On enhanced KKT optimality conditions for smooth nonlinear optimization*

Roberto Andreani† Maria L. Schuverdt‡ Leonardo D. Secchin§

December 6, 2022

Abstract

The Fritz-John (FJ) and KKT conditions are a fundamental tool to characterize minimizers and lie in the root of almost any method 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 first proposed by Hestenes in the 1970s and later extensively studied by Bertsekas and collaborators. In this work we revisited enhanced KKT stationarity for standard (smooth) nonlinear programming. We prove that every KKT point satisfies usual enhanced versions from the literature. Therefore, enhanced KKT stationarity only concerns the Lagrange multipliers. We then analyse some properties of such improved multipliers with quasi-normality (QN), showing in particular that the set of them is compact if, and only if, QN holds. In this sense, QN has the same status for enhanced KKT that Mangasarian-Fromovitz constraint qualification has for classical KKT. Also, we report some consequences of adding an extra abstract constraint to the problem. As enhanced FJ/KKT concepts are obtained by aggregating sequential conditions to FJ/KKT, we report the relevance of our results for the so-called sequential optimality conditions, that have been crucial to generalize global convergence of a consolidate safeguarded augmented Lagrangian method. Finally, we apply our theory to mathematical programs with complementarity constraints and multi-objective problems, improving and clarifying previous results from the literature.

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

1 Introduction

In this paper we consider the general constrained optimization problem

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

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 no convexity assumption is assumed. Practical algorithms rely on find a reasonable stationary point, that is, a computable point that has characteristics of minimizers. In this sense, the most important tool to characterize minimizers of (P) is the Karush-Kuhn-Tucker (KKT) conditions. They are used to assert theoretical convergence of practically every method in

*This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2018/24293-0 and 2017/18308-2), CNPq (grants 306988/2021-6 and 309136/2021-0) and PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016).

†Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil. Email: andreani@ime.unicamp.br

‡CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina. Email: schuverd@mate.unlp.edu.ar

§Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, ES, Brazil. Email: leonardo.secchin@ufes.br

constrained optimization and have been specialized/adapted to different specific contexts, such as multi-objective optimization, nonsmooth optimization among many others. Also, KKT conditions inspire practical stopping criteria for several algorithms [5, 7, 12, 19, 20, 23, 25, 37].

Contrary to the KKT conditions, the well known Fritz-John (FJ) [28] ones do not need the fulfilment of any constraint qualification (CQ) to hold at minimizers in general. Such conditions gained much attention with the works of Hestenes [27], Bertsekas, Ozdaglar and collaborators [13, 14, 15, 16], Ye [38, 39] and others, which deal with extensions of the original FJ conditions, the so-called *enhanced FJ conditions*. Hestenes [27] was the first to propose such an extension, showing the existence of special (Lagrange) multipliers satisfying certain property associated with infeasible sequences (see Theorem 1, item 3(a)). This enhanced FJ version is intimately linked to the pure external penalty method [13], and puts the quasi-normality CQ (Definition 3) in evidence when attesting the validity of KKT conditions. Later on, other enhanced versions were proposed by adding extra properties mainly related to the sensitivity of constraints associated with null multipliers. Also, enhanced KKT conditions were naturally derived by imposing a non-null multiplier for the objective.

All enhanced FJ/KKT conditions consist of usual FJ/KKT ones together extra conditions. So, one might expect that enhanced stationarity characterizes minima better than the usual KKT conditions. Surprisingly, this is not true from the primal point of view. Specifically, we prove in this paper that not only local minimizers of (P) satisfy enhanced KKT conditions, as done in previous works, but also *every* KKT point. That is, from the primal point of view, enhanced and usual KKT are equivalent. Thus, the improvements from enhanced stationarity concern the multipliers only. We then investigate new properties of these multipliers, called *quasinormal* in the literature. In [38] it was proved that the set of quasinormal multipliers $M_Q(x^*)$ associated with a KKT point x^* is bounded if the quasi-normality (QN) CQ holds at x^* (such connections justify the name “quasinormal”). Here, we prove that, indeed, $M_Q(x^*)$ is always closed, and compact if, and only if, QN holds. In this sense, QN has the same status as the Mangasarian-Fromovitz CQ for usual multipliers set [24]. Contrary to the last set, we show that $M_Q(x^*)$ is not convex in general. We also discuss the relation between quasinormal and *informative* multipliers [14], which are associated with the most stringent enhanced FJ version. The developed theory here is applied to mathematical programs with complementarity constraints and multi-objective optimization, improving and clarifying previous results from the literature.

Another relevant issue is the consequences of enhanced FJ/KKT conditions and QN for constrained nonlinear optimization methods. In fact, for many years these concepts remained restricted to the pure external method, and so their applicability to practical algorithms were limited. Only recently we proved [4] that the safeguarded augmented Lagrangian method, known as ALGENCAN [3], converges to KKT points under QN. ALGENCAN has an “official” general purpose implementation that has been used successfully in several applications [18]. The theory developed in [4] is supported by a sequential optimality condition, called *positive approximate KKT* (PAKKT), that captures the connection between the signs of the multipliers generated by the method with the infeasibility of its primal sequence. PAKKT is based on the enhanced FJ conditions of Hestenes. One interesting consequence of PAKKT is that the sequences of multipliers generated by ALGENCAN are bounded if QN holds at the limit point. This result is surprisingly since QN does not imply the boundedness of the standard multipliers set. The validity of enhanced stationarity at all KKT points that we prove in this paper is an important step to study adequately practical methods that potentially generates quasinormal multipliers, at least under QN. We give some insights in this line concerning ALGENCAN, paving the way to improve the previous results on boundedness of the penalty parameter [3, 17] that require e.g. uniqueness of usual multipliers. This is an important issue regarding the stability of the method.

This paper is organized as follows. In Section 2 we present enhanced FJ/KKT stationarity. In Section 3 we prove the equivalence between enhanced and classical KKT conditions. Section 4 is devoted to discussing/proving properties of the set of quasinormal multipliers such as compactness and convexity. Also, we consider the presence of abstract constraints. Section 5 is dedicated to discussing the consequences of enhanced stationarity on the ALGENCAN method. In Section 6 we apply our theory to the widely studied classes of problems cited above, namely, mathematical programs with complementarity constraints and multi-objective optimization. Finally, Section 7 brings our conclusions and future work.

Notation $\operatorname{sgn} a$ will denote de sign function, that is, $\operatorname{sgn} a = 1$ if $a > 0$ and $\operatorname{sgn} a = -1$ if $a < 0$. We define $I_g(z) = \{j \mid g_j(z) = 0\}$. $\|\cdot\|$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are an arbitrary norm, the Euclidean norm and the sup-norm, respectively. We use the "o and O notations", common in the computational complexity theory: 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 $b_k \le m_k \cdot a_k$ (respectively $b_k \le M \cdot a_k$) for all k.

2 Enhanced stationarity for smooth problems

Enhanced FJ/KKT versions have been proposed in the literature since at least the 1970s [27], and gained prominence with the work of Bertsekas and his collaborators [13, 14, 15, 16]. The most general enhanced FJ conditions for (P) is presented in the next theorem.

Theorem 1 ([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* $w(x) = \min\left\{\min_{i \in I_{\neq}} |h_i(x)|,\ \min_{i \in J_+} g_j(x)_+\right\}$.

Item 3(a) implies the slackness complementary 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 non-null multipliers ($I_{\neq} \cup J_+ \neq \emptyset$), it is possible to reach x^* "from the outside" of the feasible set so that $f(x^k)$ is smaller than the optimal value; this is a typical behaviour of external penalty approaches. Finally, item 3(c) carries sensitivity properties of the constraints associated with null multipliers, including inactive inequality ones.

Enhanced KKT conditions are obtained by setting $\sigma = 1$ in Theorem 1 and one or more conditions of item 3. Such conditions were treated in the literature since the late 1990s. In [13, 38], only items 1, 3(a) are considered; in [15, 14], items 3(b,c) are aggregated. These works state such conditions only on local minimizers, like Theorem 1, and their proofs were conducted applying 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,b), and changing 3(c) putting "$O(w^k)$" instead of "$o(w^k)$". The "O" notation is weaker than "o", so it results in a less stringent sensitivity measure of the constraints associated with null multipliers.

Definition 1. *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*

$$I_{\neq} = \{i \mid \lambda_i \neq 0\} \quad \text{and} \quad J_+ = \{j \mid \mu_j > 0\}, \tag{1}$$

   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*

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

Given a feasible point x^* for (P), we define the set of usual associated Lagrange multipliers and the set of those multipliers satisfying the E-KKT conditions 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

$$M_Q(x^*) = \left\{(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^p_+ \;\middle|\; \begin{array}{l}\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0, \quad \exists\{x^k\}\to x^*\\ \text{such that items 2(a–c) of Definition 1 hold}\end{array}\right\},$$

respectively. Clearly $M_Q(x^*) \subset M(x^*)$, but the contrary inclusion is not true in general; see for instance Examples 1 and 2 below. Of course, as occurs with KKT, some additional hypotheses (constraint qualifications) are necessary to ensure that a local minimizer satisfies enhanced KKT stationarity.

