Kanzow, Christian; Raharja, Andreas B.; Schwartz, Alexandra

Sequential optimality conditions for cardinality-constrained optimization problems with applications

Computational Optimization and Applications

Provided in Cooperation with:
Springer Nature

Suggested Citation: Kanzow, Christian; Raharja, Andreas B.; Schwartz, Alexandra (2021) : Sequential optimality conditions for cardinality-constrained optimization problems with applications, Computational Optimization and Applications, ISSN 1573-2894, Springer US, New York, NY, Vol. 80, Iss. 1, pp. 185-211, https://doi.org/10.1007/s10589-021-00298-z

Computational Optimization and Applications (2021) 80:185–211
https://doi.org/10.1007/s10589-021-00298-z

Sequential optimality conditions for cardinality-constrained optimization problems with applications

Christian Kanzow[1] · **Andreas B. Raharja**[1] · **Alexandra Schwartz**[2]

Received: 12 February 2021 / Accepted: 27 June 2021 / Published online: 22 July 2021

Abstract
Recently, a new approach to tackle cardinality-constrained optimization problems based on a continuous reformulation of the problem was proposed. Following this approach, we derive a problem-tailored sequential optimality condition, which is satisfied at every local minimizer without requiring any constraint qualification. We relate this condition to an existing M-type stationary concept by introducing a weak sequential constraint qualification based on a cone-continuity property. Finally, we present two algorithmic applications: We improve existing results for a known regularization method by proving that it generates limit points satisfying the aforementioned optimality conditions even if the subproblems are only solved inexactly. And we show that, under a suitable Kurdyka–Łojasiewicz-type assumption, any limit point of a standard (safeguarded) multiplier penalty method applied directly to the reformulated problem also satisfies the optimality condition. These results are stronger than corresponding ones known for the related class of mathematical programs with complementarity constraints.

Keywords Cardinality constraints · Sequential optimality condition · Cone-continuity type constraint qualification · Relaxation method · Augmented Lagrangian method

✉ Christian Kanzow
kanzow@mathematik.uni-wuerzburg.de

Andreas B. Raharja
andreas.raharja@mathematik.uni-wuerzburg.de

Alexandra Schwartz
alexandra.schwartz@tu-dresden.de

[1] Institute of Mathematics, University of Würzburg, Campus Hubland Nord, Emil-Fischer-Straße 30, 97074 Würzburg, Germany

[2] Fakultät Mathematik, Technische Universität Dresden, Zellescher Weg 12-14, 01069 Dresden, Germany

1 Introduction

We consider *cardinality-constrained (CC) optimization problems* of the form

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \le 0,\ h(x) = 0,\ \|x\|_0 \le s, \tag{1.1}$$

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. Throughout this paper, we assume that $s < n$ since the cardinality constraint would otherwise be superfluous.

This class of problems has attracted great interest in recent years due to its abundance of applications including portfolio optimization [8, 9, 11] and statistical regression [8, 14]. It should be noted, however, that these problems are difficult to solve, mainly due to the presence of the cardinality constraint defined by the mapping $\|\cdot\|_0$ which, in spite of the notation used here, does not define a norm and is not even continuous. Even testing the feasibility of (1.1) is known to be NP-complete [8].

One way to attack these problems is to reformulate them as mixed-integer problems. This reformulation is the backbone of many algorithms employing ideas from discrete optimization, see for example [8, 9, 13, 24, 30, 32].

A new approach to solve this type of problems was introduced recently in [12], see also [15] for a similar approach in the context of *sparse optimization*. There, (1.1) is reformulated as a continuous optimization problem with orthogonality-type constraints, for which first-order stationarity concepts called CC-M- and CC-S-stationarity are derived. However, in order to guarantee that these stationarity conditions hold at a local minimizer of (1.1), one needs a constraint qualification. The regularization method from [17] is adapted to solve the reformulated problem and it is shown that any limit point of this method satisfies the CC-M-stationarity condition provided that a constraint qualification called CC-CPLD holds at this limit point. Nevertheless, this convergence result is only proven for the exact case, i.e., under the assumption that an exact KKT point of the regularized subproblem can be computed in each iteration. Numerically, however, this is rarely the case. In the context of *mathematical programs with complementarity constraints* (MPCC for short), it is known that, if we take inexactness into account, then the convergence theory for the this regularization method (like for most other regularization techniques) is weakened significantly [18].

Let us now describe the contributions of our paper. We first derive a sequential optimality condition called CC-AM-stationarity for (1.1), which is the CC-analogue of the *approximate Karush-Kuhn-Tucker* (AKKT) condition for standard nonlinear optimization problems (NLP) introduced in [3, 10, 25], see also [6, 26] for similar concepts in the context of MPCCs. We show that this first-order necessary optimality condition is satisfied at every local minimizer of (1.1) without requiring a constraint qualification. In order to establish the relationship between CC-AM-stationarity and the CC-M-stationarity condition introduced in [12, 31], we then propose a constraint qualification called CC-AM-regularity based on a cone-continuity property. This constraint qualification is the CC-analogue of the *AKKT-regularity* introduced in [4, 5, 10]. Same as CC-M-stationarity, both new concepts CC-AM-stationarity and CC-AM-regularity

depend only on the original cardinality-constrained problem (1.1) and not on the auxiliary variable introduced in the continuous reformulation. Subsequently, we prove that any limit point of the regularization method introduced in [12, 17] satisfies the CC-AM-stationarity condition in both the exact and inexact case, i.e., also in the situation where the resulting NLP-subproblems are solved only inexactly. This indicates that the application of these methods for CC does not suffer from any drawback when we take inexactness into account, in contrast to the MPCC case. Finally, we show that, under a suitable Kurdyka-Łojasiewicz-type assumption, any limit point of a standard (safeguarded) augmented Lagrangian method [1, 10] applied directly to the reformulated problem also satisfies CC-AM-stationarity, see also [6] for a similar result obtained in the context of MPCCs. Since numerical results for the methods investigated here can already be found in some other papers [12, 21], our focus is on the theoretical background of these approaches.

The paper is organized as follows: We first recall some basic definitions and results in Sect. 2. Then we introduce the problem-tailored sequential optimality condition and related constraint qualification in Sects. 3 and 4, respectively. These sequential optimality conditions are then applied, in Sects. 5 and 6, to the regularization method and the augmented Lagrangian approach. We close with some final remarks in Sect. 7. There is also an appendix where we compare our sequential optimality condition with an existing one from [22], which is formulated specifically for the continuous reformulation, see Sect. 1.

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

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

Clearly we have $\{1, \dots, n\} = I_{\pm}(x) \cup I_0(x)$ and $I_{\pm}(x) \cap I_0(x) = \emptyset$. Note that these definitions imply $\|x\|_0 = |I_{\pm}(x)|$. Given a set $C \subseteq \mathbb{R}^n$, we denote the corresponding polar cone by $C^\circ := \{y \in \mathbb{R}^n \mid y^T x \leq 0 \text{ for all } x \in C\}$. We write $B_r(x)$ and $\overline{B}_r(x)$ for an open and closed ball with radius $r > 0$ around x.

2 Preliminaries

We first recall some basic definitions, cf. [27] for more details. For a multifunction $\Gamma : \mathbb{R}^l \rightrightarrows \mathbb{R}^q$ the *Painlevé-Kuratowski outer/upper limit* of $\Gamma(z)$ at $\hat{z} \in \mathbb{R}^l$ is defined as

$$\limsup_{z \to \hat{z}} \Gamma(z) := \{\hat{w} \in \mathbb{R}^q \mid \exists \{(z^k, w^k)\} \to (\hat{z}, \hat{w}) \text{ with } w^k \in \Gamma(z^k) \text{ for all } k \in \mathbb{N}\}.$$

For a nonempty and closed set $A \subseteq \mathbb{R}^n$ and a point $\hat{x} \in A$ the *Bouligand tangent cone* and the *Fréchet normal cone* to A at $\hat{x}$ are given by

$$\mathcal{T}_A(\hat{x}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \exists \{x^k\} \subseteq A, \{t_k\} \subseteq \mathbb{R}_+ : \; \{x^k\} \to \hat{x}, \{t_k\} \downarrow 0, \left\{\tfrac{x^k - \hat{x}}{t_k}\right\} \to d \right\},$$

$$\mathcal{N}_A^F(\hat{x}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \limsup_{x \to \hat{x},\, x \in A} \tfrac{d^T(x - \hat{x})}{\|x - \hat{x}\|} \leq 0 \right\} = T_A(\hat{x})^\circ.$$

The Fréchet normal cone for a set of particular interest in our framework is stated in the following result, whose proof follows from straightforward computations.

