

An Augmented Lagrangian Method for Cardinality-Constrained Optimization Problems

Christian Kanzow* Andreas B. Raharja* Alexandra Schwartz†

December 14, 2020

Abstract

A reformulation of cardinality-constrained optimization problems into continuous nonlinear optimization problems with an orthogonality-type constraint has gained some popularity during the last few years. Due to the special structure of the constraints, the reformulation violates many standard assumptions and therefore is often solved using specialized algorithms. In contrast to this, we investigate the viability of using a standard safeguarded multiplier penalty method without any problem-tailored modifications to solve the reformulated problem. We prove global convergence towards an (essentially strongly) stationary point under a suitable problem-tailored quasinormality constraint qualification. Numerical experiments illustrating the performance of the method in comparison to regularization-based approaches are provided.

1 Introduction

In recent years, *cardinality-constrained optimization problems* (CC) have received an increasing amount of attention due to their far-reaching applications, including portfolio optimization [10, 11, 14] and statistical regression [10, 20]. Unfortunately, these problems are notoriously difficult to solve, even testing feasibility is already NP-complete [10].

A recurrent strategy in mathematics is to cast a difficult problem into a simpler one, for which well-established solution techniques already exist. For CC, the recent paper [16] was written precisely in this spirit. There, the authors reformulate the problem as a continuous optimization problem with orthogonality-type constraints. This approach parallels the one made in the context of *sparse optimization problems* [21]. It should be noted, however, that due to its similarities with *mathematical programs with complementarity constraints* (MPCC), the proposed reformulation from [16] is, unfortunately, also highly degenerate in the sense that even the standard *Abadie constraint qualification* (ACQ), which is a very weak CQ, is often violated at points of interest. In addition, sequential optimality conditions like AKKT (approximate KKT) are known to be satisfied at *any* feasible point of cardinality-constrained problems, see [30], and therefore are also useless to identify suitable candidates for local minima in this context.

*Institute of Mathematics, University of Würzburg, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany; E-mails: {kanzow, andreas.raharja}@mathematik.uni-wuerzburg.de

†Fachbereich Mathematik, Technische Universität Darmstadt, Dolivostraße 15, 64293 Darmstadt, Germany; E-mail: alexandra.schwartz@tu-darmstadt.de

These observations make a direct application of most standard *nonlinear programming* (NLP) methods to solve the reformulated problem rather challenging, since they typically require the fulfillment of a CQ stronger than ACQ at a limit point to ensure stationarity. To overcome difficulties with CQs, CC-tailored CQs were introduced in [16, 34]. Regularization methods, which are standard techniques in attacking MPCC, were subsequently proposed in [14, 16], where convergence towards a stationarity point is proved using these CQs. This is not the path that we shall walk on here. In this paper, we are interested in the viability of ALGENCAN [2, 3, 12], a well established and open-source standard NLP solver based on an *augmented Lagrangian method* (ALM), to solve the reformulated problem directly without any problem-specific modifications.

ALMs belong to one of the classical solution methods for NLPs. However, up to the mid 2000s, their popularity was largely overshadowed by other techniques, in particular, the *sequential quadratic programming* methods (SQP) and the *interior point methods*. Since then, beginning with [2, 3], a particular variant of ALMs, which employs the *Powell-Hestenes-Rockafellar* (PHR) augmented Lagrangian function as well as safeguarded multipliers, has been experiencing rejuvenated interest. The aforementioned ALGENCAN implements this variant. For NLPs it has been shown that this variant possesses strong convergence properties even under very mild assumptions [4, 5]. It has since been applied to solve various other problems, including MPCC [6, 24], *quasi-variational inequalities* [25], *generalized Nash equilibrium problems* [15, 27], and *semidefinite programming* [7, 13].

Due to the structure of the reformulated problems, particularly relevant to us is, of course, the paper [24]. There, the authors proved the global convergence of the method towards an MPCC-C-stationarity point under MPCC-LICQ. However, even though the problems with orthogonality-type constraints resulting from the reformulation of CC can be viewed as MPCC in case nonnegativity constraints are present [34], we would like to stress that the results obtained in our paper are not simple corollaries of [24]. For one, we do not assume the presence of nonnegativity constraints here, making our results applicable in the general setting. Moreover, even in the presence of nonnegativity constraints, it was shown in [34, Remark 5.7 (f)] that MPCC-LICQ, which was used to guarantee convergence to a stationary point in [24], is often violated at points of interests for the reformulated problems. Instead, we therefore employ a CC-analogue of the quasinormality CQ [9], which is weaker than CC-CPLD introduced in [16], to prove the global convergence of the method.

To this end, we first recall some important properties of the CC-reformulation in Section 2 and define a CC-version of the quasinormality CQ. The ALM algorithm is introduced in Section 3, and its convergence properties under said quasinormality CQ are analyzed in Section 4. Numerical experiments illustrating the performance of ALGENCAN for the reformulated problem are then presented in Section 5. We close with some final remarks in Section 6.

Notation: For a given vector $x \in \mathbb{R}^n$, we define the two index sets

$$I_{\pm}(x) := \{i \in \{1, \dots, n\} \mid x_i \neq 0\} \quad \text{and} \quad I_0(x) := \{i \in \{1, \dots, n\} \mid x_i = 0\}.$$

Clearly, both sets are disjoint and we have $\{1, \dots, n\} = I_{\pm}(x) \cup I_0(x)$. For two vectors $a, b \in \mathbb{R}^n$, the terms $\max\{a, b\}, \min\{a, b\} \in \mathbb{R}^n$ denote the componentwise maximum/minimum of these vectors. A frequently used special case hereof is $a_+ := \max\{a, 0\} \in \mathbb{R}^n$. We denote the Hadamard product of two vectors $x, y \in \mathbb{R}^n$ with $x \circ y$, and we define $e := (1, \dots, 1)^T \in \mathbb{R}^n$.

2 Preliminaries

In this paper, we consider *cardinality-constrained optimization problems* of the form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad & g(x) \leq 0, \quad h(x) = 0, \\ & \|x\|_0 \leq s, \end{aligned} \tag{2.1}$$

where $f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, $h \in C^1(\mathbb{R}^n, \mathbb{R}^p)$, and $\|x\|_0$ denotes the number of nonzero components of a vector x . Occasionally, this problem is also called a *sparse optimization problem* [35], but sparse optimization typically refers to programs, which have a sparsity term within the objective function.

Throughout this paper, we assume $s < n$, since the cardinality constraint would be redundant otherwise. Following the approach from [16], by introducing an auxiliary variable $y \in \mathbb{R}^n$, we obtain the following relaxed program

$$\begin{aligned} \min_{x, y \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad & g(x) \leq 0, \quad h(x) = 0, \\ & n - e^T y \leq s, \\ & y \leq e, \\ & x \circ y = 0. \end{aligned} \tag{2.2}$$

Observe that the relaxed reformulation we use here is slightly different from the one in [16], because we omit the constraint $y \geq 0$, leading to a larger feasible set. Nonetheless, one can easily see that all results obtained in [16, Section 3] are applicable for (2.2) as well. We shall now gather some of these results, which are relevant for this paper. Their proofs can be found in [16].

Theorem 2.1. *Let $\hat{x} \in \mathbb{R}^n$. Then the following statements hold:*

- (a) \hat{x} is feasible for (2.1) if and only if there exists $\hat{y} \in \mathbb{R}^n$ such that (\hat{x}, \hat{y}) is feasible for (2.2).
- (b) \hat{x} is a global optimizer of (2.1) if and only if there exists $\hat{y} \in \mathbb{R}^n$ such that (\hat{x}, \hat{y}) is a global minimizer of (2.2).
- (c) If $\hat{x} \in \mathbb{R}^n$ is a local minimizer of (2.1), then there exists $\hat{y} \in \mathbb{R}^n$ such that (\hat{x}, \hat{y}) is a local minimizer of (2.2). Conversely, if (\hat{x}, \hat{y}) is a local minimizer of (2.2) satisfying $\|\hat{x}\|_0 = s$, then \hat{x} is a local minimizer of (2.1).

Theorem 2.1 shows that the relaxed problem (2.2) is equivalent to the original problem (2.1) in terms of feasible points and global minima, whereas the equivalence of local minima requires some extra condition (namely the cardinality constraint to be active). Hence, essentially, the two problems (2.1) and (2.2) may be viewed as being equivalent, and it is therefore natural to solve the given cardinality problem (2.1) via the relaxed program (2.2).

Let us now recall the stationarity concepts introduced in [16].

