A practical second-order optimality condition for cardinality-constrained problems with application to an augmented Lagrangian method*

Jean C. A. Medeiros† Ademir A. Ribeiro‡ Mael Sachine‡

Leonardo D. Secchin§

April 19, 2022 (Updated March 20, 2023)

Abstract

This paper addresses the *mathematical programs with cardinality constraints* (MPCaC). We first define two new tailored (*strong* and *weak*) *second-order necessary conditions*, MPCaC-SSONC and MPCaC-WSONC. We then propose a constraint qualification (CQ), namely, *MPCaC-relaxed constant rank constraint qualification* (MPCaC-RCRCQ), and establish the validity of MPCaC-SSONC at minimizers under this new CQ. All the concepts proposed here are based on the so-called M-stationarity, which is a suitable first-order stationarity for MPCaC. Furthermore, they are defined using only original variables, without the help of auxiliary variables of an augmented problem commonly considered in this context. This makes the proposed second-order stationarity concepts suitable for MPCaC. We illustrate the applicability of MPCaC-WSONC to derive global convergence for a second-order augmented Lagrangian algorithm on MPCaCs under MPCaC-RCRCQ. The relationship between the tailored MPCaC-WSONC and the standard WSONC (applied to a reformulated problem) are treated, showing that MPCaC-WSONC is a strong condition to study global convergence of algorithms in the MPCaC context.

1 Introduction

In this work, we deal with *mathematical programs with cardinality constraints* (MPCaC) of the form

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad g(x) \leq 0, \quad h(x) = 0, \quad \|x\|_0 \leq \alpha, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions, $0 < \alpha < n$ is a given natural number and $\|x\|_0$ is the cardinality of the vector $x \in \mathbb{R}^n$, that is, the number of nonzero components of x. It should be mentioned that $\|\cdot\|_0$ is not a norm and the cardinality constraint is not continuous neither convex. This makes the above problem distinct, more degenerated and difficult than a standard nonlinear programming problem (NLP). It has applications in several areas such as portfolio optimization, compressing sense and subset selection in regression. See [15, 30] and references there in.

Although it is possible to completely reformulate (1) by a mixed-integer model, it is common in the literature dealing with (1) through the continuous NLP

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \quad h(x) = 0, \\ & n - e^T y \leq \alpha, \quad y \leq e, \quad x * y = 0, \end{array} \tag{2}$$

*The fourth author was financially supported by National Council for Scientific and Technological Development – CNPq (Project 309136/2021-0).

†State University of Campinas - SP, Brazil. jeancarlos@ime.unicamp.br

‡Federal University of Paraná - PR, Brazil. ademir.ribeiro@ufpr.br, mael@ufpr.br

§Department of Applied Mathematics, Federal University of Espírito Santo - ES, Brazil. leonardo.secchin@ufes.br (Corresponding author)

see for instance [16, 17, 21]. Note that the relaxed reformulation we use here is slightly different from that of [16, 17] since we do not impose $y \geq 0$, as done in [21]. This, however, does not interfere with the results. It is well known that we are able to recover global solutions of (1) through (2), while local minimizers of both problems are related only if $\|x\|_0 = \alpha$ [16]. Nevertheless, problem (2) is widely used because, unlike (1) and mixed-integer models, it is well solvable by standard optimization algorithms. In this sense, stationary points of (2) are considered good candidates for the solutions of (1). Inspired in the related *mathematical programs with complementarity constraints* (MPCC) [28], the authors of [16] propose tailored stationarity concepts, namely S- and M-stationarity. S-stationarity is equivalent to the KKT conditions for (2), and thus it can be considered the best stationary concept if we study (1) through (2). Note that it involves both x and y variables. In turn, similar to the MPCC context, M-stationarity can be viewed as the KKT conditions for the NLP model obtained directly from (1) by fixing all zero x_i at the target point as equality constraints (the *tightened nonlinear program*). Despite the apparent weakness of M-stationarity compared to S-stationarity, it was recently established that these two concepts are actually equivalent if we perform a simple adjustment on the components of y [21, Proposition 2.3]. So, M-stationary becomes the "standard" first order concept for MPCaC due to its simplicity (it does not involve the auxiliary variable y). See also [24] for a unified approach of S- and M-stationarity. It should be mentioned that, despite the MPCaC-machinery is inspired in the MPCC-context, MPCaC has a lower level of degeneracy which justifies a particular study. See [16].

Second-order optimality conditions were widely studied for standard NLP; they are present in several works and classical books on nonlinear programming. The most common of them are the *strong* and *weak second-order necessary condition* (SSONC and WSONC for shortly), which basically consist of the non-negativity of the Hessian of the Lagrangian over a critical cone. However, in the context of cardinality problems only the work [15] treats the subject, to the best of our knowledge. It is worth mentioning that their conditions are constructed with a special, MPCaC-tailored, linearization of the feasible set, which differs from the standard one for (2). A similar linearization is used for MPCCs [18]. In both contexts, such linearizations are necessary due to the degeneracy of cardinality/complementarity constraints.

In this paper we define two new tailored second-order necessary conditions for MPCaC that we refer by MPCaC-SSONC and MPCaC-WSONC. The main goal is that these conditions are based on M-stationarity, which, as we already mentioned, is the most stringent and preferable first-order stationarity concept. In particular, they are defined using only the original variable x. It should be mentioned that the second-order conditions in [15] require S-stationary points, and consequently the auxiliary variable y, to attest their validity at minimizers. In this sense, our theory is clearer and better fits the original problem (1). Furthermore, we prove the convergence of a well established practical second-order augmented Lagrangian (AL) algorithm [1] to MPCaC-WSONC points. To the best of our knowledge, it is the first time that such a result has been proved for MPCaC. This is in accordance with standard NLP, where WSONC points are considered those that can be reached by algorithms [11, 19]. Curiously, although the authors of [15] consider the convergence of regularization methods, they did not use their weak necessary condition for this purpose; they focus on second-order sufficiency as a hypothesis to derive stability and local convergence results. This is the same path adopted in the MPCC context, where such methods usually come from. See e.g. [29].

It is well known that M-stationarity is not necessarily valid at minimizers of cardinality problems; to this purpose, constraint qualifications (CQ) are needed. Evidently, the same is true for MPCaC-S/WSONC since they are defined over M-stationary points. In [15], an MPCaC-tailored constant rank CQ (which we call MPCaC-CRCQ) for second-order optimality was proposed. It consists of adapting the constant rank CQ (CRCQ) for NLP proposed in [20], which is known to be a constraint qualification associated with SSONC [3]. In [26], a generalization of CRCQ was proposed, called *relaxed constant rank CQ* (RCRCQ), which later was proved to be also a CQ for SSONC [5]. Based on this fact, we proposed in this paper the *MPCaC-relaxed constant rank CQ* (MPCaC-RCRCQ), and we prove that it is a constraint qualification associated with MPCaC-SSONC (consequently, also with MPCaC-WSONC and M-stationarity). Despite the similarities with NLP, MPCaC-RCRCQ and MPCaC-CRCQ can be quite different in the MPCaC context. In fact, we show that MPCaC-RCRCQ is *always* valid at the origin, whereas MPCaC-CRCQ does not. This is coherent with the fact that the origin, whenever feasible, is always M-stationary independently of any regularity [22].

The paper is organized as follows: in Sect. 2 we define our second-order necessary conditions MPCaC-SSONC and MPCaC-WSONC, and the constraint qualification MPCaC-RCRCQ; we also prove that under this CQ, minimizers of cardinality problems are M-stationary and satisfy MPCaC-W/SSONC. In Sect. 3 we discuss the application of the second-order AL algorithm, defined in [1], for solving problem (1) and prove its global convergence to MPCaC-WSONC points under the new MPCaC-RCRCQ. In Sect. 4 we present the relationship between MPCaC-WSONC and standard WSONC condition on the reformulated problem (2). On the one hand, we show that MPCaC-WSONC has the same strength as WSONC, so MPCaC-WSONC is a suitable condition for studying global convergence of algorithms. But, on the other hand, we exhibit an example where we lose the positivity of the Hessian of the Lagrangian in WSONC when $x_i = y_i = 0$, justifying why the tailored MPCaC-WSONC condition is adequate for the MPCaC context. Conclusions and future work possibilities are presented in Sect. 5.

Notation. Throughout this paper, the Hadamard product between two vectors $x, y \in \mathbb{R}^n$, that is, the vector obtained by the componentwise product of x and y, is denoted by $x * y$. Given $\delta > 0$ and $z \in \mathbb{R}^n$, z_+ is the vector of entries $\max\{0, z_i\}$, $\mathrm{diag}(z)$ is the $n \times n$ diagonal matrix with diagonal z and $B_\delta(z)$ denotes the open ball centered at z with radius δ. We also use the following sets of indices: $I_g(x) = \{i \mid g_i(x) = 0\}$, $I_g^+(\bar{x}) = \{j \in I_g(\bar{x}) \mid \mu_j > 0\}$ and $I_0(x) = \{i \mid x_i = 0\}$.

2 Second-order optimality condition for MPCaC problems

Consider the Lagrangian function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ associated with (1), given by

$$\mathcal{L}(x, \mu, \lambda, \gamma) = f(x) + \mu^T g(x) + \lambda^T h(x) + \gamma^T x.$$

We have $\nabla_x \mathcal{L}(x, \mu, \lambda, \gamma) = \nabla f(x) + \nabla g(x)\mu + \nabla h(x)\lambda + \gamma$ and

$$\nabla^2_{xx} \mathcal{L}(x, \mu, \lambda, \gamma) = \nabla^2 f(x) + \sum_{i=1}^{m} \mu_i \nabla^2 g_i(x) + \sum_{i=1}^{p} \lambda_i \nabla^2 h_i(x). \tag{3}$$

M-stationarity is defined as follows.

Definition 1. *Let $\bar{x}$ be a feasible point for* (1). *We say that $\bar{x}$ is* M-stationary *if there exists a vector $(\mu, \lambda, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$ such that*

$$\nabla_x \mathcal{L}(\bar{x}, \mu, \lambda, \gamma) = 0, \tag{4a}$$

$$\mu^T g(\bar{x}) = 0, \tag{4b}$$

$$\gamma * \bar{x} = 0. \tag{4c}$$

Local minimizers of (1) are M-stationary points under very mild MPCaC-tailored CQs. See for instance [22, Theorems 3.2 and 4.7].

It should be noticed that the M-stationarity concept previously introduced in the literature refers to the reformulated problem (2), while our Definition 1 is given for the original problem (1). In relation to the original variable x, these definitions are exactly the same as they do not depend on y. So referring to (1) or (2) is, in our opinion, a matter of choice. As (4) does not contain y, we prefer to refer to the original problem (1). On the other hand, the S-stationarity concept defined in the literature (see e.g. [15, 16, 17]) only makes sense for (2) since it is the KKT conditions for this problem. In the statements of our results we choose to say explicitly “KKT for (2)” instead of dealing with an additional stationarity definition.

