GLOBAL CONVERGENCE OF AUGMENTED LAGRANGIAN METHODS APPLIED TO OPTIMIZATION PROBLEMS WITH DEGENERATE CONSTRAINTS, INCLUDING PROBLEMS WITH COMPLEMENTARITY CONSTRAINTS*

A. F. IZMAILOV†, M. V. SOLODOV‡, AND E. I. USKOV§

Abstract. We consider global convergence properties of the augmented Lagrangian methods on problems with degenerate constraints, with a special emphasis on mathematical programs with complementarity constraints (MPCC). In the general case, we show convergence to stationary points of the problem under an error bound condition for the feasible set (which is weaker than constraint qualifications), assuming that the iterates have some modest features of approximate local minimizers of the augmented Lagrangian. For MPCC, we first argue that even weak forms of general constraint qualifications that are suitable for convergence of the augmented Lagrangian methods, such as the recently proposed relaxed positive linear dependence condition, should not be expected to hold and thus special analysis is needed. We next obtain a rather complete picture, showing that under the usual in this context MPCC-linear independence constraint qualification feasible accumulation points of the iterates are guaranteed to be C-stationary for MPCC (better than weakly stationary), but in general need not be M-stationary (hence, neither strongly stationary). However, strong stationarity is guaranteed if the generated dual sequence is bounded, which we show to be the typical numerical behaviour even though the multiplier set itself is unbounded. Experiments with the ALGENCAN augmented Lagrangian solver on the MacMPEC and DEGEN collections are reported, with comparisons to the SNOPT and filterSQP implementations of the SQP method, to the MINOS implementation of the linearly constrained Lagrangian method, and to the interior-point solvers IPOPT and KNITRO. We show that ALGENCAN is a very good option if one is primarily interested in robustness and quality of computed solutions.

Key words. Augmented Lagrangian, method of multipliers, mathematical programs with complementarity constraints, degenerate constraints, stationarity.

AMS subject classifications. 90C30, 90C33, 90C55, 65K05.

1. Introduction. Consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{1.1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are smooth. The goal of this paper is to clarify the behavior of the augmented Lagrangian methods in the cases of degenerate constraints. To the best of our knowledge, this question had not been studied in the literature, neither for general degenerate problems nor for the specific case of complementarity constraints (at least when it comes to direct applications of established augmented Lagrangian solvers to the latter; more on this in the sequel). Generally, by degeneracy we mean violation of (more-or-less) standard constraint qualifications at some (or all) feasible points of

*Research of the first and third authors is supported by the Russian Foundation for Basic Research Grants 10-01-00251. The second author is supported in part by CNPq Grant 302637/2011-7, by PRONEX–Optimization, and by FAPERJ.

†Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia. (izmaf@ccas.ru)

‡IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. (solodov@impa.br)

§Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia. (ydoom@narod.ru)

(1.1). In such cases the set of Lagrange multipliers associated with a stationary point of (1.1) need not be a singleton, and can even be unbounded.

An important instance of intrinsically degenerate problems is the class of the so-called mathematical programs with complementarity constraints (MPCC) [42, 46]:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G(x) \geq 0,\ H(x) \geq 0,\ \langle G(x),\, H(x)\rangle \leq 0, \end{array} \tag{1.2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function and $G,\, H : \mathbb{R}^n \to \mathbb{R}^s$ are smooth mappings. As is well known, the fundamental Mangasarian–Fromovitz constraint qualification is violated at every feasible point of (1.2), which means that the multiplier set is always unbounded (assuming it is nonempty). There can be additional "usual" constraints in (1.2), which we omit here since all the essential difficulties are associated with the complementarity constraints. That said, we did check that all the results presented below have evident valid counterparts in the presence of additional constraints. Also, various equivalent forms of the last constraint in (1.2) are possible (see (4.1)–(4.4) below). It was observed in [22] that the form in (1.2) seems to have some numerical advantages. In our computational experiments we tried other options too, but eventually also found (1.2) preferable overall. For this reason, we consider (1.2) from the beginning and later focus our analysis on this form as well.

Augmented Lagrangian methods, also known as methods of multipliers, date back to [30] and [47]; some other key references are [11, 16, 2]. The augmented Lagrangian approach belongs to optimization classic and is the basis for a number of successful solvers, such as LANCELOT [38] and ALGENCAN [1]. At the same time, some improvements in its global [6, 12, 4] and local [19] convergence analysis are very recent. Moreover, these improvements appear quite relevant for problems with degenerate constraints, which in part motivated this paper. For example, in [4] global convergence to stationary points is shown under the so-called relaxed positive linear dependence condition, which is a rather weak constraint qualification allowing various forms of degeneracy and unbounded multiplier sets. Another feature of the augmented Lagrangian methods that looks appealing from the point of view of both global and local convergence in the degenerate setting is that subproblems are unconstrained. This removes at least one difficulty that often shows up in the degenerate cases for methods that linearize constraints, such as SQP [28] and the linearly constrained Lagrangian methods [44] – linearized constraints can be inconsistent.

Moreover, it was recently established in [19] that no constraint qualifications of any kind are needed for local *primal-dual* linear/superlinear convergence rate of the augmented Lagrangian methods, as long as the second-order sufficient optimality condition (SOSC) holds. Our expectation for good global behaviour of these methods on degenerated problems actually comes in part from this local convergence result, as it demonstrates that methods of this class possess an intrinsic dual stabilization property: even when the multiplier set is unbounded, dual sequences tend to converge (at least if SOSC holds). In particular, dual sequences usually remain bounded, which is confirmed by our numerical results in Section 4.

It should be mentioned here that a number of special methods for degenerate problems were developed and analyzed in the last 15 years or so [56, 29, 55, 21, 54, 34, 53, 20, 36], with stabilized SQP being perhaps the most prominent. According to the analysis in [20], stabilized SQP has the same local convergence properties as the augmented Lagrangian methods, and in particular, it possesses local superlinear convergence under the sole SOSC. Moreover, the augmented Lagrangian methods and the stabilized SQP appear to be intrinsically related (see [10, 55] and [11, p. 210]). They also may potentially suffer from similar deficiencies: their subproblems become ill-conditioned if the penalty parameter tends to $+\infty$ for the former, and as the stabilization parameter tends to 0 for the latter. However, all the special methods mentioned above (including stabilized SQP) are essentially local, as at this time no natural globalizations are known (although various hybrid-type strategies are possible that couple these local methods with globally convergent schemes). At the same time, the augmented Lagrangian methods have attractive established global convergence properties. All those considerations motivated us to take a closer look at the behavior of the augmented Lagrangian methods in the degenerate cases. The main goal of this paper is to figure out whether these methods can be a good global strategy in the context of potential degeneracy. Therefore, we concentrate on theoretical global convergence properties, robustness, and quality of the outcome of the augmented Lagrangian methods when applied to degenerate problems and to MPCCs, leaving special modifications intended for increasing efficiency (probably employing the special structure in the case of MPCC) for future research.

In Section 2, we briefly recall the global convergence theory for the ALGENCAN version of the augmented Lagrangian method when applied to general optimization problems, and we discuss some peculiarities of this theory related specifically to the degeneracy issues. To complement the picture we also prove a new global convergence result, assuming that an error bound holds for the feasible set (this is weaker than constraint qualifications, including the relaxed positive linear dependence condition) and assuming that the iterates have some modest features of approximate local minimizers (rather than being merely stationary points of the augmented Lagrangian).

Section 3 is devoted to global convergence analysis in the special case of MPCC. We first put in evidence that special analysis is required indeed. This is because the relaxed positive linear dependence condition, that does the job for general problems [4], should not be expected to hold for MPCC. Under the MPCC-linear independence constraint qualification (standard in this setting), we then show that accumulation points of the augmented Lagrangian iterates are guaranteed to be C-stationary, and they are strongly stationary if the generated dual sequence is bounded. In Section 4, we observe that this is in fact the typical numerical behaviour. Of course, it would have been desirable to obtain strong stationarity of accumulation points without additional assumptions. Some comments are in order. In the case of the specific method in consideration, this is not possible: by examples, we show that accumulation points may not be M-stationary, and thus neither strongly stationary (and also that the MPCC-linear independence constraint qualification cannot be relaxed). That said, to the best of our knowledge there currently exist no *established practical* methods for MPCC with provable global convergence to points possessing stronger properties than C-stationarity under the sole MPCC-linear independence constraint qualification, and without some additional assumptions regarding the accumulation points and the generated iterates (like solving subproblems to second-order optimality); see, e.g., [50, 49, 7, 8]. The only exceptions are some active-set or decomposition methods [25, 26, 15, 27] for the special case of *linear* complementarity constraints, which maintain (approximate) feasibility of the iterates. Thus our global convergence theory is competitive with any other established practical alternative, or is stronger. The only two algorithms with better global convergence properties for general MPCCs appear to be those proposed in [41, 39], but iterations of these methods have some combinatorial features and there is currently no numerical evidence showing that they are competitive, say on MacMPEC [43]. The method in [9] is also combinatorial, not fully general because of some convexity assumptions, and is known to be not very practical anyway.

We further mention that some type of augmented Lagrangian methods for MPCC have been considered in [57, 31]. However, we are interested in behavior of the established solvers (e.g., ALGENCAN [1]) when applied to MPCC (1.2), and the references above are not relevant for this analysis for the following reasons. In [57] the constraints of (1.2) are first reformulated into a nonsmooth system of equations and then the augmented Lagrangian methodology is applied. Solvers such as LANCELOT or ALGENCAN do not handle nonsmooth functions or constraints. In [31], only the last constraint of (1.2), written as an equality, is penalized in the augmented Lagrangian, while the first two constraints are maintained as constraints of the subproblem. Again, the solvers in consideration pass to subproblems only simple bounds or linear constraints, and not general nonlinear constraints like the first two in (1.2). Another somewhat related analysis is [5], where an application of a generic optimization algorithm to MPCC is considered. The assumed convergence properties of this algorithm for general problems are similar to those that hold for augmented Lagrangians. However, the analysis for MPCC in [5] uses rather strong assumptions, like convexity of constraint functions (when it comes to feasibility of accumulation points of the iterates) and lower-level strict complementarity (when it comes to stationarity of feasible accumulation points).

In Section 4, we present numerical results comparing ALGENCAN [1] with the SNOPT [28] and filterSQP [23] implementations of SQP, the MINOS [44] implementation of the linearly constrained Lagrangian method and with two interior-point solvers, IPOPT [32, 52] and KNITRO [37, 14], on two test collections. The first is MacMPEC [43], a well-established test collection of MPCCs. The second is DEGEN [17], which contains various small degenerate optimization problems. We note that a number of optimization solvers were tested on MacMPEC in [22], and SQP was found to work quite well in practice. But augmented Lagrangian and linearly constrained Lagrangian methods had not been tested before, as far as we are aware, at least not on the full MacMPEC. KNITRO and IPOPT-C [48] are two solvers that exploit the special structure of complementarity constraints. Our experiments show that ALGENCAN is at least as robust as the other solvers, or is better. Only IPOPT-C has slightly fewer failures, but ALGENCAN generally terminates with better objective function values than all the other solvers including IPOPT-C, which is important in the sense of "global" optimization. It should also be mentioned that IPOPT-C [48] apparently is not supported by any global convergence theory. ALGENCAN definitely outperforms its competitors in terms of major iterations count, which is the sign of higher convergence rate, in agreement with the local rate of convergence results in [19] that allow any kind of degeneracy. On the other hand, the cost of (especially late) iterations of ALGENCAN is rather high, and so its convergence rate does not translate into saved CPU time. As a result, ALGENCAN may lose in terms of computational costs (e.g., functions evaluations). The ultimate acceleration procedure currently used in ALGENCAN [13] does not seem to help here. Therefore, further development will be needed to increase the efficiency of the method on degenerate problems such as MPCC. However, if robustness and guarantees of convergence to good solutions are of the main concern as opposed to speed (this is certainly the case in some applications) then ALGENCAN is a very good option.

We now describe our main notation. By $\|\cdot\|$ we denote the Euclidean norm, where the space is always clear from the context. When other norms are used, they are specified explicitly, e.g., $\|\cdot\|_\infty$. Given a vector $y \in \mathbb{R}^p$ and an index set $I \subset \{1, \ldots, p\}$, by $y_I \in \mathbb{R}^{|I|}$ we mean a vector comprised by the components of y indexed by I, where $|I|$ is the cardinality of I. For a matrix A of any dimensions, $\ker A$ stands for its null space, and for a linear subspace M of any linear space, $M^\perp$ stands for its orthogonal complement.