Definition 2.2. Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be feasible for (2.2). Then (\hat{x}, \hat{y}) is called

(a) CC-M-stationary, if there exist multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that

- $0 = \nabla f(\hat{x}) + \nabla g(\hat{x})\lambda + \nabla h(\hat{x})\mu + \gamma$,
- $\lambda \geq 0$ and $\lambda_i g_i(\hat{x}) = 0$ for all $i = 1, \dots, m$,
- $\gamma_i = 0$ for all $i \in I_{\pm}(\hat{x})$.

(b) CC-S-stationary, if (\hat{x}, \hat{y}) is CC-M-stationary with $\gamma_i = 0$ for all $i \in I_0(\hat{y}) := \{i \mid \hat{y}_i = 0\}$.

As remarked in [16], CC-S-stationarity corresponds to the KKT condition of (2.2). In contrast, CC-M-stationarity does not depend on the auxiliary variables y and is the KKT condition of the following *tightened nonlinear program* TNLP(\hat{x})

$$\begin{aligned} \min_x f(x) \quad \text{s.t.} \quad & g(x) \leq 0, \quad h(x) = 0, \\ & x_i = 0 \quad \forall i \in I_0(\hat{x}). \end{aligned} \tag{2.3}$$

Observe that every local minimizer of (2.1) is also a local minimizer of (2.3). This justifies the definition of CC-M-stationarity. Suppose now that $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ is feasible for (2.2). By the orthogonality constraint, we clearly have $I_{\pm}(\hat{x}) \subseteq I_0(\hat{y})$ (with equality if $\|\hat{x}\|_0 = s$). Hence, if (\hat{x}, \hat{y}) is a CC-S-stationary point, then it is also CC-M-stationary. The converse is not true in general, see [16, Example 4].

It was shown in [34] that the standard Guignard CQ (GCQ) for (2.2) is equivalent to the tailored CC-GCQ, and thus local minima of (2.2) are CC-S-stationary under all CQs, that imply CC-GCQ. This is a major difference to MPCCs, where one typically needs MPCC-LICQ to guarantee S-stationarity of local minima and has to rely on M-stationarity under weaker MPCC-CQs. Since local minima of (2.2) are CC-S-stationary under CC-CQs, CC-M-stationary points seem to be undesirable solution candidates. Fortunately, if (\hat{x}, \hat{y}) is CC-M-stationary, one can simply replace \hat{y} with another auxiliary variable $\hat{z} \in \mathbb{R}^n$ such that (\hat{x}, \hat{z}) is CC-S-stationary, as the next proposition shows. Note that the proof of this result is constructive.

Proposition 2.3. Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be feasible for (2.2). If (\hat{x}, \hat{y}) is a CC-M-stationary point, then there exists $\hat{z} \in \mathbb{R}^n$ such that (\hat{x}, \hat{z}) is CC-S-stationary.

Proof. By Theorem 2.1, \hat{x} is feasible for (2.1). Now define $\hat{z} \in \mathbb{R}^n$ such that

$$\hat{z}_i := \begin{cases} 0 & \text{if } i \in I_{\pm}(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Then (\hat{x}, \hat{z}) is obviously feasible for (2.2), cf. also the proof of [16, Theorem 3.1]. By assumption, there exists $(\lambda, \mu, \gamma) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n$ such that (\hat{x}, \hat{y}) is CC-M-stationary. And since $I_{\pm}(\hat{x}) = I_0(\hat{z})$, using Definition 2.2, we can conclude that (\hat{x}, \hat{z}) is CC-S-stationary with (λ, μ, γ) from before as corresponding multipliers. ■

This shows that the difference between S- and M-stationarity in this setting is not as big as for MPCCs. More precisely, a feasible point \hat{x} of (2.1) is CC-M-stationary if and only if there exists \hat{z} such that the pair (\hat{x}, \hat{z}) is CC-S-stationary. Consequently, any constraint qualification which guarantees that a local minimum \hat{x} of (2.1) satisfies CC-M-stationarity, also yields the existence of a CC-S-stationary point (\hat{x}, \hat{z}) . Numerically, it implies that any method which generates a sequence converging to a CC-M-stationary point only, essentially gives a CC-S-stationary point.

Utilizing (2.3), CC-tailored CQs were introduced in [16]. We shall now follow this approach and introduce a CC-tailored *quasinormality* condition.

Definition 2.4. A point $\hat{x} \in \mathbb{R}^n$, feasible for (2.1), satisfies the CC-quasinormality condition, if there exist $\mathbf{no} (\lambda, \mu, \gamma) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n \setminus \{(0, 0, 0)\}$ such that the following conditions are satisfied:

- (a) $0 = \nabla g(\hat{x})\lambda + \nabla h(\hat{x})\mu + \gamma$,
- (b) $\lambda \geq 0$ and $\lambda_i g_i(\hat{x}) = 0$ for all $i = 1, \dots, m$,
- (c) $\gamma_i = 0$ for all $i \in I_{\pm}(\hat{x})$,
- (d) $\exists \{x^k\} \subseteq \mathbb{R}^n$ with $\{x^k\} \rightarrow \hat{x}$ such that, for all $k \in \mathbb{N}$, we have
 - $\forall i \in \{1, \dots, m\}$ with $\lambda_i > 0$: $\lambda_i g_i(x^k) > 0$,
 - $\forall i \in \{1, \dots, p\}$ with $\mu_i \neq 0$: $\mu_i h_i(x^k) > 0$,
 - $\forall i \in \{1, \dots, n\}$ with $\gamma_i \neq 0$: $\gamma_i x_i^k > 0$.

Obviously, CC-quasinormality corresponds to the (standard) quasinormality CQ of (2.3). By [1], CC-CPLD introduced in [16] thus implies CC-quasinormality.

3 An Augmented Lagrangian Method

Let us now describe the algorithm. For a given penalty parameter $\alpha > 0$ the *PHR augmented Lagrangian function* for (2.2) is given by

$$L((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha) := f(x) + \alpha \pi((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha)$$

with $(\lambda, \mu, \zeta, \eta, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+^n \times \mathbb{R}^n$ and

$$\pi((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha) := \frac{1}{2} \left\| \begin{pmatrix} (g(x) + \frac{\lambda}{\alpha})_+ \\ h(x) + \frac{\mu}{\alpha} \\ (n - e^T y - s + \frac{\zeta}{\alpha})_+ \\ (y - e + \frac{\eta}{\alpha})_+ \\ x \circ y + \frac{\gamma}{\alpha} \end{pmatrix} \right\|_2^2,$$

is the shifted quadratic penalty term, cf. [12, Chapter 4]. The algorithm is then stated below.

Algorithm 3.1 (Safeguarded Augmented Lagrangian Method).

(S₀) **Initialization:** Choose parameters $\lambda_{\max} > 0$, $\mu_{\min} < \mu_{\max}$, $\zeta_{\max} > 0$, $\eta_{\max} > 0$, $\gamma_{\min} < \gamma_{\max}$, $\tau \in (0, 1)$, $\sigma > 1$ and $\{\epsilon_k\} \subseteq \mathbb{R}_+$ such that $\{\epsilon_k\} \downarrow 0$.

Choose initial values $\bar{\lambda}^1 \in [0, \lambda_{\max}]^m$, $\bar{\mu}^1 \in [\mu_{\min}, \mu_{\max}]^p$, $\bar{\zeta}^1 \in [0, \zeta_{\max}]$, $\bar{\eta}^1 \in [0, \eta_{\max}]^n$, $\bar{\gamma}^1 \in [\gamma_{\min}, \gamma_{\max}]^n$, $\alpha_1 > 0$, and set $k \leftarrow 1$.

(S₁) **Update of the iterates:** Compute (x^k, y^k) as an approximate solution of

$$\min_{(x,y) \in \mathbb{R}^{2n}} L((x, y), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k)$$

satisfying

$$\|\nabla_{(x,y)} L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k)\| \leq \epsilon_k. \quad (3.1)$$

(S₂) **Update of the approximate multipliers:**

$$\begin{aligned} \lambda^k &:= (\alpha_k g(x^k) + \bar{\lambda}^k)_+ \\ \mu^k &:= \alpha_k h(x^k) + \bar{\mu}^k \\ \zeta^k &:= (\alpha_k (n - e^T y^k - s) + \bar{\zeta}^k)_+ \\ \eta^k &:= (\alpha_k (y^k - e) + \bar{\eta}^k)_+ \\ \gamma^k &:= \alpha_k x^{k \circ} y^k + \bar{\gamma}^k \end{aligned}$$

(S₃) **Update of the penalty parameter:** Define

$$U^k := \min \left\{ -g(x^k), \frac{\bar{\lambda}^k}{\alpha_k} \right\}, \quad V_k := \min \left\{ -(n - e^T y^k - s), \frac{\bar{\zeta}^k}{\alpha_k} \right\}, \quad W^k := \min \left\{ (y^k - e), \frac{\bar{\eta}^k}{\alpha_k} \right\}.$$

If $k = 1$ or

$$\begin{aligned} &\max \left\{ \|U^k\|, \|h(x^k)\|, \|V_k\|, \|W^k\|, \|x^{k \circ} y^k\| \right\} \\ &\leq \tau \max \left\{ \|U^{k-1}\|, \|h(x^{k-1})\|, \|V_{k-1}\|, \|W^{k-1}\|, \|x^{k-1 \circ} y^{k-1}\| \right\}, \end{aligned} \quad (3.2)$$

set $\alpha_{k+1} = \alpha_k$. Otherwise set $\alpha_{k+1} = \sigma \alpha_k$.

