x^*)$.

The boundedness of $M_Q(x^*)$ under QNCQ follows from [41, Theorem 3]. ∎

One might expect the converse of the last statement in the previous theorem to be valid. However, this is not the case, as the next example shows.

*Example* 3. Let us consider the problem

$$\min x_1 \quad \text{subject to} \quad x_1 \leq 0, \quad -x_1 \leq 0, \quad x_1^3 + x_2 \leq 0, \quad x_1^3 - x_2 \leq 0$$

and the feasible point $x^* = (0, 0)$. At this point, item 1 of Definition 2.2 takes the form

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \mu_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \mu_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \mu_4 \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

from which follow $1 + \mu_1 - \mu_2 = 0$ and $\mu_3 = \mu_4$. We affirm that $M_Q(x^*) = \{(0, 1, 0, 0)\}$, and thus is compact. In fact, we must have $\mu_2 > 0$, and $\mu_1$ cannot be positive; otherwise the sequence $\{x^k\}$ in item 2(a) of Definition 2.2 should satisfy $-x_1^k > 0$ and $x_1^k > 0$ for all $k$. If $\mu_3 = \mu_4 > 0$, $x^k$ should satisfy $(x_1^k)^3 + x_2^k > 0$ and $(x_1^k)^3 - x_2^k > 0$, which is impossible since $\mu_1 > 0$ implies $x_1^k < 0$ for all $k$.

On the other side, note that QNCQ does not hold at $x^*$ since $\mu = (0, 0, 1, 1)$ and $x^k = (1/k, 0)$, $k \geq 1$, fulfill the conditions of Definition 4.1.

It is clear that QNCQ implies the boundedness of the enhanced multipliers set since $M_E(x^*) \subset M_Q(x^*)$, but the converse is not true in general due to Example 3. In general, $M_E(x^*)$ is not even closed, at least in highly degenerate problems, as the next example illustrates.

*Example* 4. For the problem of minimizing $x^2$ subject to $x^4 = 0$, $x^2 = 0$, it is easy to see that $(1, \delta)$ is an enhanced multiplier vector associated with $x^* = 0$ for all $\delta > 0$, but $(1, 0) = \lim_{\delta \to 0}(1, \delta)$ is not.

The set $M(x^*)$ may be unbounded even if $M_Q(x^*)$ is a singleton: in Example 3 we have $M(x^*) = \{(s, 1 + s, t, t) \mid s, t \geq 0\}$ while $M_Q(x^*) = \{(0, 1, 0, 0)\}$. We can also have $M(x^*)$ unbounded while $M_Q(x^*)$ is bounded (but not a singleton).

*Example* 5. For the problem of minimizing $x$ subject to $x = 0$, $x = 0$, we have $M(0) = \{(t, -1 - t) \mid t \in \mathbb{R}\}$ and $M_Q(0) = \{(t, -1 - t) \mid t \in [-1, 0]\}$.

It is well known that $M(x^*) = \{(\lambda, \mu)\}$ is equivalent to the validity of the so-called *strict MFCQ* at $x^*$ for $\mu$ [34]. In this sense, we can define the strict MFCQ condition exactly as "MFCQ and the uniqueness of the Lagrange multiplier vector." Actually, MFCQ is redundant in this statement, since the uniqueness of the Lagrange multiplier implies MFCQ (see Theorem 5.1). Analogously, we can define the *strict QNCQ* condition exactly as "the uniqueness of the quasinormal multiplier vector." Unlike MFCQ, the uniqueness of the quasinormal multiplier does not imply QNCQ as illustrated by Example 3, which makes strict QNCQ and QNCQ independent of each other. It is worth noting that strict QNCQ and strict MFCQ are not CQs in the usual sense, since they involve the objective function. Finally, note that the uniqueness of the quasinormal multiplier occurs more often than the uniqueness of the usual multiplier (see Example 5), and therefore strict QNCQ is less stringent than strict MFCQ.

**4.2. On the convexity of $M_Q(x^*)$ and $M_E(x^*)$.** It is easy to see that the set of usual multipliers $M(x^*)$ is polyhedral, regardless of the validity of any CQ. In turn, the convexity of $M_Q(x^*)$ and $M_E(x^*)$ is not a trivial matter since the nonnull multipliers may change along the segment between two elements $(\lambda^1, \mu^1), (\lambda^2, \mu^2) \in M_Q(x^*)$, and thus a sequence $\{x^k\}$ satisfying item 2(a) of Definition 2.2 for $(\lambda^1, \mu^1)$ may not work for $(\lambda^2, \mu^2)$. Under MFCQ, $M_Q(x^*)$ and $M_E(x^*)$ are indeed polyhedral according to Theorem 3.9. The next example shows that, unfortunately, these sets are not convex in general.

*Example* 6. Consider the bidimensional problem

$$\min x_2 \quad \text{subject to} \quad x_2 + x_1^3 \leq 0, \quad -x_2 - x_1^3 \leq 0, \quad -x_2 \leq 0$$

and the feasible point $x^* = (0,0)$. It is easy to see that $\widehat{\mu} = (0,1,0)$ and $\bar{\mu} = (1,0,2)$ are enhanced (hence, quasinormal) multipliers associated with $x^*$ using the sequences $\widehat{x}^k = (0, -1/k)$ and $\bar{x}^k = (1/k, -1/k^4)$, $k \geq 2$, respectively. Now, consider the convex combination

$$\mu = \frac{1}{2}\widehat{\mu} + \frac{1}{2}\bar{\mu} = \left(\frac{1}{2}, \frac{1}{2}, 1\right).$$

This $\mu$ is clearly a usual multiplier, but not quasinormal. In fact, if there were a sequence $\{x^k\}$ satisfying item 2(a) of Definition 2.2 for $\mu$, we should have $x_2^k + (x_1^k)^3 < 0 < x_2^k + (x_1^k)^3$ since $\mu_1 = \mu_2 > 0$. Thus, $\mu \notin M_Q(x^*)$ and consequently $\mu \notin M_E(x^*)$. By similar reasoning, no convex combination of $\widehat{\mu}$ and $\bar{\mu}$ distinct of these vectors is in $M_Q(x^*)$ or $M_E(x^*)$.

Note that in the above example, $x^*$ does not satisfy any known CQ, including the Guignard CQ. However, if we add the constraint $x_1 \leq 0$ with null multiplier, $M_Q(x^*)$ remains nonconvex and the Abadie CQ [1] becomes valid. Nevertheless, we do not know whether $M_Q(x^*)$ is convex under stronger CQs. Although we are not able to give a proof, we conjecture that $M_Q(x^*)$ is polyhedral, at least under QNCQ.

**4.3. Informative multipliers.** We back our attention to the condition 3(c) of Theorem 2.1. We can define an enhanced KKT stationarity concept by changing item 2(c) of Definition 2.2 to that condition, that is, by using "$o$" instead of "$O$". Specifically, item 2(c) of Definition 2.2 is replaced by

$$|h_i(x^k)| = o(w(x^k)) \;\; \forall i \notin I_{\neq} \;\; \text{and} \;\; g_j(x^k)_+ = o(w(x^k)) \;\; \forall j \notin J_+, \tag{4.1}$$

where $w(x)$ is the same. As we already mentioned, the corresponding multipliers are known as informative [15]. Let us denote the set of them by $M_I(x^*)$, that is,

$$M_I(x^*) = \left\{ (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \;\middle|\; \begin{array}{l} \nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0, \quad \exists \{x^k\} \to x^* \\ \text{s.t. items 2(a), (b) of Definition 2.2 and (4.1) hold} \end{array} \right\}.$$

$M_I(x^*)$ can be a proper subset of $M_E(x^*)$. In fact, we always have $M_I(x^*) \subset M_E(x^*)$ since $a_k = o(b_k)$ implies $a_k = O(b_k)$, and in the problem presented in Example 5 we have $(-1,0) \in M_E(0) \backslash M_I(0)$; otherwise $(-1,0) \in M_I(0)$ would arrive at $|x_k| = o(|x_k|)$, which is impossible. Furthermore, it is easy to see that $M_I(0) = \{(t, -1-t) \mid t \in (-1,0)\}$, which shows that $M_I(x^*)$ is not closed in general although it is obviously bounded under QNCQ just as $M_Q(x^*)$.