If $\bar{x}$ is M-stationary, then taking $\bar{y}_i = 1$ for $i \in I_0(\bar{x})$ and $\bar{y}_i = 0$ otherwise, the point $(\bar{x}, \bar{y})$ is KKT for (2) [21, Proposition 2.3]. So, M-stationarity can be considered the strongest first-order stationarity concept for (1). Our aim is to derive a second-order (necessary) stationary concept for (1) from M-stationarity that does not use the auxiliary variable y. Also, it is desirable algorithms reach such points under suitable assumptions.

In standard nonlinear programming, second-order conditions request the positive semi-definiteness of the Hessian of the Lagrangian function over the set of non-ascending directions (the critical cone). Let us consider the feasible set of (1),

$$\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0,\ h(x) = 0,\ \|x\|_0 \leq \alpha\}.$$

The *tangent cone* to Ω at $\bar{x} \in \Omega$ is the set

$$\mathcal{T}(\bar{x}) = \big\{d \in \mathbb{R}^n \mid \exists (x^k) \subset \Omega, \{t_k\} \subset \mathbb{R}_+ \text{ with } t_k \to 0 \text{ and } (x^k - \bar{x})/t_k \to d\big\}.$$

Another cone associated with Ω is defined by means of the relaxed problem (2). Like for MPCCs, the main difficulty to linearize the feasible set relies on the constraint $x*y = 0$. When $\|x\|_0 < \alpha$, this expression has a combinatorial nature that describes different faces of the feasible set depending on the choice of y. In [15], it has been shown that the union of such faces relies on the following linearization[1]:

$$\mathcal{T}_{\text{lin}}^{BS}(\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_i(\bar{x})^T d \leq 0,\, i \in I_g(\bar{x}), \\ \nabla h(\bar{x})^T d = 0, \\ |I_0(\bar{x}) \cap I_0(d)| \geq n - \alpha \end{array} \right\}.$$

Note that this set does not depend on y, it encompasses y "implicitly". Also, $|I_0(\bar{x}) \cap I_0(d)| \geq n-\alpha$ captures all directions in the tangent space of every active face at $\bar{x}$, as showed in the next lemma.

Lemma 1. *Let $\bar{x}$ be a feasible point for (1). Then $\mathcal{T}(\bar{x}) \subset \mathcal{T}_{\text{lin}}^{BS}(\bar{x})$.*

Proof. Let $d \in \mathcal{T}(\bar{x})$ be arbitrary. Then, there exist sequences $\{x^k\} \subset \Omega$ and $\{t_k\} \subset \mathbb{R}_+$ such that $t_k \to 0$ and $(x^k - \bar{x})/t_k \to d$. The proof that $\nabla g_i(\bar{x})^T d \leq 0$ for $i \in I_g(\bar{x})$ and $\nabla h(\bar{x})^T d = 0$ is straightforward. Now, since $\|x^k\|_0 \leq \alpha$, the set $J_k = \{i \mid x_i^k = 0\}$ satisfies $|J_k| \geq n - \alpha$ for all $k \in \mathbb{N}$. Moreover, there are only finitely many possible choices of index sets J_k and hence at least one of these sets occurs infinitely often in the sequence, say $J_k = J$ for all $k \in \mathbb{N}' \subset \mathbb{N}$. Thus, $x_i^k = 0$ for all $i \in J$ and $k \in \mathbb{N}'$, which yields $\bar{x}_i = \lim_{k\in\mathbb{N}'} x_i^k = 0$ for all $i \in J$. This means that $J \subset I_0(\bar{x})$. To conclude the proof, note that $|J| \geq n - \alpha$ and $d_i = \lim_{k\in\mathbb{N}'}(x_i^k - \bar{x}_i)/t_k = 0$ for all $i \in J$. □ □

Remark 1. *As in NLP, we have $\mathcal{T}(\bar{x}) \not\supset \mathcal{T}_{\text{lin}}^{BS}(\bar{x})$ in general. For instance, consider the constraints $-x_1^3 + x_2 \leq 0$, $-x_2 \leq 0$, $\|x\|_0 \leq 1$, the point $\bar{x} = (0,0)$, and $d = (-1,0)$. Note that $d \in \mathbb{R} \times \{0\} = \mathcal{T}_{lin}^{BS}(\bar{x})$, but $d \notin \mathcal{T}(\bar{x})$. In fact, every sequence $\{x^k\}$ within Ω must satisfy $(x_1^k)^3 \geq x_2^k \geq 0$ and hence, $x_1^k/t_k \geq 0$, which means that $(x^k - \bar{x})/t_k \to d$ does not occur.*

In [15] the authors provide a second-order necessary optimality condition for critical directions obtained from $\mathcal{T}_{\text{lin}}^{BS}(\bar{x})$, assuming that the minimizer is S-stationary. However, as far as we are aware, this condition does not hold if one changes S- by M-stationarity. See Example 3 and its related discussion.

In order to establish a second-order optimality condition associated with M-stationarity, we propose the following *linearized cone:*

$$\mathcal{T}_{\text{lin}}(\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_i(\bar{x})^T d \leq 0,\, i \in I_g(\bar{x}), \\ \nabla h(\bar{x})^T d = 0, \\ d_i = 0,\, i \in I_0(\bar{x}) \end{array} \right\}.$$

It is clear that $\mathcal{T}_{\text{lin}}(\bar{x}) \subset \mathcal{T}_{\text{lin}}^{BS}(\bar{x})$, but the contrary inclusion is not true in general. Indeed, Example 1 shows that we do not have even $\mathcal{T}(\bar{x}) \subset \mathcal{T}_{\text{lin}}(\bar{x})$ in general. The only difference between these linearized sets relies on the null components of the directions: in $\mathcal{T}_{\text{lin}}(\bar{x})$, d_i must be null whenever $x_i = 0$, while in $\mathcal{T}_{\text{lin}}(\bar{x})$ some of components may be nonzero. See Figure 1.

Example 1. *Consider only the cardinality constraint $\|x\|_0 \leq 2$, where $x \in \mathbb{R}^3$, and the point $\bar{x} = (1,0,0)$. The feasible set Ω is the union of the three coordinate planes, and its tangent set $\mathcal{T}(\bar{x})$ at $\bar{x}$ is the union of the planes $x_2 = 0$ and $x_3 = 0$. But $\mathcal{T}_{lin}(\bar{x}) = \{(x_1, 0, 0) \mid x_1 \in \mathbb{R}\} \not\supset \mathcal{T}(\bar{x})$.* □

In the MPCaC literature, it is known that there is a one-to-one correspondence between local minimizers of problems (1) and (2) when $\|x\|_0 = \alpha$ [16]. In such a case, there is a slightly smaller level of degeneracy since an Abadie-type CQ holds [24]. Next we show that $\mathcal{T}_{\text{lin}}^{BS}(\bar{x})$ and $\mathcal{T}_{\text{lin}}(\bar{x})$ coincides if $\|x\|_0 = \alpha$. This encourages the use of $\mathcal{T}_{\text{lin}}(\bar{x})$ to derive a suitable second-order optimality.

[1]BS stands for Bucher and Schwartz, the authors of [15].

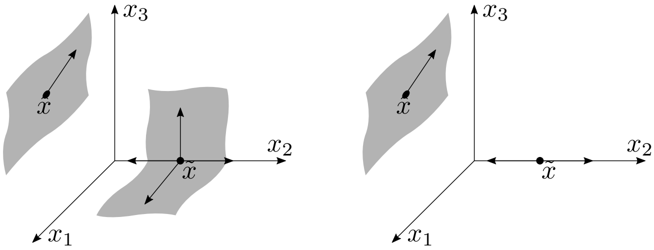

Figure 1: Representation of $\mathcal{T}_{\text{lin}}^{BS}(x)$ (left figure) and $\mathcal{T}_{\text{lin}}(x)$ (right figure) for $\|x\|_0 \leq \alpha = 2$, $x \in \mathbb{R}^3$. There are two points: $\hat{x}$ with $\hat{x}_1, \hat{x}_3 \neq 0$, $\hat{x}_2 = 0$; and $\tilde{x}$ with $\tilde{x}_2 \neq 0$, $\tilde{x}_1 = \tilde{x}_3 = 0$. At $\hat{x}$, $\|\hat{x}\|_0 = 1 = n - \alpha$, and both linearized sets are composed by directions d with $d_2 = 0$; thus they coincide (see Theorem 1). At $\tilde{x}$, the larger set $\mathcal{T}_{\text{lin}}^{BS}(x)$ encompasses directions with *at least* one component d_1 or d_3 zero, while $\mathcal{T}_{\text{lin}}(x)$ requires *both* d_1 and d_3 null.

Theorem 1. *Consider $\bar{x} \in \Omega$. If $\|\bar{x}\|_0 = \alpha$ then $\mathcal{T}_{\text{lin}}^{BS}(\bar{x}) = \mathcal{T}_{\text{lin}}(\bar{x})$.*

Proof. In fact, given $d \in \mathcal{T}_{\text{lin}}^{BS}(\bar{x})$ we have $|I_0(\bar{x})| = n - \|\bar{x}\|_0 = n - \alpha \leq |I_0(\bar{x}) \cap I_0(d)|$, which implies $I_0(\bar{x}) \subset I_0(d)$. In particular, $d_i = 0$ for all $i \in I_0(\bar{x})$, and thus $\mathcal{T}_{\text{lin}}^{BS}(\bar{x}) \subset \mathcal{T}_{\text{lin}}(\bar{x})$. The opposite inclusion is trivial. □ □

Using $\mathcal{T}_{\text{lin}}(\bar{x})$, we are able to define our strong second-order optimality necessary condition which does not depend on auxiliary variable y. This is in accordance with M-stationarity (Definition 1).