(S₄) **Update of the safeguarded multipliers:** Choose $\bar{\lambda}^{k+1} \in [0, \lambda_{\max}]^m$, $\bar{\mu}^{k+1} \in [\mu_{\min}, \mu_{\max}]^p$, $\bar{\zeta}^{k+1} \in [0, \zeta_{\max}]$, $\bar{\eta}^{k+1} \in [0, \eta_{\max}]^n$, $\bar{\gamma}^{k+1} \in [\gamma_{\min}, \gamma_{\max}]^n$.

(S₅) Set $k \leftarrow k + 1$ and go to (S₁).

Note that Algorithm 3.1 is exactly the safeguarded augmented Lagrangian method from [12]. The only difference to the classical augmented Lagrangian, see, e.g., [8, 32], is in the more careful updating of the Lagrange multipliers: The safeguarded method contains the bounded auxiliary sequences $\bar{\lambda}^k, \bar{\mu}^k, \dots$, which replace the multiplier estimates λ^k, μ^k, \dots in certain places. Note that these bounded auxiliary sequences are chosen by the user and that there is quite some freedom for their choice. In principle, one can simply take $\bar{\lambda}^k = 0, \bar{\mu}^k = 0, \dots$ for all $k \in \mathbb{N}$, in which case Algorithm 3.1 boils down to the classical quadratic penalty method. A more practical choice is to compute $\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}, \dots$ by taking the projections of the multiplier

estimates λ^k, μ^k, \dots onto the respective sets $[0, \lambda_{\max}]^m, [\mu_{\min}, \mu_{\max}]^p, \dots$. This implies that, for sufficiently loose parameters $\lambda_{\max}, \mu_{\min}, \mu_{\max}, \dots$ the safeguarded ALM often coincides with the classical ALM. Differences occur, however, in those situations where the classical ALM generates unbounded Lagrange multiplier estimates. This has a significant influence on the (global) convergence theory of both methods: While there is a very satisfactory theory for the safeguarded method, see [12], a counterexample from [28] shows that these properties do not hold for the classical approach.

We have not specified a termination condition for the algorithm here. However, the convergence analysis in the next section suggests to stop the algorithm, e.g., if the M-stationarity conditions are satisfied up to a given tolerance.,

In the subsequent discussion of the convergence properties of this algorithm, we often make use of the fact that the PHR augmented Lagrangian function is continuously differentiable with the gradient

$$\begin{aligned}\nabla_x L((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha) &= \nabla f(x) + \alpha \left[\nabla g(x) \left(g(x) + \frac{\lambda}{\alpha} \right)_+ + \nabla h(x) \left(h(x) + \frac{\mu}{\alpha} \right) + \left(x \circ y + \frac{\gamma}{\alpha} \right) \circ y \right], \\ \nabla_y L((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha) &= \alpha \left[- \left(n - e^T y - s + \frac{\zeta}{\alpha} \right)_+ e + \left(y - e + \frac{\eta}{\alpha} \right)_+ e + \left(x \circ y + \frac{\gamma}{\alpha} \right) \circ x \right],\end{aligned}$$

where $\nabla g(x)$ and $\nabla h(x)$ denote the transposed Jacobian matrices of g and h at x , respectively. Consequently, the multipliers in (S_2) are chosen exactly such that

$$\begin{aligned}\nabla_x L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) &= \nabla f(x) + \nabla g(x^k) \bar{\lambda}^k + \nabla h(x^k) \bar{\mu}^k + \gamma^k \circ y^k, \\ \nabla_y L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) &= -\bar{\zeta}^k e + \bar{\eta}^k + \gamma^k \circ x^k\end{aligned}$$

holds for all $k \in \mathbb{N}$.

4 Convergence Analysis

The aim of this section is to prove global convergence of Algorithm 3.1 to CC-M-stationary points under the fairly mild CC-quasinormality condition. To this end, we begin with an auxiliary result, which states that the sequence $\{y^k\}$ remains bounded on any subsequence, where $\{x^k\}$ itself is bounded. In particular, if $\{x^k\}$ converges on a subsequence, this then allows us to extract a limit point of the sequence $\{(x^k, y^k)\}$.

Proposition 4.1. *Let $\{x^k\} \subseteq \mathbb{R}^n$ be a sequence generated by Algorithm 3.1. Assume that $\{x^k\}$ is bounded on a subsequence. Then the auxiliary sequence $\{y^k\}$ is bounded on the same subsequence.*

Proof. In order to avoid taking further subsequences, let us assume that the entire sequence $\{x^k\}$ remains bounded. We then show that also the whole sequence $\{y^k\}$ is bounded. Define, for each $k \in \mathbb{N}$,

$$B^k := \nabla_y L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) = -\bar{\zeta}^k e + \bar{\eta}^k + \gamma^k \circ x^k. \quad (4.1)$$

By (3.1), we know that $\{B^k\} \rightarrow 0$. We first show that the sequence $\{y^k\}$ is bounded from above and then verify that it is also bounded from below.

$\{y^k\}$ is bounded above: We claim that there exists a $c \in \mathbb{R}$ such that $y^k \leq ce$ for all $k \in \mathbb{N}$. Suppose, by contradiction, that there is an index $j \in \{1, \dots, n\}$ and a subsequence $\{y_j^{k_l}\}$ such that $\{y_j^{k_l}\} \rightarrow +\infty$. Since $\alpha_k \geq \alpha_1 > 0$ for all $k \in \mathbb{N}$ and $\bar{\eta}_j^{k_l}$ is bounded by definition, we then obtain

$$\{\alpha_{k_l}(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l}\} \rightarrow +\infty. \quad (4.2)$$

This implies $\eta_j^{k_l} = \alpha_{k_l}(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l}$ for all $l \in \mathbb{N}$ sufficiently large and, hence, by (4.2), we have $\{\eta_j^{k_l}\} \rightarrow +\infty$. Observe that, for each $l \in \mathbb{N}$ sufficiently large, we have

$$y_j^{k_l} x_j^{k_l} = (\alpha_{k_l} x_j^{k_l} y_j^{k_l} + \bar{y}_j^{k_l}) x_j^{k_l} = \alpha_{k_l} (x_j^{k_l})^2 y_j^{k_l} + \bar{y}_j^{k_l} x_j^{k_l} \geq \bar{y}_j^{k_l} x_j^{k_l}.$$

From (4.1), we then obtain for these $l \in \mathbb{N}$ that $B_j^{k_l} = -\zeta^{k_l} + \eta_j^{k_l} + y_j^{k_l} x_j^{k_l} \geq -\zeta^{k_l} + \eta_j^{k_l} + \bar{y}_j^{k_l} x_j^{k_l}$, which is equivalent to $\zeta^{k_l} \geq \eta_j^{k_l} + \bar{y}_j^{k_l} x_j^{k_l} - B_j^{k_l}$. Since $\{B_j^{k_l}\} \rightarrow 0$ and $\{\bar{y}_j^{k_l} x_j^{k_l}\}$ is bounded, the right-hand side converges to $+\infty$. Consequently, we have $\{\zeta^{k_l}\} \rightarrow +\infty$. The definition of $\{\zeta^{k_l}\}$ therefore yields $\{\alpha_{k_l}(n - e^T y^{k_l} - s) + \bar{\zeta}_{k_l}\} \rightarrow +\infty$. Since $\{\bar{\zeta}_{k_l}\}$ is a bounded sequence, we get $\{\alpha_{k_l}(n - e^T y^{k_l} - s)\} \rightarrow +\infty$. We therefore have

$$n - e^T y^{k_l} - s > 0 \quad \forall l \in \mathbb{N} \text{ sufficiently large.} \quad (4.3)$$

