Augmented Lagrangian methods with telescopic subproblems for MPCCs and cardinality constraints with general constraints: constraint-preserving swaps and B-stationarity*

E. G. Birgin† J. M. Martínez‡

November 3, 2026

Abstract

We consider optimization problems with partitioned cardinality constraints, which include complementarity constraints, and general equality and inequality constraints. In the augmented Lagrangian method that we consider, the bounds and the cardinality constraints are kept in the subproblems, which are solved by the Telescopic Method, introduced in a previous work for problems with bounds and partitioned cardinality constraints only. The limit points of this augmented Lagrangian method are B-stationary, without constraint qualifications, if the penalty parameters remain bounded, but not in general, as we show by means of a counterexample. If the penalty parameters are unbounded and the stopping criterion of the subproblems includes a derivative test outside the current face, the feasible limit points at which a linear independence constraint qualification holds are M-stationary for MPCCs. The reason is that the swaps of the Telescopic Method pass through points at which the constraints have different values, and the corresponding variation of the penalty term grows with the penalty parameter. We modify the swaps, by means of a scaled Newton restoration, so that they preserve the values of the constraints. We prove that, with this modification, every feasible limit point at which a linear independence constraint qualification holds is B-stationary (S-stationary for MPCCs), whatever the behavior of the penalty parameters and of the safeguarded multipliers. We show, by means of counterexamples, that the modification is necessary, even under the linear independence constraint qualification and with exact solutions of the subproblems, and that the constraint qualification cannot be weakened to one of Mangasarian–Fromovitz type. In the MacMPEC collection, the method found certified S- or B-stationary points in 185 of the 193 instances, more than Algencan and the regularization method of Scholtes. In experiments with cardinality-constrained portfolios and sparse principal component analysis, the constraint-preserving swaps improved the original swaps and the methods applied to complementarity reformulations.

Keywords: augmented Lagrangian methods, partitioned cardinality constraints, complementarity constraints, B-stationarity, constraint-preserving swaps, Telescopic Method.

AMS subject classifications: 90C30, 90C33, 65K05, 49M37.

*This work was supported by FAPESP (grants 2022/05803-3 and 2023/08706-1) and CNPq (grants 302073/2022-1 and 305058/2024-0).

†Department of Computer Science, Institute of Mathematics, Statistics, and Computer Science (IME), University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. e-mail: egbirgin@ime.usp.br

‡Department of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing (IMECC), State University of Campinas, 13083-859 Campinas SP, Brazil. e-mail: martinez@ime.unicamp.br

1 Introduction

In [8] we introduced the Telescopic Method for the problem

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega, \tag{1}$$

where Ω is defined by bounds and partitioned cardinality constraints (PCC). This means that the variables are divided into blocks and that, in each block, at most a given number of variables may be non-null. Bound-constrained mathematical programs with complementarity constraints (MPCCs) are the particular case in which each block is a complementarity pair, with at most one non-null variable. Every limit point generated by the method is B-stationary, without constraint qualifications. In this paper we consider additional constraints $h(x) = 0$ and $c(x) \leq 0$.

Without constraint qualifications, we do not know any method that guarantees B-stationarity for problems with general nonlinear constraints, not even in smooth nonlinear programming. In that generality, augmented Lagrangian, sequential quadratic programming, and interior point methods guarantee sequential optimality conditions such as AKKT [3] (see [24] for cardinality constraints), which may hold at points that are not B-stationary. For example, for Minimize $-x_1 - x_2$ subject to $x_1 x_2 = 0$ and $x \geq 0$, AKKT holds at $x^* = 0$ (take $x^k = (1/k, 1/k)$ and the multiplier k for $x_1 x_2 = 0$), which is not B-stationary.

For MPCCs with general constraints, regularization methods, such as Scholtes' relaxation [33], and penalty methods converge, under MPEC constraint qualifications, to C- or M-stationary points, which need not be B-stationary. With inexact solutions of the subproblems, even these guarantees may be weakened [22]. Relaxation methods are compared, theoretically and on MacMPEC, in [19]. SQP methods converge locally superlinearly near strongly stationary points [15], and interior-penalty methods converge globally to strongly stationary points under standard assumptions [27]. Augmented Lagrangian methods applied to the nonlinear programming reformulation of MPCCs were analyzed in [20, 5] and, for cardinality-constrained problems (see also [6, 12]), in [23], where convergence to strongly stationary points is proved under a problem-tailored quasinormality condition. Guo and Deng [18] keep the complementarity constraints, written with slack variables, in the subproblems of an augmented Lagrangian method, solve the subproblems by a nonmonotone projected gradient method, and obtain M-stationary limit points under MPCC-RCPLD. For this method, Andreani, da Rosa and Secchin [2] proved the boundedness of the multipliers under quasi-normality-type conditions, and observed that keeping the complementarity constraints in the subproblems improves the numerical stability. Closest to our approach, Jia, Kanzow, Mehlitz and Wachsmuth [21] keep nonconvex sets with inexpensive projections, such as complementarity or cardinality constraints, in the subproblems of an augmented Lagrangian method, and obtain M-stationary limit points under an asymptotic regularity condition. The straightforward augmented Lagrangian method with telescopic subproblems also keeps the nonconvex constraints in the subproblems and, if their stopping criterion includes a derivative test outside the current face, its feasible limit points are M-stationary under MPEC-LICQ (Section 4.4). In the method proposed in this paper, the subproblems are also solved by the Telescopic Method, but its swaps are modified so that they preserve the values of the constraints, and the limit points are B-stationary (S-stationary for MPCCs) under PCC-LICQ (MPEC-LICQ for MPCCs). Active-set and hybrid methods, such as the one of Lin and Fukushima [29], assume MPEC-LICQ and use the signs of the multipliers to change the active set or to certify S-stationarity. The method of Nurkanović and Leyffer [30] solves a sequence of nonlinear programs associated with branches of the feasible set (BNLPs) and of linear programs with equilibrium constraints (LPECs). The LPECs either certify B-stationarity or provide a new branch. Global convergence to B-stationary points is proved assuming MPEC-MFCQ at

every feasible point and that the BNLP solver returns stationary points with a uniform sufficient decrease. For bound-constrained MPCCs, the method of Kirches, Larson, Leyffer and Manns [25] converges to B-stationary points without constraint qualifications. Other constraints may be handled by an augmented Lagrangian method, but, as in Section 4, the B-stationarity of the subproblems is not inherited by the limit points. To our knowledge, no method guarantees B-stationarity for PCC problems with general constraints. For MPCCs, the closest result is the one of [30], which requires MPEC-MFCQ at every feasible point and solves LPECs, whereas our result requires PCC-LICQ only at the limit point, uses only swaps, and applies to general partitioned cardinality constraints. The two assumptions are not comparable.

The contributions of this paper are the following.

- The straightforward safeguarded augmented Lagrangian method [1, 11] with telescopic subproblems (Algorithm 4.1). Its limit points are B-stationary, without constraint qualifications, if the penalty parameters remain bounded (Theorem 4.3), but not in general (counterexample of Proposition 4.4). If the penalty parameters tend to infinity, its feasible limit points at which PCC-LICQ holds are M-stationary for MPCCs (Theorem 4.6 and Corollary 4.7), thanks to condition (c) of the stopping criterion of the subproblems, which cannot be dropped (counterexample of Proposition 4.8).
- A characterization of B-stationarity for PCC problems with general constraints (Proposition 3.4).
- The constraint-preserving swap (Algorithm 5.2), based on a scaled Newton restoration with a computable tolerance (Lemma 5.1).
- The modified augmented Lagrangian method (Algorithm 5.3), whose feasible limit points at which PCC-LICQ holds are B-stationary, also when the penalty parameters tend to infinity and for the quadratic penalty method (Theorem 6.1).
- The modification that gives rise to Algorithm 5.3 is necessary, even under MPEC-LICQ and with exact solutions of the subproblems (counterexamples of Propositions 6.4, 6.5 and 6.6), and PCC-LICQ cannot be replaced by a condition of Mangasarian–Fromovitz type (counterexample of Proposition 6.7).
- Numerical confirmation of the counterexamples (Section 7.1), and experiments with the 193 instances of the MacMPEC collection (Section 7.2), in which Algorithm 5.3 certified more solutions than Algencan [1, 11] applied to the nonlinear programming reformulation and the regularization method of Scholtes [33]. In the 191 problems used by Nurkanović and Leyffer [30], the success rate of Algorithm 5.3 was 97.4%, while the highest success rate reported in [30], obtained by MPECopt, was 94.24%.
- Experiments with cardinality-constrained portfolios and sparse principal component analysis (Section 7.3). In these problems, the constraint-preserving swaps improved the original swaps. In most problems, Algorithm 5.3 also found better points than Algencan and the regularization method of Scholtes, applied to the complementarity-type reformulation of [12]. In sparse principal component analysis, it also found better points than the truncated power method of [34].

Section 2 gives an overview of the paper, and Section 3 describes the problem and the stationarity concepts. Section 4 presents the augmented Lagrangian method with telescopic subproblems, its convergence when the penalty parameters remain bounded, the counterexample, and the properties of the limit points when the penalty parameters are unbounded. Constraint-preserving swaps are introduced in Section 5, and the augmented Lagrangian method that uses them is defined in Section 5.2. Section 6 contains its convergence analysis and shows that

the modification is necessary and that MPEC-LICQ cannot be replaced by MPEC-MFCQ. Numerical experiments are reported in Section 7, and conclusions are given in Section 8.

Notation. Throughout the paper, $\|\cdot\|$ is the Euclidean norm, $\|\cdot\|_\infty$ is the sup norm and, for a vector v, $\|v\|_0$ is the number of non-null entries of v. For $t \in \mathbb{R}$, $(t)_+ = \max\{0, t\}$, applied componentwise to vectors. P_C denotes the Euclidean projection onto a closed convex set C, $T_C(x)$ is the (Bouligand) tangent cone of a set C at $x \in C$, and $e^{i,\nu}$ (or e^v) is the canonical vector associated with the variable $x_{i,\nu}$ (or x_v). For a differentiable function ϕ, $\partial_v \phi(x) = \partial\phi(x)/\partial x_v$ (or $\partial_{i,\nu}\phi(x) = \partial\phi(x)/\partial x_{i,\nu}$).

2 Overview

We consider the problem

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0,\ c(x) \le 0,\ x \in \Omega, \tag{2}$$

where f, h and c are smooth and Ω is defined by bounds on the variables and by partitioned cardinality constraints. This means that the variables are divided into blocks and, in each block, at most a given number of variables are allowed to be non-null. Complementarity constraints $0 \le y \perp z \ge 0$ are a particular case of partitioned cardinality constraints, in which each pair (y_i, z_i) is a block with at most one non-null variable. Binary variables are also a particular case of partitioned cardinality constraints (Section 3).

In [8] we considered problem (2) without the constraints $h(x) = 0$ and $c(x) \le 0$, and we introduced the Telescopic Method. For this purpose, the set Ω is divided into faces. Each face is defined by the variables that are allowed to be non-null on it and is a box. The Telescopic Method alternates iterations of a bound-constrained algorithm on the current face with *swaps*, in which a variable of the face leaves it (it is set to zero) and a variable outside the face enters it, moving away from zero along a one-dimensional search. A swap is accepted if it decreases f sufficiently, and then the method continues on the new face. Every limit point of the sequence generated by the method is B-stationary, without constraint qualifications.

To handle the additional constraints, it is natural to use an augmented Lagrangian method, as Algencan [1, 11], in which the constraints $h(x) = 0$ and $c(x) \le 0$ are penalized and the set Ω is kept in the subproblems. At each outer iteration, the subproblem (minimization of the augmented Lagrangian on Ω) is solved by the Telescopic Method. We will show that, if the penalty parameters remain bounded, every limit point is B-stationary, without constraint qualifications (Section 4). However, if the penalty parameters tend to infinity, the limit points may fail to be B-stationary, even if every subproblem is solved exactly (counterexample of Proposition 4.4). The reason is that a swap could move the point to places where the constraints have different values than the ones at the current point, in such a way that, due to a large penalty parameter, a promising swap could be rejected, in spite of having decreased the objective function. When the penalty parameters tend to infinity and the stopping criterion of the subproblems also looks outside the current face, the feasible limit points are M-stationary for MPCCs under MPEC-LICQ (Section 4.4), but not necessarily B-stationary. Moreover, in this case, every limit point is stationary for the minimization of the infeasibility measure $\|h(x)\|^2 + \|c(x)_+\|^2$ on a face of Ω (Proposition 4.5). In particular, this necessarily occurs if the feasible set is empty.

In order to remedy the inconvenience described above, we introduce *constraint-preserving swaps* (Section 5). In a constraint-preserving swap, the leaving variable is set to zero and the entering variable is moved away from zero by a backtracking search on its value. For each trial value, the other free variables of the face are corrected by a Newton-type method, so that

the resulting point w satisfies $h(w) \approx h(x^k)$. If L_k is the augmented Lagrangian of the k-th subproblem and $h(w) = h(x^k)$, then

$$L_k(w) - L_k(x^k) = f(w) - f(x^k)$$

for all values of the penalty parameter and of the multipliers. Therefore, decreases of L_k along such swaps correspond to decreases of f, and the acceptance of promising swaps is not impaired by large penalty parameters. We will prove that, under a linear independence constraint qualification at the limit point, the correction is well defined and small. Then the analysis of the Telescopic Method can be repeated with the Lagrangian of problem (2) in the place of L_k. With this modification, every feasible limit point at which this constraint qualification holds is B-stationary, even if the penalty parameters tend to infinity (Sections 5.2 and 6). The stationarity of the infeasible limit points for the infeasibility measure (Proposition 4.5) is also preserved. Finally, we show that the constraint-preserving swap is necessary and that the constraint qualification cannot be weakened to a Mangasarian–Fromovitz type condition (Section 6.2).

3 Problem and stationarity

We consider problem (2), where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $c : \mathbb{R}^n \to \mathbb{R}^q$, and Ω is defined by bounds and partitioned cardinality constraints, as follows. The vector of variables is partitioned into disjoint blocks $\vec{x}_\nu \in \mathbb{R}^{n_\nu}$, $\nu = 1, \ldots, N$, with $n_1 + \cdots + n_N = n$, and we write $x_{i,\nu}$ for the i-th variable of block ν and x_v, $v = 1, \ldots, n$, when the block structure is not needed. Given a box $\mathcal{B} = [a, b]$ and integers $p_\nu \in \{1, \ldots, n_\nu\}$,

$$\Omega = \{x \in \mathcal{B} \mid \|\vec{x}_\nu\|_0 \leq p_\nu,\ \nu = 1, \ldots, N\},$$

that is, at most p_ν variables of block ν may be non-null. We denote by $\mathcal{N}_c = \{\nu \mid p_\nu < n_\nu\}$ the set of blocks in which the cardinality constraint is effective. Mathematical programs with complementarity constraints (MPCCs) are an important particular case, which we use in the examples and in the experiments, in which every $\nu \in \mathcal{N}_c$ is a complementarity pair, with $n_\nu = 2$, $p_\nu = 1$ and $\vec{a}_\nu = 0$. Problem (2) also includes binary variables, because $x_j \in \{0, 1\}$ is equivalent to $x_j + \bar{x}_j = 1$, where $\bar{x}_j$ is an additional variable and $(x_j, \bar{x}_j)$ is a complementarity pair, that is, a block with $n_\nu = 2$ and $p_\nu = 1$, with bounds $[0, 2]$ for both variables. (With the bounds $[0, 1]$, the variable that is equal to one would be at its upper bound, and PCC-LICQ, defined below, would fail at every feasible point.) All the results of this paper apply to this reformulation.

Writing $c(x) + s = 0$ with slack variables $s \in [0, s_{\max}]$, where $s_{\max} \geq \max_{x \in \mathcal{B}} \|(-c(x))_+\|_\infty$, problem (2) becomes a problem of the same form with equality constraints only, in which the slacks form an additional block ν with $p_\nu = n_\nu$, so that $\nu \notin \mathcal{N}_c$. Its feasible set is the graph of $x \mapsto -c(x)$ on the feasible set of (2), so that B-stationarity is preserved, and so are the constraint qualifications (see the comment after Definition 3.3). Therefore, except in Section 4, where the inequality constraints are handled directly by the augmented Lagrangian method, as in Algencan, we consider

$$\text{Minimize } f(x) \text{ subject to } h(x) = 0,\ x \in \Omega, \tag{3}$$

where, with some abuse of notation, x includes the slack variables, $h : \mathbb{R}^n \to \mathbb{R}^m$ includes the constraints $c(x) + s = 0$, and we denote by $X = \{x \in \Omega \mid h(x) = 0\}$ its feasible set. We denote by $J(x) \in \mathbb{R}^{m \times n}$ the Jacobian of h, by $J_v(x)$ its v-th column and, for $S \subseteq \{1, \ldots, n\}$, by $J_S(x)$ the submatrix formed by the columns with indices in S.

We now define the faces of Ω and the projected derivatives. A *face index* is an N-tuple $I = (I_1, \ldots, I_N)$ such that $I_\nu \subseteq \{1, \ldots, n_\nu\}$ and $\#I_\nu = p_\nu$ for all ν. The corresponding *face* is the box

$$F_I = \{x \in \mathcal{B} \mid x_{i,\nu} = 0 \text{ whenever } i \notin I_\nu,\ \nu = 1, \ldots, N\},$$

and Ω is the union of the finitely many faces. For a continuously differentiable function ϕ, $x \in \mathcal{B}$, $\nu = 1, \ldots, N$ and $i = 1, \ldots, n_\nu$, the *projected derivative* is

$$g^\phi_{i,\nu}(x) = P_{[a_{i,\nu}, b_{i,\nu}]}\left(x_{i,\nu} - \frac{\partial \phi}{\partial x_{i,\nu}}(x)\right) - x_{i,\nu}, \tag{4}$$

which vanishes if and only if $x_{i,\nu}$ satisfies the first-order condition for minimizing ϕ with respect to $x_{i,\nu} \in [a_{i,\nu}, b_{i,\nu}]$, the other variables being fixed. For $x \in F_I$, the projected gradient $P_{F_I}(x - \nabla\phi(x)) - x$ is the vector $g^\phi_I(x) \in \mathbb{R}^n$ with components $g^\phi_{i,\nu}(x)$ if $i \in I_\nu$ and 0 otherwise.

We use the stationarity concept of [8], given in Definition 3.1. By Proposition 3.2, it is equivalent to B-stationarity on Ω. Assumption 1, stated after that proposition, is used in the rest of the paper.

Definition 3.1. A point $x^* \in \Omega$ is *stationary for the minimization of ϕ on Ω* if, for all $\nu = 1, \ldots, N$,

$$x^*_{i,\nu} \neq 0 \ \Rightarrow\ g^\phi_{i,\nu}(x^*) = 0, \quad i = 1, \ldots, n_\nu, \tag{5}$$

and

$$\|\vec{x}^*_\nu\|_0 < p_\nu \ \Rightarrow\ g^\phi_{i,\nu}(x^*) = 0 \text{ for all } i = 1, \ldots, n_\nu. \tag{6}$$

Condition (5) concerns the non-null variables, whereas condition (6) says that, if a block has room for one more non-null variable, then no variable of the block can decrease ϕ by moving away from zero. In the complementarity case, if both components of a pair vanish, (6) says that both partial derivatives are non-negative.

Proposition 3.2. *A point $x^* \in \Omega$ is stationary for the minimization of ϕ on Ω if and only if $\nabla\phi(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$.*

Proof. See Proposition 2.3 of [8], with ϕ in the place of f. □

Assumption 1. For all v, $a_v \leq b_v$ and, if v is in a block $\nu \in \mathcal{N}_c$, also $a_v \leq 0 \leq b_v$ and $a_v < b_v$. The box $\mathcal{B}$ is bounded. The functions f and h are continuously differentiable on an open set containing $\mathcal{B}$, and ∇f and J are Lipschitz continuous on $\mathcal{B}$.

The constraint qualification of our main result involves the set of *free variables* at $x^* \in \mathcal{B}$,

$$\mathcal{F}(x^*) = \{v \in \{1, \ldots, n\} \mid a_v < x^*_v < b_v, \text{ and } x^*_v \neq 0 \text{ if } v \text{ belongs to a block } \nu \in \mathcal{N}_c\}.$$

Definition 3.3. We say that *PCC-LICQ* holds at $x^* \in X$ if the matrix $J_{\mathcal{F}(x^*)}(x^*)$ has rank m.

For example, let $N = 1$, $n = 5$, $p_1 = 3$, $m = 2$, and $x^* = (4, 0, 0, 2, 0)^T$, with $a_v < x^*_v < b_v$ for $v = 1, 4$. Although the block admits one more non-null variable, the faces that contain x^* differ in the variable that they allow to move, and only x_1 and x_4 are free in all of them. Thus $\mathcal{F}(x^*) = \{1, 4\}$, and PCC-LICQ holds if and only if $\nabla h_1(x^*)$, $\nabla h_2(x^*)$, e^2, e^3, and e^5 are linearly independent. This choice of $\mathcal{F}(x^*)$ is what the proof of Proposition 3.4 requires, and it agrees with the convention of MPEC-LICQ (see below). In a block $\nu \notin \mathcal{N}_c$, such as the block of slacks, a variable is free if it is strictly between its bounds. If $\mathcal{F}(x^*) = \emptyset$, then PCC-LICQ holds only if $m = 0$.

For problem (2), written with slack variables, the slack of an inactive inequality constraint is a free variable, and the slack of an active one is at its lower bound. The column of the slack s_i in the Jacobian of the constraints is the canonical vector of the row of c_i, and it belongs to $J_{\mathcal{F}(x^*)}$ exactly when c_i is inactive. Therefore, PCC-LICQ is equivalent to the linear independence of the vectors of partial derivatives, with respect to the free components of x, of $h_1, \ldots, h_m$ and of the active inequality constraints c_i of (2). Equivalently, the gradients of $h_1, \ldots, h_m$ and of the active c_i, together with the canonical vectors e^v of the components v of x that are not free, are linearly independent. Each variable that is not free contributes a single canonical vector, even if it is null and at a bound at the same time. In the MPCC case, a null component of a complementarity pair is at its lower bound, so that $\mathcal{F}(x^*)$ is the set of variables that are strictly between their bounds, and PCC-LICQ says that the gradients of $h_1, \ldots, h_m$ and the canonical vectors of the variables at a bound are linearly independent. This is the usual MPEC-LICQ, that is, the LICQ of the nonlinear program in which the null components of the pairs are fixed at zero (the tightened nonlinear program of [32]), when the bounds are considered as constraints. In this case we keep the name MPEC-LICQ.

Proposition 3.4 below relates the two stationarity concepts used in this paper. A point $x^* \in X$ is B-stationary for (3) if $\nabla f(x^*)^T d \geq 0$ for all $d \in T_X(x^*)$. For $\lambda \in \mathbb{R}^m$, let $\ell_\lambda(x) = f(x) + \lambda^T h(x)$. By Proposition 3.2, x^* is stationary for the minimization of ℓ_λ on Ω (Definition 3.1) if and only if $\nabla \ell_\lambda(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$. The second concept is the one that the methods of this paper can reach, because the constraints $h(x) = 0$ are penalized and only Ω is kept in the subproblems.