Definition 2. *Let $\bar{x}$ be an M-stationary point of* (1). *We say that $\bar{x}$ fulfills the* strong second-order optimality necessary condition *(MPCaC-SSONC) if there is a multiplier vector $(\mu, \lambda, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}^n$ associated with $\bar{x}$ such that*

$$d^T \nabla_{xx}^2 \mathcal{L}(\bar{x}, \mu, \lambda, \gamma) d \geq 0 \quad \textit{for all} \quad d \in \mathcal{C}^S(\bar{x}),$$

where $\mathcal{C}^S(\bar{x})$ is the strong MPCaC critical cone *defined as*

$$\mathcal{C}^S(\bar{x}) = \{d \in \mathcal{T}_{\text{lin}}(\bar{x}) \mid \nabla f(\bar{x})^T d \leq 0\}.$$

Remark 2. *As in standard NLP, if $\bar{x}$ is M-stationary we can rewrite $\mathcal{C}^S(\bar{x})$ without the use the gradient of the objective function but instead using the multipliers. To be precise, if $(\mu, \lambda, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}^n$ is the vector of the corresponding multipliers and $I_g^+(\bar{x}) = \{i \in I_g(\bar{x}) \mid \mu_i > 0\}$, then*

$$\mathcal{C}^S(\bar{x}) = \left\{d \in \mathcal{T}_{\text{lin}}(\bar{x}) \mid \nabla g_i(\bar{x})^T d = 0, \ i \in I_g^+(\bar{x})\right\}. \tag{5}$$

Indeed, given $d \in \mathcal{C}^S(\bar{x})$, using (4a) *and* (4c) *we obtain*

$$0 \leq -\sum_{i \in I_g(\bar{x})} \mu_i \nabla g_i(\bar{x})^T d = \nabla f(\bar{x})^T d \leq 0,$$

giving $\nabla g_i(\bar{x})^T d = 0$ for $i \in I_g^+(\bar{x})$. On the other hand, if d satisfies the restrictions in (5)*, using again* (4a) *and* (4c) *we conclude that*

$$\nabla f(\bar{x})^T d = -\sum_{i \in I_g(\bar{x})} \mu_i \nabla g_i(\bar{x})^T d = -\sum_{i \in I_g^+(\bar{x})} \mu_i \nabla g_i(\bar{x})^T d = 0,$$

implying $d \in \mathcal{C}^S(\bar{x})$.

It is well known from standard NLP that the strong second-order necessary condition cannot be expected at the limit points of practical algorithms [11, 19]. Instead, the adequate concept in this context is the weak second-order necessary condition, which consists of relaxing the requirements on ∇g in (5). We then derive an MPCaC-tailored *weak second-order optimality necessary condition*

(MPCaC-WSONC), which substitutes the strong critical cone $\mathcal{C}^S(\bar{x})$, in Definition 2, by a smaller set $\mathcal{C}^W(\bar{x}) \subset \mathcal{C}^S(\bar{x})$ called *weak critical cone* and defined as

$$\mathcal{C}^W(\bar{x}) = \left\{ d \in \mathcal{T}_{\text{lin}}(\bar{x}) \mid \nabla g_i(\bar{x})^T d = 0, \ i \in I_g(\bar{x}) \right\}. \tag{6}$$

We know that the first order concept, M-stationarity, is not necessarily valid at minimizers unless some constraint qualification holds. The same occurs with second-order conditions, which requests the discussion on constraint qualifications for cardinality problems. Like in the MPCC context, a usual way to define MPCaC-tailored CQs is to require a standard CQ on the *tightened nonlinear problem* (TNLP($\bar{x}$))

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & g(x) \le 0, \quad h(x) = 0, \\ & x_i = 0, \quad i \in I_0(\bar{x}). \end{array} \tag{7}$$

Using this strategy, several standard CQs were adapted for MPCaC, ensuring M-stationarity. Next, we recall two of them (see [17]). Also, note that M-stationarity for $\bar{x}$ is exactly the KKT conditions for TNLP($\bar{x}$), and $\mathcal{T}_{\text{lin}}(\bar{x})$ is just the standard linearized cone of (7).

Definition 3. *We say that a point $\bar{x} \in \Omega$ conforms to* MPCaC-linear independence constraint qualification *(MPCaC-LICQ) if the gradients*

$$\nabla g_i(\bar{x}), \ i \in I_g(\bar{x}), \quad \nabla h_j(\bar{x}), \ j = 1, \ldots, p, \quad e_i, \ i \in I_0(\bar{x})$$

are linearly independent.

Definition 4. *We say that* MPCaC-constant rank constraint qualification *(MPCaC-CRCQ) holds at a feasible point $\bar{x} \in \Omega$ if there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ such that for every $\mathcal{I} \subset I_g(\bar{x})$, $\mathcal{J} \subset \{1, \ldots, p\}$ and $\mathcal{I}_0 \subset I_0(\bar{x})$, the family of gradients*

$$\nabla g_i(x), \ i \in \mathcal{I}, \quad \nabla h_j(x), \ j \in \mathcal{J}, \quad e_i, \ i \in \mathcal{I}_0$$

has the same rank for every $x \in N(\bar{x})$.

It is worth mentioning that MPCaC-CRCQ was defined in [17], but, instead of established through the constant rank of the gradients as in Definition 4, it was stated saying that such gradients remain linearly dependent in a neighborhood of $\bar{x}$. These two definitions are equivalent, see [5]. Definition 4 is in line with how CRCQ was originally defined for NLP [20], and is better suited to introduce its relaxed version, as we do next.

In standard NLP, the corresponding CQs above attest second-order optimality at local minimizers. Another CQ with this property is the *relaxed constant rank CQ* (RCRCQ), defined in [26]. This condition relaxes the requirements in CRCQ over equality constraints. Next, we define its MPCaC counterpart by imposing RCRCQ condition on TNLP($\bar{x}$).

Definition 5. *We say that* MPCaC-relaxed constant rank constraint qualification *(MPCaC-RCRCQ) holds at a feasible point $\bar{x} \in \Omega$ if there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ such that for every $\mathcal{I} \subset I_g(\bar{x})$, the family of gradients*

$$\nabla g_i(x), \ i \in \mathcal{I}, \quad \nabla h_j(x), \ j \in \{1, \ldots, p\}, \quad e_i, \ i \in I_0(\bar{x})$$

has the same rank for every $x \in N(\bar{x})$.

MPCaC-RCRCQ is strictly implied by MPCaC-CRCQ even in the absence of equality constraints $h(x) = 0$, due to the difference related to the canonical vectors e_i's. The next example illustrates this with an interesting situation.

Example 2. *For any constraints $g(x) \le 0$, $h(x) = 0$, $\|x\|_0 \le \alpha$ for which $\bar{x} = 0$ is feasible, MPCaC-RCRCQ is valid at $\bar{x}$. This is because e_i, $i \in I_0(\bar{x})$, form the canonical basis of $\mathbb{R}^n$. On the other hand, MPCaC-CRCQ does not necessarily hold at $\bar{x}$, for example, $g_1(x_1, x_2) = x_1^2 + x_2 \le 0$ and $\|x\|_0 \le 1$. In fact, the conditions in Definition 4 do not hold at $\bar{x}$ when $\mathcal{I} = \{1\}$ and $\mathcal{I}_0 = \{2\}$.* □

Example 2 says that MPCaC-RCRCQ always holds at the origin if it is feasible. This is in accordance with the validity of the M-stationary condition: in fact, it has been observed that the origin, whenever feasible for (1), is always M-stationary regardless of the fulfillment of any MPCaC-CQ (see [22, Example 4.8]). More generally, the components $\bar{x}_i = 0$ do not impose restrictions on the fulfillment of the M-stationarity. To see this, observe that we can take $\gamma_i = -(\nabla f(\bar{x}) + \nabla g(\bar{x})\mu + \nabla h(\bar{x})\lambda)_i$ in Definition 1 for any fixed μ and λ, whenever $\bar{x}_i = 0$. This is an intrinsic property of M-stationarity. So, MPCaC-RCRCQ is somehow closer to M-stationary points than MPCaC-CRCQ.

Very recently, a new MPCaC-tailored CQ for M-stationarity, called CC-AM-regular CQ, was defined [22]. To maintain the compatibility with our notation, we will refer to it as *MPCaC-AM-regular* CQ. It was established in the context of sequential optimality conditions, similar to what is done in NLP [7]. Actually, MPCaC-AM-regularity is just the *cone continuity property* (CCP) [7] on TNLP($\bar{x}$), and it is implied by other CQs such as MPCaC-CRCQ (see the discussion before Corollary 4.10 of [22]). Let us present it formally. Given $\bar{x} \in \Omega$, we define the set

$$K(x, \bar{x}) = \left\{ \nabla g(x)\mu + \nabla h(x)\lambda + \gamma \;\middle|\; \begin{array}{l} \mu \geq 0, \;\; \mu_j = 0, \; j \notin I_g(\bar{x}), \\ \gamma_i = 0, \; \forall i \notin I_0(\bar{x}) \end{array} \right\}$$

for each $x \in \mathbb{R}^n$. The *Painlevé-Kuratowski outer limit* [27] of the mapping $x \rightrightarrows K(x, \bar{x})$ as $x \to \bar{x}$ is the set

$$\limsup_{x \to \bar{x}} K(x, \bar{x}) = \{\bar{y} \in \mathbb{R}^n \mid \exists \{(x^k, y^k)\} \to (\bar{x}, \bar{y}) \text{ such that } y^k \in K(x^k, \bar{x}), \forall k\}.$$

Definition 6. *We say that $\bar{x} \in \Omega$ conforms to* MPCaC-AM-regular CQ *if $K(\cdot, \bar{x})$ is outer semi-continuous at $\bar{x}$, that is, if $\limsup_{x \to \bar{x}} K(x, \bar{x}) \subset K(\bar{x}, \bar{x})$.*

By [22, Theorems 3.2 and 4.7], MPCaC-AM-regularity is a CQ for M-stationarity as summarized below.

Theorem 2. *Let $\bar{x}$ be a local minimizer of the problem (1) that satisfies the MPCaC-AM-regularity condition. Then $\bar{x}$ is M-stationary.*

Next, we prove that MPCaC-RCRCQ is a CQ for M-stationarity by showing that it implies MPCaC-AM-regularity.

Theorem 3. *If $\bar{x} \in \Omega$ satisfies MPCaC-RCRCQ, then MPCaC-AM-regularity holds at $\bar{x}$. In particular, every local minimizer of (1) that conforms to MPCaC-RCRCQ is an M-stationary point.*

Proof. Note that $\bar{x}$ is feasible for problem (7). Moreover, by Definition 5, we see that this point satisfies the standard RCRCQ condition for the nonlinear problem (7). Thus, by the relations between the standard constraints qualifications established in [7], we conclude that CCP for the tightened problem (7) holds at $\bar{x}$. But this is just the meaning of MPCaC-AM-regularity given in Definition 6. The last statement follows directly from Theorem 2. □ □

Other constraint qualifications introduced in the literature are *MPCaC-constant positive linear dependence (MPCaC-CPLD) CQ* [16] and *MPCaC-quasinormality (MPCaC-QN) CQ* [21] (we rename them to fit our nomenclature). The second was used in the convergence analysis of the first-order augmented Lagrangian method defined in [2], which we will present in the next section. As before, these CQs consist of imposing their NLP counterparts on TNLP($\bar{x}$). So, the relationship among all the cited MPCaC-CQs follows that of standard NLP, see [7, Fig. 2], which is completely summarized in Figure 2.