We now claim that

$$\exists i \in \{1, \dots, n\} \setminus \{j\} : \{y_i^{k_l}\} \text{ is unbounded from below.} \quad (4.4)$$

Assume there exist $d \in \mathbb{R}$ such that $y_i^{k_l} \geq d$ for all $i \in \{1, \dots, n\} \setminus \{j\}$ and all $l \in \mathbb{N}$. We then obtain

$$n - e^T y^{k_l} - s = n - \sum_{i=1, i \neq j}^n y_i^{k_l} - y_j^{k_l} - s \leq n - (n-1)d - y_j^{k_l} - s \rightarrow -\infty.$$

We therefore get $n - e^T y^{k_l} - s < 0$ for all $l \in \mathbb{N}$ sufficiently large, but this contradicts (4.3), hence (4.4) holds. For this particular index i , we can construct a subsequence $\{y_i^{k_{l_t}}\}$ such that $\{y_i^{k_{l_t}}\} \rightarrow -\infty$. Since $\{\bar{\eta}_i\}^{k_{l_t}}$ is bounded, we then have $\{\alpha_{k_{l_t}}(y_i^{k_{l_t}} - 1) + \bar{\eta}_i^{k_{l_t}}\} \rightarrow -\infty$. This implies $\eta_i^{k_{l_t}} = 0$ for all $t \in \mathbb{N}$ sufficiently large. We therefore obtain from (4.1) that

$$\begin{aligned} B_i^{k_{l_t}} &= -\zeta^{k_{l_t}} + \eta_i^{k_{l_t}} + y_i^{k_{l_t}} x_i^{k_{l_t}} = -\zeta^{k_{l_t}} + y_i^{k_{l_t}} x_i^{k_{l_t}} = -\zeta^{k_{l_t}} + (\alpha_{k_{l_t}} x_i^{k_{l_t}} y_i^{k_{l_t}} + \bar{y}_i^{k_{l_t}}) x_i^{k_{l_t}} \\ &= -\zeta^{k_{l_t}} + \alpha_{k_{l_t}} (x_i^{k_{l_t}})^2 y_i^{k_{l_t}} + \bar{y}_i^{k_{l_t}} x_i^{k_{l_t}} \leq -\zeta^{k_{l_t}} + \bar{y}_i^{k_{l_t}} x_i^{k_{l_t}} \end{aligned}$$

for all $t \in \mathbb{N}$ large enough. Since $\{\bar{y}_i^{k_{l_t}} x_i^{k_{l_t}}\}$ is a bounded sequence and $\{\zeta^{k_{l_t}}\} \rightarrow +\infty$, we get $\{B_i^{k_{l_t}}\} \rightarrow -\infty$, which leads to a contradiction. Thus, $\{y^k\}$ is bounded above.

$\{y^k\}$ is bounded below: We claim that there exists a $d \in \mathbb{R}$ such that $y^k \geq de$ for all $k \in \mathbb{N}$. Assume, by contradiction, that there is an index $j \in \{1, \dots, n\}$ such that $\{y_j^{k_l}\} \rightarrow -\infty$ on a suitable subsequence. Then, we have $y_j^{k_l} < 0$ and $\eta_j^{k_l} = 0$ for all $l \in \mathbb{N}$ large enough, and similar to the previous case, it therefore follows that $B_j^{k_l} \leq -\zeta^{k_l} + \bar{y}_j^{k_l} x_j^{k_l}$. This can be rewritten as $\zeta^{k_l} \leq \bar{y}_j^{k_l} x_j^{k_l} - B_j^{k_l}$. Since $\{\bar{y}_j^{k_l} x_j^{k_l}\}$ is bounded and $\{B_j^{k_l}\} \rightarrow 0$, the sequence $\{\bar{y}_j^{k_l} x_j^{k_l} - B_j^{k_l}\}$ is bounded. This implies, in particular, that $\{\zeta^{k_l}\}$ is bounded above, i.e.,

$$\exists r \in \mathbb{R} \forall l \in \mathbb{N} : \zeta^{k_l} \leq r. \quad (4.5)$$

On the other hand, we already know $y^k \leq ce$ for all $k \in \mathbb{N}$. We therefore get

$$n - e^T y^{k_l} - s \geq n - (n-1)c - y_j^{k_l} - s \rightarrow +\infty.$$

This implies

$$\{\alpha_{k_l} (n - e^T y^{k_l} - s) + \bar{\zeta}_{k_l}\} \rightarrow +\infty$$

due to the boundedness of the sequence $\{\bar{\zeta}_{k_l}\}$ and $\alpha_k \geq \alpha_1 > 0$ for all $k \in \mathbb{N}$. The definition of ζ^{k_l} there yields

$$\zeta^{k_l} = \alpha_{k_l} (n - e^T y^{k_l} - s) + \bar{\zeta}_{k_l} \rightarrow +\infty,$$

which contradicts (4.5). Hence, $\{y^k\}$ is bounded below. ■

As for all penalty-type methods, one has to distinguish two aspects in a corresponding global convergence theory, namely the feasibility issue and an optimality statement. Without further assumptions, feasibility of the limit point cannot be guaranteed (for nonconvex constraints). However, there is a standard result in [12], which shows that the limit point of our stationary sequence is at least a stationary point of the constraint violation. To this end, we measure the infeasibility of a point $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ for (2.2) by using the unshifted quadratic penalty term

$$\pi_{0,1}(x, y) := \pi((x, y), 0, 0, 0, 0, 0; 1).$$

Clearly (\hat{x}, \hat{y}) is feasible for (2.2) if and only if $\pi_{0,1}(\hat{x}, \hat{y}) = 0$. This, in turn, implies that (\hat{x}, \hat{y}) minimizes $\pi_{0,1}(x, y)$. In particular, we then ought to have $\nabla \pi_{0,1}(\hat{x}, \hat{y}) = 0$.

Theorem 4.2. *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a limit point of the sequence $\{(x^k, y^k)\}$ generated by Algorithm 3.1. Then $\nabla \pi_{0,1}(\hat{x}, \hat{y}) = 0$.*

We omit the proof here, since it is identical to [12, Theorem 6.3] and [29, Theorem 6.2]. Instead, we turn to an optimality result for Algorithm 3.1. Suppose that the sequence $\{x^k\}$ generated by Algorithm 3.1 has a limit point \hat{x} . Proposition 4.1 then suggests that we can extract a limit point (\hat{x}, \hat{y}) of the sequence $\{(x^k, y^k)\}$. Under the additional assumptions that \hat{x} satisfies CC-quasinormality and (\hat{x}, \hat{y}) is feasible for (2.2), we can show that (\hat{x}, \hat{y}) is a CC-M-stationary point.

Theorem 4.3. *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a limit point of $\{(x^k, y^k)\}$ that is feasible for (2.2) and where \hat{x} satisfies CC-quasinormality. Then (\hat{x}, \hat{y}) is a CC-M-stationary point.*

Proof. To simplify the notation, we assume, throughout this proof, that the entire sequence $\{(x^k, y^k)\}$ converges to (\hat{x}, \hat{y}) . For each $k \in \mathbb{N}$, we define

$$A^k := \nabla_x L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}^k, \bar{\eta}^k, \bar{y}^k; \alpha_k) = \nabla f(x^k) + \nabla g(x^k) \lambda^k + \nabla h(x^k) \mu^k + \gamma^{k \circ} y^k.$$

Furthermore, let B^k be given as in (4.1). By (3.1) and since $\{\epsilon_k\} \downarrow 0$, we know that $\{A^k\} \rightarrow 0$ and $\{B^k\} \rightarrow 0$. Observe that, by (S_2) , we have $\{\lambda^k\} \subseteq \mathbb{R}_+^m$. Furthermore, by (S_3) , the sequence

of penalty parameters $\{\alpha_k\}$ satisfies $\alpha_k \geq \alpha_1 > 0$ for all $k \in \mathbb{N}$. Let us now distinguish two cases.