Proposition 3.4. *Let Assumption 1 hold and $x^* \in X$.*

1. *If there exists $\lambda \in \mathbb{R}^m$ such that x^* is stationary for the minimization of ℓ_λ on Ω, then x^* is B-stationary for (3).*
2. *If PCC-LICQ holds at x^* and x^* is B-stationary for (3), then there exists a unique $\lambda^* \in \mathbb{R}^m$ such that x^* is stationary for the minimization of ℓ_{λ^*} on Ω. Moreover, λ^* is the unique solution of $\nabla_{\mathcal{F}} f(x^*) + J_{\mathcal{F}}(x^*)^T \lambda = 0$, where $\mathcal{F} = \mathcal{F}(x^*)$.*

Proof. Item 1. Let $d \in T_X(x^*)$, and let $t_j \downarrow 0$ and $d^j \to d$ be such that $x^* + t_j d^j \in X$. Since $X \subseteq \Omega$, $d \in T_\Omega(x^*)$ and, since $h(x^* + t_j d^j) = h(x^*) = 0$, $J(x^*)d = 0$. Therefore, $\nabla f(x^*)^T d = \nabla \ell_\lambda(x^*)^T d \geq 0$.

Item 2. Let $\mathcal{I}(x^*)$ be the set of face indices I such that $x^* \in F_I$ and, for $I \in \mathcal{I}(x^*)$, let $X_I = F_I \cap \{x \mid h(x) = 0\} \subseteq X$. Since $T_{X_I}(x^*) \subseteq T_X(x^*)$, x^* is B-stationary for the minimization of f on X_I. The constraints of this problem are $h(x) = 0$ and the bounds of the box F_I, and the variables that are strictly inside the bounds of F_I at x^* include those of $\mathcal{F}$, because a variable x_v with $v \in \mathcal{F}$ that belongs to a block of $\mathcal{N}_c$ is non-null at x^* and, therefore, is not fixed at zero in F_I. Thus, PCC-LICQ implies the LICQ for this problem at x^* and, therefore, B-stationarity implies the KKT conditions, that is, there exists λ^I such that $g_I^{\ell_{\lambda^I}}(x^*) = 0$. Since, for every $v \in \mathcal{F}$, the variable x_v is not fixed at zero in any face F_I with $I \in \mathcal{I}(x^*)$ and $a_v < x_v^* < b_v$, we have that $\partial \ell_{\lambda^I}(x^*)/\partial x_v = 0$. Hence, $\nabla_{\mathcal{F}} f(x^*) + J_{\mathcal{F}}(x^*)^T \lambda^I = 0$ and, since $J_{\mathcal{F}}(x^*)$ has rank m, $\lambda^I = \lambda^*$ for all $I \in \mathcal{I}(x^*)$. Thus, $g_I^{\ell_{\lambda^*}}(x^*) = 0$ for all $I \in \mathcal{I}(x^*)$ which, as in the proof of Proposition 2.3 of [8], is equivalent to the stationarity of x^* for the minimization of ℓ_{λ^*} on Ω. Uniqueness follows from the same argument. □

In the MPCC case, stationarity of x^* for the minimization of ℓ_{λ^*} on Ω means the following. For the variables that are not null components of complementarity pairs, the partial derivatives of ℓ_{λ^*} at x^* satisfy the usual sign conditions at the bounds. There is no condition on the derivative with respect to a null component whose partner is non-null, and both derivatives are non-negative for a biactive pair. This is S-stationarity [32].

4 Augmented Lagrangian method with telescopic subproblems

In this section we consider problem (2), in which, besides $x \in \Omega$, there are general equality and inequality constraints $h(x) = 0$ and $c(x) \leq 0$. A natural way of handling this problem is the augmented Lagrangian method of Algencan [1, 11], in which the constraints $h(x) = 0$ and $c(x) \leq 0$ are penalized, while $x \in \Omega$ (the bounds and the partitioned cardinality constraints) is kept in the subproblems, which are solved by the Telescopic Method. We call it the *straightforward* method. We prove that, if the penalty parameters remain bounded, every limit point of the sequence generated by the straightforward method is B-stationary, without constraint qualifications (Theorem 4.3). However, if the penalty parameters tend to infinity, the limit points may fail to be B-stationary, even if every subproblem is solved exactly (counterexample of Proposition 4.4, illustrated numerically at the end of Section 4.3). In this case, we show in Section 4.4 that the infeasible limit points are stationary for the infeasibility measure on a face and that, in the MPCC case, the feasible limit points at which MPEC-LICQ holds are M-stationary. This is why, in Section 5.2, we introduce a less obvious method, in which the swaps of the Telescopic Method are replaced by constraint-preserving swaps (Section 5).

4.1 The Telescopic Method

In this subsection we summarize the Telescopic Method of [8] and the results of [8] that are used in this paper. The method minimizes a function ϕ on Ω, where the bounds satisfy the first part of Assumption 1, $\mathcal{B}$ is bounded, and ϕ is continuously differentiable on an open set containing $\mathcal{B}$, with $\nabla\phi$ Lipschitz continuous on $\mathcal{B}$ with constant L_ϕ. (In [8], these are Assumptions A1–A3, the last one being the boundedness of a level set, which holds because $\mathcal{B}$ is bounded.) The method alternates swaps, which change the face, with iterations of a bound-constrained algorithm on the current face.

The swaps use a one-dimensional *escaping procedure* (Algorithm 3.1 of [8]). Let $\bar{y} \in \mathcal{B}$ with $\bar{y}_{j,\nu} = 0$ and $g^\phi_{j,\nu}(\bar{y}) \neq 0$, and let $\psi(t) = \phi(\bar{y} + te^{j,\nu})$ for $t \in [a_{j,\nu}, b_{j,\nu}]$. Given $\alpha_{esc} > 0$ and $\sigma_{\min} > 0$, the procedure computes, for $\sigma = \sigma_{\min}, 2\sigma_{\min}, 4\sigma_{\min}, \ldots$, the minimizer $t = P_{[a_{j,\nu},b_{j,\nu}]}(-\psi'(0)/\sigma)$ of the model $\psi(0) + \psi'(0)t + \sigma t^2/2$ on $[a_{j,\nu}, b_{j,\nu}]$, and stops with $t_{out} = t$ at the first σ for which $\psi(t) \leq \psi(0) - \alpha_{esc}t^2$. By Lemma 3.1 of [8], it stops after finitely many trials, and

$$\phi(\bar{y} + t_{out}e^{j,\nu}) \leq \phi(\bar{y}) - \theta\, g^\phi_{j,\nu}(\bar{y})^2, \quad \text{where} \quad \theta = \frac{\alpha_{esc}}{\max\{1, \bar{\sigma}\}^2} \quad \text{and} \quad \bar{\sigma} = \max\{\sigma_{\min}, L_\phi + 2\alpha_{esc}\}. \tag{7}$$

Thus, the decrease is bounded below in terms of the projected derivative of the entering variable, with a constant θ that depends only on the parameters and on L_ϕ.

On a face F_I, the minimization of ϕ is a bound-constrained problem, and $x \in F_I$ is a stationary point of this problem if and only if $g^\phi_I(x) = 0$. An iteration of the method used on the faces, called Algorithm MIA in [8] (Gencan in our implementation), computes a point $\hat{w} \in F_I$ from a point $w \in F_I$. Since the Telescopic Method leaves a face after one iteration of Algorithm MIA and may return to it later at a different point, the assumptions on Algorithm MIA concern the following sequences. A sequence $\{w^j\} \subseteq F_I$ is an *MIA sequence with replacements* if, for all j, $\phi(w^{j+1}) \leq \phi(\hat{w}^{j+1})$, where $\hat{w}^{j+1}$ is computed by an iteration of Algorithm MIA from w^j.

Assumption 2. (Assumptions A4–A6 of [8].) For every face F_I, (i) an iteration of Algorithm MIA from $w \in F_I$ computes $\hat{w} \in F_I$ with $\phi(\hat{w}) \leq \phi(w)$, (ii) every limit point w^* of an MIA sequence with replacements satisfies $g^\phi_I(w^*) = 0$, and (iii) if $\{w^j\}$ is an MIA sequence with replacements and $\lim_{j\in J} w^j = w^*$ for an infinite set J, then $\lim_{j\in J} \|\hat{w}^{j+1} - w^j\| = 0$.

These are natural properties of bound-constrained methods, and (iii) holds, for example, if $\|\hat{w} - w\| \leq C\|g_I^\phi(w)\|$ for some constant C.

The Telescopic Method (Algorithm 5.1 of [8]) is given $\delta \in (0, +\infty]$, $\eta_{\rm br} \in (0, +\infty]$, the parameters of the escaping procedure, and $x^0 \in F_{I^0}$. At iteration k, it performs a *swap search* at $x^k \in F_{I^k}$. For every $\nu \in \mathcal{N}_c$, every $i \in I_\nu^k$ with $|x_{i,\nu}^k| \leq \delta$ and every $j \notin I_\nu^k$ such that $g_{j,\nu}^\phi(\bar{y}) \neq 0$, where $\bar{y} = x^k - x_{i,\nu}^k e^{i,\nu}$ is x^k with the leaving variable set to zero, the escaping procedure gives the point $w = \bar{y} + t_{out} e^{j,\nu}$, which belongs to the face obtained from I^k by replacing i with j in I_ν^k, and satisfies $\phi(w) < \phi(\bar{y})$. The best of these points and its value f_{best} are kept, and the search is interrupted as soon as $f_{best} \leq \phi(x^k) - \eta_{\rm br}$ (brute acceptance). If $f_{best} < \phi(x^k)$, the best point becomes z^k and its face index becomes J^k. Otherwise, $z^k = x^k$ and $J^k = I^k$. Then $x^{k+1} \in F_{J^k}$ is computed by one iteration of Algorithm MIA from z^k, and $I^{k+1} = J^k$. The analysis uses only that $z^k \in F_{J^k}$ and $\phi(z^k) \leq f_{best}$. Therefore, as observed in Remark 5.1 of [8], z^k may be the best of several candidates improved by Algorithm MIA in their new faces (the *look-ahead* strategy of our implementation), and any finite number of iterations of Algorithm MIA may be performed on each visited face. The following lemma collects the results of [8] used in this paper.

Lemma 4.1. *Let the assumptions above on $\mathcal{B}$ and ϕ and Assumption 2 hold, and let $\{x^k\}$ and $\{z^k\}$ be generated by the Telescopic Method applied to the minimization of ϕ on Ω.*

1. *For all k, $\phi(x^{k+1}) \leq \phi(z^k) \leq \phi(x^k)$, and $\lim_{k\to\infty}[\phi(x^k) - \phi(z^k)] = 0$.*
2. *If $J^k = I$ for all k in an infinite set K, then every limit point z^* of $\{z^k\}_{k\in K}$ belongs to F_I and satisfies $g_I^\phi(z^*) = 0$ and, if $\lim_{k\in K} z^k = z^*$, then $\lim_{k\in K} x^{k+1} = z^*$.*
3. *Every limit point of $\{x^k\}$ is stationary for the minimization of ϕ on Ω (Definition 3.1) or, equivalently, B-stationary.*
4. *For every $\varepsilon > 0$, there exists k such that the swap search at iteration k is not interrupted by the brute acceptance test, $\|g_{I^k}^\phi(x^k)\|_\infty \leq \varepsilon$, and $\phi(w) \geq \phi(x^k) - \varepsilon$ for every point w computed in that swap search.*

Proof. Items 1 and 2 are Lemmas 6.1, 6.2 and 6.3 of [8], item 3 is Theorem 6.4 of [8], and item 4 is Lemma 6.5 of [8]. The proof of item 3 combines items 1 and 2 with (7). If $J^k = I$ and $z^k \to x^*$ for $k \in K$, and a block $\nu \in \mathcal{N}_c$ with $\|\vec{x}_\nu^*\|_0 < p_\nu$ had a variable $x_{j,\nu}$, $j \notin I_\nu$, with $g_{j,\nu}^\phi(x^*) \neq 0$, then, since I_ν contains an index i with $x_{i,\nu}^* = 0$, by (7), the swap $x_{i,\nu} \to x_{j,\nu}$ at x^{k+1} (or the brute acceptance test, if it interrupts the search first) would give $\phi(z^{k+1}) \leq \phi(x^{k+1}) - c$ for a constant $c > 0$ and all $k \in K$ large enough, contradicting item 1. □

4.2 Algorithm and convergence with bounded penalty parameters

We denote by $X = \{x \in \Omega \mid h(x) = 0,\ c(x) \leq 0\}$ the feasible set of (2), and say that $x^* \in X$ is B-stationary for (2) if $\nabla f(x^*)^T d \geq 0$ for all $d \in T_X(x^*)$. We denote by $\nabla h(x)$ and $\nabla c(x)$ the $n \times m$ and $n \times q$ matrices whose columns are the gradients of the components of h and c, respectively. For $\rho > 0$, $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}_+^q$, the (Powell–Hestenes–Rockafellar) augmented Lagrangian is

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

In this section we use the following assumption, under which the augmented Lagrangian has the regularity required by the Telescopic Method.

Assumption 3. The bounds satisfy the first part of Assumption 1, $\mathcal{B}$ is bounded, and f, h and c are continuously differentiable on an open set containing $\mathcal{B}$, with Lipschitz continuous gradients on $\mathcal{B}$.

Under Assumption 3, for all ρ, λ and μ, the function $L_\rho(\cdot, \lambda, \mu)$ satisfies the assumptions on ϕ of Section 4.1, since $t \mapsto (t)_+^2$ is continuously differentiable with Lipschitz continuous derivative. For such functions ϕ, the projected derivatives $g_{i,\nu}^\phi$ and g_I^ϕ are defined in Section 3. We denote by $\bar{\varepsilon}$ half of the smallest absolute value of the non-null bounds a_v, b_v of the variables of the blocks of $\mathcal{N}_c$.

The following algorithm uses the safeguarded multiplier updates and the penalty parameter update of Algencan [1, 11].

Algorithm 4.1 (Augmented Lagrangian with telescopic subproblems)
Let $\tau \in (0,1)$, $\gamma > 1$, $\rho_1 > 0$, $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^1 \in [0, \mu_{\max}]^q$, a sequence $\{\varepsilon_k\}$ with $\varepsilon_k > 0$ and $\lim_{k\to\infty} \varepsilon_k = 0$, the parameters δ and η_{br} of the Telescopic Method, the parameters of the escaping procedure, and $x^0 \in \Omega$ be given. Set $k \leftarrow 1$.

Step 1 (Subproblem). Let $L_k(x) = L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k)$. Starting from x^{k-1}, apply the Telescopic Method to the minimization of L_k on Ω until a point $x^k \in \Omega$ and a face index I^k with $x^k \in F_{I^k}$ are found such that

(a) $\|g_{I^k}^{L_k}(x^k)\|_\infty \leq \varepsilon_k$;

(b) no original swap decreases L_k significantly from x^k: for every $\nu \in \mathcal{N}_c$, $i \in I_\nu^k$ with $|x_{i,\nu}^k| \leq \delta$ and $j \notin I_\nu^k$, let $\bar{y} = x^k - x_{i,\nu}^k e^{i,\nu}$ (x^k with the leaving variable set to zero) and, if $g_{j,\nu}^{L_k}(\bar{y}) \neq 0$, the point $w = \bar{y} + t_{out} e^{j,\nu}$, where t_{out} is computed by the one-dimensional escaping procedure of the Telescopic Method (Section 4.1) applied to $t \mapsto L_k(\bar{y} + te^{j,\nu})$, $t \in [a_{j,\nu}, b_{j,\nu}]$, satisfies $L_k(w) \geq L_k(x^k) - \varepsilon_k$;

(c) for every $\nu \in \mathcal{N}_c$, $i \in I_\nu^k$ with $|x_{i,\nu}^k| \leq \min\{\varepsilon_k, \bar{\varepsilon}\}$ and $j \notin I_\nu^k$, $|\partial_{i,\nu} L_k(x^k)| \leq \varepsilon_k$ or $|g_{j,\nu}^{L_k}(x^k)| \leq \varepsilon_k$.

Step 2 (Penalty parameter). Let $V^k = \max\{c(x^k), -\bar{\mu}^k/\rho_k\}$ (componentwise) and $M_k = \max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\}$. If $k > 1$ and $M_k > \tau M_{k-1}$, set $\rho_{k+1} = \gamma \rho_k$. Otherwise, set $\rho_{k+1} = \rho_k$.

Step 3 (Multipliers). Set

$$\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k), \quad \mu^{k+1} = (\bar{\mu}^k + \rho_k c(x^k))_+,$$

choose $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ and $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^q$ (for example, the projections of λ^{k+1} and μ^{k+1} onto these boxes, which are used in our implementation), set $k \leftarrow k+1$, and go to Step 1.

Conditions (a) and (b) say that x^k is approximately stationary in its face (the stopping criterion of the bound-constrained solver) and that no swap examined by the Telescopic Method decreases L_k significantly, which are the properties used in the proof of item 3 of Lemma 4.1. The practical stopping criterion of Section 5 of [8] (no swap found and (14) of [8] with tolerance ε_k) implies (a) and (b). Condition (c) says that, if a variable of the current face is almost null, then either it is approximately stationary also as an interior variable, or the null variables of its block cannot decrease L_k by moving away from zero. Conditions (a) and (c) together are an approximate version, in terms of derivatives, of the stationarity of x^k for the minimization of L_k on Ω (Definition 3.1), and condition (c) is the part of (6) that looks outside the current face.

Unlike the decrease guaranteed by (7), which is used in (b), its tolerance does not deteriorate when ρ_k grows. Condition (c) is not used in Theorem 4.3, and it is what gives M-stationarity when $\rho_k \to \infty$ (Section 4.4). The following lemma shows that Step 1 is well defined, and Theorem 4.3 shows that, if the penalty parameters remain bounded, the limit points are B-stationary.

Lemma 4.2. *Let Assumption 3 hold and let $\phi = L_\rho(\cdot, \lambda, \mu)$ for some $\rho > 0$, λ and $\mu \geq 0$. Let the Telescopic Method be applied to the minimization of ϕ on Ω and let $\varepsilon > 0$. Then, after a finite number of iterations, the point x^j at which the swap search of the Telescopic Method is performed satisfies (a), (b) and (c) with ϕ, x^j, I^j and ε in the places of L_k, x^k, I^k and ε_k, respectively.*

Proof. Since ϕ satisfies the assumptions on ϕ of Section 4.1, the proofs of items 3 and 4 of Lemma 4.1 (Theorem 6.4 and Lemma 6.5 of [8]) apply to ϕ. By the proof of item 4, (b) holds for all j large enough. By the proof of item 3, there exist an infinite set K, a face index I and a point x^*, stationary for the minimization of ϕ on Ω, such that $\lim_{j\in K} x^j = x^*$ and $I^j = I$ for all $j \in K$. By the continuity of g_I^ϕ, (a) holds for $j \in K$ large enough. Let $\bar{\varepsilon}_1 = \min\{\varepsilon, \bar{\varepsilon}\}$, $\nu \in \mathcal{N}_c$, $i \in I_\nu$ and $\ell \notin I_\nu$. If $x^*_{i,\nu} = 0$, then $\|\vec{x}^*_\nu\|_0 < p_\nu$, because $\#I_\nu = p_\nu$ and $x^*_{\ell',\nu} = 0$ for all $\ell' \notin I_\nu$. Then, by (6), $g^\phi_{\ell,\nu}(x^*) = 0$ and, by continuity, $|g^\phi_{\ell,\nu}(x^j)| \leq \varepsilon$ for $j \in K$ large enough. If $0 < |x^*_{i,\nu}| \leq \bar{\varepsilon}_1$, then $x^*_{i,\nu}$ is not a bound, because the non-null bounds have absolute values at least $2\bar{\varepsilon}$. Thus, by (5), $\partial_{i,\nu}\phi(x^*) = 0$ and, by continuity, $|\partial_{i,\nu}\phi(x^j)| \leq \varepsilon$ for $j \in K$ large enough. Finally, if $|x^*_{i,\nu}| > \bar{\varepsilon}_1$, then $|x^j_{i,\nu}| > \bar{\varepsilon}_1$ for $j \in K$ large enough, and (c) imposes no condition on i. Therefore, (a), (b) and (c) hold for $j \in K$ large enough. □

Theorem 4.3. *Let Assumption 3 hold, and let $\{x^k\}$ be generated by Algorithm 4.1. If $\{\rho_k\}$ is bounded, then every limit point x^* of $\{x^k\}$ is feasible for (2) and B-stationary. Moreover, there exist $\hat{\lambda} \in \mathbb{R}^m$ and $\hat{\mu} \in \mathbb{R}^q_+$, with $\hat{\mu}_i = 0$ whenever $c_i(x^*) < 0$, such that x^* is stationary for the minimization of $f + \hat{\lambda}^T h + \hat{\mu}^T c$ on Ω.*

Proof. Since $\{\rho_k\}$ is bounded and non-decreasing, there exist $\bar{\rho}$ and k_1 such that $\rho_k = \bar{\rho}$ for all $k \geq k_1$. By Step 2, $M_k \leq \tau^{k-k_1} M_{k_1}$ for $k \geq k_1$, so that $h(x^k) \to 0$ and $V^k \to 0$. Since $c(x^k) \leq V^k$, every limit point of $\{x^k\}$ is feasible.

Let x^* be a limit point. Since the safeguarded multipliers are bounded and there are finitely many face indices, there exist an infinite set K, $\tilde{\lambda}$, $\tilde{\mu}$ and a face index I such that $\lim_{k\in K} x^k = x^*$, $\lim_{k\in K} \bar{\lambda}^k = \tilde{\lambda}$, $\lim_{k\in K} \bar{\mu}^k = \tilde{\mu}$, and $I^k = I$ for all $k \in K$. If $c_i(x^*) < 0$, then $c_i(x^k) < c_i(x^*)/2$ for $k \in K$ large enough and, since $V^k_i \to 0$, $\bar{\mu}^k_i/\bar{\rho} \to 0$, so that $\tilde{\mu}_i = 0$. Define $\Phi(x) = L_{\bar{\rho}}(x, \tilde{\lambda}, \tilde{\mu})$.

Since $(x, \lambda, \mu) \mapsto L_{\bar{\rho}}(x, \lambda, \mu)$ and its gradient with respect to x are continuous, and $\mathcal{B}$ and the safeguarding boxes are compact, $L_k \to \Phi$, $\nabla L_k \to \nabla\Phi$ and, consequently, $g^{L_k}_{i,\nu} \to g^\Phi_{i,\nu}$ uniformly on $\mathcal{B}$ for $k \in K$. Moreover, there exist G and $\hat{L}$ such that, for all $k \geq k_1$, $\|\nabla L_k(x)\| \leq G$ for all $x \in \mathcal{B}$ and ∇L_k is Lipschitz continuous on $\mathcal{B}$ with constant $\hat{L}$. Hence (7) holds for all the functions L_k, $k \geq k_1$, with the same constant $\hat{\theta} = \alpha_{esc}/\max\{1, \hat{\sigma}\}^2$, where $\hat{\sigma} = \max\{\sigma_{\min}, \hat{L} + 2\alpha_{esc}\}$.

