Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints*

R. Andreani† L.D. Secchin‡ P.J.S. Silva†

April 13, 2017 (revised December 05, 2017)

Abstract

Mathematical Programs with Complementarity Constraints (MPCCs) are difficult optimization problems that do not satisfy the majority of the usual constraint qualifications (CQs) for standard nonlinear optimization. Despite this fact, classical methods behaves well when applied to MPCCs. Recently, Izmailov, Solodov and Uskov proved that first order augmented Lagrangian methods, under a natural adaption of the Linear Independence Constraint Qualification to the MPCC setting (MPCC-LICQ), converge to strongly stationary (S-stationary) points, if the multiplier sequence is bounded. If the multiplier sequence is not bounded, only Clarke stationary (C-stationary) points are recovered. In this paper we improve this result in two ways. For the case of bounded multipliers we are able replace the MPCC-LICQ assumption by the much weaker MPCC-Relaxed Positive Linear Dependence condition (MPCC-RCLPD). For the case with unbounded multipliers, building upon results from Scholtes, Anitescu, and others, we show that a *second order* augmented Lagrangian method converges to points that are at least Mordukhovich stationary (M-stationary) but we still need the more stringent MPCC-LICQ assumption. Numerical tests, validating the theory, are also presented.

Keywords: Mathematical programs with complementarity constraints, second-order methods, M-stationarity.

1 Introduction

In this paper we are interested in the *Mathematical Program with Complementarity Constraint*

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & H(x) = 0, \quad G(x) \leq 0 \\ & h(x) \geq 0, \quad g(x) \geq 0, \quad h(x)^t g(x) \leq 0, \end{aligned} \tag{MPCC}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $H : \mathbb{R}^n \to \mathbb{R}^q$, $G : \mathbb{R}^n \to \mathbb{R}^s$, and h, $g : \mathbb{R}^n \to \mathbb{R}^m$ are C^2 functions. The last (inequality) constraint, which ensure that g and h are complementary, gives name to problem.

MPCCs appear frequently in the literature. They are, for example, related to bilevel optimization problems and Stackelberg games (see [16, 41] and references therein). They have also been used by Andreani et al. to devise an alternative formulation for Order-Value Optimization problems that have applications in portfolio optimization and risk analysis [3]. MPCCs are also present in many applications like urban traffic control, economy, problems arising from the electrical sector, etc. See [18, 25, 41, 42, 51] and references therein.

*This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (Grant 2013/05475-7) and CNPq (Grants 303013/2013-3, 306986/2016-7, and 302915/2016-8).

†Institute of Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda, 651, 13083-859, Campinas, São Paulo, Brazil. E-mail: andreani@ime.unicamp.br, pjssilva@ime.unicamp.br

‡Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, Espírito Santo, Brazil. E-mail: leonardo.secchin@ufes.br

MPCCs are highly degenerate problems: they do not satisfy the majority of the established Constraint Qualifications (CQ). In particular, no feasible point conforms to Mangasarian-Fromovitz Constraint Qualification (MFCQ) and only pathological examples conform to the Relaxed Constant Positive Linear Dependency (RCPLD) condition [31]. Moreover, in the absence of strict complementarity, only Guignard's condition can be expected with certain generality [19]. Since there is not any computational method that converge to KKT points using only Guignard's condition, there is not any guarantee that traditional optimization methods applied to MPCC converge to KKT points. The main difficulty resides in the sign of the multipliers associated to the constraints g_i, h_i at points where both vanish, that is when strict complementarity fails.

Specific CQs for the MPCC, like MPCC-LICQ, were introduced to try to overcome this difficulty [24, 27, 28, 34, 46, 48, 50, 54]. Moreover, different notions of stationarity, allowing for "wrong" signs of the multipliers whenever strict complementarity fails, were also introduced. The main stationarity concepts for MPCC, in order of strength, are W/C/M/S-stationarity [48], see Figure 1. Using these concepts the behavior of general nonlinear optimization algorithms for the MPCC was studied. For example Interior Point Methods [37], penalty approaches [30, 31, 52, 53], and Sequential Quadratic Programming [23]. Such methods can only recover C-stationary points. Particularly, the SQP method can converge to any feasible point even under MPCC-LICQ [23].

Also, many specific methods for the MPCC were developed, see [17, 32, 33, 35, 38, 39, 49]. Such methods have good convergence properties, converging to M-stationary points, but require exact computation of KKT points in the subproblems which is not computationally feasible. If one relaxes such exactness requirements, only C-stationary points can be ensured unless strict extra assumptions are used [36]. Such convergence is not better than what can be achieved with classical nonlinear optimization algorithms. This lead some authors to put in check the efficiency of such specific methods when compared to general algorithms [22, 28].

In this work we are specially interested in the results from Izmailov, Solodov and Uskov showing the augmented Lagrangians methods can only converge to C-stationary points when applied to MPCC [2, 31]. To illustrate this consider the bidimensional MPCC

$$\min \frac{1}{2}\left[(x-1)^2+(y-1)^2\right] \ \text{ s.t. } \ x \geq 0, \ y \geq 0, \ xy \leq 0.$$

Its global minima are $(1,0)$ and $(0,1)$. The origin is a C-stationary point where MPCC-LICQ holds, however it is a strict local maximizer that may be approximated by a sequence computed by some augmented Lagrangian methods. In fact, this is the problem `scholtes3` of MacMPEC collection maintained by Sven Leyffer, and in Section 4 we observe this behavior with Algencan [2] algorithm (see Table 1).

In this paper we show that the second order augmented Lagrangian method from Andreani et al. [1], called Algencan-second, is able to ensure convergence to points that are at least M-stationary, a concept much stronger than C-stationarity. In particular, in the example above, the origin is not M-stationary and hence Algencan-second is able to avoid it. This is a case where the second order information is able to unveil better first-order stationary points by better controlling the sign of the multipliers. We note that it is not the first time in the literature that second order information is used to obtain algorithmic theoretical convergence to M-stationary points. In fact, in a seminar paper, Scholtes [49] used it to show that his regularization scheme converges to M-stationarity points under MPCC-LICQ. Later on, in two important papers, Anitescu showed a similar result using an "elastic mode" approach [8, 9]. This result was then improved, in collaboration with Tseng and Wright, in [10], where an algorithm which allows for inexact second order computations that still converges do M-stationary points still using MPCC-LICQ was presented. Second order information was also considered in a partially-augmented Lagrangian approach in [40]. We improve upon this results by establishing the convergence properties for a "pure"augmented Lagrangian method, in which all the constraints can be penalized. This is the situation especially well suited for nonlinear constraints. We also were able to replace the MPCC-LICQ condition by the much weaker MPCC-RCPLD whenever the multiplier sequence is bounded. MPCC-RCPLD holds whenever the contraints are linear without any further assumption, hence it is adequate to unify results for the cases of nonlinear and linear constraints.

This paper is organized as follows: Section 2 revises basic concepts about the MPCC and augmented Lagrangian methods. Section 3 presents the convergence of Algencan-second for

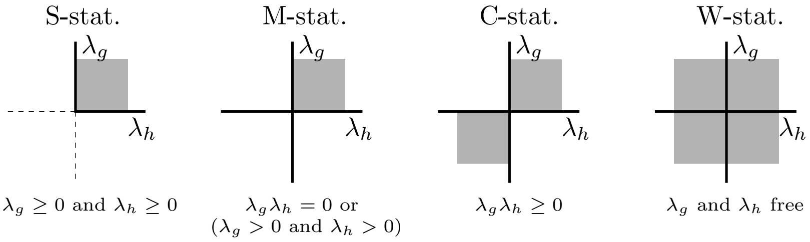

Figure 1: Different notions of stationarity. λ_g and λ_h are the multipliers of the active constraints $g(x) \geq 0$ and $h(x) \geq 0$.

the MPCC. Section 4 presents some numerical tests that validate the convergence results and, finally, Section 5 closes with some final remarks and future directions.

Notation: The symbol $\|\cdot\|$ will denote the Euclidean norm. If $q : \mathbb{R}^n \to \mathbb{R}^m$ then we write $\nabla q(x)$ to denote the $n \times m$ matrix whose columns are $\nabla q_i(x)$, $i = 1, \ldots, m$. We denote by $\lambda_0(A)$ the smallest eigenvalue of a symmetric matrix A. Given a set S of vectors, $\operatorname{span} S$ is the space spanned by the vectors of S, and $\operatorname{span} S^\perp$ is its orthogonal (Euclidean) complement. If $z \in \mathbb{R}^n$, the components of z_+ are defined by $(z_+)_i = \max\{0, z_i\}$, $i = 1, \ldots, n$.

2 Basic concepts

2.1 Augmented Lagrangian methods

Let us consider the general nonlinear optimization problem

$$\min F(x) \;\text{ s.t. }\; H(x) = 0,\; G(x) \leq 0,\; x \in \Omega, \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}$, $H : \mathbb{R}^n \to \mathbb{R}^q$ and $G : \mathbb{R}^n \to \mathbb{R}^s$ are smooth functions and $\Omega \subset \mathbb{R}^n$ is compact. We denote the index set of active inequality constraints at x by $I_G(x) = \{i \mid G_i(x) = 0\}$. The Lagrangian function associated with this problem is defined by

$$L(x, \mu) = F(x) + (\mu^G)^t G(x) + (\mu^H)^t H(x)$$

for all $x \in \Omega$ and $\mu = (\mu^G, \mu^H) \in \mathbb{R}^s_+ \times \mathbb{R}^q$. We will also consider the (Powell-Hestenes-Rockafellar) augmented Lagrangian [26, 45, 47], defined in

$$L_\rho(x, \mu) = F(x) + \frac{\rho}{2} \left\{ \left\| H(x) + \frac{\mu^H}{\rho} \right\|^2 + \left\| \left(G(x) + \frac{\mu^G}{\rho} \right)_+ \right\|^2 \right\},$$

where $x \in \Omega$, $\mu = (\mu^G, \mu^H) \in \mathbb{R}^s_+ \times \mathbb{R}^q$ and $\rho > 0$.

One of the most important classes of algorithms to solve (1) are the augmented Lagrangian methods. These methods are based on successive (partial) minimizations of the augmented Lagrangian function with a fixed multiplier followed by a multiplier update. Among them, we are particularly interested in the ALGENCAN algorithm, an augmented Lagrangian method developed in [2]. ALGENCAN is the basis for the second order augmented Lagrangian method named ALGENCAN-SECOND [1].

Both variations consider that the abstract set Ω only contains box constraints, as these are simple enough to be handled by the inner minimization procedure. Thus, we assume from now on that Ω only contains box constraints, i.e., $\Omega = \{x \mid \ell \leq x \leq u\}$, and the augmented Lagrangian subproblems consist in minimizing $L_\rho(\cdot, \mu)$ over Ω for fixed μ. To solve these subproblems ALGENCAN uses GENCAN [13], a box constrained solver based on an active-set strategy and in Spectral Projected Gradient (SPG) steps, while ALGECAN-SECOND uses a variation of GENCAN that is able to deal with directions of negative curvature [1]. In the rest of this section, we briefly introduce both variations of the ALGENCAN algorithm.

Let us define the *continuous projected gradient* as

$$\mathcal{G}_P(x) = P_\Omega\left(x - \nabla_x L_\rho(x,\mu)\right) - x,$$

where $P_\Omega(\cdot)$ is the orthogonal projection onto Ω. The multiplier μ and the penalty parameter ρ are fixed, and they will be clear from the context. When the continuous projected gradient is nonzero, it is a descent direction for the augmented Lagrangian function $L_\rho(\cdot,\mu)$, and it is feasible with respect to the box Ω (see Figure 2). Thus $\|\mathcal{G}_P(\cdot)\|$ is a measure for the first order (KKT) optimality of the augmented Lagrangian subproblem, which is used in step 1 of the ALGENCAN algorithm below.

Algorithm 1 ALGENCAN

Let $\mu^H_{\min} < \mu^H_{\max}$, $\mu^G_{\max} > 0$, $\gamma > 1$, $0 < \tau < 1$ and $\{\varepsilon_k\}$ be a sequence of positive scalars with $\lim_{k\to\infty} \varepsilon_k = 0$. Let $(\mu^H)^1_i \in [\mu^H_{\min}, \mu^H_{\max}]$, $i = 1,\ldots,q$, $(\mu^G)^1_i \in [0, \mu^G_{\max}]$, $i = 1,\ldots,s$, and $\rho_1 > 0$. Initialize $k \leftarrow 1$.