2. General global convergence theory. In this section we consider the general mathematical programming problem (1.1). Let $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ be the usual Lagrangian of problem (1.1), i.e.,

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle. \tag{2.1}$$

Then stationary points and associated Lagrange multipliers of problem (1.1) are characterized by the Karush–Kuhn-Tucker optimality system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0, \tag{2.2}$$

with respect to $x \in \mathbb{R}^n$ and $(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$.

Given a penalty parameter $c > 0$, the augmented Lagrangian $L_c : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ for this problem is defined by

$$L_c(x, \lambda, \mu) = f(x) + \frac{1}{2c}(\|\lambda + ch(x)\|^2 + \|\max\{0, \mu + cg(x)\}\|^2), \tag{2.3}$$

where the max operation is applied componentwise. If among the constraints of (1.1) there are simple bounds (or more generally, linear constraints), in practice these are often left out of the augmented Lagrangian and are treated directly, i.e., at each iteration of the algorithm the current augmented Lagrangian is minimized subject to those constraints (see, e.g., [16, 2]). We shall not deal with this generalization, for the sake of keeping the presentation technically simpler. For the same reasons, we shall not talk much about the practically important feature of solving the subproblems approximately, and shall not consider the option of using different penalty parameters for different constraints or nonmonotone choice of penalty parameters. None of this has any specific significance when talking about degeneracy of constraints. We next state the algorithm that will be used for our theoretical analysis. With the simplifications mentioned above, it mimics the one behind the ALGENCAN solver [1] (see [2, Algorithm 3.1], and some very recent modifications/improvements described in [4, 12]).

ALGORITHM 2.1. Choose the following scalar parameters: $\bar{\lambda}_{\min}$ and $\bar{\lambda}_{\max}$ such that $\bar{\lambda}_{\min} \leq \bar{\lambda}_{\max}$, $\bar{\mu}_{\max} \geq 0$, $c_0 > 0$, $\theta \in [0, 1)$, and $\rho > 1$. Set $k = 0$.

1. Choose $(\bar{\lambda}^k, \bar{\mu}^k) \in \mathbb{R}^l \times \mathbb{R}^m$ such that $\bar{\lambda}_{\min} \leq \bar{\lambda}_j^k \leq \bar{\lambda}_{\max}$ $\forall\, j = 1, \ldots, l$, $0 \leq \bar{\mu}_i^k \leq \bar{\mu}_{\max}$ $\forall\, i = 1, \ldots, m$. (For $k > 0$, the typical option is to take $(\bar{\lambda}^k, \bar{\mu}^k)$ as the Euclidian projection of (λ^k, μ^k), defined in step 2, onto the box $\bigotimes_{j=1}^l [\bar{\lambda}_{\min}, \bar{\lambda}_{\max}] \times \bigotimes_{i=1}^m [0, \bar{\mu}_{\max}]$.)
   Compute $x^{k+1} \in \mathbb{R}^n$ as a stationary point of the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & L_{c_k}(x, \bar{\lambda}^k, \bar{\mu}^k) \\ \text{subject to} & x \in \mathbb{R}^n. \end{array} \tag{2.4}$$

2. Set

$$\lambda^{k+1} = \bar{\lambda}^k + c_k h(x^{k+1}), \quad \mu^{k+1} = \max\{0,\, \bar{\mu}^k + c_k g(x^{k+1})\}, \tag{2.5}$$

$$\tau^{k+1} = \min\{\mu^{k+1},\, -g(x^{k+1})\}. \tag{2.6}$$

3. If $k = 0$ or

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

select any $c_{k+1} \geq c_k$. Otherwise select $c_{k+1} \geq \rho c_k$. Adjust k by 1, and go to step 1.

Global convergence analysis of Algorithm 2.1 considers separately the cases of feasible and infeasible accumulations points. As is well understood, convergence of the augmented Lagrangian methods to infeasible points cannot be ruled out, in general. However, this has nothing to do with possible degeneracy of constraints *at solutions or feasible points*, which is the main issue in this work. The possibility of convergence to infeasible points is a general weakness of augmented Lagrangian algorithms. One may look for more special conditions giving feasibility of accumulation points when the constraints have some structure; e.g., [5]. We next consider the case of feasible accumulation points.

The following discussion essentially corresponds to [2, Theorems 4.1, 4.2], taking into account some very recent improvements in [4]. The key ingredients are the following.

From (2.1), (2.3) and (2.5) one can immediately see that for each $k = 1, 2, \ldots$, the iterates generated by Algorithm 2.1 satisfy the conditions

$$\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) = \frac{\partial L_{c_{k-1}}}{\partial x}(x^k,\, \bar{\lambda}^{k-1},\, \bar{\mu}^{k-1}) = 0, \quad \mu^k \geq 0. \tag{2.7}$$

Consider the case when the sequence $\{x^k\}$ has a feasible accumulation point. From (2.5) and (2.6), from the boundedness of $\{\bar{\mu}^k\}$, and from the rule for updating the penalty parameter in step 3 of Algorithm 2.1, it follows that if for some $K \subset \{0, 1, \ldots\}$ the subsequence $\{x^k \mid k \in K\}$ converges to a feasible $\bar{x}$ then

$$\{\mu^k_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} \mid k \in K\} \to 0 \quad \text{as} \quad k \to \infty. \tag{2.8}$$

Therefore, if the sequence $\{(\lambda^k,\, \mu^k) \mid k \in K\}$ has an accumulation point $(\bar{\lambda},\, \bar{\mu})$ (in particular, when this sequence is bounded) then passing onto the limit in (2.7) along the appropriate subsequence, and using (2.8), one immediately derives that $\bar{x}$ is a stationary point of problem (1.1), while $(\bar{\lambda},\, \bar{\mu})$ is an associated Lagrange multiplier.

Observe also that since $\{(\bar{\lambda}^k,\, \bar{\mu}^k)\}$ is bounded, according to (2.5) the entire sequence $\{(\lambda^k,\, \mu^k)\}$ is bounded whenever the sequence of penalty parameters $\{c_k\}$ is bounded. Therefore, any possible difficulties in convergence analysis are concerned with the case when $c_k \to +\infty$ and $\{(\lambda^k,\, \mu^k)\}$ is unbounded. And this is the point where CQs come into play.

Let $A(\bar{x}) = \{i = 1,\, \ldots,\, m \mid g_i(\bar{x}) = 0\}$ be the index set of active inequality constraints. The Mangasarian–Fromovitz constraint qualification (MFCQ) holds at a feasible point $\bar{x}$ of problem (1.1) if the system

$$(h'(\bar{x}))^{\mathrm{T}}\eta + (g'_{A(\bar{x})}(\bar{x}))^{\mathrm{T}}\tilde{\zeta} = 0, \quad \tilde{\zeta} \geq 0,$$

in the variables $(\eta, \tilde{\zeta}) \in \mathbb{R}^l \times \mathbb{R}^{|A(\bar{x})|}$ has only the zero solution. (This is the dual form of MFCQ; from it one can easily see that the nonempty set of multipliers is bounded if and only if MFCQ holds.)

The consequence for Algorithm 2.1 is that MFCQ immediately excludes the case of unbounded $\{(\lambda^k, \mu^k) \mid k \in K\}$ (dividing the equality in (2.7) by $\|(\lambda^k, \mu^k)\|$, passing onto the limit and taking into account (2.8), gives a contradiction with MFCQ).

A more involved argument in [4] shows that the following (weaker) CQ can be used in the analysis of Algorithm 2.1.

For a feasible point $\bar{x}$ of problem (1.1), let $J \subset \{1, \ldots, l\}$ be such that $\{h'_j(\bar{x}) \mid j \in J\}$ is a basis in the linear subspace spanned by $\{h'_j(\bar{x}) \mid j = 1, \ldots, l\}$. We say that the relaxed constant positive linear dependence constraint qualification (RCPLD) [4] holds at a feasible point $\bar{x}$ of problem (1.1) if there exists a neighborhood U of $\bar{x}$ such that:

1. It holds that $\operatorname{rank} h'(x)$ is constant for all $x \in U$.
2. For every $I \subset A(\bar{x})$, if there exist $\tilde{\eta} \in \mathbb{R}^{|J|}$ and $\tilde{\zeta} \in \mathbb{R}^{|I|}$, not all equal to zero and such that

$$(h'_J(\bar{x}))^{\mathrm{T}}\tilde{\eta} + (g'_I(\bar{x}))^{\mathrm{T}}\tilde{\zeta} = 0, \quad \tilde{\zeta} \geq 0, \tag{2.9}$$

then

$$\operatorname{rank}\begin{pmatrix} h'_J(x) \\ g'_I(x) \end{pmatrix} < |J| + |I| \quad \forall\, x \in U. \tag{2.10}$$

This condition is indeed a relaxed version of the constant positive linear dependence constraint qualification (CPLD) at a feasible point $\bar{x}$ of problem (1.1), which consists of saying that there exists a neighborhood U of $\bar{x}$ such that for any $J \subset \{1, \ldots, l\}$ and $I \subset A(\bar{x})$, if there exist $\tilde{\eta} \in \mathbb{R}^{|J|}$ and $\tilde{\zeta} \in \mathbb{R}^{|I|}$ not all equal to zero and such that (2.9) holds, then (2.10) holds as well. It is easy to see that MFCQ implies CPLD and, hence, RCPLD. (As an aside, we note that CPLD is also implied by the well-known constant rank constraint qualification.)

The next statement summarizes convergence properties of Algorithm 2.1 when it comes to feasible accumulation points [2, 4].

THEOREM 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable on $\mathbb{R}^n$. Let $\{x^k\} \subset \mathbb{R}^n$ be a sequence generated by Algorithm 2.1, and let $\bar{x} \in \mathbb{R}^n$ be an accumulation point of this sequence.*

If the point $\bar{x}$ is feasible in problem (1.1) *and RCPLD holds at $\bar{x}$, then $\bar{x}$ is a stationary point of problem* (1.1).

Moreover, if MFCQ holds at $\bar{x}$ then for any infinite set $K \subset \{0, 1, \ldots\}$ such that the subsequence $\{x^k \mid k \in K\}$ converges to $\bar{x}$, the corresponding subsequence $\{(\lambda^k, \mu^k) \mid k \in K\}$ is bounded and each of its accumulation points is a Lagrange multiplier associated to $\bar{x}$.

In the absence of RCPLD the assertion of this theorem is not valid, which can be seen from very simple examples.

EXAMPLE 2.1. Let $n = l = 1$, $m = 0$, $f(x) = x$, $h(x) = x^2$. Problem (1.1) with this data has the unique feasible point (hence, the unique solution) $\bar{x} = 0$, which is nonstationary and violates RCPLD. We have

$$L_c(x, \lambda) = x + \frac{1}{2c}\lambda^2 + \lambda x^2 + \frac{c}{2}x^4, \quad \frac{\partial L_c}{\partial x}(x, \lambda) = 1 + 2\lambda x + 2cx^3.$$

Taking, e.g., $\bar{\lambda}^k = 0$ and any $c_k > 0$, we obtain that the point $x^{k+1} = -1/(2c_k)^{1/3}$ is the unique stationary point and the global solution of the subproblem (2.4). If $c_k \to +\infty$, the sequence $\{x^k\}$ tends to $\bar{x}$.

In principle, the convergence results outlined above are fully relevant in the presence of degenerate solutions: the theory allows for such solutions to be accumulation points of the iterates generated by the augmented Lagrangian methods, and this is good news. The problems arise when there exist *multiple* feasible points which are degenerate in the sense of violating even weak conditions like RCPLD, and which are not optimal/stationary. The situation becomes even worse when all feasible points are expected to be degenerate. Theorem 2.1 allows for convergence to any of such feasible points, which of course is not satisfactory. In the case of MPCC, each feasible point violates MFCQ. It turns out that (R)CPLD can be satisfied. However, in Section 3 we shall show that RCPLD is an extremely atypical property in the context of MPCC and should not be expected to hold. Thus, independent analysis would be presented for MPCC, based on the special structure of this problem.