We prove that x^* is stationary for the minimization of Φ on Ω in the sense of Definition 3.1. By (a) and the uniform convergence, $g^\Phi_{i,\nu}(x^*) = 0$ for all $i \in I_\nu$, $\nu = 1, \ldots, N$, which, since $x^* \in F_I$, implies condition (5). Let $\nu \in \mathcal{N}_c$ be such that $\|\vec{x}^*_\nu\|_0 < p_\nu$ and let $\ell \notin I_\nu$ (for $\nu \notin \mathcal{N}_c$, condition (6) follows from the previous observation). As in the proof of item 3 of Lemma 4.1, there exists $i \in I_\nu$ with $x^*_{i,\nu} = 0$. Assume, by contradiction, that $g^\Phi_{\ell,\nu}(x^*) = \kappa \neq 0$, and let $\bar{y}^k = x^k - x^k_{i,\nu} e^{i,\nu}$, so that $\lim_{k\in K} \bar{y}^k = x^*$. For $k \in K$ large enough, $|x^k_{i,\nu}| \leq \delta$, $|g^{L_k}_{\ell,\nu}(\bar{y}^k)| \geq |\kappa|/2$, $L_k(\bar{y}^k) \leq L_k(x^k) + G|x^k_{i,\nu}| \leq L_k(x^k) + \hat{\theta}\kappa^2/8$, and $\varepsilon_k < \hat{\theta}\kappa^2/8$. Then the pair (i, ℓ) is examined

in the swap search from x^k and, by (7), the corresponding point w satisfies

$$L_k(w) \leq L_k(\bar{y}^k) - \hat{\theta}\kappa^2/4 \leq L_k(x^k) - \hat{\theta}\kappa^2/8 < L_k(x^k) - \varepsilon_k,$$

which contradicts (b). Therefore, condition (6) holds and, by Proposition 3.2 (applied to Φ), $\nabla\Phi(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$.

Since $h(x^*) = 0$, $\nabla\Phi(x^*) = \nabla f(x^*) + \nabla h(x^*)\hat{\lambda} + \nabla c(x^*)\hat{\mu}$, where $\hat{\lambda} = \tilde{\lambda}$ and $\hat{\mu} = (\tilde{\mu} + \bar{\rho}c(x^*))_+ \geq 0$, with $\tilde{\mu}_i = \hat{\mu}_i = 0$ if $c_i(x^*) < 0$. This proves the second statement of the theorem. Let $d \in T_X(x^*)$, and let $t_j \downarrow 0$ and $d^j \to d$ be such that $x^* + t_j d^j \in X$ for all j. Since $X \subseteq \Omega$, $d \in T_\Omega(x^*)$. Since $h(x^* + t_j d^j) = h(x^*) = 0$, $\nabla h(x^*)^T d = 0$ and, if $c_i(x^*) = 0$, since $c_i(x^* + t_j d^j) \leq 0$, $\nabla c_i(x^*)^T d \leq 0$. Therefore,

$$\nabla f(x^*)^T d = \nabla\Phi(x^*)^T d - \hat{\lambda}^T \nabla h(x^*)^T d - \sum_{i\,:\,c_i(x^*)=0} \hat{\mu}_i \nabla c_i(x^*)^T d \geq 0,$$

so that x^* is B-stationary for (2). □

No constraint qualification is used in Theorem 4.3. The only property of the constraints used in the last part of the proof is that the tangent cone of X is contained in the intersection of the tangent cone of Ω with the linearized cone of the constraints $h(x) = 0$ and $c(x) \leq 0$, which is always true. The hypothesis of the theorem is the boundedness of the penalty parameters. For nonlinear programming problems with general lower-level constraints, the boundedness of the penalty parameters of Algencan was proved in [1] under the linear independence of the gradients of the active constraints, a second-order condition, and the assumption that the Lagrange multipliers lie in the interior of the safeguarding boxes. Whether similar conditions guarantee the boundedness of $\{\rho_k\}$ in Algorithm 4.1, in which the lower-level set Ω is not convex, is an open question.

4.3 Unbounded penalty parameters: a counterexample

We first describe the mechanism of the failure. B-stationarity with respect to Ω is not a closed condition. At an iterate x^k at which a complementarity pair has one non-null component, the stationarity condition says nothing about the derivative with respect to the null component, but at the limit point both components may vanish, and B-stationarity then requires the corresponding partial derivative (of the Lagrangian) to be non-negative. In the Telescopic Method, this is taken care of by the swaps, in which the non-null component is set to zero and the null one is moved away from zero by a one-dimensional procedure whose decrease is bounded below by θg^2, where g is the projected derivative of the entering variable. For equality constraints only, the augmented Lagrangian is

$$L_k(x) = f(x) + \frac{\rho_k}{2}\left\|h(x) + \frac{\bar{\lambda}^k}{\rho_k}\right\|^2,$$

and both parts of a swap are penalized. Setting a variable x_i to zero changes h by $O(|x_i|)$, which costs $O(\rho_k x_i^2)$ in the penalty term. Moreover, the constant θ deteriorates as ρ_k grows, because the curvature of L_k along the entering variable contains the term $\rho_k(\partial h/\partial x_j)^2$. In the counterexample below, the cost of the first part of the swap tends to infinity.

When $\rho_k \to \infty$, the argument of the proof of Theorem 4.3 breaks down. The Lipschitz constants of ∇L_k grow with ρ_k, so that the constant θ of (7) tends to zero, and the difference $L_k(\bar{y}^k) - L_k(x^k)$, which involves the penalty of the (possibly infeasible) point $\bar{y}^k$, need not tend to zero. The following example shows that, in this case, the limit points of Algorithm 4.1 may fail to be B-stationary, even if every subproblem is solved exactly.

Example 1. Let $n = 3$, $N = 2$, $\vec{x}_1 = (x_1, x_2)$ with $p_1 = 1$, $\vec{x}_2 = x_3$ with $p_2 = 1$ (so that $\mathcal{N}_c = \{1\}$), $\mathcal{B} = [0,2] \times [0,1] \times [0,2]$, $c \equiv 0$ (no inequality constraints), and

$$f(x) = -\frac{x_2}{2} - \frac{x_1^2}{2} + \frac{x_3^2}{2} + \frac{x_3^4}{4}, \qquad h(x) = \frac{(x_1 - x_3)^2}{2} + x_2 x_3^2. \tag{8}$$

Since $h(x) = 0$ with $x \geq 0$ implies $x_1 = x_3$ and $x_2 x_3 = 0$, and $x_1 x_2 = 0$ in Ω, the feasible set is $X = \{(t,0,t) \mid t \in [0,2]\} \cup \{(0,t,0) \mid t \in [0,1]\}$, and the global minimizer is $(0,1,0)$, with $f = -1/2$. In the iterations of Algorithm 4.1, $L_k(x) = f(x) + \bar{\lambda}^k h(x) + \rho_k h(x)^2/2$ up to an additive constant.

Proposition 4.4. *Consider problem (2) with the data (8) and Algorithm 4.1 with $\lambda_{\min} \leq 0 < \lambda_{\max} < \infty$, $\bar{\lambda}^1 \in [0, \lambda_{\max}]$, $\rho_1 > 4$, $\bar{\lambda}^{k+1} = P_{[\lambda_{\min}, \lambda_{\max}]}(\lambda^{k+1})$, and arbitrary τ, γ, δ, η_{br} and $\{\varepsilon_k\}$. For each k, let s_k be the unique solution in $(0,1)$ of*

$$\bar{\lambda}^k + \frac{\rho_k s^6}{2} = \frac{1 + s^2}{s^2}, \tag{9}$$

and let $x^k = (s_k + s_k^3, 0, s_k)$ and $I^k = (\{1\}, \{1\})$. Then:

1. *x^k is a strict local minimizer of L_k on Ω and satisfies (a), (b) and (c) with $\varepsilon_k = 0$. Moreover, condition (b) holds trivially, because no swap is a candidate ($g_{j,\nu}^{L_k}(\bar{y}) = 0$ for the only possible swap).*
2. *$\lim_{k\to\infty} \rho_k = \infty$ and $\lim_{k\to\infty} x^k = x^* = 0$, which is feasible.*
3. *x^* is not B-stationary for (2).*

Proof. For all k, $\bar{\lambda}^k \in [0, \lambda_{\max}]$ and $\rho_k \geq \rho_1 > 4$. Indeed, $\{\rho_k\}$ is non-decreasing and, if $\bar{\lambda}^k \geq 0$, then, as shown below, $\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k) = (1 + s_k^2)/s_k^2 > 0$, so that $\bar{\lambda}^{k+1} = \min\{\lambda^{k+1}, \lambda_{\max}\} \in (0, \lambda_{\max}]$. The left-hand side of (9) is increasing and the right-hand side is decreasing in $s > 0$, the right-hand side tends to $+\infty$ when $s \to 0$, and at $s = 1$ the left-hand side, $\bar{\lambda}^k + \rho_k/2 > 2$, is larger than the right-hand side. Thus, $s_k \in (0,1)$ is well defined and $x^k \in F_{I^k}$, with $0 < x_1^k, x_3^k < 2$.

Item 1. Let $s = s_k$. Then $x_1^k - x_3^k = s^3$, $h(x^k) = s^6/2$ and, by (9), $\bar{\mu} := \bar{\lambda}^k + \rho_k h(x^k) = (1 + s^2)/s^2$. Since $\nabla L_k(x) = \nabla f(x) + (\bar{\lambda}^k + \rho_k h(x)) \nabla h(x)$, with $\nabla f(x) = (-x_1, -1/2, x_3 + x_3^3)^T$ and $\nabla h(x) = (x_1 - x_3, x_3^2, -(x_1 - x_3) + 2x_2 x_3)^T$, we have that $\partial L_k(x^k)/\partial x_1 = -(s + s^3) + \bar{\mu} s^3 = 0$ and $\partial L_k(x^k)/\partial x_3 = s + s^3 - \bar{\mu} s^3 = 0$. Since x_1^k and x_3^k are strictly between their bounds, $g_{I^k}^{L_k}(x^k) = 0$, so (a) holds. In a neighborhood of x^k, $x_1 > 0$ and, therefore, Ω coincides with $F_{I^k} = \{x \in \mathcal{B} \mid x_2 = 0\}$. The Hessian of L_k with respect to (x_1, x_3) at x^k is

$$\begin{pmatrix} P - 1 & -P \\ -P & P + 1 + 3s^2 \end{pmatrix}, \quad \text{where} \quad P = \bar{\lambda}^k + \frac{3}{2}\rho_k s^6 \geq \bar{\mu} = \frac{1 + s^2}{s^2}.$$

Its determinant is $3s^2 P - 1 - 3s^2 \geq 3s^2 \bar{\mu} - 1 - 3s^2 = 2$ and $P - 1 \geq 1/s^2 > 0$, so it is positive definite and x^k is a strict local minimizer of L_k on Ω. In the swap search from x^k, the only possibility is $i = 1$, $j = 2$ in block $\nu = 1$, with $\bar{y} = (0, 0, s)$ and $h(\bar{y}) = s^2/2$. By (9), $\rho_k s^4/2 = (1 + s^2)/s^4 - \bar{\lambda}^k/s^2$ and, since $\bar{\lambda}^k \leq \bar{\mu} = (1 + s^2)/s^2$,

$$\frac{\partial L_k}{\partial x_2}(\bar{y}) = -\frac{1}{2} + \left(\bar{\lambda}^k + \frac{\rho_k s^2}{2}\right) s^2 = -\frac{1}{2} + \frac{1 + s^2}{s^4} - \bar{\lambda}^k \left(\frac{1}{s^2} - s^2\right) \geq -\frac{1}{2} + 1 + s^2 > 0.$$

Since $\bar{y}_2 = 0$ is the lower bound of x_2, $g_{2,1}^{L_k}(\bar{y}) = 0$. Therefore, no point w is computed in (b), and (b) holds trivially. Condition (c) also holds, because $\partial L_k(x^k)/\partial x_1 = 0$.

Item 2. Assume that $\{\rho_k\}$ is bounded. Then there exist $\bar{\rho}$ and k_1 such that $\rho_k = \bar{\rho}$ and $h(x^k) \leq \tau h(x^{k-1})$ for all $k > k_1$. Thus, $s_k^6/2 = h(x^k) \to 0$ and, by (9), $\bar{\lambda}^k + \bar{\rho}\, h(x^k) = (1 + s_k^2)/s_k^2 \to \infty$, which contradicts $\bar{\lambda}^k \leq \lambda_{\max}$. Therefore, $\rho_k \to \infty$. Since $\bar{\lambda}^k \geq 0$ and $s_k < 1$, (9) implies that $\rho_k s_k^6/2 \leq 2/s_k^2$, i.e., $s_k \leq (4/\rho_k)^{1/8} \to 0$. Hence, $x^k \to 0$, and $h(0) = 0$.

Item 3. For all $t \in [0, 1]$, $(0, t, 0) \in X$. Thus, $e^2 = (0, 1, 0)^T \in T_X(0)$ and $\nabla f(0)^T e^2 = -1/2 < 0$. □

At the limit point of this example, $\nabla h(0) = 0$, so that no constraint qualification holds. In Section 6.2 we show that the method may also fail under MPEC-LICQ, with the quadratic penalty method (counterexample of Proposition 6.4) or with inexact solutions of the subproblems (counterexample of Proposition 6.5). It may fail even with exact solutions of the subproblems and multipliers in the interior of the safeguarding box (counterexample of Proposition 6.6).

Thus, the iterates of Algorithm 4.1 may be strict local minimizers of the subproblems, which are B-stationary for the subproblems and are accepted by the Telescopic Method with any tolerance, and yet converge to a point that is not B-stationary for (2). (Whether the Telescopic Method started at x^{k-1} actually returns the point x^k of Proposition 4.4 depends on the algorithm that it uses for the iterations within a face, called Algorithm MIA in [8]. In [8], Algorithm MIA is the bound-constrained solver Gencan [10, 11], and the numerical results below show that, with Gencan, it does.) The same conclusion holds without safeguards ($\lambda_{\max} = +\infty$, i.e., $\bar{\lambda}^{k+1} = \lambda^{k+1}$). Indeed, if $\rho_k = \bar{\rho}$ for $k > k_1$, then $\bar{\lambda}^{k+1} = \bar{\lambda}^k + \bar{\rho} h(x^k)$, so that $\{\bar{\lambda}^k\}$ is bounded, because $\{h(x^k)\}$ converges to zero linearly, and the contradiction of Item 2 follows as before. Since no swap is a candidate in (b), the look-ahead strategy of Section 4.1 does not change the iterates either.

In the example, the multiplier estimates $\lambda^{k+1} = (1 + s_k^2)/s_k^2$ tend to infinity. In nonlinear programming, in the absence of constraint qualifications, augmented Lagrangian methods are known to generate sequences that satisfy sequential (approximate) optimality conditions, which hold at every local minimizer but are weaker than the KKT conditions [11]. Here, each x^k is exactly B-stationary for the minimization of $f + \lambda^{k+1} h$ on Ω (by Item 1, since $\nabla L_k(x^k) = \nabla f(x^k) + \lambda^{k+1} \nabla h(x^k)$) and $h(x^k) \to 0$, but the limit point is not B-stationary. As described at the beginning of this subsection, the swap that would detect the descent along x_2 passes through the point $\bar{y}^k$, which is far from feasible ($h(\bar{y}^k) = s_k^2/2$, whereas $h(x^k) = s_k^6/2$), and its penalty grows with ρ_k, since $L_k(\bar{y}^k) - L_k(x^k) \geq \rho_k [h(\bar{y}^k)^2 - h(x^k)^2]/2 = \rho_k s_k^4 (1 - s_k^8)/8$.

To illustrate this behavior numerically, we applied Algorithm 4.1 to problem (2) with the data (8), using the parameters of Algencan ($\tau = 0.5$, $\gamma = 10$, $\lambda_{\min} = -10^{20}$, $\lambda_{\max} = 10^{20}$), $\bar{\lambda}^1 = 0$, $\rho_1 = 10$, and $x^0 = (1, 0, 0.9)$. The subproblems were solved by our implementation of the Telescopic Method with Gencan, exact second derivatives and the default parameters of Section 7 of [8] (in particular, $\varepsilon = 10^{-8}$, $\delta = \eta_{\text{br}} = +\infty$, and look-ahead), started at x^{k-1}. Table 1 shows the results of the first 40 iterations. Every subproblem ended at an approximately B-stationary point of the subproblem, with no swaps, and $x_2^k = 0$ for all k. The iterates coincide, to about six significant digits, with the points of Proposition 4.4, computed independently by solving (9) with the same updates of $\bar{\lambda}^k$ and ρ_k. The last two columns show that the swap that would lead to the global minimizer becomes increasingly unprofitable. Its first step (setting x_1 to zero) increases L_k by an amount that grows with ρ_k, and at $\bar{y}^k$ the derivative of L_k with respect to x_2 is positive, so that x_2 is not a candidate to enter the face. With a feasibility tolerance of 10^{-8}, an implementation of the augmented Lagrangian method would stop at iteration 21, at $x^{21} \approx (0.049, 0, 0.049)$, with $f(x^{21}) \approx -4.3 \times 10^{-6}$, whereas $f(0, 1, 0) = -1/2$. Computing the points of Proposition 4.4 with the same updates in 120-digit arithmetic, we verified that the same behavior persists up to iteration 80, with $\rho_{80} = 10^{40}$, $x^{80} \approx (9.5 \times 10^{-6}, 0, 9.5 \times 10^{-6})$,

$L_{80}(\bar{y}^{80}) - L_{80}(x^{80}) \approx 1.0 \times 10^{19}$ and $\partial L_{80}(\bar{y}^{80})/\partial x_2 \approx 4.1 \times 10^{19}$, and that the conclusions are the same with $\lambda_{\max} = 10$.

k	ρ_k	λ^k	x_1^k	x_3^k	$h(x^k)$	$f(x^k)$	$L_k(\bar{y}^k) - L_k(x^k)$	$\partial L_k(\bar{y}^k)/\partial x_2$
1	10^1	0	1.556e0	8.784e-1	2.3e-1	−6.8e-1	1.7e0	2.5e0
2	10^1	2.30e0	1.079e0	7.143e-1	6.6e-2	−2.6e-1	1.3e0	2.0e0
4	10^2	3.35e0	6.863e-1	5.340e-1	1.2e-2	−7.3e-2	1.7e0	4.5e0
8	10^4	8.45e0	3.271e-1	3.001e-1	3.7e-4	−6.4e-3	1.1e1	4.1e1
12	10^6	2.46e1	1.736e-1	1.688e-1	1.2e-5	−6.2e-4	1.0e2	4.1e2
16	10^8	7.55e1	9.575e-2	9.490e-2	3.7e-7	−6.1e-5	1.0e3	4.1e3
20	10^{10}	2.37e2	5.352e-2	5.336e-2	1.2e-8	−6.1e-6	1.0e4	4.1e4
21	10^{10}	3.52e2	4.897e-2	4.885e-2	6.8e-9	−4.3e-6	7.1e3	2.8e4
24	10^{12}	7.46e2	3.004e-2	3.001e-2	3.7e-10	−6.1e-7	1.0e5	4.1e5
28	10^{14}	2.36e3	1.688e-2	1.688e-2	1.2e-11	−6.1e-8	1.0e6	4.1e6
32	10^{16}	7.45e3	9.491e-3	9.490e-3	3.7e-13	−6.1e-9	1.0e7	4.1e7
36	10^{18}	2.36e4	5.337e-3	5.336e-3	1.2e-14	−6.1e-10	1.0e8	4.1e8
40	10^{20}	7.45e4	3.001e-3	3.001e-3	3.7e-16	−6.1e-11	1.0e9	4.1e9

Table 1: Augmented Lagrangian method with telescopic subproblems applied to problem (2) with the data (8). In all the iterations, $x_2^k = 0$, the Telescopic Method performed no swaps, and $\bar{y}^k = (0, 0, x_3^k)$.

4.4 Unbounded penalty parameters: properties of the limit points

When the penalty parameters tend to infinity, Theorem 4.3 does not apply, and the counterexample of Section 4.3 shows that the feasible limit points of Algorithm 4.1 need not be B-stationary. In this subsection we show what can be said in this case. Infeasible limit points are stationary for the infeasibility measure on a face (Proposition 4.5). Feasible limit points at which PCC-LICQ holds are KKT points of the branch problem defined by a face and satisfy an additional condition, given by condition (c) of Algorithm 4.1, which, in the MPCC case, means M-stationarity (Theorem 4.6 and Corollary 4.7). Proposition 4.8 shows that, without condition (c), the limit points may be A-stationary and not M-stationary.

Augmented Lagrangian methods applied to MPCCs written as nonlinear programs converge, under MPCC-tailored constraint qualifications, to C-stationary points [20, 4]. M-stationarity has been obtained with second-order information in the subproblems [5] and, with first-order methods, by keeping the complementarity constraints in the subproblems [18, 21, 2]. In [2], the subproblems of the method of [18] are solved up to approximate M-stationarity, measured with the limiting normal cone of the complementarity set at the current iterate, and the feasible limit points are M-stationary under constraint qualifications weaker than MPEC-LICQ. For cardinality constraints, an augmented Lagrangian method that keeps the cardinality constraints in the subproblems was analyzed in [23]. Algorithm 4.1 also keeps the complementarity (and, in general, the cardinality) constraints in the subproblems. Conditions (a) and (b) alone do not imply approximate M-stationarity of the subproblems, because (a) concerns only the current face and the decrease guaranteed by the escaping procedure in (b), given by (7), is of order $1/\rho_k^2$, since the Lipschitz constant of ∇L_k grows like ρ_k. Condition (c) fills this gap.

Let $\Psi(x) = \frac{1}{2}\left(\|h(x)\|^2 + \|c(x)_+\|^2\right)$ be the infeasibility measure.

Proposition 4.5. *Let Assumption 3 hold and let $\{x^k\}$ be generated by Algorithm 4.1. If $\{\rho_k\}$ is bounded, every limit point is feasible. If $\rho_k \to \infty$, $\lim_{k\in K} x^k = x^*$ and $I^k = I$ for all $k \in K$,*

then $g_I^\Psi(x^*) = 0$, that is, x^* is stationary for the minimization of the infeasibility measure Ψ on the face F_I.

Proof. If $\{\rho_k\}$ is bounded, the feasibility of the limit points follows as in the first paragraph of the proof of Theorem 4.3, which only uses Step 2. Assume that $\rho_k \to \infty$, and let $\Phi_k = L_k/\rho_k$, so that

$$\nabla\Phi_k(x) = \frac{\nabla f(x)}{\rho_k} + \nabla h(x)\left(h(x) + \frac{\bar\lambda^k}{\rho_k}\right) + \nabla c(x)\left(c(x) + \frac{\bar\mu^k}{\rho_k}\right)_+ .$$

Since the safeguarded multipliers are bounded and f, h and c are continuously differentiable on the compact set $\mathcal{B}$, $\nabla\Phi_k \to \nabla\Psi$ uniformly on $\mathcal{B}$. Since, for an interval $[a,b]$, $x \in [a,b]$ and $d \in \mathbb{R}$, $|P_{[a,b]}(x - td) - x|$ is non-decreasing in $t \geq 0$, taking $t = 1/\rho_k \leq 1$ and $d = \partial L_k(x^k)/\partial x_{i,\nu}$, this gives $|g_{i,\nu}^{\Phi_k}(x^k)| \leq |g_{i,\nu}^{L_k}(x^k)|$, so that, by (a), $\|g_I^{\Phi_k}(x^k)\|_\infty \leq \varepsilon_k$ for $k \in K$ large. By the uniform convergence and the continuity of the projection, $g_I^\Psi(x^*) = 0$. □

Proposition 4.5 is the analogue, on a face, of the usual result for augmented Lagrangian methods at infeasible limit points. We do not know whether x^* is B-stationary for the minimization of Ψ on Ω. The argument of Theorem 4.3 does not apply, because, on the scale of Φ_k, the escaping procedure only guarantees a decrease of order $1/\rho_k$, while setting the leaving variable $x_{i,\nu}$ to zero may cost a quantity of order $x_{i,\nu}^2$, as in Proposition 4.4.