Lemma 2.1 *Let* $C := \{(a,b) \in \mathbb{R}^2 \mid ab = 0\}$ *and* $(x,y) \in C$. *Then we have*

$$\mathcal{N}_C^F((x,y)) = \begin{cases} \mathbb{R} \times \{0\} & \text{if } x = 0, y \neq 0, \\ \{0\} \times \mathbb{R} & \text{if } x \neq 0, y = 0, \\ \{(0,0)\} & \text{if } x = 0, y = 0. \end{cases}$$

Next, let us take a closer look at (1.1) and follow the approach introduced in [12]. To simplify the notation, we define the set $X := \{x \in \mathbb{R}^n \mid g(x) \leq 0,\ h(x) = 0\}$. Now consider $x \in \mathbb{R}^n$, and define a corresponding $y \in \mathbb{R}^n$ by setting $y_i := 0$ for $i \in I_{\pm}(x)$ and $y_i := 1$ for $i \in I_0(x)$. Then $\|x\|_0 = n - e^T y$, where $e := (1, \ldots, 1)^T \in \mathbb{R}^n$. This leads to the following mixed-integer problem

$$\min_{x,y} f(x) \quad \text{s.t.} \quad x \in X,\ n - e^T y \leq s,\ y \in \{0,1\}^n,\ x \circ y = 0 \tag{2.1}$$

and its relaxation

$$\min_{x,y} f(x) \quad \text{s.t.} \quad x \in X,\ n - e^T y \leq s,\ y \leq e,\ x \circ y = 0, \tag{2.2}$$

where $\circ$ denotes the Hadamard product. Note that (2.2) slightly differs from the continuous reformulation in [12] since we drop the constraint $y \geq 0$ here, which leads to a larger feasible set. Nevertheless, it is easy to see that all results obtained in Sect. 3 of [12] are applicable to our reformulation here as well. Let us now gather these results, cf. [12] for the proofs.

Theorem 2.2 *Let* $\hat{x} \in \mathbb{R}^n$. *Then the following statements hold:*

(a) $\hat{x}$ *is feasible for* (1.1) *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 minimizer of* (1.1) *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* (1.1), *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* $\|\hat{y}\|_0 = s$, *then* $\hat{x}$ *is a local minimizer of* (1.1).

Note that the extra condition for the converse statement in Theorem 2.2 (c) is necessary, in general, see [12, Example 3] for a counterexample.

We close this section by noting that, occasionally, some constraint qualifications defined in [31, Definition 3.5] will play some role within this paper. In particular, this includes the CC-ACQ and CC-GCQ condition, which are problem-tailored modifications of the standard Abadie and Guignard CQs, respectively. Since their exact definitions require some overhead and the details are not relevant in our

context, we refrain from stating their definitions here. We only stress that these are fairly mild constraint qualifications.

3 A sequential optimality condition

Sequential optimality conditions like the AKKT conditions for NLPs have become very popular during the last few years, see [10]. In principle, these AKKT conditions can also be applied to the optimization problem (2.2) by viewing this program as an NLP. But then too many points satisfy the AKKT property, see [22, Thm. 4.1], so that the AKKT conditions turn out to be an optimality condition, which is too weak for this problem (i.e., besides the local minima, many other feasible points satisfy the standard AKKT conditions). This means that suitable problem-tailored sequential optimality conditions are required for cardinality-constrained and related problems with "difficult" constraints.

This was done, for example, in [23] for a very general class of problems. The concept there is based on the limiting normal cone and can, in principle, be specialized to our setting. Instead of recalling this general theory and then specializing the corresponding concepts, we decided to use a direct and very elementary approach in this (and the subsequent) section. We stress that our definition is based on the original problem (1.1) in the x-space. The recent report [22] also introduces a sequential optimality condition for cardinality-constrained programs which, however, is essentially based on the reformulated problem (2.2) in the (x, y)-space. Nevertheless, it turns out that our formulation is, in some sense, equivalent to the notion from [22]. Since this equivalence is not exploited in our subsequent analysis, we discuss the details in an appendix, see Sect. 1.

Definition 3.1 Let $\hat{x} \in \mathbb{R}^n$ be a feasible point of (1.1). We say that $\hat{x}$ is *CC approximately M-stationary* (CC-AM-stationary), if there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that

(a) $\{x^k\} \to \hat{x}$ and $\{\nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k\} \to 0$,
(b) $\lambda_i^k = 0$ for all $i \notin I_g(\hat{x})$ as well as $\gamma_i^k = 0$ for all $i \in I_{\pm}(\hat{x})$ for all $k \in \mathbb{N}$.

Note that the two requirements $\lambda_i^k = 0$ and $\gamma_i^k = 0$ are assumed to hold for all $k \in \mathbb{N}$. Subsequencing if necessary, it is easy to see that this is equivalent to forcing these multiplier estimates to be zero only for all $k \in \mathbb{N}$ sufficiently large. We further stress that Definition 3.1 makes no assumptions regarding the boundedness of the multiplier estimates.

If we define $W := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid x \circ y = 0\}$, then the feasible set Z of (2.2) has the form

$$Z = \left\{ (x, y) \in W \,\middle|\, \begin{array}{ll} g(x) \leq 0, & h(x) = 0, \\ n - e^T y - s \leq 0, & y - e \leq 0 \end{array} \right\}.$$

The following theorem shows that CC-AM-stationarity is a first-order necessary optimality condition for (1.1) without the need for some kind of constraint qualification.

Theorem 3.2 *Let $\hat{x} \in \mathbb{R}^n$ be a local minimizer of (1.1). Then $\hat{x}$ is a CC-AM-stationary point.*

Proof Since $\hat{x}$ is a local minimizer of (1.1), by Theorem 2.2, there exists $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a local minimizer of (2.2). Hence, we can find an $\epsilon > 0$ such that

$$f(\hat{x}) \le f(x) \quad \forall (x, y) \in \overline{B}_\epsilon((\hat{x}, \hat{y})) \cap Z.$$

Obviously $(\hat{x}, \hat{y})$ is then the unique global minimizer of

$$\min_{x,y} f(x) + \frac{1}{2}\|(x, y) - (\hat{x}, \hat{y})\|_2^2 \quad \text{s.t.} \quad (x, y) \in \overline{B}_\epsilon((\hat{x}, \hat{y})) \cap Z. \tag{3.1}$$

Now pick a sequence $\{\alpha_k\} \subseteq \mathbb{R}_+$ such that $\{a_k\} \uparrow \infty$, and consider for each $k \in \mathbb{N}$ the partially penalized and localized problem

$$\min_{x,y} f(x) + \alpha_k \pi(x, y) + \frac{1}{2}\|(x, y) - (\hat{x}, \hat{y})\|_2^2 \quad \text{s.t.} \quad (x, y) \in \overline{B}_\epsilon((\hat{x}, \hat{y})) \cap W, \tag{3.2}$$

where

$$\begin{aligned} \pi(x, y) &:= \frac{1}{2}\left\|(g(x)_+, h(x), (n - e^T y - s)_+, (y - e)_+)\right\|_2^2 \\ &= \frac{1}{2}\Big(\sum_{i=1}^{m} \max\{0, g_i(x)\}^2 \\ &\quad + \sum_{i=1}^{p} h_i(x)^2 + \max\{0, n - e^T y - s\}^2 + \sum_{i=1}^{n} \max\{0, y_i - 1\}^2 \Big). \end{aligned}$$

The objective function of (3.2) is continuously differentiable for all $k \in \mathbb{N}$. Furthermore, the feasible set $\overline{B}_\epsilon((\hat{x}, \hat{y})) \cap W$ is nonempty and compact. Hence, for each $k \in \mathbb{N}$, (3.2) admits a global minimizer (x^k, y^k). We thus have a sequence $\{(x^k, y^k)\}$ in the compact set $\overline{B}_\epsilon((\hat{x}, \hat{y})) \cap W$ and can thus assume w.l.o.g. that $\{(x^k, y^k)\}$ converges, i.e., there exists $(\bar{x}, \bar{y}) \in \overline{B}_\epsilon((\hat{x}, \hat{y})) \cap W$ such that $\{(x^k, y^k)\} \to (\bar{x}, \bar{y})$. We now want to show that $(\bar{x}, \bar{y}) = (\hat{x}, \hat{y})$. Since $(\hat{x}, \hat{y}) \in Z$, it is a feasible point of (3.2) for each $k \in \mathbb{N}$ with $\pi(\hat{x}, \hat{y}) = 0$. Thus, we obtain for each $k \in \mathbb{N}$ that

$$\begin{aligned} f(x^k) + \alpha_k \pi(x^k, y^k) + \frac{1}{2}\left\|(x^k, y^k) - (\hat{x}, \hat{y})\right\|_2^2 &\le f(\hat{x}) + \alpha_k \pi(\hat{x}, \hat{y}) \\ + \frac{1}{2}\|(\hat{x}, \hat{y}) - (\hat{x}, \hat{y})\|_2^2 &= f(\hat{x}). \end{aligned} \tag{3.3}$$