Case 1: $\{\alpha_k\}$ is bounded. Then $\{\alpha_k\}$ is eventually constant, say $\alpha_k = \alpha_K$ for all $k \geq K$ with some sufficiently large $K \in \mathbb{N}$. Now, let us take a closer look at (S_2) . The boundedness of $\{\alpha_k\}$ immediately implies that the sequences $\{\mu^k\}$ and $\{\gamma^{k \circ} y^k\}$ are bounded. By passing onto subsequences if necessary, we can assume w.l.o.g. that these sequences converge, i.e. $\{\mu^k\} \rightarrow \hat{\mu}$ and $\{\gamma^{k \circ} y^k\} \rightarrow \hat{y}$. For all $i \in I_{\pm}(\hat{x})$ the feasibility of (\hat{x}, \hat{y}) implies $\hat{y}_i = 0$. Since, in this case, we have $\{y_i^k\} \rightarrow 0$, it follows that

$$\hat{y}_i = \lim_{k \rightarrow \infty} \gamma_i^k y_i^k = \lim_{k \rightarrow \infty} \alpha_k x_i^k (y_i^k)^2 + \lim_{k \rightarrow \infty} \bar{\gamma}_i^k y_i^k = \alpha_K \cdot 0 + \lim_{k \rightarrow \infty} \bar{\gamma}_i^k y_i^k = 0 \quad \forall i \in I_{\pm}(\hat{x}).$$

Next, observe that, for each $i \in \{1, \dots, m\}$, we have $0 \leq \lambda_i^k \leq |\alpha_k g_i(x^k) + \bar{\lambda}_i^k|$ for all $k \in \mathbb{N}$. Thus, $\{\lambda_i^k\}$ is bounded as well and has a convergent subsequence. Thus, we can assume w.l.o.g. that $\{\lambda^k\} \rightarrow \hat{\lambda}$ on the whole sequence. Now, the boundedness of $\{\alpha_k\}$ and (S_3) also imply $\{\|U^k\|\} \rightarrow 0$. Let $i \notin I_g(\hat{x})$. Since, by definition, $\{\bar{\lambda}^k\}$ is bounded, $\left\{\frac{\bar{\lambda}_i^k}{\alpha_k}\right\}$ is bounded as well and therefore has a convergent subsequence. Assume w.l.o.g. that this sequence converges to some limit point a_i . Then

$$0 = \lim_{k \rightarrow \infty} \|U_i^k\| = \|\min\{-g_i(\hat{x}), a_i\}\| \quad \Rightarrow \quad \min\{-g_i(\hat{x}), a_i\} = 0.$$

Since $-g_i(\hat{x}) > 0$, we get $a_i = 0$. This implies

$$\left\{g_i(x^k) + \frac{\bar{\lambda}_i^k}{\alpha_k}\right\} \rightarrow g_i(\hat{x}) + a_i = g_i(\hat{x}) < 0.$$

Thus, by (S_2) we have

$$\lambda_i^k = \max\{0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k\} = 0 \quad \forall k \in \mathbb{N} \text{ sufficiently large.} \quad (4.6)$$

As its limit, we then also have $\hat{\lambda}_i = 0$. Letting $k \rightarrow \infty$, the definition of A^k then yields

$$0 = \nabla f(\hat{x}) + \nabla g(\hat{x})\hat{\lambda} + \nabla h(\hat{x})\hat{\mu} + \hat{y}.$$

Altogether, it follows that (\hat{x}, \hat{y}) is a CC-M-stationary point.

Case 2: $\{\alpha_k\}$ is unbounded. Then, we have $\{\alpha_k\} \rightarrow +\infty$. Now define, for each $k \in \mathbb{N}$,

$$\tilde{\gamma}_i^k := \gamma_i^k y_i^k \quad \forall i \in \{1, \dots, n\}.$$

We claim that the sequence $\{(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\}$ is bounded. By contradiction, assume that $\{(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\} \rightarrow \infty$, w.l.o.g. on the whole sequence. The corresponding normalized sequence $\left\{\frac{(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|}\right\}$ is bounded and therefore, again w.l.o.g. on the whole sequence, convergent to a (nontrivial) limit, i.e.

$$\left\{\frac{(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|}\right\} \rightarrow (\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}, \tilde{\zeta}, \tilde{\eta}) \neq 0.$$

We show that this limit, together with the sequence $\{x^k\}$, contradicts CC-quasinormality in \hat{x} : Since $\lambda^k \geq 0$ for all k , it follows that $\tilde{\lambda} \geq 0$. Now, take an index $i \notin I_g(\hat{x})$, i.e. $g_i(\hat{x}) < 0$. Since

$\{\bar{\lambda}_i^k\}$ is bounded, it follows that $\{\alpha_k g_i(x^k) + \bar{\lambda}_i^k\} \rightarrow -\infty$. This implies $\lambda_i^k = 0$ for all $k \in \mathbb{N}$ sufficiently large, hence we get

$$\tilde{\lambda}_i = \lim_{k \rightarrow \infty} \frac{\lambda_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} = 0 \quad \forall i \notin I_g(\hat{x}). \quad (4.7)$$

Next take an index $i \in I_{\pm}(\hat{x})$. Since (\hat{x}, \hat{y}) is feasible, we then have $\hat{y}_i = 0$. The boundedness of $\{\bar{\eta}_i^k\}$ therefore yields $\{\alpha_k (y_i^k - 1) + \bar{\eta}_i^k\} \rightarrow -\infty$. Consequently, we obtain

$$\eta_i^k = 0 \quad \forall i \in I_{\pm}(\hat{x}) \quad \forall k \in \mathbb{N} \text{ sufficiently large.} \quad (4.8)$$

Now, we claim that $\tilde{\gamma}_i = 0$ holds for such an index i . Suppose not. Then $\tilde{\gamma}_i^k \neq 0$ for all $k \in \mathbb{N}$ sufficiently large. Since $\tilde{\gamma}_i^k = \gamma_i^k y_i^k$, this implies $y_i^k \neq 0$ for all $k \in \mathbb{N}$ large enough. We then have

$$B_i^k = -\zeta^k + \eta_i^k + \gamma_i^k x_i^k \stackrel{(4.8)}{=} -\zeta^k + \gamma_i^k x_i^k = -\zeta^k + \frac{\tilde{\gamma}_i^k}{y_i^k} x_i^k. \quad (4.9)$$

Rearranging and dividing (4.9) by $\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|$ then gives

$$\frac{B_i^k + \zeta^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} = \frac{\tilde{\gamma}_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \cdot x_i^k \cdot \frac{1}{y_i^k}. \quad (4.10)$$

Observe that the left-hand side of (4.10) converges. On the other hand, since

$$\left\{ \frac{\tilde{\gamma}_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} x_i^k \right\} \rightarrow \tilde{\gamma}_i \hat{x}_i \neq 0$$

and $\{y_i^k\} \rightarrow 0$, the right-hand side diverges. This contradiction shows that

$$\tilde{\gamma}_i = 0 \quad \forall i \in I_{\pm}(\hat{x}). \quad (4.11)$$

Now, we claim that $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}) \neq 0$. Suppose not. Then, since $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}, \tilde{\zeta}, \tilde{\eta}) \neq 0$, it follows that $(\tilde{\zeta}, \tilde{\eta}) \neq 0$. Consider an index $i \in I_0(\hat{y})$. Since $\{y_i^k\} \rightarrow \hat{y}_i$ and $\{\bar{\eta}_i^k\}$ is a bounded sequence, we have $\{\alpha_k (y_i^k - 1) + \bar{\eta}_i^k\} \rightarrow -\infty$. Just like before, we can then assume w.l.o.g. that

$$\eta_i^k = 0 \quad \forall i \in I_0(\hat{y}) \quad \forall k \in \mathbb{N} \quad (4.12)$$

which implies $\tilde{\eta}_i = 0$. Hence, we have

$$(\tilde{\zeta}, \tilde{\eta}_i \ (i \in I_{\pm}(\hat{y}))) \neq 0. \quad (4.13)$$

Now let $i \in I_{\pm}(\hat{y})$. Since $\hat{y}_i \neq 0$ and $\{y_i^k\} \rightarrow \hat{y}_i$, we can assume w.l.o.g. that $y_i^k \neq 0$ for all $k \in \mathbb{N}$. We then get, for each $k \in \mathbb{N}$, that

$$B_i^k = -\zeta^k + \eta_i^k + \gamma_i^k x_i^k = -\zeta^k + \eta_i^k + \frac{\tilde{\gamma}_i^k}{y_i^k} x_i^k. \quad (4.14)$$

Rearranging and dividing (4.14) by $\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|$ yields

$$\frac{B_i^k + \zeta^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} = \frac{\eta_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} + \frac{\tilde{\gamma}_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \cdot x_i^k \cdot \frac{1}{y_i^k}. \quad (4.15)$$

By assumption, $\tilde{y}_i = 0$. Consequently, letting $k \rightarrow \infty$ in (4.15) yields

$$\tilde{\zeta} = \tilde{\eta}_i + 0 \cdot \hat{x}_i \cdot \frac{1}{\hat{y}_i} = \tilde{\eta}_i. \quad (4.16)$$