Similar to standard NLP [3, 5], MPCaC-RCRCQ is a CQ for strong second-order stationarity as we show next.

Theorem 4. *Let x^* be a local minimizer of (1) and suppose that MPCaC-RCRCQ is satisfied at x^*. Then x^* is M-stationary and, for any associated multiplier vector (μ, λ, γ), x^* fulfills the MPCaC-SSONC condition, that is,*

$$d^T \nabla^2_{xx} \mathcal{L}(x^*, \mu, \lambda, \gamma) d \geq 0$$

for all $d \in \mathcal{C}^S(x^*)$, where $\mathcal{C}^S(x^*)$ is the strong MPCaC critical cone (5).

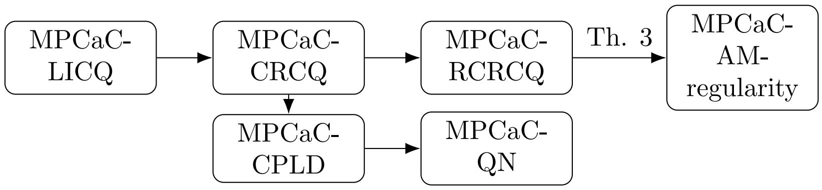

Figure 2: Relations among the MPCaC-tailored CQs. An arrow indicates a strict implication between two conditions.

Proof. By Theorem 3, x^* is M-stationary. Let (μ, λ, γ) be an arbitrary multiplier vector associated with it and take a vector $d \in \mathcal{C}^S(x^*)$, $d \neq 0$. Applying [3, Proposition 3.2] to the tightened problem (7), we conclude that there exists a twice continuously differentiable arc $\zeta : [0, \delta) \to \mathbb{R}^n$ such that $\zeta(0) = x^*$, $\zeta'(0) = d$, $\zeta([0, \delta))$ is feasible for (7), and $g_j(\zeta(t)) = 0$ for all $t \in [0, \delta)$ and $j \in I_g^+(x^*)$. Note that, in particular, $\zeta_\ell(t) = 0$ for all $t \in [0, \delta)$ and $\ell \in I_0(x^*)$. Define $\beta : [0, \delta) \to \mathbb{R}$ by $\beta(t) = f(\zeta(t))$. Since

$$\nabla f(x^*) + \sum_{j \in I_g^+(x^*)} \mu_j \nabla g_j(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{\ell \in I_0(x^*)} \gamma_\ell e_\ell = 0, \tag{8}$$

we have $\beta'(0) = \nabla f(x^*)^T d = 0$. Moreover, x^* is a local minimizer of (7) and hence $\beta(0) \leq \beta(t)$ for all $t \geq 0$ sufficiently small. Therefore,

$$d^T \nabla^2 f(x^*) d + \nabla f(x^*)^T \zeta''(0) = \beta''(0) \geq 0. \tag{9}$$

On the other hand, we have

$$d^T \nabla^2 g_j(x^*) d + \nabla g_j(x^*)^T \zeta''(0) = (g_j \circ \zeta)''(0) = 0 \tag{10}$$

for $j \in I_g^+(x^*)$,

$$d^T \nabla^2 h_i(x^*) d + \nabla h_i(x^*)^T \zeta''(0) = (h_i \circ \zeta)''(0) = 0 \tag{11}$$

for $i = 1, \ldots, p$ and

$$e_\ell^T \zeta''(0) = \zeta_\ell''(0) = 0 \tag{12}$$

for $\ell \in I_0(x^*)$. Thus, multiplying (10) by μ_j, (11) by λ_i, (12) by γ_ℓ and summing the resulting expressions over $j \in I_g^+(x^*)$, $i = 1, \ldots, p$, $\ell \in I_0(x^*)$ together with (9), we obtain the desired result in view of (8). $\square$ $\square$

Since $\mathcal{C}^W(x^*) \subset \mathcal{C}^S(x^*)$, an immediate consequence of the above theorem is that MPCaC-RCRCQ is also a CQ for weak second-order stationarity.

Corollary 1. *Let x^* be a local minimizer of (1) and suppose that MPCaC-RCRCQ holds at x^*. Then x^* is M-stationary and fulfills MPCaC-WSONC for any multiplier vector (μ, λ, γ) associated with x^*.*

Remark 3. *Despite Theorem 4 deals with strong second-order stationarity, condition MPCaC-WSONC in Corollary 1 has a more practical appeal. In fact, even in standard NLP we do not expect that strong stationarity is associated with the convergence of practical algorithms [11, 19].*

In [15], a second-order necessary optimality condition linked with S-stationarity (KKT for (2)) is defined. So, this condition depends on y. Also, the authors remark that it is not possible to extend it to encompass M-stationary points due to the freedom in which d can be taken. This is disappointing since M-stationarity is as good as KKT, while it is simpler since it does not use y. Instead, by tightening the requirements on d regarding to the cardinality $\|x\|_0 \leq \alpha$, our second-order stationarity is suitable for M-stationarity, and y is not used. Furthermore, MPCaC-WSONC actually strengthens the quality of stationarity beyond M-stationary, see Example 4.

A second-order condition using the linearization $\mathcal{T}_{\text{lin}}^{BS}(\bar{x})$ instead of $\mathcal{T}_{\text{lin}}(\bar{x})$, as in Definition 2, was stated in [15, Corollary 3.1]. This condition still involves the auxiliary variable y, as it requires

S-stationarity. But since $\mathcal{T}_{\text{lin}}^{BS}(\bar{x})$ does not explicitly have y, one might ask whether it is reasonable to define a (weak) second-order stationarity associated with M-stationary points using $\mathcal{T}_{\text{lin}}^{BS}(\bar{x})$ instead of $\mathcal{T}_{\text{lin}}(\bar{x})$; that is, using

$$\mathcal{C}^{BS}(\bar{x}) = \left\{d \in \mathcal{T}_{\text{lin}}^{BS}(\bar{x}) \mid \nabla g_i(\bar{x})^T d = 0,\ i \in I_g(\bar{x})\right\} \tag{13}$$

in Definition 2. In view of Theorem 1, the resulting concept is the same as MPCaC-WSONC if $\|\bar{x}\| = \alpha$. If $\|\bar{x}\| < \alpha$, using (13) gives more freedom to the directions d and so, in principle, would lead to a stronger stationarity. However, (13) can lead to an undesirable concept due to the cardinality constraint, as the next example shows. In particular, Corollary 1 would no longer be valid.

Example 3. *Let us consider the three-dimensional problem*

$$\text{minimize} \quad \tfrac{1}{2}\left[(x_1-1)^2 - x_2^2 - x_3^2\right] \quad \text{subject to} \quad x_2 - x_3^3 = 0,\ \|x\|_0 \leq 2.$$

The feasible point $\bar{x} = (1,0,0)$ is an isolated local minimizer since $(1,t^3,t)$, $t \neq 0$, is not feasible. It is clear that MPCaC-LICQ is valid at $\bar{x}$, and so $\bar{x}$ is M-stationary with multipliers λ and $\gamma = (0,-\lambda,0)$ (see Definition 1). For any $\lambda \in \mathbb{R}$ we have

$$\nabla_{xx}^2 \mathcal{L}(\bar{x},\lambda,\gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad \textit{and} \qquad C^{BS}(\bar{x}) = \mathbb{R} \times \{0\} \times \mathbb{R}.$$

Taking $d = (0,0,1) \in C^{BS}(\bar{x})$, we have $d^T \nabla_{xx}^2 \mathcal{L}(\bar{x},\lambda,\gamma) d = -1 < 0$. On the other hand, $C^W(\bar{x}) = \mathbb{R} \times \{0\}^2$ and so MPCaC-WSONC is valid at $\bar{x}$.

To end this section, we briefly comment on the validity of MPCaC-WSONC at the origin. As we already mentioned in the introduction, the origin, if feasible, is always M-stationary independently of the validity of any CQ. MPCaC-WSONC, as long as defined over M-stationary points, has the same property (note that $d_i = 0$ in $C^W(\bar{x})$ if $\bar{x}_i = 0$). In this sense, MPCaC-WSONC acts only on nonzero components of $\bar{x}$. This is not a serious drawback, for instance, when a sparse solution is required, a typical situation modelled by (1).

3 An augmented Lagrangian method that converges to MPCaC-WSONC points

In this section we consider the application of the second-order (safeguarded) AL method known as Algencan-second [1] to solve (1). It is based on the well established first-order Algencan method [2]. As we mentioned in the introduction, the reformulated problem (2) is suitable to be solved by optimization methods that require smooth data, such as Algencan (see [21]) and Algencan-second. Thus, the algorithm deals with (2), but its convergence will be established using stationary concepts not involving the auxiliary variable y.

The Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian function associated with (2) is given by

$$L_\rho(x,y,\bar{\mu},\bar{\lambda},\bar{\mu}^e,\bar{\mu}^y,\bar{\gamma}) = f(x) + \frac{\rho}{2}\left[\left\|\left(g(x)+\frac{\bar{\mu}}{\rho}\right)_+\right\|_2^2 + \left\|h(x)+\frac{\bar{\lambda}}{\rho}\right\|_2^2 + \left(n - e^T y - \alpha + \frac{\bar{\mu}^e}{\rho}\right)_+^2 + \left\|\left(y - e + \frac{\bar{\mu}^y}{\rho}\right)_+\right\|_2^2 + \left\|x * y + \frac{\bar{\gamma}}{\rho}\right\|_2^2\right],$$

where $\rho > 0$, and the subproblem is stated as

$$\underset{x,y}{\text{minimize}}\ L_\rho(x,y,\bar{\mu},\bar{\lambda},\bar{\mu}^e,\bar{\mu}^y,\bar{\gamma}) \tag{14}$$

for given $\rho > 0$, $\bar{\mu} \in \mathbb{R}_+^p$, $\bar{\lambda} \in \mathbb{R}^m$, $\bar{\mu}^e \in \mathbb{R}_+$, $\bar{\mu}^y \in \mathbb{R}_+^n$ and $\bar{\gamma} \in \mathbb{R}^n$.

In order to obtain MPCaC-WSONC points, we require that the subproblems are solved up to second-order stationarity. We say that $(x,y) \in \mathbb{R}^{2n}$ is a *second-order ε-stationary point* for (14) if the following conditions are satisfied:

- **first-order condition**

$$\left\|\nabla_{(x,y)} L_\rho(x, y, \bar{\mu}, \bar{\lambda}, \bar{\mu}^e, \bar{\mu}^y, \bar{\gamma})\right\|_\infty \le \varepsilon. \tag{15}$$

This is in accordance with ALGENCAN method described in [2]. See also [21].