Dividing (3.3) by α_k and letting $k \to \infty$ yields $\pi(\bar{x}, \bar{y}) \leq 0$. This implies that $(\bar{x}, \bar{y}) \in \overline{B_\epsilon((\hat{x}, \hat{y}))} \cap Z$ and therefore, it is feasible for (3.1). Furthermore, we also obtain from (3.3) that

$$f(x^k) + \frac{1}{2}\left\|(x^k, y^k) - (\hat{x}, \hat{y})\right\|_2^2 \leq f(\hat{x})$$

and hence, by letting $k \to \infty$,

$$f(\bar{x}) + \frac{1}{2}\|(\bar{x}, \bar{y}) - (\hat{x}, \hat{y})\|_2^2 \leq f(\hat{x}) = f(\hat{x}) + \frac{1}{2}\|(\hat{x}, \hat{y}) - (\hat{x}, \hat{y})\|_2^2.$$

Since $(\hat{x}, \hat{y})$ is the unique global solution of (3.1), we then necessarily have $(\bar{x}, \bar{y}) = (\hat{x}, \hat{y})$. This shows that $\{(x^k, y^k)\} \to (\hat{x}, \hat{y})$. We can thus assume w.l.o.g. that $(x^k, y^k) \in B_\epsilon((\hat{x}, \hat{y})) \cap W$ for each $k \in \mathbb{N}$. This, in turn, implies that, for each $k \in \mathbb{N}$, (x^k, y^k) is a local minimizer of

$$\min_{x,y} f(x) + \alpha_k \pi(x, y) + \frac{1}{2}\|(x, y) - (\hat{x}, \hat{y})\|_2^2 \quad \text{s.t.} \quad (x, y) \in W.$$

By [27, Theorem 6.12], we then have for each $k \in \mathbb{N}$ that

$$-\left(\begin{bmatrix}\nabla f(x^k) \\ 0\end{bmatrix} + \alpha_k \begin{bmatrix}\nabla_x \pi(x^k, y^k) \\ \nabla_y \pi(x^k, y^k)\end{bmatrix} + \begin{bmatrix}x^k - \hat{x} \\ y^k - \hat{y}\end{bmatrix}\right) \in \mathcal{N}_W^F((x^k, y^k)), \tag{3.4}$$

where

$$\nabla_x \pi(x^k, y^k) = \sum_{i=1}^m \max\{0, g_i(x^k)\} \nabla g_i(x^k) + \sum_{i=1}^p h_i(x^k) \nabla h_i(x^k) \qquad \text{and}$$

$$\nabla_y \pi(x^k, y^k) = -\max\{0, n - e^T y^k - s\} e + \sum_{i=1}^n \max\{0, y_i^k - 1\} e_i.$$

Observe that $W = C^n$, where C is the set from Lemma 2.1, and hence, by [27, Proposition 6.41], we obtain $\mathcal{N}_W^F((x^k, y^k)) = \bigtimes_{i=1}^n \mathcal{N}_C^F((x_i^k, y_i^k))$. Now define, for each $k \in \mathbb{N}$,

$$\begin{aligned} \gamma^k &:= -\left(\nabla f(x^k) + \alpha_k \nabla_x \pi(x^k, y^k) + x^k - \hat{x}\right), \\ \lambda_i^k &:= \alpha_k \max\{0, g_i(x^k)\} && \forall i = 1, \ldots, m, \\ \mu_i^k &:= \alpha_k h_i(x^k) && \forall i = 1, \ldots, p. \end{aligned}$$

Clearly, we have $\{\gamma^k\} \subseteq \mathbb{R}^n$. Now suppose that $i \in I_\pm(\hat{x})$. Since $\{x_i^k\} \to \hat{x}_i$, we can assume w.l.o.g. that $x_i^k \neq 0$ for each $k \in \mathbb{N}$. Lemma 2.1 then implies (recall that (x^k, y^k) is feasible) that $\mathcal{N}_C^F((x_i^k, y_i^k)) = \{0\} \times \mathbb{R}$ and thus, by the definition of γ^k and (3.4), we have $\gamma_i^k = 0$ for all $k \in \mathbb{N}$.

Now observe that, by definition, we have $\{\lambda^k\} \subseteq \mathbb{R}_+^m$. Suppose that $i \notin I_g(\hat{x})$. Then $g_i(\hat{x}) < 0$. Since $\{g_i(x^k)\} \to g_i(\hat{x})$, we can assume w.l.o.g. that, for each $k \in \mathbb{N}$, we have $g_i(x^k) < 0$, which in turn implies that $\max\{0, g_i(x^k)\} = 0$ and hence, in particular, $\lambda_i^k = 0$ for all $k \in \mathbb{N}$.

Using the definition of γ^k and $\{x^k\} \to \hat{x}$, we obtain

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i = \hat{x} - x^k \to 0.$$

This completes the proof. □

It is also possible to bypass the continuous reformulation (2.2) and prove Theorem 3.2 directly based on the original problem (1.1), using techniques from variational analysis. The reason why we did not do that here, is that the above proof also shows that every local minimizer of (2.2) is a CC-AM-stationary point. Now recall that (2.2) can have local minimizers, which are not a local minimizers of (1.1), see e.g. [12, Example 3]. This immediately implies that CC-AM-stationary points are not necessarily local minimizers of (1.1), i.e. the converse of Theorem 3.2 is false in general.

We close this section by considering the special case $X = \mathbb{R}^n$, i.e., we have the problem

$$\min_x f(x) \quad \text{s.t.} \quad \|x\|_0 \le s. \tag{3.5}$$

In [7], a first-order necessary optimality condition for (3.5) called *basic feasibility* was introduced, see the reference for details. Here we only note that the notion of basic feasibility can be shown to be identical to our CC-AM-stationarity at any feasible point $\hat{x}$ satisfying $\|\hat{x}\|_0 = s$, i.e., these two optimality conditions coincide in the interesting case, where the cardinality constraint is active.

4 A cone-continuity-type constraint qualification

Let $\hat{x} \in \mathbb{R}^n$ be feasible for (1.1). Then we define for each $x \in \mathbb{R}^n$ the cone

$$K_{\hat{x}}(x) := \left\{ \nabla g(x)\lambda + \nabla h(x)\mu + \gamma \;\middle|\; \begin{array}{c} (\lambda, \mu, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n, \\ \lambda_i = 0 \; \forall i \notin I_g(\hat{x}), \\ \gamma_i = 0 \; \forall i \in I_{\pm}(\hat{x}) \end{array} \right\}. \tag{4.1}$$

Note that the index sets $I_g(\hat{x})$ and $I_{\pm}(\hat{x})$ depend on $\hat{x}$ and not x. With this cone, we can translate Definition 3.1 into the language of variational analysis, see also [4].

Theorem 4.1 *$\hat{x}$ feasible for (1.1) is CC-AM-stationary* $\iff$ $-\nabla f(\hat{x}) \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x)$.

Proof "⇒": By assumption, there exist sequences $\{x^k\}, \{\gamma^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, and $\{\mu^k\} \subseteq \mathbb{R}^p$ such that the conditions in Definition 3.1 hold. Now define

$$u^k := \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k.$$

Then we have $\{u^k\} \to 0$. Next we define

$$w^k := u^k - \nabla f(x^k) = \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k.$$

Clearly, we have $\{w^k\} \to -\nabla f(\hat{x})$. Moreover, by the last two conditions in Definition 3.1, we also have $w^k \in K_{\hat{x}}(x^k)$ for each $k \in \mathbb{N}$. Hence, we have $-\nabla f(\hat{x}) \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x)$.

"$\Leftarrow$": By assumption, there exist sequences $\{x^k\}, \{w^k\} \subseteq \mathbb{R}^n$ such that $\{x^k\} \to \hat{x}$, $\{w^k\} \to -\nabla f(\hat{x})$, and $w^k \in K_{\hat{x}}(x^k)$ for each $k \in \mathbb{N}$. Now, by (4.1), for each $k \in \mathbb{N}$, there exist $(\lambda^k, \mu^k, \gamma^k) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$ such that

$$w^k = \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k$$

with $\lambda^k_i = 0$ for all $i \notin I_g(\hat{x})$ and $\gamma^k_i = 0$ for all $i \in I_{\pm}(\hat{x})$. For these multipliers, we obtain

$$\begin{aligned}\{\nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k\} &= \{\nabla f(x^k) + w^k\} \\ &\to \nabla f(\hat{x}) - \nabla f(\hat{x}) = 0.\end{aligned}$$

Thus, $\hat{x}$ is a CC-AM-stationary point. □

Let us now recall the CC-M-stationary concept introduced in [12], where it was shown to be a first-order optimality condition for (1.1) under suitable assumptions.