Step 1. Find, using GENCAN, an approximate minimizer x^k of the problem $\min_{x\in\Omega} L_{\rho_k}(x,\mu^k)$ such that

$$\left\|\mathcal{G}_P(x^k)\right\| \le \varepsilon_k.$$

Step 2. Define

$$V_i^k = \max\left\{G_i(x^k), -\frac{(\mu^G)^k_i}{\rho_k}\right\}, \quad i = 1,\ldots,s.$$

If $k > 1$ and $\max\{\|H(x^k)\|_\infty, \|V^k\|_\infty\} \le \tau \max\{\|H(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}$, define $\rho_{k+1} = \rho_k$. Otherwise, define $\rho_{k+1} = \gamma\rho_k$.

Step 3. Compute $(\mu^H)^{k+1}_i \in [\mu^H_{\min}, \mu^H_{\max}]$, $i = 1,\ldots,q$ and $(\mu^G)^{k+1}_i \in [0, \mu^G_{\max}]$, $i = 1,\ldots,s$. Take $k \leftarrow k+1$ and go to the Step 1.

In order to have more stability on solving subproblems, the penalty parameter ρ is increased only if feasibility and complementarity measures do not improve (step 2 of Algorithm 1). In step 3, a practical choice of new multipliers estimates is to project $(\mu^H)^k$ and $(\mu^G)^k$ onto boxes $[\mu^H_{\min}, \mu^H_{\max}]^q$ and $[0, \mu^G_{\max}]^s$, respectively.

Originally, the convergence theory of ALGENCAN was established under the *Constant Positive Linear Dependence* (CPLD) constraint qualification [2], which states that if a subset of gradients of the active constraints is positive linearly dependent (PLD) at a feasible x^* then these gradients remain LD at the points in a neighborhood of x^*. Recently, this result was improved with the introduction of the *Cone Continuity Property* (CCP) [7]: we say that a feasible x^* conforms to CCP if the multifunction $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by

$$K(x) = \left\{\nabla H(x)\mu^H + \nabla G(x)\mu^G \mid \mu^H \in \mathbb{R}^q,\ \mu^G \in \mathbb{R}^s_+,\ \mu^G_j = 0 \text{ for } j \notin I_G(x^*)\right\}$$

is outer semicontinuous at x^*, i.e., if $\limsup_{x\to x^*} K(x) \subset K(x^*)$. We note that the KKT conditions can be written as $-\nabla F(x^*) \in K(x^*)$. In [7], the authors showed that CCP is the weakest constraint qualification that ensures that an *Approximate KKT* (AKKT) [5] point is KKT. As ALGENCAN generates AKKT sequences [6], its convergence is automatically established with the CCP condition.

Theorem 2.2. *Let $\{x^k\}$ be a sequence generated by Algorithm 1. Then,*

- *$\{x^k\}$ admits at least one limit point and any limit x^* is a KKT point of the infeasibility problem*

$$\min \|H(x)\|^2 + \|G(x)_+\|^2 \ \ s.t.\ x \in \Omega. \tag{2}$$

- *If a limit point x^* is feasible and satisfies the CCP condition then it is a KKT point.*

For more details on ALGENCAN, see [12, 11, 14].

Next we present ALGENCAN-SECOND and GENCAN-SECOND algorithms. For all $x \in \Omega$, we define

$$\mathcal{F}(x) = \{z \in \Omega \mid z_i = \ell_i, \text{ if } x_i = \ell_i, \ z_i = u_i, \text{ if } x_i = u_i, \ \ell_i < z_i < u_i, \text{ otherwise}\}.$$

Geometrically, $\mathcal{F}(x)$ is the smallest face of Ω containing x, and we say that it is the *open face* to which x belongs. The variables z_i such that $\ell_i < z_i < u_i$ are the *free* variables, while the remaining are called *fixed* variables. Let $\mathcal{V}(x)$ be the minimal affine subspace that contains $\mathcal{F}(x)$, and $\mathcal{S}(x)$ the parallel subspace to $\mathcal{V}(x)$. The dimension of $\mathcal{F}(x)$, denoted by $\dim \mathcal{F}(x)$, is the dimension of $\mathcal{S}(x)$ and coincides with the number of free variables in the open face $\mathcal{F}(x)$. We define the *inner projected gradient* as

$$\mathcal{G}_I(x) = P_{\mathcal{S}(x)}\left(\mathcal{G}_P(x)\right).$$

This vector plays a similar role to the continuous projected gradient. It is a measure for the first order (KKT) optimality of the augmented Lagrangian subproblem, but now, within the open face. Figure 2 illustrates the geometry of the gradients $\mathcal{G}_P$ and $\mathcal{G}_I$.

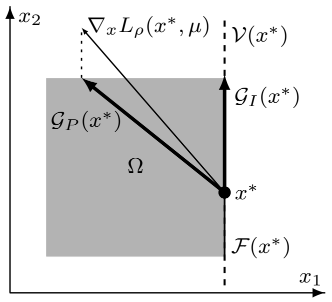

Figure 2: Geometry of $\mathcal{G}_P$ and $\mathcal{G}_I$. The open face $\mathcal{F}(x^*)$ is the right border of Ω, where x_1 is fixed and x_2 is free. If $\mathcal{G}_P(x^*) \neq 0$ then $\mathcal{G}_P(x^*)$ is a feasible descent direction for $L_\rho(\cdot, \mu)$ at x^*. On the other hand, $\mathcal{G}_I(x^*) \neq 0$ is a feasible descent direction within $\mathcal{F}(x^*)$.

The Hessian of L_ρ may not be defined at points where a component of $\mu^G + \rho G$ is zero. We then define, for each $x \in \Omega$ and $\varepsilon > 0$, the *approximate ε-Hessian* of L_ρ (with respect to x) as

$$\begin{aligned}\nabla^2_\varepsilon L_\rho(x, \mu) = \nabla^2 F(x) + &\sum_{i=1}^{q} \left(\mu_i^H + \rho H_i(x)\right) \nabla^2 H_i(x) + \rho \sum_{i=1}^{q} \nabla H_i(x) \nabla H_i(x)^t \\ + &\sum_{i=1}^{s} \left(\mu_i^G + \rho G_i(x)\right)_+ \nabla^2 G_i(x) + \rho \sum_{i \in I_\varepsilon(x)} \nabla G_i(x) \nabla G_i(x)^t\end{aligned}$$

where $I_\varepsilon(x)$ is a relaxed version of the index set of zero components of $\mu^G + \rho G$, defined by

$$I_\varepsilon(x) = \left\{ j \;\middle|\; \frac{1}{\sqrt{\rho}} \left(\mu_j^G + \rho G_j(x)\right) \geq -\varepsilon \right\}.$$

Observe that $\nabla^2_0 L_\rho(x, \mu)$ is the true Hessian of L_ρ where it exists. At these points the eigenvalues of $\nabla^2_\varepsilon L_\rho$ give upper bounds to the eigenvalues of the true Hessian. Furthermore, if the true Hessian of L_ρ is positive semidefinite then $\nabla^2_\varepsilon L_\rho$ is also positive semidefinite.

Now, let $\mathcal{F}$ be an open face and $\varepsilon > 0$. For each $x \in \mathcal{F}$, $\rho > 0$ and $\mu \in \mathbb{R}^q \times \mathbb{R}^s_+$, we define the *reduced ε-Hessian* $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x, \mu)$ as the matrix whose entry (i, j) is the entry (i, j) of $\nabla^2_\varepsilon L_\rho(x, \mu)$ if x_i and x_j are free in $\mathcal{F}$, and the entry (i, j) of the identity matrix otherwise. The matrix $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x, \mu)$ gives a second order optimality measure for the augmented Lagrangian subproblem within the open face $\mathcal{F}$. In fact, if x_i is fixed in $\mathcal{F}$ then the directions pointing out the affine subspace $\mathcal{V}$ have a nonzero i-component (see Figure 2). By the construction of $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x, \mu)$, this i-component does not affect its positive semidefiniteness.

We are now ready to state the second order algorithms that will play a pivotal role in this paper (Algorithms 2 and 3). Compared with ALGENCAN, step 1 of Algorithm 2 contains an extra second order condition which attests that the current point x^k is a vertex of the box Ω (i.e., $\dim \mathcal{F}(x^k) = 0$) or satisfy approximately a necessary second order optimality condition for the augmented Lagrangian subproblem (namely, WSONC, defined below). In step 2 of Algorithm 3, we abandon the current face only when the first order optimality measure $\|\mathcal{G}_I(x^k)\|$, within the face, is small enough compared to the optimality measure $\|\mathcal{G}_P(x^k)\|$ on the entire box Ω (condition (3)), and additionally, when a second order optimality condition is satisfied. That is, we explore the current open face $\mathcal{F}(x^k)$ to exhaustion before moving to another face.

Algorithm 2 ALGENCAN-SECOND

Let $\mu^H_{\min} < \mu^H_{\max}$, $\mu^G_{\max} > 0$, $\gamma > 1$ and $0 < \tau < 1$. Let $\{\varepsilon_k^{\text{fun}}\}$, $\{\varepsilon_k^{\text{grad}}\}$ and $\{\varepsilon_k^{\text{hess}}\}$ be sequences of positive scalars such that $\lim_{k\to\infty} \varepsilon_k^{\text{fun}} = \lim_{k\to\infty} \varepsilon_k^{\text{grad}} = \lim_{k\to\infty} \varepsilon_k^{\text{hess}} = 0$. Let $(\mu^H)_i^1 \in [\mu^H_{\min}, \mu^H_{\max}]$, $i = 1, \ldots, q$, $(\mu^G)_i^1 \in [0, \mu^G_{\max}]$, $i = 1, \ldots, s$, and $\rho_1 > 0$. Initialize $k \leftarrow 1$.

Step 1. Find an approximate minimizer x^k of the problem $\min_{x\in\Omega} L_{\rho_k}(x, \mu^k)$ using Algorithm 3 (GENCAN-SECOND) with $\overline{\rho} = \rho_k$, $\overline{\mu} = \mu^k$, $\varepsilon^{\text{fun}} = \varepsilon_k^{\text{fun}}$, $\varepsilon^{\text{grad}} = \varepsilon_k^{\text{grad}}$ and $\varepsilon^{\text{hess}} = \varepsilon_k^{\text{hess}}$. The iterate x^k satisfies

$$\|\mathcal{G}_P(x^k)\| \leq \varepsilon_k^{\text{grad}}$$

and

$$\dim \mathcal{F}(x^k) = 0 \quad \text{or} \quad \lambda_0 \left(\mathcal{H}_{[\mathcal{F}(x^k), \varepsilon_k^{\text{fun}}, \rho_k]}(x^k, \mu^k)\right) \geq -\varepsilon_k^{\text{hess}}.$$

Step 2. Define

$$V_i^k = \max\left\{G_i(x^k), -\frac{(\mu^G)_i^k}{\rho_k}\right\}, \quad i = 1, \ldots, s.$$

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 $(\mu^H)_i^{k+1} \in [\mu^H_{\min}, \mu^H_{\max}]$, $i = 1, \ldots, q$ and $(\mu^G)_i^{k+1} \in [0, \mu^G_{\max}]$, $i = 1, \ldots, s$. Take $k \leftarrow k + 1$ and go to the Step 1.