- **second-order condition**

$$d^T \nabla^2_{(x,y)} L_\rho(x, y, \bar{\mu}, \bar{\lambda}, \bar{\mu}^e, \bar{\mu}^y, \bar{\gamma}) d \ge -\varepsilon, \quad \forall d \in \mathbb{R}^{2n}, \tag{16}$$

where the second derivatives of terms $\frac{\rho}{2}(p(z)+\eta/\rho)_+^2$ related to inequality constraints $p(z) \le 0$ are computed as

$$(\rho p(z) + \eta)_+ \nabla^2 p(z) + \delta \rho \nabla p(z) \nabla p(z)^T,$$

$\delta = 1$ if $\frac{1}{\sqrt{\rho}}(\rho p(z) + \eta) \ge -\varepsilon$ and zero otherwise. This "smooth" second derivative was employed in the ALGENCAN-SECOND method described in [1]. In that paper it was also presented an implementable algorithm, namely GENCAN-SECOND, that ensures conditions (15) and (16).

We present the second-order AL method in Algorithm 1.

Algorithm 1 (Safeguarded) Second-order AL method applied to MPCaC

Set the parameters $\tau \in [0, 1)$, $\theta > 1$, $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max}, \mu^e_{\max}, \mu^y_{\max} > 0$ and $\rho_1 > 0$. Let $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^1 \in [0, \mu_{\max}]^p$, $\bar{\mu}^{e,1} \in [0, \mu^e_{\max}]$, $\bar{\mu}^{y,1} \in [0, \mu^y_{\max}]^n$ and $\bar{\gamma}^1 \in [\gamma_{\min}, \gamma_{\max}]^n$ be the initial Lagrange multipliers estimates, and $\{\varepsilon_k\} \downarrow 0$. Initialize $k \leftarrow 1$.

Step 1. (Solving the subproblem) Compute a second-order ε_k-approximate stationary point (x^k, y^k) of the subproblem $\min_{x,y} L_{\rho_k}(x, y, \bar{\mu}^k, \bar{\lambda}^k, \bar{\mu}^{e,k}, \bar{\mu}^{y,k}, \bar{\gamma}^k)$.

Step 2. (Estimate new multipliers) Compute

$$\mu^k = \left(\bar{\mu}^k + \rho_k g(x^k)\right)_+, \qquad \lambda^k = \bar{\lambda}^k + \rho_k h(x^k), \qquad \gamma^k = \bar{\gamma}^k + \rho_k (x^k * y^k).$$
$$\mu^{e,k} = \left(\bar{\mu}^{e,k} + \rho_k(n - e^T y^k - \alpha)\right)_+, \quad \mu^{y,k} = \left(\bar{\mu}^{y,k} + \rho_k(y^k - e)\right)_+$$

Step 3. (Update the penalty parameter) Define

$$V^k = \min\left\{-g(x^k), \frac{\bar{\mu}^k}{\rho_k}\right\}, \quad V^k_e = \min\left\{e^T y^k + \alpha - n, \frac{\bar{\mu}^{e,k}}{\rho_k}\right\}, \quad V^k_y = \min\left\{e - y^k, \frac{\bar{\mu}^{y,k}}{\rho_k}\right\}.$$

If $k = 1$ or

$$\max\left\{\|h(x^k)\|_\infty, \|(V^k, V^k_e, V^k_y)\|_\infty\right\} \le \tau \max\left\{\|h(x^{k-1})\|_\infty, \|(V^{k-1}, V^{k-1}_e, V^{k-1}_y)\|_\infty\right\},$$

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

Step 4. (Update multipliers estimates) Compute $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^p$, $\bar{\mu}^{e,k+1} \in [0, \mu^e_{\max}]$, $\bar{\mu}^{y,k+1} \in [0, \mu^y_{\max}]^n$ and $\bar{\gamma}^{k+1} \in [\gamma_{\min}, \gamma_{\max}]^n$.

Step 5. Set $k \leftarrow k + 1$ and go to step 1.

Algorithm 1 generates a sequence $\{(x^k, y^k)\}$. However, in view of Definition 1, we are interested in the accumulation points of $\{x^k\}$ only. The next result says that we can take the limit of the auxiliary sequence $\{y^k\}$ whenever $\{x^k\}$ converges. Due to this lemma, we will take accumulation points of the auxiliary sequence without further details.

Lemma 2 ([21, Proposition 4.1]). *Let $\{x^k\}$ be a sequence generated by Algorithm 1. If $\{x^k\}$ has a bounded subsequence, then $\{y^k\}$ is bounded on the same subsequence.*

In [21], it was shown that the usual first-order safeguarded AL method converges to a stationary point of the infeasibility without additional assumptions. That is, only approximate first-order (15) guarantees that every (possibly non feasible) limit point of Algorithm 1 is stationary for the infeasibility measure

$$\Phi(x,y) = \|g(x)_+\|_2^2 + \|h(x)\|_2^2 + (n - e^T y - \alpha)_+^2 + \|y - e\|_2^2 + \|x * y\|_2^2. \tag{17}$$

This is in accordance with standard results in NLP for Algencan [2, 13]. So, in the sequel we deal only with feasible accumulation points.

Convergence of the first-order counterpart of Algorithm 1 through sequential optimality conditions was addressed in [22]. The authors showed that Algorithm 1, with subproblems solved only up to (15), converges to M-stationary points under a *generalized Kurdyka-Lojasiewicz* (GKL) [9] inequality and the MPCaC-AM-regular CQ (Definition 6). The GKL inequality is a common hypothesis in convergence results of Algencan on degenerate problems. See [6, 22]. We say that a smooth function $F : \mathbb{R}^q \to \mathbb{R}$ satisfies the GKL inequality at $\bar{z}$ if there exist $\delta > 0$ and $\varphi : B_\delta(\bar{z}) \to \mathbb{R}$ such that $\lim_{z\to\bar{z}} \varphi(z) = 0$ and for each $z \in B_\delta(\bar{z})$ we have $|F(z) - F(\bar{z})| \le \varphi(z)\|\nabla F(z)\|_2$. With this previously established result, the convergence to M-stationary points under MPCaC-RCRCQ plus GKL is a consequence of Theorem 3 and [22, Theorems 4.7 and 6.3]. We summarize it in the next theorem.

Theorem 5. *Let (x^*, y^*) be a feasible accumulation point of a sequence generate by Algorithm 1 with subproblems solved only up to satisfy (15). Suppose that x^* conforms to MPCaC-AM-regularity (or, in particular, MPCaC-RCRCQ) and that $\Phi(x,y)$ (see (17)) satisfies the GKL inequality at (x^*, y^*). Then x^* is M-stationary.*

Next, we discuss the global convergence for the second-order AL method, where subproblems are solved in order to satisfy (15) and (16). Similar to standard NLP [1], we need the following technical result:

Lemma 3. *Let $\bar{x} \in \Omega$. Define the set*

$$C^W(x,\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_i(x)^T d = 0,\, i \in I_g(\bar{x}), \\ \nabla h(x)^T d = 0, \\ d_i = 0,\, i \in I_0(\bar{x}) \end{array} \right\}$$

and let $\bar{d} \in C^W(\bar{x},\bar{x})$. Suppose that $\bar{x}$ conforms to MPCaC-RCRCQ. Also, let $\{x^k\}$ be a sequence, not necessarily feasible, converging to $\bar{x}$. Then there exists a sequence $\{d^k\}$ converging to $\bar{d}$ such that $d^k \in C^W(x^k,\bar{x})$ for all k.

Proof. This lemma was proved for standard NLP in [8] assuming, instead of a CQ, that the gradients of the active (equality and inequality) constraints at $\bar{x}$ have the same rank in an open neighborhood of $\bar{x}$. Thus, the statement will follow from [8, Lemma 3.1] applied to TNLP($\bar{x}$) provided the following trivial assertions: (*i*) the set $T(x)$ in [8] with $x^* = \bar{x}$ for TNLP($\bar{x}$) is exactly $C^W(x,\bar{x})$; (*ii*) MPCaC-RCRCQ guarantees that all the gradients of the active constraints of TNLP($\bar{x}$) at $\bar{x}$, namely $\nabla g_i(\bar{x})$, $i \in I_g(\bar{x})$, $\nabla h_j(\bar{x})$, $j = 1,\dots,p$, e_i, $i \in I_0(\bar{x})$, have constant rank in an open neighborhood of $\bar{x}$. □ □

We prove next the main result of this section.

Theorem 6. *Let (x^*, y^*) be a feasible accumulation point of a sequence generate by Algorithm 1. Suppose that x^* conforms to MPCaC-RCRCQ and that $\Phi(x,y)$ satisfies the GKL inequality at (x^*, y^*). Then x^* is an MPCaC-WSONC point.*

Proof. Taking a subsequence if necessary and using Lemma 2, we can suppose w.l.o.g. that the entire sequence $\{(x^k, y^k)\}$ generated by the method converges to (x^*, y^*). From (15), the estimate given by the method at iteration k for M-stationary multipliers as in Definition 1 is

$$(\tilde{\mu}^k, \tilde{\lambda}^k, \tilde{\gamma}^k) = (\mu^k, \lambda^k, \gamma^k * y^k). \tag{18}$$

Considering (3), condition (16) says that (we eventually omit the index k for simplicity)

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x,\tilde{\mu},\tilde{\lambda},\tilde{\gamma}) & 0 \\ 0 & 0 \end{bmatrix} + \rho \begin{bmatrix} \sum_{i=1}^p \nabla h_i \nabla h_i^T + \sum_{j|\frac{1}{\sqrt{\rho}}(\rho g_j+\bar{\mu}_j)\geq -\varepsilon_k} \nabla g_j \nabla g_j^T & 0 \\ 0 & 0 \end{bmatrix}$$
$$+ \rho \begin{bmatrix} Y^2 & 2XY + \mathrm{diag}(\bar{\gamma})/\rho \\ 2XY + \mathrm{diag}(\bar{\gamma})/\rho & I^e + I^y + X^2 \end{bmatrix} \succeq -\varepsilon_k I, \tag{19}$$

where the ∇h_i, ∇g_j and g_j are evaluated at x, $X_k = \mathrm{diag}(x^k)$, $Y_k = \mathrm{diag}(y^k)$,

$$I^e = \begin{cases} E, & \frac{1}{\sqrt{\rho}}\big(\rho(n-\alpha-e^T y) + \bar{\mu}^e\big) \geq -\varepsilon_k \\ 0, & \text{otherwise,} \end{cases}$$

E is the square matrix of all ones, and

$$I^y_{ij} = \begin{cases} 1, & \frac{1}{\sqrt{\rho}}\big(\rho(y_i - e_i) + \bar{\mu}^y_i\big) \geq -\varepsilon_k \quad \text{and} \quad i = j \\ 0, & \text{otherwise} \end{cases}.$$

If $\{\rho_k\}$ is bounded, then $\{(\mu^k,\lambda^k,\gamma^k)\}$ is bounded too and, by Lemma 2, (18) is also bounded. Thus, taking a subsequence if necessary, the last sequence converges, let us say, $\lim_{k\in K}(\tilde{\mu}^k,\tilde{\lambda}^k,\tilde{\gamma}^k) = (\mu^*,\lambda^*,\gamma^*)$. We affirm that this limit is an M-multiplier vector for x^*. In fact, (4a) follows from (15) and $\varepsilon_k \downarrow 0$. When $g_j(x^*) < 0$, step 3 of Algorithm 1 ensures that $V^k \to 0$, and so $-\bar{\mu}^k_j/\rho_k \to 0$, which implies $\bar{\mu}^k_j \to 0$. In this case, $\mu^* = \lim_{k\in K}\tilde{\mu}^k_j = \big(\bar{\mu}^k + \rho_k g(x^k)\big)_+ = 0$, leading to (4b). Condition (4c) holds since $\gamma^*_i x^*_i = \lim_{k\in K}\tilde{\gamma}^k_i x^k_i = \lim_{k\in K}\gamma^k_i(x^k_i y^k_i)$, which is zero by feasibility and the boundedness of $\{\gamma^k\}$.