Definition 4.2 Let $\hat{x} \in \mathbb{R}^n$ be a feasible point of (1.1). We then say that $\hat{x}$ is *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

(a) $0 = \nabla f(\hat{x}) + \nabla g(\hat{x})\lambda + \nabla h(\hat{x})\mu + \gamma$,
(b) $\lambda_i = 0$ for all $i \notin I_g(\hat{x})$ as well as $\gamma_i = 0$ for all $i \in I_{\pm}(\hat{x})$.

The following translation is then obvious.

Lemma 4.3 *$\hat{x}$ feasible for (1.1) is CC-M-stationary* $\iff$ $-\nabla f(\hat{x}) \in K_{\hat{x}}(\hat{x})$.

This implies that CC-AM-stationarity is a weaker optimality condition than CC-M-stationarity.

Lemma 4.4 *$\hat{x}$ feasible for (1.1) is CC-M-stationary* $\implies$ *$\hat{x}$ is CC-AM-stationary.*

Proof Since $\hat{x}$ is CC-M-stationary, Lemma 4.3 implies $-\nabla f(\hat{x}) \in K_{\hat{x}}(\hat{x}) \subseteq \limsup_{x \to \hat{x}} K_{\hat{x}}(x)$. The assertion then follows from Theorem 4.1. □

The reverse implication is not true in general as the following example shows.

Example 4.5 ([12], Page 423) Consider the problem

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

with the unique global minimizer $(\frac{1}{2}, 0)^T$. By Theorem 3.2, this point is CC-AM-stationary. On the other hand, we have for any $(\lambda, \gamma) \in \mathbb{R}_+ \times \mathbb{R}$ that

$$\begin{bmatrix} 1 \\ 10 \end{bmatrix} + \lambda \begin{bmatrix} 2(\frac{1}{2} - \frac{1}{2}) \\ 2(0 - 1) \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 - 2\lambda + \gamma \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Hence, it cannot be CC-M-stationary.

The following cone-continuity type condition is sufficient to bridge that gap.

Definition 4.6 A feasible point $\hat{x}$ of (1.1) satisfies the *CC-AM-regularity condition* if

$$\limsup_{x\to\hat{x}} K_{\hat{x}}(x) \subseteq K_{\hat{x}}(\hat{x}).$$

Theorem 4.7 *Let $\hat{x} \in \mathbb{R}^n$ be a CC-AM-stationary point of (1.1) and satisfy the CC-AM-regularity condition. Then $\hat{x}$ is CC-M-stationary.*

Proof Since $\hat{x}$ is a CC-AM-stationary point, Theorem 4.1 yields $-\nabla f(\hat{x}) \in \limsup_{x\to\hat{x}} K_{\hat{x}}(x)$. By Definition 4.6, we then have $-\nabla f(\hat{x}) \in K_{\hat{x}}(\hat{x})$. Hence, $\hat{x}$ is CC-M-stationary by Lemma 4.3. □

The following example shows that the origin, whenever it belongs to the feasible set, is always a CC-M-stationary point and satisfies CC-AM-regularity.

Example 4.8 Suppose that $0 \in \mathbb{R}^n$ is feasible for (1.1) and $f \in C^1(\mathbb{R}^n, \mathbb{R})$ an arbitrary objective function. Then we have

$$I_0(0) = \{1, \dots, n\} \quad \text{and} \quad I_\pm(0) = \emptyset.$$

This implies $K_0(0) = \mathbb{R}^n$ and thus $-\nabla f(0) \in K_0(0)$. By Lemma 4.3 and Lemma 4.4, 0 then is CC-M-stationary and CC-AM-stationary. Moreover, we have $\limsup_{x\to 0} K_0(x) \subseteq \mathbb{R}^n = K_0(0)$, and therefore, 0 satisfies CC-AM-regularity as well.

Borrowing terminology from [4], Theorem 4.7 proves that CC-AM-regularity is a "strict constraint qualification" in the sense that it yields the implication "CC-AM-stationarity $\Longrightarrow$ CC-M-stationarity". The next result shows that CC-AM-regularity is actually the weakest condition, which guarantees that CC-AM-stationary points are already CC-M-stationary.

Theorem 4.9 *Let $\hat{x} \in \mathbb{R}^n$ be feasible for (1.1). Suppose that, for every continuously differentiable function $f \in C^1(\mathbb{R}^n, \mathbb{R})$, the following implication holds:*

$$\hat{x} \text{ is CC-AM-stationary} \quad \Longrightarrow \quad \hat{x} \text{ is CC-M-stationary.} \tag{4.2}$$

Then $\hat{x}$ satisfies CC-AM-regularity.

Proof By Theorem 4.1 and Lemma 4.3, for each $f \in C^1(\mathbb{R}^n, \mathbb{R})$, the assumed implication (4.2) is equivalent to

$$-\nabla f(\hat{x}) \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x) \quad \Longrightarrow \quad -\nabla f(\hat{x}) \in K_{\hat{x}}(\hat{x}).$$

To prove CC-AM-regularity, we need to show $\limsup_{x \to \hat{x}} K_{\hat{x}}(x) \subseteq K_{\hat{x}}(\hat{x})$. To this end, consider an arbitrary $\hat{w} \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x)$ and define $f(x) := -\hat{w}^T x$. Then $f \in C^1(\mathbb{R}^n, \mathbb{R})$ with $-\nabla f(\hat{x}) = \hat{w} \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x)$. By assumption, this implies $-\nabla f(\hat{x}) = \hat{w} \in K_{\hat{x}}(\hat{x})$. This shows $\limsup_{x \to \hat{x}} K_{\hat{x}}(x) \subseteq K_{\hat{x}}(\hat{x})$ and completes the proof. □

Suppose now that $\hat{x} \in \mathbb{R}^n$ is a feasible point of (1.1). As noted in [12], $\hat{x}$ is a CC-M-stationary point if and only if it is a KKT point of the tightened nonlinear program TNLP($\hat{x}$)

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \leq 0,\ h(x) = 0,\ x_i = 0 \quad (i \in I_0(\hat{x})). \tag{4.3}$$

Taking a closer look at Definition 3.1 and using [10, Theorem 3.2] one can see that $\hat{x}$ is a CC-AM-stationary point if and only if it is an AKKT-stationary point of (4.3). Moreover, it follows from (4.1) that $\hat{x}$ satisfies CC-AM-regularity if and only if it satisfies AKKT-regularity with respect to (4.3). Recall from [12, Definition 4.8] that a feasible point $\hat{x}$ of (1.1) is said to satisfy CC-CPLD, if it satisfies the corresponding CPLD for TNLP($\hat{x}$). Using this observation and combining it with some existing results and implications regarding constraint qualifications for standard nonlinear programs, cf. [4, 5], we immediately obtain the following statements.

Corollary 4.10

(a) *If CC-CPLD holds in $\hat{x}$ feasible for* (1.1)*, then so does CC-AM-regularity.*
(b) *If g and h are affine-linear, then CC-AM-regularity holds in every feasible point of* (1.1).

Observe that both CC-CPLD and CC-AM-regularity do not depend on the auxiliary variable y. In contrast to this, CC-ACQ and CC-GCQ are defined using (2.2) and thus depend on both $(\hat{x}, \hat{y})$. The following implications

$$\text{CC-CPLD in } \hat{x} \quad \Longrightarrow \quad \text{CC-ACQ in } (\hat{x}, \hat{y}) \quad \Longrightarrow \quad \text{CC-GCQ in } (\hat{x}, \hat{y})$$

are known from [31] for a feasible point $(\hat{x}, \hat{y})$ of (2.2). For standard NLPs, it is known that AKKT-regularity implies ACQ, cf. [4, Theorem 4.4]. However, as the following example illustrates, for cardinality-constrained problems CC-AM-regularity does not even imply CC-GCQ.

Example 4.11 ([12], Example 4) We consider

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

Then $\hat{x} := (0, 0)^T$ is the unique global minimizer of the problem. By Example 4.8, it also satisfies CC-AM-regularity. On the other hand, if we choose $\hat{y} := (0, 1)^T$, it follows from [12] that $(\hat{x}, \hat{y})$ does not satisfy CC-GCQ, even though $(\hat{x}, \hat{y})$ is a global minimizer of the corresponding reformulated problem.

To close this section, let us remark on the relationship between CC-AM-stationarity and another stationarity concept introduced in [12] called CC-S-stationarity. We first recall the definition of CC-S-stationarity.