To complete the picture concerning feasible accumulation points, we next prove a new result, which provides some further insight. We make two assumptions, which are not standard in this context but are very reasonable and should hold in many cases of interest. One assumption is the Lipschitzian error bound (see (2.12) below) on the distance to the feasible set of problem (1.1) in terms of constraints violations. Note that this is weaker than assuming a constraint qualification at $\bar{x}$: RCPLD (and hence, MFCQ) implies the error bound (2.12) [4], while the error bound does not imply RCPLD [4, Counter-example 3]. Our second assumption is that the iterates computed by the algorithm are not merely stationary points of subproblems (2.4) but have at least some modest features of minimizers, even if approximate minimizers, in the sense of (2.11) below. This requirement seems not stringent at all computationally, considering that x^k is generated minimizing $L_{c_{k-1}}(\cdot,\, \bar{\lambda}^{k-1},\, \bar{\mu}^{k-1})$. It is satisfied automatically if x^k is a solution of the subproblem, which is global in some neighbourhood of the accumulation point $\bar{x}$. The latter, in turn, holds automatically under any of a number of conditions ensuring local convexity of the augmented Lagrangian on some fixed (independent of k) ball around the accumulation point in question.

THEOREM 2.2. *Let D stand for the feasible set of problem* (1.1). *Under the assumptions of Theorem 2.1, suppose that there exists an infinite set $K \subset \{1, 2, \ldots\}$ such that $\{x^k \mid k \in K\}$ converges to $\bar{x} \in D$, and for each $k \in K$, some $\beta > 0$, $\theta \in (0, 1/2)$, and some projection $\bar{x}^k$ of x^k onto D, it holds that*

$$L_{c_{k-1}}(x^k,\, \bar{\lambda}^{k-1},\, \bar{\mu}^{k-1}) \leq L_{c_{k-1}}(\bar{x}^k,\, \bar{\lambda}^{k-1},\, \bar{\mu}^{k-1}) + \beta(\|h(x^k)\| + \|\max\{0,\, g(x^k)\}\|) + \frac{\theta\|(\lambda^k,\, \mu^k)\|^2}{c_{k-1}}. \quad (2.11)$$

Assume further that the error bound

$$\mathrm{dist}(x^k,\, D) = O(\|h(x^k)\| + \|\max\{0,\, g(x^k)\}\|). \quad (2.12)$$

holds for $k \in K$.

Then the sequence $\{(\lambda^k,\, \mu^k) \mid k \in K\}$ is bounded, $\bar{x}$ is a stationary point of problem (1.1)*, and every accumulation point of $\{(\lambda^k,\, \mu^k) \mid k \in K\}$ is a Lagrange multiplier associated with $\bar{x}$.*

(Since the feasible set D need not be convex, projection of x^k onto D, i.e., a point in D closest to x^k in the norm used to define dist in (2.12), is not necessarily unique. In the statement above, any projection $\bar{x}^k$ is appropriate.)

Proof. Employing (2.3) and (2.5) we obtain that

$$
\begin{aligned}
L_{c_{k-1}}(x^k, \bar{\lambda}^{k-1}, \bar{\mu}^{k-1}) &= f(x^k) + \frac{1}{2c_{k-1}}(\|\bar{\lambda}^{k-1} + c_{k-1}h(x^k)\|^2 \\
&\quad + \|\max\{0, \bar{\mu}^{k-1} + c_{k-1}g(x^k)\}\|^2) \\
&= f(x^k) + \frac{\|(\lambda^k, \mu^k)\|^2}{2c_{k-1}}.
\end{aligned}
$$

On the other hand, since $\bar{x}^k$ is feasible in (1.1), it holds that $h(\bar{x}^k) = 0$ and also $\max\{0, \bar{\mu}^{k-1} + c_{k-1}g(\bar{x}^k)\} \leq \bar{\mu}^{k-1}$, so that

$$
\begin{aligned}
L_{c_{k-1}}(\bar{x}^k, \bar{\lambda}^{k-1}, \bar{\mu}^{k-1}) &= f(\bar{x}^k) + \frac{1}{2c_{k-1}}(\|\bar{\lambda}^{k-1} + c_{k-1}h(\bar{x}^k)\|^2 \\
&\quad + \|\max\{0, \bar{\mu}^{k-1} + c_{k-1}g(\bar{x}^k)\}\|^2) \\
&\leq f(\bar{x}^k) + \frac{\|(\bar{\lambda}^{k-1}, \bar{\mu}^{k-1})\|^2}{2c_{k-1}}.
\end{aligned}
$$

Combining the last two relations with (2.11), we obtain that

$$
\begin{aligned}
\frac{\|(\lambda^k, \mu^k)\|^2}{2c_{k-1}} &\leq f(\bar{x}^k) - f(x^k) + \frac{\|(\bar{\lambda}^{k-1}, \bar{\mu}^{k-1})\|^2}{2c_{k-1}} \\
&\quad + \beta(\|h(x^k)\| + \|\max\{0, g(x^k)\}\|) + \frac{\theta\|(\lambda^k, \mu^k)\|^2}{c_{k-1}}.
\end{aligned}
\tag{2.13}
$$

Employing local Lipschitz-continuity of a continuously differentiable function f (with some constant $\ell > 0$), the multiplier update rule (2.5), the error bound (2.12) and inequality (2.13), as well as boundedness of $\{(\bar{\lambda}^k, \bar{\mu}^k)\}$ and inequality $\bar{\mu}^k \geq 0$ for all k, we derive that there exists a constant $M > 0$ such that

$$
\begin{aligned}
(1 - 2\theta)\|(\lambda^k, \mu^k)\|^2 &\leq 2\ell c_{k-1}\|\bar{x}^k - x^k\| + \|(\bar{\lambda}^{k-1}, \bar{\mu}^{k-1})\|^2 \\
&\quad + 2\beta c_{k-1}(\|h(x^k)\| + \|\max\{0, g(x^k)\}\|) \\
&\leq M(\|c_{k-1}h(x^k)\| + \|\max\{0, c_{k-1}g(x^k)\}\| + 1) \\
&\leq M(\|\lambda^k - \bar{\lambda}^{k-1}\| + \|\mu^k\| + 1) \\
&\leq M(\|\lambda^k\| + \|\mu^k\| + \|\bar{\lambda}^{k-1}\| + 1).
\end{aligned}
$$

This chain of relations would result in a contradiction if $\{(\lambda^k, \mu^k)\}$ were to be unbounded. We therefore conclude that $\{(\lambda^k, \mu^k)\}$ is bounded. The assertion now follows immediately, by passing onto the limits along convergent subsequences in (2.7), and taking into account (2.8). ∎

In particular, under the assumption (2.11), any condition implying the Lipschitzian error bound (2.12) (e.g., RCPLD [4]) ensures boundedness of $\{(\lambda^k, \mu^k) \mid k \in K\}$ and stationarity of the accumulation point $\bar{x}$.

REMARK 2.1. In view of (2.7), the test in step 3 of Algorithm 2.1 to decide on the increase of the penalty parameter is nothing more than a linear decrease test from one iteration to the next for the natural residual measuring the violation of the KKT conditions (2.2) for problem (1.1). This test was introduced in this context in [12]. Previous versions of ALGENCAN (e.g., [2]) were using instead of τ^{k+1} given by (2.6) the following:

$$
\tau^{k+1} = \max\left\{g(x^{k+1}), -\frac{\bar{\mu}^k}{c_k}\right\}.
$$

It can be seen that all the statements and comments above, as well as in the next section on MPCC, also hold for this version of the method. In fact, the analysis simplifies somewhat, as property (2.8) would be replaced by the stronger

$$\mu^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0$$

for all $k \in K$ large enough.

3. Global convergence for MPCC. In this section we consider MPCC (1.2). Of course, the "usual" equality and inequality constraints can also appear in MPCC setting. As already mentioned, we drop them for brevity; all the statements do extend to the more general case in an obvious manner.

The first issue to settle is whether Theorem 2.1 can already be expected to provide an adequate answer for MPCC (then no further analysis is needed). Recall that in the case of MPCC each feasible point violates MFCQ. But what about the weaker RCPLD? It can be seen that it is possible to construct examples where RCPLD and even CPLD hold. At the same time, the considerations that follow make it clear that such examples are completely artificial and RCPLD cannot be expected to hold with any frequency in cases of interest.

To show that CPLD may hold in principle, consider MPCC (1.2) with the mappings $G,\, H\colon \mathbb{R}^n \to \mathbb{R}^s$ being identically zero. Then each point $\bar{x} \in \mathbb{R}^n$ is feasible and satisfies CPLD: for any $x \in \mathbb{R}^n$, the Jacobian of constraints is identically zero, and thus any subset of its rows is linearly dependent.

However, the latter example is of course extremely pathological. We next argue that for practical MPCCs (R)CPLD cannot be expected to hold. Since the constraints in (1.2) imply that s equalities

$$G_i(x)H_i(x) = 0, \quad i = 1,\, \ldots,\, s, \tag{3.1}$$

hold at any feasible point, in practical problems s typically should be no greater than $n-1$. Otherwise, there remains no "degrees of freedom" for optimization, and MPCC becomes essentially a feasibility problem (similar to the nonlinear complementarity problem). For $s \geq n$ (R)CPLD can hold in a stable way, but in the context of MPCC the relevant case is $s \leq n-1$. Considering the constraints in (1.2) as a particular case of usual inequality constraints, as in (1.1), observe that the Jacobian of these constraints at any $x \in \mathbb{R}^n$ has the form

$$\begin{pmatrix} -G'(x) \\ -H'(x) \\ (G'(x))^{\mathrm{T}} H(x) + (H'(x))^{\mathrm{T}} G(x) \end{pmatrix}. \tag{3.2}$$

According to (3.1), for a given feasible point $\bar{x} \in \mathbb{R}^n$ of MPCC (1.2) we can assume without loss of generality that $H(\bar{x}) = 0$ (this can always be achieved moving the positive components of H to G, and the corresponding zero components of G to H). Then the last row of the Jacobian (3.2) at $\bar{x}$ has the form $\sum_{i=1}^s G_i(\bar{x}) H_i'(\bar{x})$ with nonnegative coefficients $G_i(\bar{x})$, $i = 1,\, \ldots,\, s$. Hence, the rows $-H_i'(\bar{x})$, $i = 1,\, \ldots,\, s$, of this Jacobian combined with this last row are positively linearly dependent. Therefore, for (R)CPLD to hold at $\bar{x}$, the vectors $-H_i'(x)$, $i = 1,\, \ldots,\, s$, and $\sum_{i=1}^s (G_i(x)H_i'(x) + H_i(x)G_i'(x))$ must be linearly dependent for all $x \in \mathbb{R}^n$ close enough to $\bar{x}$. If $s \leq n-1$, this system contains no more than n rows. The property of no more than n vectors dependent on $x \in \mathbb{R}^n$ to be linearly dependent in $\mathbb{R}^n$ for all small perturbations of

$\bar{x}$ is *extremely* atypical (generically, in the situation at hand small perturbations of a linearly dependent system give a linearly independent one).

The conclusion is that RCPLD cannot be expected to hold for MPCC in cases of interest, and Theorem 2.1 is not relevant for MPCC. Thus, an independent analysis of MPCC is needed, that takes into account the problem structure. To this end, we need to introduce some terminology, all standard in the MPCC literature.

We define the usual Lagrangian $L : \mathbb{R}^n \times (\mathbb{R}^s \times \mathbb{R}^s \times \mathbb{R}) \to \mathbb{R}$ of problem (1.2) by

$$L(x, \mu) = f(x) - \langle \mu_G, G(x) \rangle - \langle \mu_H, H(x) \rangle + \mu_0 \langle G(x), H(x) \rangle \tag{3.3}$$

(which fully agrees with (2.3)), and the family of the augmented Lagrangians L_c : $\mathbb{R}^n \times (\mathbb{R}^s \times \mathbb{R}^s \times \mathbb{R}) \to \mathbb{R}$ by

$$L_c(x, \mu) = f(x) + \frac{1}{2c}(\| \max\{0, \mu_G - cG(x)\}\|^2 + \| \max\{0, \mu_H - cH(x)\}\|^2 + (\max\{0, \mu_0 + c\langle G(x), H(x)\rangle\})^2)$$

(which fully agrees with (2.3)), where $\mu = (\mu_G, \mu_H, \mu_0)$. We also define the so-called MPCC-Lagrangian $\mathcal{L} : \mathbb{R}^n \times (\mathbb{R}^s \times \mathbb{R}^s) \to \mathbb{R}$ by

$$\mathcal{L}(x, \lambda) = f(x) - \langle \lambda_G, G(x) \rangle - \langle \lambda_H, H(x) \rangle,$$

where $\lambda = (\lambda_G, \lambda_H)$.