As in ALGENCAN, the convergence analysis of ALGENCAN-SECOND considers separately feasible and infeasible limit points. With respect to the first order stationarity, we naturally have the same results from ALGENCAN. In what follows, we discuss the convergence to second order stationary points. If x^* is a KKT point with multipliers $\mu = (\mu^G, \mu^H) \in \mathbb{R}^s_+ \times \mathbb{R}^q$ we say that x^* satisfies the *Weak Second Order Necessary Condition* (WSONC) if the Hessian of the Lagrangian is positive semidefinite in the subspace orthogonal to gradients of the active constraints, i.e., if

$$d^t \nabla^2 L(x^*, \mu) d \geq 0, \quad \forall d \in \mathbb{R}^n \quad \text{such that} \quad \nabla G_{I_G(x^*)}(x^*) d = 0, \quad \nabla H(x^*) d = 0.$$

Recently, a new second order constraint qualification was introduced in [4]. It is called *Second Order Cone Continuity Property* (CCP2) and it can be viewed as an adaptation of CCP to take into account second order information. For each $x \in \mathbb{R}^n$, let us consider the cone

$$\mathcal{C}(x, x^*) = \left\{d \in \mathbb{R}^n \mid \nabla H_i(x)^t d = 0, i = 1, \ldots, q, \; \nabla G_j(x)^t d = 0, j \in I_G(x^*)\right\}.$$

The set $\mathcal{C}(x, x^*)$ can be viewed as a perturbation of the weak critical cone $\mathcal{C}(x^*) := \mathcal{C}(x^*, x^*)$ around a feasible point x^*. As in CCP, we can write WSONC in the compact form

$$\left(-\nabla F(x^*), -\nabla^2 F(x^*)\right) \in K_2(x^*)$$

where $K_2(x)$ denotes the convex cone

$$K_2(x) = \bigcup_{\substack{(\mu^H, \mu^G) \in \mathbb{R}^q \times \mathbb{R}^s_+, \\ \mu_j^G = 0 \text{ for } j \notin I_G(x^*)}} \left\{ \begin{array}{l} \left(\nabla H(x)\mu^H + \nabla G(x)\mu^G, \; \mathcal{M}\right) \text{ such that} \\ -\mathcal{M} + \displaystyle\sum_{i=1}^q \mu_i^H \nabla^2 H_i(x) + \sum_{j \in I_G(x^*)} \mu_j^G \nabla^2 G_j(x) \\ \text{is positive semidefinite on } \mathcal{C}(x, x^*) \end{array} \right\}.$$

Algorithm 3 GENCAN-SECOND

Let $x^0 \in \Omega$ be an approximate initial solution of the problem

$$\min_{x\in\Omega} L_{\overline{\rho}}(x, \overline{\mu})$$

($\overline{\mu}$ and $\overline{\rho}$ are fixed here). Assume that $\eta \in (0,1)$, $\varepsilon^{\rm fun}, \varepsilon^{\rm grad}, \varepsilon^{\rm hess} > 0$ and $\varepsilon^{\rm curv} \in (0, \varepsilon^{\rm hess})$. Initialize $k \leftarrow 1$.

Step 1. (Stopping criterion) If $\|\mathcal{G}_P(x^k)\| \leq \varepsilon^{\rm grad}$ and

$$\dim \mathcal{F}(x^k) = 0 \quad \text{or} \quad \lambda_0\left(\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\rm fun},\overline{\rho}]}(x^k, \overline{\mu})\right) \geq -\varepsilon^{\rm hess},$$

stop declaring convergence.

Step 2. (Decision about keeping or abandoning the current face) If

$$\|\mathcal{G}_I(x^k)\| \leq \eta \|\mathcal{G}_P(x^k)\| \tag{3}$$

and

$$\dim \mathcal{F}(x^k) = 0 \quad \text{or} \quad \lambda_0\left(\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\rm fun},\overline{\rho}]}(x^k, \overline{\mu})\right) \geq -\varepsilon^{\rm curv},$$

compute $x^{k+1} \in \Omega$ using the SPG method [15] (abandon the current face). Otherwise, compute x^{k+1} by [1, Algorithm 2.3] (inner iteration).

Step 3. Take $k \leftarrow k+1$ and go to the Step 1.

Finally, we say that a feasible x^* conforms to CCP2 if the multifunction $K_2 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semicontinuous at x^*. As in CCP, CCP2 is weakest constraint qualification that guarantees that an *Approximate Second Order Stationary Point* (AKKT2) point fulfills WSONC (see [4] and its Theorem 4.2). As ALGENCAN-SECOND generates AKKT2 sequences [4] (and consequently an AKKT sequence), the next result was established generalizing [1, Theorem 2].

Theorem 2.3. *Let $\{x^k\}$ be a sequence generated by Algorithm 2. Then*

- *$\{x^k\}$ admits at least one limit point and any limit x^* is a KKT point of the infeasibility problem (2).*
- *If a limit x^* is feasible and satisfies CCP then x^* is a KKT point.*
- *If a limit x^* is feasible and satisfies CCP2 then x^* is a KKT point and fulfills WSONC.*

2.4 Stationarity concepts and constraint qualifications for MPCC

For the sake of simplicity, we will omit in this section the box Ω and the non-complementary constraints $G(x) \leq 0$, $H(x) \leq 0$. The adaptations to take this extra constraints into account can be easily carried out by the reader. Therefore, we will focus on following particular case of (MPCC):

$$\min f(x) \ \text{ s.t. } \ g(x) \geq 0, \ h(x) \geq 0, \ g(x)^t h(x) \leq 0.$$

Its usual Lagrangian is

$$L(x, \mu) = f(x) - (\mu^g)^t g(x) - (\mu^h)^t h(x) + \mu_0 g(x)^t h(x)$$

and the *MPCC-Lagrangian* is defined by

$$\mathcal{L}(x, \lambda) = f(x) - (\lambda^g)^t g(x) - (\lambda^h)^t h(x).$$

For a feasible x we define the sets of indexes of the active constraints

$$I_g(x) = \{i \mid g_i(x) = 0\}, \quad I_h(x) = \{i \mid h_i(x) = 0\} \quad \text{and} \quad I_0(x) = I_g(x) \cap I_h(x).$$

We note that $I_g(x) \cup I_h(x) = \{1, \ldots, m\}$. For simplicity we denote $I_g = I_g(x^*)$, $I_h = I_h(x^*)$ and $I_0 = I_0(x^*)$ whenever x^* is clear from the context.

As we already mentioned, MPCCs do not satisfy the majority of the established CQs, not even the Abadie's condition [19]. Hence, specific constraint qualifications are defined. Among the various CQs, we present two of our interest.

Definition 2.5 ([48]). *We say that x^* satisfies the* MPCC-Linear Independence Constraint Qualification *(MPCC-LICQ) if the gradients*

$$\nabla g_i(x^*), i \in I_g(x^*), \quad \nabla h_i(x^*), i \in I_h(x^*)$$

are linearly independent.

It is known that MPCC-LICQ implies the classical Guignard's condition, but the same does not occur with the sightly less stringent MPCC-MFCQ [19].

Next, we present the extension of RCPLD [6] to the MPCC that was introduced in [24].

Definition 2.6. *Let $\mathcal{I}_g \subset I_g \backslash I_h$ and $\mathcal{I}_h \subset I_h \backslash I_g$ (at x^*) such that*

$$\nabla g_i(x^*),\ i \in \mathcal{I}_g, \quad \nabla h_i(x^*),\ i \in \mathcal{I}_h$$

is a basis for

$$\operatorname{span}\{\nabla g_i(x^*), \nabla h_j(x^*) \mid i \in I_g \backslash I_h,\ j \in I_h \backslash I_g\}.$$

We say that x^* satisfies the MPCC-Relaxed Constant Positive Linear Dependence *(MPCC-RCPLD) constraint qualification if there is a open neighborhood $\mathcal{N}(x^*)$ of x^* such that*

- *$\{\nabla g_i(x), \nabla h_j(x) \mid i \in I_g \backslash I_h,\ j \in I_h \backslash I_g\}$ has the same rank for all $x \in \mathcal{N}(x^*)$;*
- *for each $\tilde{I}_g, \tilde{I}_h \subset I_0$, whenever there are multipliers μ^g, μ^h, not all zeros, satisfying ($\mu_i^h \mu_i^g = 0$ or $\mu_i^h, \mu_i^g > 0$) $\forall i \in I_0$ and*

$$\sum_{i \in \mathcal{I}_g \cup \tilde{I}_g} \mu_i^g \nabla g_i(x^*) + \sum_{i \in \mathcal{I}_h \cup \tilde{I}_h} \mu_i^h \nabla h_i(x^*) = 0,$$

  then

$$\{\nabla g_i(x), \nabla h_j(x) \mid i \in \mathcal{I}_g \cup \tilde{I}_g,\ j \in \mathcal{I}_h \cup \tilde{I}_h\}$$

  is linearly dependent for each $x \in \mathcal{N}(x^*)$.

Just like in nonlinear optimization, these MPCC CQs are important to assert the validity of first order stationarity conditions akin to KKT. However, in MPCC there are different notions of stationarity ([41, 43, 48]) that we describe below.

Definition 2.7. *We say that a feasible point x of MPCC is* weakly stationary *(W-stationary) if there is $\lambda = (\lambda^g, \lambda^h)$ such that $\nabla_x \mathcal{L}(x, \lambda) = 0$, $\lambda^g_{I_h(x) \backslash I_g(x)} = 0$ and $\lambda^h_{I_g(x) \backslash I_h(x)} = 0$.*

Other stationarity concepts are usual in the literature, and deal with the different possibilities for the signs of the nonzero multipliers.

Definition 2.8. *Let x be a W-stationary point with associated multipliers $\lambda = (\lambda^g, \lambda^h)$.*

- *If $\lambda_i^g \lambda_i^h \geq 0$ for all $i \in I_0(x)$ then we say that x is* Clarke stationary *(C-stationary);*
- *If for all $i \in I_0(x)$ we have $\lambda_i^g \lambda_i^h = 0$ or $\lambda_i^g > 0$, $\lambda_i^h > 0$ then we say that x is* Mordukhovich stationary *(M-stationary);*
- *If $\lambda^g_{I_0(x)} \geq 0$ and $\lambda^h_{I_0(x)} \geq 0$ then we say that x is* strongly stationary *(S-stationary).*

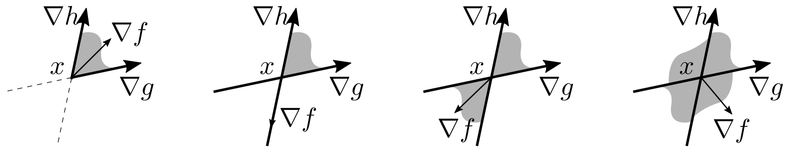

Figure 3: Geometry of the stationarity concepts. The constraints $g(x), h(x) \geq 0$ are active at x, and the equation $\nabla\mathcal{L}(x,\lambda) = 0$ is equivalent to $\nabla f(x) = \lambda^g \nabla g(x) + \lambda^h \nabla h(x)$. The signs of the multipliers λ^g and λ^h determine the regions where $\nabla f(x)$ can be. More stringent concepts result in smaller regions.

Clearly S-stationarity ⇒ M-stationarity ⇒ C-stationarity ⇒ W-stationarity. Figure 3 illustrates the geometry of these concepts. When the lower level strict complementarity does not hold (i.e., when $I_0(x) \neq \emptyset$), these concepts are not equivalent in general. In the example of the introduction, the local maximizer $(0,0)$ is a C-stationary point, but not M-stationary. In the problem

$$\min \frac{1}{2}\left[(x-1)^2 + 4y^2\right] \text{ s.t. } x \geq 0,\ y \geq 0,\ xy \leq 0$$

the origin is a M-stationary point, but not S-stationary.

M-stationarity is not a necessary optimality condition. However, if x^* is a local minimizer of MPCC and fulfills MPCC-Guignard's condition then x^* is a M-stationary point [20, 21] (i.e., "M-stationary or not MPCC-Guignard" is a necessary optimality condition). MPCC-Guignard is a variant of the classical Guignard's condition, and is the weakest known constraint qualification for M-stationarity (for a detailed description of relations between constraint qualifications for M-stationarity, see [24, 46]). The same does not occur with S-stationarity (which is equivalent to KKT, see [19, Proposition 4.2]): even MPCC-MFCQ does not guarantee that local minimizers are S-stationary points [48]. In this sense, it seems natural to build numerical methods to search for M-stationarity points, as they are guaranteed to exist under reasonable (weak) constraint qualifications. Furthermore, a M-stationary point is not S-stationary only if $\lambda_i^g \lambda_i^h = 0$ for some index in $I_0(x)$. From a numerical point of view, this situation seems to be atypical, and theoretical convergence results to M-stationary points seem reasonable.

3 Convergence results for MPCCs

In this section, we improve the convergence results for the (first order) augmented Lagrangian methods applied to MPCC presented in [31]. We also establish the corresponding result for Algencan-second [1]. As before, for the sake of simplicity, we will omit the constraints $G(x) \leq 0$ and $H(x) \leq 0$. As we already mentioned, Algencan and Algencan-second treat box constraints separately, without penalizing them. When the box Ω is present, the reduced ε-Hessian $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu)$ in Algencan-second and the approximate ε-Hessian $\nabla^2_\varepsilon L_\rho(x,\mu)$ have different entries related to fixed variables in the current open face (see Section 2.1). Hence, once again for simplicity we also omit the box Ω, which removes this difficulty ensuring that $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu) = \nabla^2_\varepsilon L_\rho(x,\mu)$. The adaptations of this theory to the general case are simple and mostly technical, and will be treated separately in remarks that follow the results and in Section 3.3.