Definition 4.12 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 *CC-S-stationary*, if it is CC-M-stationary with $\gamma_i = 0$ for all $i \in I_0(\hat{y})$.

As remarked in [12], CC-S-stationarity in $(\hat{x}, \hat{y})$ corresponds to the KKT conditions of (2.2) and implies CC-M-stationarity of $\hat{x}$. The converse is not true in general, see [12, Example 4]. However, if $(\hat{x}, \hat{y})$ is CC-M-stationary, then it is always possible to replace $\hat{y}$ with another auxiliary variable $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary, see [21, Prop. 2.3].

In [31, Theorem 4.2] it was shown that every local minimizer $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ of (2.2), where CC-GCQ holds, is a CC-S-stationary point. By Theorem 3.2, Theorem 4.7, and [21, Prop. 2.3], we obtain a similar result under CC-AM-regularity.

Corollary 4.13 *Let $\hat{x} \in \mathbb{R}^n$ be a local minimizer of (1.1) such that CC-AM-regularity holds at $\hat{x}$. Then there exists $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a CC-S-stationary point of (2.2).*

5 Application to regularization methods

Let us consider the regularization method from [17], which was adapted for (2.2) in [20]. Let $t \ge 0$ be a regularization parameter and define

$$\varphi : \mathbb{R}^2 \to \mathbb{R}, \quad \varphi((a, b); t) := \begin{cases} (a - t)(b - t) & \text{if } a + b \ge 2t, \\ -\frac{1}{2}[(a - t)^2 + (b - t)^2] & \text{if } a + b < 2t. \end{cases}$$

As it was shown in [17, Lemma 3.1], this function is continuously differentiable with

$$\nabla\varphi((a,b);t) = \begin{cases} \begin{pmatrix} b-t \\ a-t \end{pmatrix} & \text{if } a+b \geq 2t, \\ \begin{pmatrix} -(a-t) \\ -(b-t) \end{pmatrix} & \text{if } a+b < 2t \end{cases}$$

and $\varphi((a,b);0)$ is an NCP-function, i.e., $\varphi((a,b),0) = 0$ if and only if $a \geq 0, b \geq 0, ab = 0$.

Now, let $t > 0$ be a regularization parameter. In order to relax the constraint $x \circ y = 0$ in (2.2) in all four directions, we define the following functions for all $i \in \{1, \ldots, n\}$:

$$\begin{aligned} \Phi^{KS}_{1,i}((x,y);t) &:= \varphi((x_i,y_i);t), & \Phi^{KS}_{2,i}((x,y);t) &:= \varphi((x_i,-y_i);t), \\ \Phi^{KS}_{3,i}((x,y);t) &:= \varphi((-x_i,-y_i);t), & \Phi^{KS}_{4,i}((x,y);t) &:= \varphi((-x_i,y_i);t) \end{aligned}$$

These functions are continuously differentiable and their derivatives with respect to (x, y) can be computed using $\nabla\varphi$ and the chain rule.

For $t > 0$, we now formulate the regularized problem $NLP^{KS}(t)$ as (see Fig. 1)

$$\begin{aligned} \min_{x,y} f(x) \text{ s.t. } & g(x) \leq 0, \quad h(x) = 0, \quad n - e^T y \leq s, \quad y \leq e, \\ & \Phi^{KS}_{j,i}((x,y);t) \leq 0 \quad \forall i = 1, \ldots, n \; \forall j = 1, \ldots, 4. \end{aligned} \tag{5.1}$$

Note that our regularized problem slightly differs from the one used in [12] since we drop the constraint $y \geq 0$ here and instead use two more regularization functions $\Phi^{KS}_{2,i}$ and $\Phi^{KS}_{3,i}$. In the exact case, we obtain the following convergence result.

Theorem 5.1 *Let* $\{t_k\} \downarrow 0$, $\hat{x} \in \mathbb{R}^n$, *and* $\{((x^k, y^k), \lambda^k, \mu^k, \zeta_k, \eta^k, \gamma^{1,k}, \gamma^{2,k}, \gamma^{3,k}, \gamma^{4,k})\}$ *be a sequence of KKT-points of* $NLP^{KS}(t_k)$ *such that* $\{x^k\} \to \hat{x}$. *Then* $\hat{x}$ *is a CC-AM-stationary point of* (1.1).

The proof of this result is similar to the inexact case, which we discuss next. Hence, we omit the details and refer to the proof of the related result in Theorem 5.3. In order to tackle the inexact case, we first need to define inexactness. Consider a standard NLP

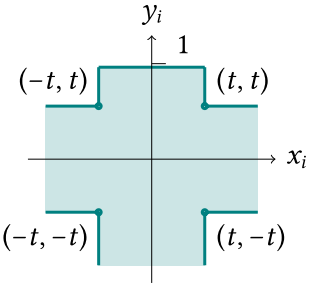

Fig. 1 Illustration of the regularization method

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \le 0,\ h(x) = 0, \tag{5.2}$$

where all functions are assumed to be continuously differentiable. The following definition of inexactness can be found e.g. in [18, Definition 1].

Definition 5.2 Let $x \in \mathbb{R}^n$ and $\epsilon > 0$. We then say that x is an *ϵ-stationary point* of (5.2), if there exists $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p$ such that

- $\left\| \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^p \mu_i \nabla h_i(x) \right\| \le \epsilon$,
- $g_i(x) \le \epsilon, \quad \lambda_i \ge -\epsilon, \quad |\lambda_i g_i(x)| \le \epsilon \quad \forall i = 1, \ldots, m$,
- $|h_i(x)| \le \epsilon \quad \forall i = 1, \ldots, p$.

In the context of MPCCs, it is known that inexactness negatively impacts the convergence theory of this relaxation method, see [18]. The following result shows that this is not the case for cardinality-constrained problems.

Theorem 5.3 *Let $\{t_k\} \downarrow 0$, $\{\epsilon_k\} \downarrow 0$, and $\{(x^k, y^k)\}$ be a sequence of ϵ_k-stationary points of $NLP^{KS}(t_k)$. Suppose that $\{x^k\} \to \hat{x}$. Then $\hat{x}$ is a CC-AM-stationary point.*

Proof By assumption, there exists $\{(\lambda^k, \mu^k, \zeta_k, \eta^k, \gamma^{1,k}, \gamma^{2,k}, \gamma^{3,k}, \gamma^{4,k})\} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \times (\mathbb{R}^n)^5$ such that

$$(Ks_1) \quad \left\| \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \sum_{i=1}^n \sum_{j=1}^4 \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) \right\| \le \epsilon_k,$$

$$(Ks_2) \quad \left\| -\zeta_k e + \eta^k + \sum_{i=1}^n \sum_{j=1}^4 \gamma_i^{j,k} \nabla_y \Phi_{j,i}^{KS}((x^k, y^k); t_k) \right\| \le \epsilon_k,$$

$$(Ks_3) \quad g_i(x^k) \le \epsilon_k, \quad \lambda_i^k \ge -\epsilon_k, \quad |\lambda_i^k g_i(x^k)| \le \epsilon_k \quad \forall i = 1, \ldots, m,$$

$$(Ks_4) \quad |h_i(x^k)| \le \epsilon_k \quad \forall i = 1, \ldots, p,$$

$$(Ks_5) \quad n - e^T y^k - s \le \epsilon_k, \quad \zeta_k \ge -\epsilon_k, \quad |\zeta_k(n - e^T y^k - s)| \le \epsilon_k,$$

$$(Ks_6) \quad y_i^k - 1 \le \epsilon_k, \quad \eta_i^k \ge -\epsilon_k, \quad |\eta_i^k (y_i^k - 1)| \le \epsilon_k \quad \forall i = 1, \ldots, n,$$

$$(Ks_7) \quad \Phi_{j,i}^{KS}((x^k, y^k); t_k) \le \epsilon_k, \quad \gamma_i^{j,k} \ge -\epsilon_k, \quad |\gamma_i^{j,k} \Phi_{j,i}^{KS}((x^k, y^k); t_k)| \le \epsilon_k \quad \forall i = 1, \ldots, n,\ \forall j = 1, \ldots, 4.$$

Let us first note that $\{y^k\}$ is bounded. In fact, by (Ks_6), we have for each $i \in \{1, \ldots, n\}$ that $y_i^k \le 1 + \epsilon_k$ for all $k \in \mathbb{N}$, hence $\{y^k\}$ is bounded from above. Taking this into account and using (Ks_5), i.e., $n - s - \epsilon_k \le e^T y^k$, we also get that $\{y^k\}$ is bounded from below. Since $\{y^k\}$ is bounded, it has a convergent subsequence. By passing to a subsequence, we can assume w.l.o.g. that the whole sequence converges, say $\{y^k\} \to \hat{y}$. In particular, we then have $\{(x^k, y^k)\} \to (\hat{x}, \hat{y})$.