An issue related to informative multipliers is whether Theorem 3.5 remains valid with (4.1), that is, if for *every* KKT point $x^*$ we have $M_I(x^*) \neq \emptyset$. The answer is yes. Although Proposition 2.2(a) of [15] assumes that $x^*$ is a local minimizer, such a statement does not depend on the minimality of $x^*$. The core of its proof relies on [15, Lemma 2.1], which is presented below in a specialized version for our purposes.

LEMMA 4.3. *Let $c, a_1, \ldots, a_m, b_1, \ldots, b_r \in \mathbb{R}^n$. Suppose that*

$$M = \left\{ (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^r_+ \;\middle|\; c + \sum_{j=1}^{m} \lambda_i a_i + \sum_{j=1}^{r} \mu_j b_j = 0 \right\}$$

*is nonempty. Then there exists a sequence $\{d^k\} \subset \mathbb{R}^n$ such that*

1. $c^t d^k \to -\|(\lambda^*, \mu^*)\|_2^2$;
2. $a_i^t d^k \to \lambda_i^*$ *for all* $i = 1, \ldots, m$;
3. $(b_j^t d^k)_+ \to \mu_j^*$ *for all* $j = 1, \ldots, r$,

*where $(\lambda^*, \mu^*)$ is the element of $M$ with minimum 2-norm.*

*Proof.* The statement follows from applying [15, Lemma 2.1] with $N = \{0\}$, after rewriting each $\lambda_i a_i$ as $(\lambda_i^+ - \lambda_i^-) a_i$, $\lambda_i^+, \lambda_i^- \geq 0$, and noting that $(\lambda_i^+)^* = 0$ or $(\lambda_i^-)^* = 0$ in the minimum 2-norm element. ∎

THEOREM 4.4. *Let $x^*$ be a KKT point. Then the associated multiplier with minimum 2-norm is informative. In particular, $M_I(x^*) \neq \emptyset$ and Theorem 3.5 is valid if we define the E-KKT points using (4.1) instead of item 2(c) of Definition 2.2.*

*Proof.* The proof is similar to that of [15, Proposition 2.2(a)]. We can suppose without loss of generality that $I_g(x^*) = \{1, \ldots, r\}$. From now on, we set $c = \nabla f(x^*)$, $a_i = \nabla h_i(x^*)$, $i = 1, \ldots, m$, and $b_j = \nabla g_j(x^*)$, $j = 1, \ldots, r$, in Lemma 4.3. If $c = 0$, then $(\lambda^*, \mu^*) = 0$ is the multiplier vector with minimum 2-norm and there is nothing to prove. Suppose that $c \neq 0$. Thus $(\lambda^*, \mu^*) \neq 0$ and $d^k \neq 0$ for all $k$ taking a subsequence if necessary, let us say, with indices $k \in K$. From Lemma 4.3, the limit $d$ of the unitary convergent subsequence $\{d^k / \|d^k\|_2\}_{k \in K}$ satisfies

$$(4.2)\qquad \nabla f(x^*)^t d < 0, \quad \operatorname{sgn}(\lambda_i^*) \cdot \nabla h_i(x^*)^t d > 0 \;\; \forall i \in I_{\neq}, \quad \nabla g_j(x^*)^t d > 0 \;\; \forall j \in J_+,$$

where $I_{\neq} = \{i \mid \lambda_i^* \neq 0\}$ and $J_+ = \{j \mid \mu_j^* > 0\}$. Also, there is a sequence $\{\delta_k\}$ converging to zero such that

$$(4.3)\qquad |\nabla h_l(x^*)^t d^k| \leq \delta_k \cdot m(d^k),\; l \notin I_{\neq}, \quad (\nabla g_s(x^*)^t d^k)_+ \leq \delta_k \cdot m(d^k),\; s \notin J_+,$$

for all $k$ large enough, where

$$m(z) = \min\{|\nabla h_i(x^*)^t z|, (\nabla g_j(x^*)^t z)_+ \mid i \in I_{\neq}, j \in J_+\},$$

since the left sides of the inequalities in (4.3) go to zero and $m(d^k)$ remains bounded away from zero.

Define $x^k = x^* + d/k$ for all $k$. We have $\|x^k - x^*\|_2 = 1/k$ and, from the Taylor expansion of $f$ around $x^*$,

$$f(x^k) - f(x^*) = \|x^k - x^*\|_2 \left[ \nabla f(x^*)^t d + \frac{o(\|x^k - x^*\|_2)}{\|x^k - x^*\|_2} \right],$$

which, considering (4.2), gives $f(x^k) < f(x^*)$ for all $k$ large enough. Therefore, item 2(b) of Definition 2.2 holds. Analogously, for all $i \in I_{\neq}$ we have

$$(4.4)\qquad h_i(x^k) = h_i(x^k) - h_i(x^*) = \|x^k - x^*\|_2 \left[ \nabla h_i(x^*)^t d + \frac{o(\|x^k - x^*\|_2)}{\|x^k - x^*\|_2} \right],$$

which, together with (4.2), implies $\lambda_i^* h_i(x^k) > 0$ for all $k$ large enough. Similarly, we can conclude that $g_j(x^k) > 0$ for all $j \in J_+$ and $k$ sufficiently large. So item 2(a) of Definition 2.2 is also verified.

Now, let us prove that $\{x^k\}$ verifies (4.1). Dividing (4.3) by $\|d^k\|_2$ and taking the limit over $K$, we arrive at $\nabla h_l(x^*)^t d = 0$, which implies $|h_l(x^k)| = o(\|x^k - x^*\|_2)$ for all $l \notin I_{\neq}$ in view of (4.4). Thus, $r_k^l = |h_l(x^k)|/\|x^k - x^*\|_2 \to 0$. Notice that by (4.2) we have $m(d) = \lim_{k\in K} m(d^k)/\|d^k\|_2 > 0$, so by (4.4),

$$\frac{w(x^k)}{\|x^k - x^*\|_2} = \frac{1}{\|x^k - x^*\|_2} \min\left\{\min_{i\in I_{\neq}} |h_i(x^k)|,\ \min_{j\in J_+} g_j(x^k)_+\right\} \geq m(d)/2 > 0$$

for all $k$ large enough. Therefore,

$$\frac{|h_l(x^k)|}{\|x^k - x^*\|_2} = r_k^l = \frac{2r_k^l}{m(d)}\frac{m(d)}{2} \leq \frac{2r_k^l}{m(d)}\frac{w(x^k)}{\|x^k - x^*\|_2} \ \Rightarrow\ |h_l(x^k)| \leq \frac{2r_k^l}{m(d)} w(x^k)$$

for all $k$ large enough. As $2r_k^l/m(d) \to 0$, we prove that (4.1) is valid for all $h_l$, $l \notin I_{\neq}$. Similarly we can conclude that (4.1) is valid for all $g_s$, $s \notin J_+$.

Finally, note that the indices $i \notin I_g(x^*)$ do not interfere in the analysis since in this case $\mu_i^* = 0$ and $g_i(x^k)_+ = 0$ for all $k$ large enough. This concludes the proof. ∎

**4.4. Abstract constraints.** In this section we deal with the problem (P) with additional abstract constraints

$$(\mathrm{P_X}) \qquad \min_x f(x) \quad \text{subject to} \quad h(x) = 0, \quad g(x) \leq 0, \quad x \in X,$$

where $X$ is a nonempty, closed subset of $\mathbb{R}^n$. FJ/KKT enhanced stationarity concepts were developed to the present case. In order to introduce them, we denote the *normal cone* of $X$ at $x$ by

$$N_X(x) = \{z \in \mathbb{R}^n \mid \exists\{x^k\} \to x,\ \exists\{z^k\} \to z \ \text{ such that } \ x^k \in X,\ z^k \in T_X(x^k)^\circ\ \ \forall k\},$$

where $T_X(x)$ is the tangent cone of $X$ at $x$ and $C^\circ$ denotes the polar set of $C$ [14]. Specifically, Bertsekas, Nedic, and Ozdaglar [14] provide a version of Theorem 2.1 for problem ($\mathrm{P_X}$), which consists of replacing item 1 with

$$-\left[\sigma\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu\right] \in N_X(x^*)$$

and considering the sequence $\{x^k\}$ of item 3 in the set $X$. QNCQ is adapted accordingly (see [15]). For the sake of clarification, we present next the adaptations of Definitions 2.2 and 4.1 to this case.