Now, suppose that $\rho_k \to \infty$. As MPCaC-RCRCQ is valid at x^*, Theorem 5 ensures that x^* is M-stationary. We will show that the sequence (18) admits an associated M-multiplier vector $(\mu^*,\lambda^*,\gamma^*)$ as an accumulation point. In fact, first note that $\tilde{\mu}^k_j = \mu^k_j = 0$ for all k large enough and $j \notin I_g(x^*)$. Using the same arguments from the proof of [22, Theorem 6.3], we can show that the GKL hypothesis implies

$$\rho_k \Phi(x^k,y^k) \to 0 \tag{20}$$

(indeed, this argument is usual in this context and does not depend on whether $\{\rho_k\}$ is unbounded or not, see [6, 9, 10]). In view of (17), we have in particular that $\rho_k(x^k_i y^k_i)^2 \to 0$ for all i. So, by feasibility and the boundedness of $\{\bar{\gamma}^k\}$, we have

$$\tilde{\gamma}^k_i x^k_i = \bar{\gamma}^k_i(x^k_i y^k_i) + \rho_k(x^k_i y^k_i)^2 \to 0, \quad \forall i. \tag{21}$$

In particular, $\tilde{\gamma}^k_i \to 0$ for all $i \notin I_0(x^*)$. Therefore, expression (15) gives

$$\nabla f(x^k) + \sum_{j\in I_g(x^*)} \tilde{\mu}^k_j \nabla g_j(x^k) + \sum_{i=1}^p \tilde{\lambda}^k_i \nabla h_i(x^k) + \sum_{i\in I_0(x^*)} \tilde{\gamma}^k_i e_i \to 0.$$

At x^k, we can take index sets $I^k_h \subset \{1,\ldots,p\}$ and $I^k_0 \subset I_0(x^*)$ such that the gradients from the equality constraints $\{\nabla h_i(x^k), e_j \mid i \in I^k_h, j \in I^k_0\}$ form a basis for $\{\nabla h_i(x^k), e_j \mid i \in \{1,\ldots,p\}, j \in I_0(x^k)\}$. Also, we can assume w.l.o.g. that $\tilde{\mu}^k_j$, $\tilde{\lambda}^k_i$ and $\tilde{\gamma}^k_i$ are all non zero in the above sum. Then applying [5, Lemma 1], we can extract, for each k, a subset $I^k_g \subset I_g(x^*)$ such that the gradients

$$\nabla g_i(x^k),\ i \in I^k_g, \quad \nabla h_j(x^k),\ j \in I^k_h, \quad e_i,\ i \in I^k_0 \tag{22}$$

are linearly independent and

$$\nabla f(x^k) + \sum_{j\in I^k_g} \widehat{\mu}^k_j \nabla g_j(x^k) + \sum_{i\in I^k_h} \widehat{\lambda}^k_i \nabla h_i(x^k) + \sum_{i\in I^k_0} \widehat{\gamma}^k_i e_i = v^k$$

for some $(\widehat{\mu}^k_{I^k_g}, \widehat{\lambda}^k_{I^k_h}, \widehat{\gamma}^k_{I^k_0})$ with $\widehat{\mu}^k_j \tilde{\mu}^k_j > 0$, and where $v^k \to 0$. As there are only finitely many sets I^k_g, I^k_h and I^k_0, there is a subsequence $\{x^k\}_{k\in K}$ for which these sets repeat, let us say, $I^k_g = \mathcal{I}_g$, $I^k_h = \mathcal{I}_h$ and $I^k_0 = \mathcal{I}_0$ for all $k \in K$. So,

$$\lim_{k\in K} \nabla f(x^k) + \sum_{j\in \mathcal{I}_g} \widehat{\mu}^k_j \nabla g_j(x^k) + \sum_{i\in \mathcal{I}_h} \widehat{\lambda}^k_i \nabla h_i(x^k) + \sum_{i\in \mathcal{I}_0} \widehat{\gamma}^k_i e_i = 0. \tag{23}$$

Clearly, due to the linear independence of the gradients in (22), those of the new index sets are linearly independent too.

Define $\delta_k = \|(\widehat{\mu}^k_{\mathcal{I}_g}, \widehat{\lambda}^k_{\mathcal{I}_h}, \widehat{\gamma}^k_{\mathcal{I}_0})\|_\infty$ for all $k \in K$. If $\{\delta_k\}_{k\in K}$ has a bounded subsequence, we proceed as before taking this subsequence. Now, let us consider the case $\lim_{k\in K} \delta_k = \infty$. Dividing (23) by δ_k we arrive at

$$\lim_{k\in K} \frac{\nabla f(x^k)}{\delta_k} + \sum_{j\in\mathcal{I}_g} \frac{\widehat{\mu}^k_j}{\delta_k}\nabla g_j(x^k) + \sum_{i\in\mathcal{I}_h} \frac{\widehat{\lambda}^k_i}{\delta_k}\nabla h_i(x^k) + \sum_{i\in\mathcal{I}_0} \frac{\widehat{\gamma}^k_i}{\delta_k} e_i y_i = 0.$$

By definition of δ_k, we have $\widehat{\mu}^k_j/\delta_k = 1$, $\widehat{\lambda}^k_i/\delta_k = 1$ or $\widehat{\gamma}^k_i/\delta_k = 1$ for all $k \in K$. So, we get a linear combination of linearly independent vectors tending to the null vector, with at least one non-zero scalar. In other words, the rank of the gradients of active constraints at x^* in the neighborhood of x^* is larger than the rank of those gradients at x^*, contradicting the MPCaC-RCRCQ condition at x^*. Thus, (18) has a bounded subsequence. Note that the convergence or not of $\{\mu^{e,k}\}$ and $\{\mu^{y,k}\}$ does not matter.

We proceed to show that any accumulation point of a chosen bounded subsequence of $\{(\tilde{\mu}^k, \tilde{\lambda}^k, \tilde{\gamma}^k)\}$ is an M-multiplier associated with x^*. For simplicity, assume that the entire sequence $\{(\tilde{\mu}^k, \tilde{\lambda}^k, \tilde{\gamma}^k)\}$ converges to $(\mu^*, \lambda^*, \gamma^*)$. Clearly, (15) and $\varepsilon_k \downarrow 0$ imply (4a) for this limit. As we already mentioned, $\mu^*_j = \lim_k \tilde{\mu}^k_j = 0$ for $j \notin I_g(x^*)$. By (20), we have $\rho_k g_j(x^k)^2 \to 0$, and therefore

$$\mu^*_j g_j(x^*) = \lim_k \tilde{\mu}^k_j g_j(x^k) = \left(\bar{\mu}^k g_j(x^k) + \rho_k g_j(x^k)^2\right)_+ = 0.$$

Thus, (4b) also holds. The validity of (4c) follows from (21). Therefore, all the conditions of Definition 1 are fulfilled with x^* and $(\mu^*, \lambda^*, \gamma^*)$, as we wanted.

Now, suppose that x^* does not satisfy the MPCaC-WSONC. Thus, considering the multiplier vector $(\mu^*, \lambda^*, \gamma^*)$ just defined, there exists $d^* \in \mathcal{C}^W(x^*) = C^W(x^*, x^*)$ such that

$$(d^*)^T \nabla^2 \mathcal{L}(x^*, \mu^*, \lambda^*, \gamma^*) d^* < 0.$$

By Lemma 3, there is a sequence $\{d^k\}_{k\in K} \to d^*$ such that $d^k \in C^W(x^k, x^*)$ for all $k \in K$. In particular, $\nabla h(x^k)^T d^k = 0$ and $\nabla g_i(x^k)^T d^k = 0$ for all $i \in I_g(x^*)$ and $k \in K$. In addition, note that $\frac{1}{\sqrt{\rho_k}}(\rho_k g_j(x^k) + \bar{\mu}^k_j) \geq -\varepsilon_k \to 0$ implies $g_j(x^*) = 0$. In fact, it is trivial if $\rho_k \to \infty$ by the boundedness of $\{\bar{\mu}^k_j\}$. If $\{\rho_k\}$ is bounded, then $V^k \to 0$ by Step 3 of Algorithm 1. Thus, immediately $\bar{\mu}^k_j \to 0$ or $g_j(x^k) \to 0$. If the first case happens, it follows that $\frac{1}{\sqrt{\rho_k}}\rho_k g_j(x^k) \to 0$, which in turn implies $g_j(x^k) \to 0$.

Defining $\bar{d}^k = (d^k, 0)$ for all $k \in K$, we have from (19) and the above discussion that

$$\begin{aligned}(\bar{d}^k)^T \nabla^2 L_{\rho_k}(x^k, y^k, \bar{\mu}^k, \bar{\lambda}^k, \bar{\mu}^{e,k}, \bar{\mu}^{y,k}, \bar{\gamma}^k)\bar{d}^k \\ = (d^k)^T \nabla^2_{xx}\mathcal{L}(x^k, \tilde{\mu}^k, \tilde{\lambda}^k, \tilde{\gamma}^k) d^k + \rho_k (d^k)^T Y_k^2 d^k,\end{aligned} \tag{24}$$

and

$$\rho_k (d^k)^T Y_k^2 d^k = \sum_{i | x^*_i \neq 0} (d^k_i)^2 [\rho_k (y^k_i)^2]$$

for all $k \in K$ large enough. From (20) we have $\rho_k (x^k_i y^k_i)^2 \leq \rho_k \Phi(x^k, y^k) \to 0$ for all i. In particular, for all $i \notin I_0(x^*)$ we have $\rho_k (y^k_i)^2 \to 0$ and therefore (24) implies

$$(\bar{d}^k)^T \nabla^2 L_\rho(x^k, y^k, \bar{\mu}^k, \bar{\lambda}^k, \bar{\mu}^{e,k}, \bar{\mu}^{y,k}, \bar{\gamma}^k)\bar{d}^k \leq \frac{1}{2}(d^*)^T \nabla^2_{xx}\mathcal{L}(x^*, \mu^*, \lambda^*, \gamma^*) d^* < 0$$

for all $k \in K$ large enough. This contradicts (16), and hence x^* is an MPCaC-WSONC point. □

□

Remark 4. *The reader may ask if it is possible to obtain Theorem 6 directly from related results in NLP, specifically the third item of [1, Theorem 2] and [4, Theorems 4.2 and Section 3.3]. The answer is "no". The reason is that these results require standard CQs, but they may not hold for (2) [24].*

The next example illustrates that if the subproblems of Algorithm 1 are solved only up to the first-order condition (15), the AL method can converge to a non-MPCaC-WSONC M-stationary point even if MPCaC-LICQ and GKL hold.