In the rest of this subsection, the inequality constraints are written with slack variables, as in Section 3, so that $q = 0$ and $\nabla L_k(x^k) = \nabla f(x^k) + J(x^k)^T\lambda^{k+1} = \nabla\ell_{\lambda^{k+1}}(x^k)$.

Theorem 4.6. *Let Assumption 3 hold, with $q = 0$, and let $\{x^k\}$ be generated by Algorithm 4.1. Let x^* be a feasible limit point at which PCC-LICQ holds, and let K be an infinite set and I a face index such that $\lim_{k\in K} x^k = x^*$ and $I^k = I$ for all $k \in K$. Then $\lim_{k\in K}\lambda^{k+1} = \lambda^*$, where λ^* is the unique solution of $\nabla_{\mathcal{F}} f(x^*) + J_{\mathcal{F}}(x^*)^T\lambda = 0$, with $\mathcal{F} = \mathcal{F}(x^*)$, and $g_I^{\ell^*}(x^*) = 0$, where $\ell^* = \ell_{\lambda^*}$, that is, x^* is a KKT point of the branch problem of minimizing f subject to $h(x) = 0$ and $x \in F_I$. Moreover, for every $\nu \in \mathcal{N}_c$, $i \in I_\nu$ with $x_{i,\nu}^* = 0$ and $j \notin I_\nu$,*

$$\partial_{i,\nu}\ell^*(x^*) = 0 \quad \text{or} \quad g_{j,\nu}^{\ell^*}(x^*) = 0. \tag{10}$$

Proof. Let $v \in \mathcal{F}$. Since $x^* \in F_I$ and $x_v^* \neq 0$ if v belongs to a block of $\mathcal{N}_c$, the variable x_v is not fixed at zero in F_I. Since $a_v < x_v^* < b_v$, for $k \in K$ large enough x_v^k is at distance larger than ε_k from its bounds, so that, by (a) and (4), $|\partial_v L_k(x^k)| \leq \varepsilon_k$. Therefore, $\lim_{k\in K}[\nabla_{\mathcal{F}} f(x^k) + J_{\mathcal{F}}(x^k)^T\lambda^{k+1}] = 0$. By PCC-LICQ, $J_{\mathcal{F}}(x^k)$ has rank m for $k \in K$ large enough and $J_{\mathcal{F}}(x^k) \to J_{\mathcal{F}}(x^*)$, so that $\{\lambda^{k+1}\}_{k\in K}$ converges to the unique solution λ^* of $\nabla_{\mathcal{F}} f(x^*) + J_{\mathcal{F}}(x^*)^T\lambda = 0$. Hence, $\lim_{k\in K}\nabla L_k(x^k) = \nabla\ell^*(x^*)$ and, by (a) and the continuity of the projection, $g_I^{\ell^*}(x^*) = 0$.

Let $\nu \in \mathcal{N}_c$, $i \in I_\nu$ with $x_{i,\nu}^* = 0$, and $j \notin I_\nu$. If $|x_{i,\nu}^k| > \min\{\varepsilon_k, \bar\varepsilon\}$ for infinitely many $k \in K$, then, for these k large enough, $x_{i,\nu}^k$ is at distance larger than ε_k from its bounds, because the null bound, if any, is at distance $|x_{i,\nu}^k| > \varepsilon_k$ and the non-null bounds are far from $x_{i,\nu}^k$, which tends to zero. As in the first part of the proof, $|\partial_{i,\nu}L_k(x^k)| \leq \varepsilon_k$ for these k, and $\partial_{i,\nu}\ell^*(x^*) = 0$. Otherwise, $|x_{i,\nu}^k| \leq \min\{\varepsilon_k, \bar\varepsilon\}$ for all $k \in K$ large enough and, by (c), one of the two inequalities of (c) holds for infinitely many $k \in K$. Taking limits, by the continuity of the projection, one of the two equalities of (10) holds. □

The penalty parameters play no role in Theorem 4.6, which also holds when $\{\rho_k\}$ is bounded, but then Theorem 4.3 gives B-stationarity. Condition (b) is not used in the proof. In the MPCC case, we use the usual terminology for a feasible point x^* with a multiplier λ^* such that the partial derivatives of ℓ^* satisfy the sign conditions at the bounds for all the variables that are

not null components of complementarity pairs. There is no condition on a null component whose partner is non-null. For a biactive pair (x_a, x_b), x^* is *A-stationary* if $\partial_a \ell^*(x^*) \geq 0$ or $\partial_b \ell^*(x^*) \geq 0$, and *M-stationary* if both derivatives are non-negative or one of them is null. For partitioned cardinality constraints, (10) is the natural extension of M-stationarity. It is stronger than the M-stationarity of the complementarity reformulation of cardinality constraints [12, 23], which only requires the derivatives with respect to the non-null variables to vanish.

Corollary 4.7. *In the MPCC case, under the assumptions of Theorem 4.6, x^* is M-stationary with multiplier λ^*.*

Proof. For the variables that are not null components of pairs, the sign conditions follow from $g_I^{\ell^*}(x^*) = 0$, because these variables are not fixed at zero in F_I. Let (x_a, x_b) be a biactive pair with $a \in I_\nu$ and $b \notin I_\nu$. Since $x_a^* = 0$ is the lower bound of x_a, $g_a^{\ell^*}(x^*) = 0$ gives $\partial_a \ell^*(x^*) \geq 0$. By (10), either $\partial_a \ell^*(x^*) = 0$ or $g_b^{\ell^*}(x^*) = 0$, which means $\partial_b \ell^*(x^*) \geq 0$. In both cases, the pair satisfies the condition of M-stationarity. □

In Example 1, $\nabla h(0) = 0$, so that PCC-LICQ fails at $x^* = 0$ and $\lambda^{k+1} \to \infty$. At the limit points of the counterexamples of Propositions 6.4, 6.5 and 6.6 (Section 6.2), which satisfy MPEC-LICQ, $\lambda^* = 0$ and the derivatives of ℓ^* with respect to the biactive pair are 0 and $-1/4$, so that these points are M-stationary, as stated by Corollary 4.7, but not S-stationary. The following proposition shows that, without condition (c), the feasible limit points of Algorithm 4.1 may be A-stationary and not M-stationary, even under MPEC-LICQ and with the multipliers in the interior of the safeguarding box.

Proposition 4.8. *Let $n = 3$, let (x_1, x_2) be a complementarity pair, let x_3 form a block without cardinality constraints, $\mathcal{B} = [0,1]^2 \times [-1,1]$, $h(x) = x_3 - x_2$, and $f(x) = x_1 - x_2 + x_3^2/2$. Consider Algorithm 4.1 without condition (c), with $\tau = 1/2$, $\gamma = 10$, $\rho_1 \geq 10$, $\lambda_{\min} = -1$, $\lambda_{\max} = 1$, $\bar\lambda^1 = 1$, $\bar\lambda^{k+1} = P_{[\lambda_{\min}, \lambda_{\max}]}(\lambda^{k+1})$, arbitrary δ and η_{br}, and $\varepsilon_k = 1/k$. For $k \geq 1$, let $x^k = (0, 0, s_k)$, where $s_k = (1/(2k) - \bar\lambda^k)/\rho_k$, and let I^k be the face index in which x_1 is not fixed at zero. Then x^k satisfies conditions (a) and (b) of Algorithm 4.1, $\bar\lambda^{k+1} = \lambda^{k+1} = 1/(2k) \to 0$, which belongs to the interior of the safeguarding box, $\rho_k \to \infty$, and $x^k \to 0$, which satisfies MPEC-LICQ and is A-stationary, but not M-stationary.*

Proof. For $k = 1$, $s_1 = -1/(2\rho_1)$ and $\lambda^2 = \bar\lambda^1 + \rho_1 s_1 = 1/2$. If $\bar\lambda^k = 1/(2(k-1))$ for some $k \geq 2$, then $h(x^k) = s_k = -1/(2k(k-1)\rho_k)$ and $\lambda^{k+1} = \bar\lambda^k + \rho_k s_k = 1/(2k)$, so that, by induction, $\bar\lambda^{k+1} = \lambda^{k+1} = 1/(2k)$ for all k. Moreover, $|s_k| \leq 1/(2\rho_1) < 1$.

Condition (a). Since $x_1^k = 0$ and $\partial_1 L_k(x^k) = 1$, $g_1^{L_k}(x^k) = 0$. Also, $\partial_3 L_k(x^k) = s_k + \bar\lambda^k + \rho_k s_k = s_k + \lambda^{k+1}$, whose absolute value is at most $1/2 \leq \varepsilon_1$ for $k = 1$ and at most $1/(2k) + 1/(2k(k-1)\rho_1) \leq 1/k = \varepsilon_k$ for $k \geq 2$. Since $|g_3^{L_k}(x^k)| \leq |\partial_3 L_k(x^k)|$, (a) holds.

Penalty parameters. We have $M_k = |s_k|$, $|s_2|/|s_1| = 1/2$ and $|s_3|/|s_2| = 1/3$, so that $\rho_4 = \rho_3 = \rho_2 = \rho_1$. For $k \geq 4$, let $r_k = |s_k|/|s_{k-1}| = (k-2)\rho_{k-1}/(k\rho_k)$. Then $r_4 = 1/2$, so that $\rho_5 = \rho_1$ and, for $k \geq 5$, if $\rho_k = \rho_{k-1}$, then $r_k = (k-2)/k > 1/2$ and $\rho_{k+1} = 10\rho_k$, whereas, if $\rho_k = 10\rho_{k-1}$, then $r_k < 1/2$ and $\rho_{k+1} = \rho_k$. Hence, $\rho_k = \rho_1 10^{\lfloor (k-4)/2 \rfloor}$ for $k \geq 4$, and $\rho_k \to \infty$. In particular, $\rho_k \geq 9k/8$ for all k.

Condition (b). The only swap is $x_1 \to x_2$ and, since $x_1^k = 0$, $\bar y = x^k$. The function $\psi(t) = L_k(0, t, s_k)$ is a quadratic with $\psi''(t) = \rho_k$ and $\psi'(0) = -1 - \bar\lambda^k - \rho_k s_k = -1 - \lambda^{k+1}$. Therefore, for all t, $\psi(t) \geq \psi(0) - (1 + \lambda^{k+1})^2/(2\rho_k) \geq L_k(x^k) - 9/(8\rho_k) \geq L_k(x^k) - \varepsilon_k$, and (b) holds.

The limit point. Clearly, $x^k \to 0$, which is feasible. Since $\mathcal{F}(0) = \{3\}$ and $\partial_3 h = 1$, MPEC-LICQ holds at 0, and $\lambda^* = -\partial_3 f(0) = 0$. Then $\partial_1 \ell^*(0) = 1$ and $\partial_2 \ell^*(0) = -1$, so that 0 is

A-stationary, but not M-stationary. It is not B-stationary either, because $(0, t, t)$ is feasible for $t \in [0, 1]$ and $f(0, t, t) = -t + t^2/2$. □

In Proposition 4.8, condition (c) fails at x^k for $i = 1$ and $j = 2$, because $|\partial_1 L_k(x^k)| = 1$ and $|g_2^{L_k}(x^k)| = 1$. The swap $x_1 \to x_2$ decreases L_k by at most $9/(8\rho_k)$, which is smaller than ε_k because $\rho_k \varepsilon_k \to \infty$. Thus, a decrease test with a tolerance that does not depend on ρ_k, such as (b), cannot replace (c).

5 Augmented Lagrangian with constraint-preserving swaps

In the Telescopic Method, a swap sets a nonzero variable of a constrained block to zero (the leaving variable) and lets a null variable of the same block move (the entering variable). In the augmented Lagrangian method, the swaps are applied to the minimization of L_k, and a swap usually changes the value of h. When ρ_k is large, the corresponding change of the penalty term of L_k may hide a decrease of f, and the swap is rejected although it leads to a better branch of the feasible set. This is what happens in the counterexample of Section 4.3 and in Proposition 6.4 (see Figure 2). The aim of this section is to define a swap that does not change h. A *constraint-preserving swap* (CPS) performs the same exchange of variables and then corrects the other free variables of the face, so that the resulting point w satisfies $h(w) = h(y)$, where y is the current point. Then, $L_k(w) - L_k(y) = f(w) - f(y)$ for all values of the penalty parameter and of the multipliers, and the swap is accepted or rejected according to the variation of f alone. The correction is computed by a Newton-type method (Algorithm 5.1), which, under PCC-LICQ, converges and moves the variables by an amount proportional to the change of h produced by the exchange (Lemma 5.1). The CPS itself is Algorithm 5.2. In practice, $h(w)$ is computed with a tolerance, and $L_k(w) - L_k(y)$ differs from $f(w) - f(y)$ by a controlled amount (see (12) below). The augmented Lagrangian method that uses the constraint-preserving swaps (Algorithm 5.3) is defined in Section 5.2.

5.1 Scaled Newton restoration and constraint-preserving swaps

Algorithm 5.1 below solves the system $h(z) = \eta$, which, in general, has more unknowns (the variables of S) than equations, starting from a point z^0. Among its solutions, we want one close to z^0, so that the swap is changed as little as possible. Each iteration is a Newton step, in which h is linearized at z^j and, among the solutions Δ_S of $J_S(z^j)\Delta_S = -r^j$, the one that minimizes the scaled norm $\sum_{v \in S} \Delta_v^2/d_v^2$ (with $\Delta_v = 0$ if $d_v = 0$) is taken. The weights are the (truncated) distances of the variables to their bounds, that is, for $x \in \mathcal{B}$ and $v = 1, \ldots, n$,

$$d_v(x) = \min\{1, x_v - a_v, b_v - x_v\} \geq 0,$$

and $|d_v(x) - d_v(y)| \leq |x_v - y_v|$. The Newton step, $\Delta_S = -D^2 J_S(z^j)^T M_j^{-1} r^j$, is a linear combination of the scaled gradients $D^2 \nabla_S h_i(z^j)$, $i = 1, \ldots, m$, so that the variables at their bounds are not moved and those close to them move little. In Algorithm 5.2, z^0 is the point obtained from the current point y by the exchange of variables and $\eta = h(y)$, so that $h(w) = \eta$ means that the swap preserves the values of the constraints. Figure 1 illustrates the iteration.

Algorithm 5.1 (Scaled Newton restoration)
Let $z^0 \in \mathcal{B}$, $S \subseteq \{1, \ldots, n\}$, $D = \operatorname{diag}(d_v)_{v \in S}$ with $d_v \in [0, 1]$, $\eta \in \mathbb{R}^m$, a tolerance $\zeta > 0$ and a maximum number of iterations $j_{\max} \in \{1, 2, \ldots\}$ be given. For $j = 0, 1, 2, \ldots$:

Step 1. Let $r^j = h(z^j) - \eta$. If $\|r^j\| \leq \zeta$, stop and return $w = z^j$. If $j = j_{\max}$, stop without returning a point.

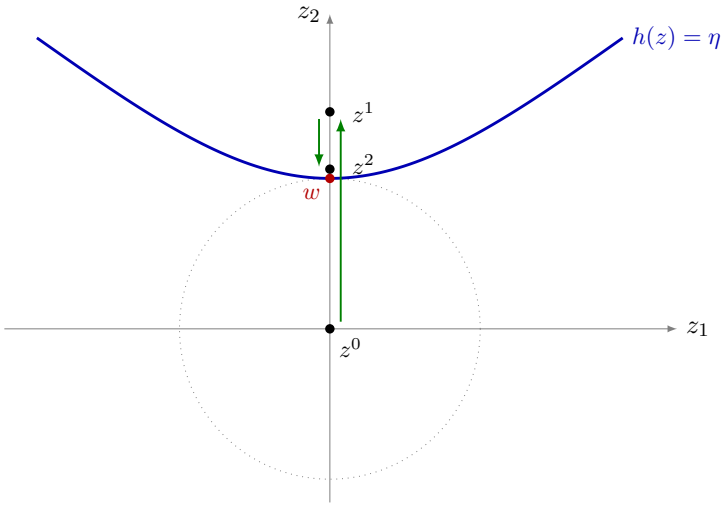


Figure 1: Algorithm 5.1 with $S = \{1, 2\}$, $D = I$, $h(z) = e^{z_2} - z_1^2$, $\eta = 2$ and $z^0 = (0, 0)$. The iterates are $z^1 = (0, 1)$, $z^2 \approx (0, 0.736)$, $z^3 \approx (0, 0.694)$, ..., and converge to $w = (0, \ln 2)$, which, in this example, is the solution of $h(z) = \eta$ closest to z^0 (dotted circle).

Step 2. Let $M_j = J_S(z^j) D^2 J_S(z^j)^T$. If M_j is singular, stop without returning a point.

Step 3. Define z^{j+1} by $z_S^{j+1} = z_S^j - D^2 J_S(z^j)^T M_j^{-1} r^j$ and $z_v^{j+1} = z_v^j$ for $v \notin S$. If $z^{j+1} \notin \mathcal{B}$, stop without returning a point.

The next lemma shows that, near a point x^* at which PCC-LICQ holds, Algorithm 5.1 behaves as Newton's method near a regular solution. If z^0 is close to x^* and the initial residual $\|h(z^0) - \eta\|$ is small, the algorithm is well defined, the iterates remain in the box, and the residual decreases quadratically, and at least by half at each iteration. Therefore, a point w with $h(w) = \eta$ (with tolerance ζ) is found within a number of iterations that can be computed in advance. Moreover, w is close to z^0, and the variables close to their bounds move very little. The lemma does not assume that $h(z) = \eta$ has a solution. As in Kantorovich's theorem for Newton's method (see [13, Section 5.3, Theorem 5.3.1] and [31], and, for Newton's method with generalized inverses for systems with more unknowns than equations, [7]), the existence of a solution follows from the nonsingularity of the matrix near x^* and from the smallness of the initial residual.

Lemma 5.1. *Let Assumption 1 hold and let $x^* \in \mathcal{B}$ be such that $J_{\mathcal{F}}(x^*)$ has rank m, where $\mathcal{F} = \mathcal{F}(x^*)$. Then there exist $r > 0$, $\bar{r} \in (0, 1]$ and $C > 0$ with the following property. Let $\hat{x} \in \mathcal{B}$ with $\|\hat{x} - x^*\| \le r$, $S \subseteq \{1, \ldots, n\}$ with $\mathcal{F} \subseteq S$, $d_v = d_v(\hat{x})$ for $v \in S$, $z^0 \in \mathcal{B}$ with $\|z^0 - x^*\| \le r$ and $z_S^0 = \hat{x}_S$, and $\eta \in \mathbb{R}^m$ with $\|h(z^0) - \eta\| \le \bar{r}$. Then Algorithm 5.1 is well defined and, while it does not stop, $\|r^{j+1}\| \le C\|r^j\|^2$ and $\|r^j\| \le 2^{-j}\|r^0\|$. If $j_{\max} \ge \log_2(1/\zeta)$, it returns a point w that satisfies $w \in \mathcal{B}$, $\|h(w) - \eta\| \le \zeta$, $w_v = z_v^0$ for $v \notin S$,*

$$|w_v - z_v^0| \le C\, d_v(\hat{x})^2\, \|h(z^0) - \eta\| \text{ for } v \in S, \quad \text{and} \quad \|w - z^0\| \le C\|h(z^0) - \eta\|.$$

The constants r, $\bar{r}$ and C depend only on x^* and on the problem (through σ, $\bar{d}$, L_J and G_J, defined in the proof), and not on $\hat{x}$, z^0 or η.

Proof. Let $\sigma > 0$ be the smallest singular value of $J_{\mathcal{F}}(x^*)$, $\bar{d} = \min_{v \in \mathcal{F}} d_v(x^*) > 0$, L_J a Lipschitz constant of J on $\mathcal{B}$, and $G_J = \max_{x \in \mathcal{B}} \|J(x)\|$. Let $r = \min\{\bar{d}/2, \sigma/(6L_J)\}$. If $\|\hat{x} - x^*\| \le r$,

then $d_v(\hat{x}) \geq \bar{d}/2$ for $v \in \mathcal{F}$, and, if $z \in \mathcal{B}$ and $\|z - x^*\| \leq 3r$, the smallest singular value of $J_{\mathcal{F}}(z)$ is at least $\sigma/2$. For such z, since $\mathcal{F} \subseteq S$,

$$J_S(z)D^2J_S(z)^T = \sum_{v \in S} d_v^2 J_v(z)J_v(z)^T \succeq \frac{\bar{d}^2}{4} J_{\mathcal{F}}(z)J_{\mathcal{F}}(z)^T \succeq \left(\frac{\bar{d}\sigma}{4}\right)^2 I.$$

Hence the matrix is nonsingular and the norm of its inverse is at most $c_1 = (4/(\bar{d}\sigma))^2$. The step $\Delta^j = z^{j+1} - z^j$ satisfies, for $v \in S$, $|\Delta_v^j| = d_v^2 |J_v(z^j)^T M_j^{-1} r^j| \leq c_2 d_v^2 \|r^j\|$ and, since $d_v \leq 1$, $\|\Delta^j\| \leq c_2\|r^j\|$, where $c_2 = G_J c_1$. Since $J(z^j)\Delta^j = J_S(z^j)\Delta_S^j = -r^j$, if the segment $[z^j, z^{j+1}]$ is contained in $\mathcal{B}$, then

$$\|r^{j+1}\| = \|h(z^j + \Delta^j) - h(z^j) - J(z^j)\Delta^j\| \leq \frac{L_J}{2}\|\Delta^j\|^2 \leq c_3\|r^j\|^2, \quad c_3 = \frac{L_J c_2^2}{2}.$$

Let $\bar{r} = \min\{1, 1/(2c_3), r/(2c_2), 1/(2c_2)\}$ and $C = \max\{2c_2, c_3\}$. We prove by induction that, while the algorithm does not stop, $z^j \in \mathcal{B}$, $\|z^j - x^*\| \leq 2r$, $\|r^j\| \leq 2^{-j}\|r^0\|$ and $|z_v^j - z_v^0| \leq 2c_2 d_v^2 (1 - 2^{-j})\|r^0\|$ for $v \in S$. This is true for $j = 0$. If it is true for j, then M_j is nonsingular, $|z_v^{j+1} - z_v^0| \leq 2c_2 d_v^2 (1 - 2^{-j})\|r^0\| + c_2 d_v^2 2^{-j}\|r^0\| = 2c_2 d_v^2 (1 - 2^{-j-1})\|r^0\|$, and $\|z^{j+1} - z^0\| \leq 2c_2\|r^0\| \leq r$, so that $\|z^{j+1} - x^*\| \leq 2r$. Since $2c_2\|r^0\| \leq 1$ and $d_v \leq 1$, we have that $|z_v^{j+1} - z_v^0| \leq d_v$, and $z_v^0 = \hat{x}_v$ is at distance at least $d_v = d_v(\hat{x})$ from its bounds, so that $z^{j+1} \in \mathcal{B}$ and, since $\mathcal{B}$ is convex, the segment $[z^j, z^{j+1}]$ is contained in $\mathcal{B}$. Thus, $\|r^{j+1}\| \leq c_3\|r^j\|^2 \leq c_3 2^{-j}\|r^0\|\,\|r^j\| \leq \|r^j\|/2$, which completes the induction. Since $\|r^0\| \leq \bar{r} \leq 1$, the test of Step 1 holds at the latest at the first $j \geq \log_2(1/\zeta)$, so the algorithm returns a point, and the bounds of the statement follow from the induction. □

As stated by Lemma 5.1, the convergence of Algorithm 5.1 is quadratic. However, the constant C is not known, and the computable bound $j_{\max} \geq \log_2(1/\zeta)$ on the number of iterations, which is used in Algorithm 5.3, comes from the linear bound $\|r^j\| \leq 2^{-j}\|r^0\|$. Lemma 5.1 also holds, with a linear rate, for the simplified version of Algorithm 5.1 in which the matrix is not recomputed, that is, M_j is replaced by $M = J_S(\hat{x})D^2J_S(\hat{x})^T$ in Step 2 and $J_S(z^j)$ is replaced by $J_S(\hat{x})$ in Step 3. (In Algorithm 5.2, $\hat{x}$ is the point y at which the swap is performed, and this is, up to a small regularization, the version used in our implementation, described in Section 7.2.) In the proof, since $\|\hat{x} - x^*\| \leq r$, the matrix M is nonsingular with $\|M^{-1}\| \leq c_1$, and $|\Delta_v^j| \leq c_2 d_v^2 \|r^j\|$ as before. Since now $J_S(\hat{x})\Delta_S^j = -r^j$, the bound on the residual becomes

$$\|r^{j+1}\| \leq \frac{L_J}{2}\|\Delta^j\|^2 + L_J\|z^j - \hat{x}\|\,\|\Delta^j\| \leq \left(c_3\|r^j\| + L_J c_2\|z^j - \hat{x}\|\right)\|r^j\|.$$

By the induction hypothesis, $\|r^j\| \leq \|r^0\|$ and $\|z^j - \hat{x}\| \leq \|z^0 - \hat{x}\| + 2c_2\|r^0\|$. Since $L_J c_2^2 = 2c_3$, it follows that

$$\|r^{j+1}\| \leq q\|r^j\|, \quad q = 5c_3\|r^0\| + L_J c_2\|z^0 - \hat{x}\|.$$

Since $\|z^0 - \hat{x}\| \leq 2r$, if r also satisfies $r \leq 1/(12L_J c_2)$ and $\bar{r}$ also satisfies $\bar{r} \leq 1/(15c_3)$, then $q \leq 1/2$, and the rest of the proof is unchanged. Thus, the convergence is linear, with a ratio q that tends to zero when r^0 tends to zero and z^0 approaches $\hat{x}$ (in Algorithm 5.2, $\|z^0 - y\| \leq |y_{i,\nu}| + t$). The bound $j_{\max} \geq \log_2(1/\zeta)$ and the results of Section 6 are not affected.