DEFINITION 4.5. *We say that a feasible $x^*$ for* ($\mathrm{P_X}$) *is an E-KKT point if there are $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ such that*

1. $-\left[\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu\right] \in N_X(x^*)$;
2. *if $I_{\neq} \cup J_+ \neq \emptyset$, then there is a sequence $\{x^k\} \subset X$ converging to $x^*$ such that, for all $k$, the following conditions are valid:*
   (a) $\lambda_i h_i(x^k) > 0\ \ \forall i \in I_{\neq}$, *and* $\mu_j g_j(x^k) > 0\ \ \forall j \in J_+$;
   (b) $f(x^k) < f(x^*)$;
   (c) $|h_i(x^k)| = O(w(x^k))\ \ \forall i \notin I_{\neq}$, *and* $g_j(x^k)_+ = O(w(x^k))\ \ \forall j \notin J_+$,

   *where $I_{\neq}$, $J_+$, and $w(x)$ are defined as in Definition* 2.2.

DEFINITION 4.6. *We say that a feasible $x^*$ for* ($\mathrm{P_X}$) *satisfies QNCQ if there are no $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^{|I_g(x^*)|}$, $\mu \geq 0$, such that*

1. $-\left[\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j\in I_g(x^*)} \mu_j \nabla g_j(x^*)\right] \in N_X(x^*)$;
2. $(\lambda, \mu) \neq 0$;

3. *there is* $\{x^k\} \subset X$ *converging to* $x^*$ *such that, for each* $k$, $\lambda_i h_i(x^k) > 0$ *for all* $i \in I_{\neq}$ *and* $\mu_j g_j(x^k) > 0$ *for all* $j \in J_+$, *where* $I_{\neq}$ *and* $J_+$ *are defined as in Definition* 2.2.

Similar to what is done in the literature, we say that $x^*$ is a KKT point when item 1 of Definition 4.5 and the complementarity slackness $\mu_j g_j(x^*) = 0 \ \forall j \in I_g(x^*)$ hold. The sets $M(x^*)$, $M_Q(x^*)$, and $M_E(x^*)$ of usual, quasinormal, and enhanced multipliers are defined accordingly.

The approach by normal cones is common in nonlinear optimization, where a typical (perhaps necessary) regularity assumption on $X$ is imposed. We say that $X$ is *regular* at $x \in X$ if

$$N_X(x) = T_X(x)^\circ.$$

It is worth mentioning that if $X$ is convex, then it is regular at all $x \in X$ [38]. In particular, $X = \mathbb{R}^n$ and $X = \{x \mid \ell \le x \le u\}$ are regular at all their feasible points. The last constraints are commonly used to ensure the well-definiteness of several algorithms and are present in popular computational optimization packages like Algencan [3] (in section 5 we present this method without abstract constraints).

Next, we analyze the validity of the main results in the presence of abstract constraints.

**4.4.1. Nonemptiness and compactness of $M_Q(x^*)$.** As mentioned in subsection 4.3, the same arguments for proving [15, Proposition 2.2] are still valid for any KKT point $x^*$. So, in the case where $T_X(x^*)$ is convex, Theorem 4.4 (and consequently Theorem 3.5) is valid when dealing with the abstract constraints through its normal cone as in Definition 4.5. Note that $T_X(x^*)$ is convex if $X$ is regular at $x^*$, which in turn is true if $X$ is convex [15, 38]. In particular, the first statement of Theorem 4.2 ($M_Q(x^*) \neq \emptyset$) is valid in this case. The closedness of $M_Q(x^*)$ can be proved analogously to what is done in Theorem 4.2 because $N_X(x^*)$ is closed.

In the proof of Theorem 4.2, we claim that the boundedness of $M_Q(x^*)$ under QNCQ follows from [41, Theorem 3]. The cited theorem from [41] remains valid in the presence of $x \in X$ even if $X$ is not regular at $x^*$. Thus, $M_Q(x^*)$ (considering abstract constraints) is bounded under QNCQ in the sense of Definition 4.6. So, by the closedness of $N_X(x^*)$, the proof of Theorem 4.2 remains valid.

Theorem 4.7. *Let* $x^*$ *be a KKT point for* ($\mathrm{P_X}$). *Then* $M_Q(x^*)$ *is always nonempty and closed. Also, if QNCQ (in the sense of Definition* 4.6*) holds at* $x^*$, *then* $M_Q(x^*)$ *is compact.*

**4.4.2. Relationship between $M(x^*)$, $M_Q(x^*)$, and $M_E(x^*)$ under MFCQ.** In what follows, we show that Theorem 3.9 is no longer valid in the presence of abstract constraints even if $X$ is regular at $x^*$. In [15], a general extension of MFCQ (Definition 3.1) was defined, and specialized cases are discussed. The most natural of them is the following.

Definition 4.8. *We say that a feasible* $x^*$ *for* ($\mathrm{P_X}$) *satisfies MFCQ if there is no nonnull* $\lambda \in \mathbb{R}^m$ *such that*

$$-\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*),$$

*and there exists* $d \in N_X(x^*)^\circ$ *satisfying*

$$\nabla h_i(x^*)^t d = 0, \ i = 1, \ldots, m, \qquad \nabla g_j(x^*)^t d < 0 \ \forall j \in I_g(x^*).$$

Note that the above definition reduces to Definition 3.1 if $X = \mathbb{R}^n$. The next example shows that the sets $M(x^*)$ and $M_Q(x^*)$ may be distinct in a very simple case where $X$ is regular at $x^*$ satisfying MFCQ in the sense of Definition 4.8.

*Example* 7. Consider the problem

$$\min f(x) = -x_2 \quad \text{subject to} \quad h(x) = x_1 - x_2 = 0, \quad g(x) = x_1 \leq 0, \quad x \in X,$$

where $X = \{x \in \mathbb{R}^2 \mid x_1 \leq 0\}$, and the feasible point $x^* = (0,0)$. It is easy to see that $N_X(x^*) = \mathbb{R}_+ \times \{0\}$ and $N_X(x^*)^\circ = X$. We have $-\nabla h(x^*) = (-1,1) \notin N_X(x^*)$, and taking $d = (-1,-1) \in N_X(x^*)^\circ$ we arrive at $\nabla h(x^*)^t d = 0$ and $\nabla g(x^*)^t d = -1 < 0$. Thus, MFCQ holds at $x^*$. Note that

$$-\left[\nabla f(x^*) - 1 \cdot \nabla h(x^*) + \mu \cdot \nabla g(x^*)\right] = (1-\mu, 0) \in N_X(x^*) \quad \forall \mu \in [0,1];$$

therefore $x^*$ is a KKT point with multipliers $\lambda = -1$, $\mu \in [0,1]$. However, it is clear that there is no $\{x^k\} \subset X$ such that $g(x^k) > 0$, and therefore $M_E(x^*) = M_Q(x^*) = \{(-1,0)\} \neq \{-1\} \times [0,1] = M(x^*)$.

**5. On augmented Lagrangian methods.** In this section, we consider the safeguarded (Powell–Hestenes–Rockafellar (PHR)) augmented Lagrangian method defined in [3], namely ALGENCAN, which we recall in Algorithm 5.1. This method consists in successively minimizing the PHR augmented Lagrangian function associated with (P),

$$L_\rho(x, \bar{\lambda}, \bar{\mu}) = f(x) + \frac{\rho}{2}\left[\left\| h(x) + \frac{\bar{\lambda}}{\rho}\right\|_2^2 + \left\| \left(g(x) + \frac{\bar{\mu}}{\rho}\right)_+ \right\|_2^2\right],$$

with respect to $x$ for a fixed *penalty parameter* $\rho > 0$ and fixed *projected multipliers* estimates $\bar{\lambda} \in \mathbb{R}^m$, $\bar{\mu} \in \mathbb{R}^p_+$, which are computed within a predefined compact set (*safeguards*). Note that Algorithm 5.1 generates the multiplier estimates

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k), \qquad \mu^k = (\bar{\mu}^k + \rho_k g(x^k))_+ \tag{5.1}$$

at each iteration $k$ when solving the subproblem of Step 1.