For a point $\bar{x} \in \mathbb{R}^n$ feasible in problem (1.2), define the index sets

$$I_G(\bar{x}) = \{i = 1, \ldots, s \mid G_i(\bar{x}) = 0\}, \quad I_H(\bar{x}) = \{i = 1, \ldots, s \mid H_i(\bar{x}) = 0\},$$
$$I_0(\bar{x}) = I_G(\bar{x}) \cap I_H(\bar{x}).$$

Observe that by necessity $I_G(\bar{x}) \cup I_H(\bar{x}) = \{1, \ldots, s\}$. Recall that the MPCC-linear independence constraint qualification (MPCC-LICQ) consists of saying that

$$G_i'(\bar{x}),\ i \in I_G(\bar{x}), \quad H_i'(\bar{x}),\ i \in I_H(\bar{x}) \quad \text{are linearly independent.}$$

A feasible point $\bar{x}$ of problem (1.2) is referred to as weakly stationary if there exists $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H) \in \mathbb{R}^s \times \mathbb{R}^s$ satisfying

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \bar{\lambda}) = 0, \quad (\bar{\lambda}_G)_{I_H(\bar{x}) \setminus I_G(\bar{x})} = 0,\ (\bar{\lambda}_H)_{I_G(\bar{x}) \setminus I_H(\bar{x})} = 0.$$

If, in addition,

$$(\bar{\lambda}_G)_i (\bar{\lambda}_H)_i \geq 0 \quad \forall\, i \in I_0(\bar{x}),$$

then $\bar{x}$ is a C-stationary point. If

$$\forall\, i \in I_0(\bar{x}) \text{ either } (\bar{\lambda}_G)_i (\bar{\lambda}_H)_i = 0 \text{ or } (\bar{\lambda}_G)_i > 0,\ (\bar{\lambda}_H)_i > 0,$$

then $\bar{x}$ is an M-stationary point. And if

$$(\bar{\lambda}_G)_{I_0(\bar{x})} \geq 0, \quad (\bar{\lambda}_H)_{I_0(\bar{x})} \geq 0,$$

then $\bar{x}$ is a strongly stationary point. Evidently, all these stationarity concepts are the same in the case of lower-level strict complementarity, i.e., when $I_0(\bar{x}) = \emptyset$.

In the case of MPCC (1.2), the counterpart of (2.5) is the following:

$$\mu_G^{k+1} = \max\{0,\, \bar{\mu}_G^k - c_k G(x^{k+1})\}, \quad \mu_H^{k+1} = \max\{0,\, \bar{\mu}_H^k - c_k H(x^{k+1})\}, \tag{3.4}$$

$$\mu_0^{k+1} = \max\{0,\, \bar{\mu}_0^k + c_k \langle G(x^{k+1}),\, H(x^{k+1})\rangle\}. \tag{3.5}$$

The key relations (2.7) and (2.8) take the following form:

$$\frac{\partial L}{\partial x}(x^k,\, \mu^k) = 0, \quad \mu^k = (\mu_G^k,\, \mu_H^k,\, \mu_0^k) \geq 0 \tag{3.6}$$

for all $k = 1, 2, \ldots$, and if for some $K \subset \{0, 1, \ldots\}$ the subsequence $\{x^k \mid k \in K\}$ converges to a feasible $\bar{x}$ then

$$\{(\mu_G^k)_{I_H(\bar{x})\setminus I_G(\bar{x})} \mid k \in K\} \to 0, \quad \{(\mu_H^k)_{I_G(\bar{x})\setminus I_H(\bar{x})} \mid k \in K\} \to 0 \tag{3.7}$$

for $k \in K$ as $k \to \infty$.

We start the analysis of Algorithm 2.1 with a technical lemma, to be used in the sequel.

LEMMA 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be differentiable in some neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being continuous at $\bar{x}$. Assume that* $\operatorname{rank} h'(\bar{x}) = l$. *Let $\{x^k\} \subset \mathbb{R}^n$ be a sequence which converges to $\bar{x}$. Suppose that the equality*

$$\alpha_k f'(x^k) - (h'(x^k))^{\mathrm{T}} \lambda^k = \omega^k \tag{3.8}$$

holds for some $\alpha_k \in \mathbb{R}$, $\lambda^k \in \mathbb{R}^l$ and $\omega^k \in \mathbb{R}^n$ for all k.

If $\alpha_k \to \bar{\alpha}$ and $\{\omega^k\} \to 0$ as $k \to \infty$ then there exists the unique $\bar{\lambda} \in \mathbb{R}^s$ such that $\{\lambda^k\} \to \bar{\lambda}$ as $k \to \infty$ and

$$\bar{\alpha} f'(\bar{x}) - (h'(\bar{x}))^{\mathrm{T}} \bar{\lambda} = 0. \tag{3.9}$$

In particular, if $\bar{\alpha} = 0$ then $\{\lambda^k\} \to 0$.

Proof. Assume first that the sequence $\{\lambda^k\}$ is unbounded. Then there exists an infinite set $K \subset \{0, 1, \ldots\}$ such that $\{\|\lambda^k\| \mid k \in K\} \to +\infty$. Hence, dividing (3.8) by $\|\lambda^k\|$ and considering any convergent subsequence of the bounded sequence $\{\lambda^k/\|\lambda^k\| \mid k \in K\}$, we obtain a contradiction with the condition $\operatorname{rank} h'(\bar{x}) = l$.

Therefore $\{\lambda^k\}$ is bounded. Then, passing onto the limit along any convergent subsequence, we obtain the equality (3.9) for some $\bar{\lambda} \in \mathbb{R}^l$. The uniqueness of $\bar{\lambda}$ follows from the assumption that $\operatorname{rank} h'(\bar{x}) = l$. ◻

The following theorem establishes that feasible accumulation points of Algorithm 2.1 applied to MPCC are guaranteed to be C-stationary (better than weakly stationary) provided MPCC-LICQ holds, and they are strongly stationary if a certain dual sequence is bounded. Together with examples below that show that in general M-stationarity (and thus also strong stationarity) may not hold, this gives a complete picture of global convergence properties of the augmented Lagrangian method when applied to MPCC. While these properties are not ideal, as discussed in the Introduction, they are currently fully competitive with any other practical algorithm applied to MPCC. Also, Section 4 shows that strong stationarity is in fact usually achieved in practice.

THEOREM 3.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $G, H : \mathbb{R}^n \to \mathbb{R}^s$ be continuously differentiable on $\mathbb{R}^n$. Let $\{x^k\} \subset \mathbb{R}^n$ be a sequence generated by Algorithm 2.1 applied to* *problem* (1.2)*, and let* $\bar{x} \in \mathbb{R}^n$ *be an accumulation point of this sequence. Suppose that the point* $\bar{x}$ *is feasible in problem* (1.2)*, and that MPCC-LICQ holds at* $\bar{x}$*.*

If the sequence $\{\mu_0^k\}$ *generated by Algorithm 2.1 has a bounded subsequence such that the corresponding subsequence of* $\{x^k\}$ *converges to* $\bar{x}$ *(in particular, if the sequence* $\{c_k\}$ *is bounded) then* $\bar{x}$ *is a strongly stationary point of problem* (1.2)*. Otherwise* $\bar{x}$ *is at least a C-stationary point of problem* (1.2)*.*

Proof. We assume without loss of generality that $G(\bar{x}) = 0$ (as already commented above, this can always be achieved by moving the positive components of G to H and the corresponding zero components of H to G). We next define the index sets

$$I_0 = \{\, i = 1, \ldots, s \mid H_i(\bar{x}) = 0 \,\}, \quad I_+ = \{\, i = 1, \ldots, s \mid H_i(\bar{x}) > 0 \,\}. \tag{3.10}$$

It is clear that in the assumed case of $G(\bar{x}) = 0$, we have that $I_0 = I_0(\bar{x})$ and $I_+ = I_G(\bar{x}) \setminus I_H(\bar{x})$, where $I_0(\bar{x})$, $I_G(\bar{x})$ and $I_H(\bar{x})$ were defined above. In this case MPCC-LICQ at $\bar{x}$ takes the form

$$\operatorname{rank} \begin{pmatrix} G'(\bar{x}) \\ H'_{I_0}(\bar{x}) \end{pmatrix} = s + |I_0|.$$

From (3.3) and (3.6) we have that

$$\begin{aligned} 0 &= f'(x^k) - (G'(x^k))^{\mathrm{T}}\big(\mu_G^k - \mu_0^k H(x^k)\big) - (H'(x^k))^{\mathrm{T}}\big(\mu_H^k - \mu_0^k G(x^k)\big) \\ &= f'(x^k) - \big(G'_{I_0}(x^k)\big)^{\mathrm{T}}(\lambda_G^k)_{I_0} - \big(G'_{I_+}(x^k)\big)^{\mathrm{T}}(\lambda_G^k)_{I_+} \\ &\quad - \big(H'_{I_0}(x^k)\big)^{\mathrm{T}}(\lambda_H^k)_{I_0} - \big(H'_{I_+}(x^k)\big)^{\mathrm{T}}(\lambda_H^k)_{I_+}, \end{aligned} \tag{3.11}$$

where

$$\lambda_G^k = \mu_G^k - \mu_0^k H(x^k), \quad \lambda_H^k = \mu_H^k - \mu_0^k G(x^k). \tag{3.12}$$

Let $K \subset \{0, 1, \ldots\}$ be an infinite index set such that the sequence $\{x^k \mid k \in K\}$ converges to $\bar{x}$. As discussed above, if $\{c_k\}$ is bounded then $\bar{x}$ is a stationary point in the usual KKT sense. For MPCC this is equivalent to strong stationarity of $\bar{x}$. Therefore, from now on we consider the case $c_k \to +\infty$. In this case from (3.4) and (3.10) we obtain that

$$(\mu_H^k)_{I_+} = 0 \tag{3.13}$$

for all k large enough.

From the second condition in (3.12) and from (3.13) it follows that if $\{\mu_0^k \mid k \in K_0\}$ is bounded for some infinite set $K_0 \subset K$ then the sequence $\{(\lambda_H^k)_{I_+} \mid k \in K_0\}$ tends to zero. Applying Lemma 3.1 to (3.11) with $\alpha_k = 1$, we obtain that $\{\lambda_G^k \mid k \in K_0\}$ and $\{(\lambda_H^k)_{I_0} \mid k \in K_0\}$ converge to some $\bar{\lambda}_G$ and $(\bar{\lambda}_H)_{I_0}$, respectively, and

$$f'(\bar{x}) - (G'(\bar{x}))^{\mathrm{T}}\bar{\lambda}_G - (H'_{I_0}(\bar{x}))^{\mathrm{T}}(\bar{\lambda}_H)_{I_0} = 0. \tag{3.14}$$

Passing onto the limit in (3.12) along K_0, we conclude that $(\bar{\lambda}_G)_{I_0} \geq 0$ and $(\bar{\lambda}_H)_{I_0} \geq 0$ and therefore $\bar{x}$ is a strongly stationary point. Note that the above argument is also applicable if $I_0 = \emptyset$ or $I_+ = \emptyset$.

It remains to consider the case when $\{\mu_0^k \mid k \in K\} \to +\infty$. We first prove that in this case I_0 is nonempty and

$$\frac{c_{k-1}}{\mu_0^k}\langle G_{I_0}(x^k),\, H_{I_0}(x^k)\rangle \to M > 0 \tag{3.15}$$

for $k \in K$ tending to infinity.

Since $\{\mu_0^k \mid k \in K\} \to +\infty$, it follows from (3.5) that

$$\mu_0^k = \bar{\mu}_0^{k-1} + c_{k-1}\langle G(x^k),\, H(x^k)\rangle.$$

Therefore,

$$\langle G(x^k),\, H(x^k)\rangle > 0 \tag{3.16}$$

for all $k \in K$ large enough, and

$$\frac{c_{k-1}}{\mu_0^k}\langle G(x^k),\, H(x^k)\rangle = 1 - \frac{\bar{\mu}_0^{k-1}}{\mu_0^k} \to 1 \tag{3.17}$$

as $k \to \infty$. It is also evident that in this case

$$\frac{\mu_0^k}{c_{k-1}} = \langle G(x^k),\, H(x^k)\rangle + \frac{\bar{\mu}_0^{k-1}}{c_{k-1}},$$

and therefore

$$\frac{\mu_0^k}{c_{k-1}} \to 0. \tag{3.18}$$

It is clear that if $I_+ = \emptyset$ (and, evidently, $I_0 \neq \emptyset$) then (3.15) immediately follows from (3.17).