Ye and Zhang [38] consider the set $M_Q(x^*)$ without conditions 2(b–c). The multipliers in this set are called *quasinormal* because their intimate connection with the quasi-normality CQ (see Definition 3). We show in the next sections that $M_Q(x^*)$ has also strongly connections with QN, so we refer to its elements by *quasinormal multipliers* too. We establish in the next sections new connections between E-KKT and KKT, and between $M_Q(x^*)$ and QN.

3 Equivalence between enhanced and classical KKT

The main purpose of this section is to highlight the primal 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 version of MFCQ for (P).

Definition 2. *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*

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

The next lemma contains an interesting geometric property of smooth constraints. Roughly speaking, it says that when the gradients of active constraints at x^* are linearly independent, it is possible to approaches x^* by "any side of each constraint". 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 1. *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 a sequence $\{x^k\}\subset\mathbb{R}^n\backslash\{x^*\}$ converging to x^* and $\gamma > 0$ such that $s_i c_i(x^k) \geq \gamma\|x^k - x^*\|_2$ for all k and $i = 1,\ldots,q$.*

Proof. As in the proof of [4, Lemma 2.6], we consider the set

$$\mathcal{S}_i^\gamma = \left\{ x\in\mathbb{R}^n\backslash\{x^*\} \;\middle|\; \begin{array}{ll}\nabla c_i(x^*)^t\dfrac{x-x^*}{\|x-x^*\|_2} \geq 2\gamma & \text{if } s_i = 1\\[2ex] \nabla c_i(x^*)^t\dfrac{x-x^*}{\|x-x^*\|_2} \leq -2\gamma & \text{if } s_i = -1\end{array}\right\}$$

for each $i = 1,\ldots,q$ and $\gamma > 0$. As $s_i\nabla c_i(x^*) + x^* \in \mathcal{S}_i^\gamma$ for all $0 < \gamma \leq \|\nabla c_i(x^*)\|_2/2$, we have $\mathcal{S}_i^\gamma \neq \emptyset$ for such γ's. For each $\gamma > 0$ small enough and $i = 1,\ldots,q$, the first order Taylor expansion of c_i around x^*, together the fact that $c_i(x^*) = 0$, implies the existence of an open neighbourhood $\mathcal{B}^\gamma(x^*)$ of x^* such that

$$s_i c_i(x) \geq \gamma\|x - x^*\|_2, \quad \forall x \in \mathcal{S}_i^\gamma \cap \mathcal{B}^\gamma(x^*). \tag{2}$$

Let

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

where

$$\mathcal{I}_+ = \{i \mid s_i = 1\} \quad \text{and} \quad \mathcal{I}_- = \{i \mid s_i = -1\}.$$

By hypotheses, the gradients of all constraints in D evaluated at x^* are linearly independent and thus MFCQ with respect to D holds at x^*. So, there is an unitary $d \in \mathbb{R}^n$ such that

$$\nabla c_j(x^*)^t d > 0, \quad \nabla c_l(x^*)^t d < 0, \quad j \in \mathcal{I}_+,\ l \in \mathcal{I}_-. \tag{3}$$

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. Furthermore, by the differentiability of c_i at x^*, $i = 1, \ldots, q$, we have

$$c_i(x^k) = c_i(x^* + d/k) = c_i(x^*) + 1/k \nabla c_i(x^*)^t d + r_i(d/k), \tag{4}$$

where $k \cdot r_i(d/k)$ tends to zero as $k \to \infty$. As $c(x^*) = 0$, dividing the above expression by $\|x^k - x^*\|_2 = 1/k$ arrives at

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

for all k large enough. Therefore, $x = x^k$ satisfies (2) for all k taking

$$\gamma = 1/2 \cdot \min\left\{\nabla c_j(x^*)^t d,\ -\nabla c_l(x^*)^t d \mid j \in \mathcal{I}_+,\ l \in \mathcal{I}_-\right\} > 0$$

(note that d does not depends on γ, and therefore the above minimum is well defined). This concludes the proof. □

The next lemma is a consequence of the well known Carathéodory lemma, and it will be useful to our purposes. It is a simplified version of [8, Lemma 1].

Lemma 2. *Let $\mathcal{Z}$ be a set of indexes and suppose that $z = \sum_{i \in \mathcal{Z}} \alpha_i v_i$ is a linear combination of vectors $v_i \in \mathbb{R}^n$. If $\alpha_i \neq 0$ for all $i \in \mathcal{B}$, 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 will show that the stationarity concept of Definition 1 is equivalent to the classical KKT conditions. This is an interesting issue since

- the classical KKT conditions lies in the root of nonlinear programming. They have been used for decades, serve as the theoretical basis for practically all existing methods, and were adapted to non-standard contexts such as multi-objective and nonsmooth optimization;
- enhanced FJ/KKT conditions have been proposed 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 characterizes minimizers of (P) as good as its enhanced versions. It is particularly interesting when we confront item 2(a) of Definition 1 with interior point strategies.

Theorem 2. *Every KKT point x^* is E-KKT and vice-versa.*

Proof. It is trivial that every E-KKT point is KKT. Let us prove the converse. Although part of the proof is based on the arguments of [4, Lemma 2.6], we provide a complete proof for the sake of clarification.

Let x^* be a KKT point, λ and μ any associated multipliers and $I_{\neq}$ and J_+ be the induced sets of indexes of nonzero multipliers as in (1). Clearly it is sufficient to consider only the nontrivial case $\nabla f(x^*) \neq 0$, which implies $I_{\neq} \cup J_+ \neq \emptyset$. Thus, Lemma 2 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

$$\nabla f(x^*) + \sum_{i \in \mathcal{I}} \widehat{\lambda}_i \nabla h_i(x^*) + \sum_{j \in \mathcal{J}} \widehat{\mu}_j \nabla g_j(x^*) = 0, \tag{5}$$

$\widehat{\lambda}_i \neq 0$ for all $i \in \mathcal{I}$, $\widehat{\mu}_j > 0$ for all $j \in \mathcal{J}$, and such that the gradients of constraints in (5) are linearly independent. Applying Lemma 1 on such gradients, we obtain $\{x^k\}$ satisfying

$$\operatorname{sgn}(\widehat{\lambda}_i) \cdot h_i(x^k) \geq \gamma \|x^k - x^*\|_2, \ \forall i \in \mathcal{I} \quad \text{and} \quad g_j(x^k) \geq \gamma \|x^k - x^*\|_2, \ \forall j \in \mathcal{J} \tag{6}$$

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 1 are valid with the sequence $\{x^k\}$ and multipliers $\widehat{\lambda}$, $\widehat{\mu}$.

As in the proof of Lemma 1, we have $x^k = x^* + d/k$ where d is unitary satisfying

$$\operatorname{sgn}(\widehat{\lambda}_i) \cdot \nabla h_i(x^*)^t d > 0, \quad \nabla g_j(x^*)^t d > 0, \quad i \in \mathcal{I}, \ j \in \mathcal{J}. \tag{7}$$

Multiplying (5) by d and using (7) we conclude that $\nabla f(x^*)^t d < 0$. Thus, by the first order Taylor expansion of f around x^* with increment d/k, as in (4), we obtain item 2(b) of Definition 1 using x^k for all k sufficiently large.

Finally, let us prove that the same sequence $\{x^k\}$ also satisfies item 2(c) of Definition 1, at least 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

$$|h_i(x^k)| = |h_i(x^k) - h_i(x^*)| \leq L\|x^k - x^*\|_2 \tag{8}$$

and

$$g_j(x^k)_+ \leq \left[g_j(x^k) - g_j(x^*)\right]_+ \leq L\|x^k - x^*\|_2 \tag{9}$$

for all k sufficiently large, $i \notin \mathcal{I}$ and $j \notin \mathcal{J}$. Considering the multipliers $\widehat{\lambda}$ and $\widehat{\mu}$ in (5), $w(x^k)$ 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 (6), (8) and (9) we have, for all k large enough,

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

Thus $\{x^k\}$ fulfils item 2(c) of Definition 1 as we wanted, concluding the proof. □

Remark 1. *By the proof of Theorem 1, a sequence $\{x^k\}$ fulfilling item 2(a) of Definition 1 has a subsequence satisfying also items 2(b) and 2(c). So, the only really important condition is 2(a), at least in the smooth setting. In other words, the set of multipliers as defined e.g. in [15, 33, 38], without 2(b,c) (called* strong multipliers *in [15]), coincides with $M_Q(x^*)$.*

Remark 2. *In [26], the equivalence between KKT points and those satisfying items 1 and 2(a,b) of Definition 1 are proved in the context of multi-objective optimization. On the one hand, Theorem 2 includes item 2(c) of Definition 1.*

It is easy to exhibit examples where a multiplier vector serves to the classical KKT conditions, but not to the enhanced version (see Example 1 below). From the proof of Theorem 2, given an arbitrary multiplier vector (λ, μ), the vector $(\widehat{\lambda}, \widehat{\mu})$ obtained by the application of Lemma 2 on (λ, μ) is suitable for both KKT and E-KKT stationarity. This lemma also ensures that these multipliers preserve signs, that is,

$$\lambda_i \widehat{\lambda}_i \geq 0, \ \forall i \quad \text{and} \quad \mu_j \widehat{\mu}_j \geq 0, \ \forall j.$$

To put multipliers like $(\widehat{\lambda}, \widehat{\mu})$ in perspective, we say that a multiplier vector has *linearly independent support* if the gradients of equality and inequality constraints associated with non-null multipliers are linearly independent. Note that a possible result of an application of Lemma 2 on a multiplier vector (λ, μ) with linearly independent support is (λ, μ) itself. Also, it is clear that every KKT point admits such multipliers.