The constraint-preserving swap is defined as follows. Let $y \in F_I$, $\nu \in \mathcal{N}_c$, $i \in I_\nu$ (the leaving variable) and $j \notin I_\nu$ (the entering variable, so that $y_{j,\nu} = 0$). Let J be the face index I with I_ν replaced by $(I_\nu \setminus \{i\}) \cup \{j\}$, and let

$$S = \{(u, \omega) \mid u \in I_\omega, \text{ and } y_{u,\omega} \neq 0 \text{ or } \omega \notin \mathcal{N}_c\} \setminus \{(i, \nu)\}$$

be the set of indices of the variables that are not fixed at zero in F_I, other than the leaving one, excluding the null variables of the constrained blocks (which, therefore, remain null). The entering variable moves in a direction $\varsigma \in \{-1, +1\}$, which is *admissible* if $\varsigma = 1$ and $b_{j,\nu} > 0$, or $\varsigma = -1$ and $a_{j,\nu} < 0$. (In the MPCC case, only $\varsigma = 1$ is admissible.)

In words, Algorithm 5.2 starts from the swap of the Telescopic Method, in which the leaving variable is set to zero and the entering variable is moved by a step t, which gives the point z^0. This point usually changes the values of the constraints, and Algorithm 5.1 corrects it by moving the other free variables, so that the values of the constraints at the current point are recovered. Then, a sufficient decrease condition is tested on the merit function Φ, which, in Algorithm 5.3, is the augmented Lagrangian. If the correction fails or the decrease is not sufficient, the step of the entering variable is halved (backtracking), which modifies z^0, and the process is repeated.

Algorithm 5.2 (Constraint-preserving swap, CPS)
Let y, ν, i, j, J, S, and an admissible ς be as above, let Φ be a merit function, and let $t_{\max} > 0$, $t_{\min} \in (0, t_{\max}]$, $\alpha > 0$, $\zeta > 0$ and $j_{\max}$ be given. Let $\eta = h(y)$, $d_v = d_v(y)$ for $v \in S$, and $t = \min\{t_{\max}, b_{j,\nu}\}$ if $\varsigma = 1$ or $t = \min\{t_{\max}, -a_{j,\nu}\}$ if $\varsigma = -1$.

Step 1. Let $z^0 = y - y_{i,\nu} e^{i,\nu} + \varsigma t e^{j,\nu}$. Apply Algorithm 5.1 with z^0, S, $D = \operatorname{diag}(d_v)$, η, ζ and $j_{\max}$. If it returns a point w and $\Phi(w) \le \Phi(y) - \alpha t^2$, stop and return w.

Step 2. Set $t \leftarrow t/2$. If $t < t_{\min}$, stop without returning a point (the swap is discarded). Otherwise, go to Step 1.

If Algorithm 5.2 returns a point w, then $w \in F_J \subseteq \Omega$, because $w_{i,\nu} = 0$, $w_{j,\nu} = \varsigma t \in [a_{j,\nu}, b_{j,\nu}]$, the variables fixed at zero in F_I are not moved, and the variables of S remain in the box. Moreover, $\|h(w) - h(y)\| \le \zeta$. For the augmented Lagrangian $\Phi(x) = f(x) + (\rho/2)\|h(x) + \bar\lambda/\rho\|^2$, writing $\mu = \bar\lambda + \rho h(y)$ and $e = h(w) - h(y)$, we have the identity

$$\Phi(w) - \Phi(y) = f(w) - f(y) + \mu^T e + \frac{\rho}{2}\|e\|^2.$$

Given $\varepsilon > 0$, let

$$\zeta(y) = \min\left\{ \frac{\varepsilon}{4(1 + \|\mu\|)},\ \sqrt{\frac{\varepsilon}{2\rho}} \right\}. \tag{11}$$

If the tolerance of Algorithm 5.1 satisfies $\zeta \le \zeta(y)$, then

$$|\Phi(w) - \Phi(y) - (f(w) - f(y))| \le \frac{\varepsilon}{2}. \tag{12}$$

Thus, up to $\varepsilon/2$, the penalty parameter and the multipliers disappear from the swap test. Note that $\zeta(y)$ is computable at y. Algorithm 5.2 performs at most $1 + \log_2(t_{\max}/t_{\min})$ restorations.

5.2 The modified augmented Lagrangian method

Algorithm 5.3 is the safeguarded augmented Lagrangian method of Algorithm 4.1 applied to problem (3), with the same updates of the multipliers and of the penalty parameter. The only difference is in the stopping criterion of the subproblem. Condition (b) of Step 1 concerns the constraint-preserving swaps of Algorithm 5.2 instead of the original swaps, so that the sufficient decrease is tested at points at which the constraints have, up to a tolerance, the same values as at the current point. Condition (c) is not required.

Algorithm 5.3 (Augmented Lagrangian with telescopic subproblems and CPS)
Let $\tau \in (0, 1)$, $\gamma > 1$, $\rho_1 > 0$, $\lambda_{\min} \le \lambda_{\max}$, $\bar\lambda^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\delta > 0$, $t_{\max} > 0$, $\alpha > 0$, sequences $\varepsilon_k \downarrow 0$ and $t_{\min,k} \downarrow 0$ with $t_{\min,k} \le t_{\max}$, and $x^0 \in \Omega$ be given. Set $k \leftarrow 1$.

Step 1 (Subproblem). Let $L_k(x) = f(x) + (\rho_k/2)\|h(x) + \bar{\lambda}^k/\rho_k\|^2$. Starting from x^{k-1}, compute $x^k \in \Omega$ and a face index I^k with $x^k \in F_{I^k}$ such that

(a) $\|g_{I^k}^{L_k}(x^k)\|_\infty \le \varepsilon_k$, and

(b) no CPS decreases L_k significantly from x^k: for every $\nu \in \mathcal{N}_c$, $i \in I_\nu^k$ with $|x_{i,\nu}^k| \le \delta$, $j \notin I_\nu^k$ and admissible ς, Algorithm 5.2 applied with $y = x^k$, $\Phi = L_k$, $t_{\min} = t_{\min,k}$, tolerance $\zeta = \zeta_k(x^k) > 0$, where $\zeta_k(x^k)$ is (11) with $y = x^k$, $\varepsilon = \varepsilon_k$, $\rho = \rho_k$ and $\bar{\lambda} = \bar{\lambda}^k$, and $j_{\max} = \max\{1, \lceil \log_2(1/\zeta_k(x^k)) \rceil\}$ (or any larger integer), either returns no point or returns a point w with $L_k(w) \ge L_k(x^k) - \varepsilon_k$.

Step 2 (Penalty parameter). If $k > 1$ and $\|h(x^k)\|_\infty > \tau\|h(x^{k-1})\|_\infty$, set $\rho_{k+1} = \gamma\rho_k$. Otherwise, set $\rho_{k+1} = \rho_k$.

Step 3 (Multipliers). Set $\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k)$ and choose $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ (for example, $P_{[\lambda_{\min},\lambda_{\max}]^m}(\lambda^{k+1})$, used in our implementation). Set $k \leftarrow k+1$ and go to Step 1.

All the inner processes of Step 1 are finite. Algorithm 5.1 performs at most $j_{\max}$ iterations, Algorithm 5.2 at most $1 + \log_2(t_{\max}/t_{\min,k})$ restorations, condition (b) involves finitely many swaps, and, by Lemma 5.2, Step 1 is completed after finitely many iterations of the Telescopic Method. Conditions (a) and (b) are the stopping criterion of the subproblem, and the points w computed by Algorithm 5.2 in (b) serve only to test x^k, and they are not used in Steps 2 and 3. If some CPS returns w with $L_k(w) < L_k(x^k) - \varepsilon_k$, then x^k is not acceptable, and the method used at Step 1 must go on (in the implementation described below, from w or from a better candidate). Algorithm 5.3 is stated for problem (3), in which the inequality constraints are written with slack variables, because the constraint-preserving swaps keep the values of h. In our implementation, both methods use slack variables. With $\lambda_{\min} = \lambda_{\max} = 0$, Algorithm 5.3 is the quadratic penalty method. Since a CPS that returns a point w satisfies $L_k(w) \le L_k(x^k) - \alpha t^2$, condition (b) says that every CPS at x^k that returns a point has $\alpha t^2 < \varepsilon_k$. Note that $\mu = \lambda^{k+1}$ in (11) when $y = x^k$.

Algorithm 5.2 returns no point when, for every tried step $t \ge t_{\min,k}$, the restoration (Algorithm 5.1) returns no point or returns a point w with $L_k(w) > L_k(x^k) - \alpha t^2$. The restoration returns no point when M_j is singular, when an iterate leaves $\mathcal{B}$, or when the tolerance is not reached in $j_{\max}$ iterations. In this case, the swap is discarded, as in a line search that does not find a sufficient decrease, and x^k passes the test (b) for this swap. By Lemma 5.1, if x^k is close to a feasible point at which PCC-LICQ holds and $|x_{i,\nu}^k| + t$ is small, the restoration returns a point, and the swap is discarded only if the decrease is not sufficient. Since $L_k(w) - L_k(x^k) \approx f(w) - f(x^k)$ by (12), and $\varepsilon_k \to 0$, condition (b) says, asymptotically, that no CPS with a step t bounded away from zero decreases f by an amount bounded away from zero. This is what is used in Section 6. If a limit point x^* were not B-stationary, there would be a descent direction of f along an entering variable, and, for k large, the CPS along it with a fixed step would return a point with a decrease larger than ε_k, contradicting (b). Failures of the restoration for large steps do not affect this argument, because Algorithm 5.2 reduces t down to $t_{\min,k}$, which tends to zero, and therefore reaches the region in which Lemma 5.1 guarantees that the restoration returns a point.

Step 1 can be implemented by the Telescopic Method (Section 4.1) applied to the minimization of L_k on Ω, with the swap search of Step 1 of that algorithm extended by the CPS. In the notation of Section 4.1, the swap search is performed at the point y^j computed by the last iteration of Algorithm MIA (the point x^k of Section 4.1, with j in the place of k), in the face F_{I^j}. For every $\nu \in \mathcal{N}_c$, $i \in I_\nu^j$ with $|y_{i,\nu}^j| \le \delta$, $j' \notin I_\nu^j$ and admissible ς, Algorithm 5.2 is

applied with $y = y^j$ and $\zeta = \zeta_k(y^j)$. If it returns a point w with $L_k(w) < f_{best}$, then w becomes the best candidate, and the face of w becomes the candidate face. The point z^j from which the next iteration of Algorithm MIA starts (the point z^k of Section 4.1) is the best candidate, which belongs to its face and satisfies $L_k(z^j) \leq f_{best} \leq L_k(y^j)$. The convergence analysis of the Telescopic Method uses only these two properties of z^j. Since Algorithm MIA starts from z^j and does not increase L_k, the values $L_k(y^j)$ do not increase, and the iterations of Algorithm MIA performed on a fixed face behave, in the limit, as if no swaps had been performed between them (items 1 and 2 of Lemma 4.1). Therefore, the analysis is not affected when the CPS are added to the swap search. (The original swaps may be kept, although they are not needed in the analysis below.) The next lemma shows that, with this extended swap search, Step 1 is well defined.

Lemma 5.2. *Let Assumption 1 hold and assume that Algorithm MIA satisfies Assumption 2. For every k, the Telescopic Method with the extended swap search, applied to the minimization of L_k on Ω, finds after finitely many iterations a point y^j that satisfies (a) and (b).*

Proof. Since $\mathcal{B}$ is bounded, L_k satisfies the assumptions on ϕ of Section 4.1, and we can repeat the proof of Lemma 6.5 of [8], with the points returned by the CPS in addition to the points computed by the escaping procedure in the swap search of the Telescopic Method. If, at iteration j, the swap search is interrupted by the brute acceptance test, or a CPS returns a point w with $L_k(w) < L_k(y^j) - \varepsilon_k$, then $L_k(y^{j+1}) \leq L_k(z^j) \leq f_{best} \leq L_k(y^j) - \min\{\eta_{\rm br}, \varepsilon_k\}$, because Algorithm MIA does not increase L_k (Assumption 2). Since $\{L_k(y^j)\}$ is non-increasing and bounded below, this happens only for finitely many j. If $\|g^{L_k}_{I^j_j}(y^j)\|_\infty > \varepsilon_k$ for all j large enough, the argument of the proof of Lemma 6.5 of [8] (based on item 2 of Lemma 4.1) leads to a contradiction. Therefore, for some j, the search is not interrupted, $\|g^{L_k}_{I^j_j}(y^j)\|_\infty \leq \varepsilon_k$ and every CPS at y^j that returns a point returns w with $L_k(w) \geq L_k(y^j) - \varepsilon_k$, that is, y^j satisfies (a) and (b). □

Remark 5.3. The Telescopic Method may be used at Step 1 of Algorithms 4.1 and 5.3 because it computes, after finitely many iterations, points that satisfy conditions (a) and (b), and also (c) in Algorithm 4.1 (Lemmas 4.2 and 5.2). The convergence results of Sections 4 and 6 use only these conditions, so any method that guarantees them could be used instead. However, it is not enough to compute B-stationary points of the subproblems, because condition (b) is not a stationarity condition for the minimization of L_k on Ω, but a test on specific moves (the swaps of the Telescopic Method in Algorithm 4.1 and the CPS in Algorithm 5.3), as the counterexample of Proposition 4.4 shows.

6 Convergence and necessity of the modification

In this section we analyze the modified augmented Lagrangian method (Algorithm 5.3). In Section 6.1 we prove that its feasible limit points at which PCC-LICQ holds are B-stationary, whatever the behavior of the penalty parameters, and we consider infeasible limit points. In Section 6.2 we show, by means of counterexamples, that the modification is necessary, even under MPEC-LICQ, and that MPEC-LICQ cannot be replaced by MPEC-MFCQ.

6.1 B-stationarity of the limit points

In this subsection we prove the main result of the paper (Theorem 6.1), state its consequence for MPCCs (Corollary 6.2), and consider infeasible limit points (Proposition 4.5).

Theorem 6.1. *Let Assumption 1 hold, let $\{x^k\}$ be generated by Algorithm 5.3, and let x^* be a limit point of $\{x^k\}$ such that $h(x^*) = 0$ and PCC-LICQ holds at x^*. Let K be an infinite set such that $\lim_{k\in K} x^k = x^*$. Then $\lim_{k\in K} \lambda^{k+1} = \lambda^*$, where λ^* is the unique solution of $\nabla_{\mathcal{F}} f(x^*)+J_{\mathcal{F}}(x^*)^T\lambda = 0$, $\mathcal{F} = \mathcal{F}(x^*)$, and x^* is stationary for the minimization of $\ell^* = f+\lambda^{*T}h$ on Ω. Consequently, x^* is B-stationary for (3). No assumption is made on $\{\rho_k\}$.*

Proof. Since there are finitely many face indices, we may assume that $I^k = I$ for all $k \in K$, and then $x^* \in F_I$ because F_I is closed. Let $\bar{d} = \min_{v\in\mathcal{F}} d_v(x^*) > 0$.

Step 1: multipliers. We have that $\nabla L_k(x) = \nabla f(x)+J(x)^T(\bar{\lambda}^k+\rho_k h(x))$, so that $\nabla L_k(x^k) = \nabla f(x^k)+J(x^k)^T\lambda^{k+1}$. Let $v \in \mathcal{F}$. Then x_v is not fixed at zero in F_I (either $x^*_v \neq 0$ or v belongs to a block $\nu \notin \mathcal{N}_c$) and, for $k \in K$ large enough, x^k_v is at distance at least $\bar{d}/2$ from its bounds. Since $g^{L_k}_v(x^k) = P_{[a_v,b_v]}(x^k_v - \partial_v L_k(x^k)) - x^k_v$, we have that $|g^{L_k}_v(x^k)| = |\partial_v L_k(x^k)|$ if $|\partial_v L_k(x^k)| \leq \bar{d}/2$ and $|g^{L_k}_v(x^k)| \geq \bar{d}/2$ otherwise. By (a), for $k \in K$ large enough (with $\varepsilon_k < \bar{d}/2$), $|\partial_v L_k(x^k)| \leq \varepsilon_k$. Thus, $\nabla_{\mathcal{F}} f(x^k) + J_{\mathcal{F}}(x^k)^T\lambda^{k+1} \to 0$ for $k \in K$. If $\{\lambda^{k+1}\}_{k\in K}$ were unbounded, dividing by $\|\lambda^{k+1}\|$ and taking a convergent subsequence we would obtain a unit vector u with $J_{\mathcal{F}}(x^*)^T u = 0$, which contradicts PCC-LICQ. Therefore, $\{\lambda^{k+1}\}_{k\in K}$ is bounded and each of its limit points solves $\nabla_{\mathcal{F}} f(x^*)+J_{\mathcal{F}}(x^*)^T\lambda = 0$, whose unique solution is λ^*. Hence, $\lim_{k\in K}\lambda^{k+1} = \lambda^*$, $\lim_{k\in K}\nabla L_k(x^k) = \nabla\ell^*(x^*)$, and, by (a) and the continuity of the projection, $g^{\ell^*}_I(x^*) = 0$. In particular, $\partial_v\ell^*(x^*) = 0$ for every v such that x_v is not fixed at zero in F_I and $a_v < x^*_v < b_v$. These indices include $\mathcal{F}$ and the indices of the null variables of blocks of $\mathcal{N}_c$ that are not fixed at zero in F_I and whose bounds satisfy $a_v < 0 < b_v$.

Step 2: stationarity. As in the proof of item 3 of Lemma 4.1, $g^{\ell^*}_I(x^*) = 0$ implies condition (5) for ℓ^*, and condition (6) for the blocks without cardinality constraints. Let $\nu \in \mathcal{N}_c$ be such that $\|\vec{x}^*_\nu\|_0 < p_\nu$. As in the proof of item 3 of Lemma 4.1, there exists $i \in I_\nu$ such that $x^*_{i,\nu} = 0$, and it remains to prove that $g^{\ell^*}_{j,\nu}(x^*) = 0$ for all $j \notin I_\nu$. Assume, by contradiction, that $g^{\ell^*}_{j,\nu}(x^*) \neq 0$ for some $j \notin I_\nu$. Since $x^*_{j,\nu} = 0 \in [a_{j,\nu}, b_{j,\nu}]$, this means that $\kappa = |\partial_{j,\nu}\ell^*(x^*)| > 0$ and that $\varsigma = -\mathrm{sign}(\partial_{j,\nu}\ell^*(x^*))$ is admissible.

Step 3: the CPS returns a point with a large decrease. Let S_k be the set S defined before Algorithm 5.2 for the swap (ν, i, j) at $y = x^k$, that is, the indices of the variables that may be moved by the restoration, and let $S = S_k$. For $k \in K$ large enough, $\mathcal{F} \subseteq S_k$, because $x^*_{i,\nu} = 0$ and the variables x_v, $v \in \mathcal{F}$, that belong to blocks of $\mathcal{N}_c$ are non-null at x^* and, hence, at x^k. Let r, $\bar{r}$ and C be the constants of Lemma 5.1, G an upper bound of $\|\nabla\ell^*\|_\infty$ and of $\|J\|$ on $\mathcal{B}$, and L_ℓ a Lipschitz constant of $\nabla\ell^*$ on $\mathcal{B}$. Let t_0 be the initial value of t in Algorithm 5.2 for (ν, i, j, ς), which does not depend on k. For $k \in K$ and $t \in (0, t_0]$, let $z^0(t) = x^k - x^k_{i,\nu}e^{i,\nu} + \varsigma t e^{j,\nu}$ (recall that $x^k_{j,\nu} = 0$). Then $z^0(t)_S = x^k_S$, $\|z^0(t) - x^k\| \leq |x^k_{i,\nu}| + t$, and $\|h(z^0(t)) - h(x^k)\| \leq G(|x^k_{i,\nu}| + t)$. Let

$$\hat{t} = \min\left\{t_0,\ \frac{r}{2},\ \frac{\bar{r}}{2G},\ \frac{\kappa}{8L_\ell C_1^2},\ \frac{\kappa}{16\alpha}\right\}, \quad \text{where} \quad C_1 = 1 + CG.$$

Let $k \in K$ be large enough so that $\|x^k - x^*\| + |x^k_{i,\nu}| \leq r/2$, $G|x^k_{i,\nu}| \leq \bar{r}/2$, and $|x^k_{i,\nu}| \leq \delta$. Then, for all $t \in (0, \hat{t}]$, Lemma 5.1 applies with $\hat{x} = x^k$, $z^0 = z^0(t)$, $\eta = h(x^k)$, $\zeta = \zeta_k(x^k)$ and $j_{\max} \geq \log_2(1/\zeta)$: Algorithm 5.1 returns a point $w = w(t)$ that satisfies $\|h(w) - h(x^k)\| \leq \zeta_k(x^k) \leq \varepsilon_k/4$, $w_v = x^k_v$ for $v \notin S \cup \{(i,\nu),(j,\nu)\}$, $|w_v - x^k_v| \leq C d_v(x^k)^2 G(|x^k_{i,\nu}| + t)$ for $v \in S$, and $\|w - x^k\| \leq C_1(|x^k_{i,\nu}| + t)$.