Suppose that $I_+ \neq \emptyset$. From the second equality in (3.12) and from (3.13) we obtain that

$$\frac{(\lambda_H^k)_{I_+}}{\mu_0^k} = -G_{I_+}(x^k) \tag{3.19}$$

for all $k \in K$ large enough. Hence, $\{(\lambda_H^k)_{I_+}/\mu_0^k \mid k \in K\} \to 0$ as $k \to \infty$. Dividing (3.11) by μ_0^k and applying Lemma 3.1 with $\alpha_k = 1/\mu_0^k$, we get

$$\left\{\frac{\lambda_G^k}{\mu_0^k} \;\middle|\; k \in K\right\} \to 0, \quad \left\{\frac{(\lambda_H^k)_{I_0}}{\mu_0^k} \;\middle|\; k \in K\right\} \to 0. \tag{3.20}$$

From (3.12) we have

$$\frac{(\lambda_G^k)_{I_+}}{\mu_0^k} = \frac{(\mu_G^k)_{I_+}}{\mu_0^k} - H_{I_+}(x^k),$$

and since $H_{I_+}(\bar{x}) > 0$, the first condition in (3.20) implies that $\{(\mu_G^k)_i \mid k \in K\} \to +\infty$ for all $i \in I_+$. Then it follows from (3.4) that $G_{I_+}(x^k) < 0$ for all $k \in K$ large enough. Thus, from the first equality in (3.12), we obtain that for such k,

$$\begin{aligned}(\lambda_G^k)_{I_+} &= \max\{0,\, (\bar{\mu}_G^{k-1})_{I_+} - c_{k-1}G_{I_+}(x^k)\} - \mu_0^k H_{I_+}(x^k)\\ &= (\bar{\mu}_G^{k-1})_{I_+} - c_{k-1}G_{I_+}(x^k) - \mu_0^k H_{I_+}(x^k).\end{aligned}$$

Therefore

$$\frac{c_{k-1}}{\mu_0^k}G_{I_+}(x^k) + H_{I_+}(x^k) = \frac{(\bar{\mu}_G^{k-1})_{I_+}}{\mu_0^k} - \frac{(\lambda_G^k)_{I_+}}{\mu_0^k}.$$

Then from the first relation in (3.20) we obtain

$$\left\{ \frac{c_{k-1}}{\mu_0^k} G_{I_+}(x^k) + H_{I_+}(x^k) \;\Big|\; k \in K \right\} \to 0.$$

It follows from this relation that

$$\left\{ -\frac{c_{k-1}}{\mu_0^k} G_{I_+}(x^k) \;\Big|\; k \in K \right\} \to H_{I_+}(\bar{x}), \tag{3.21}$$

and therefore

$$\frac{c_{k-1}}{\mu_0^k} \langle G_{I_+}(x^k),\, H_{I_+}(x^k)\rangle \to -\|H_{I_+}(\bar{x})\|^2 \tag{3.22}$$

for $k \in K$ as $k \to \infty$.

Now from (3.16) and (3.22) we obtain that since $I_+ \neq \emptyset$, the index set I_0 must be nonempty as well. Combining (3.22) with (3.17), we obtain that

$$\frac{c_{k-1}}{\mu_0^k} \langle G_{I_0}(x^k),\, H_{I_0}(x^k)\rangle \to 1 + \|H_{I_+}(\bar{x})\|^2 \tag{3.23}$$

which is (3.15) with $M = 1 + \|H_{I_+}(\bar{x})\|^2$. From (3.19) and (3.21) we also get the following:

$$\left\{ \frac{c_{k-1}}{(\mu_0^k)^2} (\lambda_H^k)_{I_+} \;\Big|\; k \in K \right\} \to H_{I_+}(\bar{x}). \tag{3.24}$$

Next, we prove that there exist an infinite index set $K_0 \subset K$ and $\bar{\lambda}_G, \bar{\lambda}_H \in \mathbb{R}^s$ such that

$$\{\lambda_G^k \mid k \in K_0\} \to \bar{\lambda}_G, \quad \{\lambda_H^k \mid k \in K_0\} \to \bar{\lambda}_H, \tag{3.25}$$

$(\bar{\lambda}_H)_{I_+} = 0$, and the condition (3.14) is fulfilled. Note that this implies, in particular, weak stationarity of $\bar{x}$.

From (3.12) we have

$$(\lambda_G^k)_{I_0} = \max\{0,\, (\bar{\mu}_G^{k-1})_{I_0} - c_{k-1} G_{I_0}(x^k)\} - \mu_0^k H_{I_0}(x^k),$$
$$(\lambda_H^k)_{I_0} = \max\{0,\, (\bar{\mu}_H^{k-1})_{I_0} - c_{k-1} H_{I_0}(x^k)\} - \mu_0^k G_{I_0}(x^k).$$

Suppose that for some $i \in I_0$ there exists an infinite index set $K_i \subset K$ such that the condition

$$(\bar{\mu}_G^{k-1})_i - c_{k-1} G_i(x^k) \geq 0 \tag{3.26}$$

is satisfied for all $k \in K_i$. In this case it holds that

$$(\lambda_G^k)_i = (\bar{\mu}_G^{k-1})_i - c_{k-1} G_i(x^k) - \mu_0^k H_i(x^k),$$

which implies

$$\frac{c_{k-1}}{\mu_0^k} G_i(x^k) = \frac{(\bar{\mu}_G^{k-1})_i}{\mu_0^k} - \frac{(\lambda_G^k)_i}{\mu_0^k} - H_i(x^k) \to 0,$$

and therefore

$$\frac{c_{k-1}}{\mu_0^k} G_i(x^k) H_i(x^k) \to 0 \tag{3.27}$$

for $k \in K_i$. It is evident that (3.27) is also true if the condition

$$(\bar{\mu}_H^{k-1})_i - c_{k-1} H_i(x^k) \geq 0 \tag{3.28}$$

holds for some infinite index set $K_i \subset K$.

Now we construct the index set J_0 in the following way. First we set $J_0 = \emptyset$ and $K_1 = K$. Then if for some $i \in I_0 \setminus J_0$ there exists an infinite index set $K_2 \subset K_1$ for which (3.26) or (3.28) are fulfilled for all $k \in K_1$ then we add i to J_0 and repeat the above process with the index set K_2 instead of K_1. We repeat this process until we get J_0 and K_q such that for all $i \in I_0 \setminus J_0$ the inequalities (3.26), (3.28) hold only for a finite number of points x^k, $k \in K_q$.

Let J_+ stand for the set $I_0 \setminus J_0$. It is evident that the condition (3.27) is fulfilled for all $i \in J_0$ for $k \in K_q$. Combining the latter with (3.15), we obtain that $J_+ \neq \emptyset$ and the condition

$$\frac{c_{k-1}}{\mu_0^k} \langle G_{J_+}(x^k),\, H_{J_+}(x^k)\rangle \to M \tag{3.29}$$

holds for $k \in K_q$ as $k \to \infty$. Since for $i \in J_+$ the inequalities (3.26), (3.28) are satisfied only for a finite number of elements of the sequence $\{x^k \mid k \in K_q\}$, it follows from (3.4) and (3.12) that

$$(\lambda_G^k)_{J_+} = -\mu_0^k H_{J_+}(x^k), \quad (\lambda_H^k)_{J_+} = -\mu_0^k G_{J_+}(x^k)$$

for all $k \in K_q$ large enough. The latter condition and (3.29) imply that

$$\frac{c_{k-1}}{(\mu_0^k)^3} \langle (\lambda_G^k)_{J_+},\, (\lambda_H^k)_{J_+}\rangle \to M \tag{3.30}$$

for $k \in K_q$ as $k \to \infty$.

We now show that either $I_+ = \emptyset$ or

$$\{(\lambda_H^k)_{I_+} \mid k \in K_q\} \to 0. \tag{3.31}$$

Indeed, suppose that $I_+ \neq \emptyset$ and for some $\varepsilon > 0$ and some infinite set $K_q^0 \subset K_q$ the condition

$$\|(\lambda_H^k)_{I_+}\| \geq \varepsilon \tag{3.32}$$

is fulfilled for all $k \in K_q^0$. Then, recalling (3.24) which implies

$$\frac{c_{k-1}^2}{(\mu_0^k)^4} \|(\lambda_H^k)_{I_+}\|^2 \to \|H_{I_+}(\bar{x})\|^2,$$

for $k \in K$, and combining the latter with (3.30), we get the condition

$$\frac{\mu_0^k}{c_{k-1}} \frac{\langle (\lambda_G^k)_{J_+},\, (\lambda_H^k)_{J_+}\rangle}{\|(\lambda_H^k)_{I_+}\|^2} \to \frac{M}{\|H_{I_+}(\bar{x})\|^2} > 0$$

for $k \in K_q$ as $k \to \infty$. Therefore it follows from (3.18) that

$$\frac{\langle (\lambda_G^k)_{J_+},\, (\lambda_H^k)_{J_+} \rangle}{\|(\lambda_H^k)_{I_+}\|^2} \to +\infty$$

for $k \in K_q^0$ as $k \to \infty$. Then the chain of inequalities

$$\langle (\lambda_G^k)_{J_+},\, (\lambda_H^k)_{J_+} \rangle \le \|(\lambda_G^k)_{J_+}\|\, \|(\lambda_H^k)_{J_+}\| \le \|(\lambda_G^k, (\lambda_H^k)_{I_0})\|^2$$

implies

$$\frac{\|(\lambda_G^k, (\lambda_H^k)_{I_0})\|}{\|(\lambda_H^k)_{I_+}\|} \to \infty$$

for $k \in K_q^0$. Combined with (3.32), it follows that $\{\|(\lambda_G^k, (\lambda_H^k)_{I_0})\| \mid k \in K_q^0\} \to \infty$. Dividing (3.11) by $\|(\lambda_G^k, (\lambda_H^k)_{I_0})\|$ and applying Lemma 3.1 for $k \in K_q^0$ with

$$\alpha_k = \frac{1}{\|(\lambda_G^k, (\lambda_H^k)_{I_0})\|}, \quad \omega^k = \frac{(\lambda_H^k)_{I_+}}{\|(\lambda_G^k, (\lambda_H^k)_{I_0})\|},$$

(evidently, $\{\alpha_k \mid k \in K_q^0\} \to 0$ and $\{\omega_k \mid k \in K_q^0\} \to 0$) we obtain that

$$\left\{ \frac{(\lambda_G^k, (\lambda_H^k)_{I_0})}{\|(\lambda_G^k, (\lambda_H^k)_{I_0})\|} \;\middle|\; k \in K_q^0 \right\} \to 0,$$

which is evidently impossible.

Therefore, (3.31) holds (it becomes trivial if $I_+ = \emptyset$). Applying Lemma 3.1 with $\alpha_k = 1$ to (3.11) for $k \in K_q$, we conclude that

$$\{\lambda_G^k \mid k \in K_q\} \to \bar\lambda_G, \quad \{(\lambda_H^k)_{I_0} \mid k \in K_q\} \to (\bar\lambda_H)_{I_0}$$

for some $\bar\lambda_G$ and $(\bar\lambda_H)_{I_0}$, and the condition (3.14) is fulfilled. In view of (3.31), to obtain (3.25) it remains to take $(\bar\lambda_H)_{I_+} = 0$ and $K_0 = K_q$.

Finally, we prove that $\bar x$ is C-stationary. Since the conditions (3.14), (3.25) hold for some $K_0 \subset K$ and some $\bar\lambda_G, \bar\lambda_H \in \mathbb{R}^s$ with $(\bar\lambda_H)_{I_+} = 0$, it remains to show that $(\bar\lambda_G)_i(\bar\lambda_H)_i \ge 0$ for all $i \in I_0$.

We construct the index sets J_0 and J_+ and the set K_q the same way as above, but taking K_0 instead of K. It is clear that the condition (3.30) is also true for new J_0, J_+ and K_q. Since the sequence $\{(\lambda_G^k, \lambda_H^k) \mid k \in K_0\}$ is bounded and hence

$$\left\{ \frac{\langle (\lambda_G^k)_{J_+},\, (\lambda_H^k)_{J_+} \rangle}{\mu_0^k} \;\middle|\; k \in K_0 \right\} \to 0,$$

the condition (3.30) implies

$$\frac{c_{k-1}}{(\mu_0^k)^2} \to +\infty \tag{3.33}$$

for $k \in K_q$ as $k \to \infty$.