Corollary 1. *Let x^* be a KKT point. Then every multiplier vector with linearly independent support fulfils all conditions of Definition 1. In other words, these multipliers are quasinormal.*

Remark 3. *Bertsekas [14] proved Theorem 1 for the case where all functions f, h and g are continuously differentiable. Ye and Zhang [38] provided a nonsmooth version of this result, without items 2(b,c), where all data functions are only Lipschitz continuous around x^*. It is straightforward to verify that Theorem 2 remains valid if we suppose that f, h and g are differentiable, not necessarily with continuous derivatives (remember that Lemma 1 is valid with differentiability only). As consequence, Corollary 1 is still valid in this case.*

In a reasoning similar to Corollary 1, we may ask what known CQs ensure that *any* Lagrange multiplier is quasinormal. Certainly, the *linear independence of the gradients of the active constraints* (LICQ) is one of them, since in this case the unique multipliers verify those of Corollary 1. The same does not occur with linear constraints and, consequently, with any implied CQ as the next example shows.

Example 1. *The point $x^* = 0$ is KKT for the problem of minimizing x subject to $g_1(x) = x \leq 0$ and $g_2(x) = -x \leq 0$ 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 the MFCQ, as LICQ, ensures that any multiplier vector works for E-KKT.

Theorem 3. *Let x^* be a KKT point. If MFCQ holds at x^* then any associated multiplier vector (μ, λ) is quasinormal, that is, $M(x^*) = M_Q(x^*)$.*

Proof. The proof is an adaptation of that of Theorem 2. Let $(\lambda, \mu) \in M(x^*)$ and $I_{\neq}$, J_+ as in (1), that is,

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

As in the proof of Theorem 2, we consider only the nontrivial case $\nabla f(x^*) \neq 0$. As MFCQ holds, it is also valid at x^* with respect to the set

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

where $I_+ = \{i \in I_{\neq} \mid \lambda_i > 0\}$ and $I_- = \{i \in I_{\neq} \mid \lambda_i < 0\}$. Therefore, there is an 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 1, defining $x^k = x^* + d/k$ for all $k \geq 1$ we can conclude that x^k satisfy (6) with $\widehat{\lambda}_i = \lambda_i$ for all $i \in I_{\neq}$ and $j \in J_+$ if we take

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

Thus, items 2(a,b) of Definition 1 can be obtained using x^k for all k sufficiently large. Now, we take $L > \max\{|\nabla h_i(x^*)^t d|, |\nabla g_j(x^*)^t d| \mid i \notin I_{\neq},\ j \notin J\}$ to obtain (8) and (9) for all $i \notin I_{\neq}$ and $j \notin J_+$. In the same way, the last inequalities of the proof of Theorem 2 hold with $\mathcal{I} = I_{\neq}$ and $\mathcal{J} = J_+$, and then item 2(c) of Definition 1 is valid. This concludes the proof. □

Recently [31], it has been shown that some CQs such as the *(relaxed) constant positive linear dependence* ((R)CPLD) [8], the *(relaxed) constant rank CQ* ((R)CRCQ) [32] 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 1: we can rewrite the feasible set by the unique equality constraint $x = 0$. However, we do not known *a priori* how such a reformulation can be done. It is worth mentioning that Example 1 satisfies all the cited CQs, quasi-normality (see Definition 3 below) and also the pseudo-normality CQ defined in [15]. Thus, Theorem 3 can not be improved using other known CQ from the literature (see [4, Figure 4]).

4 Properties of the quasinormal multipliers

4.1 Compactness of $M_Q(x^*)$ under quasi-normality

It is well known that MFCQ is necessary and sufficient to the set of multipliers $M(x^*)$ be bounded [24]. Due to the peculiar nature of quasinormal multipliers, a question that arises is what condition on x^* is equivalent to the boundedness of $M_Q(x^*)$.

The connection of enhanced stationarity with external penalty approaches brings the quasi-normality CQ to the discussion. In fact, it is design to eliminate possible "wrong" multipliers generated by this method. It was introduced for the first time by Hestenes in [27] and generalized in [15] to the case where additional abstract constraints are present.

Definition 3. *We say that a feasible x^* for* (P) *satisfies the* quasi-normality CQ *if there is 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 as in Definition 1.*

In [38], it was proved that the set of quasinormal multipliers $M_Q(x^*)$ is bounded when x^* satisfies QN (actually, only condition 2(a) is considered in [38], but 2(b,c) is a consequence of it by Remark 1). Next we prove that $M_Q(x^*)$ is always closed and that QN is not only sufficient to the boundedness of $M_Q(x^*)$, but also necessary.

Theorem 4. *Let x^* be a KKT point. Then $M_Q(x^*) \neq \emptyset$ and closed. Also, $M_Q(x^*)$ is compact if, and only if, quasi-normality holds at x^*.*

Proof. By Theorem 2, $M_Q(x^*) \neq \emptyset$. Let us show that $M_Q(x^*)$ is closed. Let 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 1 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_+. Items 2(b,c) follow taking a subsequence if necessary (see Remark 1). Thus $(\lambda, \mu) \in M_Q(x^*)$, from which we conclude the closedness of $M_Q(x^*)$.

The boundedness of $M_Q(x^*)$ under QN follows from [38, Theorem 3] and Remark 1. Let us prove the converse. Suppose that x^* does not conform to the QN CQ, and take $0 \neq (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$, $\{x^k\}$ satisfying all conditions in Definition 3; that is, $\nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$, $\mu_j = 0$ if $j \notin I_g(x^*)$, and $\{x^k\} \to x^*$ such that, for each k,

$$\lambda_i h_i(x^k) > 0, \ \forall i \in I_{\neq} \quad \text{and} \quad \mu_j g_j(x^k) > 0, \ \forall j \in J_+, \tag{10}$$

where $I_{\neq} = \{i \mid \lambda_i \neq 0\}$ and $J_+ = \{j \mid \mu_j > 0\}$ are not both empty. For each k, we apply Lemma 2 with $\mathcal{Z} = I_{\neq} \cup J_+$ to obtain sets $I^k_{\neq} \subset I_{\neq}$ and $J^k_+ \subset J_+$, and vectors λ^k and μ^k such that

$$\sum_{i \in I^k_{\neq}} \lambda^k_i \nabla h_i(x^*) + \sum_{i \in J^k_+} \mu^k_j \nabla g_j(x^*) = 0, \tag{11}$$

$\lambda^k_i \lambda_i > 0$ for all $i \in I^k_{\neq}$, $\lambda^k_i = 0$ otherwise, $\mu^k_j \mu_j > 0$ for all $j \in J^k_+$, $\mu^k_j = 0$ otherwise, and the gradients in (11) are linearly independent for each fixed k. Dividing (11) by $\|(1, \lambda^k, \mu^k)\|_\infty$ and taking the limit, we obtain a vector $0 \neq (\widehat{\lambda}, \widehat{\mu}) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that

$$\sum_{i=1}^{m} \widehat{\lambda}_i \nabla h_i(x^*) + \sum_{j \in I_g(x^*)} \widehat{\mu}_j \nabla g_j(x^*) = 0, \qquad \widehat{\mu}_j = 0,\ \forall j \notin I_g(x^*), \tag{12}$$

$$\widehat{\lambda}_i \lambda_i > 0,\ \forall i \in \widehat{I}_{\neq} = \{i \mid \widehat{\lambda}_i \neq 0\} \quad \text{and} \quad \widehat{\mu}_j \mu_j > 0,\ \forall j \in \widehat{J}_+ = \{j \mid \widehat{\mu}_j > 0\},$$

where $\{\nabla h_i(x^*), \nabla g_j(x^*) \mid i \in \widehat{I}_{\neq},\ j \in \widehat{J}_+\}$ is linearly independent. By the last condition about the signs of these non-null multipliers, $\widehat{I}_{\neq} \subset I_{\neq}$, $\widehat{J}_+ \subset J_+$ and (10), the same sequence $\{x^k\}$ associated with (λ, μ) satisfies

$$\widehat{\lambda}_i h_i(x^k) > 0,\ \forall i \in \widehat{I}_{\neq} \quad \text{and} \quad \widehat{\mu}_j g_j(x^k) > 0,\ \forall j \in \widehat{J}_+.$$

In other words, $(\widehat{\lambda}, \widehat{\mu})$ satisfies all conditions in Definition 3 as does (λ, μ).