From (4.13) we then obtain $\tilde{\zeta} \neq 0$ and $\tilde{\eta}_i = \tilde{\zeta} \neq 0$ for all $i \in I_{\pm}(\hat{y})$. Since $\zeta^k \geq 0$ for all $k \in \mathbb{N}$, we have $\tilde{\zeta} \geq 0$ and, therefore, $\tilde{\zeta} > 0$. Hence, we can assume w.l.o.g. that $\zeta^k > 0$ for all $k \in \mathbb{N}$. This implies $\zeta^k = \alpha_k (n - e^T y^k - s) + \bar{\zeta}^k$. We then have

$$\begin{aligned} 0 < \tilde{\zeta} &= \lim_{k \rightarrow \infty} \frac{\zeta^k}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k (n - e^T y^k - s)}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|} + \lim_{k \rightarrow \infty} \frac{\bar{\zeta}^k}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k (n - e^T y^k - s)}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|}, \end{aligned}$$

since $\{\bar{\zeta}^k\}$ is bounded by definition. Consequently, we can assume w.l.o.g. that

$$n - e^T y^k - s > 0 \quad \forall k \in \mathbb{N}. \quad (4.17)$$

By assumption, (\hat{x}, \hat{y}) is feasible and, hence, $n - e^T \hat{y} - s \leq 0$. Thus, we obtain from (4.17) that $n - e^T y^k - s > n - e^T \hat{y} - s$ and, therefore,

$$e^T \hat{y} > e^T y^k \quad \forall k \in \mathbb{N}. \quad (4.18)$$

Furthermore, since $\tilde{\zeta} > 0$, by (4.16), we also have that $\tilde{\eta}_i > 0$ for all $i \in I_{\pm}(\hat{y})$. This implies $\eta_i^k > 0$ for all sufficiently large $k \in \mathbb{N}$. Consequently, we have $\eta_i^k = \alpha_k (y_i^k - 1) + \bar{\eta}_i^k$ for all $k \in \mathbb{N}$ large enough. We then obtain

$$\begin{aligned} 0 < \tilde{\eta}_i &= \lim_{k \rightarrow \infty} \frac{\eta_i^k}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k (y_i^k - 1)}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|} + \lim_{k \rightarrow \infty} \frac{\bar{\eta}_i^k}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k (y_i^k - 1)}{\|(\lambda^k, \mu^k, \tilde{y}^k, \zeta^k, \eta^k)\|}, \end{aligned}$$

since $\{\bar{\eta}_i^k\}$ is bounded by definition. Hence, we can assume w.l.o.g. that $y_i^k > 1$ for all $k \in \mathbb{N}$. On the other hand, the feasibility of (\hat{x}, \hat{y}) implies $\hat{y}_i \leq 1$ for all $i \in \{1, \dots, n\}$. Consequently, we obtain

$$\hat{y}_i < y_i^k \quad \forall i \in I_{\pm}(\hat{y}) \quad \forall k \in \mathbb{N}. \quad (4.19)$$

Together, this implies

$$\sum_{i \in I_{\pm}(\hat{y})} \hat{y}_i = e^T \hat{y} > e^T y^k = \sum_{i \in I_{\pm}(\hat{y})} y_i^k + \sum_{i \in I_0(\hat{y})} y_i^k \geq \sum_{i \in I_{\pm}(\hat{y})} \hat{y}_i + \sum_{i \in I_0(\hat{y})} y_i^k \implies \sum_{i \in I_0(\hat{y})} y_i^k < 0$$

for all $k \in \mathbb{N}$. By passing to a subsequence, we can therefore assume w.l.o.g. that there exists a $j \in I_0(\hat{y})$ with $y_j^k < 0$ for all $k \in \mathbb{N}$. Since $j \in I_0(\hat{y})$, by (4.12), we have $\eta_j^k = 0$ for all $k \in \mathbb{N}$ and, hence, $B_j^k = -\zeta^k + \gamma_j^k x_j^k$ or, equivalently, $B_j^k + \zeta^k = \gamma_j^k x_j^k$. Since $y_j^k \leq 0$, we then have

$$\gamma_j^k x_j^k = (\alpha_k x_j^k y_j^k + \bar{\gamma}_j^k) x_j^k = \alpha_k (x_j^k)^2 y_j^k + \bar{\gamma}_j^k x_j^k \leq \bar{\gamma}_j^k x_j^k.$$

Consequently, we have $B_j^k + \zeta^k \leq \tilde{\gamma}_j^k x_j^k$ and, therefore,

$$\frac{B_j^k + \zeta^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \leq \frac{\tilde{\gamma}_j^k x_j^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|}.$$

Since $\{\tilde{\gamma}_j^k x_j^k\}$ is bounded, letting $k \rightarrow \infty$ then yields the contradiction $0 < \tilde{\zeta} \leq 0$. Hence we have $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}) \neq 0$.

Dividing A^k by $\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|$ and letting $k \rightarrow \infty$ then yields

$$0 = \sum_{i=1}^m \tilde{\lambda}_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \tilde{\mu}_i \nabla h_i(\hat{x}) + \sum_{i=1}^n \tilde{\gamma}_i e_i$$

where $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}) \neq 0$ and, in view of (4.7) and (4.11), $\tilde{\lambda} \in \mathbb{R}_+^m$, $\tilde{\lambda}_i = 0$ for all $i \notin I_g(\hat{x})$, and $\tilde{\gamma}_i = 0$ for all $i \in I_{\pm}(\hat{x})$. This shows that \hat{x} satisfies properties (a)–(c) from Definition 2.4. We now verify that also the three conditions from part (d) hold.

For this purpose, let $i \in \{1, \dots, m\}$ such that $\tilde{\lambda}_i > 0$ holds. Then, we can assume w.l.o.g. that $\lambda_i^k > 0$ for all $k \in \mathbb{N}$ and, thus, $\lambda_i^k = \alpha_k g_i(x^k) + \bar{\lambda}_i^k$. Consequently, we have

$$\begin{aligned} 0 < \tilde{\lambda}_i &= \lim_{k \rightarrow \infty} \frac{\lambda_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k g_i(x^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} + \lim_{k \rightarrow \infty} \frac{\bar{\lambda}_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k g_i(x^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \end{aligned}$$

by the boundedness of $\{\bar{\lambda}_i^k\}$. Thus, we have $g_i(x^k) > 0$ for all $k \in \mathbb{N}$ sufficiently large and, therefore, also $\tilde{\lambda}_i g_i(x^k) > 0$ for all these $k \in \mathbb{N}$.

Next consider an index $i \in \{1, \dots, p\}$ such that $\tilde{\mu}_i \neq 0$. The boundedness of $\{\tilde{\mu}_i^k\}$ then implies

$$\begin{aligned} \tilde{\mu}_i &= \lim_{k \rightarrow \infty} \frac{\mu_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} = \lim_{k \rightarrow \infty} \frac{\alpha_k h_i(x^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} + \lim_{k \rightarrow \infty} \frac{\tilde{\mu}_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k h_i(x^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|}. \end{aligned}$$

Since $\alpha_k > 0$, this implies that $\tilde{\mu}_i \neq 0$ and $h_i(x^k)$ have the same sign for all $k \in \mathbb{N}$ sufficiently large, i.e. $\tilde{\mu}_i h_i(x^k) > 0$.

Finally, consider an index $i \in \{1, \dots, n\}$ such that $\tilde{\gamma}_i \neq 0$. The boundedness of $\{\tilde{\gamma}_i^k\}$ yields

$$\begin{aligned} \tilde{\gamma}_i &= \lim_{k \rightarrow \infty} \frac{\tilde{\gamma}_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} = \lim_{k \rightarrow \infty} \frac{\gamma_i^k y_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} = \lim_{k \rightarrow \infty} \frac{(\alpha_k x_i^k y_i^k + \tilde{\gamma}_i^k) y_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \\ &= \lim_{k \rightarrow \infty} \frac{\alpha_k x_i^k (y_i^k)^2}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} + \lim_{k \rightarrow \infty} \frac{\tilde{\gamma}_i^k y_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} = \lim_{k \rightarrow \infty} \frac{\alpha_k x_i^k (y_i^k)^2}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|}. \end{aligned}$$

Hence, $\tilde{\gamma}_i \neq 0$ and x_i^k also have the same sign for all $k \in \mathbb{N}$ large, i.e. $\tilde{\gamma}_i x_i^k > 0$.