An important consequence of the enhanced FJ conditions is that they motivate the definition of QNCQ, which was one of the first generalizations of MFCQ. Since then, several new CQs have emerged, most of them linked to the global convergence of practical optimization methods, especially ALGENCAN. See [6, Figure 1] for a comprehensive overview of such CQs. Since their introduction in the 1970s, enhanced FJ conditions and QNCQ have been adapted to different problems [26, 27, 31, 33, 42]

---

**Algorithm 5.1.** ALGENCAN.

Let $\mu_{\max} > 0$, $\lambda_{\min} < \lambda_{\max}$, $\gamma > 1$, $0 < \tau < 1$, and $\{\varepsilon_k\} \to 0$ such that $\varepsilon_k > 0$ $\forall k$.

Let $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^1 \in [0, \mu_{\max}]^p$, and $\rho_1 > 0$. Initialize $k \leftarrow 1$.

*Step* 1. Find an approximate first-order stationary point $x^k$ of the unconstrained problem $\min_x L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k)$, that is, $\left\|\nabla_x L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\right\| \leq \varepsilon_k$.

*Step* 2. Define $V^k = \min\{-g(x^k), \bar{\mu}^k/\rho_k\}$. If $k > 1$ and

$$\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\},$$

define $\rho_{k+1} = \rho_k$. Otherwise, define $\rho_{k+1} = \gamma\rho_k$.

*Step* 3. Compute $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^p$ and $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$, take $k \leftarrow k+1$, and go to Step 1.

---

(we discuss some of them in section 6). However, a connection between these concepts and the convergence of a practical method beyond pure external penalty, in this case Algencan, was established only recently in our work [4], through the sequential optimality condition *positive approximate KKT* (PAKKT). In addition to strengthening the theoretical global convergence of Algencan, a surprisingly practical consequence of PAKKT is that the multiplier sequences (5.1) are bounded if QNCQ holds at the feasible limit point [4, Corollary 4.8]. This fact was crucial in attesting convergence of Algencan using a scaled stopping criteria, which improves the numerical performance of the method [7] in cases where the standard one fails. The purpose of this section is to indicate possible contributions of the theory on enhanced stationarity and quasinormal multipliers to improve the performance of Algencan in critical situations where it suffers from numerical instabilities.

From our numerical experience while writing [7], Algencan generally fails when the penalty parameter $\rho_k$ explodes. This is consistent with the fact that in this case Algencan's subproblems become increasingly ill-conditioned. Thus, establishing sufficient conditions to bound $\{\rho_k\}$ is of interest. In [3], the boundedness of $\{\rho_k\}$ was proved under strong hypotheses such as second-order sufficiency and LICQ, which implies the uniqueness of the usual multiplier. Later on [17], LICQ was slightly relaxed to the following assumption:

A0. MFCQ holds at $x^*$, and there is only one multiplier vector associated with $x^*$.

From [34, Proposition 1.1], we conclude that MFCQ is redundant in the above statement. For the sake of completeness, we provide below a simple proof of this fact by means of the well-known relation between the boundedness of $M(x^*)$ and MFCQ [24].

Theorem 5.1. *Let $x^*$ be a KKT point for* (P)*, and suppose that $M(x^*)$ is bounded. Then MFCQ holds at $x^*$. In particular, if there is only one associated multiplier vector $(\lambda^*, \mu^*)$, then MFCQ holds at $x^*$.*

*Proof.* If MFCQ is not valid, there is $(\bar{\lambda}, \bar{\mu}) \neq 0$ such that $\nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$, $\mu \geq 0$, by Theorem 3.2. Then $(t\bar{\lambda} + \lambda^*, t\bar{\mu} + \mu^*) \neq (\lambda^*, \mu^*)$ is a multiplier vector associated with $x^*$ for all $t > 0$. ∎

In view of the above theorem, the sufficient hypotheses in [17] for the boundedness of $\{\rho_k\}$ in Algorithm 5.1 can be stated as follows:

A1. $\{x^k\}$ converges to a feasible point $x^*$.

A2. There is only one (usual) multiplier vector $(\lambda^*, \mu^*)$ associated with $x^*$.

A3. $f, g, h$ are twice continuously differentiable at $x^*$, and the second-order sufficient condition (SOSC) holds at $x^*$.

A4. The tolerance $\varepsilon_k$ is driven to zero fast enough in the sense of

$$\varepsilon_k = o\big( \|( \nabla_x L(x^{k-1}, \lambda^{k-1}, \mu^{k-1}), h(x^{k-1}), \min\{-g(x^{k-1}), \mu^{k-1}\} )\| \big),$$

where $L(x, \lambda, \mu)$ is the Lagrangian function.

A5. There is $k_0 \in \mathbb{N}$ such that $\bar{\lambda}^{k+1} = \lambda^k$ and $\bar{\mu}^{k+1} = \mu^k$ in Step 3 for all $k \geq k_0$.

Assumption A4 can be fulfilled in practice since $\varepsilon_k$ is computed after $(x^{k-1}, \lambda^{k-1}, \mu^{k-1})$. For large safeguards $\mu_{\max} > 0$, $\lambda_{\min} < 0$, and $\lambda_{\max} > 0$, we can expect the validity of A5 under QNCQ and A1, as the entire sequence $\{(\lambda^k, \mu^k)\}$ is bounded in this case [4]. It is worth mentioning that, in practice, the projected multipliers $\bar{\lambda}_i^{k+1}$ and $\bar{\mu}_j^{k+1}$ are chosen as the projection of $\lambda_i^k$ and $\mu_j^k$ onto $[\lambda_{\min}, \lambda_{\max}]$ and $[0, \mu_{\max}]$, respectively [7, 18]. So, we have $\bar{\lambda}^{k+1} = \lambda^k$ and $\bar{\mu}^{k+1} = \mu^k$ whenever these estimates are within the safeguard intervals. This somehow recovers the classical augmented

Lagrangian method without safeguards, which exhibits favorable properties in convex problems [37].

To the best of our knowledge, A1–A5 are the least stringent assumptions so far that guarantee that $\{\rho_k\}$ is bounded. However, extensive numerical experience [7] suggests that $\{\rho_k\}$ remains bounded in a wider range of situations. We believe that the specificity of enhanced/quasinormal multipliers can help to elucidate this behavior and, mainly, inspire modifications in the algorithm that mitigate failures. The following facts support our belief:

F1 As we already mentioned, the multiplier sequences (5.1) generated by ALGENCAN are bounded if QNCQ holds at the feasible limit point [4]. Although this does not imply the boundedness of the penalty parameter sequence $\{\rho_k\}$, it serves as an indication of the necessity of QNCQ for this purpose.

F2 As discussed in subsection 4.1, the uniqueness of the quasinormal multiplier occurs more often than the uniqueness of the usual multiplier.

F3 If ALGENCAN generates a quasinormal multiplier vector, the boundedness of $\{\rho_k\}$ could be established by changing A2 to the following assumption, using the same arguments as in [17]:

A6. QNCQ holds at $x^*$, and there is only one quasinormal multiplier vector $(\lambda^*, \mu^*)$ associated with $x^*$.

*Remark* 5.2. Theorem 5.1 says that if for *some objective function* $f$ the set $M(x^*)$ is bounded, then MFCQ holds at $x^*$. Thus, the condition "there is a function $f$ for which $M(x^*)$ is bounded" is a CQ, which in fact does not depend on the objective function of the problem considered. In particular, hypothesis A2 is a CQ. On the other hand, the boundedness/uniqueness of the quasinormal, enhanced, or informative multiplier for some objective does not constitute a CQ, given that for the constraints $x_1 \leq 0$ and $-x_1 \leq 0$, $M_Q(0,0) = \{(0,1)\}$ if $f(x_1,x_2) = x_1$, but $(0,0)$ is not a KKT point if $f(x_1,x_2) = x_2$.