Izmailov, Solodov and Uskov [31] showed that under MPCC-LICQ, augmented Lagrangian methods (in particular Algencan) converge to points that are at least C-stationary. The authors showed that when a certain dual sequence is bounded, augmented Lagrangian methods converge to S-stationary points under MPCC-LICQ. We improve last statement showing that the same occurs with the weaker MPCC-RCPLD condition. We apply the same technique that was used in [6].

Theorem 3.1. *Let $\{x^k\}$ be a sequence generated by* Algencan *(Algorithm 1) and x^* be a feasible limit point with associated infinite index set K, i.e. $\lim_{k\in K} x^k = x^*$. If the sequence of complementarity multipliers is such that $\liminf_{k\in K}(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) < \infty$ and MPCC-RCPLD holds at x^*, then x^* is S-stationary. Otherwise, if MPCC-LICQ holds at x^*, it is at least C-stationary.*

Proof. We only need to prove the first assertion of the theorem. The second statement follows from the proof of [31, Theorem 3.2]. Hence, we assume from now on that MPCC-RCPLD holds at x^* and that $\liminf_{k\in K}(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) < \infty$.

ALGENCAN generates a sequence $\{x^k, \mu^k, \rho_k\}$ satisfying

$$\nabla L_{\rho_k}(x^k, \mu^k) = \nabla f(x^k) - \sum_{i=1}^m \left((\mu_i^g)^k - \rho_k g_i(x^k)\right)_+ \nabla g_i(x^k) - \sum_{i=1}^m \left((\mu_i^h)^k - \rho_k h_i(x^k)\right)_+ \nabla h_i(x^k) + \left(\mu_k^0 + \rho_k g(x^k)^t h(x^k)\right)_+ v^k \to 0, \quad (4)$$

where $v^k = \sum_{i=1}^m \left(\nabla g_i(x^k) h_i(x^k) + \nabla h_i(x^k) g_i(x^k)\right)$. We can write

$$\nabla f(x^k) - \sum_{i=1}^m (\lambda_i^g)^k \nabla g_i(x^k) - \sum_{i=1}^m (\lambda_i^h)^k \nabla h_i(x^k) \to 0, \quad (5)$$

where

$$(\lambda_i^g)^k = \left((\mu_i^g)^k - \rho_k g_i(x^k)\right)_+ - \lambda_k^0 h_i(x^k), \quad (6)$$
$$(\lambda_i^h)^k = \left((\mu_i^h)^k - \rho_k h_i(x^k)\right)_+ - \lambda_k^0 g_i(x^k), \quad (7)$$

and $\lambda_k^0 = (\mu_k^0 + \rho_k g(x^k)^t h(x^k))_+$. We will call $\{\lambda_k^0\}$ the sequence of complementarity multipliers.

If $\{\rho_k\}$ is bounded, it is straightforward, by taking limits in K, to show that x^* is a KKT point or, in other words, S-stationary. Thus, we restrict our attention to the case $\rho_k \to \infty$. As we are under the assumption that $\liminf_{k\in K} \lambda_k^0 < \infty$, there is an infinite index set $K_1 \subset K$ where $\{\lambda_k^0\}$ is bounded. Thus, (6) and (7) imply that $\lim_{k\in K_1}(\lambda^g_{I_h\backslash I_g})^k = 0$ and $\lim_{k\in K_1}(\lambda^h_{I_g\backslash I_h})^k = 0$. Hence, (5) may be rewritten as

$$\lim_{k\in K_1}\left[\nabla f(x^k) - \sum_{i\in I_g}(\lambda_i^g)^k \nabla g_i(x^k) - \sum_{i\in I_h}(\lambda_i^h)^k \nabla h_i(x^k)\right] = 0. \quad (8)$$

Again, by expressions (6) and (7) we have

$$\lim_{k\in K_1}(\lambda^g_{I_0})^k \geq 0 \quad \text{and} \quad \lim_{k\in K_1}(\lambda^h_{I_0})^k \geq 0. \quad (9)$$

The last expression allow us to perturb $(\lambda^g_{I_0})^k$ and $(\lambda^h_{I_0})^k$ in order to obtain 8 with $(\lambda^g_{I_0})^k \geq 0$ and $(\lambda^h_{I_0})^k \geq 0$ for all $k \in K_1$ large enough. Consider $\mathcal{I}_g \subset I_g\backslash I_h$ and $\mathcal{I}_h \subset I_h\backslash I_g$ as in the definition of MPCC-RCPLD, i.e., such that

$$\mathcal{B}(x^*) = \{\nabla g_i(x^*), \nabla h_j(x^*) \mid i \in \mathcal{I}_g,\ j \in \mathcal{I}_h\}$$

is a basis to

$$\mathcal{S}(x^*) = \operatorname{span}\{\nabla g_i(x^*), \nabla h_j(x^*) \mid i \in I_g\backslash I_h,\ j \in I_h\backslash I_g\}.$$

From the first condition of MPCC-RCPLD, $\mathcal{B}(x^k)$ is a basis to $\mathcal{S}(x^k)$ for all $k \in K_1$ sufficiently large. Let us say for all $k \in K_2 \subset K_1$. Then, there are sequences $\{(\hat\lambda^g_{\mathcal{I}_g})^k\}$ and $\{(\hat\lambda^h_{\mathcal{I}_h})^k\}$ such that

$$\sum_{i\in I_g\backslash I_h}(\lambda_i^g)^k \nabla g_i(x^k) + \sum_{i\in I_h\backslash I_g}(\lambda_i^h)^k \nabla h_i(x^k) = \sum_{i\in\mathcal{I}_g}(\hat\lambda_i^g)^k \nabla g_i(x^k) + \sum_{i\in\mathcal{I}_h}(\hat\lambda_i^h)^k \nabla h_i(x^k),$$

and (8) gives

$$\lim_{k\in K_2}\left[\nabla f(x^k) - \sum_{i\in\mathcal{I}_g}(\hat\lambda_i^g)^k \nabla g_i(x^k) - \sum_{i\in\mathcal{I}_h}(\hat\lambda_i^h)^k \nabla h_i(x^k) - \sum_{i\in I_0}(\lambda_i^g)^k \nabla g_i(x^k) - \sum_{i\in I_0}(\lambda_i^h)^k \nabla h_i(x^k)\right] = 0.$$

The gradients of the first two sums are linearly independent and the multipliers of the last two sums are nonnegative. By [6, Lemma 1] there are, for each $k \in K_2$, a set $I_0^k \subset I_0$ and vectors $(\hat{\lambda}^g_{I_0^k})^k$, $(\hat{\lambda}^h_{I_0^k})^k \geq 0$ such that

$$\lim_{k\in K_2}\Bigg[\nabla f(x^k) - \sum_{i\in\mathcal{I}_g}(\hat{\lambda}^g_i)^k\nabla g_i(x^k) - \sum_{i\in\mathcal{I}_h}(\hat{\lambda}^h_i)^k\nabla h_i(x^k) - \sum_{i\in I_0^k}(\hat{\lambda}^g_i)^k\nabla g_i(x^k) - \sum_{i\in I_0^k}(\hat{\lambda}^h_i)^k\nabla h_i(x^k)\Bigg] = 0,$$

where all gradients of constraints are linearly independent. As there are only a finite number of sets I_0^k, there is an infinite index set $K_3 \subset K_2$ and $\hat{I}_0 \subset I_0$ such that

$$\lim_{k\in K_3}\Bigg[\nabla f(x^k) - \sum_{i\in\mathcal{I}_g\cup\hat{I}_0}(\hat{\lambda}^g_i)^k\nabla g_i(x^k) - \sum_{i\in\mathcal{I}_h\cup\hat{I}_0}(\hat{\lambda}^h_i)^k\nabla h_i(x^k)\Bigg] = 0, \tag{10}$$

where all gradients of constraints remain linearly independent.

Let $S_k = \max\{\|(\hat{\lambda}^g_{\mathcal{I}_g\cup\hat{I}_0})^k\|_\infty, \|(\hat{\lambda}^h_{\mathcal{I}_h\cup\hat{I}_0})^k\|_\infty\}$. If $\liminf_{k\in K_3} S_k < \infty$ then there is an infinite set $K_4 \subset K_3$ such that $\lim_{k\in K_4}(\hat{\lambda}^g_{\mathcal{I}_g\cup\hat{I}_0})^k = \hat{\lambda}^g_{\mathcal{I}_g\cup\hat{I}_0}$ and $\lim_{k\in K_4}(\hat{\lambda}^h_{\mathcal{I}_h\cup\hat{I}_0})^k = \hat{\lambda}^h_{\mathcal{I}_h\cup\hat{I}_0}$. As $(\hat{\lambda}^g_{\hat{I}_0})^k, (\hat{\lambda}^h_{\hat{I}_0})^k \geq 0$ for all $k \in K_4$, taking $(\hat{\lambda}^g_{I_0\setminus\hat{I}_0})^k = (\hat{\lambda}^h_{I_0\setminus\hat{I}_0})^k = 0$ we conclude that x^* is S-stationary.

Now we suppose that $\lim_{k\in K_3} S_k = \infty$. Dividing (10) by S_k we obtain

$$\lim_{k\in K_3}\Bigg[\sum_{i\in\mathcal{I}_g\cup\hat{I}_0}\frac{(\hat{\lambda}^g_i)^k}{S_k}\nabla g_i(x^k) + \sum_{i\in\mathcal{I}_h\cup\hat{I}_0}\frac{(\hat{\lambda}^h_i)^k}{S_k}\nabla h_i(x^k)\Bigg] = 0.$$

All the sequences $(\hat{\lambda}^g_i)^k/S_k$ and $(\hat{\lambda}^h_i)^k/S_k$ are bounded, and hence there are convergent subsequences with indices in a infinite set $K_5 \subset K_3$. By the definition of S_k, one of these limits is equal to 1, and thus the gradients of the constraints with indexes in $\mathcal{I}_g \cup \hat{I}_0$, $\mathcal{I}_h \cup \hat{I}_0$ are linearly dependent at x^*. But this contradicts the second item of the MPCC-RCPLD definition, completing the proof. □

Remark 1: *Box and non-complementarity constraints do not offer major difficulties. For these constraints, the MPCC-RCPLD definition is stated as in [24]. The adaptation of Section 3.1 also follows the strategy adopted in [6].*

We now proceed to study the application of Algencan-second to MPCC. In the method, we consider the subproblem

$$\min_x f(x) + \frac{\rho}{2}\left\{\left\|\left(\frac{\mu^g}{\rho} - g(x)\right)_+\right\|^2 + \left\|\left(\frac{\mu^h}{\rho} - h(x)\right)_+\right\|^2 + \left[\left(\frac{\mu^0}{\rho} + g(x)^t h(x)\right)_+\right]^2\right\}.$$

Algencan-second generates sequences $\{x^k\} \subset \mathbb{R}^n$, $\{((\mu^g)^k, (\mu^h)^k, \mu^0_k)\} \subset \mathbb{R}^m_+ \times \mathbb{R}^m_+ \times \mathbb{R}_+$, $\{\rho_k\} \subset \mathbb{R}_+$ and $\{\varepsilon_k = \varepsilon_k^{\text{fun}}\} \subset \mathbb{R}_+$ satisfying the following first and second order conditions (for simplicity, we will omit "(x^k)"):

First order condition:

$$\nabla L_{\rho_k} = \nabla f - \sum_{i=1}^m\left((\mu^g_i)^k - \rho_k g_i\right)_+\nabla g_i - \sum_{i=1}^m\left((\mu^h_i)^k - \rho_k h_i\right)_+\nabla h_i + (\mu^0_k + \rho_k g^t h)_+\sum_{i=1}^m(g_i\nabla h_i + h_i\nabla g_i) \to 0. \tag{11}$$

Second order condition: The smallest eigenvalue of the matrix sequence $\{\mathcal{H}^k = \mathcal{H}_{[\mathcal{F}(x^k),\varepsilon_k,\rho_k]}\}$ tends to a non-negative value. The entries of $\mathcal{H}^k$ related to free variables are the entries of

$$\begin{aligned}\nabla^2_{\varepsilon_k} L_{\rho_k} = \nabla^2 f &- \sum_{i=1}^{m} \left[((\mu_i^g)^k - \rho_k g_i)_+ - (\mu_k^0 + \rho_k g^t h)_+ h_i\right] \nabla^2 g_i \\ &- \sum_{i=1}^{m} \left[((\mu_i^h)^k - \rho_k h_i)_+ - (\mu_k^0 + \rho_k g^t h)_+ g_i\right] \nabla^2 h_i \\ &+ \rho_k \sum_{i\in I^g_{\varepsilon_k}(x^k)} \nabla g_i \nabla g_i^t + \rho_k \sum_{i\in I^h_{\varepsilon_k}(x^k)} \nabla h_i \nabla h_i^t \\ &+ (\mu_k^0 + \rho_k g^t h)_+ \left[\sum_{i=1}^{m} (\nabla h_i \nabla g_i^t + \nabla g_i \nabla h_i^t)\right] \\ &+ z_k \rho_k \left[\sum_{i=1}^{m} (g_i \nabla h_i + h_i \nabla g_i)\right] \left[\sum_{i=1}^{m} (g_i \nabla h_i + h_i \nabla g_i)\right]^t. \end{aligned} \tag{12}$$

where

$$I^r_{\varepsilon_k}(x^k) = \left\{ i \;\middle|\; \frac{1}{\sqrt{\rho_k}} \left((\mu_i^r)^k - \rho_k r_i(x^k)\right) \geq -\varepsilon_k \right\},$$

for $r = g, h$ and

$$z_k = \begin{cases} 1 & \text{if } (1/\sqrt{\rho_k})(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) \geq -\varepsilon_k \\ 0 & \text{if } (1/\sqrt{\rho_k})(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) < -\varepsilon_k \end{cases}.$$

Observe that as we are supposing that there is no box-constraints, $\mathcal{H}^k = \nabla^2_{\varepsilon_k} L_{\rho_k}$.