Now, let $(\bar{\lambda}, \bar{\mu})$ be any element in $M_Q(x^*)$. Do the following steps:

1. Set $\mathcal{I} \leftarrow \emptyset$, $\mathcal{J} \leftarrow \emptyset$ and define $\mathcal{G} = \{\nabla h_i(x^*), \nabla g_j(x^*) \mid i \in \widehat{I}_{\neq} \cup \mathcal{I},\ j \in \widehat{J}_+ \cup \mathcal{J}\}$;
2. if $\bar{\lambda}_1 \neq 0$ and $\nabla h_1(x^*) \in \operatorname{span} \mathcal{G}$, rewrite $\nabla h_1(x^*)$ as a linear combination of the gradients in $\mathcal{G}$. If $\bar{\lambda}_1 \neq 0$ and $\nabla h_1(x^*) \notin \operatorname{span} \mathcal{G}$, set $\mathcal{I} \leftarrow \mathcal{I} \cup \{1\}$ and update $\mathcal{G}$. Note that $\mathcal{G}$ remains linearly independent. Repeat this procedure inductively on $\bar{\lambda}_2, \dots, \bar{\lambda}_m$;
3. Apply inductively a procedure analogous to the previous step on $\bar{\mu}_1, \dots, \bar{\mu}_p$, updating $\mathcal{J}$ and $\mathcal{G}$ accordingly.

After applying the above process, and using the fact that $(\bar{\lambda}, \bar{\mu}) \in M_Q(x^*)$, we can write

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

where a_i's and b_j's are the aggregated coefficients of all the linear combinations made during the process. Multiplying (12) by $t > 0$ and summing with the above expression, we arrive at

$$\begin{aligned} \nabla f(x^*) &+ \sum_{i \in \widehat{I}_{\neq}} (a_i + t\widehat{\lambda}_i) \nabla h_i(x^*) + \sum_{i \in \mathcal{I}} (a_i + \bar{\lambda}_i) \nabla h_i(x^*) \\ &+ \sum_{j \in \widehat{J}_+} (b_j + t\widehat{\mu}_j) \nabla g_j(x^*) + \sum_{j \in \mathcal{J}} (b_j + \bar{\mu}_j) \nabla g_j(x^*) = 0. \end{aligned} \tag{13}$$

For a fixed $\bar{t} > 0$ large enough, we have

$$(a_i + t\widehat{\lambda}_i)\widehat{\lambda}_i > 0,\ i \in \widehat{I}_{\neq}, \quad \text{and} \quad (b_i + t\widehat{\mu}_j)\widehat{\mu}_j > 0,\ j \in \widehat{J}_+, \quad \text{for all } t \geq \bar{t} \tag{14}$$

since $\widehat{\lambda}_i \neq 0$ for all $i \in \widehat{I}_{\neq}$ and $\widehat{\mu}_j > 0$ for all $j \in \widehat{J}_+$. Now, the gradients of constraints in (13) are linearly independent by construction. So we apply Lemma 1 to obtain a sequence $\{\widehat{x}^k\}$ converging to x^* such that, together with the coefficients of the gradients of constraints in (13) with $t = \bar{t}$, satisfies items 1 and 2(a) of Definition 1 (items 2(b,c) follow from Remark 1 taking a subsequence if necessary). Also, note that, due to (14), this sequence works for all $t \geq \bar{t}$. In other words, we prove that the multiplier vector within (13) is in $M_Q(x^*)$ for all $t \geq \bar{t}$. So taking $t \to \infty$ we conclude the unboundedness of $M_Q(x^*)$, and the statement is proved. □

Obviously $M(x^*)$ may be unbounded even if $M_Q(x^*)$ is bounded, since MFCQ is strictly stronger than QN. The next example illustrates this fact.

Example 2. *For the problem*

$$\min_x x \quad \textit{subject to} \quad x = 0, \quad x = 0,$$

we have $M(0) = \{(t, 1-t) \mid t \in \mathbb{R}\}$ *and* $M_Q(0) = \{(t, 1-t) \mid t \in [0, 1]\}$.

Even more, $M_Q(x^*)$ may be a singleton while $M(x^*)$ is unbounded: for the problem of minimizing x subject to $-x \leq 0$ and $x \leq 0$, we have $M_Q(0) = \{(1, 0)\}$ and $M(0) = \{(t, t-1) \mid t \geq 1\}$.

4.2 On the convexity of $M_Q(x^*)$

It easy to see that the set of usual multipliers $M(x^*)$ is always polyhedral, independently of the validity of any constraint qualification. In turn, the convexity of $M_Q(x^*)$ is not a trivial issue since the non-null multipliers may change along the segment between two elements $(\lambda^1, \mu^1), (\lambda^2, \mu^2) \in M_Q(x^*)$, and so a sequence $\{x^k\}$ satisfying item 2(a) of Definition 1 for (λ^1, μ^1) may not work for (λ^2, μ^2).

Under MFCQ, $M_Q(x^*)$ is indeed polyhedral by Theorem 3. The next example shows that, unfortunately, $M_Q(x^*)$ is not convex in general.

Example 3. *Consider the bi-dimensional problem*

$$\min x_2 \quad \textit{subject to} \quad x_2 + x_1^5 \le 0, \quad -x_2 - x_1^3 \le 0, \quad -x_2 \le 0$$

and the unique feasible point $x^* = (0,0)$. It is easy to see that $\widehat{\mu} = (0,1,0)$ and $\bar{\mu} = (1,0,2)$ are quasinormal multipliers associated with x^*. Now, consider the convex combination

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

This μ is clearly a usual multiplier, but not quasinormal. In fact, if there were a sequence $\{x^k\}$ satisfying item 2(a) of Definition 1 for μ, we should have $x_2^k < 0$ for all k by the third constraint and its associated positive multiplier. On the other hand, as the two first multipliers are both positive we should have $x_2^k + (x_1^k)^5 > 0$ and $x_2^k + (x_1^k)^3 < 0$ for all k, which would imply $(x_1^k)^3 < -x_2^k < (x_1^k)^5$, a contradiction. Thus, $\mu \notin M_Q(x^*)$. By similar reasoning, no convex combination of $\widehat{\mu}$ and $\bar{\mu}$ distinct of these vectors is in $M_Q(x^*)$.

Note that in the above example x^* does not conform to any known CQ (in particular the Guignard CQ is not valid). However, if we add the constraint $x_1 = 0$ with null multiplier, the example remains valid and the Abadie CQ [1] becomes valid. Nevertheless, the convexity of $M_Q(x^*)$ under stronger CQs remains an open question. Although we are not able to give a proof, we conjecture that $M_Q(x^*)$ is polyhedral at least under quasi-normality.

4.3 Quasinormal and informative multipliers

We back our attention to the condition 3(c) of Theorem 1. We can define an enhanced KKT stationarity concept by changing item 2(c) of Definition 1 to that condition, i.e., by using "o" instead of "O". That is, item 2(c) of Definition 1 is replaced by

$$|h_i(x^k)| = o(w(x^k)), \ \forall i \notin I_{\neq}, \text{ and } g_j(x^k)_+ = o(w(x^k)), \ \forall j \notin J_+, \tag{15}$$

where $w(x)$ is the same. The corresponding multipliers are called *informative* by Bertsekas [14], and we denote the set of them by $M_I(x^*)$. Specifically,

$$M_I(x^*) = \left\{ (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \;\middle|\; \begin{array}{l} \nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0, \quad \exists \{x^k\} \to x^* \\ \text{s.t. items 2(a–b) of Definition 1 and (15) hold} \end{array} \right\}.$$

As pointed out in [14], the set of informative multipliers can be a proper subset of $M_Q(x^*)$. In fact, note that we always have

$$M_I(x^*) \subset M_Q(x^*) \subset M(x^*)$$

since $a_k = o(b_k)$ implies $a_k = O(b_k)$. On the other hand, for the problem in Example 2 we have $(1,0) \in M_Q(0) \backslash M_I(0)$, otherwise $(1,0) \in M_I(0)$ would arrive at $|x_k| = o(|x_k|)$, which is impossible. In this example, it is easy to see that $M_I(0) = \{(t, 1-t) \mid t \in (0,1)\}$, which shows that $M_I(x^*)$ is not closed in general although it is obviously bounded under QN just as $M_Q(x^*)$.