A question that immediately arises is whether the dual accumulation points generated by Algorithm 5.1 are quasinormal when $\rho_k \to \infty$ and QNCQ holds. The answer does not seem trivial due to the projected multipliers. We ran the ALGENCAN implementation available from the TANGO project (https://www.ime.usp.br/∼egbirgin/tango/codes.php) on the problem

$$\min\ x_1^2 - 2x_2 \quad \text{subject to} \quad x_2 e^{x_1} \leq 0, \quad -x_2 \leq 0. \tag{5.2}$$

This problem has a unique minimizer $x^* = (0,0)$, which satisfies QNCQ but not MFCQ (see [11]). We have $M(x^*) = \{(t, t-2) \mid t \geq 2\}$ and $M_Q(x^*) = \{(2,0)\}$. We start the algorithm with $\bar{\mu}^1 = (0,0)$ and $\rho_1 = 10$. Condition "$\rho_k \to \infty$" was forced by modifying the code properly to multiply $\rho_k$ by 10 every iteration (so the test in Step 2 is neglected) and by requiring a feasibility tolerance equal to $10^{-40}$ (in our test, the last $\rho$ was $10^{28}$). The problem was solved with final multiplier $\mu^* \approx (2.02109, 0.02109)$. This suggests that the method may converge to a non-quasinormal multiplier vector even when $\rho_k \to \infty$.

To characterize a multiplier vector as quasinormal, we should exhibit a sequence $\{x^k\}$ in the spirit of item 2(a) of Definition 2.2. Such a sequence can be the one generated by Algorithm 5.1 itself, although this is not mandatory. Items F1–F3 listed above motivate the study of modifications in Algorithm 5.1 that try to adjust the projected multipliers in order to avoid an explosion of $\rho_k$ whenever it might seem to happen, that is, when the algorithm appears to be heading toward failure. We enumerate some possibilities:

- Since every multiplier vector with a minimum 2-norm is quasinormal (Theorem 4.4), compute the project multipliers as those of minimum 2-norm whenever (i) the feasibility is almost reached, (ii) $\rho$ is large, and (iii) the sequences $\{(\lambda^k, \mu^k)\}$, $\{x^k\}$ do not satisfy item 2(a) of Definition 2.2, with $I_{\neq}$ and $J_+$ computed approximately. In this case, $\rho$ is reset to a smaller value.
- Eventually nullify the projected multipliers associated with indices $i$, $j$ such that $\lambda_i^k h_i(x^k) < 0$ or $\mu_j^k g_j(x^k) < 0$. Using such a strategy on (5.2) with the same previous adaptations, the algorithm reached $\mu^* \approx (2, 10^{-15})$, the unique quasinormal multiplier.
- Use additional criteria to increase $\rho$ beyond the feasibility/complementarity test of Step 2 of Algorithm 5.1. For example, if item 2(a) of Definition 2.2 holds, $\rho_k$ is left unchanged during a few consecutive iterations (this may increase the number of outer iterations, but preliminary tests suggest that the inner solver reoptimizes the subproblem faster when $\rho$ is the same, while the optimality and feasibility measures improve due to the update of the multipliers). In this scenario, using a different $\rho_i$ for each constraint can be helpful. Another criterion can be based on heuristic checking of local convexity (for example, using second-order information and directions defined by the last iterates); we know that a similar augmented Lagrangian method converges using a constant $\rho$ on convex problems [37].
- Adjust the parameter $\tau$ of Step 2 dynamically so that the reduction in feasibility is alleviated near the solution $x^*$ whenever the optimality measure improves. As noted in [20, section 3.4.2], good multiplier estimates can prevent $\rho$ from increasing when $x^k$ is close to a solution.

We stress that the above proposals should be seen as strategies to improve the performance of ALGENCAN on problems where it fails to converge, in the spirit of [7].

## 6. Enhanced KKT-type conditions for special optimization problems.

In this section, we recall enhanced stationarity from the literature for some special classes of problems.

### 6.1. Mathematical programs with complementarity constraints and related problems.

*Mathematical programs with complementarity (or equilibrium) constraints* (MPCCs) constitute a class of problems that has been extensively studied in the literature. Due to their high level of degeneracy compared to standard NLP, they deserve special treatment (see [10] and references therein). This class of problems is formulated as

$$\text{(MPCC)} \quad \begin{aligned} \min_x f(x) \quad \text{subject to} \quad & h(x)=0, \quad g(x)\le 0, \quad G(x)\ge 0, \quad H(x)\ge 0, \\ & G_i(x)H_i(x)=0,\ i=1,\dots,q, \end{aligned}$$

where $G, H : \mathbb{R}^n \to \mathbb{R}^q$ are continuously differentiable functions. The last constraints $G_i(x)H_i(x) = 0$, $i = 1, \dots, q$, are called *complementarity constraints.*

It is well known that the KKT conditions cannot be expected to hold at local minimizers $x^*$ of (MPCC) if we do not have (lower level) strict complementarity [39], i.e., if $G_i(x^*) = H_i(x^*) = 0$ for some $i$. Thus, weaker stationarity concepts are defined in the literature. One of the most stringent among them is the Mordukhovich-stationarity (M-stationarity). Given a feasible $z$, we define $I_g(z)$ as before and

$$\begin{aligned} I_{00}(z) &= \{i \mid G_i(z)=0,\ H_i(z)=0\}, \\ I_{0+}(z) &= \{i \mid G_i(z)=0,\ H_i(z)>0\}, \\ I_{+0}(z) &= \{i \mid G_i(z)>0,\ H_i(z)=0\}. \end{aligned}$$

Definition 6.1. *We say that a feasible* $x^*$ *for* (MPCC) *is an* M-stationary *point if there are* $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p_+$, *and* $\gamma^G, \gamma^H \in \mathbb{R}^q$ *such that*

1. $\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla G(x^*)\gamma^G - \nabla H(x^*)\gamma^H = 0$;
2. $\mu_j = 0\ \forall j \notin I_g(x^*)$, $\gamma_i^G = 0\ \forall i \in I_{+0}(x^*)$, $\gamma_i^H = 0\ \forall i \in I_{0+}(x^*)$, *and*

$$\text{either} \quad \gamma_i^G \gamma_i^H = 0 \quad \text{or} \quad \gamma_i^G > 0,\ \gamma_i^H > 0\ \ \forall i \in I_{00}(x^*). \tag{6.1}$$

In [31], enhanced FJ-type conditions linked with M-stationarity were proposed in the spirit of items 1, 2, and 3(a) and (b) of Theorem 2.1. As in standard NLP, enhanced KKT-type conditions can be derived by taking the multiplier associated with the objective equal to one, as done from Theorem 2.1 to Definition 2.2. So, inspired by [31, Theorem 3.1], we define the counterpart of Definition 2.2 to the MPCC context.

Definition 6.2. *We say that a feasible* $x^*$ *for* (MPCC) *is an* enhanced M-stationary *(EM-stationary) point if there are* $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p_+$, *and* $\gamma^G, \gamma^H \in \mathbb{R}^q$ *such that*

1. $\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla G(x^*)\gamma^G - \nabla H(x^*)\gamma^H = 0$;
2. *either* $\gamma_i^G \gamma_i^H = 0$ *or* $\gamma_i^G > 0$, $\gamma_i^H > 0$ *for all* $i \in I_{00}(x^*)$ *(condition* (6.1)*)*;
3. *if* $I_{\neq} \cup J_+ \cup V^G_{\neq} \cup V^H_{\neq} \neq \emptyset$, *where* $I_{\neq} = \{i \mid \lambda_i \neq 0\}$, $J_+ = \{j \mid \mu_j > 0\}$, $V^G_{\neq} = \{i \mid \gamma_i^G \neq 0\}$, *and* $V^H_{\neq} = \{i \mid \gamma_i^H \neq 0\}$, *then there is a sequence* $\{x^k\} \subset \mathbb{R}^n$ *converging to* $x^*$ *such that, for all* $k$, *the following are valid:*
   (a) $\lambda_i h_i(x^k) > 0\ \forall i \in I_{\neq}$, $\mu_j g_j(x^k) > 0\ \forall j \in J_+$, $\gamma_i^G G_i(x^k) < 0\ \forall i \in V^G_{\neq}$, $\gamma_i^H H_i(x^k) < 0\ \forall i \in V^H_{\neq}$;
   (b) $f(x^k) < f(x^*)$;
   (c) $|h_i(x^k)| = O(w(x^k))\ \forall i \notin I_{\neq}$, $g_j(x^k)_+ = O(w(x^k))\ \forall j \notin J_+$, $(-G_i(x^k))_+ = O(w(x^k))\ \forall i \notin V^G_{\neq}$, *and* $(-H_i(x^k))_+ = O(w(x^k))\ \forall i \notin V^H_{\neq}$, *where*

$$w(x) = \min\left\{ \min_{i \in I_{\neq}} |h_i(x)|,\ \min_{i \in J_+} g_j(x)_+,\ \min_{i \in V^G_{\neq}} |G_i(x)|,\ \min_{i \in V^H_{\neq}} |H_i(x)| \right\}.$$

Note that items 2 and 3(a) of the above definition imply item 2 of Definition 6.1. It is worth mentioning that the enhanced stationarity established in [31] was extended to nonsmooth MPCCs in [42], but we do not treat this case.