Theorem 3.2. *Let $\{x^k\}$ be a sequence generated by* Algencan-second *(Algorithm 2) for solving (MPCC) and let x^* be a feasible limit point with associated infinite index set K, i.e. $\lim_{k\in K} x^k = x^*$. If the sequence of complementarity multipliers is such that $\liminf_{k\in K}(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) < \infty$ and MPCC-RCPLD holds at x^*, then x^* is S-stationary. Otherwise, if MPCC-LICQ holds at x^* and $\varepsilon_k = O(1/\sqrt{\rho_k})$, then x^* is at least M-stationary.*

Proof. As we stated in the beginning of this section we will focus on the case where there are only complementarity constraints. The general case, allowing for non-complementarity constraints will be treated in the next section.

The first statement is a direct consequence of Section 3.1. We proceed to prove the second one. That is, we are interested in the case where MPCC-LICQ holds at x^* and $\lim_{k\in K}(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) = \infty$. In particular, as the multiplier estimates computed by Algencan-second are bounded, $\rho_k \to \infty$.

The iterates satisfy

$$\lim_{k\in K} \left[\nabla f - \sum_{i=1}^{m} (\lambda_i^g)^k \nabla g_i - \sum_{i=1}^{m} (\lambda_i^h)^k \nabla h_i\right] = 0 \tag{13}$$

where $(\lambda_i^g)^k = ((\mu_i^g)^k - \rho_k g_i)_+ - \lambda_k^0 h_i$, $(\lambda_i^h)^k = ((\mu_i^h)^k - \rho_k h_i)_+ - \lambda_k^0 g_i$ and $\lambda_k^0 = (\mu_k^0 + \rho_k g^t h)_+$. Following the proof of [31, Theorem 3.2] and passing to a subsequence if necessary, we can assume that $\{(\lambda^g)^k\}_{k\in K}$ and $\{(\lambda^h)^k\}_{k\in K}$ converge to a $\overline{\lambda}^g$ and $\overline{\lambda}^h$ with $\overline{\lambda}^g_{I_h\setminus I_g} = 0$, $\overline{\lambda}^h_{I_g\setminus I_h} = 0$ and $\overline{\lambda}^g_i \overline{\lambda}^h_i \geq 0$ for all $i \in I_0$. That is, x^* is C-stationary with

$$\nabla f(x^*) - \sum_{i\in I_g} \overline{\lambda}^g_i \nabla g_i(x^*) - \sum_{i\in I_h} \overline{\lambda}^h_i \nabla h_i(x^*) = 0. \tag{14}$$

We note that MPCC-LICQ at x^* guarantees the uniqueness of $\overline{\lambda}^g_{I_g}$ and $\overline{\lambda}^h_{I_h}$ in the expression above. This result depends only on the first order information. Our objective is to use the second order properties of Algencan-second to prove that x^* is actually M-stationary.

Suppose, by contradiction, that x^* is not a M-stationary point. Thus, there is an index $j \in I_0$ such that $\overline{\lambda}_j^g \overline{\lambda}_j^h \neq 0$ and ($\overline{\lambda}_j^g < 0$ or $\overline{\lambda}_j^h < 0$). As $\overline{\lambda}_j^g \overline{\lambda}_j^h \geq 0$, we have $\overline{\lambda}_j^g < 0$ and $\overline{\lambda}_j^h < 0$. Hence, for all $k \in K$ sufficiently large we have

$$(\lambda_j^g)^k = ((\mu_j^g)^k - \rho_k g_j)_+ - (\mu_k^0 + \rho_k g^t h)_+ h_j \leq -\delta \tag{15}$$

and

$$(\lambda_j^h)^k = ((\mu_j^h)^k - \rho_k h_j)_+ - (\mu_k^0 + \rho_k g^t h)_+ g_j \leq -\delta \tag{16}$$

for a certain fixed $\delta > 0$. As $h_j \to 0$, (15) implies

$$\mu_k^0 + \rho_k g^t h \to \infty. \tag{17}$$

We affirm that $(\mu_j^g)^k - \rho_k g_j \to -\infty$ and $(\mu_j^h)^k - \rho_k h_j \to -\infty$. In fact, if $(\mu_j^g)^k - \rho_k g_j \to \infty$ then $g_j < 0$ and (15) implies $(\mu_k^0 + \rho_k g^t h)_+ h_j \to \infty$. By (16) we have $(\mu_k^0 + \rho_k g^t h)_+ \to -\infty$, which is impossible. Moreover, if $(\mu_j^g)^k - \rho_k g_j \to a \in \mathbb{R}$ then $(\mu_k^0 + \rho_k g^t h)_+ h_j \to b = (a)_+ - \overline{\lambda}_j^g > 0$. As $\mu_k^0 h_j \to 0$ we obtain $(g^t h)\rho_k h_j \to b > 0$, which implies $|\rho_k h_j| \to \infty$ because $g^t h \to 0$. Now, if $\rho_k h_j \to -\infty$ then $g^t h < 0$ which implies the boundedness of $\{(\mu_k^0 + \rho_k g^t h)_+\}$, contradicting (17). If $\rho_k h_j \to \infty$ then $g^t h > 0$ and $((\mu_j^h)^k - \rho_k h_j)_+ \to 0$. By (16) $(\mu_k^0 + \rho_k g^t h) g_j = (\mu_k^0 + \rho_k g^t h)_+ g_j \to -\overline{\lambda}_j^h > 0$, which implies $\rho_k g_j \to \infty$. Then $(\mu_j^g)^k - \rho_k g_j \to -\infty$, which contradicts the convergence to a. We conclude that $(\mu_j^g)^k - \rho_k g_j \to -\infty$, as we wanted. This analysis is valid for all $k \in K$ sufficiently large, taking successive subsequences if necessary. The proof of $(\mu_j^h)^k - \rho_k h_j \to -\infty$ is analogous.

We write

$$v^k = \sum_{i=1}^{m} \left(g_i(x^k) \nabla h_i(x^k) + h_i(x^k) \nabla g_i(x^k) \right)$$

and for each $k \in K$ we take a unitary vector

$$d^k \in \text{span}\{v^k, \nabla g_{I_g \backslash \{j\}}(x^k), \nabla h_{I_h \backslash \{j\}}(x^k)\}^{\perp}. \tag{18}$$

Such d^k exists since MPCC-LICQ holds at x^* and thus there is no more than $n-1$ vectors in this spanned subspace. Multiplying (11) by d^k we obtain

$$\nabla f^t d^k - \sum_{i=1}^{m} \left((\mu_i^g)^k - \rho_k g_i \right)_+ \nabla g_i^t d^k - \sum_{i=1}^{m} \left((\mu_i^h)^k - \rho_k h_i \right)_+ \nabla h_i^t d^k \to 0. \tag{19}$$

For all $k \in K$ large enough we have

- $\nabla g_i(x^k)^t d^k = 0, \forall i \in I_g \backslash \{j\}$, by the definition of d^k;
- $\left((\mu_i^g)^k - \rho_k g_i \right)_+ = 0, \forall i \notin I_g$, by the feasibility of x^*;
- $\left((\mu_j^g)^k - \rho_k g_j \right)_+ = 0$ since $(\mu_j^g)^k - \rho_k g_j \to -\infty$.

The same holds for h and then by (19) we have $\nabla f^t d^k \to 0$. As $\{d^k\}_{k \in K}$ is a sequence of unitary vectors, there is a infinite set $K_1 \subset K$ such that $\lim_{k \in K_1} d^k = d$. By continuity,

$$d \in S = \text{span}\{\nabla f(x^*), \nabla g_{I_g \backslash \{j\}}(x^*), \nabla h_{I_h \backslash \{j\}}(x^*)\}^{\perp}.$$

Since MPCC-LICQ holds at x^* and as we can choose d^k in (18) converging to an arbitrary $d \in S$, we can suppose that $\nabla h_j(x^*)^t d \neq 0$. In fact, if $\nabla h_j(x^*)^t d = 0$ for all $d \in S$, i.e., if

$$\nabla h_j(x^*) \in S^{\perp} = \text{span}\{\nabla f(x^*), \nabla g_{I_g \backslash \{j\}}(x^*), \nabla h_{I_h \backslash \{j\}}(x^*)\},$$

then we can write

$$\nabla h_j(x^*) = \alpha \nabla f(x^*) + \sum_{i \in I_g \backslash \{j\}} \alpha_i^g \nabla g_i(x^*) + \sum_{i \in I_h \backslash \{j\}} \alpha_i^h \nabla h_i(x^*) + 0 \cdot \nabla g_j(x^*).$$

We have $\alpha \neq 0$, since the gradients of these (active) constraints are linearly independent by MPCC-LICQ. Dividing this expression by α we obtain $\nabla f(x^*)$ as a combination of gradients of active constraints where the factor of $\nabla g_j(x^*)$ is zero. But this contradicts the uniqueness of the multipliers that appear in (14), since $\overline{\lambda}^g_j \leq -\delta$.

Multiplying (14) by d we obtain

$$\nabla g_j(x^*)^t d = -\frac{\overline{\lambda}^h_j}{\overline{\lambda}^g_j} \nabla h_j(x^*)^t d \neq 0.$$

As $\overline{\lambda}^g_j \leq -\delta$ and $\overline{\lambda}^h_j \leq -\delta$ we have $\overline{\lambda}^h_j / \overline{\lambda}^g_j > 0$, and thus

$$d^t \left[\nabla h_j(x^*) \nabla g_j(x^*)^t + \nabla g_j(x^*) \nabla h_j(x^*)^t \right] d = 2(\nabla g_j(x^*)^t d)(\nabla h_j(x^*)^t d) < 0. \tag{20}$$

Multiplying (12) by d_k, for $k \in K_1$, and remembering that $I_g \cup I_h = \{1, \ldots, m\}$ we obtain

$$\begin{aligned} d_k^t \nabla^2_{\varepsilon_k} L_{\rho_k} d_k = & \underbrace{d_k^t \left[\nabla^2 f(x^k) - \sum_{i=1}^m (\lambda^g_i)^k \nabla^2 g_i(x^k) - \sum_{i=1}^m (\lambda^h_i)^k \nabla^2 h_i(x^k) \right] d_k}_{A} \\ & \underbrace{+ \rho_k \sum_{i \in I^g_{\varepsilon_k}(x^k)} \left(d_k^t \nabla g_i(x^k) \right)^2 + \rho_k \sum_{i \in I^h_{\varepsilon_k}(x^k)} \left(d_k^t \nabla h_i(x^k) \right)^2}_{B} \\ & + \underbrace{\left(\mu^0_k + \rho_k g(x^k)^t h(x^k) \right)_+ \left[d_k^t \left(\nabla h_j(x^k) \nabla g_j(x^k)^t + \nabla g_j(x^k) \nabla h_j(x^k)^t \right) d_k \right]}_{C}. \end{aligned} \tag{21}$$