Since $I_0 = J_0 \cup J_+$, we need to show that $(\bar\lambda_G)_i(\bar\lambda_H)_i \ge 0$ for $i \in J_0$ and for $i \in J_+$. It follows from the definition of J_0 that for all $i \in J_0$ either (3.26) is satisfied for all $k \in K_q$ or (3.28) is satisfied for all $k \in K_q$. It is clear that for any $i \in J_0$ there are two possible cases: either there exists an infinite set $K_q^1 \subset K_q$ such that both (3.26) and (3.28) hold for all $k \in K_q^1$ or there exists an infinite set $K_q^2 \subset K_q$ such that only one of these conditions is fulfilled for all $k \in K_q^2$, while the other is violated for all $k \in K_q^2$.

First we consider the case when for some $i \in J_0$ and an infinite set $K_q^1 \subset K_q$ both (3.26) and (3.28) hold for all $k \in K_q^1$. In this case from (3.4) and (3.12) we have

$$(\lambda_G^k)_i = (\bar{\mu}_G^{k-1})_i - c_{k-1} G_i(x^k) - \mu_0^k H_i(x^k), \tag{3.34}$$

$$(\lambda_H^k)_i = (\bar{\mu}_H^{k-1})_i - c_{k-1} H_i(x^k) - \mu_0^k G_i(x^k), \tag{3.35}$$

for $k \in K_q^1$. Evaluating $G_i(x^k)$ from the second equation and putting it into the first, we obtain

$$(\lambda_G^k)_i = (\bar{\mu}_G^{k-1})_i - \mu_0^k H_i(x^k) - \frac{c_{k-1}}{\mu_0^k}\big((\bar{\mu}_H^{k-1})_i - c_{k-1} H_i(x^k) - (\lambda_H^k)_i\big),$$

and therefore

$$H_i(x^k) = \frac{\mu_0^k((\lambda_G^k)_i - (\bar{\mu}_G^{k-1})_i) - c_{k-1}((\lambda_H^k)_i - (\bar{\mu}_H^{k-1})_i)}{c_{k-1}^2 - (\mu_0^k)^2},$$

$$c_{k-1} H_i(x^k) = \frac{c_{k-1}^2}{c_{k-1}^2 - (\mu_0^k)^2}\left(\frac{\mu_0^k}{c_{k-1}}\big((\lambda_G^k)_i - (\bar{\mu}_G^{k-1})_i\big) - \big((\lambda_H^k)_i - (\bar{\mu}_H^{k-1})_i\big)\right).$$

From (3.18) and (3.25) it follows that for $k \in K_q^1$ the first term of the product in the right-hand side of the last equation tends to one while the second term of this product is bounded. Therefore the sequence $\{c_{k-1} H_i(x^k) \mid k \in K_q^1\}$ is bounded. Repeating the same arguments for $G_i(x^k)$, we obtain boundedness of $\{c_{k-1} G_i(x^k) \mid k \in K_q^1\}$. In view of (3.18), the last two conditions imply that $\mu_0^k G_i(x^k)$ and $\mu_0^k H_i(x^k)$ tend to zero for $k \in K_q^1$. Combined with (3.26) and (3.28) and applied to (3.34), (3.35), this results in $(\bar{\lambda}_G)_i \geq 0$ and $(\bar{\lambda}_H)_i \geq 0$.

Now we consider the second case when for some $i \in J_0$ and an infinite set $K_q^2 \subset K_q$ one of the two conditions (3.26), (3.28) holds for all $k \in K_q^2$ while the other does not hold for all $k \in K_q^2$. We assume that the first condition is fulfilled. In this case it follows from (3.4) and (3.12) that for all $k \in K_q^2$, the condition (3.34) holds and, instead of (3.35), the condition

$$(\lambda_H^k)_i = -\mu_0^k G_i(x^k) \tag{3.36}$$

holds. If $(\bar{\lambda}_H)_i = 0$ then the equality $(\bar{\lambda}_G)_i(\bar{\lambda}_H)_i = 0$ holds automatically. Suppose that $(\bar{\lambda}_H)_i \neq 0$. From (3.34) and (3.36) we obtain

$$(\lambda_G^k)_i = (\bar{\mu}_G^{k-1})_i + \frac{c_{k-1}}{\mu_0^k}(\lambda_H^k)_i - \mu_0^k H_i(x^k),$$

$$\frac{c_{k-1}}{(\mu_0^k)^2}(\lambda_H^k)_i = H_i(x^k) - \frac{(\lambda_G^k)_i - (\bar{\mu}_G^{k-1})_i}{\mu_0^k},$$

and hence,

$$\frac{c_{k-1}}{(\mu_0^k)^2}(\lambda_H^k)_i \to 0$$

as $k \to \infty$. Since $(\bar{\lambda}_H)_i \neq 0$, the last condition contradicts (3.33).

It remains to show that $(\bar{\lambda}_G)_i(\bar{\lambda}_H)_i \geq 0$ for all $i \in J_+$. The definition of J_+ implies the existence of an infinite index set $K_q^3 \subset K_q$ such that the conditions (3.26), (3.28) are violated for all $k \in K_q^3$. Therefore from (3.4) and (3.12) we obtain that for all $i \in J_+$ and $k \in K_q^3$

$$(\lambda_G^k)_i = -\mu_0^k H_i(x^k), \quad (\lambda_H^k)_i = -\mu_0^k G_i(x^k). \tag{3.37}$$

If $(\bar{\lambda}_G)_i = 0$ or $(\bar{\lambda}_H)_i = 0$ for some i then the condition $(\bar{\lambda}_G)_i(\bar{\lambda}_H)_i \geq 0$ is satisfied. Suppose that $(\bar{\lambda}_G)_i \neq 0$ and $(\bar{\lambda}_H)_i \neq 0$. In this case from (3.18) and (3.37) we obtain that $|c_{k-1}G_i(x^k)|$ and $|c_{k-1}H_i(x^k)|$ tend to infinity for $k \in K_q^3$. Then the equalities (3.4) and violation of (3.26) and (3.28) imply $G_i(x^k) > 0$ and $H_i(x^k) > 0$ for all k large enough. Applied to (3.37), these conditions result in $(\bar{\lambda}_G)_i < 0$ and $(\bar{\lambda}_H)_i < 0$ (since these are nonzero). Therefore $(\bar{\lambda}_G)_i(\bar{\lambda}_H)_i > 0$. ☐

The next two examples demonstrate that under the assumptions of Theorem 3.2 accumulation points of the iterates generated by Algorithm 2.1 need not be strongly stationary or even M-stationary. In the first example all the indices belong to I_0, while in the second both I_0 and its complement are nonempty.

EXAMPLE 3.1. This is problem `scholtes3` from MacMPEC [43]; similar effects are also observed for `scale4` and `scale5` from the same test collection. Let $n = 2$, $s = 1$, $f(x) = ((x_1 - 1)^2 + (x_2 - 1)^2/2$, $G(x) = x_1$, $H(x) = x_2$. Problem (1.2) with this data has two solutions (1, 0) and (0, 1), both strongly stationary, and also one nonoptimal C-stationary (but not M-stationary) point $\bar{x} = (0, 0)$ satisfying MPCC-LICQ.

When started from default initial points ($x^0 \in \mathbb{R}^2$ close to 0 and $\mu^0 = (0, 0, 0)$), ALGENCAN solver (with disabled acceleration step and tools for solving subproblems "to second-order optimality", i.e., the pure Algorithm 2.1 above) converges to the specified non-strongly stationary $\bar{x}$, with $\mu_G^k = 0$ and $\mu_H^k = 0$ for all k, and with $c_k \to +\infty$. What happens is that the unconstrained minimization method used to solve subproblems of the form (2.4) in Algorithm 2.1 picks up the saddle point $x^{k+1} = (t_k, t_k)$ of $L_{c_k}(\cdot, \bar{\mu}^k)$, where $t_k \approx 1/c_k^{1/3}$ is defined by the equation

$$t - 1 + (\bar{\mu}_0^k + c_k t^2)t = 0.$$

(Of course, this may happen only for very special starting points used by the the unconstrained minimization method, in this case points on the straight line given by $x_1 = x_2$.) Note that according to (3.5), $\mu_0^{k+1} = \bar{\mu}_0^k + c_k t_k^2 \approx \bar{\mu}_0^k + c_k^{1/3} \to +\infty$.

Observe again that ALGENCAN is treating nonnegativity constraints as bounds, not including them in the augmented Lagrangian. However, since $\mu_G^k = 0$ and $\mu_H^k = 0$, this partial augmented Lagrangian coincides with the full augmented Lagrangian on $\mathbb{R}_+^2$, and therefore, the above gives an adequate understanding of ALGENCAN behavior: the algorithm involving the full augmented Lagrangian behaves similarly on this problem.

Similar observations as in the example above were reported in [3, Example 2] for the problem with the same constraints, but with $f(x) = -x_1 - x_2$. For this problem ALGENCAN converges to the C-stationary point $\bar{x} = 0$ which is in fact a global maximizer rather than minimizer. However, it is important to emphasize that small perturbations of the starting point give convergence to strong stationarity, which means that the undesirable phenomenon is actually *not* typical.

EXAMPLE 3.2. Consider the problem (1.2) with the following data: $n = 4$, $s = 2$, $f(x) = -\sqrt{2}(x_1 + x_3) - x_2 + (x_4 - 1)^2/2$, $G(x) = (x_1, x_2)$, $H(x) = (x_3, x_4)$. Consider the feasible point $\bar{x} = (0, 0, 0, 1)$ which is nonoptimal and satisfies MPCC-LICQ. It is easy to see that this point is C-stationary but not M-stationary.

For this problem, equality in (3.6) for each k is equivalent to the following system:

$$-\sqrt{2} - (\mu_G^k)_1 + \mu_0^k x_3^k = 0, \quad -1 - (\mu_G^k)_2 + \mu_0^k x_4^k = 0,$$
$$-\sqrt{2} - (\mu_H^k)_1 + \mu_0^k x_1^k = 0, \quad x_4^k - 1 - (\mu_H^k)_2 + \mu_0^k x_2^k = 0,$$

where μ_G^k, μ_H^k and μ_0^k are defined according to (3.4), (3.5). Suppose that for all $k = 1, 2, \ldots$ we take $\bar{\mu}_G^{k-1} = \bar{\mu}_H^{k-1} = 0$ and $\bar{\mu}_0^{k-1} = 1$. Then $x_1^k = x_3^k = \sqrt{2}/t_k$, $x_2^k = -1/t_k^2$, $x_4^k = 1 + 1/t_k$, and $\mu_0^k = t_k$, $\mu_G^k = (0, t_k)$, $\mu_H^k = 0$, where $t_k = c_{k-1}^{1/3}$, satisfy all the needed relations. It is clear that if $c_k \to +\infty$ then the sequence $\{x^k\}$ converges to $\bar{x}$.

REMARK 3.1. It can be shown that the first part of the assertion of Theorem 3.2 (that is, strong stationarity of the point $\bar{x}$ in the case when the sequence $\{\mu_0^k\}$ has a bounded subsequence such that the corresponding subsequence of $\{x^k\}$ converges to $\bar{x}$) remains valid if MPCC-LICQ at $\bar{x}$ is replaced by MFCQ for the so-called relaxed nonlinear programming problem (RNLP) associated to MPCC (1.2):

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G_{I_G(\bar{x}) \setminus I_H(\bar{x})}(x) = 0, \ H_{I_H(\bar{x}) \setminus I_G(\bar{x})}(x) = 0, \\ & G_{I_0(\bar{x})}(x) \geq 0, \ H_{I_0(\bar{x})}(x) \geq 0. \end{array}$$

MFCQ for RNLP is evidently weaker than MPCC-LICQ.

However, MPCC-LICQ is essential for the second part of the assertion of Theorem 3.2. The following example demonstrates that if $\{\mu_0^k\}$ tends to infinity then MFCQ for RNLP may not imply even weak stationarity of the accumulation point.

EXAMPLE 3.3. Consider the problem (1.2) with the following data: $n = 6$, $s = 3$, $f(x) = -3x_1/2 + x_3(-1 - x_4^2 + x_5^2) - x_6$, $G(x) = (|x_1|^{3/2} + x_2, -|x_1|^{3/2} + x_2, x_3)$, $H(x) = (x_4, x_5, x_6)$. It can be easily seen that the feasible point $\bar{x} = 0$ satisfies MFCQ for RNLP, though it is nonoptimal and not weakly stationary.