We have that $f(w) - f(x^k) = \ell^*(w) - \ell^*(x^k) - \lambda^{*T}(h(w) - h(x^k)) \leq \ell^*(w) - \ell^*(x^k) + \|\lambda^*\|\varepsilon_k/4$.

By the Lipschitz continuity of $\nabla\ell^*$, $\ell^*(w) - \ell^*(x^k) \le \nabla\ell^*(x^k)^T(w - x^k) + (L_\ell/2)\|w - x^k\|^2$, and

$$\begin{aligned}\nabla\ell^*(x^k)^T(w - x^k) = & - x^k_{i,\nu}\,\partial_{i,\nu}\ell^*(x^k) + \varsigma t\,\partial_{j,\nu}\ell^*(x^k) \\ & + \sum_{v\in S_1} \partial_v\ell^*(x^k)(w_v - x^k_v) + \sum_{v\in S_2} \partial_v\ell^*(x^k)(w_v - x^k_v),\end{aligned}$$

where S_1 is the set of the variables $v \in S$ such that $a_v < x^*_v < b_v$ and $S_2 = S \setminus S_1$. The first term is at most $G|x^k_{i,\nu}|$. By Step 1 and continuity, for $k \in K$ large, $\varsigma\partial_{j,\nu}\ell^*(x^k) \le -\kappa/2$ and $|\partial_v\ell^*(x^k)| \le \omega_k$ for $v \in S_1$, with $\omega_k \to 0$. If $v \in S_2$, then x^*_v is at one of its bounds, so that $d_v(x^k) \le |x^k_v - x^*_v| \to 0$. Therefore, there exists $\omega'_k \to 0$ such that the last two sums are bounded by $\omega'_k(|x^k_{i,\nu}| + t)$. Using $\|w - x^k\|^2 \le 2C_1^2(|x^k_{i,\nu}|^2 + t^2)$ and (12), we obtain

$$L_k(w) - L_k(x^k) \le -\frac{\kappa}{2}t + \omega'_k t + L_\ell C_1^2 t^2 + \beta_k,$$

where $\beta_k = (G + \omega'_k)|x^k_{i,\nu}| + L_\ell C_1^2|x^k_{i,\nu}|^2 + \|\lambda^*\|\varepsilon_k/4 + \varepsilon_k/2 \to 0$. For $k \in K$ large enough, $\omega'_k \le \kappa/8$ and $\beta_k \le \kappa\hat{t}/32$. Then, for all $t \in [\hat{t}/2, \hat{t}]$, using $L_\ell C_1^2 t \le \kappa/8$ and $\beta_k \le \kappa t/16$,

$$L_k(w(t)) - L_k(x^k) \le -\frac{\kappa}{2}t + \frac{\kappa}{8}t + \frac{\kappa}{8}t + \frac{\kappa}{16}t = -\frac{3\kappa}{16}t \le -\alpha t^2,$$

because $\alpha t \le \alpha\hat{t} \le \kappa/16$.

Step 4: contradiction. For $k \in K$ large, $t_{\min,k} < \hat{t}/2$. Algorithm 5.2 applied with $y = x^k$ tries the values $t_0 2^{-q}$, $q = 0, 1, \ldots$, while $t_0 2^{-q} \ge t_{\min,k}$, and, since $t_0 \ge \hat{t}$, one of them belongs to $[\hat{t}/2, \hat{t}]$. Therefore, the algorithm returns a point with some $t \ge \hat{t}/2$, either at a value $t > \hat{t}$ or, at the latest, at the first value in $[\hat{t}/2, \hat{t}]$, by Step 3. The returned point satisfies $L_k(w) \le L_k(x^k) - \alpha t^2 \le L_k(x^k) - \alpha\hat{t}^2/4$. Since $\varepsilon_k < \alpha\hat{t}^2/4$ for k large, this contradicts (b). Hence, $g^{\ell^*}_{j,\nu}(x^*) = 0$ for all $j \notin I_\nu$, and x^* is stationary for the minimization of ℓ^* on Ω. By Proposition 3.4, x^* is B-stationary for (3). □

Corollary 6.2. *In the MPCC case, every limit point x^* of the sequence generated by Algorithm 5.3 such that $h(x^*) = 0$ and MPEC-LICQ holds at x^* is B-stationary and S-stationary, with multiplier $\lambda^* = \lim_{k\in K} \lambda^{k+1}$ for every infinite set K such that $\lim_{k\in K} x^k = x^*$.*

Remark 6.3.

1. The penalty parameters appear in the proof only through (12), which is guaranteed by the computable tolerance (11) of the restoration. The multiplier estimates λ^{k+1} converge by PCC-LICQ, independently of ρ_k and of the safeguarding box.

2. The scaling $d_v(x^k)^2$ is what makes the variables that tend to their bounds irrelevant in Step 3, because they are moved by $O(d_v(x^k)^2 t)$ and, therefore, their (possibly non-null) partial derivatives do not spoil the decrease. Without scaling, the restoration could move a variable that tends to a bound at which the derivative of ℓ^* is positive, or take it out of the box. Variables of constrained blocks that tend to zero in the interior of their bounds may be moved by $O(t)$, but the derivative of ℓ^* with respect to them vanishes at x^*.

3. The tolerance (11) behaves like $\sqrt{\varepsilon_k/\rho_k}$ when ρ_k is large. In floating point arithmetic it cannot be smaller than the accuracy with which h is evaluated, and in our tests, with $\varepsilon_k = 10^{-8}$ and $\rho_k \le 10^{20}$, it is at least of the order of 10^{-14}.

4. When $\rho_k \to \infty$, the guarantees of Algorithm 4.1 (without CPS) are weaker. Its feasible limit points at which PCC-LICQ holds satisfy (10), that is, they are M-stationary in the MPCC case (Theorem 4.6 and Corollary 4.7), but they need not be B-stationary, even under MPEC-LICQ and with exact solutions of the subproblems (Propositions 6.4–6.6).

As usual for augmented Lagrangian methods, the limit points may be infeasible. Proposition 4.5 (Section 4.4) also holds for Algorithm 5.3, under Assumption 1 and without the constraints c, with the same proof, which uses only Step 2 and condition (a). The constraint-preserving swaps, which keep h constant, cannot decrease the infeasibility measure Ψ.

6.2 Is the modification necessary? Can MPEC-LICQ be weakened?

Theorem 4.3 (original swaps) requires bounded penalty parameters but no constraint qualification, while Theorem 6.1 (CPS) requires PCC-LICQ but nothing on the penalty parameters. In this subsection we consider two questions in the MPCC case. Does the original method fail under MPEC-LICQ? Can MPEC-LICQ be replaced by MPEC-MFCQ in Theorem 6.1? The answer to the first question is yes, even with exact solutions of the subproblems and multipliers in the interior of the safeguarding box (counterexamples of Propositions 6.4, 6.5 and 6.6). The answer to the second one is no, at least with swaps in a single block (counterexample of Proposition 6.7).

Example 2. As in Example 1 (Section 4.3), $n = 3$, $\vec{x}_1 = (x_1, x_2)$ is a complementarity pair, x_3 forms a block without cardinality constraints, $x_1 \in [0, 2]$, $x_2 \in [0, 1]$, but with $x_3 \in [-2, 2]$, $h(x) = x_1 - x_3$, and $f(x) = -x_2/4 - x_1^2/2 + x_3^2/2 + x_3^4/4$. The feasible set is $X = \{(t, 0, t) \mid t \in [0, 2]\} \cup \{(0, t, 0) \mid t \in [0, 1]\}$, and the global minimizer is $(0, 1, 0)$, with $f = -1/4$. MPEC-LICQ holds at every feasible point other than $(2, 0, 2)$, because $\partial h/\partial x_3 = -1$ and x_3 is free at these points. The following proposition shows that the quadratic penalty method with the original swaps (Algorithm 4.1 with $\lambda_{\min} = \lambda_{\max} = 0$) may converge to a point that is not B-stationary even under MPEC-LICQ.

Proposition 6.4. *Consider Example 2 and Algorithm 4.1 with $\bar{\lambda}^k = 0$ for all k and $\rho_1 \geq 3$. For each k, let $s_k = (\rho_k - 1)^{-1/2}$, $t_k = s_k\rho_k/(\rho_k - 1)$, and $x^k = (t_k, 0, s_k)$. Then x^k is a strict local minimizer of L_k on Ω and satisfies conditions (a), (b) and (c) of Algorithm 4.1 with $\varepsilon_k = 0$, also with the look-ahead strategy of our implementation. Moreover, $\rho_k \to \infty$, $x^k \to 0$, and 0 is not B-stationary, although MPEC-LICQ holds at 0.*

Proof. Here $L_k(x) = f(x) + \rho_k(x_1 - x_3)^2/2$. Let $\rho = \rho_k$, $s = s_k$ and $t = t_k$. Then $t - s = s/(\rho - 1)$, $\partial L_k(x^k)/\partial x_1 = -t + \rho(t - s) = 0$ and $\partial L_k(x^k)/\partial x_3 = s + s^3 - \rho(t - s) = s + s^3 - t = s^3 - s/(\rho - 1) = 0$. Since $0 < t, s < 2$ (because $\rho \geq 3$), (a) holds and, since $\partial L_k(x^k)/\partial x_1 = 0$, so does (c). Near x^k, $x_1 > 0$ forces $x_2 = 0$, and the Hessian of L_k with respect to (x_1, x_3) is $\begin{pmatrix} \rho - 1 & -\rho \\ -\rho & \rho + 1 + 3s^2 \end{pmatrix}$, with determinant $3s^2\rho - 1 - 3s^2 = 2$ and positive diagonal, so x^k is a strict local minimizer of L_k on Ω. The only swap is $x_1 \to x_2$, with $\bar{y} = (0, 0, s)$, and, since $L_k(\bar{y} + ue^2) = L_k(\bar{y}) - u/4$, every point w produced by the escaping procedure satisfies $L_k(w) \geq L_k(\bar{y}) - 1/4$. Moreover,

$$L_k(\bar{y}) - L_k(x^k) = \frac{t^2}{2} + \frac{\rho}{2}\left(s^2 - (t - s)^2\right) \geq \frac{\rho}{2(\rho - 1)} - \frac{\rho}{2(\rho - 1)^3} \geq \frac{1}{2} - \frac{3}{16},$$

because $\rho/(\rho - 1) > 1$ and $\rho/(\rho - 1)^3 \leq 3/8$ for $\rho \geq 3$. Hence $L_k(w) - L_k(x^k) \geq 1/2 - 3/16 - 1/4 = 1/16 > 0$, and (b) holds. Since no swap decreases L_k, there are no candidates for the look-ahead strategy, which in our implementation only improves swaps that decrease the objective function. Since $h(x^k) = t - s = (\rho_k - 1)^{-3/2}$ depends only on ρ_k, the penalty parameter is increased at least every other iteration, so $\rho_k \to \infty$ and $x^k \to 0$. Finally, $(0, u, 0) \in X$ for $u \in [0, 1]$ and $\nabla f(0)^T e^2 = -1/4 < 0$, so 0 is not B-stationary. □

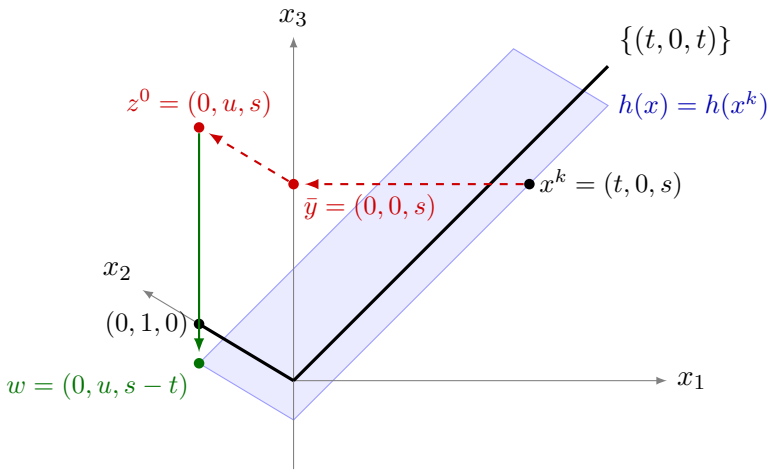


Figure 2: Example 2 at an iterate $x^k = (t, 0, s)$ of Proposition 6.4, close to the branch $\{(t, 0, t)\}$ of the feasible set (thick lines, not to scale). The original swap $x_1 \to x_2$ (dashed) goes to $\bar{y}$ and then along x_2, through points that are far from the plane $h(x) = h(x^k)$, because $h(\bar{y}) = -s$, so that L_k includes the penalty $\rho_k s^2/2$, larger than the decrease $u/4$ of f along x_2, and the swap is rejected. The CPS (Algorithm 5.2) starts from the same point $z^0 = (0, u, s)$ and restores the free variable x_3 with Algorithm 5.1, so that w lies on the plane $h(x) = h(x^k)$ and $L_k(w) - L_k(x^k) = f(w) - f(x^k) \approx -u/4 < 0$ when s and t are small. Thus, x^k does not satisfy condition (b) of Algorithm 5.3.

Theorem 6.1 applies to Algorithm 5.3 in Example 2, also with $\lambda_{\min} = \lambda_{\max} = 0$, so that no limit point can be 0. As Figure 2 shows, near 0 the CPS finds the descent along x_2 that the original swap misses. In Proposition 6.4, the safeguarding box is $\{0\}$, so that $\bar{\lambda}^k = 0$ for all k, the method is the quadratic penalty method, and $\lambda^* = 0$ lies on the boundary of this box. If the safeguarding box contains λ^* in its interior (for example, $\lambda_{\min} = -10^{20}$ and $\lambda_{\max} = 10^{20}$, as in Section 7.1) and the subproblems are solved accurately, the augmented Lagrangian method with the original swaps does not fail in this example. The multiplier estimate $\bar{\lambda}^2 = \lambda^2 > 0$ makes the subproblem of the second iteration move to the branch $x_1 = 0$ (see Section 7.1). In Proposition 6.6 below we show, with another example, that the method may fail also with such a safeguarding box and exact solutions of the subproblems. The next proposition shows, however, that the original method may fail under MPEC-LICQ with multipliers in the interior of the safeguarding box if the subproblems are solved inexactly, with tolerances $\varepsilon_k \to 0$ as allowed by Algorithm 4.1.

Proposition 6.5. *Consider Example 2 and Algorithm 4.1 with $\tau = 1/2$, $\gamma = 10$, $\rho_1 \geq 16$, $\lambda_{\min} < 0$, $\lambda_{\max} \geq 1/2$, $\bar{\lambda}^1 = 0$, $\bar{\lambda}^{k+1} = P_{[\lambda_{\min},\lambda_{\max}]}(\lambda^{k+1})$, arbitrary δ and η_{br}, $\varepsilon_1 = 1$, and $\varepsilon_k = 8/(k+1)^3 + 1/(k(k+1)\rho_1)$ for $k \geq 2$. Let $t_0 = 0$ and, for $k \geq 1$, $t_k = 1/(k+1)$, $h_k = (t_k - t_{k-1})/\rho_k$, and $x^k = (t_k, 0, t_k - h_k)$. Then x^k satisfies conditions (a), (b) and (c) of Algorithm 4.1, $\bar{\lambda}^{k+1} = \lambda^{k+1} = t_k \to 0 = \lambda^*$, which belongs to the interior of the safeguarding box, $\rho_k \to \infty$, and $x^k \to 0$, which satisfies MPEC-LICQ and is not B-stationary.*

Proof. Note that $h(x^k) = x_1^k - x_3^k = h_k$. By induction, $\bar{\lambda}^k = t_{k-1}$. This is true for $k = 1$ and, if it is true for k, then $\lambda^{k+1} = \bar{\lambda}^k + \rho_k h_k = t_k \in (0, 1/2]$, so that $\bar{\lambda}^{k+1} = t_k$. Let $t = t_k$, $s = x_3^k = t - h_k$, and $\mu = \lambda^{k+1} = t$.

Condition (a). We have $\partial L_k(x^k)/\partial x_1 = -t + \mu = 0$ and $\partial L_k(x^k)/\partial x_3 = s + s^3 - \mu = s^3 - h_k$. For $k = 1$, $s = 1/2 - 1/(2\rho_1)$ and $|s^3 - h_1| < 1 = \varepsilon_1$. For $k \geq 2$, $h_k = -1/(k(k+1)\rho_k) < 0$ and

$s \le t + 1/(k(k+1)\rho_1) \le 2t$, so that $|s^3 - h_k| \le 8/(k+1)^3 + 1/(k(k+1)\rho_1) = \varepsilon_k$. Since x_1^k and x_3^k are at distance at least $1 > \varepsilon_k$ from their bounds (for $k \ge 2$) and x_2 is fixed in the face, (a) holds (for $k = 1$, the projected derivatives are bounded by the derivatives). Since $\partial L_k(x^k)/\partial x_1 = 0$, condition (c) also holds.

Penalty parameters. For $k \ge 3$, let $r_k = |h_k|/|h_{k-1}| = (k-1)\rho_{k-1}/((k+1)\rho_k)$, and note that $|h_2|/|h_1| = 1/3$. Thus $\rho_3 = \rho_2 = \rho_1$, $r_3 = 1/2$, so that $\rho_4 = \rho_1$, and, for $k \ge 4$, if $\rho_k = \rho_{k-1}$, then $r_k = (k-1)/(k+1) > 1/2$ and $\rho_{k+1} = 10\rho_k$, whereas, if $\rho_k = 10\rho_{k-1}$, then $r_k < 1/2$ and $\rho_{k+1} = \rho_k$. Hence $\rho_k = \rho_1 10^{\lfloor (k-3)/2 \rfloor}$ for $k \ge 3$, and $\rho_k \to \infty$. Clearly, $x^k \to 0$.

Condition (b). The only possible swap is $x_1 \to x_2$, with $\bar{y} = (0, 0, s)$. Since $L_k(\bar{y} + ue^2) = L_k(\bar{y}) - u/4$, every point w produced by the escaping procedure satisfies $L_k(w) \ge L_k(\bar{y}) - 1/4$. Using $h(\bar{y}) = -s$, $s + h_k = t$ and $s^2 - h_k^2 = t(s - h_k)$,

$$L_k(\bar{y}) - L_k(x^k) = \frac{t^2}{2} - \bar{\lambda}^k t + \frac{\rho_k}{2} t(s - h_k).$$

For $k = 1$, this is $1/8 + (\rho_1 - 2)/8 \ge 1/4$. For $k \ge 2$, $\bar{\lambda}^k = 1/k \le 2t$ and $s - h_k = t - 2h_k \ge t$, so that $L_k(\bar{y}) - L_k(x^k) \ge t^2(\rho_k - 3)/2 \ge 1/4$ if $\rho_k \ge 3 + (k+1)^2/2$. This holds for $k \le 4$ because $\rho_k = \rho_1 \ge 16$, and for $k \ge 5$ because $\rho_k \ge 16 \cdot 10^{(k-4)/2}$, whose ratio between consecutive values, $\sqrt{10}$, is larger than the one of $3 + (k+1)^2/2$. Therefore, $L_k(w) \ge L_k(x^k)$ and (b) holds. Finally, MPEC-LICQ holds at 0, $\lambda^* = 0$ by Proposition 3.4, and 0 is not B-stationary, as in Proposition 6.4. □

In Proposition 6.5 the tolerances ε_k are much larger than $|h(x^k)|$, so that the condition $\varepsilon_k \le \eta_k \|h(x^k)\|$, with $\eta_k \to 0$, used in [1] to prove the boundedness of the penalty parameters, does not hold. Moreover, the branch problem Minimize f subject to $h(x) = 0$, $x_2 = 0$, does not satisfy the second-order sufficient condition at 0 (on the feasible branch, $f(t, 0, t) = t^4/4$). On the other hand, Theorem 6.1 holds for every sequence of tolerances $\varepsilon_k \to 0$, and the points of Proposition 6.5 do not satisfy condition (b) of Algorithm 5.3 for k large, because the CPS returns $w = (0, t, -h_k)$ with $f(w) - f(x^k) \approx -t/4$.

The following example shows that the original method may fail under MPEC-LICQ also with multipliers in the interior of the safeguarding box and exact solutions of the subproblems (strict local minimizers), with the parameters $\tau = 1/2$ and $\gamma = 10$ of Algencan. As expected from the results of [1], the second-order sufficient condition fails at the limit point (on the feasible branch, f is of order six). The iterates alternate between two branches of local minimizers of the subproblems, on which h has opposite signs.

Example 3. $n = 4$, $\vec{x}_1 = (x_1, x_2)$ is a complementarity pair, (x_3, x_4) forms a block without cardinality constraints, $\mathcal{B} = [0, 1]^2 \times [-1, 1]^2$, $h(x) = x_1 - x_4$, and

$$f(x) = -\frac{x_2}{4} + \frac{1}{6}\left[\left(\frac{x_1 + x_3}{2}\right)^6 + \left(\frac{x_1 - x_3}{2}\right)^6\right] + x_1 x_3 (x_1 - x_4).$$

MPEC-LICQ holds at $x^* = 0$, because x_4 is free and $\partial h/\partial x_4 = -1$, and x^* is not B-stationary, because $(0, u, 0, 0)$ is feasible for $u \in [0, 1]$ and $\nabla f(0)^T e^2 = -1/4$.

Proposition 6.6. *Consider Example 3 and Algorithm 4.1 with $\tau = 1/2$, $\gamma = 10$, $\lambda_{\min} = -1$, $\lambda_{\max} = 1$, $\rho_1 = 100$, $\bar{\lambda}^1 = 1/100$, $\bar{\lambda}^{k+1} = P_{[\lambda_{\min}, \lambda_{\max}]}(\lambda^{k+1})$, and $\varepsilon_k = 0$ for all k. For each k, let $y_k > 0$ be the positive root of $\rho_k y^2 - \sqrt{2}\, y - |\bar{\lambda}^k| = 0$, $s_k = (2y_k^2)^{1/4}$, and $x^k = (s_k, 0, s_k, s_k + y_k^2)$ if $\bar{\lambda}^k > 0$, $x^k = (s_k, 0, -s_k, s_k - y_k^2)$ if $\bar{\lambda}^k < 0$. Then x^k is a strict local minimizer of L_k on*

Ω *that satisfies conditions (a), (b) and (c) of Algorithm 4.1 with* $\varepsilon_k = 0$, $\bar\lambda^{k+1} = \lambda^{k+1} \to 0$, *which belongs to the interior of the safeguarding box,* $\rho_k \to \infty$, *and* $x^k \to 0$, *which satisfies MPEC-LICQ and is not B-stationary.*

Proof. In the face $x_2 = 0$, we use the coordinates $t_1 = (x_1 + x_3)/2$, $t_2 = (x_1 - x_3)/2$ and $\omega = h(x) = x_1 - x_4$, in which $L_k = (t_1^6 + t_2^6)/6 + (t_1^2 - t_2^2)\omega + \bar\lambda^k\omega + \rho_k\omega^2/2$. Let $\rho = \rho_k$, $y = y_k$, $s = s_k$, and assume that $\bar\lambda^k > 0$ (the other case is symmetric, with t_1 and t_2 exchanged). Then x^k corresponds to $t_1 = s$, $t_2 = 0$, $\omega = -y^2$, with $s^4 = 2y^2$, and $\partial L_k/\partial t_1 = s(s^4 + 2\omega) = 0$, $\partial L_k/\partial t_2 = 0$, and $\partial L_k/\partial\omega = s^2 + \bar\lambda^k - \rho y^2 = \sqrt{2}\,y + \bar\lambda^k - \rho y^2 = 0$. The Hessian is block diagonal, with the entry $5t_2^4 - 2\omega = 2y^2 > 0$ and the block $\left(\begin{smallmatrix} 8y^2 & 2s \\ 2s & \rho \end{smallmatrix}\right)$ in (t_1, ω), whose determinant $8\rho y^2 - 4\sqrt{2}\,y$ is positive because $\rho y > \sqrt{2}$ (the positive root satisfies $y > \sqrt{2}/\rho$). Since $x_1^k > 0$, Ω coincides with the face near x^k, so that x^k is a strict local minimizer of L_k on Ω and (a) holds with $\varepsilon_k = 0$ (the iterates belong to the interior of the box, see below). Since x_1^k is strictly between its bounds, $\partial L_k(x^k)/\partial x_1 = 0$, so that (c) also holds.