Let us now prove that $(\hat{x}, \hat{y})$ is feasible for (2.2). By (Ks_3) - (Ks_6), we obviously have $g(\hat{x}) \le 0$, $h(\hat{x}) = 0$, $n - e^T \hat{y} \le s$, and $\hat{y} \le e$. Hence, it remains to prove that $\hat{x} \circ \hat{y} = 0$. Suppose that this is not the case. Then there exists an index $i \in \{1, \ldots, n\}$ such that $\hat{x}_i \hat{y}_i \ne 0$. W.l.o.g. let us assume $\hat{x}_i > 0 \wedge \hat{y}_i > 0$, the other three possibilities can be treated analogously. Since $\{x_i^k + y_i^k\} \to \hat{x}_i + \hat{y}_i > 0$ and

$\{t_k\} \downarrow 0$, we can assume w.l.o.g. that $x_i^k + y_i^k \geq 2t_k$ for all $k \in \mathbb{N}$. Hence we have $\Phi_{1,i}^{KS}((x^k, y^k); t_k) = (x_i^k - t_k)(y_i^k - t_k)$. From (Ks_7), we then obtain $\hat{x}_i \hat{y}_i \leq 0$ for the limit, which yields a contradiction since $\hat{x}_i \hat{y}_i > 0$ in this case. Altogether, we can conclude that $\hat{x} \circ \hat{y} = 0$ and, therefore, $(\hat{x}, \hat{y})$ is feasible for (2.2).

By Theorem 2.2, $\hat{x}$ is then feasible for (1.1). Now define

$$w^k := \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \sum_{i=1}^{n}\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k). \tag{5.3}$$

By (Ks_1), we know that $\{w^k\} \to 0$. For all $i \notin I_g(\hat{x})$ we know $\{g_i(x^k)\} \to g_i(\hat{x}) < 0$ and thus can assume w.l.o.g.

$$0 \leq |\lambda_i^k| \leq \frac{\epsilon_k}{|g_i(x^k)|} \quad \forall k \in \mathbb{N},$$

cf. (Ks_3). Letting $k \to \infty$, we then get $\{\lambda_i^k\} \to 0$ and, therefore, $\{\lambda_i^k \nabla g_i(x^k)\} \to 0$. Reformulating (5.3), we then obtain for each $k \in \mathbb{N}$

$$\begin{aligned} w^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) =& \nabla f(x^k) + \sum_{i \in I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \nabla h(x^k)\mu^k \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k), \end{aligned} \tag{5.4}$$

where the left-hand side tends to 0. Now define for each $k \in \mathbb{N}$

$$\hat{\lambda}_i^k := \begin{cases} \lambda_i^k + \epsilon_k & \text{if } i \in I_g(\hat{x}), \\ 0 & \text{else.} \end{cases} \tag{5.5}$$

By (Ks_3), we then have $\{\hat{\lambda}^k\} \subseteq \mathbb{R}_+^m$. Since $\{\epsilon_k\} \to 0$, we also have $\{\epsilon_k \nabla g_i(x^k)\} \to 0$ for each $i \in I_g(\hat{x})$ and reformulating (5.4) yields

$$\begin{aligned} w^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i \in I_g(\hat{x})} \epsilon_k \nabla g_i(x^k) =& \nabla f(x^k) + \nabla g(x^k)\hat{\lambda}^k + \nabla h(x^k)\mu^k \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k), \end{aligned}$$

where the left-hand side converges to 0. For all $i \in I_{\pm}(\hat{x})$ we know $\hat{y}_i = 0$ from the feasibility of $(\hat{x}, \hat{y})$ for (2.2). Assume first that $\hat{x}_i > 0$. Since $\{t_k\} \downarrow 0$ and $\{x_i^k \pm y_i^k\} \to \hat{x}_i \pm \hat{y}_i = \hat{x}_i > 0$, we can assume w.l.o.g. that for all $k \in \mathbb{N}$ the following is true:

$x_i^k + y_i^k \geq 2t_k$:	$\Phi_{1,i}^{KS}((x^k, y^k); t_k) = (x_i^k - t_k)(y_i^k - t_k)$,	$\nabla_x \Phi_{1,i}^{KS}((x^k, y^k); t_k) = (y_i^k - t_k)e_i$
$x_i^k - y_i^k \geq 2t_k$:	$\Phi_{2,i}^{KS}((x^k, y^k); t_k) = (x_i^k - t_k)(-y_i^k - t_k)$,	$\nabla_x \Phi_{2,i}^{KS}((x^k, y^k); t_k) = -(y_i^k + t_k)e_i$
$-x_i^k - y_i^k < 2t_k$:	$\Phi_{3,i}^{KS}((x^k, y^k); t_k) = -\frac{1}{2}[(x_i^k + t_k)^2 + (y_i^k + t_k)^2]$,	$\nabla_x \Phi_{3,i}^{KS}((x^k, y^k); t_k) = -(x_i^k + t_k)e_i$
$-x_i^k + y_i^k < 2t_k$:	$\Phi_{4,i}^{KS}((x^k, y^k); t_k) = -\frac{1}{2}[(x_i^k + t_k)^2 + (y_i^k - t_k)^2)]$,	$\nabla_x \Phi_{4,i}^{KS}((x^k, y^k); t_k) = -(x_i^k + t_k)e_i$

Using (Ks_7), we obtain the following for each case:

- $|\gamma_i^{1,k}\Phi_{1,i}^{KS}((x^k,y^k);t_k)| = |\gamma_i^{1,k}(x_i^k - t_k)(y_i^k - t_k)| \le \epsilon_k$. Since $\{x_i^k - t_k\} \to \hat{x}_i > 0$, we can assume w.l.o.g. that $|x_i^k - t_k| > 0$ for all $k \in \mathbb{N}$ and thus

$$0 \le |\gamma_i^{1,k}(y_i^k - t_k)| \le \frac{\epsilon_k}{|x_i^k - t_k|} \to 0.$$

  Hence $\{\gamma_i^{1,k}(y_i^k - t_k)\} \to 0$, which implies $\{\gamma_i^{1,k}\nabla_x\Phi_{1,i}^{KS}((x^k,y^k);t_k)\} \to 0$.
- $|\gamma_i^{2,k}\Phi_{2,i}^{KS}((x^k,y^k);t_k)| = |\gamma_i^{2,k}(x_i^k - t_k)(y_i^k + t_k)| \le \epsilon_k$. As above, we obtain $\{\gamma_i^{2,k}(-y_i^k - t_k)\} \to 0$ and, $\{\gamma_i^{2,k}\nabla_x\Phi_{2,i}^{KS}((x^k,y^k);t_k)\} \to 0$.
- $|\gamma_i^{3,k}\Phi_{3,i}^{KS}((x^k,y^k);t_k)| = |\gamma_i^{3,k}(-\frac{1}{2})[(x_i^k + t_k)^2 + (y_i^k + t_k)^2]| \le \epsilon_k$. If we define $\beta_k := |-\frac{1}{2}[(x_i^k + t_k)^2 + (y_i^k + t_k)^2]|$, then $\{\beta_k\} \to \frac{1}{2}\hat{x}_i^2 > 0$. Hence, we can assume w.l.o.g. that $\beta_k > 0$ for each $k \in \mathbb{N}$ and obtain

$$0 \le |\gamma_i^{3,k}| \le \frac{\epsilon_k}{\beta_k}.$$

  This implies $\{\gamma_i^{3,k}\} \to 0$ and thus $\{\gamma_i^{3,k}\nabla_x\Phi_{3,i}^{KS}((x^k,y^k);t_k)\} = \{\gamma_i^{3,k}(-(x_i^k + t_k))e_i\} \to 0$.
- $|\gamma_i^{4,k}\Phi_{4,i}^{KS}((x^k,y^k);t_k)| = |\gamma_i^{4,k}(-\frac{1}{2})[(x_i^k + t_k)^2 + (y_i^k - t_k)^2]| \le \epsilon_k$. As above, we obtain $\{\gamma_i^{4,k}\} \to 0$ and thus $\{\gamma_i^{4,k}\nabla_x\Phi_{4,i}^{KS}((x^k,y^k);t_k)\} \to 0$.