Altogether, this contradicts the assumed CC-quasinormality of \hat{x} . Thus, $\{(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\}$ is bounded and therefore has a convergent subsequence. Assume w.l.o.g. that the whole sequence converges, i.e.,

$$\exists (\hat{\lambda}, \hat{\mu}, \hat{\gamma}, \hat{\zeta}, \hat{\eta}) : \{(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\} \rightarrow (\hat{\lambda}, \hat{\mu}, \hat{\gamma}, \hat{\zeta}, \hat{\eta}).$$

Since $\{\lambda^k\} \subseteq \mathbb{R}_+^m$, we also have $\hat{\lambda} \in \mathbb{R}_+^m$. Consider an index $i \notin I_g(\hat{x})$. Then, just like for $\tilde{\lambda}_i$, one can show that $\hat{\lambda}_i = 0$. Similarly, for $i \in I_\pm(\hat{x})$, following the argument for \tilde{y}_i , one also gets $\hat{y}_i = 0$. Taking $k \rightarrow \infty$ in the definition of A^k , we then obtain

$$0 = \nabla f(\hat{x}) + \nabla g(\hat{x})\hat{\lambda} + \nabla h(\hat{x})\hat{\mu} + \hat{y},$$

where $\hat{\lambda}_i = 0$ for all $i \notin I_g(\hat{x})$ and $\hat{y}_i = 0$ for all $i \in I_\pm(\hat{x})$. Thus, we conclude that (\hat{x}, \hat{y}) is CC-M-stationary. \blacksquare

Note that, in the previous proof, the CC-quasinormality of \hat{x} is only needed in case the penalty parameter $\{\alpha_k\}$ is unbounded. A direct consequence of Proposition 2.3 is the following:

Corollary 4.4. *Under the assumptions of Theorem 4.3, there exists $\hat{z} \in \mathbb{R}^n$ such that (\hat{x}, \hat{z}) is a CC-S-stationary point, where \hat{x} denotes a limit point generated by Algorithm 3.1.*

5 Numerical Results

In this section, we compare the performance of ALGENCAN with the Scholtes regularization method from [14] as well as the Kanzow-Schwartz regularization method from [16]. All experiments were conducted using Python together with the Numpy library. We used ALGENCAN 2.4.0 compiled with MA57 library [23] and called through its Python interface with user-supplied gradients of the objective functions, sparse Jacobian of the constraints, as well as sparse Hessian of the Lagrangian. As a subsolver for the two regularization methods, we used the (for academic use) freely available SQP solver WORHP version 1.14 [17] called through its Python interface. For the Scholtes regularization method, WORHP was called with user-supplied sparse gradients of the objective functions, sparse Jacobian of the constraints, as well as the sparse Hessian of the Lagrangian. On the other hand, for the Kanzow-Schwartz regularization method, since the analytical Hessian does not exist as the corresponding NCP-function is not twice differentiable, we called WORHP with user-supplied sparse gradients of the objective functions and sparse Jacobian of the constraints only. The Hessian of the Lagrangian was then approximated using the BFGS method. Throughout the experiments, both ALGENCAN and WORHP were called using their respective default settings.

We applied ALGENCAN directly to the relaxed reformulation of the test problems as in (2.2), i.e. without a lower bound for the auxiliary variable y . In contrast, following [14] and [16], for both regularization methods, we bounded y from below by 0. For each test problem, we started both regularization methods with an initial regularization parameter $t_0 = 1.0$ and decreased t_k in each iteration by a factor of 0.01. The regularization methods were terminated, if either $t_k < 10^{-8}$ or $\|x^k \circ y^k\|_\infty \leq 10^{-6}$.

5.1 Pilot Test

Let us begin by considering the following academic example

$$\min_{x \in \mathbb{R}^2} x_1 + 10x_2 \quad \text{s.t.} \quad \left(x_1 - \frac{1}{2}\right)^2 + (x_2 - 1)^2 \leq 1, \quad \|x\|_0 \leq 1$$

which is taken from [16]. This problem has a local minimizer in $(0, 1 - \frac{1}{2}\sqrt{3})$ and an isolated global minimizer in $(\frac{1}{2}, 0)$. Following [16], we discretised the rectangle $[-1, \frac{3}{2}] \times [-\frac{1}{2}, 2]$ resulting in 441 starting points for the considered methods. For each of these starting points, ALGENCAN converged towards the global minimizer $(\frac{1}{2}, 0)$. The same behaviour was also observed for the Scholtes regularization method. On the other hand, the Kanzow-Schwartz regularization method was slightly less succesful, converging in 437 cases towards the global minimizer. In the other 4 cases, the method converged towards the local minimizer. This behaviour might be due to the performance of the BFGS method used by WORHP in approximating the Hessian of the Lagrangian. Indeed, running the Scholtes regularization method without user-supplied Hessian of the Lagrangian, letting the Hessian be approximated by the BFGS method instead, yielded in a convergence towards the global minimizer in only 394 cases. In the other 47 cases, the Scholtes regularization method only managed to find the local minimizer.

5.2 Portfolio Optimization Problems

Following [16], we consider a classical portfolio optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} x^T Q x \quad \text{s.t.} \quad & \mu^T x \geq \rho, \quad e^T x \leq 1, \quad 0 \leq x \leq u, \\ & \|x\|_0 \leq s, \end{aligned} \tag{5.1}$$

where Q and μ are the covariance matrix and the mean of n possible assets and $e^T x \leq 1$ is the budget constraint, see [11, 18]. We generated the test problems using the data from [22], considering $s = 5, 10, 20$ for each dimension $n = 200, 300, 400$, which resulted in 270 test problems, see also [16]. Here, we considered six total approaches:

- ALGENCAN without a lower bound on y
- ALGENCAN with an additional lower bound $y \geq 0$
- Scholtes and Kanzow-Schwartz regularization for cardinality-constrained problems [14, 16] with a regularization of both upper quadrants
- Scholtes and Kanzow-Schwartz regularization for MPCCs [26, 33] with a regularization of the upper right quadrant only.

As discussed before, introducing a lower bound $y \geq 0$ in (2.2) is possible without changing the theoretical properties of the reformulation. Similarly, due to the constraint $x \geq 0$ in (5.1), the feasible set of the reformulated problem actually has the classical MPCC structure, and thus only one regularization function in the first quadrant suffices. This motivates the modifications of both ALGENCAN and the two regularization methods described above, which should theoretically not have any effect on the performance of the solution algorithms.

For each test problem, we used the initial values $x^0 = 0$ and $y^0 = e$. As a performance measure for the considered methods we compared the attained objective function values and generated a performance profile as suggested in [19], where we set the objective function value of a method for a problem to be ∞ , if the method failed to find a feasible point of the problem within a tolerance of 10^{-6} .

As can be seen from Figure 1, ALGENCAN worked very reliable with regards to feasibility of the solutions. It often outperformed the regularization methods in terms of objective function value of the solution, especially for larger values of s . Although introducing the lower