Given a feasible $x^*$ we denote the set of *M-multipliers* $(\lambda, \mu, \gamma^G, \gamma^H)$ satisfying Definition 6.1 by $M^{\mathrm{M}}(x^*)$, those satisfying items 1, 2, and 3(a) of Definition 6.2 by $M_Q^{\mathrm{M}}(x^*)$ (*quasinormal M-multipliers*), and those satisfying also items 3(b) and (c) by $M_E^{\mathrm{M}}(x^*)$ (*enhanced M-multipliers*). A question that arises is whether a local minimizer of (MPCC) is EM-stationary. In [31], some MPCC-tailored CQs are provided so that every local minimizer satisfies items 1, 2, and 3(a) and (b) of Definition 6.2. The validity of item 3(c) follows from a reasoning similar to that of the final part of the proof of Theorem 3.5.

In this section, we do not focus on MPCC-CQs but on the equivalence between enhanced stationarity from Definition 6.2 and the usual M-stationarity. Given a feasible point $x^*$ for (MPCC), let us consider the *tightened nonlinear problem*

$$(\text{TNLP}(x^*)) \qquad \begin{aligned} \min_x f(x) \quad \text{subject to} \quad & h(x) = 0, \quad g(x) \le 0, \\ & G_i(x) = 0, \quad i \in I_{0+}(x^*) \cup I_{00}(x^*), \\ & G_i(x) \ge 0, \quad i \in I_{+0}(x^*), \\ & H_i(x) = 0, \quad i \in I_{+0}(x^*) \cup I_{00}(x^*), \\ & H_i(x) \ge 0, \quad i \in I_{0+}(x^*), \end{aligned}$$

which consists of fixing the active complementary constraints at $x^*$ as equalities. It is easy to verify that M-stationarity is exactly the KKT conditions for (TNLP($x^*$)) (the so-called *weak stationarity,* or *W-stationarity*) together with (6.1).

THEOREM 6.3. *Every M-stationary point $x^*$ is EM-stationary and vice versa.*

*Proof.* Applying Theorem 3.5 to (TNLP($x^*$)) we obtain the aforementioned equivalence, except for (6.1), possibly with different M-multipliers. But in the proof of Theorem 3.5, the multipliers (in this case, multipliers for (TNLP($x^*$))) are adjusted using Lemma 3.4, which preserves the signs of nonnull ones. So (6.1) is still valid when we redefine the M-multipliers to satisfy Definition 6.2. ∎

Most of the standard CQs are not valid for (MPCC), so this problem deserves MPCC-tailored CQs. A common way to define such MPCC-CQs is to impose a standard CQ on (TNLP($x^*$)), possibly together with (6.1). See [10] and references therein. In [39], a Mangasarian–Fromovitz-type CQ was defined as MFCQ on (TNLP($x^*$)). In [31], it was generalized by adding (6.1). We refer to the last CQ as *MMPCC-MFCQ.* Still in [31], a generalized MPCC-QNCQ was defined in the same way, which we will refer to as *MMPCC-QNCQ.* We enunciate them below.

DEFINITION 6.4. *Let $x^*$ be a feasible point for* (MPCC). *We say that*
1. *$x^*$ satisfies* MMPCC-MFCQ *if it conforms to MFCQ regarding* (TNLP($x^*$)) *and* (6.1) *holds;*
2. *$x^*$ satisfies* MMPCC-QNCQ *if there is no nonnull M-multiplier* $(\lambda, \mu, \gamma^G, \gamma^H)$, $\mu \geq 0$, *such that*
$$\nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla G(x^*)\gamma^G - \nabla H(x^*)\gamma^H = 0$$
*and items* 2, *and* 3(a) *of Definition* 6.2 *hold.*

THEOREM 6.5. *Let $x^*$ be an M-stationary point.*
1. *If MMPCC-MFCQ holds at $x^*$, then $M^M(x^*) = M_Q^M(x^*) = M_E^M(x^*)$.*
2. *$M_E^M(x^*) \subset M_Q^M(x^*)$ are nonempty, and $M_Q^M(x^*)$ is closed.*
3. *$M_Q^M(x^*)$ is compact if MMPCC-QNCQ holds at $x^*$.*

*Proof.* Item 1 is a consequence of Theorem 3.9 applied on (TNLP($x^*$)), since (6.1) does not interfere in the analysis. Item 2 follows from Theorem 6.3 and the reasoning in the first part of the proof of Theorem 4.2, noting that $\{(\gamma^G, \gamma^H) \in \mathbb{R}^{2q} \mid (6.1)\}$ is closed. Let us prove the third item. By Theorem 4.2, the set $M_Q(x^*)$ of multipliers associated with (TNLP($x^*$)) is bounded under MMPCC-QNCQ, since this MPCC-CQ implies QNCQ for (TNLP($x^*$)). As $M_Q^M(x^*)$ is contained in $M_Q(x^*)$, it is also bounded. ∎

There are other stationarity concepts for (MPCC) besides M-stationarity in the literature. The widely used ones are *strong, weak*, and *Clarke* stationarity (S/W/C-stationarity, respectively). As in Definition 6.2, we can define their enhanced versions by adding item 3, although to the best of our knowledge they have not yet been defined. W-stationarity is just the KKT conditions for (TNLP($x^*$)). So all the results of sections 3 and 4 are valid, considering MPCC-MFCQ and MPCC-QNCQ defined just imposing the standard CQs on (TNLP($x^*$)). S-stationarity is equivalent to the KKT conditions for (MPCC) viewed as a standard nonlinear problem, so Theorem 3.5 attests to the equivalence between S-stationarity and its enhanced version. However, no feasible point of (MPCC) satisfies MFCQ [22], while QNCQ can only be expected at points where the strict complementarity holds [4, Lemma 4.4]. Therefore, Theorems 3.9 and 4.2 cannot hold in general for S-stationarity. C-stationarity is defined as in Definition 6.1 but weakening (6.1) to

$$\gamma_i^G \gamma_i^H \geq 0 \quad \forall i \in I_{00}(x^*) \tag{6.2}$$

[39]. This concept is widely used for the convergence of several methods, see, e.g., [10, 29, 32]. Due to its importance, we state the definitions and results related to C-stationarity.

Definition 6.6. *We say that a feasible $x^*$ for* (MPCC) *is a* C-stationary *point if there are $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}_+^p$, and $\gamma^G, \gamma^H \in \mathbb{R}^q$ such that*

1. $\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla G(x^*)\gamma^G - \nabla H(x^*)\gamma^H = 0$;
2. $\mu_j = 0 \; \forall j \notin I_g(x^*)$, $\gamma_i^G = 0 \; \forall i \in I_{+0}(x^*)$, $\gamma_i^H = 0 \; \forall i \in I_{0+}(x^*)$, *and* (6.2) *holds.*

Definition 6.7. *We say that a feasible $x^*$ for* (MPCC) *is an* enhanced C-stationary *(EC-stationary) point if there are $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}_+^p$, and $\gamma^G, \gamma^H \in \mathbb{R}^q$ such that* (6.2) *and items* 1 *and* 3 *of Definition* 6.2 *hold.*

Analogously to M-stationarity, we define $M^{\mathrm{C}}(x^*)$, $M_Q^{\mathrm{C}}(x^*)$, and $M_E^{\mathrm{C}}(x^*)$ as the sets of *C-multipliers, quasinormal C-multipliers*, and *enhanced C-multipliers*, respectively. We can define the associated CMPCC-MFCQ and CMPCC-QNCQ by imposing the corresponding CQs on (TNLP($x^*$)) together with (6.2).