Multipliers and penalty parameters. We have $\lambda^{k+1} = \bar\lambda^k + \rho\omega = -\sqrt{2}\,y$ (and $+\sqrt{2}\,y$ if $\bar\lambda^k < 0$). Let $a_k = |\bar\lambda^k|$ and $z_k = \rho_k|\lambda^{k+1}|$. Then $z_k = 1 + \sqrt{1 + 2\rho_k a_k}$ and $|h(x^k)| = y_k^2 = z_k^2/(2\rho_k^2)$. If $|\lambda^{k+1}| \le 1$, then $\bar\lambda^{k+1} = \lambda^{k+1}$, $a_{k+1} = z_k/\rho_k$, and $z_{k+1} = 1 + \sqrt{1 + 2(\rho_{k+1}/\rho_k)z_k}$ with $\rho_{k+1}/\rho_k \in \{1, 10\}$. Since $z \mapsto 1 + \sqrt{1 + 20z}$ is increasing and its fixed point is 22, and $z_1 = 1 + \sqrt{3}$, by induction $2 < z_k \le 22$ and $|\lambda^{k+1}| = z_k/\rho_k \le 22/100 < 1$ for all k. Thus the multipliers alternate in sign, $\bar\lambda^{k+1} = \lambda^{k+1}$, and $|\lambda^{k+1}| \to 0$ if $\rho_k \to \infty$. If $\{\rho_k\}$ were bounded, then $\rho_k = \bar\rho$ for $k \ge k_1$, $z_{k+1} = 1 + \sqrt{1 + 2z_k}$, $z_k \to 4$, and $|h(x^k)|/|h(x^{k-1})| = (z_k/z_{k-1})^2 \to 1 > \tau$, so that the penalty parameter would be increased, a contradiction. Hence $\rho_k \to \infty$, $y_k \to 0$, $s_k \to 0$ and $x^k \to 0$. Moreover, $y_k = |\lambda^{k+1}|/\sqrt{2} \le 0.16$ and $s_k \le 0.48$, so that the iterates belong to the interior of $\mathcal{B}$.

Condition (b). The only swap is $x_1 \to x_2$, with $\bar y = (0, 0, x_3^k, x_4^k)$. Since $L_k(\bar y + ue^2) = L_k(\bar y) - u/4$, every point w produced by the escaping procedure satisfies $L_k(w) \ge L_k(\bar y) - 1/4$. Using $s^2 = \sqrt{2}\,y$, $h(\bar y) - h(x^k) = -s$ and $|\bar\lambda^k| = \rho y^2 - \sqrt{2}\,y$, a direct computation gives

$$L_k(\bar y) - L_k(x^k) = \frac{65\sqrt{2}}{96}y^3 + \frac{\rho y}{\sqrt{2}} \pm \sqrt{2}\,ys,$$

with the sign $+$ if $\bar\lambda^k > 0$ and $-$ otherwise. Since $\rho y > \sqrt{2}$ and $\sqrt{2}\,ys = 2^{3/4}y^{3/2} < 0.12$, we obtain $L_k(\bar y) - L_k(x^k) > 1 - 0.12 > 1/4$, so that $L_k(w) > L_k(x^k)$ and (b) holds. □

The following example shows that Theorem 6.1 does not hold if MPEC-LICQ is replaced by MPEC-MFCQ. The reason is that a descent direction may require two simultaneous swaps.

Example 4. $n = 6$, with complementarity pairs $(y_1, z_1) = (x_1, x_2)$ and $(y_2, z_2) = (x_3, x_4)$, a block $(w_1, w_2) = (x_5, x_6)$ without cardinality constraints, $\mathcal{B} = [0, 1]^6$, and

$$f(x) = y_1 + y_2 - z_1 - z_2 + 3w_1 + 3w_2, \qquad h(x) = z_1 - z_2 + w_1 - w_2.$$

The global minimizer is $(0, 1, 0, 1, 0, 0)$, with $f = -2$. The next proposition shows that the method may fail in this example.

Proposition 6.7. *In Example 4, MPEC-MFCQ holds at* $x^* = 0$, *MPEC-LICQ does not hold, and* x^* *is not B-stationary. Let* I *be the face index in which* y_1, y_2, w_1 *and* w_2 *are the variables that are not fixed at zero in* F_I. *Then, for every* $\rho > 0$ *and* $\bar\lambda$ *with* $-3 \le \bar\lambda \le 3$, *the point* $x^* = 0$ *satisfies conditions (a) and (b) of Algorithm 5.3 with* $\varepsilon_k = 0$. *More strongly, for every*

single swap ($y_1 \to z_1$ or $y_2 \to z_2$) and every point w of the new face with $h(w) = h(0)$, we have $f(w) \geq f(0)$. Therefore, the constant sequence $x^k = 0$ (with $\bar{\lambda}^1 = 0$) is generated by Algorithm 5.3, and its limit is not B-stationary.

Proof. At $x^* = 0$ all the variables are at their lower bounds and both pairs are biactive. The equality constraints of the tightened problem are $h = 0$ and $y_1 = z_1 = y_2 = z_2 = 0$, whose gradients are linearly independent, because ∇h has non-null components in w_1 and w_2. The direction d with $d_{w_1} = d_{w_2} = 1$ and the other components null satisfies $\nabla h^T d = 0$ and is strictly feasible for the active bounds $w \geq 0$. Thus, MPEC-MFCQ holds. MPEC-LICQ fails because $\mathcal{F}(0) = \emptyset$. The points $(0, u, 0, u, 0, 0)$, $u \in [0, 1]$, are feasible, with $f = -2u$, so 0 is not B-stationary. Since $h(0) = 0$, $\nabla L_k(0) = \nabla f(0) + \bar{\lambda}\nabla h(0)$, whose components corresponding to the variables that are not fixed at zero in F_I (y_1, y_2, w_1, w_2) are 1, 1, $3 + \bar{\lambda}$ and $3 - \bar{\lambda}$. Since these variables are at their lower bounds, $g_I^{L_k}(0) = 0$ and (a) holds. Also, $\lambda^{k+1} = \bar{\lambda}^k$, so that, with $\bar{\lambda}^1 = 0$, $\bar{\lambda}^k = 0$ and $\rho_k = \rho_1$ for all k. Since $h(w) = h(0)$ implies $L_k(w) - L_k(0) = f(w) - f(0)$, it suffices to consider f. In the swap $y_1 \to z_1$, the points of the new face have $z_2 = 0$, and $h(w) = 0$ with $z_1 = t \geq 0$ gives $w_2 = t + w_1$, so that $f(w) = y_2 - t + 3w_1 + 3(t + w_1) \geq 2t \geq 0$. The swap $y_2 \to z_2$ is analogous ($w_1 = t + w_2$). Hence no CPS returns a point, whatever the restoration procedure, and (b) holds. (In Algorithm 5.2, the restoration returns no point because the moving variables w_1, w_2 have $d_v = 0$.) □

A descent direction at 0 requires z_1 and z_2 to enter simultaneously. Under MPEC-LICQ, the free variables can always compensate the variation of h caused by a single entering variable, which is why swaps in one block at a time suffice. Under MPEC-MFCQ, a complete test would have to consider simultaneous swaps in several blocks, guided by the linearization of the problem, which leads to combinatorial subproblems such as the linear programs with equilibrium constraints of [30]. In Example 4, the original swaps of the Telescopic Method, which do not preserve h, do escape from 0, because, from $\bar{y} = 0$ along z_1, $L_k(ue^{z_1}) = -u + \rho u^2/2 < 0$ for small u. Indeed, our implementation, which keeps both kinds of swaps, finds the global minimizer in the first outer iteration (Section 7.1).

Remark 6.8. In Example 4 (MPEC-MFCQ), the limit point $x^* = 0$ of Algorithm 5.3 in the counterexample of Proposition 6.7 is A-stationary but not M-stationary, because M-stationarity of the first pair requires $\lambda \geq 1$, and that of the second, $\lambda \leq -1$. The methods of [18, 21] and [2], which also keep the complementarity constraints in the subproblems, obtain M-stationary limit points under their constraint qualifications, whereas, with the constraint-preserving swaps, Theorem 6.1 gives B-stationarity under MPEC-LICQ.

7 Numerical experiments

We report experiments with the examples of Sections 4 and 6 (Section 7.1), with the MacMPEC collection (Section 7.2), and with problems with partitioned cardinality constraints (Section 7.3).

We implemented Algorithm 5.3 in Fortran, as a driver around the implementation of the Telescopic Method of [8], the subroutine `pcc_telescopic`. Algorithm MIA is Gencan, with exact second derivatives and its default parameters. As in Step 3 of Algorithm 5.1 of [8], a single iteration of Gencan is performed on each visited face ($G = 1$, in the notation of Section 7.2). We used $\varepsilon = 10^{-8}$, $\delta = \eta_{\text{br}} = +\infty$, and look-ahead with 10 candidates, in whose faces Gencan is run until convergence. As in Algencan, the safeguarded multipliers are the projections $\bar{\lambda}^{k+1} = P_{[\lambda_{\min}, \lambda_{\max}]^m}(\lambda^{k+1})$. In all the experiments, $\varepsilon_k = 10^{-8}$ for all k. The subroutine `pcc_telescopic` stops when condition (a) holds with this tolerance and the swap test of [8] is satisfied, and

the CPS search described below checks condition (b). For Algorithm 4.1, `pcc_telescopic` also checks condition (c) when it stops. If (c) does not hold, the variables of the face that violate it are set to zero, provided that this decreases L_k, and the method goes on. Step 1 is implemented as follows. First, `pcc_telescopic` is applied to L_k. When it stops, a CPS search is performed over all the blocks with cardinality constraints, leaving and entering variables, and admissible directions, with $\delta = +\infty$, $t_{\max} = 1$, $t_{\min} = 10^{-8}$, $\alpha = 10^{-4}$, $\varepsilon = 10^{-8}$ in (11), and $j_{\max} = 50$. The restoration used in the other experiments is described in Section 7.2. If some CPS returns w with $L_k(w) < L_k(x^k) - 10^{-8}$, the best such w is taken as the new initial point and `pcc_telescopic` is called again, in the face of w. The outer iterations use $\tau = 0.5$ and $\gamma = 10$. The codes were compiled with gfortran 15.2 (option `-O3`), and all the experiments were run on an Apple M1 Max computer (8 performance cores, 64 GB of memory, macOS 13.7), with eight sequential runs in parallel. The codes, the test problems and the complete results (final points, residuals, stationarity measures and objective values of all the methods) are available at `https://www.ime.usp.br/~egbirgin/`. Tables with the results by instance of all the methods are given in the supplementary material.

7.1 Examples

In the examples of Sections 4 and 6, we used $\rho_1 = 10$, $\bar{\lambda}^1 = 0$, $x^0 = (1, 0, 0.9)$ in Examples 1 and 2 and $x^0 = (0.5, 0, 0.5, 0, 0, 0)$ in Example 4 (Section 6.2), and stop when $\|h(x^k)\|_\infty \leq 10^{-8}$. Since $\nabla L_k(x^k) = \nabla f(x^k) + \nabla h(x^k)\lambda^{k+1}$, condition (a) says that, at the final iterate, the projected gradient of the Lagrangian $f + (\lambda^{k+1})^T h$ on the face is at most 10^{-8}. We used two settings of the multipliers. In the augmented Lagrangian method (AL), $\lambda_{\min} = -10^{20}$ and $\lambda_{\max} = 10^{20}$. In the quadratic penalty method (QP), $\lambda_{\min} = \lambda_{\max} = 0$, so that $\bar{\lambda}^k = 0$ for all k. To show that the CPS works independently of the penalty parameter, we also ran the method with the CPS activated only from a prescribed outer iteration k_{cps} on.

Example 1 (Section 4.3). By Proposition 4.4, Algorithm 4.1 (without CPS) may generate iterates that converge to $x^* = 0$, which is not B-stationary, and this is what our implementation does. MPEC-LICQ does not hold at $x^* = 0$ (all the variables are at their bounds, so $\mathcal{F}(x^*) = \emptyset$) nor at $(0, 1, 0)$, where $\nabla h = 0$, so Theorem 6.1 does not apply. Nevertheless, the CPS works in this example. At $x^k = (s_k + s_k^3, 0, s_k)$, the restoration moves only x_3, with $d_3(x^k) = s_k > 0$ and $\partial h(0, t, s_k)/\partial x_3 = s_k(1 + 2t) \neq 0$, and returns $w = (0, t, x_3)$ with $(1 + 2t)x_3^2 = s_k^6$, at which $f(w) \approx -t/2$.

Table 2 summarizes the runs. Telescopic AL and CPS AL denote Algorithms 4.1 (original swaps only) and 5.3 (both kinds of swaps) with the AL setting of the multipliers, and Telescopic QP and CPS QP denote the same algorithms with the QP setting. In the runs of the original method that failed, the iterates behaved as predicted by Proposition 4.4 and Proposition 6.4 (in Example 2, because of the tolerance of the subproblems, the iterates are only approximations of the points of Proposition 6.4 when ρ_k is large). In Examples 1 and 2, in all the runs with CPS, exactly one CPS was performed, with $t = 1$, after which the Telescopic Method converged in the new face to an approximation of the global minimizer $(0, 1, 0)$, where the outer iterations stopped. In Example 1 with the CPS activated at $k = 40$ ($\rho_{40} = 10^{20}$), the CPS was performed at $x^{40} = (3.0010 \times 10^{-3}, 0, 3.0009 \times 10^{-3})$, with $f(x^{40}) = -6.1 \times 10^{-11}$, and returned $w = (0, 1, 4.8 \times 10^{-8})$ with $f(w) = -0.5$ and $L_{40}(w) = -0.5$. Thus, the penalty parameter played no role. The restoration tolerance (11) was 7×10^{-15}. In Example 1 with $k_{cps} = 1$, the CPS was rejected at $k = 1$ (the iterate was far from feasible, with $f(x^1) = -0.68$) and accepted at $k = 2$, at $x^2 = (1.079, 0, 0.714)$ with $f(x^2) = -0.26$, returning $w = (0, 1, 0.210)$ with $f(w) = -0.477$. In

Example 4 (Section 6.2), started in the face of y_1 and y_2, the original swaps of `pcc_telescopic` found the global minimizer $(0, 1, 0, 1, 0, 0)$ in the first outer iteration, with or without CPS.

Example	Method	k_{cps}	Outer it.	CPS	Final point	f	$\|h\|_\infty$
1	Telescopic AL	–	21	–	(4.9e-2, 0, 4.9e-2)	−4.3e-6	6.8e-9
	CPS AL	1	2	1	(0, 1, 7.3e-20)	−0.5	8.0e-39
	CPS AL	21	21	1	(0, 1, 4.2e-12)	−0.5	2.6e-23
	CPS AL	40	40	1	(0, 1, 6.9e-9)	−0.5	7.1e-17
2	Telescopic QP	–	11	–	(2.3e-3, 0, 2.3e-3)	1.5e-12	2.3e-9
	CPS QP	1	1	1	(0, 1, −1.4e-16)	−0.25	1.4e-16
	CPS QP	11	11	1	(0, 1, 0)	−0.25	0
	Telescopic AL	–	6	–	(0, 1, 3.2e-10)	−0.25	3.2e-10
4	Telescopic AL	–	1	–	(0, 1, 0, 1, 0, 0)	−2	0
	CPS AL	1	1	0	(0, 1, 0, 1, 0, 0)	−2	0

Table 2: Results. "Outer it." is the number of outer iterations, "CPS" the number of constraint-preserving swaps performed. With $k_{cps} > 1$, the stopping test is applied only for $k \geq k_{cps}$. The penalty parameters at $k = 21$ and $k = 40$ of Example 1 are 10^{10} and 10^{20}, and at $k = 11$ of Example 2 it is 10^6.

7.2 The MacMPEC collection

MacMPEC [28] is a collection of AMPL models of MPCCs with general constraints. We used the model and data files distributed with MacMPEC.jl [26], which also lists the best objective value known for each instance (most of them obtained with the filter SQP method of [14]). Each of the 193 model/data combinations was translated by AMPL into a `.nl` file, read by the AMPL Solver Library (ASL) [17]. Instance `ex9.1.2` has a binary variable, which was reformulated as in Section 3 for all the methods. Four of them (`pack-rig2-16`, `pack-rig2c-16`, `pack-rig2-32` and `pack-rig2c-32`), marked as infeasible in the list of [26], are counted as failures of all the methods in the columns "Feas." and "Stat." of Table 3 (see item 2 below). We call *small* the 116 instances with at most 100 variables in the AMPL model and *large* the other 77 (the largest has 3436 variables).

With the option `cc_simplify` of the ASL, every complementarity condition becomes $0 \leq x_a \perp x_b \geq 0$ for two variables (the ASL adds auxiliary variables and equality constraints when a side of the condition is an expression, and splits mixed complementarity conditions), and the other constraints have the form $\ell_i \leq c_i(x) \leq u_i$. Every inequality constraint is written as $c_i(x) - s_i = 0$, with a slack $s_i \in [\ell_i, u_i]$. In this way the problem has the form considered in Section 3, in which every complementarity pair is a block with $p = 1$ and the remaining variables (original and auxiliary variables and slacks) form a block without cardinality constraints. Functions and first and second derivatives are evaluated by the ASL. When an evaluation fails (for example, outside the domain of a function), the value 10^{99} is returned for the objective function of the subproblem, so that the trial point is rejected.

We compared variations of the augmented Lagrangian method with telescopic subproblems with two methods from the literature. All of them use Gencan (the bound-constrained solver of Algencan 4.0.0) for the subproblems:

- CPS AL: Algorithm 5.3 (original swaps and constraint-preserving swaps);
- Telescopic AL: the same method with the original swaps only (Algorithm 4.1);
- Algencan: Algencan 4.0.0 [9] applied to the nonlinear programming reformulation in which each pair is replaced by $x_a \geq 0$, $x_b \geq 0$, $x_a x_b \leq 0$ (augmented Lagrangian methods applied

to this reformulation were analyzed in [20, 5]);

- Scholtes: the regularization method of Scholtes [33], with $x_a x_b \leq t$ for $t = 1, 10^{-1}, \ldots, 10^{-8}$ and finally $t = 0$, each problem being solved by Algencan 4.0.0 from the solution and the multipliers of the previous one.

In CPS AL and Telescopic AL, the subproblems are solved by the implementation described at the beginning of Section 7, with three differences. First, at most $G = 10$ iterations of Gencan are performed on each visited face before the next swap search (Algorithm 5.1 of [8] performs a single one, and Remark 5.1 of [8] allows any number). Other values of G are considered below. Second, there is no look-ahead. Third, we use simultaneous swaps. When improving swaps are found in several blocks, the best swap of each block is performed at the same time. If the resulting point is not better than the best single swap, halves of the list are tried and, at last, the best single swap is performed. This is allowed by Remark 5.1 of [8], by which any point of the new face with a functional value not larger than f_{best} may be chosen. The values of the constraints are updated incrementally after a swap. The CPS search is performed when the Telescopic Method stops at a point with $\|h(x)\|_\infty \leq 10^{-5}$, since condition (b) is required in the proof of Theorem 6.1 only for the iterates close to a feasible limit point. Its restoration uses simplified Newton iterations, in which the matrix of Algorithm 5.1 is replaced by the matrix $J_S D^2 J_S^T + \kappa_{\text{reg}} I$, computed at the starting point y and kept fixed in all the iterations. Here J_S is the Jacobian of h with respect to the moving variables, $D = \text{diag}(d_v)$ and $\kappa_{\text{reg}} = 10^{-12} \max\{1, \max_i (J_S D^2 J_S^T)_{ii}\}$, and the matrix is factorized once per search by MA57, through the linear solver interface of Algencan. Lemma 5.1 also holds for this version, as shown after its proof. The column of the leaving variable is removed by the Sherman–Morrison formula. At most 50 iterations are performed, and the restoration also stops without returning a point if an iterate leaves $\mathcal{B}$ or if the residual $\|h(z^j) - \eta\|$ becomes larger than twice its initial value. The candidates (pair, entering variable) are sorted by the first-order prediction of the variation of L_k, only candidates with negative prediction are tried, and the first CPS that returns a point is accepted ($\alpha = 10^{-4}$, $t_{\max} = \max\{1, \|x\|_\infty\}$, $t_{\min} = 10^{-2}$, $\varepsilon = 10^{-8}$). Since $t_{\min}$ and ε are fixed in the implementation, Theorem 6.1 does not apply literally, and the final points are verified a posteriori, as described below.

The initial point of CPS AL and Telescopic AL is computed by a light Phase I, in which Algencan, with at most 10 outer iterations and tolerances 10^{-4}, is applied to the reformulation used by Algencan from the initial point of the AMPL model. If it stops at a point whose sup-norm of the constraint violation is larger than 10^{-4}, it is applied again from a random perturbation of the AMPL point (fixed seed), until such a point is found or the time limit is reached. The best point found, in which the smaller variable of each pair is set to zero and the slacks are set to the projections of the constraint values, is the initial point of the method, and its multipliers are the initial safeguarded multipliers $\bar{\lambda}^1$. The outer iterations use the parameters of Algencan, $\tau = 0.5$, $\gamma = 10$, $\lambda_{\min} = -10^{20}$, $\lambda_{\max} = 10^{20}$ and $\rho_1 = \max\{10^{-8}, \min\{10 \max\{1, |f(x^0)|\} / \max\{1, \frac{1}{2}\|h(x^0)\|^2\}, 10^8\}\}$. There are at most 50 outer iterations, and the method stops when $\|h(x^k)\|_\infty \leq 10^{-8}$ and the subproblem was solved with the required tolerance, when $\|h(x^k)\|_\infty \leq 10^{-8}$ at three consecutive iterations in which the subproblem solver did not reach the required tolerance (as in Algencan), or when $\rho_k > 10^{20}$. If a subproblem is unbounded below, its solution is discarded and the subproblem is solved again from the previous iterate with ρ_k multiplied by γ. To assess the CPS independently of Phase I, CPS AL and Telescopic AL were also run without Phase I, from the initial point of the AMPL model projected onto the bounds (with the same treatment of pairs and slacks), in variants denoted by CPS AL_0 and Telescopic AL_0.