Assuming that $x_1 \geq 0$, the first equality in (3.6) for each k is equivalent to the following system:

$$-1 - x_1^{1/2}((\mu_G^k)_1 - (\mu_G^k)_2 - \mu_0^k(x_4 - x_5)) = 0, \quad -(\mu_G^k)_1 - (\mu_G^k)_2 + \mu_0^k(x_4 + x_5) = 0,$$
$$-1 - x_4^2 + x_5^2 - (\mu_G^k)_3 + \mu_0^k x_6 = 0, \quad -2x_3x_4 - (\mu_H^k)_1 + \mu_0^k(x_1^{3/2} + x_2) = 0,$$
$$2x_3x_5 - (\mu_H^k)_2 + \mu_0^k(-x_1^{3/2} + x_2) = 0, \quad -1 - (\mu_H^k)_3 + \mu_0^k x_3 = 0,$$

where μ_G^k, μ_H^k and μ_0^k are defined according to (3.4), (3.5).

Suppose that for all $k = 1, 2, \ldots$ we take $\bar{\mu}_G^{k-1} = \bar{\mu}_H^{k-1} = 0$ and $\bar{\mu}_0^{k-1} = 0$. Then $x^k = (t_k^6, 0, t_k^4, t_k/2, t_k/2, t_k^4)$ and $\mu_0^k = 1/t_k^4$, $\mu_G^k = (0, 1/t_k^3, 0)$, $\mu_H^k = 0$, where $t_k = 1/c_{k-1}^{1/12}$, satisfy all the needed relations. It is clear that if $c_k \to +\infty$ then the sequence $\{x^k\}$ converges to $\bar{x}$.

We complete this section by mentioning the example provided in [24, Section 7.2], showing that SQP applied to MPCC may converge to an *arbitrary* feasible point, not even weakly stationary, even if it satisfies MPCC-LICQ. As demonstrated above, such arbitrary accumulation points are ruled out for Algorithm 2.1.

4. Numerical results. In this section, we report on the performance of the ALGENCAN [1] implementation of the augmented Lagrangian method, compared with two well-established implementations of SQP, namely, SNOPT [28] and filterSQP

[23]; with the linearly constrained Lagrangian method implemented in MINOS [44]; and also with two interior-point solvers IPOPT [32, 52] and KNITRO [37, 14]. For MPCCs, we used IPOPT-C [48] which is a modification of IPOPT making use of the special structure of complementarity constraints. KNITRO also has some special features for treating MPCCs [40]. We invoke SQP methods because it is known [22] that they are quite robust and effective when applied to MPCC. The use of the linearly constrained (augmented) Lagrangian algorithm implemented in MINOS is motivated by the fact that (in a certain sense) it is related to both, augmented Lagrangian methods and SQP; see [33, 35].

We used ALGENCAN 2.3.7 with AMPL interface, compiled with the use of MA57 library [51]. The latter provides effective software for solving sparse symmetric linear systems, making ALGENCAN much faster on large-scale problems. The newer versions of ALGENCAN have an ultimate acceleration option [13], which consists of identifying the active constraints and switching to a Newton-type method for the resulting system of equations. In our experiments we first tried skipping the ALGENCAN's acceleration steps. The reason for this is that our experiments are concerned with degenerate problems, and standard Newton-like steps could be harmful in degenerate cases, at least potentially. However, for our type of problems we did not observe any serious advantage in skipping the acceleration steps, while in some cases they were still helpful. Below we report results with the acceleration steps being active.

Regarding the other solvers, we used SNOPT 7.2-8 and MINOS 5.51 coming with AMPL, and IPOPT 3.8.0. Solvers KNITRO 8.0.0 and filterSQP 20020316 were run on the NEOS server [45]. For MPCCs we used the latest available version 2.2.1e of IPOPT-C. For all the solvers, including ALGENCAN, we used the default values of all the parameters with one exception: we increased the maximum number of major iterations for MINOS from 50 to 1000 because this significantly improved its robustness when applied to degenerate problems.

As discussed above, our main concern in this paper is robustness and quality of the outcome (the value of the objective function when terminated at a feasible point). This is one reason why we do not systematically study the relative efficiency of the solvers, in particular the CPU times. The other (obvious) reason is that some solvers were run on the NEOS server, and so comparing CPU times is simply not meaningful. Nevertheless, for some additional information we compare the solvers by the evaluations of the objective function and of the constraints, as this partial indicator of efficiency is available for all the solvers (except constraints evaluations of KNITRO, which are not reported by the solver). Concerning function evaluations, one subtle point is that in MINOS and SNOPT linear functions do not affect the evaluation counts reported by the solvers, while all the other solvers report on the numbers of all evaluations, linear and nonlinear. To make the comparison more-or-less consistent, in those cases when the objective function or all constraints are linear, we simply put the corresponding number of evaluations equal to 1 for all solvers. Note that MacMPEC problems always have at least one nonlinear constraint (corresponding to the last constraint in (1.2)); therefore, there is no need for such manipulations of constraints evaluations counts for this collection.

Finally, we examine some properties of the solvers from the viewpoint of the quality of the outcome. To this end, apart from failures (that is, the cases when the solver terminates with any exit flag other than "Optimal solution found" or equivalent), we also report on the cases of convergence to nonoptimal objective values, and we provide the analysis of boundedness of dual sequences generated by ALGENCAN. Recall that according to Theorem 3.2, boundedness of the dual sequence of ALGENCAN implies strong stationarity of primal accumulation points.

4.1. Numerical results for MacMPEC. This section contains the results of experiments on MacMPEC [43], which is an established AMPL collection of 180 MPCCs. We used 161 of these problems (13 were excluded because they involve mixed complementarity constraints, 1 because it has a binary variable, and 5 because they are infeasible; according to [43]).

It should be mentioned that MacMPEC problems are written using the AMPL operator `complements` for complementarity conditions. This required special treatment in our experiments. First, KNITRO and IPOPT-C are the only solvers in our set whose AMPL interfaces understand this operator. Interfaces of all the other solvers simply ignore the constraints involving `complements`; for these solvers complementarity constraints had to be stated as the usual equality and inequality constraints. Second, even KNITRO cannot deal with many of MacMPEC models directly, since it requires complementarity constraints to be stated via slack variables. We therefore had to introduce slacks in the MacMPEC problems formulations submitted to KNITRO and IPOPT-C.

For all the solvers except KNITRO and IPOPT-C (that use `complements`) we tried four (equivalent) formulations, with the last constraint in (1.2) given by either of

$$\langle G(x),\, H(x)\rangle \leq 0; \tag{4.1}$$

$$\langle G(x),\, H(x)\rangle = 0; \tag{4.2}$$

$$G_i(x)H_i(x) \leq 0, \quad i = 1,\, \dots,\, s; \tag{4.3}$$

$$G_i(x)H_i(x) = 0, \quad i = 1,\, \dots,\, s. \tag{4.4}$$

Moreover, each of the forms (4.1)–(4.4) has a counterpart employing slack variables, and this leads to eight different formulations of complementarity constraints. According to our numerical experience, for ALGENCAN the form (4.1) is slightly preferable, though the differences are not significant. But for both MINOS and SNOPT, the forms (4.1) and (4.2) turn out to be seriously preferable. Furthermore, introducing slacks slightly improves the performance of all the solvers, which agrees with previous numerical experience in [22, 24]. For this reason, for ALGENCAN, SNOPT, filterSQP and MINOS we report the results for the formulation of MPCC corresponding to (4.1) (i.e., that in (1.2)), re-stated using slack variables when appropriate. Observe that filterSQP applied to this reformulation is supposed to be equivalent to the solver filtermpec [22] applied to the original MacMPEC models.

For each of MacMPEC problems, we performed a single run from the default starting point specified in the collection. The diagram in Figure 4.1 reports on the numbers of failures and of the cases of successful convergence but to nonoptimal values. The objective function value at termination is regarded nonoptimal if its difference with the best known value exceeds 1e-1. The optimal values were taken from [43], except for some cases where better feasible solutions were found in the course of our experiments. The diagram shows that ALGENCAN has fewer cases of convergence to nonoptimal points than any of its competitors, while only IPOPT-C has (slightly) fewer failures than ALGENCAN. Note, however, that IPOPT-C does not have any theoretical global convergence guarantees to support it [48].

Other comparisons are presented in the form of performance profiles [18]. For each solver the value of the plotted function at $\tau \in [1,\, \infty)$ is the portion of test

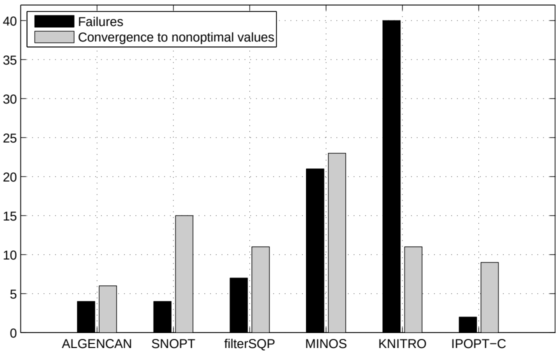
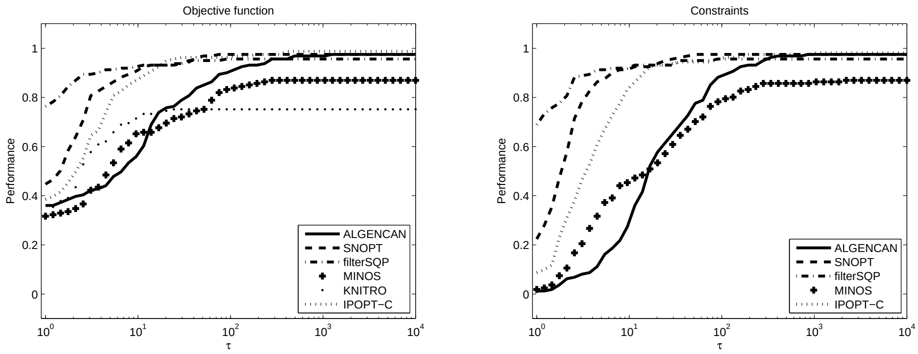


Fig. 4.1. *Failures and cases of convergence to nonoptimal values on MacMPEC.*



(a) Objective function

(b) Constraints

Fig. 4.2. *Nonlinear evaluations on MacMPEC.*

problems for which the result of this solver was no more than τ times worse than the best result over all solvers. Failure is regarded as infinitely many times worse than any other result.

Figures 4.2(a) and 4.2(b) report on evaluations of the objective function and of the constraints, respectively. According to these figures, ALGENCAN, SNOPT, filterSQP and IPOPT-C demonstrated similarly high robustness (about 97% of problems solved), while MINOS and KNITRO are less robust on MacMPEC (about 85% and 75% of problems solved, respectively).

It should be mentioned that in terms of major iterations ALGENCAN outperforms all the other solvers: it has the best result for about 60% of problems while filterSQP wins in less than 40% of the cases, and for the other solvers this percentage is less than 15%. Figure 4.2 shows that in terms of objective function and constraints evaluations, ALGENCAN is comparable with MINOS, but both are outperformed by all the other solvers.

Finally, Table 4.1 summarizes all MacMPEC problems on which some kind of "anomalies" were observed for ALGENCAN: failures (columns "F"), cases of convergence to nonoptimal values (columns "NO"), and cases of unbounded dual trajectory ("UB"). The dual trajectory is regarded unbounded if the infinity norm of the dual

TABLE 4.1
MacMPEC problems with "anomalies" for ALGENCAN

Problem	F	NO	UB	Problem	F	NO	UB
bilevel1		•		qpec-200-4		•	
bilevel3		•		qpec2			•
bilin		•		ralphmod	•		•
design-cent-2	•		•	scale4			•
design-cent-4		•		scholtes4			•
ex9.2.2			•	siouxfls1	•		•
hakonsen	•		•	water-FL		•	•
pack-comp2p-8			•	water-net			•

iterate at termination is greater than 1e+4. As can be seen from Table 4.1, in some of the cases unboundedness actually was not a problem, as convergence to optimal values still occurred, and some cases of unboundedness are actually also failures of some other nature. More importantly, it can be concluded that boundedness of dual sequences is clearly a typical scenario for ALGENCAN even when the multiplier set is unbounded, and thus strong stationarity for MPCC is guaranteed (assuming MPCC-LICQ). And even in cases when the dual sequence is unbounded, it is not necessarily a problem for convergence to optimal values.

Overall, the conclusion is that if the main requirement for a given application is robustness and guarantees of convergence to good solutions, then ALGENCAN is a good choice. If speed is important, then some acceleration is needed, which would be a subject of our future research.