Definition 6.8. *Let $x^*$ be a feasible point for* (MPCC). *We say that*

1. *$x^*$ satisfies* CMPCC-MFCQ *if it conforms to MFCQ regarding* (TNLP($x^*$)) *and* (6.2) *holds;*
2. *$x^*$ satisfies* CMPCC-QNCQ *if there is no nonnull C-multiplier $(\lambda, \mu, \gamma^G, \gamma^H)$, $\mu \geq 0$, such that*

$$\nabla h(x^*)\lambda + \nabla g(x^*)\mu - \nabla G(x^*)\gamma^G - \nabla H(x^*)\gamma^H = 0,$$

(6.2)*, and item* 3(a) *of Definition* 6.2 *hold.*

Everything we did for M-stationarity remains valid for C-stationarity after straightforward adaptations. We summarize this in the next result.

Theorem 6.9. *Every C-stationary point is EC-stationary and vice versa. Let $x^*$ be a C-stationary point.*

1. *If CMPCC-MFCQ holds at $x^*$, then $M^C(x^*) = M_Q^C(x^*) = M_E^C(x^*)$.*
2. *$M_E^C(x^*) \subset M_Q^C(x^*)$ are nonempty, and $M_Q^C(x^*)$ is closed.*
3. *$M_Q^C(x^*)$ is compact if CMPCC-QNCQ holds at $x^*$.*

A well-known problem related to MPCC is the *mathematical program with vanishing constraints* (MPVC). It was proposed in [2] and consists of (MPCC) changing the constraints involving $G$ and $H$ by

$$H(x) \geq 0, \quad G_i(x)H_i(x) \leq 0, \; i = 1, \ldots, q.$$

Inspired by what was done for MPCCs, enhanced FJ-type conditions were defined for MPVCs in [33]. By comparing the works [31] and [33], all the above discussion is applicable to MPVCs by making straightforward adaptations.

Enhanced stationarity was developed in [27] for another much more general related problem that includes MPCC, MPVC, cone-constrained programming, semidefinite programming, and mathematical programs with semidefinite cone complementarity constraints, among others. The authors considered the *mathematical program with geometric constraints* (MPGC) defined as

$$\min_x f(x) \quad \text{subject to} \quad x \in \Omega, \quad F(x) \in \Lambda, \tag{MPGC}$$

where $\mathbb{X}$ is a Banach space, $\mathbb{Y}$ is a finite dimensional Hilbert space, $f : \mathbb{X} \to \mathbb{R}$ and $F : \mathbb{X} \to \mathbb{Y}$ are Lipschitz functions near the point of interest, and $\Omega \subset \mathbb{X}$, $\Lambda \subset \mathbb{Y}$ are nonempty and closed sets. An enhanced FJ optimality necessary condition associated with items 2(a) and (b) of Definition 2.2 was proposed for (MPGC), from which an enhanced FJ optimality for nonsmooth standard nonlinear optimization was derived. The application of our ideas to this general scenario is beyond the scope of this work and should be considered in future research.

**6.2. Multiobjective optimization.** Multiobjective problems deal with the minimization of multiple objective functions simultaneously. They are written as

$$\text{(MOP)} \qquad \min_x \, (f_1(x), \ldots, f_q(x)) \quad \text{subject to} \quad h(x) = 0, \quad g(x) \leq 0,$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, q$, are continuously differentiable functions. Here "min" means that we search for an *efficient* (or *Pareto* optimal) point $x^*$ in the sense that there is no feasible $\bar{x}$ such that $f_i(\bar{x}) \leq f_i(x^*)$ for all $i$ and $f_j(\bar{x}) < f_j(x^*)$ for some $j$. A weaker concept, but suitable for establishing convergence theory of algorithms, is the *weak efficiency* (or *weak Pareto* optimality). A feasible point $x^*$ is said to be *weakly Pareto* if there is no feasible $\bar{x}$ such that $f_i(\bar{x}) < f_i(x^*)$ for all $i$. It is worth noting that these concepts are standard in the literature [21].

It is well known that every optimal solution of the weighted-sum scalarization

$$\min_x \sum_{i=1}^{q} \sigma_i f_i(x) \quad \text{subject to} \quad h(x) = 0, \quad g(x) \leq 0, \tag{6.3}$$

is a weakly Pareto point for (MOP), for each fixed $\sigma \in \mathbb{R}^q$ such that $\sum_{i=1}^{q} \sigma_i = 1$ and $\sigma_i \geq 0$ for all $i$ [21]. Therefore, (6.3) is commonly used to address (MOP), including for deriving FJ/KKT-type stationarity. In [26], enhanced FJ conditions for (MOP) are established by just changing item 1 of Theorem 2.1 by

$$\sum_{i=1}^{q} \sigma_i \nabla f_i(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0,$$

where $(\sigma_1, \ldots, \sigma_q, \lambda, \mu) \neq 0$ and $\sigma \geq 0$. The other conditions are exactly the same, and thus it is straightforward to adapt our theory to (MOP) using the following enhanced stationarity (note that CQs do not depend on the objectives, so they are the standard ones).

DEFINITION 6.10. *We say that a feasible $x^*$ for* (MOP) *is an E-KKT point if there are $\sigma \in \mathbb{R}^q$, $\lambda \in \mathbb{R}^m$, and $\mu \in \mathbb{R}^p_+$ such that*
1. $\sum_{i=1}^{q} \sigma_i \nabla f_i(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$*;*
2. $\sum_{i=1}^{q} \sigma_i = 1$ *and* $\sigma_i \geq 0$ *for all* $i$*;*
3. *items* 2(a) and (c) *of Definition* 2.2 *hold.*

**7. Conclusions.** Since the seminal work of Fritz John [30], the related FJ optimality conditions were improved by adding extra conditions verified by minimizers [13, 14, 15, 28, 41]. These enhanced concepts carry an additional sequential condition that connects the signs of nonnull multipliers with a primal infeasible sequence, from which the notion of *quasinormal multipliers* (see item 2(a) of Definition 2.2) derives. In this paper we deepen the study of enhanced FJ/KKT conditions for smooth nonlinear optimization. A new type of multiplier is defined, which we call *enhanced multipliers.* Similar to the informative multipliers defined in [15], they carry information about the sensitivity of the constraints around the KKT point. We argue that such extra conditions do not strengthen the set of KKT points. Our result differs from previous ones [13, 15] that only tackled this equivalence for minimizers. We also analyze the properties of the enhanced/quasinormal multiplier set such as compactness and convexity. Furthermore, we apply our theory to mathematical programs with complementarity constraints and multiobjective optimization, extending and clarifying previous results from the literature.

Since several methods work with a primal-dual pair, the fact that KKT points satisfy an enhanced stationarity draws attention to the study of algorithms that generate quasinormal multipliers. In this sense, we present some insights about the multipliers generated by the augmented Lagrangian method ALGENCAN defined in [3, 18]. We discuss the implications for the boundedness of the penalty parameter sequence, which should be addressed in future work. Another interesting topic to be investigated is the effect of enhanced/quasinormal multipliers on the second-order stationarity; in fact, we can define it using only such multipliers. By narrowing the multipliers, we can potentially obtain suitable second-order conditions for showing the convergence of algorithms.

**Acknowledgment.** We would like to thank the two anonymous referees for their valuable suggestions.

REFERENCES

[1] J. Abadie, *On the Kuhn-Tucker theorem*, in Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam; John Wiley & Sons, New York, 1967, pp. 19–36.

[2] W. Achtziger and C. Kanzow, *Mathematical programs with vanishing constraints: Optimality conditions and constraint qualifications*, Math. Program., 114 (2008), pp. 69–99, https://doi.org/10.1007/s10107-006-0083-3.