In Algencan and Scholtes, the feasibility, complementarity and optimality tolerances of Algencan were set to 10^{-8}, with at most 50 outer iterations and the default values of the other parameters, from the initial point of the AMPL model projected onto the bounds. Each run had a time limit of one hour. In CPS AL and Telescopic AL, the limit is checked by the code between subproblems (and between Phase I attempts), and the best point found is returned. A run still active after 3800 seconds was interrupted and counted as a failure. Algencan and Scholtes were interrupted after one hour.

The final points were checked independently of the methods. The original AMPL model (without simplifications) was read by the ASL, and we computed at the final point the maximum violation of the bounds and the constraints and the complementarity residual. For a condition $\ell \leq x_j \leq u$ complementary to $c_i(x)$, the residual is $|x_j - P_{[\ell,u]}(x_j - c_i(x))|$, which is $\min\{|x_j - \ell|, |c_i(x)|\}$ when $u = +\infty$. A point is *feasible* when both quantities are at most 10^{-6}. At feasible points, stationarity was checked in the simplified problem by solving, with the bounded-variable least-squares method, $\nabla f(x) = \sum_{i\in\mathcal{A}} \mu_i \nabla c_i(x) + \sum_{j\in\mathcal{V}} \pi_j e^j$, where $\mathcal{A}$ and $\mathcal{V}$ are the active constraints and bounds (relative tolerance 10^{-6}), with the signs given by the activity (free multipliers for equalities and for the null variable of a nondegenerate pair). The point is *S-stationary* when the relative residual is at most 10^{-6} with nonnegative multipliers for both variables of every biactive pair. Otherwise, if there are at most 10 biactive pairs, all the branches were checked (in each biactive pair, the multiplier of one of the variables is free and the other is nonnegative), and the point is *B-stationary* when the residual is at most 10^{-6} in every branch (under MPEC-LICQ, this is B-stationarity). A run is a *success* when its final point is feasible and S- or B-stationary. At infeasible final points, we checked the stationarity of the infeasibility measure Ψ (Proposition 4.5), with the slacks at their best values. The point is *stationary for* Ψ when the projected gradient of Ψ on the face of the point (for Algencan and Scholtes, on the box, with the complementarity constraints written as $x_a x_b \leq 0$) is at most $10^{-6}\|\nabla\Psi\|_\infty$. The relative test excludes nearly feasible points, at which $\nabla\Psi$ is tiny. Finally, the objective value of a feasible point was compared with the best known value f_{best}, and it *reaches* f_{best} when it is at most $f_{best} + 10^{-3}\max\{1, |f_{best}|\}$ (for minimization).

Table 3 summarizes the results.

	All (193)				Small (116)			Large (77)		
Method	Feas.	Stat.	Ψ	Best	Feas.	Stat.	Best	Feas.	Stat.	Best
CPS AL	185	185	2	165	115	115	110	70	70	55
Telescopic AL	185	185	2	165	115	115	110	70	70	55
CPS AL_0	175	174	13	148	108	108	99	67	66	49
Telescopic AL_0	175	174	13	145	108	108	97	67	66	48
Algencan	179	171	8	156	110	106	102	69	65	54
Scholtes	168	165	8	150	109	106	102	59	59	48

Table 3: MacMPEC: number of instances in which the final point is feasible ("Feas."), is feasible and S- or B-stationary ("Stat."), and reaches the best known value ("Best"), with $G = 10$ in CPS AL, Telescopic AL, CPS AL_0 and Telescopic AL_0. "Ψ" is the number of instances in which the final point is infeasible and stationary for Ψ. The four infeasible instances are counted as failures in "Feas." and "Stat.".

We highlight the following facts.

1. Stationarity. All the feasible final points of CPS AL and Telescopic AL, and all but one of CPS AL_0 and of Telescopic AL_0 (both in `bem-milanc30-s`), were certified S- or B-stationary, against 171 of the 179 feasible points of Algencan and 165 of the 168 of

Scholtes. Four points of CPS AL are B-stationary with biactive pairs (`ex9.2.2`, `qpec2`, `ralph1`, `scholtes4`). In these instances, the complementarity residuals of Algencan are between 3.7×10^{-5} and 7.5×10^{-5}, a typical consequence of the degeneracy of the reformulation $x_a x_b \leq 0$, and, with tolerance 10^{-4} for this residual, Algencan and Scholtes would be feasible in 184 and 182 instances. In `bem-milanc30-s`, CPS AL, Telescopic AL and Algencan found S-stationary points with objective value 0.0944, much smaller than the value 83.28 of the list. In the 179 instances in which both CPS AL and Algencan found feasible points, the objective values coincide (up to 10^{-3}) in 168, CPS AL is better in 7 and Algencan in 4.

2. Failures and infeasible points. The failures of CPS AL are the four infeasible instances and `TSC-7`, `TSC-8`, `TSC-12` and `taxmcp`, in which Phase I did not find a point with violation at most 10^{-4} within the time limit. In the four infeasible instances, all the runs that stopped before the time limit found the same point, stationary for Ψ, with $\|h\|_\infty = 1.0 \times 10^{-4}$ (`pack-rig2-16`, `pack-rig2c-16`) and 2.3×10^{-4} (`pack-rig2-32`, `pack-rig2c-32`), which supports their infeasibility. Without Phase I, CPS AL_0 and Telescopic AL_0 also stopped with $\rho_k > 10^{20}$ at infeasible points stationary for Ψ in nine other instances, as predicted by Proposition 4.5.
3. Phase I and CPS. With Phase I, no CPS was accepted, and CPS AL and Telescopic AL returned the same points (the CPS search found candidates in 113 instances, but the Telescopic Method had already stopped at S-stationary points). Without Phase I, the CPS improved the final point of Telescopic AL_0 in five instances, namely `ex9.2.5` ($f = 9 \to 5$), `kth3` ($1 \to 0.5$) and `qpec-100-2` ($-4.2517 \to -6.5907$), reaching the best known values, and `qpec-200-4` ($-3.2184 \to -4.9845$) and `qpec-100-1` ($0.2828 \to 0.2782$). Phase I increased the number of successes from 174 to 185. In `ex9.1.4`, for example, the first subproblem without Phase I, with a small penalty parameter, chooses branches of the pairs with no feasible point near the limit, and the CPS, which preserve the values of the constraints, cannot correct this.
4. Time. In the 167 instances in which all four methods found feasible points, the total times were 814 s (CPS AL), 730 s (Telescopic AL), 599 s (Algencan) and 1321 s (Scholtes), so that the CPS searches increased the time by about 12%. The four runs of CPS AL that reached the time limit account for 14401 of its 18906 seconds. CPS AL_0 and Telescopic AL_0 took 4099 seconds (2569 in `bem-milanc30-s`) and 1782 seconds.
5. Iterations of Gencan per visited face. With $G = 1$, the swap search is performed after every iteration of Gencan, and its cost dominates in the instances with many pairs (1835 seconds in `incid-set1c-32`, with 961 pairs, against 52 seconds with $G = 10$). CPS AL had 183, 184, 185 and 184 successes with $G = 1$, 5, 10 and 50000, and the total times in the 183 instances solved by the four versions were 4924, 1273, 1102 and 1263 seconds.

Nurkanović and Leyffer [30] solved 191 MacMPEC problems (translated to CasADi, with IPOPT as NLP solver) and counted as a success the computation of a B-stationary point (certified a posteriori for the methods other than MPECopt). Their highest success rate, 94.24%, was obtained by MPECopt with regularization in Phase I, followed by 93.2% for a regularization method, 92.14% for an ℓ_∞ penalty method, 90.0% for IPOPT applied to the NLP reformulation and 72.52% for a mixed-integer reformulation solved by Bonmin. With the same criterion (feasible and S- or B-stationary), the success rates of our experiments in their list of 191 problems are 97.4% (CPS AL and Telescopic AL), 91.6% (CPS AL_0 and Telescopic AL_0), 90.1% (Algencan) and 85.9% (Scholtes). Their list contains 14 repeated instances under different names (for example, `pack-rig-k` and `pack-rig2p-k`), and the rate of CPS AL in its 177 distinct instances

is 97.2%. The comparison is only indicative, because the formulations, NLP solvers, tolerances, time limits and computers are different. Infeasible final points are counted as failures, as in [30], although MPECopt also has a convergence result for infeasible points (Theorem 4.2 of [30]).

In summary, as predicted by the theory, the feasible final points of CPS AL were S- or B-stationary, and CPS AL and Telescopic AL succeeded in 97.4% of the problems of [30]. With Phase I, the original swaps were already sufficient in this collection, and the CPS added the guarantee of Theorem 6.1 at a small cost. When the initial point does not lie in a good branch, the CPS improve the points at which the original swaps stop (CPS AL_0 versus Telescopic AL_0). The remaining failures are related to Phase I, not to the CPS.

7.3 Partitioned cardinality constraints

In this subsection the blocks have more than two variables and, in two of the three families, more than one non-null variable is allowed in each block. We used the implementation of Section 7.2, extended to general blocks, with the same variants. CPS AL and Telescopic AL start from the point given by Phase I, that is, by Algencan applied to the problem without cardinality constraints, followed by keeping the largest entries of each block. CPS AL_0 and Telescopic AL_0 are the same methods without Phase I. When the variables are non-negative, we also ran Algencan and the regularization method of Scholtes applied to the reformulation of [12]. The time limit was 600 seconds. Every final point was verified independently for feasibility with tolerance 10^{-6}, cardinality (entries with absolute value at most 10^{-8} are considered null), and stationarity in the sense of Definition 3.1 for the Lagrangian, with tolerance 10^{-6} (under PCC-LICQ, B-stationarity). We then say that the final point is *verified*. For each instance, we report the relative difference between the objective function value of each method and the best value found by all the methods (the gap). In the three families described below, the constraint-preserving swaps improved the original swaps, never worsening the final point, and the resulting method was better than the methods applied to the reformulation of [12] and close to the reference values computed by CPLEX.

7.3.1 Portfolio selection

Mean-variance portfolios. The problem is

$$\text{Minimize } x^T Q x \text{ subject to } \bar{r}^T x \geq r_{\min},\ \textstyle\sum_{j=1}^n x_j = 1,\ 0 \leq x \leq u,\ \|x\|_0 \leq s,$$

where Q is the covariance matrix of the returns of n assets, $\bar{r}$ is the vector of mean returns, $r_{\min}$ is the minimum required return and u is the vector of upper bounds (one block, with $p_1 = s$). We used the 90 instances of Frangioni and Gentile [16] (available at `https://commalab.di.unipi.it/datasets/MV/`), with $n = 200$, 300 and 400, ignoring the minimum buy-in constraints, as in [23, 21], and $s \in \{5, 10, 20\}$ (270 problems). The initial point of CPS AL_0 and Telescopic AL_0 is $x = 0$. As a reference, we used CPLEX 22.2 applied to the mixed-integer formulation, with one thread and a time limit of 600 seconds, which it reached in 266 problems, so that its values are not certified to be optimal. Table 4 shows, for each method, the number of verified final points and the mean gap. All the final points of Telescopic AL, Telescopic AL_0, Scholtes and CPLEX were verified.

Comparing CPS AL with Telescopic AL problem by problem, CPS AL found a better point in 194 problems and the same point in 74, and never a worse one. CPS AL was better than Scholtes in 245 problems and worse in 13, and better than Algencan in 188 problems and worse in 29. The eight failures of CPS AL and CPS AL_0 occurred for $n = 400$ and $s = 20$, by the time

	$s = 5$	$s = 10$	$s = 20$
CPS AL	90 / 0.35%	90 / 0.32%	88 / 0.07%
CPS AL_0	90 / 0.48%	90 / 0.24%	84 / 0.25%
Telescopic AL	90 / 5.65%	90 / 4.51%	90 / 4.05%
Telescopic AL_0	90 / 12.9%	90 / 14.7%	90 / 11.2%
Algencan	81 / 10.1%	76 / 7.44%	71 / 5.99%
Scholtes	90 / 3.66%	90 / 3.59%	90 / 3.54%
CPLEX	90 / 0.01%	90 / 0.00%	90 / 0.01%

Table 4: Mean-variance portfolios (90 problems for each s): number of verified final points and mean gap.

limit. CPLEX found better points than CPS AL in 157 problems. In all of them, the two points have different sets of non-null variables, that is, CPS AL stops at another local minimizer. The median relative difference was 0.2% in these problems and 0.03% over all the problems. The constraint-preserving swaps were accepted in most problems, and they are responsible for the difference between CPS AL and Telescopic AL. They make the method slower. The total time of CPS AL was 11580 seconds, against 2962 for Telescopic AL, because there are $s(n-s)$ candidate swaps.

The instances with $n = 200$ were also used in [21], with the same cardinalities. A direct comparison is not possible, because the code of [21] is not available and the results are reported only graphically, relative to the values obtained by CPLEX with a time limit of 30 minutes. According to [21], the values obtained by their augmented Lagrangian method are usually not too far from those of CPLEX, and those of the boosted version are usually very close to them. In these 90 problems, the values obtained by CPS AL were equal to or better than those of CPLEX (with 600 seconds) in 41 problems. The median relative difference was 0.02% (mean 0.25%, maximum 4.8%).

Portfolios with a risk constraint and sector cardinality. The problem is

$$\text{Maximize } \bar{r}^T x \text{ subject to } x^T C x \le \sigma_{\max}^2,\ \textstyle\sum_{j=1}^n x_j = 1,\ 0 \le x \le u,\ \|\vec{x}_\nu\|_0 \le p,\ \nu = 1, \ldots, N,$$

where $\bar{r}$ and C are the mean returns and the covariance matrix of the assets, and each block $\vec{x}_\nu$ is a sector, in which at most p assets may be chosen. We used the data of the S&P 500 index used in [8] ($n = 497$ assets in $N = 11$ sectors, with $\bar{r}$ and C computed from 10,000 scenarios), with $p \in \{1, 2, 3\}$, $u \in \{0.2, 1\}$ (the same bound for all the assets) and $\sigma_{\max} \in \{0.15, 0.20, 0.25\}$ (18 problems). We excluded $\sigma_{\max} = 0.10$, for which no method, CPLEX included, found a feasible point. The reference values were computed by CPLEX applied to the mixed-integer quadratically constrained formulation (600 seconds). All the final points of CPS AL and Telescopic AL were verified, as were 17 of the 18 points of CPS AL_0, Telescopic AL_0 and Scholtes, and only 6 of Algencan. Because of its integrality tolerances, the points of CPLEX were checked only for feasibility and cardinality, and 17 of them passed. Table 5 shows the number of verified points and the mean gap over them (the best value is the best one among the verified points of all the methods, CPLEX included).

CPS AL was better than Telescopic AL in 9 problems and equal in the other 9. It was better than Scholtes in 8 problems, equal in 8 and worse in 1. CPS AL and Telescopic AL took about 4 and 3 seconds per problem.

	Verified points	Mean gap (%)
CPS AL	18	0.15
CPS AL_0	17	0.08
Telescopic AL	18	0.50
Telescopic AL_0	17	3.8
Algencan	6	6.8
Scholtes	17	0.94
CPLEX	17*	0.02

Table 5: Portfolios with a risk constraint and sector cardinality (18 problems): number of verified final points and mean gap. *Feasibility and cardinality only.

7.3.2 Sparse principal component analysis

The problem is

$$\text{Minimize} -x^T Ax \text{ subject to } \|x\|^2 = 1,\ -1 \leq x \leq 1,\ \|x\|_0 \leq p,$$

where $A \in \mathbb{R}^{n\times n}$ is a covariance matrix, so that $x^T Ax$ is the variance explained by the direction x, and at most p entries of x may be non-null (one block). We used the sample covariance matrices A of n observations of a normal vector in $\mathbb{R}^n$ with covariance $I + \theta vv^T$, where v has 10 non-null entries $\pm 1/\sqrt{10}$, for $n \in \{50, 100, 200\}$, $\theta \in \{1, 3\}$ and five random seeds, and $p \in \{5, 10, 20\}$ (90 problems). Since $x = 0$ is stationary for the augmented Lagrangian, CPS AL_0 and Telescopic AL_0 start from the canonical vector of the largest diagonal entry of A. We compared with the truncated power method (TP) of Yuan and Zhang [34], started from the leading eigenvector truncated to its p largest entries. All the final points of all the methods were verified. Table 6 shows the mean gap of each method and the number of problems in which it found the best value.

	Mean gap (%)	Best value
CPS AL	0.30	70
CPS AL_0	1.5	55
Telescopic AL	0.52	60
Telescopic AL_0	2.0	45
TP	3.3	21

Table 6: Sparse principal component analysis (90 problems): mean gap and number of problems in which each method found the best value.

CPS AL was better than Telescopic AL in 13 problems and equal in the other 77, and better than TP in 65 and worse in 2. The constraint-preserving swaps were accepted 22 times. All the runs took less than one second.

8 Conclusions

We considered the augmented Lagrangian method of Algencan in which the bounds and the partitioned cardinality constraints (in particular, complementarity constraints) are kept in the subproblems, which are solved by the Telescopic Method. The B-stationarity that the Telescopic Method guarantees for the subproblems is inherited by the limit points when the penalty

parameters remain bounded, but not in general. The reason is that the variation of the penalty term along a swap grows with the penalty parameter, so that the swaps of the subproblems may stop detecting descent directions of the original problem. Constraint-preserving swaps remove the dependence on the penalty parameter. With them, every feasible limit point at which PCC-LICQ (MPEC-LICQ for MPCCs) holds is B-stationary (S-stationary for MPCCs), whatever the behavior of the penalty parameters and of the safeguarded multipliers, which includes the quadratic penalty method. Without them, if the stopping criterion of the subproblems also includes a derivative test outside the current face, the feasible limit points are still M-stationary for MPCCs under MPEC-LICQ, as for other augmented Lagrangian methods that keep the complementarity constraints in the subproblems. We also showed that the modification is necessary under MPEC-LICQ, even with exact solutions of the subproblems and multipliers in the interior of the safeguarding box, and that this assumption cannot be replaced by MPEC-MFCQ, because descent may require simultaneous swaps in several blocks.

In the MacMPEC collection, with a light Phase I, the original swaps already stopped at S- or B-stationary points, and no constraint-preserving swap was accepted. Without Phase I, the constraint-preserving swaps improved the final points of the original swaps in five instances, three of which reached the best known value. The method found certified S- or B-stationary points in 185 of the 193 instances, more than the other methods tested, and its failures are related to Phase I, not to the swaps. With cardinality constraints (Section 7.3), on the other hand, the constraint-preserving swaps were often accepted and improved the final points of the original swaps, at the price of a larger number of candidate swaps.

Declaration of generative AI use

During the preparation of this work, the authors used Claude (Anthropic) to assist in drafting the manuscript, developing and checking the proofs, implementing the codes and running the numerical experiments. The authors reviewed and edited all the content and take full responsibility for the content of the publication.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, On augmented Lagrangian methods with general lower-level constraints, *SIAM Journal on Optimization* 18, pp. 1286–1309, 2008.

[2] R. Andreani, M. da Rosa, and L. D. Secchin, On the boundedness of multipliers in augmented Lagrangian methods for mathematical programs with complementarity constraints, Optimization Online, 2025.

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

[4] R. Andreani, G. Haeser, L. D. Secchin, and P. J. S. Silva, New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences, *SIAM Journal on Optimization* 29, pp. 3201–3230, 2019.

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

[6] A. Beck and Y. C. Eldar, Sparsity constrained nonlinear optimization: optimality conditions and algorithms, *SIAM Journal on Optimization* 23, pp. 1480–1509, 2013.

[7] A. Ben-Israel, A Newton–Raphson method for the solution of systems of equations, *Journal of Mathematical Analysis and Applications* 15, pp. 243–252, 1966.

[8] E. G. Birgin and J. M. Martínez, A Telescopic Method with guaranteed B-stationarity for partitioned cardinality and complementarity constraints, arXiv:2611.00897 [math.OC], 2026.

[9] E. G. Birgin and J. M. Martínez, Complexity and performance of an augmented Lagrangian algorithm, *Optimization Methods and Software* 35, pp. 885–920, 2020.

[10] E. G. Birgin and J. M. Martínez, Large-scale active-set box-constrained optimization method with spectral projected gradients, *Computational Optimization and Applications* 23, pp. 101–125, 2002.

[11] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, Philadelphia, 2014.

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

[13] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics in Applied Mathematics 16, SIAM, Philadelphia, 1996.

[14] R. Fletcher and S. Leyffer, Solving mathematical programs with complementarity constraints as nonlinear programs, *Optimization Methods and Software* 19, pp. 15–40, 2004.

[15] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, Local convergence of SQP methods for mathematical programs with equilibrium constraints, *SIAM Journal on Optimization* 17, pp. 259–286, 2006.

[16] A. Frangioni and C. Gentile, SDP diagonalizations and perspective cuts for a class of nonseparable MIQP, *Operations Research Letters* 35, pp. 181–185, 2007.

[17] D. M. Gay, Hooking your solver to AMPL, Technical Report 97-4-06, Computing Sciences Research Center, Bell Laboratories, 1997.

[18] L. Guo and Z. Deng, A new augmented Lagrangian method for MPCCs—theoretical and numerical comparison with existing augmented Lagrangian methods, *Mathematics of Operations Research* 47, pp. 1229–1246, 2022.

[19] T. Hoheisel, C. Kanzow, and A. Schwartz, Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints, *Mathematical Programming* 137, pp. 257–288, 2013.

[20] 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 Journal on Optimization* 22, pp. 1579–1606, 2012.

[21] X. Jia, C. Kanzow, P. Mehlitz, and G. Wachsmuth, An augmented Lagrangian method for optimization problems with structured geometric constraints, *Mathematical Programming* 199, pp. 1365–1415, 2023.

[22] C. Kanzow and A. Schwartz, The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited, *Mathematics of Operations Research* 40, pp. 253–275, 2015.

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

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

[25] C. Kirches, J. Larson, S. Leyffer, and P. Manns, Sequential linearization method for bound-constrained mathematical programs with complementarity constraints, *SIAM Journal on Optimization* 32, pp. 75–99, 2022.

[26] B. Legat, MacMPEC.jl, `https://github.com/blegat/MacMPEC.jl`.

[27] S. Leyffer, G. López-Calva, and J. Nocedal, Interior methods for mathematical programs with complementarity constraints, *SIAM Journal on Optimization* 17, pp. 52–77, 2006.

[28] S. Leyffer, MacMPEC: AMPL collection of MPECs, Argonne National Laboratory, 2000, `https://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC`.

[29] G.-H. Lin and M. Fukushima, Hybrid approach with active set identification for mathematical programs with complementarity constraints, *Journal of Optimization Theory and Applications* 128, pp. 1–28, 2006.

[30] A. Nurkanović and S. Leyffer, A globally convergent method for computing B-stationary points of mathematical programs with equilibrium constraints, *Computational Optimization and Applications*, 2026, DOI 10.1007/s10589-026-00832-x.

[31] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

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

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

[34] X.-T. Yuan and T. Zhang, Truncated power method for sparse eigenvalue problems, *Journal of Machine Learning Research* 14, pp. 899–925, 2013.