The term A is bounded since $\lim_{k \in K_1} \lambda^g = \overline{\lambda}^g$ and $\lim_{k \in K_1} \lambda^h = \overline{\lambda}^h$. As $\lim_{k \in K_1} ((\mu^g_j)^k - \rho_k g_j) = \lim_{k \in K_1} ((\mu^h_j)^k - \rho_k h_j) = -\infty$, we conclude that $j \notin I^g_{\varepsilon_k}(x^k) \cup I^h_{\varepsilon_k}(x^k)$ for all $k \in K_1$ if ϵ_k is small enough to ensure that

$$0 < \varepsilon_k < -\max \left\{ \frac{1}{\sqrt{\rho_k}} ((\mu^g_j)^k - \rho_k g_j), \frac{1}{\sqrt{\rho_k}} ((\mu^h_j)^k - \rho_k h_j) \right\}. \tag{22}$$

Furthermore, we have, for all $k \in K_1$ large enough, $(\mu^g_j)^k - \rho_k g_j \leq -\delta$ and $(\mu^h_j)^k - \rho_k h_j \leq -\delta$ for a fixed $\delta > 0$. Thus, the algorithmic condition $\varepsilon_k = O(1/\sqrt{\rho_k})$ is sufficient to guarantees the validity of (22). Recall that for all $i \notin I_g$, if $k \in K_1$ is large enough, g_j will be positive and bounded away from zero. Hence, as $\rho_k \to \infty$, it follows that

$$\frac{1}{\sqrt{\rho_k}} ((\mu^g_j)^k - \rho_k g_j) = \frac{(\mu^g_j)^k}{\sqrt{\rho_k}} - \sqrt{\rho_k} g_j \underset{K_1}{\to} -\infty.$$

Therefore $(I_h \backslash I_g) \cap I^g_{\varepsilon_k}(x^k) = \emptyset$ for these indices k. Analogously, $(I_g \backslash I_h) \cap I^h_{\varepsilon_k}(x^k) = \emptyset$ for all $k \in K_1$ large enough. We conclude that the term B in (21) does not appear. Finally, by (17) and (20), the term C in (21) tends to $-\infty$. This shows that $d_k^t \nabla^2_{\varepsilon_k} L_{\rho_k} d_k \to -\infty$, contradicting the second order test (12). We conclude that x^* is a M-stationary point, completing the proof. □

Figure 4 gives a geometric interpretation for the ALGENCAN-SECOND convergence.

Remark 2 (box constraints): *We can adapt the proof of Section 3.2 to cover box-constraints considering*

$$d^k \in span\{v^k, \nabla g_{I_g \backslash \{j\}}(x^k), \nabla h_{I_h \backslash \{j\}}(x^k), e_{\mathcal{F}}\}^\perp$$

instead of (18), where $e_{\mathcal{F}}$ is the matrix formed by canonical vectors e_i's of $\mathbb{R}^n$ that indices are the ones of fixed variables in the open face $\mathcal{F} = \mathcal{F}(x^*)$ (this is possible by MPCC-LICQ condition). Thus, d^k will also be orthogonal to $e_{\mathcal{F}}$ and by the first order expression we will obtain $\nabla f(x^k)^t d^k \to$

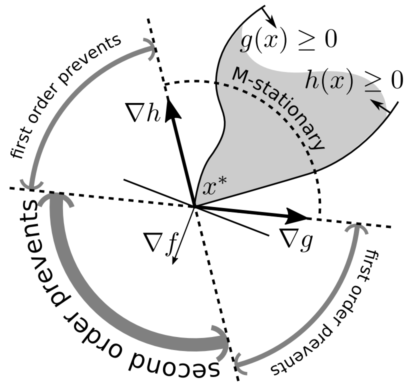

Figure 4: ALGENCAN-SECOND convergence. The two constraints $g(x) \geq 0$ and $h(x) \geq 0$ are active at x^*. When $\nabla f(x^*)$ is in the regions delimited by thin solid arcs, x^* is not C-stationary, while the thick solid arc indicate the region for which x^* is C-stationary, but not M-stationary. When $\nabla f(x^*)$ is in the regions indicated by dotted lines, x^* is M-stationary. ALGENCAN does not converge to x^* when $\nabla f(x^*)$ is on thin arcs, while ALGENCAN-SECOND also avoids this undesirable situation when $\nabla f(x^*)$ is on the thick arc. In the last case, a orthogonal vector d to $\nabla f(x^*)$ fulfills $d^t(\nabla h \nabla g^t + \nabla g \nabla h^t)d = 2(\nabla h^t d)(\nabla g^t d) < 0$.

0. *Analogously we conclude that* $(d^k)^t \nabla^2_{\varepsilon_k} L_{\rho_k} d^k \to -\infty$. *Now, as* d^k *is orthogonal to* $e_{\mathcal{F}}$ *then* $d_i = 0$ *for all* i *such that* x_i *is fixed in* $\mathcal{F}$. *Therefore, we can write*

$$d^k = \sum_{i;\, x_i\, free} d_i^k e_i$$

and hence

$$(d^k)^t \mathcal{H}^k d^k = \sum_{\substack{(i,j);\\ x_i, x_j\ free}} d_i^k d_j^k \mathcal{H}_{ij}^k = \sum_{\substack{(i,j);\\ x_i, x_j\ free}} d_i^k d_j^k \left(\nabla^2_{\varepsilon_k} L_{\rho_k}\right)_{ij} = (d^k)^t \nabla^2_{\varepsilon_k} L_{\rho_k} d^k \to -\infty,$$

completing the argument.

Remark 3: *In the example of introduction,* ALGENCAN *may converges to* $(0,0)$ *with the sequences* $(x_k, y_k) = (1/k, 1/k)$ *and* $\rho_k = k^3$, *but not* ALGENCAN-SECOND *since the origin is not M-stationary. In the example of Section 2.4,* ALGENCAN-SECOND *(and consequently* ALGENCAN*) may converge to the origin with the sequences* $(x_k, y_k) = (1/k^2, 1/k)$ *and* $\rho_k = k^4$, *in which case* $\mu_k^0 + \rho_k x^k y^k \to \infty$. *This point is M-stationary but not S-stationary. Therefore we can not expect more than convergence to M-stationary points.*

3.3 General constraints

Let us show briefly that Section 3.2 also holds in the presence of the non-complementarity constraints $G(x) \leq 0$ and $H(x) = 0$ (box constraints are neglected here by simplicity; remark 2 is still applicable). In order to prove the Theorem in this case let us recall that ALGECAN-SECOND subproblem is

$$\min f(x) + \frac{\rho}{2}\left\{ \left\| \left(\frac{\mu^G}{\rho} + G(x)\right)_+ \right\|^2 + \left\| \frac{\mu^H}{\rho} + H(x) \right\|^2 \right.$$
$$\left. + \left\| \left(\frac{\mu^g}{\rho} - g(x)\right)_+ \right\|^2 + \left\| \left(\frac{\mu^h}{\rho} - h(x)\right)_+ \right\|^2 + \left[\left(\frac{\mu^0}{\rho} + g(x)^t h(x)\right)_+ \right]^2 \right\}.$$

The algorithm then generates sequences satisfying the following first and second order conditions:

First order condition:

$$\begin{aligned}\nabla L_{\rho_k} =& \nabla f + \sum_{i=1}^{s}((\mu_i^G)^k + \rho_k G_i)_+ \nabla G_i + \sum_{i=1}^{q}((\mu_i^H)^k + \rho_k H_i)\nabla H_i \\ &- \sum_{i=1}^{m}((\mu_i^g)^k - \rho_k g_i)_+ \nabla g_i - \sum_{i=1}^{m}((\mu_i^h)^k - \rho_k h_i)_+ \nabla h_i \\ &+ (\mu_k^0 + \rho_k g^t h)_+ \sum_{i=1}^{m}(\nabla h_i g_i + \nabla g_i h_i) \to 0.\end{aligned} \tag{23}$$

Second order condition: The smallest eigenvalue of the matrix below converges to a nonnegative value.

$$\begin{aligned}\nabla^2_{\varepsilon_k} L_{\rho_k} = \nabla^2 f \;&+ \sum_{i=1}^{s}((\mu_i^G)^k + \rho_k G_i)_+ \nabla^2 G_i + \sum_{i=1}^{q}((\mu_i^H)^k + \rho_k H_i)\nabla^2 H_i \\ &+ \rho_k \sum_{i \in I^G_{\varepsilon_k}(x^k)} \nabla G_i \nabla G_i^t + \rho_k \sum_{i=1}^{q} \nabla H_i \nabla H_i^t \\ &- \sum_{i=1}^{m}[((\mu_i^g)^k - \rho_k g_i)_+ - (\mu_k^0 + \rho_k g^t h)_+ h_i]\nabla^2 g_i \\ &- \sum_{i=1}^{m}[((\mu_i^h)^k - \rho_k h_i)_+ - (\mu_k^0 + \rho_k g^t h)_+ g_i]\nabla^2 h_i \\ &+ \rho_k \sum_{i \in I^g_{\varepsilon_k}(x^k)} \nabla g_i \nabla g_i^t + \rho_k \sum_{i \in I^h_{\varepsilon_k}(x^k)} \nabla h_i \nabla h_i^t \\ &+ (\mu_k^0 + \rho_k g^t h)_+ \left[\sum_{i=1}^{m}(\nabla h_i \nabla g_i^t + \nabla g_i \nabla h_i^t)\right] \\ &+ z_k \rho_k \left[\sum_{i=1}^{m}(\nabla h_i g_i + \nabla g_i h_i)\right]\left[\sum_{i=1}^{m}(\nabla h_i g_i + \nabla g_i h_i)\right]^t.\end{aligned}$$

where the new set

$$I^G_{\varepsilon_k}(x^k) = \left\{ i \subset \{1, \ldots, p\} \;\middle|\; \frac{1}{\sqrt{\rho_k}}\left((\mu_i^G)^k + \rho_k G_i(x^k)\right) \geq -\varepsilon_k \right\}.$$

As previously, we consider the additional set of indices $I_G(x) = \{i \mid G_i(x) = 0\}$ and denote $I_G = I_G(x^*)$, $I_g = I_g(x^*)$ and $I_h = I_h(x^*)$.

Again, if $\liminf(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) < \infty$, we can use only the first order properties to ensure convergence to S-stationary points under MPCC-RCPLD. We now focus on the case $\liminf(\mu_k^0 + \rho_k g(x^k)^t h(x^k)) = \infty$.

Once more, considering only the first order properties we can argue that, Algencan-second converges to C-stationary points, i.e.,

$$\nabla f(x^*) + \sum_{i \in I_G} \overline{\lambda}_i^G \nabla G_i(x^*) + \sum_{i=1}^{q} \overline{\lambda}_i^H \nabla H_i(x^*) - \sum_{i \in I_g} \overline{\lambda}_i^g \nabla g_i(x^*) - \sum_{i \in I_h} \overline{\lambda}_i^h \nabla h_i(x^*) = 0. \tag{24}$$

We proceed supposing that x^* is not M-stationary, and then an index $j \in I_0$ exists. If MPCC-LICQ holds on x^* then the gradients

$$\nabla H(x^*), \quad \nabla G_{I_G}(x^*), \quad \nabla g_{I_g}(x^*) \quad \text{and} \quad \nabla h_{I_h}(x^*)$$

are linearly independent. Therefore, we can take

$$d_k \in \operatorname{span}\left\{v^k, \nabla h_{I_h \setminus \{j\}}(x^k), \nabla g_{I_g \setminus \{j\}}(x^k), \nabla G_{I_G}(x^k), \nabla H(x^k)\right\}^\perp.$$

Multiplying the first order expression (23) by d_k we obtain

$$\nabla f^t d_k + \sum_{i=1}^{s}((\mu_i^G)^k + \rho_k G_i)_+ \nabla G_i^t d_k + \sum_{i=1}^{q}((\mu_i^H)^k + \rho_k H_i)\nabla H_i^t d_k \\ - \sum_{i=1}^{m}((\mu_i^g)^k - \rho_k g_i)_+ \nabla g_i^t d_k - \sum_{i=1}^{m}((\mu_i^h)^k - \rho_k h_i)_+ \nabla h_i^t d_k \to 0,$$

and analogously $\nabla f^t d_k \to 0$. The limit $d = \lim_{k\in K_1} d_k$ satisfies

$$d \in \operatorname{span}\{\nabla f(x^*), \nabla h_{I_h \setminus \{j\}}(x^*), \nabla g_{I_g \setminus \{j\}}(x^*), \nabla G_{I_G}(x^*), \nabla H(x^*)\}^{\perp}.$$