[3] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, *On augmented Lagrangian methods with general lower-level constraints*, SIAM J. Optim., 18 (2007), pp. 1286–1309, https://doi.org/10.1137/060654797.

[4] R. Andreani, N.S. Fazzio, M. L. Schuverdt, and L. D. Secchin, *A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences*, SIAM J. Optim., 29 (2019), pp. 743–766, https://doi.org/10.1137/17M1147330.

[5] R. Andreani, G. Haeser, and J. M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641, https://doi.org/10.1080/02331930903578700.

[6] R. Andreani, G. Haeser, L. M. Mito, A. Ramos, and L. D. Secchin, *On the best achievable quality of limit points of augmented Lagrangian schemes*, Numer. Algorithms, 90 (2022), pp. 851–877, https://doi.org/10.1007/s11075-021-01212-8.

[7] R. Andreani, G. Haeser, M. L. Schuverdt, L. D. Secchin, and P. J. S. Silva, *On scaled stopping criteria for a safeguarded augmented Lagrangian method with theoretical guarantees*, Math. Program. Comput., 14 (2022), pp. 121–146, https://doi.org/10.1007/s12532-021-00207-9.

[8] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, *A relaxed constant positive linear dependence constraint qualification and applications*, Math. Program., 135 (2012), pp. 255–273, https://doi.org/10.1007/s10107-011-0456-0.

[9] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, *Two new weak constraint qualifications and applications*, SIAM J. Optim., 22 (2012), pp. 1109–1135, https://doi.org/10.1137/110843939.

[10] R. Andreani, G. Haeser, L. D. Secchin, and P. J. S. Silva, *New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences*, SIAM J. Optim., 29 (2019), pp. 3201–3230, https://doi.org/10.1137/18M121040X.

[11] R. Andreani, J. M. Martínez, and M. L. Schuverdt, *On the relation between constant positive linear dependence condition and quasinormality constraint qualification*, J. Optim. Theory Appl., 125 (2005), pp. 473–483, https://doi.org/10.1007/s10957-004-1861-9.

[12] R. Andreani, J. M. Martínez, and B. F. Svaiter, *A new sequential optimality condition for constrained optimization and algorithmic consequences*, SIAM J. Optim., 20 (2010), pp. 3533–3554, https://doi.org/10.1137/090777189.

[13] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, 1999.

[14] D. P. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, 2003.

[15] D. P. Bertsekas and A. E. Ozdaglar, *Pseudonormality and a Lagrange multiplier theory for constrained optimization*, J. Optim. Theory Appl., 114 (2002), pp. 287–343, https://doi.org/10.1023/A:1016083601322.

[16] D. P. Bertsekas, A. E. Ozdaglar, and P. Tseng, *Enhanced Fritz John conditions for convex programming*, SIAM J. Optim., 16 (2006), pp. 766–797, https://doi.org/10.1137/040613068.

[17] E. G. Birgin, D. Fernández, and J. M. Martínez, *The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems*, Optim. Methods Softw., 27 (2012), pp. 1001–1024, https://doi.org/10.1080/10556788.2011.556634.

[18] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, Philadelphia, PA, 2014, https://doi.org/10.1137/1.9781611973365.

[19] C. Büskens and D. Wassel, *The ESA NLP solver WORHP*, in Modeling and Optimization in Space Engineering, Springer, New York, 2013, pp. 85–110, https://doi.org/10.1007/978-1-4614-4469-5_4.

[20] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Lancelot: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*, Springer Ser. Comput. Math. 17, Springer-Verlag, Berlin, Heidelberg, 1992, https://doi.org/10.1007/978-3-662-12211-2.

[21] M. Ehrgott, *Multicriteria Optimization*, 2nd ed., Springer-Verlag, 2005, https://doi.org/10.1007/3-540-27659-9.

[22] M. L. Flegel and C. Kanzow, *On the Guignard constraint qualification for mathematical programs with equilibrium constraints*, Optimization, 54 (2005), pp. 517–534, https://doi.org/10.1080/02331930500342591.

[23] R. Fletcher and S. Leyffer, *User Manual for filterSQP*, Technical report NA/181, University of Dundee, Dundee, Scotland, 1998, https://wiki.mcs.anl.gov/leyffer/images/5/58/SQP_manual.pdf.

[24] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Program., 12 (1977), pp. 136–138, https://doi.org/10.1007/BF01593777.

[25] P. E. Gill, W. Murray, and M. A. Saunders, *SNOPT: An SQP algorithm for large-scale constrained optimization*, SIAM Rev., 47 (2005), pp. 99–131, https://doi.org/10.1137/S0036144504446096.

[26] G. Giorgi, B. Jiménez, and V. Novo, *Approximate Karush–Kuhn–Tucker condition in multiobjective optimization*, J. Optim. Theory Appl., 171 (2016), pp. 70–89, https://doi.org/10.1007/s10957-016-0986-y.

[27] L. Guo, J. J. Ye, and J. Zhang, *Mathematical programs with geometric constraints in Banach spaces: Enhanced optimality, exact penalty, and sensitivity*, SIAM J. Optim., 23 (2013), pp. 2295–2319, https://doi.org/10.1137/130910956.

[28] M. R. Hestenes, *Optimization Theory: The Finite Dimensional Case*, John Wiley & Sons, New York, 1975.

[29] A. F. Izmailov, M. V. Solodov, and E. I. Uskov, *Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints*, SIAM J. Optim., 22 (2012), pp. 1579–1606, https://doi.org/10.1137/120868359.

[30] F. John, *Extremum problems with inequalities as subsidiary conditions*, in Traces and Emergence of Nonlinear Programming, Springer, Basel, 2013, pp. 197–215, https://doi.org/10.1007/978-3-0348-0439-4_9.

[31] C. Kanzow and A. Schwartz, *Mathematical programs with equilibrium constraints: Enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results*, SIAM J. Optim., 20 (2010), pp. 2730–2753, https://doi.org/10.1137/090774975.

[32] C. Kanzow and A. Schwartz, *The price of inexactness: Convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited*, Math. Oper. Res., 40 (2015), pp. 253–275, https://doi.org/10.1287/moor.2014.0667.

[33] A. Khare and T. Nath, *Enhanced Fritz John stationarity, new constraint qualifications and local error bound for mathematical programs with vanishing constraints*, J. Math. Anal. Appl., 472 (2019), pp. 1042–1077, https://doi.org/10.1016/j.jmaa.2018.11.063.

[34] J. Kyparisis, *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming*, Math. Program., 32 (1985), pp. 242–246, https://doi.org/10.1007/bf01586095.

[35] L. Minchenko, *Note on Mangasarian–Fromovitz-like constraint qualifications*, J. Optim. Theory Appl., 182 (2019), pp. 1199–1204, https://doi.org/10.1007/s10957-019-01519-6.

[36] L. Minchenko and S. Stakhovski, *On relaxed constant rank regularity condition in mathematical programming*, Optimization, 60 (2011), pp. 429–440, https://doi.org/10.1080/02331930902971377.

[37] R. Rockafellar, *The multiplier method of Hestenes and Powell applied to convex programming*, J. Optim. Theory Appl., 12 (1973), pp. 555–562, https://doi.org/10.1007/bf00934777.

[38] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, 1st ed., Grundlehren Math. Wiss. 317, Springer-Verlag, Berlin, Heidelberg, 1998, https://doi.org/10.1007/978-3-642-02431-3.

[39] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22, https://doi.org/10.1287/moor.25.1.1.15213.

[40] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57, https://doi.org/10.1007/s10107-004-0559-y.

[41] J. J. Ye and J. Zhang, *Enhanced Karush-Kuhn-Tucker condition and weaker constraint qualifications*, Math. Program., 139 (2013), pp. 353–381, https://doi.org/10.1007/s10107-013-0667-7.

[42] J. J. Ye and J. Zhang, *Enhanced Karush-Kuhn-Tucker conditions for mathematical programs with equilibrium constraints*, J. Optim. Theory Appl., 163 (2014), pp. 777–794, https://doi.org/10.1007/s10957-013-0493-3.