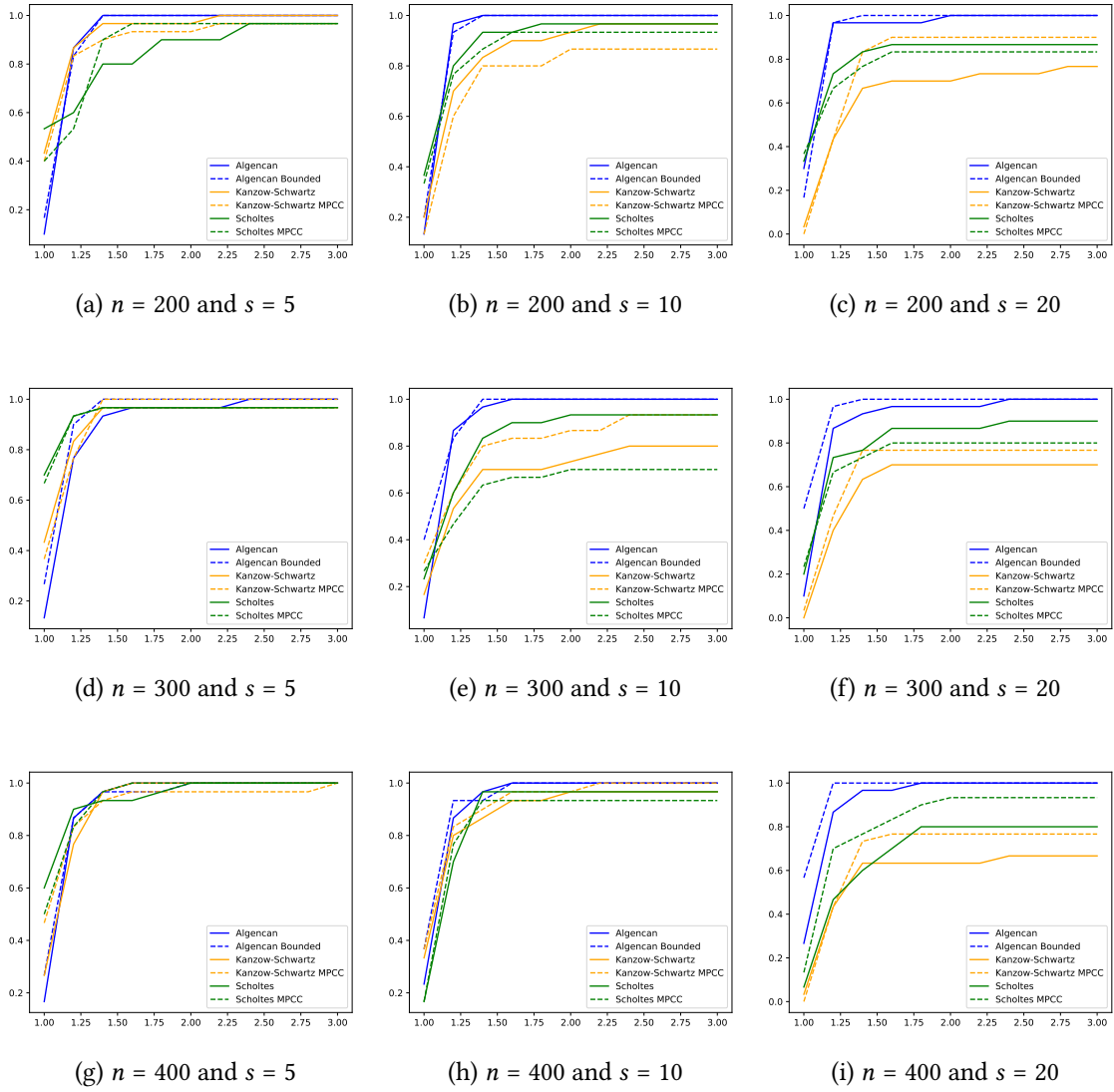


Figure 1: Comparing the performance of ALGENCAN and the regularization methods for (5.1)

bound $y \geq 0$ does not have any theoretical effect on ALGENCAN, the numerical results suggest that it could bring slight improvements to ALGENCAN's performance.

6 Final Remarks

This paper shows that the safeguarded augmented Lagrangian method applied directly and without problem-specific modifications to the continuous reformulation of cardinality-constrained problems converges to suitable (M-, essentially even S-) stationary points under a weak problem-tailored CQ called CC-quasinormality. On the other hand, it is known that this safeguarded ALM generates so-called AKKT sequences (AKKT = approximate KKT) which, under suitable constraint qualifications, lead to KKT points and, hence, to S-stationary points. In the context of cardinality constraints, however, the AKKT concept is useless as an optimality criterion since any feasible point is known to be an AKKT point, cf. [30].

On the other hand, there are some recent reports, which define a problem-tailored AKKT-type condition for cardinality constrained problems, see [30] and [31] (the latter in a more general context). Algorithmic applications of these AKKT-type conditions are not discussed in these papers. We therefore plan to investigate this topic within our future research. Note that a corresponding convergence theory based on AKKT-type conditions for cardinality constrained problems will be different from our current theory, based on CC-quasinormality, since it is already known from standard NLPs that quasinormality and AKKT regularity conditions are two independent concepts, cf. [4].

References

- [1] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *J. Optim. Theory Appl.*, 125(2):473–485, 2005.
- [2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM J. Optim.*, 18(4):1286–1309, 2007.
- [3] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Math. Program.*, 111(1-2, Ser. B):5–32, 2008.
- [4] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM J. Optim.*, 26(1):96–110, 2016.
- [5] R. Andreani, N. S. Fazzio, M. L. Schuverdt, and L. D. Secchin. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM J. Optim.*, 29(1):743–766, 2019.
- [6] R. Andreani, G. Haeser, L. D. Secchin, and P. J. S. Silva. New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. *SIAM J. Optim.*, 29(4):3201–3230, 2019.
- [7] R. Andreani, G. Haeser, and D. S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Math. Program.*, 180(1-2, Ser. A):203–235, 2020.

- [8] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1982.
- [9] D. P. Bertsekas and A. E. Ozdaglar. Pseudonormality and a Lagrange multiplier theory for constrained optimization. *J. Optim. Theory Appl.*, 114(2):287–343, 2002.
- [10] D. Bertsimas and R. Shioda. Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.*, 43(1):1–22, 2009.
- [11] D. Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Math. Programming*, 74(2, Ser. A):121–140, 1996.
- [12] E. G. Birgin and J. M. Martínez. *Practical augmented Lagrangian methods for constrained optimization*, volume 10 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.
- [13] E. G. Birgin, W. Gómez, G. Haeser, L. M. Mito, and D. O. Santos. An augmented Lagrangian algorithm for nonlinear semidefinite programming applied to the covering problem. *Comput. Appl. Math.*, 39(1):Paper No. 10, 21, 2020.
- [14] M. Branda, M. Bucher, M. Červinka, and A. Schwartz. Convergence of a Scholtes-type regularization method for cardinality-constrained optimization problems with an application in sparse robust portfolio optimization. *Comput. Optim. Appl.*, 70(2):503–530, 2018.
- [15] L. F. Bueno, G. Haeser, and F. N. Rojas. Optimality conditions and constraint qualifications for generalized Nash equilibrium problems and their practical implications. *SIAM J. Optim.*, 29(1):31–54, 2019.
- [16] O. P. Burdakov, C. Kanzow, and A. Schwartz. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. *SIAM J. Optim.*, 26(1):397–425, 2016.
- [17] C. Büskens and D. Wassel. The ESA NLP solver WORHP. In *Modeling and optimization in space engineering*, volume 73 of *Springer Optim. Appl.*, pages 85–110. Springer, New York, 2013.
- [18] D. Di Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen, and M. Sciandrone. A concave optimization-based approach for sparse portfolio selection. *Optim. Methods Softw.*, 27(6):983–1000, 2012.
- [19] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2, Ser. A):201–213, 2002.
- [20] H. Dong, M. Ahn, and J.-S. Pang. Structural properties of affine sparsity constraints. *Math. Program.*, 176(1-2, Ser. B):95–135, 2019.
- [21] M. Feng, J. E. Mitchell, J.-S. P., X. Shen, and A. Wächter. Complementarity formulations of ℓ_0 -norm optimization problems. *Pac. J. Optim.*, 14(2):273–305, 2018.
- [22] A. Frangioni and C. Gentile. SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Oper. Res. Lett.*, 35(2):181–185, 2007.

- [23] HSL. *A collection of Fortran codes for large scale scientific computation*. <http://www.hsl.rl.ac.uk/>.
- [24] A. F. Izmailov, M. V. Solodov, and E. I. Uskov. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM J. Optim.*, 22(4):1579–1606, 2012.
- [25] C. Kanzow. On the multiplier-penalty-approach for quasi-variational inequalities. *Math. Program.*, 160(1-2, Ser. A):33–63, 2016.
- [26] C. Kanzow and A. Schwartz. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM J. Optim.*, 23(2):770–798, 2013.
- [27] C. Kanzow and D. Steck. Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems. *SIAM J. Optim.*, 26(4):2034–2058, 2016.
- [28] C. Kanzow and D. Steck. An example comparing the standard and safeguarded augmented Lagrangian methods. *Oper. Res. Lett.*, 45(6):598–603, 2017.
- [29] C. Kanzow, D. Steck, and D. Wachsmuth. An augmented Lagrangian method for optimization problems in Banach spaces. *SIAM J. Control Optim.*, 56(1):272–291, 2018.
- [30] E. H. M. Krulikowski, A. A. Ribeiro, and M. Sachine. A sequential optimality condition for mathematical programs with cardinality constraints. *ArXiv e-prints*, 2020. ArXiv:2008.03158.
- [31] P. Mehlitz. Asymptotic stationarity and regularity for nonsmooth optimization problems. *Journal of Nonsmooth Analysis and Optimization*. doi: 10.46298/jnsao-2020-6575.
- [32] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.
- [33] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.
- [34] M. Červinka, C. Kanzow, and A. Schwartz. Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. *Math. Program.*, 160(1-2, Ser. A):353–377, 2016.
- [35] C. Zhao, N. Xiu, H.-D. Qi, and Z. Luo. A Lagrange-Newton algorithm for sparse nonlinear programming. *ArXiv e-prints*, 2020. ArXiv:2004.13257.