With similar arguments on (24) we obtain $\nabla g_j(x^*)^t d = -(\overline{\lambda}_j^h / \overline{\lambda}_j^g)\nabla h_j(x^*)^t d \neq 0$, and (20) holds. Multiplying the second order expression by d_k, $k \in K_1$, we obtain

$$\begin{aligned} d_k^t \nabla^2_{\varepsilon_k} L_{\rho_k} d_k = d_k^t \Bigg[&\nabla^2 f(x^k) + \sum_{i=1}^{s} \left((\mu_i^G)^k + \rho_k G_i(x^k)\right)_+ \nabla^2 G_i(x^k) \\ &+ \sum_{i=1}^{q} \left((\mu_i^H)^k + \rho_k H_i(x^k)\right) \nabla^2 H_i(x^k) - \sum_{i=1}^{m} (\lambda_i^g)^k \nabla^2 g_i(x^k) - \sum_{i=1}^{m} (\lambda_i^h)^k \nabla^2 h_i(x^k) \Bigg] d_k \\ &+ \underbrace{\rho_k \sum_{i \in I^g_{\varepsilon_k}(x^k)} \left(d_k^t \nabla g_i(x^k)\right)^2 + \rho_k \sum_{i \in I^h_{\varepsilon_k}(x^k)} \left(d_k^t \nabla h_i(x^k)\right)^2 + \rho_k \sum_{i \in I^G_{\varepsilon_k}(x^k)} \left(d_k^t \nabla G_i(x^k)\right)^2}_{B} \\ &+ \underbrace{\left(\mu_k^0 + \rho_k g(x^k)^t h(x^k)\right)_+ \left[d_k^t \left(\nabla h_j(x^k) \nabla g_j(x^k)^t + \nabla g_j(x^k) \nabla h_j(x^k)^t\right) d_k\right]}_{C}. \end{aligned}$$

The term between brackets is bounded and C tends to $-\infty$. In B, the two first sums do not appear when ε_k is sufficiently small. Now, if $i \notin I_G(x^*)$ then, for all k sufficiently large, $G_i(x^k) \leq -\gamma < 0$ and

$$\frac{1}{\sqrt{\rho_k}} \left((\mu_i^G)^k + \rho_k G_i(x^k)\right) \leq \frac{(\mu_i^G)^k}{\sqrt{\rho_k}} - \sqrt{\rho_k}\gamma \to -\infty$$

which implies $i \notin I^G_{\varepsilon_k}(x^k)$. Thus the third sum in B will not appear either. This completes the proof of Section 3.2 considering non-complementarity constraints.

4 Numerical experiments

In this section we report and discuss the behavior of our implementation of Algencan-second. The implementation was based on the Algencan 3.0.0 beta package provided by the TANGO project (http://www.ime.usp.br/~egbirgin/tango). Analogously, our implementation of Gencan-second was based on the Gencan code supplied with the Algencan package. To compute eigenvalues, we used the *Jacobian Preconditioned Conjugate Gradients* method (JPCG) [44] implemented in the HSL EA19 package (1.4.0 version) [29]. In order to force Algencan to behave like a pure augmented Lagrangian method we turned off an acceleration option that consists in trying to apply a Newton-type method to the non-scaled KKT system obtained by fixing active constraints as equalities [12]. This acceleration routine was not considered in our analysis.

Regarding to our Algencan-second implementation we highlight:

- Gencan employs a quadratic interpolation backtracking. We keep this backtracking strategy if a first order direction is chosen in our implementation of Gencan-second. When a second order direction is chosen, we perform a simple backtracking $t \leftarrow 0.9 \times t$. Observe that the sufficient decrease criteria for the second order in Algencan-second is not of Armijo-type [1];

- Gencan employs extrapolation steps [13] that improve the global performance of Algencan. We maintain this strategy for Gencan-second;
- As we stated in Section 3.2, the tolerance ε^{fun} for the approximate ε-Hessian must be of order $O(1/\sqrt{\rho})$. In our implementation, we take the nonincreasing sequence

$$\varepsilon_k^{\text{fun}} = \max\left\{10^{-14}, \min\{\varepsilon_{k-1}^{\text{fun}}, 1/\sqrt{\rho_k}\}\right\};$$

- At each iteration of Gencan, the decrease of the objective function, the decrease of the projected gradient norm, and the step size are monitored. Gencan stops by lack of progress if for `maxinnitnp` consecutive iterations at least `itnplevel` of these values do not improve. In Gencan-second we additionally monitor the smallest eigenvalue of the reduced ε-Hessian. More specifically, we stop Gencan-second if during `maxinnitnp` consecutive iterations at least `itnplevel` original measures do not improve and, additionally, if the smallest eigenvalue remains negative. We use the default Gencan values for `maxinnitnp` $= 3$ and `itnplevel` $=$ 2;
- Gencan-second starts searching for directions of negative curvature only near first order stationary points. More specifically, eigenvalues are computed once both conditions below hold:
  - the projected gradient norm is less or equal than $\varepsilon^{\text{grad}}$ (step 1 of Gencan-second);
  - the gradient norm is less or equal than a threshold $\varepsilon^0 > 0$. This criterion is related to the choice of the second order directions in the inner iterations of step 2 of Gencan-second (see [1, Algorithm 2.3]).

  Furthermore, eigenvalues may be computed to verify the lack of progress condition described in the last item;
- Augmented Lagrangian methods usually start from arbitrary multiplier values, typically the zero vector. Only after a few outer iterations, the multiplier estimates start to carry real information about the problem. Hence, we avoid to use the second information from the approximate ε-Hessian $\nabla_\varepsilon^2 L_\rho$ prematurely:
  - A Gencan-second iteration is never executed in the first outer iteration of Algencan-second;
  - Gencan-second is turned on after the multiplier estimates λ^k settled. That is, if $\|\lambda^k - \lambda^{k-1}\|_\infty \leq \varepsilon^\lambda$ by N_λ consecutive iterations;
  - Gencan-second is also turned on if the first order convergence criteria of Algencan-second are satisfied with its tolerance multiplied by 10. That is, if the complementarity, the feasibility and the projected gradient norm are at most 10 times the desired first order tolerance;
  - After the first execution of Gencan-second, all future iterations of Algencan-second will use it. This ensures that we realize an "infinite" number of second order iterations, in which case Section 3.2 is valid;
  - We do not declare convergence in Algencan-second before performing the first second order iteration. Thus, we avoid cases where the convergence to stationary points occurs rapidly only with first order directions, before at least one second order test takes place. Hence, we guarantee that we have a legitimate second order method.
- On the other hand, we start to use Gencan-second whenever we can detect that Algencan-second may fail:
  - Algencan-second may fail declaring that it found a stationary point of the feasibility measure. Therefore, we turn on Gencan-second if the gradient of the feasibility measure at the current point is less than two times the tolerance that Algencan-second uses to declare failure;

– The method may also fail if the penalty parameter ρ_k increases too much. Therefore, we turn on GENCAN-SECOND if ρ_k is larger than fixed large $\rho_{2\text{nd}}$. Hopefully the use of second order information may help to preclude the failure;
– GENCAN-SECOND is also started after 2/3 of the maximum number of iterations of ALGENCAN-SECOND is reached. This is an undesirable situation, but aims to execute second order iterations before the first order method halts.

The computational tests was performed using the MacMPEC collection which is available in the AMPL modelling language at http://www.mcs.anl.gov/~leyffer/MacMPEC. From the 193 test problems, 19 were either multi-objective (`ralph1`), infeasible (`pack-rig2-16/32`, `pack--rig2c-16/32`), had integer variables (`ex9.1.2`) or had mixed complementarity constraints (`bilevel1m`, `bilevel2m`, `gnash10m–19m`, `taxmcp`), and were hence ignored. Another 15 tests problems were not considered because they cause unknown fatal errors in the AMPL interface of ALGENCAN (`gnash14`, `incid-set-1-8/16/32`, `incid-set-1c-8/16/32`, `incid-set-2-8/16/32`, `incidset-2c-8/16/32`, `water-FL`, `water-net`). Problem `pack-rig3c-32` causes memory overflow in the second-order method, while `bem-milanc30-s`, `tollmpec`, and `tollmpec1` exceed five hours on the second order computations. These problems were then also ignored. We note that the original instances use the directive `complements` to represent complementarity constraints. This directive is not understood by ALGENCAN's AMPL interface. Therefore we rewrote these constraints using explicit, functional, complementarities. The AMPL option `presolve` was turned off in order avoid changes in the structure of the problems. A computer with an Intel i7 2.6 Ghz processor, 8 Gb RAM and GNU/Linux 64 bits system (Ubuntu 16.04) was used. For ALGENCAN and ALGENCAN-SECOND we used $\varepsilon^{\text{opt}} = \varepsilon^{\text{feas}} = 10^{-6}$ and the maximum number of outer iterations equal to 100. Regarding the parameters that are exclusive the second order methods, we set $\varepsilon^0 = 10^{-4}$, $\varepsilon^{\text{curv}} = 0.99 \times 10^{-6}$, $\varepsilon^{\text{hess}} = 10^{-6}$, $\kappa = 1$, $\eta = 0.1$, $\varepsilon_0^{\text{fun}} = 10^{-6}$, $\varepsilon^{\lambda} = \sqrt{\varepsilon^{\text{feas}}} = 10^{-3}$, $N_\lambda = 3$ and $\rho_{2\text{nd}} = 10^8$. We initialize the multipliers as zero. When not supplied, the initial point was taken as the origin.

The use of slacks in ALGENCAN is not recommended in general, since the method can benefit from active inequality constraints [11]. Hence the use of slacks in ALGENCAN is disabled by default. However, the numerical experience of SQP in MPCCs suggests that it is better to write complementarity in the form $g(x)^t h(x) + s = 0$, $s \geq 0$ [23]. Izmailov, Solodov and Uskov [31] adopt this form in their computational tests with ALGENCAN. In order to compare the behavior of ALGENCAN and ALGENCAN-SECOND in these two situations, we realized tests with complementarity in both forms $g(x)^t h(x) \leq 0$ and $g(x)^t h(x) + s = 0, s \geq 0$. In both cases, ALGENCAN and ALGENCAN-SECOND behave similary in most problems, without significant differences.

We consider that ALGENCAN/ALGENCAN-SECOND did converged if one the following situations occurred:

- the algorithm declares convergence to a stationary point;
- the algorithm stops with a large penalty parameter ρ (greater then 10^8, the standard value in ALGENCAN), but with feasibility and with the final functional value not worse than the best objective value known in the literature, as reported in MacMPEC collection. This is a typical situation where the limit is a C/M-stationary point only.

Otherwise, we consider that the algorithm failed.

Results for complementarity without slack

In this case, complementarity takes the form $g(x)^t h(x) \leq 0$. From the 155 considered problems, both methods fail in 15 (9,68%) of them.

Table 1 brings the problems where there was convergence for at least one method, and different limits were reached. The column "Best Obj" contains the best objective function value known in the literature, as described in MacMPEC collection. The columns "Obj" and "Infeas" contain, respectively, the final objective function value and the measure of infeasibility [11] given by

$$\max\left\{\|H(x^k)\|_\infty,\ \|G(x^k)_+\|_\infty,\ \|\left(-h(x^k)\right)_+\|_\infty,\ \|\left(-g(x^k)\right)_+\|_\infty,\ \left(g(x^k)^t h(x^k)\right)_+\right\}.$$

Problem	Best Obj	Algencan Obj	Algencan Status	Algencan Infeas	Algencan-second Obj	Algencan-second Status	Algencan-second Infeas
pack-comp1c-32	6.61e-01	6.01e-01	$\rho \gg 1$	1.14e-03	6.61e-01	conv./S	3.47e-08
scale4	1.00e+00	1.71e+00	conv./C	5.57e-07	1.00e+00	conv./S	1.58e-08
scale5	1.00e+02	2.00e+02	conv./C	5.56e-07	1.00e+02	conv./S	3.00e-09
scholtes3	5.00e-01	9.99e-01	conv./C	5.56e-07	5.00e-01	conv./S	3.00e-10

Table 1: Computational tests, complementarity as $g(x)^t h(x) \leq 0$. Differences between Algencan and Algencan-second.