An issue related with informative multipliers is whether Theorem 2 is valid using them, that is, if for *every* KKT point x^* we have $M_I(x^*) \neq \emptyset$. The answer is yes. Although in Proposition 2.2(a) of [15] x^* is suppose to be a local minimizer, such 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 to our purposes.

Lemma 3. *Let $c, a_1, \dots, a_m, b_1, \dots, b_r \in \mathbb{R}^n$. 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 non-empty. 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* $i = 1, \dots, m$*;*
3. $(b_j^t d^k)_+ \to \mu_j^*$ *for all* $j = 1, \dots, r$*,*

where (λ^*, μ^*) *is the element of* M *with minimum 2-norm.*

Proof. The proof follows 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 5. *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 2 is valid if we define E-KKT points using* (15) *instead of item 2(c) in Definition 1.*

Proof. The proof is similar to that of [15, Proposition 2.2(a)]. We can suppose without loss of generality that $I_g(x^*) = \{1, \dots, r\}$. From now on, we set $c = \nabla f(x^*)$, $a_i = \nabla h_i(x^*)$, $i = 1, \dots, m$, and $b_j = \nabla g_j(x^*)$, $j = 1, \dots, r$, in Lemma 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. From Lemma 3, the limit d of an unitary convergent subsequence $\{d^k / \|d^k\|_2\}_{k \in K}$ satisfies

$$\nabla f(x^*)^t d < 0, \qquad \operatorname{sgn}(\lambda_i^*) \cdot \nabla h_i(x^*)^t d > 0, \ \forall i \in I_{\neq}, \qquad \nabla g_j(x^*)^t d > 0, \ \forall j \in J_+, \tag{16}$$

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

$$|\nabla h_l(x^*)^t d^k| \leq \delta_k \cdot m(d^k), \ l \notin I_{\neq}, \quad (\nabla g_s(x^*)^t d^k)_+ \leq \delta_k \cdot m(d^k), \ s \notin J_+ \tag{17}$$

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 side of the inequalities in (17) goes 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 (16), gives $f(x^k) < f(x^*)$ for all k large enough. Then item 2(b) of Definition 1 holds. Analogously, for all i we have

$$h_i(x^k) = h_i(x^k) - h_i(x^*) = \|x^k - x^*\|_2 \left[\nabla h_i(x^*)^t d + \frac{o(\|x^k - x^*\|_2)}{\|x^k - x^*\|_2} \right], \tag{18}$$

which, together (16), implies $\lambda_i^* h_i(x^k) > 0$ for all k large enough whenever $i \in I_{\neq}$. 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 1 is also verified.

Now, let us prove that $\{x^k\}$ verifies (15). Dividing (17) by $\|d^k\|_2$ and take 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 (18). Thus,

$r_k^l = |h_l(x^k)|/\|x^k - x^*\|_2 \to 0$. Notice that by (16) we have $m(d) = \lim_{k\in K} m(d^k)/\|d^k\|_2 > 0$, so by (18)

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

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 have proved that (15) is valid for all h_l, $l \notin I_{\neq}$. Similarly we can conclude that (15) 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. □

Example 2 suggests that $\operatorname{cl} M_I(x^*) = M_Q(x^*)$ for all KKT point x^*, where "cl" denotes the closure of a set. We have $\operatorname{cl} M_I(x^*) \subset M_Q(x^*)$ since $M_I(x^*) \subset M_Q(x^*)$ and $M_Q(x^*)$ is closed by Theorem 4. The contrary inclusion is more delicate, but we conjecture that this is always true. The equality $\operatorname{cl} M_I(x^*) = M_Q(x^*)$, if it holds in general, leads to some interesting conclusions: (i) that quasinormal multipliers in some sense give the same sensitivity measure as the informative ones; (ii) as every multiplier vector with minimum 2-norm is informative (Theorem 5), quasinormal multipliers are as good as informative ones to estimate those with minimum 2-norm; and (iii) that multiplier vectors with minimum number of non-zero entries are arbitrary close to $M_I(x^*)$ (these type of multipliers were considered in [15], and are quasinormal by [15, Proposition 2.2(b)] and Remark 1).

4.4 Abstract constraints

In this section we deal with the problem (P) with additional abstract constraints

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

where X is a non-empty closed subset of $\mathbb{R}^n$. FJ/KKT enhanced stationarity concepts were developed to the present case. In order to discuss them, we denote the normal cone of X at x by $N_X(x)$. Specifically, Bertsekas [14] provides a version of Theorem 1 for the problem (P_X) which consists of replacing item 1 by

$$-\left[\sigma\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*)\right] \in N_X(x^*)$$

and taking the sequence $\{x^k\}$ of item 3 feasible with respect to X, that is, $x^k \in X$ for all k. The quasi-normality CQ is adapted accordingly (see [15]). For the sake of clarification, we present next the adaptations of Definitions 1 and 3 to this case.

Definition 4. *We say that a feasible x^* for (P_X) is an E-KKT point if there are $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ such that*

1. $-\big[\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu\big] \in N_X(x^*)$;
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 as in Definition 1.

Definition 5. *We say that a feasible* x^* *for* $(\mathrm{P_X})$ *satisfies the* quasi-normality CQ *if there is no* $\lambda \in \mathbb{R}^m$ *and* $\mu \in \mathbb{R}^{|I_g(x^*)|}$, $\mu \geq 0$, *such that*

1. $-\big[\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j\in I_g(x^*)} \mu_j \nabla g_j(x^*)\big] \in N_X(x^*)$;
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 as in Definition 1.*

Similarly to that is done in the literature, we say that x^* is a KKT point when item 1 of Definition 4 and the usual complementarity slackness $\mu_j g_j(x^*) = 0$, $\forall j \in I_g(x^*)$, hold. The sets $M(x^*)$ and $M_Q(x^*)$ of usual and quasinormal multipliers are defined accordingly.

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

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

where $T_X(x)$ is the tangent cone of X at x (see [14]) and C° denotes the polar set of C. It worth be mentioned that if X is convex then X is regular at all $x \in X$. In particular, $X = \mathbb{R}^n$ and $X = \{x \mid \ell \leq x \leq u\}$ are regular at all their feasible points. The last type of constraints are common to ensure the well-definiteness of several algorithms and they are used in popular computational optimization packages such as Algencan [3] (in the next section we will present this method without the box-constraints).

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

4.4.1 Non-emptiness and closedness of $M_Q(x^*)$ at KKT points

As mentioned in Section 4.3, the same arguments for proving Proposition 2.2 of [15] still valid for any KKT point x^*. So, in the case of $T_X(x^*)$ is convex, Theorem 5 (and consequently Theorem 2) is valid when we deal with the abstract set constraints X through its normal cone as in Definition 4. Note that $T_X(x^*)$ is convex if X is regular at x^*, which in turn is true if X is convex [15, 35]. In particular, the first statement of Theorem 4 ($M_Q(x^*) \neq \emptyset$) is valid in this case. The closedness of $M_Q(x^*)$ can be proved in an analogous way to that done in Theorem 4 because $N_X(x^*)$ is closed.

When $T_X(x^*)$ is not convex, the presence of abstract constraints changes drastically the relations between usual and quasinormal multipliers. In particular, Theorem 2 is not valid in this case (see [15, Example 2.2]).

4.4.2 Boundedness of $M_Q(x^*)$ under QN

In the proof of Theorem 4, we claim that the boundedness of $M_Q(x^*)$ under quasi-normality (without abstract constraints) follows from [38, Theorem 3] and Remark 1. The cited theorem from [38] refers to multipliers satisfying only items 1 and 2(a) of Definition 4, and remains valid in the presence of $x \in X$ even if X is not regular at x^*. As the set of these multipliers are possibly larger than $M_Q(x^*)$, the set $M_Q(x^*)$ is bounded under QN in the sense of Definition 6.

4.4.3 Relationship between $M(x^*)$ and $M_Q(x^*)$ under MFCQ

In the next, we show that Theorem 3 is no longer valid in the presence of abstract constraints even if X is regular at x^*. In [15], an general extension of MFCQ (Definition 2) was defined and specialized cases are discussed. The most natural of them is the following:

Definition 6. *We say that a feasible* x^* *for* $(\mathrm{P_X})$ *satisfies MFCQ if there is no nonzero* $\lambda \in \mathbb{R}^m$ *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 MFCQ reduces to that of Definition 2 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^* and satisfies MFCQ in the sense of Definition 6.