4.2. Numerical results for DEGEN. In this section we present the results of the experiments on DEGEN [17], which is an AMPL collection of 109 test problems with degenerate constraints. Most problems are very small, but many are in some ways difficult. We used all DEGEN problems except for the following 9 instances: problems 20205, 20223, 30204 and 40207 are unbounded (though they have degenerate stationary points); problems 2DD01-500h and 2DD01-500v are too large with respect to the other problems in the collection; problems 20201, 40210 and 40211 actually have only bound constraints, and this is identified by ALGENCAN which does not perform any major iterations at all on these problems.

DEGEN AMPL models include the mechanism for choosing random primal and dual starting points in the domain of a specified size. We used the default size equal to 100, and for each test problem we performed 100 runs from random starting points.

The results below are presented in the form of performance profiles which is a slightly modified version of the original proposal in [18]. For each solver the plotted function $\pi\colon [1, \infty) \to [0, 1]$ is defined as follows. For a given characteristic (e.g., the iteration count), let k_p stand for the average result of a given solver per one successful run for problem p. Let s_p denote the portion of successful runs on this problem. Let r_p be equal to the best (minimum) value of k_p over all solvers. Then

$$\pi(\tau) = \frac{1}{P} \sum_{p \in R(\tau)} s_p,$$

where P is the number of problems in the test set (100 in our case) and $R(\tau)$ is the

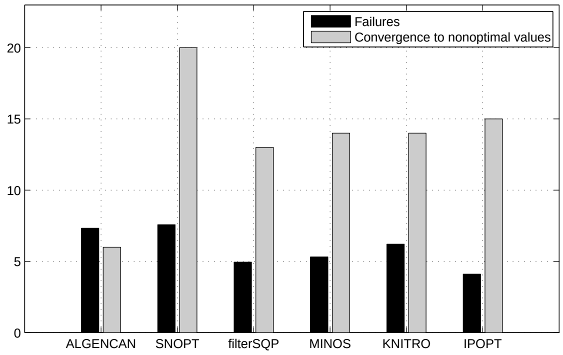
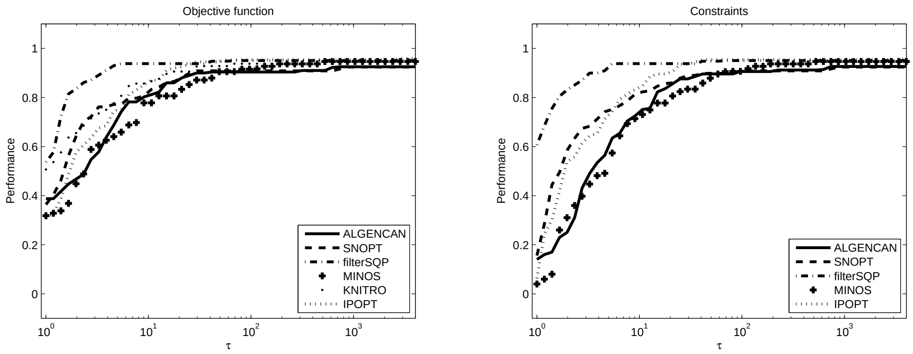


Fig. 4.3. *Failures and cases of convergence to nonoptimal values on DEGEN.*



(a) Objective function (b) Constraints

Fig. 4.4. *Nonlinear evaluations on DEGEN.*

set of problems for which k_p is no more than τ times worse (larger) than the best result r_p:

$$R(\tau) = \{p = 1, \ldots, P \mid k_p \leq \tau r_p\}, \quad \tau \in [1, \infty).$$

In particular, the value $\pi(1)$ corresponds to the portion of runs for which the given solver demonstrated the best result. The values of $\pi(\tau)$ for large τ characterize robustness, that is, the portion of successful runs.

In order to single out the cases of convergence to nonoptimal values, we used the threshold 1e-2 for the distance to the optimal values reported within DEGEN models. The diagram in Figure 4.3 reports on the numbers of failures and of convergence to nonoptimal values. For ALGENCAN, the percentage of the cases of convergence to optimal value over successful runs is about 95%, while for all the other solvers it is no greater than 90%. Moreover, the percentage of convergence to optimal value over all runs (including those which ended in failures) is also the highest for ALGENCAN.

In almost all cases of failure, the output flag of ALGENCAN was "The penalty parameter is too large". This usually happens in the cases of convergence to an infeasible point, and rarely because of slow convergence.

Comparisons by evaluations of the objective function and the constraints are presented in Figures 4.4(a) and 4.4(b), respectively. According to these figures, all the solvers demonstrate similar robustness on DEGEN collection, with about 94% of successful runs.

In terms of major iterations, ALGENCAN is again significantly better than all the other solvers: it has the best result for almost all the problems. Moreover, the result of all the other solvers is more than 4 times worse for about 50% of problems.

Regarding objective function and constraints evaluation counts (see Figure 4.4), the picture is similar to that for MacMPEC: ALGENCAN is again somewhat more effective than MINOS and less effective than the other solvers. However, the difference is less significant than on MacMPEC.

Finally, the cases of unbounded dual sequence were detected for 9.1% of runs. Moreover, there were only 12 problems for which these sequences were unbounded for at least 20% of runs. Therefore, we can conclude again that despite the fact that most of DEGEN problems have unbounded multiplier sets, dual trajectories of ALGENCAN usually remain bounded.

Overall, the conclusions are similar to those for MacMPEC test problems. ALGENCAN is a good choice when computing good solutions (rather than speed) is the primary concern.

Acknowledgments. We thank Ernesto Birgin for his prompt and patient assistance with compiling and tuning the ALGENCAN solver.

REFERENCES

[1] ALGENCAN. `http://www.ime.usp.br/~egbirgin/tango/`.
[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 (2007), 1286–1309.
[3] 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 (2010), 209–236.
[4] R. Andreani, G. Haeser, M.L. Schuverdt, and P.J.S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. (2011). DOI: 10.1007/s10107-011-0456-0.
[5] R. Andreani and J.M. Martinez. On the solution of mathematical programming problems with equilibrium constraints. Math. Meth. Oper. Res. 54 (2001), 345–358.
[6] R. Andreani, J.M. Martínez, and M.L. Schuverdt. On second-order optimality conditions for nonlinear programming. Optimization. 56 (2007), 529–542.
[7] M. Anitescu. Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints. SIAM J. Optim. 16 (2005), 120–145.
[8] M. Anitescu, P. Tseng, and S.J. Wright. Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties. Math. Program. 110 (2007), 337–371.
[9] J.F. Bard. Convex two-level optimization. Math. Program. 40 (1988), 15–27.
[10] M.C. Bartholomew-Biggs. Recursive quadratic programming methods based on the augmented Lagrangian. Math. Program. Study 31 (1987), 21–41.
[11] D.P. Bertsekas. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, 1982.
[12] E.G. Birgin, D. Fernández, and J.M. Martínez. The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. Optim. Meth. Software. 27 (2012), 1001–1024.
[13] E.G. Birgin and J.M. Martínez. Improving ultimate convergence of an augmented Lagrangian method. Optim. Meth. Software 23 (2008), 177–195.
[14] R.H. Byrd, J. Nocedal, and R.A. Waltz. KNITRO: an integrated package for nonlinear optimization. In Large-Scale Nonlinear Optimization, G. di Pillo and M. Roma, eds., Springer-Verlag, New York, NY, 2006, 35–59.
[15] L. Chen and D. Goldfarb. An active set method for mathematical programs

with linear complementarity constraints. Manuscript (2007). Available at http://www.columbia.edu/~goldfarb/goldfarbs_site/goldfarb_papers.html.
[16] A.R. Conn, N. Gould, A. Sartenaer, and P.L. Toint. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. SIAM J. Optim. 6 (1996), 674–703.
[17] DEGEN. http://w3.impa.br/~optim/DEGEN_collection.zip.
[18] E.D. Dolan and J.J. Morè. Benchmarking optimization software with performance profiles. Math. Program. 91 (2002), 201–213.
[19] D. Fernández and M.V. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. 22 (2012), 384–407.
[20] D. Fernández and M.V. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. 125 (2010), 47–73.
[21] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. 94 (2002), 91–124.
[22] R. Fletcher and S. Leyffer. Solving mathematical programs with equilibrium constraints as nonlinear programs. Optim. Methods Softw. 19 (2004), 15–40.
[23] R. Fletcher and S. Leyffer. User manual for filterSQP. University of Dundee Numerical Analysis Report NA 181, April 1998.
[24] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim. 17 (2006), 259–286.
[25] M. Fukushima and P. Tseng. An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints. SIAM J. Optim. 12 (2002), 724–739.
[26] M. Fukushima and P. Tseng. An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints: Erratum. SIAM J. Optim. 12 (2007), 1253–1257.
[27] G. Giallombardo and D. Ralph. Multiplier convergence in trust-region methods with application to convergence of decomposition methods for MPECs Math. Program. 112 (2008), 335–369.
[28] P.E. Gill, W. Murray, and M.A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM J. Optim. 12 (2002), 979–1006.
[29] W.W. Hager. Stabilized sequential quadratic programming. Comp. Optim. Appl. 12 (1999), 253–273.
[30] M.R. Hestenes. Multiplier and gradient methods. J. Optim. Theory Appl. 4 (1969), 303–320.
[31] X.X. Huang, X.Q. Yang, and K.L. Teo. Partial augmented Lagrangian method and mathematical programs with complementarity constraints. J. Global Optim. 35 (2006), 235–254.
[32] IPOPT. https://projects.coin-or.org/Ipopt.
[33] A.F. Izmailov and M.V. Solodov. A truncated SQP method based on inexact interior-point solutions of subproblems. SIAM J. Optim. 20 (2010), 2584–2613.
[34] A.F. Izmailov and M.V. Solodov. Newton-type methods for optimization problems without constraint qualifications. SIAM J. Optim. 15 (2004), 210–228.
[35] A.F. Izmailov and M.V. Solodov. On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. Math. Program. 126 (2011), 231–257.
[36] A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. Math. Program. 133 (2012), 93–120.
[37] KNITRO. http://www.ziena.com/knitro.htm.
[38] LANCELOT. http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml.
[39] S. Leyffer and T.S. Munson. A globally convergent filter method for MPECs. Preprint ANL/MCS-P1457-0907. Mathematics and Computer Science Division, Argonne National Laboratories, Argonne, 2007.
[40] S. Leyffer, G. López-Calva, and J. Nocedal. Interior methods for mathematical programs with complementarity constraints. SIAM J. Optim. 17 (2006), 52–77.
[41] G.S. Liu and J.J. Ye. Merit-function piecewise SQP algorithm for mathematical programs with equilibrium constraints. J. Optim. Theory Appl. 135 (2007), 623–641.
[42] Z.-Q. Luo, J.-S. Pang, and D. Ralph. Mathematical Programs with Equilibrium Constraints. Cambridge Univ. Press, Cambridge, 1996.
[43] MacMPEC. http://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC.
[44] B.A. Murtagh and M.A. Saunders. MINOS 5.0 user's guide. Technical Report SOL 83.20,

Stanford University, December 1983.
[45] NEOS server. http://neos-server.org/neos/.
[46] J.V. Outrata, M. Kocvara, and J. Zowe. Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results. Kluwer Academic Publishers, Boston, 1998.
[47] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In Optimization, R. Fletcher, ed., Academic Press, New York, NY, 1969, 283–298.
[48] A.U. Ranghunathan and L.T. Biegler. An interior point method for mathematical programs with complementarity constraints (MPCCs). SIAM J. Optim. 15 (2005), 720–750.
[49] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM J. Optim. 11 (2001), 918–936.
[50] S. Scholtes and M. Stöhr. Exact penalization of mathematical programs with equilibrium constraints. SIAM J. Control Optim. 37 (1999), 617–652.
[51] The HSL Mathematical Software Library. http://www.hsl.rl.ac.uk.
[52] A. Wächter and L.T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Math. Program. 106 (2006), 25–57.
[53] S.J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim. 15 (2005), 673–696.
[54] S.J. Wright. Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math. Program. 95 (2003), 137–160.
[55] S.J. Wright. Modifying SQP for degenerate problems. SIAM J. Optim. 13 (2002), 470–497.
[56] S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. and Appl. 11 (1998), 253–275.
[57] X.Q. Yang and X.X. Huang. Convergence analysis of an augmented Lagrangian method for mathematical programs with complementarity constraints. Nonlinear Analysis 63 (2005), 2247–2256.