In the column "Status", "$\rho \gg 1$" means that the method failed with a large penalty parameter. The entry "conv./C" means that the method declared convergence but at least one pair of multipliers associated to complementarity constraints seems to be converging to negative values. That is, the point found is approximately C-stationary only. The entry "conv./S" means convergence to a point that is approximately S-stationary, in particular all multipliers associated to complementarity constraints are approximately non-negative. We can see that in the `pack-comp1c-32` problem, Algencan-second reached feasibility, while Algencan did not achieve the required precision. This lead to failure due to large penalty parameter for the first-order method. In the others three problems, Algencan-second converges to a point with smaller functional value than the ones attained by the first order method. It is possible to verify that in each case Algencan converged to a point that is only C-stationary, while Algencan-second converged to the best known solution which is an S-stationary point, corroborating our theory. In the other 135 problems that both methods declared convergence, the same functional value was obtained by two methods. In these problems, the functional value was equal to the best value reported in MacMPEC collection with exception to the problems `ex9.2.5`, `qpecgen100-4`, `qpecgen200-3`, and `qpecgen200-4` where a S-stationary point with better objective value were found, and `bard3`, `ex9.2.3`, `pack-comp1p-8`, `qpecgen100-1`, `qpecgen100-3`, `qpecgen200-2`, and `TrafficSignalCycle-1–11/13` where the methods converged to S-stationary points with worse objective values. We note that small random pertubations of the starting points still lead to similar results, where the first-order method converges to points that are approximately C-stationary in a few problems while the second-order variation recovers approximately S-stationary points. This observation is in line with our theoretical results.

Results for complementarity with slack

When we consider complementarity as $g(x)^t h(x) + s = 0$, $s \geq 0$, the general behavior of both methods is very similar to the previous case. The number of problems considered was one less than before, 154, as detected an unknown error in AMPL interface in the `hakonsen` problem. Once again, both methods fail in only 15 (9,74%) of the tests. As before, we highlight the problems where the feasibility was reached by at least one of methods and where both of them converged to different limits in Table 2. In the previous case, Algencan and Algencan-second attained the same functional value for `pack-comp1p-8` instance which was larger than the value reported by MacMPEC. Now, both methods declare convergence, but Algencan-second reaches a better functional value, quite close to the best known value.

Note that Algencan behaves very similarly in the two ways of writing the complementarity constraints. Thus we can not assert that one form is better than the other. In [31], the authors presented a similar observation. The same comment is valid for Algencan-second.

In the computational tests performed in [31], the authors considered 161 problems from the MacMPEC collection and reported that Algencan converged in about 97% of them. In our tests, this percentage was roughly 90%. This difference is probably explained by different options and the different versions of Algencan used in both tests. In particular, we did not use the specialized solver for sparse linear systems from HSL that is suggested in Algencan's documentation for large scale problems. When such linear solver is available, Algencan employs a trust region Newton method in inner iterations of Gencan [11]. This differs from the original line search-based version of Gencan [13] and is not compatible with Gencan-second. On the other hand, the main objective of the numerical experiments described in this section is to see whether the second order

Problem	Best Obj	Algencan Obj	Algencan Status	Algencan Infeas	Algencan-second Obj	Algencan-second Status	Algencan-second Infeas
`pack-comp1c-32`	6.61e-01	8.25e-01	$\rho \gg 1$	9.32e+06	6.61e-01	conv./S	6.46e-07
`pack-comp1p-8`	6.00e-01	6.61e-01	conv./S	2.11e-07	6.01e-01	conv./S	1.50e-09
`scale4`	1.00e+00	1.71e+00	conv./C	5.57e-07	1.00e+00	conv./S	1.58e-08
`scale5`	1.00e+02	2.00e+02	conv./C	5.56e-07	1.00e+02	conv./S	3.00e-10
`scholtes3`	5.00e-01	9.99e-01	conv./C	5.56e-07	5.00e-01	conv./S	3.00e-10

Table 2: Computational tests, complementarity as $g(x)^t h(x) + s = 0$, $s \geq 0$. Differences between Algencan and Algencan-second.

version of Gencan is able to recover better MPCC-stationary points than the method that is concerned only with first order stationarity, as suggested by Section 3.2. In order to make this comparison, the original Gencan has to be used in all the tests.

In any case, everything indicates that Algencan reaches minimizers frequently. Thus, we should not consider first and second order methods competitors among themselves. Our intention is to show that there are few cases where second order information can be useful in recovering better stationary points. This is in the spirit of Algencan-second: we only resort to the second order when the first order tends to be exhausted.

5 Conclusions

In this paper, we improved the convergence results [31] for Algencan [2] when applied to MPCCs. We were able to replace the stringent MPCC-LICQ condition by the much more general MPCC-RCPLD. We also proved a new result for the second order augmented Lagrangian method Algencan-second [1]. We showed that it is able to cope with situations where unbounded multipliers appear, still asserting convergence to M-stationary points. We performed numerical tests using MacMPEC instances and showed cases for which Algencan-second converges to M-stationary points, while Algencan is only able to recover C-stationarity.

Finally, there are many other second order methods for general nonlinear optimization. We are interested in analyzing the convergence of such methods to second order stationary points for MPCC.

Acknowledgments

We would like to thank the referees for their valuable comments that greatly improved the original manuscript.

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. *Comput. Optim. Appl.*, 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 J. Optim.*, 18(4):1286–1309, 2007.

[3] R. Andreani, C. Dunder, and J. M. Martínez. Nonlinear-programming reformulation of the order-value optimization problem. *Math. Methods Oper. Res.*, 61(3):365–384, 2005.

[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 J. Numer. Anal.*, 2017.

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

[6] Roberto Andreani, Gabriel Haeser, María Laura Schuverdt, and Paulo J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Math. Program.*, 135(1):255–273, 2012.

[7] Roberto Andreani, José Mário Martínez, Alberto Ramos, and Paulo J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM J. Optim.*, 26(1):96–110, 2016.

[8] Mihai Anitescu. Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints. *SIAM J. Optim.*, 16(1):120–145, 2005.

[9] Mihai Anitescu. On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints. *SIAM J. Optim.*, 15(4):1203–1236, 2005.

[10] Mihai Anitescu, Paul Tseng, and Stephen J. Wright. Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties. *Math. Program.*, 110(2):337–371, 2007.

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

[12] E.G. Birgin and J.M. Martínez. Improving ultimate convergence of an augmented lagrangian method. *Optim. Methods Softw.*, 23(2):177–195, 2008.

[13] Ernesto G. Birgin and José Mario Martínez. Large-scale active-set box-constrained optimization method with spectral projected gradients. *Comput. Optim. Appl.*, 23(1):101–125, 2002.

[14] Ernesto G. Birgin and J. M. Martínez. Augmented lagrangian method with nonmonotone penalty parameters for constrained optimization. *Comput. Optim. Appl.*, 51(3):941–965, 2012.

[15] Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. Inexact spectral projected gradient methods on convex sets. *IMA J. Numer. Anal.*, 23(4):539, 2003.

[16] Stephan Dempe. Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. *Optimization*, 52(3):333–359, 2003.

[17] Francisco Facchinei, Houyuan Jiang, and Liqun Qi. A smoothing method for mathematical programs with equilibrium constraints. *Math. Program.*, 85(1):107–134, 1999.

[18] M. C. Ferris and J. S. Pang. Engineering and economic applications of complementarity problems. *SIAM Rev.*, 39(4):669–713, 1997.

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

[20] Michael L. Flegel and Christian Kanzow. *A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints*, pages 111–122. Springer US, Boston, MA, 2006.

[21] Michael L. Flegel, Christian Kanzow, and Jiří V. Outrata. Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints. *Set-Valued Analysis*, 15(2):139–162, 2007.

[22] R. Fletcher and S. Leyffer. Numerical experience with solving MPECs as NLPs. Technical report, University of Dundee Report NA 210, 2002.

[23] Roger Fletcher, Sven Leyffer, Danny Ralph, and Stefan Scholtes. Local convergence of sqp methods for mathematical programs with equilibrium constraints. *SIAM J. Optim.*, 17(1):259–286, 2006.

[24] Lei Guo, Gui-Hua Lin, and Jane J. Ye. Second-order optimality conditions for mathematical programs with equilibrium constraints. *J. Optim. Theory Appl.*, 158(1):33–64, 2013.

[25] Lei Guo, Gui-Hua Lin, Dali Zhang, and Daoli Zhu. An MPEC reformulation of an EPEC model for electricity markets. *Oper. Res. Lett.*, 43(3):262–267, 2015.

[26] Magnus R. Hestenes. Multiplier and gradient methods. *J. Optim. Theory Appl.*, 4(5):303–320, 1969.

[27] Tim Hoheisel, Christian Kanzow, and Alexandra Schwartz. Convergence of a local regularization approach for mathematical programmes with complementarity or vanishing constraints. *Optim. Methods Softw.*, 27(3):483–512, 2012.

[28] Tim Hoheisel, Christian Kanzow, and Alexandra Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Math. Program.*, 137(1):257–288, 2013.

[29] HSL. A collection of fortran codes for large scale scientific computation, 2016.

[30] X. M. Hu and D. Ralph. Convergence of a penalty method for math. program. with complementarity constraints. *J. Optim. Theory Appl.*, 123(2):365–390, 2004.

[31] A. F. Izmailov, M. V. Solodov, and E. I. Uskov. Global convergence of augmented lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM J. Optim.*, 22(4):1579–1606, 2012.

[32] Abdeslam Kadrani, Jean-Pierre Dussault, and Abdelhamid Benchakroun. A new regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Optim.*, 20(1):78–103, 2009.

[33] Abdeslam Kadrani, Jean Pierre Dussault, and Abdelhamid Benchakroun. A globally convergent algorithm for MPCC. *EURO J. Comput. Optim.*, 3(3):263–296, 2015.

[34] Christian Kanzow and Alexandra Schwartz. Mathematical programs with equilibrium constraints: Enhanced fritz john-conditions, new constraint qualifications, and improved exact penalty results. *SIAM J. Optim.*, 20(5):2730–2753, 2010.

[35] Christian Kanzow and Alexandra Schwartz. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM J. Optim.*, 23(2):770–798, 2013.

[36] Christian Kanzow and Alexandra Schwartz. The price of inexactness: Convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited. *Math. Oper. Res.*, 40(2):253–275, 2015.

[37] Sven Leyffer, Gabriel López-Calva, and Jorge Nocedal. Interior methods for mathematical programs with complementarity constraints. *SIAM J. Optim.*, 17(1):52–77, 2006.

[38] Gui-Hua Lin and Masao Fukushima. A modified relaxation scheme for mathematical programs with complementarity constraints. *Ann. Oper. Res.*, 133(1):63–84, 2005.

[39] Xinwei Liu and Jie Sun. Generalized stationary points and an interior-point method for mathematical programs with equilibrium constraints. *Math. Program.*, 101(1):231–261, 2004.

[40] H. Z. Luo, X. L. Sun, and Y. F. Xu. Convergence properties of modified and partially-augmented Lagrangian methods for mathematical programs with complementarity constraints. *J. Optim. Theory Appl.*, 145(3):489–506, 2010.

[41] Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[42] J. Outrata, M. Kocvara, and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, volume 28 of *Nonconvex Optimization and Its Applications*. Springer US, 1 edition, 1998.

[43] J. V. Outrata. Optimality conditions for a class of mathematical programs with equilibrium constraints: Strongly regular case. *Kybernetika*, 35(2):177–193, 1999.

[44] E. E. Ovtchinnikov. Jacobi correction equation, line search, and conjugate gradients in hermitian eigenvalue computation I: Computing an extreme eigenvalue. *SIAM J. Numer. Anal.*, 46(5):2567–2592, 2008.

[45] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, New York, 1969.

[46] A. Ramos. Mathematical programms with equilibrium constraints: A sequential optimality condition, new constraint qualifications and algorithmic consequences, abr 2016.

[47] R. Tyrrell Rockafellar. Augmented lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2):268–285, 1974.

[48] Holger Scheel and Stefan Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Math. Oper. Res.*, 25(1):1–22, 2000.

[49] Stefan Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Optim.*, 11(4):918–936, 2001.

[50] Sonja Steffensen and Michael Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM J. Optim.*, 20(5):2504–2539, 2010.

[51] Huayu Xu, Jong-Shi Pang, Fernando Ordóñez, and Maged Dessouky. Complementarity models for traffic equilibrium with ridesharing. *Transportation Research Part B: Methodological*, 81, Part 1:161–182, 2015.

[52] X. Q. Yang and X. X. Huang. Lower-order penalty methods for mathematical programs with complementarity constraints. *Optim. Methods Softw.*, 19(6):693–720, 2004.

[53] X.Q. Yang and X.X. Huang. Convergence analysis of an augmented lagrangian method for mathematical programs with complementarity constraints. *Nonlinear Anal.*, 63(5-7):e2247–e2256, 2005.

[54] Jane J. Ye. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *J. Math. Anal. Appl.*, 307(1):350–369, 2005.