Example 4. *Inspired in [4, Example 3.10], let us consider the three-dimensional problem*

$$\text{minimize} \quad -x_1 - x_2 + x_3^2 \quad \text{subject to} \quad x_1^2 x_2^2 - 1 \le 0, \quad \|x\|_0 \le 2,$$

whose relaxed reformulation is

$$\begin{array}{rl} \text{minimize} & -x_1 - x_2 + x_3^2 \\ \text{subject to} & x_1^2 x_2^2 - 1 \le 0, \\ & 3 - e^T y \le 2, \quad y \le e, \quad x * y = 0. \end{array}$$

The feasible point $x^* = (1, 1, 0)$ clearly satisfies MPCaC-LICQ. Also, $\Phi(x, y)$ satisfies the GKL inequality at all $(x, y) \in \mathbb{R}^6$ since all data functions are analytic [9].

In view of Definition 1, it is easy to verify that x^* is M-stationary with unique $\mu = 1$ and $\gamma = (0, 0, 0)$. But x^* is not an MPCaC-WSONC point since for $d = (-1, 1, 0) \in C^W(x^*) = \{(d_1, d_2, 0) \mid d_1 + d_2 = 0\}$ (see (6)) we have

$$d^T \nabla^2_{xx} \mathcal{L}(x^*, \mu, \gamma) d = d^T \begin{bmatrix} 2 & 4 & 0 \\ 4 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} d = -2 < 0.$$

Let us show that Algorithm 1 with subproblems solved only satisfying (15) can converge to x^*. To simplify, suppose that all multipliers estimates in Step 4 are chosen as zero. So Step 1 reduces to

$$\begin{bmatrix} -1 \\ -1 \\ 2x_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \rho(x_1^2 x_2^2 - 1)_+ 2x_1 x_2^2 \\ \rho(x_1^2 x_2^2 - 1)_+ 2x_1^2 x_2 \\ 0 \\ \rho(y_1 - 1)_+ - \rho(1 - e^T y)_+ \\ \rho(y_2 - 1)_+ - \rho(1 - e^T y)_+ \\ \rho(y_3 - 1)_+ - \rho(1 - e^T y)_+ \end{bmatrix} + \rho \begin{bmatrix} x_1 y_1^2 \\ x_2 y_2^2 \\ x_3 y_3^2 \\ x_1^2 y_1 \\ x_2^2 y_2 \\ x_3^2 y_3 \end{bmatrix} \to \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{25}$$

Take the sequences defined by

$$x^k = \left(1 + \frac{1}{8\rho_k}, 1 + \frac{1}{8\rho_k}, 0\right), \quad y^k = \left(\frac{1}{\rho_k^2}, \frac{1}{\rho_k^2}, 1\right) \quad \text{and} \quad \rho_k = \theta^{k-2}\rho_1,$$

$k \ge 2$ and $\rho_1 > 1$. We have $\rho_k((x_1^k)^2 (x_2^k)^2 - 1)_+ \to 1/2$, $(1 - e^T y^k)_+ = 0$, $(y_i^k - 1)_+ = 0$ and $\rho_k x_i^k y_i^k \to 0$ for $i = 1, 2, 3$. Thus these sequences satisfy (25). Also, note that all quantities $\|V_z^k\|_\infty$ do not decrease at a linear rate, so the test in Step 4 fails infinitely many times. Therefore, the above sequences are possible in Algorithm 1 when only (15) is required, that is, the method can converge to the non-MPCaC-WSONC point x^*.

On the other hand, Algorithm 1 with subproblems requiring (16) cannot converge to x^* due to Theorem 6.

As we mentioned in the introduction, the reformulated problem (2) is related to MPCCs, which are characterized by the presence of complementarity-type constraints $G(x) \ge 0$, $H(x) \ge 0$ and $G(x) * H(x) = 0$. One of the most prominent difference between MPCCs and MPCaCs relies on the fact that, for MPCaCs, M- and S-stationarity (KKT for (2)) are in some sense equivalent (see Theorem 7). This is not true for the related concepts for MPCC. Actually, S-stationarity in MPCCs can be guaranteed only under assumptions stronger than the Mangasarian-Fromovitz-type CQ or assuming that G_i and H_i cannot vanish simultaneously for each i. See [28]. As a consequence, we do not expect algorithms converging to S-stationary points of MPCCs even under strong assumptions. This scenario is clearly different for MPCaCs where, in particular, AL and regularization methods converge to KKT under very mild hypotheses. See for instance [21, 22] and Theorem 5.

Our theory for Algorithm 1 goes in the same direction. It shows that, besides first-order, MPCaCs inherit much of the second-order stationarity of standard NLP. Despite the (strong) second-order stationarity concept provided in [15], Theorem 6 illustrates that MPCaC-WSONC is suitable for convergence analysis of algorithms, as WSONC is for standard NLP [11, 19]. Indeed, the spirit of our work is to derive a concept linked to practical algorithms. So, we contribute to attest that MPCaC is somewhat closer to standard NLP than to MPCC. It is worth mentioning that it is not known whether the second-order AL method [1], which derives Algorithm 1, converges to second-order stationary points of MPCCs or not. The convergence of such method for MPCCs was considered in [12], where it was shown only that (first-order) M-stationary-type points are reached under a strong linear independence assumption. Instead, Theorem 6 attest convergence to second-order points of MPCaC under GKL and the mild MPCaC-RCRCQ condition.

4 Strength of the MPCaC-WSONC condition

A natural question is what the relationship between the specialized MPCaC-WSONC condition and the standard WSONC applied to the reformulated problem (2) viewed as an NLP. Next we address this issue.

Let us consider a feasible $(\bar{x}, \bar{y})$ for (2). The KKT conditions for this problem are

$$\begin{bmatrix} \nabla f(\bar{x}) + \nabla g(\bar{x})\mu + \nabla h(\bar{x})\lambda \\ -\mu^e e - \mu^y \end{bmatrix} + \begin{bmatrix} \lambda^c * \bar{y} \\ \lambda^c * \bar{x} \end{bmatrix} = 0, \tag{26a}$$

$$g_j(\bar{x})\mu_j = 0, \ \forall j, \quad \mu_i^y(\bar{y}_i - 1) = 0, \ \forall i, \quad \mu^e(n - e^T\bar{y} - \alpha) = 0, \tag{26b}$$

$$\mu \geq 0, \quad \mu^e \geq 0, \quad \mu^y \geq 0, \tag{26c}$$

for certain multipliers $\mu \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^p$, $\mu^e \in \mathbb{R}$, $\mu^y \in \mathbb{R}^n$ and $\lambda^c \in \mathbb{R}^n$. In turn, standard WSONC at $(\bar{x}, \bar{y})$ holds if this point is KKT (in the sense of (26)) with an associated multiplier vector $(\mu, \lambda, \mu^e, \mu^y, \lambda^c)$ satisfying

$$d^T \begin{bmatrix} \nabla^2 f(\bar{x}) + \sum_j \mu_j \nabla^2 g_j(\bar{x}) + \sum_i \lambda_i \nabla^2 h_i(\bar{x}) & \mathrm{diag}(\lambda^c) \\ \mathrm{diag}(\lambda^c) & 0 \end{bmatrix} d \geq 0 \tag{27}$$

for all $d = (d^x, d^y)$ in

$$C^W_{\mathrm{NLP}}(\bar{x}, \bar{y}) = \left\{ (d^x, d^y) \in \mathbb{R}^{2n} \;\middle|\; \begin{array}{rcl} \nabla g_i(\bar{x})^T d^x & = & 0, \ i \in I_g(\bar{x}), \\ \nabla h(\bar{x})^T d^x & = & 0, \\ e^T d^y & = & 0 \ \text{ if } \ n - e^T\bar{y} = \alpha, \\ d_i^y & = & 0, \ i : \bar{y}_i = 1, \\ \bar{y}_i d_i^x + \bar{x}_i d_i^y & = & 0, \ i = 1, \ldots, n \end{array} \right\}.$$

Note that the usual first and second-order optimality conditions (26) and (27) for (2) change as auxiliary variables y_i assume fractional values. By the proof of [21, Proposition 2.3], it was observed that when y_i is fractional at optimality, taking $y_i = 1$ maintains optimality. This is the crucial fact to elect M-stationarity as the strongest first order concept for (1), which we summarize below.

Theorem 7. *Let $\bar{x}$ be a feasible point for (1) and define $\bar{y}$ putting $\bar{y}_i = 1$ if $i \in I_0(\bar{x})$ and $\bar{y}_i = 0$ otherwise. Then $\bar{x}$ is M-stationary for (1) if, and only if, $(\bar{x}, \bar{y})$ is KKT for (2).*

A similar result is valid for our second-order optimality condition.

Theorem 8. *Let $\bar{x}$ and $\bar{y}$ as in Theorem 7. Then $\bar{x}$ is an MPCaC-WSONC point if, and only if, $(\bar{x}, \bar{y})$ conforms to the standard WSONC for (2).*

Proof. Let $\bar{x}$ be an MPCaC-WSONC point. In particular, $\bar{x}$ is M-stationary, for which there are $\mu \in \mathbb{R}^m_+$, $\lambda \in \mathbb{R}^p$ and $\gamma \in \mathbb{R}^n$ such that

$$\nabla f(\bar{x}) + \nabla g(\bar{x})\mu + \nabla h(\bar{x})\lambda + \gamma = 0, \quad \mu^T g(\bar{x}) = 0, \quad \gamma * \bar{x} = 0.$$

As $\bar{y}_i = 1$ whenever $\bar{x}_i = 0$, we have $\gamma * \bar{y} = \gamma$, and thus conditions (26) are satisfied with $\mu^e = 0$, $\mu^y = 0$ and $\lambda^c = \gamma$. That is, $(\bar{x}, \bar{y})$ is KKT for (2) associated with $(\mu, \lambda, 0, 0, \gamma)$.