For the case $\hat{x}_i < 0$, we can analogously prove

$$\{\gamma_i^{j,k}\nabla_x\Phi_{j,i}^{KS}((x^k,y^k);t_k)\} \to 0 \quad \forall j = 1,\dots,4.$$

Putting things together, we obtain

$$\lim_{k\to\infty} \sum_{i\in I_{\pm}(\hat{x})} \sum_{j=1}^{4} \gamma_i^{j,k}\nabla_x\Phi_{j,i}^{KS}((x^k,y^k);t_k) = 0.$$

Defining

$$A^k := w^k - \sum_{i\notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i\in I_g(\hat{x})} \epsilon_k \nabla g_i(x^k) - \sum_{i\in I_{\pm}(\hat{x})} \sum_{j=1}^{4} \gamma_i^{j,k}\nabla_x\Phi_{j,i}^{KS}((x^k,y^k);t_k)$$

for each $k \in \mathbb{N}$, we obtain

$$A^k = \nabla f(x^k) + \nabla g(x^k)\hat{\lambda}^k + \nabla h(x^k)\mu^k + \sum_{i\in I_0(\hat{x})} \sum_{j=1}^{4} \gamma_i^{j,k}\nabla_x\Phi_{j,i}^{KS}((x^k,y^k);t_k)$$

and $\{A^k\} \to 0$. From the structure of $\nabla_x\Phi_{j,i}^{KS}$, we know $\nabla_x\Phi_{j,i}^{KS}((x^k,y^k);t_k) \in \operatorname{span}\{e_i\}$ for all $i \in I_0(\hat{x})$ and all $j = 1,\dots,4$. Consequently, there exists $\hat{\gamma}_i^k \in \mathbb{R}$ for all $i \in I_0(\hat{x})$ such that

$$\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) = \hat{\gamma}_i^k e_i.$$

If we define $\hat{\gamma}_i^k := 0$ for each $i \in I_{\pm}(\hat{x})$, we obtain for all $k \in \mathbb{N}$,

$$A^k = \nabla f(x^k) + \nabla g(x^k)\hat{\lambda}^k + \nabla h(x^k)\mu^k + \hat{\gamma}^k.$$

By (5.5) and since $\{A^k\} \to 0$, it then follows that $\hat{x}$ is a CC-AM-stationary point. □

Combining Theorem 5.3 with Theorem 4.7 and [21, Prop. 2.3] allows the following conclusions:

Corollary 5.4 *If, in addition to the assumptions in Theorem 5.3, $\hat{x}$ also satisfies CC-AM-regularity, then $\hat{x}$ is CC-M-stationary. Furthermore, there exists a $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary.*

In light of Corollary 4.10, the results obtained in this section are stronger than the result from [12], not only because we take inexactness into account, but also we only need to assume CC-AM-regularity instead of CC-CPLD.

6 Application to augmented Lagrangian methods

In this section, we consider the applicability of a standard (safeguarded) augmented Lagrangian method for NLPs as described in [1, 10] directly to the reformulated problem (2.2). Let us first describe the algorithm. For a given penalty parameter $\alpha > 0$, the *Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian* for (2.2) is defined by

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

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

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

is the shifted quadratic penalty term, cf. the proof of Theorem 3.2. The algorithm, applied to (2.2), is then as follows.

Algorithm 6.1 (Safeguarded Augmented Lagrangian Method)

(S_0) **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 $k \leftarrow 1$.

(S_1) **Update of the iterate:** *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

$$\|\nabla_{(x,y)} L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k)\| \le \epsilon_k. \tag{6.1}$$

(S_2) **Update of the approximate multipliers:**

$$\begin{array}{ll} \lambda^k := \max\{0, \alpha_k g(x^k) + \bar{\lambda}^k\}, & \mu^k := \alpha_k h(x^k) + \bar{\mu}^k, \\ \zeta_k := \max\{0, \alpha_k (n - e^T y^k - s) + \bar{\zeta}_k\}, & \eta^k := \max\{0, \alpha_k (y^k - e) + \bar{\eta}^k\}, \\ \gamma^k := \alpha_k x^k \circ y^k + \bar{\gamma}^k & \end{array}$$

(S_3) **Update of the penalty parameter:** *Define*

$$\begin{aligned} U^k &:= \min\Big\{ -g(x^k), \tfrac{\bar{\lambda}^k}{\alpha_k} \Big\},\ V_k := \min\Big\{ -(n - e^T y^k - s), \tfrac{\bar{\zeta}_k}{\alpha_k} \Big\},\ W^k \\ &:= \min\Big\{ -(y^k - e), \tfrac{\bar{\eta}^k}{\alpha_k} \Big\}. \end{aligned}$$

If $k = 1$ *or*

$$\max\left\{ \begin{array}{c} \|U^k\|,\ \|h(x^k)\|,\ \|V_k\|, \\ \|W^k\|,\ \|x^k \circ y^k\| \end{array} \right\} \le \tau \max\left\{ \begin{array}{c} \|U^{k-1}\|,\ \|h(x^{k-1})\|,\ \|V_{k-1}\|, \\ \|W^{k-1}\|,\ \|x^{k-1} \circ y^{k-1}\| \end{array} \right\}, \tag{6.2}$$

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

(S_4) **Update of the safeguarded multipliers:** *Choose* $\bar{\lambda}^{k+1} \in [0, \lambda_{\max}]^m$, $\bar{\mu}^{k+1} \in [\mu_{\min}, \mu_{\max}]^p$, $\bar{\zeta}_{k+1} \in [0, \zeta_{\max}]$, $\bar{\eta}^{k+1} \in [0, \eta_{\max}]^n$, $\bar{\gamma}^{k+1} \in [\gamma_{\min}, \gamma_{\max}]^n$.

(S_5) *Set* $k \leftarrow k + 1$, *and go to* (S_1).

To measure the infeasibility of a point $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ for (2.2), we consider 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 $\pi_{0,1}((x,y))$ and thus $\nabla \pi_{0,1}((\hat{x}, \hat{y})) = 0$. In this respect, the following result follows from [10, Theorem 6.3] or [19, Theorem 6.2].

Theorem 6.2 *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 6.1. Then* $\nabla \pi_{0,1}((\hat{x}, \hat{y})) = 0$.

Hence, even if a limit point is infeasible, we have at least a stationary point of the constraint violation. In general, for the nonconvex case discussed here, we cannot expect more than this. The remaining part of this section therefore considers the case where a limit point if feasible.

A global convergence analysis of Algorithm 6.1 for cardinality-constrained problems can already be found in the [21], where the authors establish convergence to CC-M-stationary points under a problem-tailored quasi-normality condition. Here, our aim it to verify that Algorithm 6.1, at least under suitable assumptions, satisfies the sequential optimality conditions introduced in this paper. Note that this result is independent from the one in [21] since the quasi-normality condition from there is not related to our sequential regularity assumption, cf. [4] for a corresponding discussion in the NLP case.

Just like in [6, Theorem 5.1], we require the *generalized Kurdyka-Łojasiewicz (GKL) inequality* to be satisfied by $\pi_{0,1}$ at a feasible limit point $(\hat{x}, \hat{y})$ of Algorithm 6.1. Some comments on the GKL inequality are due. A continuously differentiable function $F \in C^1(\mathbb{R}^n, \mathbb{R})$ is said to satisfy the GKL inequality at $\hat{x} \in \mathbb{R}^n$ if there exist $\delta > 0$ and $\psi : B_\delta(\hat{x}) \to \mathbb{R}$ such that $\lim_{x \to \hat{x}} \psi(x) = 0$ and for each $x \in B_\delta(\hat{x})$ we have $|F(x) - F(\hat{x})| \leq \psi(x)\|\nabla F(x)\|$. According to [2, Page 3546], the GKL inequality is a relatively mild condition. For example, it is satisfied at every feasible point of the standard NLP (5.2) provided that all constraint functions are analytic. If we view (2.2) as a standard NLP, then all constraints involving the auxiliary variable y are polynomial in nature and therefore analytic. Thus, if the nonlinear constraints g_i and h_i are analytic, the GKL inequality is then automatically satisfied. In the rest of this section, we consider a feasible limit point $(\hat{x}, \hat{y})$ of the sequence $\{(x^k, y^k)\}$ generated by Algorithm 6.1 and prove that $\hat{x}$ is a CC-AM-stationary point, if the GKL inequality is satisfied by $\pi_{0,1}$ at $(\hat{x}, \hat{y})$.

Before we proceed, we would like to note that, just like in the case of the previous regularization method, the existence of a limit point of $\{x^k\}$ actually already guarantees the boundedness of $\{y^k\}$ on the same subsequence. Hence, we essentially only need to assume the convergence of $\{x^k\}$ and we can then extract a limit point $(\hat{x}, \hat{y})$ of $\{(x^k, y^k)\}$. A proof of this observation is given in [21]. But for simplicity, we assume in our next result that the sequence $\{(x^k, y^k)\}$ has a limit point.