Example 4. *Consider the problem*

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

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 have* $\nabla h(x^*)^t d = 0$ *and* $\nabla g(x^*)^t d = -1 < 0$*. Thus, MFCQ holds at* x^**. Note that*

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

and so x^* *is KKT in the usual sense with any 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* $M_Q(x^*) = \{(-1, 0)\} \neq \{-1\} \times [0, 1] = M(x^*)$*.*

5 On augmented Lagrangian methods

An interesting practical consequence of QN is related to the dual sequences generated by the safeguarded (Powell-Hestenes-Rockafellar – PHR) augmented Lagrangian method defined in [3], namely Algencan, which we recall in Algorithm 1. This method consists in successively minimizing the PHR augmented Lagrangian function associated to (P),

$$L_\rho(x, \bar{\lambda}, \bar{\mu}) = f(x) + \frac{\rho}{2}\left[\left\|h(x) + \frac{\bar{\lambda}}{\rho}\right\|_2^2 + \left\|\left(g(x) + \frac{\bar{\mu}}{\rho}\right)_+\right\|_2^2\right], \tag{19}$$

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

Algorithm 1 Algencan

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

Algencan is closed related to the pure external penalty method. In fact, it reduces to the second if we choose $\bar{\lambda}^{k+1} = 0$ and $\bar{\mu}^{k+1} = 0$ in Step 3. The safeguarding on $\bar{\lambda}$ and $\bar{\mu}$, that is, the boundedness of $\{(\bar{\lambda}^k, \bar{\mu}^k)\}$, plays a crucial role: on the one hand, it recovers the strong theoretical properties of the external penalty method [6]; on the other hand, the shifted penalty (19) tends to mitigate the grown of ρ, making the algorithm much more stable and reliable in practice. Under some hypotheses, $\{\rho_k\}$ can even be bounded. Of course, this is the ideal scenario for the stability of the numerical method. Thus, establishing sufficient conditions to bound $\{\rho_k\}$ is of great interest.

When applied to (P), Algencan generates the multipliers estimates

$$\lambda^k = \bar{\lambda}^k + \rho_k h(x^k), \qquad \mu^k = (\bar{\mu}^k + \rho_k g(x^k))_+ \tag{20}$$

at each iteration, provided by the subproblem of Step 1. In [4], the global convergence of ALGENCAN was improved through the sequential optimality condition positive approximate KKT (PAKKT), which was derived from enhanced FJ conditions. A consequence of PAKKT is that the multiplier sequences (20) are bounded if QN holds at the feasible limit point [4, Corollary 4.8]. This fact was crucial to attest convergence of ALGENCAN with scaled stopping criteria defined in [7] (SCALED-ALGENCAN) to KKT points, for which QN is the weakest CQ fulfilling this property. It is worth mentioning that SCALED-ALGENCAN has a superior numerical performance and so can be considered the "standard" version of ALGENCAN for general purpose optimization. However, this boundedness of the sequences (20) does not imply the boundedness of the penalty parameter sequence $\{\rho_k\}$. In [3], the boundedness of $\{\rho_k\}$ was established under additional strong hypothesis such as second-order sufficiency and LICQ, which implies the uniqueness of usual multipliers. Later on [17], LICQ was slightly relaxed to the following assumption:

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

From the relation between MFCQ and the compactness of $M(x^*)$ [24], we conclude that MFCQ is redundant in the above statement. For the sake of completeness, we provide below a simple proof of this fact.

Theorem 6. *Let x^* be a KKT point for* (P) *and suppose that there is only one associated multiplier vector (λ^*, μ^*). Then MFCQ holds at x^*.*

Proof. Instead of Definition 2, here we use the dual form of the MFCQ which states that MFCQ holds at x^* if, and only if, $(\lambda, \mu) = 0$ whenever $\nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$ and $\mu \geq 0$. So if MFCQ is not valid, there is $(\bar{\lambda}, \bar{\mu}) \neq 0$ such that $\nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$, $\mu \geq 0$. Then $(\bar{\lambda} + \lambda^*, \bar{\mu} + \mu^*) \neq (\lambda^*, \mu^*)$ is a multiplier vector associated with x^*. □

In view of the above theorem, the sufficient hypotheses in [17] for the boundedness of $\{\rho_k\}$ in Algorithm 1 can be stated as follows:

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

A2. there is only one multiplier vector (λ^*, μ^*) 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 tolerances ε_k for the subproblems are driven to zero fast enough in the sense of

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

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

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

Assumption A4 can be fulfilled in practice since ε_{k+1} are computed after (x^k, λ^k, μ^k). For large safeguards $\mu_{\max} > 0$, $\lambda_{\min} < 0$ and $\lambda_{\max} > 0$, we can expect assumption A5 holds under QN and A1, since the entire sequence $\{(\lambda^k, \mu^k)\}$ is bounded in this case [4]. It worth be mentioned that in practice the projected multipliers $\bar{\lambda}_i^{k+1}$ and $\bar{\mu}_j^{k+1}$ are chosen as the projection of λ_i^k and μ_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 fall within the safeguard intervals. This somehow recovers the classical augmented Lagrangian method without safeguards, that has favourable properties on convex problems [34].

In [4], it has been shown that the sequences $\{x^k\}$ and $\{(\lambda^k, \mu^k)\}$ (see (20)) generated by Algorithm 1 have a weak control of signs similar to condition 2(a) in Definition 1: it was proved that $\lambda_i^k h_i(x^k) > 0$ for all k large enough and the indexes i associated with sequences $\{|\lambda_i^k|\}$ that grow faster, i.e., $\lim_k |\lambda_i^k|/\|(1, \lambda^k, \mu^k)\|_\infty > 0$ (similarly for μ_j^k). This property is inherited from the pure external penalty that controls all the signs as in Definition 1. A question that arises is whether the dual accumulation points generated by Algorithm 1 are quasinormal multipliers. Unfortunately, the answer is negative in general due to the projected multipliers. For instance, it is easy to see that Algorithm 1 can get stuck if we initialize it at a KKT point $x^0 = x^*$ with an associated non-quasinormal multiplier vector $(\bar{\lambda}^0, \bar{\mu}^0) = (\lambda^*, \mu^*)$. In this case, however, $\rho_k = \rho_0$ for all k. The question remains open in the case of $\rho_k \to \infty$. If Algencan generates only quasinormal multipliers when $\rho_k \to \infty$ (possibly under additional suitable assumptions as A1 and A5), the boundedness of $\{\rho_k\}$ could be established changing A2 to the less stringent hypothesis

A7. there is only one quasinormal multiplier vector (λ^*, μ^*) associated with x^*

using the same arguments of [17], and noting that when $M_Q(x^*)$ is a singleton, thus compact, then QN is valid at x^* by Theorem 4. Note that the uniqueness of quasinormal multipliers occurs more often than the uniqueness of all multipliers, see the discussion after Example 2.

Aiming to answer if Algencan generates quasinormal multipliers under QN when $\rho_k \to \infty$, we ran its implementation available by the TANGO project (`https://www.ime.usp.br/~egbirgin/tango/codes.php`) on the problem

$$\min\ x_1^2 - 2x_2 \quad \text{subject to} \quad x_2 e^{x_1} \leq 0, \quad -x_2 \leq 0. \tag{21}$$

This problem has unique minimizer $x^* = (0, 0)$, which satisfies QN but not MFCQ (see [11]). We have $M(x^*) = \{(t, t-2) \mid t \geq 2\}$ and $M_Q(x^*) = \{(2, 0)\}$. We start with $\bar{\mu}^1 = (0, 0)$ and $\rho_1 = 10$. Condition "$\rho_k \to \infty$" was forced by modifying the code properly to multiply ρ_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 ρ was 10^{28}). The problem was solved achieving the final multiplier $\mu^* \approx (2.02109, 0.02109)$. This suggests that the method may converges to a non-quasinormal multiplier vector. Unfortunately, we can not be able to reproduce this behavior analytically, so from the theoretical point of view the question remains open.

From our numerical experience with Algencan on CUTEst problems while writing [7], the algorithm very often maintain dual sequences with moderate norm. This is in line with the generality of QN [4, Corollary 4.8]. On the other hand, it is reasonable to choose projected multipliers as (20) whenever possible since these are the estimates provided directly from the first-order optimality conditions of the subproblem in Step 1. In this sense, assumption A5 is realistic if we choose large safeguard intervals $[\lambda_{\min}, \lambda_{\max}]$ and $[0, \mu_{\max}]$; in practice, $\lambda_{\min} = -10^{20}$ and $\lambda_{\max} = \mu_{\max} = 10^{20}$ are adopted as default in Algencan.