Now, let us show that this KKT-multiplier vector fulfills WSONC for (2). Given $(d^x, d^y) \in C^W_{\text{NLP}}(\bar{x}, \bar{y})$, we have $\nabla g_i(\bar{x})^T d^x = 0$ for all $i \in I_g(\bar{x})$ and $\nabla h(\bar{x})^T d^x = 0$. As $\bar{y}_i = 1$ if $\bar{x}_i = 0$, the last condition in $C^W_{\text{NLP}}(\bar{x}, \bar{y})$ implies that $d^x_i = 0$ whenever $\bar{x}_i = 0$, and hence $d^x \in C^W(\bar{x})$ (see (6)). Again from the definition of $\bar{y}$, we have $\bar{y}_i = 0$ if and only if $\bar{x}_i \neq 0$. So, the last condition in $C^W_{\text{NLP}}(\bar{x}, \bar{y})$ implies that $d^x * d^y = 0$. Therefore, the left hand-side of (27) with $\lambda^c = \gamma$ is

$$(d^x)^T W d^x + 2\gamma^T (d^x * d^y) = (d^x)^T W d^x, \tag{28}$$

where W is the Hessian of the Lagrangian function. As $d^x \in C^W(\bar{x})$, MPCaC-WSONC implies $(d^x)^T W d^x \geq 0$, from which we conclude that WSONC for (2) holds at $(\bar{x}, \bar{y})$.

The converse is straightforward considering the M-multiplier vector (μ, λ, γ) with $\gamma = \lambda^c * \bar{y}$ from (26), the definition of $\bar{y}$, expression (28) and the fact that $d^x \in C^W(\bar{x}) \Rightarrow (d^x, 0) \in C^W_{\text{NLP}}(\bar{x}, \bar{y})$. □ □

Theorem 8 says that MPCaC-WSONC is as strong as WSONC for the reformulated problem (2) viewed as an NLP. However, C^W_{NLP} is inconsistent in the following sense: if $\bar{x}_i = 0 \leq \bar{y}_i < 1$ then d^y_i is free in $C^W_{\text{NLP}}(\bar{x}, \bar{y})$, augmenting this cone unnecessarily (actually, when $\bar{x}_i = \bar{y}_i = 0$, we even may lost KKT conditions for (2)); and if $\bar{x}_i = 0 < \bar{y}_i = 1$ then $d^y_i = 0$. Instead, MPCaC-WSONC removes this ambiguity since it works only with the original variable x. In this sense, MPCaC-WSONC is to WSONC as M-stationarity is to KKT. The next example illustrates that standard WSONC may fail if $x_i = y_i = 0$ for some i.

Example 5. *Let us consider the three-dimensional problem of minimizing $x_1^2 + x_2^2 + x_3^2$ subject to $\|x\|_0 \leq 2$, for which $\bar{x} = (0, 0, 0)$ is the global solution. Its reformulated problem is*

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 + x_3^2 \\ \text{subject to} & 3 - y_1 - y_2 - y_3 \leq 2, \quad y \leq e, \quad x_i y_i = 0, \; i = 1, 2, 3. \end{array}$$

It is straightforward to verify that $(\bar{x}, \bar{y}) = (0, 0, 0, 0, 3/4, 3/4)$ is KKT for the above problem with multipliers $\mu^e = 0$, $\mu^y = (0, 0, 0)$ and any $\lambda^c \in \mathbb{R} \times \{0\}^2$. The left hand-side of (27) reduces to

$$2\|d^x\|_2^2 + 2\lambda_1^c d_1^x d_1^y. \tag{29}$$

Note that for $(d^x, d^y) = (1, 0, 0, 1, 0, 0) \in C^W_{NLP}(\bar{x}, \bar{y})$ and $\lambda^c = (-2, 0, 0)$, expression (29) is negative. □

In Example 5, the solution $\bar{x} = (0, 0, 0)$ satisfies the most stringent constraint qualification MPCaC-LICQ (Definition 3). Also, it satisfies second-order stationarity in both senses MPCaC-WSONC and standard WSONC; in fact, it conforms to WSONC by taking $\lambda^c = (0, 0, 0)$ and/or $\bar{y}_i > 0$ for $i = 1, 2, 3$ (e.g. $\bar{y} = (1, 1, 1)$). However, if an algorithm reaches $\bar{x} = \bar{y} = 0$, it can fails to declare convergence using the standard second-order optimality. So, establishing convergence using MPCaC-WSONC is more appropriate.

Remark 5. *Theorem 8 remains valid if, for each $\bar{x}_i = 0$, we take any $\bar{y}_i \in (0, 1]$ such that $(\bar{x}, \bar{y})$ is feasible. The proof is the same, but setting $\lambda^c_i = \gamma_i / \bar{y}_i$ for all i such that $\bar{y}_i > 0$ and zero otherwise (note that condition $\bar{y}_i = 1 \Rightarrow d^y_i = 0$ in $C^W_{NLP}(\bar{x}, \bar{y})$ was never used). In particular, the same statement is valid for M-stationarity (Theorem 7). In other words, the true degeneracy of MPCaC problems viewed through the reformulated problem (2) lies on the bi-activity $\bar{x}_i = \bar{y}_i = 0$. Note that if $(\bar{x}, \tilde{y})$ is feasible for (2), then all $(\bar{x}, \bar{y})$ where $1 \geq \bar{y}_i \geq \tilde{y}_i$ if $\bar{x}_i = 0$ and $\bar{y}_i = \tilde{y}_i = 0$ otherwise are also feasible.*

5 Conclusions

In this paper we have presented a study on second-order optimality conditions, constraint qualifications and a practical algorithm for cardinality-constrained problems (MPCaC). We have defined two new tailored second-order necessary conditions, MPCaC-SSONC and MPCaC-WSONC, a constraint qualification, MPCaC-RCRCQ, and we proved that minimizers of the original problem (1) satisfies MPCaC-SSONC under this CQ. The tailored constraint qualification, based on

RCRCQ [26], had not been defined yet in the context of MPCaC. Differently from previous second-order conditions [15], ours do not use the auxiliary variable y presented in the reformulated problem (2), only the original x. This allows them to be constructed using M-stationarity, which is the strongest first-order concept in the MPCaC-context. So, our second-order conditions are simple and fit well into the structure of cardinality-constrained problems. We also compared MPCaC-WSONC and the usual WSONC applied to (2), and showed that the proposed tailored condition excludes possible problems when verifying WSONC originated from the artificial variable y.

A primal (safeguarded) augmented Lagrangian algorithm was considered and its global convergence to second-order points was established. We have proved that, besides M-stationary points, the algorithm achieves MPCaC-WSONC ones. To the best of your knowledge, this is the first time that convergence up to second-order stationarity was established for an algorithm applied to MPCaC.

As the proposed MPCaC-WSONC proves to be a second-order optimality condition associated with algorithms, a topic for future research is the convergence of regularization methods such as those defined in [14, 15, 16]. It worth mentioning that these methods usually are adaptations from the MPCC-context, from where many other potentially algorithms can be originated. See for instance [23] and references therein. Another interesting topic is, despite the lack of continuity and convexity of the cardinality constraint, the establishment of methods that deal exclusively with original variable x. We believe that an inexact restoration [25] approach may be possible in this direction.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications*, 45(2):209–236, 2010.

[2] 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.

[3] R. Andreani, C. E. Echagüe, and M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization Theory and Applications*, 146(2):255–266, 2010.

[4] R. Andreani, G. Haeser, A. Ramos, and P. J. S. Silva. A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis*, 37(4):1902–1929, 2017.

[5] 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.

[6] 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.

[7] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016.

[8] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56(5–6):529–542, 2007.

[9] 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.

[10] R. Andreani, A. Ramos, A. A. Ribeiro, L. D. Secchin, and A. R. Velazco. On the convergence of augmented Lagrangian strategies for nonlinear programming. *IMA Journal of Numerical Analysis*, 42(2):1735–1765, 2021.

[11] R. Andreani and L. D. Secchin. A note on the convergence of an augmented Lagrangian algorithm to second-order stationary points. In *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, volume 6. Sociedade Brasileira de Matemática Aplicada e Computacional, 2018.

[12] R. Andreani, L. D. Secchin, and P. J. S. Silva. Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 28(3):2574–2600, 2018.

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

[14] M. Branda, M. Bucher, M. Červinka, and A. Schwartz. Convergence of a Scholtes-type regularization method for cardinality-constrained optimization problems with an application in sparse robust portfolio optimization. *Computational Optimization and Applications*, 70(2):503–530, 2018.

[15] M. Bucher and A. Schwartz. Second-order optimality conditions and improved convergence results for regularization methods for cardinality-constrained optimization problems. *Journal of Optimization Theory and Applications*, 178(2):383–410, Aug 2018.

[16] O. P. Burdakov, C. Kanzow, and A. Schwartz. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. *SIAM Journal on Optimization*, 26(1):397–425, 2016.

[17] M. Červinka, C. Kanzow, and A. Schwartz. Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. *Mathematical Programming*, 160(1):353–377, 2016.

[18] M. L. Flegel and C. Kanzow. Abadie-type constraint qualification for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 124(3):595–614, 2005.

[19] N. I. M. Gould and P. L. Toint. A note on the convergence of barrier algorithms to second-order necessary points. *Mathematical Programming*, 85(2):433–438, 1999.

[20] R. Janin. Directional derivative of the marginal function in nonlinear programming. In A. V. Fiacco, editor, *Sensitivity, Stability and Parametric Analysis*, volume 21 of *Mathematical Programming Studies*, pages 110–126. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.

[21] C. Kanzow, A. B. Raharja, and A. Schwartz. An augmented Lagrangian method for cardinality-constrained optimization problems. *Journal of Optimization Theory and Applications*, 189(3):793–813, 2021.

[22] C. Kanzow, A. B. Raharja, and A. Schwartz. Sequential optimality conditions for cardinality-constrained optimization problems with applications. *Computational Optimization and Applications*, 80(1):185–211, 2021.

[23] C. Kanzow and A. Schwartz. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM Journal on Optimization*, 23(2):770–798, 2013.

[24] E. H. M. Krulikovski, A. A. Ribeiro, and M. Sachine. On the weak stationarity conditions for mathematical programs with cardinality constraints: a unified approach. *Appl. Math. Optim.*, 84:3451–3473, 2021.

[25] J. M. Martínez and E. A. Pilotta. Inexact-restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications*, 104(1):135–163, 2000.

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

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

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

[29] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.

[30] A. M. Tillmann, D. Bienstock, A. Lodi, and A. Schwartz. Cardinality minimization, constraints, and regularization: A survey. June 2021.