To characterize a multiplier vector as quasinormal, we should exhibit a sequence $\{x^k\}$ in the spirit of item 3 of Definition 1. Such a sequence can be the one generated by Algorithm 1 itself although this is not mandatory. Considering the sequence $\{x^k\}$ generate by the algorithm, it seems difficult to predict the sign of $\lambda_i^k h_i(x^k)$ and $\mu_j^k g_j(x^k)$ at iteration k. To simplify, consider only equality constraints. Given the initial $\bar{\lambda}^1$ we have at iteration k and take into account (20) and A5 with $k_0 = 1$,

$$\lambda_i^k h_i(x^k) = \bar{\lambda}_i^k h_i(x^k) + \rho_k h_i(x^k) h_i(x^k) = \bar{\lambda}_i^1 h_i(x^k) + \sum_{l=1}^{k} \rho_l h_i(x^l) h_i(x^k).$$

So $\lambda_i^k h_i(x^k)$ depends on all previous x^i and ρ_i, which have an intricate relationship with each other. Also x^k depends on $\bar{\lambda}^k$, which makes the sign of $\bar{\lambda}_i^k h_i(x^k)$ difficult to analyse whatever $\bar{\lambda}_i^k \neq 0$ is chosen at the iteration $k-1$. Trying to deepen the knowledge about the numerical behaviour of Algencan, we ran it on (21) with the same previous adaptations but now taking $\bar{\mu}_j^{k+1} = 0$ whenever $\mu_j^k g_j(x^k) < 0$ in Step 3. The problem was solved with last $\rho_{26} = 10^{26}$, and $\mu^* \approx (2, 10^{-15})$, the unique quasinormal multiplier. It is worth mentioning that we are not arguing that this simple modification leads to a better algorithm; on the contrary, we believe it to be worse on average. Our intention is to contribute to understand how dual iterates of Algorithm 1 behave. We believe that adjusting the dual sequence properly can lead to improved performance of Algencan in problems where it struggles to reach optimality tolerance, 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) is a class of problems widely studied in the literature. Due to its high degeneracy compared to a standard NLP, they deserve special treatment (see [10] and references therein). Its general formulation is

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

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

It is well known that the KKT conditions cannot be expected at local minimizers x^* of (MPCC) if we do not have (lower level) strict complementarity [36], i.e., if $G_i(x^*) = H_i(x^*) = 0$ for some i. Thus, other weak stationarity concepts are defined in the literature. A stringent and popular of 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 7. *We say that a feasible x^* for* (MPCC) *is an* M-stationary *point if there are multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p_+$, 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. $\mu_j = 0\ \forall j \notin I_g(x^*), \quad \gamma_i^G = 0\ \forall i \in I_{+0}(x^*), \quad \gamma_i^H = 0\ \forall i \in I_{0+}(x^*)$

and

$$\text{either } \gamma_i^G\gamma_i^H = 0 \text{ or } \gamma_i^G > 0, \gamma_i^H > 0\ \forall i \in I_{00}(x^*). \tag{22}$$

In [29], enhanced FJ-type conditions linked with M-stationarity were proposed for (MPCC) with smooth data, in the spirit of items 1, 2 and 3(a,b) of Theorem 1. As in standard NLP, enhanced KKT-type conditions are derived by taking the multiplier associated with the objective equal to one, as done from Theorem 1 to Definition 1. So, inspired in [29, Theorem 3.1], we define the counterpart of Definition 1 to the MPCC context.

Definition 8. *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 (22));*
3. *if $I_{\neq} \cup J_+ \cup V_{\neq}^G \cup V_{\neq}^H \neq \emptyset$, where $I_{\neq} = \{i \mid \lambda_i \neq 0\}$, $J_+ = \{j \mid \mu_j > 0\}$, $V_{\neq}^G = \{i \mid \gamma_i^G \neq 0\}$ and $V_{\neq}^H = \{i \mid \gamma_i^H \neq 0\}$, then there is a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to x^* such that, for all k, the following conditions are valid:*
   - *(a)* $\lambda_i h_i(x^k) > 0,\ \forall i \in I_{\neq}, \quad \mu_j g_j(x^k) > 0,\ \forall j \in J_+, \quad \gamma_i^G G_i(x^k) < 0,\ \forall i \in V_{\neq}^G,$ $\gamma_i^H H_i(x^k) < 0,\ \forall i \in V_{\neq}^H$;
   - *(b)* $f(x^k) < f(x^*)$;
   - *(c)* $|h_i(x^k)| = O(w(x^k)),\ \forall i \notin I_{\neq}, \quad g_j(x^k)_+ = O(w(x^k)),\ \forall j \notin J_+, \quad (-G_i(x^k))_+ = O(w(x^k)),\ \forall i \notin V_{\neq}^G$ *and* $(-H_i(x^k))_+ = O(w(x^k)),\ \forall i \notin V_{\neq}^H$, *where*

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

Note that item 3(a) of the above definition implies item 2 of Definition 7. It is worth mentioning that enhanced stationarity established in [29] was extended to non-smooth MPCCs in [39], but we do not treat this case since we deal only with smooth problems in this work.

A question that immediately arises is whether a local minimizer of (MPCC) is EM-stationary. In [29], some MPCC-tailored CQs are provided so that every local minimizer satisfies items 1, 2 and 3(a,b) of Definition 8. The validity of item 3(c) follows from a reasoning similar to that of the final part of the proof of Theorem 2 (see also Remark 1).

In this section, we do not focus on MPCC-CQs, but on the equivalence between enhanced stationarity from Definition 8 and the usual M-stationarity. Given a feasible point x^* for (MPCC), let us consider the *tightened nonlinear problem*

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

which consists in fixing as equalities the active complementary constraints at x^*. It is easy to verify that M-stationarity is exactly the KKT conditions for (TNLP(x^*)) (the so-called *weakly stationarity*, or *W-stationarity*) together (22).

Theorem 7. *Every M-stationary point x^* is EM-stationary and vice-versa.*

Proof. Applying Theorem 2 to (TNLP(x^*)) we obtain the aforementioned equivalence except for (22), possibly with different M-multipliers. But in the proof of Theorem 2, the multipliers (in this case, multipliers for (TNLP(x^*))) are adjusted using Lemma 2, which preserves the signs of non-null ones; see the discussion after Theorem 2. So (22) is still valid when we redefine the M-multipliers to satisfy Definition 8. □

Given a feasible x^* we denote the set of *M-multipliers* $(\lambda, \mu, \gamma^G, \gamma^H)$ satisfying Definition 7 by $M^{\text{MPCC}}(x^*)$, and those satisfying the conditions in Definition 8 by $M_Q^{\text{MPCC}}(x^*)$. We can refer to M-multipliers in $M_Q^{\text{MPCC}}(x^*)$ by *quasinormal M-multipliers* due to its relation with an MPCC's version of quasi-normality [29].

Most of standard CQs are not valid for (MPCC), so this problem deserves MPCC-tailored CQs. A usual way to define such MPCC-CQs are imposing usual CQs on (TNLP(x^*)), possibly together (22). See [10] and references therein. In [36], a Mangasarian-Fromovitz-type CQ (MPCC-MFCQ) was defined as MFCQ on (TNLP(x^*)). In [29], it was generalized by adding (22). We refer to the last definition as *generalized MPCC-MFCQ*. Still in [29], a *generalized MPCC-quasi-normality CQ* was defined in the same way. We enunciate them below.

Definition 9. *Let x^* be a feasible point for* (MPCC). *We say that*

1. *x^* satisfies the* generalized MPCC-MFCQ *if it conforms to MFCQ for* (TNLP(x^*)) *and* (22) *holds;*

2. *x^* satisfies the* generalized MPCC-quasi-normality CQ *if there is no non-null M-multiplier such that items 1, 2 and 3 of Definition 8 hold.*

Theorem 8. *Let x^* be M-stationary.*

1. *If generalized MPCC-MFCQ holds at x^* then $M^{MPCC}(x^*) = M_Q^{MPCC}(x^*)$;*

2. *$M_Q^{MPCC}(x^*) \neq \emptyset$ and closed;*

3. *$M_Q^{MPCC}(x^*)$ is compact if, and only if, generalized MPCC-quasi-normality holds at x^*.*

Proof. Item 1 is a consequence of Theorem 3 applied on (TNLP(x^*)), since (22) does not interfere in the analysis. Item 2 follows from Theorem 7 and the fact that $\{(\gamma^G, \gamma^H) \in \mathbb{R}^{2q} \mid (22)\}$ is closed; thus, the reasoning of the first part of the proof of Theorem 4 is still valid.

Let us prove the third item. By Theorem 4 the set $M_Q(x^*)$ of multipliers associated with (TNLP(x^*)) is bounded under generalized MPCC-quasi-normality CQ, since this MPCC-CQ implies the standard QN on (TNLP(x^*)). As $M_Q^{\text{MPCC}}(x^*)$ is contained in $M_Q(x^*)$, it is bounded too. For the converse, we can use the same arguments as in the proof of Theorem 4 since Lemma 2 preserves the sign of non-null multipliers. □

There are other stationarity concepts for (MPCC) besides M-stationarity in the literature. The widely use ones are *strongly*, *weakly* and *Clarke* stationarity (S/W/C-stationarity, respectively). As in Definition 8, we can define their enhanced versions by adding item 3, although to the best of our knowledge they have not been defined in the literature. W-stationarity is just the KKT conditions for (TNLP(x^*)). So all the results of Sections 3 and 4 are valid, considering that MPCC-tailored CQs, e.g. MPCC-MFCQ and MPCC-quasi-normality, are defined just imposing standard CQs on (TNLP(x^*)). S-stationarity is just KKT for (MPCC) viewed as an NLP. Thus, Theorem 2 can be applied to attest the equivalence between S-stationarity and its enhanced version. However, no feasible point of (MPCC) satisfies MFCQ [22], while QN only be expected at points where the strict complementarity holds [4, Lemma 4.4]. Therefore, Theorems 3 and 4 cannot hold in general for S-stationarity. C-stationarity is defined as in Definition 7 but weakening (22) to [36]

$$\gamma_i^G \gamma_i^H \geq 0, \quad \forall i \in I_{00}(x^*). \tag{23}$$

We can define the associated MPCC-MFCQ and MPCC-quasi-normality accordingly by imposing the correspondent CQs on (TNLP(x^*)) together (23). All we made for M- remains valid for C-stationarity under straightforward adaptations.

We end this section with comments on a problem related to MPCC, 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 made for MPCCs, enhanced FJ-type conditions were defined for MPVCs in [30]. By comparing the works [29] and [30], all the above discussion is applicable to MPVCs making straightforward adaptations.

6.2 Multi-objective optimization

Multi-objective problems deal with the minimization of multiple objectives functions simultaneously. The general problem treated here is

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

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, q$, are continuously differentiable. 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 *weakly efficiency* (or *weakly Pareto* optimality). A feasible point x^* is said to be *weak 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

$$\min_x \sum_{i=1}^{q} \sigma_i f_i(x) \quad \text{subject to} \quad h(x) = 0, \quad g(x) \leq 0, \tag{24}$$

is a weak 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, (24) is commonly used to address (MOP), including for deriving KKT-type stationarity. In [26], enhanced Fritz-John conditions for (MOP) are established just changing item 1 of Theorem 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 $\sigma \geq 0$, $\sum_i \sigma_i = 1$. The other conditions are exactly the same, and thus it is straightforward to adapt all we made for (P) using the following enhanced stationarity (note that the CQs do not depend on objectives, so they are the standard ones):

Definition 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. *item 3 of Definition 1 holds.*

7 Conclusions

The KKT conditions probably are the most important concept in constrained optimization, supporting stopping criteria and convergence analysis of almost every known method. Since the seminal work of Fritz John [28], the related FJ optimality conditions were improved by adding extra conditions verified by minimizers [13, 14, 15, 27, 38]. These enhanced concepts, particularly the E-KKT points as Definition 1, carry an additional sequential condition that connects the signs of non-null multipliers with a primal infeasible sequence. In this paper we deepen the study on enhanced FJ/KKT conditions for smooth nonlinear optimization. We show that such extra conditions do not strengthen the set of KKT points, that is, every E-KKT point is KKT and vice-versa. So, their effect concern only in the description of the Lagrange multipliers, that are called quasinormal in the sense of Definition 1. Our result differs from previous ones [13, 15] that only tackled this equivalence for minimizers. In this context, we analyse the shape of the quasinormal multipliers set such as its compactness and convexity. We also apply our theory to mathematical programs with complementarity constraints and multi-objective optimization, extending and clarifying previous results from the literature.

Since several methods generate a pair of primal-dual solution to the KKT system, the fact that every KKT point satisfies an enhanced stationarity draws attention to study/modify algorithms that/to generate quasinormal multipliers. These multipliers carry information about the sensitivity of the constraints and are more likely to be those with minimal 2-norm (see Section 4.3); therefore, they can be preferable in applications. In this sense, we present some insights about the multipliers generated by the augmented Lagrangian method ALGENCAN defined in [3, 18]. Its global convergence was studied in [4] through the positive approximate KKT (PAKKT) condition. The control of signs in enhanced FJ conditions (item 2(a) of Definition 1) was fundamental to define PAKKT, which later inspired the use of scale stopping criteria in ALGENCAN [7] with great numerical success. An important consequence of PAKKT to the numerical stability of the method is that its dual sequences are bounded if the quasi-normality CQ holds at the primal limit point. Quasi-normality is a very general CQ proposed by Hestenes [27]; it is a tailor-made condition to attest KKT optimality from enhanced FJ conditions (i.e., it ensures the existence of a non-null multiplier for the objective). If at the beginning even the relationship of quasi-normality with other CQs was not a central issue, nowadays PAKKT gives it a status of CQ linked to practical algorithms. Regarding ALGENCAN, the validity of quasi-normality can be used to improve the existing results on the boundedness of its penalty parameter [3, 17] if we can modify the method to generate only quasinormal multipliers. See Section 5. Another interesting topic to be explored is the effect of quasinormal multipliers on second-order stationarity. In fact, we can define them using only such multipliers. By narrowing the multipliers we must test the positivity of the Hessian of the Lagrangian, we potentially obtain more affordable conditions to algorithms. For instance, quasinormal multipliers are more often unique than usual ones.

References

[1] J. Abadie. On the Kuhn-Tucker theorem. In J. Abadie, editor, *Nonlinear Programming*, pages 21–36. John Wiley, 1967.

[2] W. Achtziger and C. Kanzow. Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. *Mathematical Programming*, 114(1):69–99, 2008.

[3] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2007.

[4] R. Andreani, N. Fazzio, M. Schuverdt, and L. Secchin. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM Journal on Optimization*, 29(1):743–766, 2019.

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

[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. *Numerical Algorithms*, 90(2):851–877, 2022.

[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. *Mathematical Programming Computation*, 14(1):121–146, 2022.

[8] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1):255–273, 2012.

[9] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012.

[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 Journal on Optimization*, 29(4):3201–3230, 2019.

[11] R. Andreani, J. M. Martinez, 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):473–483, 2005.

[12] R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010.

[13] D. P. Bertsekas. *Nonlinear Programming: 2nd Edition*. 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. *Journal of Optimization Theory and Applications*, 114(2):287–343, 2002.

[16] D. P. Bertsekas, A. E. Ozdaglar, and P. Tseng. Enhanced Fritz John conditions for convex programming. *SIAM Journal on Optimization*, 16(3):766–797, 2006.

[17] E. G. Birgin, D. Fernández, and J. M. Martínez. The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. *Optimization Methods and Software*, 27(6):1001–1024, 2012.

[18] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[19] C. Büskens and D. Wassel. *The ESA NLP Solver WORHP*, pages 85–110. Springer New York, New York, NY, 2013.

[20] A. R. Conn, G. I. M. Gould, and P. L. Toint. *Lancelot: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*, volume 17 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, Heidelberg, 1992.

[21] M. Ehrgott. *Multicriteria Optimization*. Springer-Verlag, 2 edition, 2005.

[22] M. L. Flegel and C. Kanzow. On the Guignard constraint qualification for mathematical programs with equilibrium constraints. *Optimization*, 54(6):517–534, 2005.

[23] R. Fletcher and S. Leyffer. User manual for filterSQP. Technical Report NA/181, University of Dundee, 1998.

[24] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12(1):136–138, 1977.

[25] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47(1):99–131, 2005.

[26] G. Giorgi, B. Jiménez, and V. Novo. Approximate karush–kuhn–tucker condition in multiobjective optimization. *Journal of Optimization Theory and Applications*, 171(1):70–89, 2016.

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

[28] F. John. Extremum problems with inequalities as subsidiary conditions. In *Traces and Emergence of Nonlinear Programming*, pages 197–215. Springer Basel, jul 2013.

[29] C. Kanzow and A. Schwartz. Mathematical programs with equilibrium constraints: enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results. *SIAM Journal on Optimization*, 20(5):2730–2753, 2010.

[30] A. Khare and T. Nath. Enhanced Fritz John stationarity, new constraint qualifications and local error bound for mathematical programs with vanishing constraints. *Journal of Mathematical Analysis and Applications*, 472(1):1042–1077, 2019.

[31] L. Minchenko. Note on Mangasarian–Fromovitz-like constraint qualifications. *Journal of Optimization Theory and Applications*, 182(1):1199–1204, 2019.

[32] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011.

[33] A. E. Ozdaglar and D. P. Bertsekas. The relation between pseudonormality and quasiregularity in constrained optimization. *Optimization Methods and Software*, 19(5):493–506, 2004.

[34] R. Rockafellar. The multiplier method of Hestenes and Powell applied to convex programming. *Journal of optimization Theory and Applications*, 12(6):555–562, 1973.

[35] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1 edition, 1998.

[36] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[37] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[38] J. J. Ye and J. Zhang. Enhanced Karush-Kuhn-Tucker condition and weaker constraint qualifications. *Mathematical Programming*, 139(1):353–381, 2013.

[39] J. J. Ye and J. Zhang. Enhanced Karush-Kuhn-Tucker conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 163(3):777–794, 2014.