Theorem 6.3 *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 6.1 that is feasible for (2.2). Assume that $\pi_{0,1}$ satisfies the GKL inequality at $(\hat{x}, \hat{y})$, i.e., there exist $\delta > 0$ and $\psi : B_\delta((\hat{x}, \hat{y})) \to \mathbb{R}$ such that $\lim_{(x,y) \to (\hat{x}, \hat{y})} \psi((x, y)) = 0$ and for each $(x, y) \in B_\delta((\hat{x}, \hat{y}))$ we have*

$$|\pi_{0,1}((x, y)) - \pi_{0,1}((\hat{x}, \hat{y}))| \leq \psi((x, y))\|\nabla \pi_{0,1}((x, y))\|. \tag{6.3}$$

Then $\hat{x}$ is a CC-AM-stationary point.

Proof Observe that by (S_2) we have for each $k \in \mathbb{N}$

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

Moreover, using (6.1) and $\{\epsilon_k\} \downarrow 0$, we have

$$\{\Omega^k\} := \{\nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k \circ y^k\} \to 0. \tag{6.4}$$

Now, by (S_2), we also have $\{\lambda^k\} \subseteq \mathbb{R}^m_+$. Furthermore, the sequence of penalty parameters $\{\alpha_k\}$ is nondecreasing and satisfies

$$\alpha_k \geq \alpha_1 > 0 \; \forall k \in \mathbb{N}. \tag{6.5}$$

We distinguish two cases.

Case 1: $\{\alpha_k\}$ is bounded. Then (S_3) implies $\alpha_k = \alpha_K$ for all $k \geq K$ with some sufficiently large $K \in \mathbb{N}$. In particular, (6.2) then holds for each $k \geq K$. This, in turn, implies that $\{\|U^k\|\} \to 0$. Consider an index $i \notin I_g(\hat{x})$. By definition, $\{\bar{\lambda}^k\}$ is bounded. Thus, $\{\frac{\bar{\lambda}^k_i}{\alpha_k}\}$ is bounded as well and therefore has a convergent subsequence. Assume w.l.o.g. that $\{\frac{\bar{\lambda}^k_i}{\alpha_k}\}$ converges and denote with a_i its limit point. We therefore get $0 = \lim_{k\to\infty} |U^k_i| = |\min\{-g_i(\hat{x}), a_i\}|$. Since $-g_i(\hat{x}) > 0$, this implies $a_i = 0$. Consequently, we have $\{g_i(x^k) + \frac{\bar{\lambda}^k_i}{\alpha_k}\} \to g_i(\hat{x}) + a_i = g_i(\hat{x}) < 0$ and thus by (S_2) for all $k \in \mathbb{N}$ large

$$\lambda^k_i = \max\{0, \alpha_k g_i(x^k) + \bar{\lambda}^k_i\} = 0. \tag{6.6}$$

Now consider an $i \in I_{\pm}(\hat{x})$. By assumption, $(\hat{x}, \hat{y})$ is feasible for (2.2), which implies $\hat{y}_i = 0$. By assumption, $\{\alpha_k\}$ is bounded. Moreover, $\{\bar{\gamma}^k_i\}$ is also bounded by definition. Hence, by (S_2), $\{\gamma^k_i\}$ is bounded as well, which implies $\{\gamma^k_i y^k_i\} \to 0$. Let us define $\hat{\gamma}^k \in \mathbb{R}^n$ for all $k \in \mathbb{N}$ by

$$\hat{\gamma}^k_i := \begin{cases} \gamma^k_i y^k_i & \text{if } i \in I_0(\hat{x}), \\ 0 & \text{if } i \in I_{\pm}(\hat{x}). \end{cases}$$

Then by (6.4) we have

$$\begin{aligned} \Omega^k - \sum_{i \in I_{\pm}(\hat{x})} \gamma^k_i y^k_i e_i &= \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \sum_{i \in I_0(\hat{x})} \gamma^k_i y^k_i e_i \\ &= \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \hat{\gamma}^k, \end{aligned}$$

where the left-hand side tends to 0. This, along with (6.6), implies that $\hat{x}$ is CC-AM-stationary.

Case 2: $\{\alpha_k\}$ is unbounded. Since $\{\alpha_k\}$ is nondecreasing, we then have $\{\alpha_k\} \uparrow \infty$. Consider an index $i \notin I_g(\hat{x})$, i.e. with $g_i(\hat{x}) < 0$. By the boundedness of $\{\bar{\lambda}^k_i\}$, we obtain $\lambda^k_i = 0$ for all $k \in \mathbb{N}$ sufficiently large. Condition (6.1) implies

$$\left\| \alpha_k \nabla_{(x,y)} \pi((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\xi}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) \right\| \leq \epsilon_k + \left\| \begin{pmatrix} \nabla f(x^k) \\ 0 \end{pmatrix} \right\|.$$

Since $\{\epsilon_k\} \to 0$ and $\{\nabla f(x^k)\} \to \nabla f(\hat{x})$, the right hand side is convergent and therefore bounded. Thus, $\{\alpha_k \nabla \pi((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k)\}$ is bounded as well. Now observe that

$$
\begin{aligned}
&\alpha_k \nabla_{(x,y)} \pi((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\xi}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) \\
&= \sum_{i=1}^{m} \alpha_k \max\left\{0, g_i(x^k) + \tfrac{\bar{\lambda}_i^k}{\alpha_k}\right\} \begin{pmatrix} \nabla g_i(x^k) \\ 0 \end{pmatrix} + \sum_{i=1}^{p} \alpha_k \left(h_i(x^k) + \tfrac{\bar{\mu}_i^k}{\alpha_k}\right) \begin{pmatrix} \nabla h_i(x^k) \\ 0 \end{pmatrix} \\
&\quad + \alpha_k \max\left\{0, n - e^T y^k - s + \tfrac{\bar{\xi}_k}{\alpha_k}\right\} \begin{pmatrix} 0 \\ -e \end{pmatrix} + \sum_{i=1}^{n} \alpha_k \max\left\{0, y_i^k - 1 + \tfrac{\bar{\eta}_i^k}{\alpha_k}\right\} \begin{pmatrix} 0 \\ e_i \end{pmatrix} \\
&\quad + \sum_{i=1}^{n} \alpha_k \left(x_i^k y_i^k + \tfrac{\bar{\gamma}_i^k}{\alpha_k}\right) \begin{pmatrix} y_i^k e_i \\ x_i^k e_i \end{pmatrix} \\
&= \alpha_k \nabla_{(x,y)} \pi_{0,1}((x^k, y^k)) + z^k
\end{aligned}
$$

where

$$
\begin{aligned}
z^k := &\sum_{i=1}^{m} \left(\max\left\{0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k\right\} - \max\left\{0, \alpha_k g_i(x^k)\right\}\right) \begin{pmatrix} \nabla g_i(x^k) \\ 0 \end{pmatrix} \\
&+ \sum_{i=1}^{p} \bar{\mu}_i^k \begin{pmatrix} \nabla h_i(x^k) \\ 0 \end{pmatrix} \\
&\quad + \left(\max\left\{0, \alpha_k(n - e^T y^k - s) + \bar{\xi}_k\right\} - \max\left\{0, \alpha_k(n - e^T y^k - s)\right\}\right) \begin{pmatrix} 0 \\ -e \end{pmatrix} \\
&+ \sum_{i=1}^{n} \left(\max\left\{0, \alpha_k(y_i^k - 1) + \bar{\eta}_i^k\right\} - \max\left\{0, \alpha_k(y_i^k - 1)\right\}\right) \begin{pmatrix} 0 \\ e_i \end{pmatrix} \\
&+ \sum_{i=1}^{n} \bar{\gamma}_i^k \begin{pmatrix} y_i^k e_i \\ x_i^k e_i \end{pmatrix}.
\end{aligned}
$$

Using the Lipschitz continuity of the mapping $t \mapsto \max\{0, t\}$, we obtain

- $\left|\max\left\{0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k\right\} - \max\left\{0, \alpha_k g_i(x^k)\right\}\right| \leq \bar{\lambda}_i^k$ for all $i = 1, \ldots, m$,
- $\left|\max\left\{0, \alpha_k(n - e^T y^k - s) + \bar{\xi}_k\right\} - \max\left\{0, \alpha_k(n - e^T y^k - s)\right\}\right| \leq \bar{\xi}_k$,
- $\left|\max\left\{0, \alpha_k(y_i^k - 1) + \bar{\eta}_i^k\right\} - \max\left\{0, \alpha_k(y_i^k - 1)\right\}\right| \leq \bar{\eta}_i^k$ for all $i = 1, \ldots, n$.

This, along with the boundedness of the safeguarded multipliers and the convergence of the sequence $\{(x^k, y^k)\}$, then implies that $\{z^k\}$ is bounded. Using the relation

$$
\alpha_k \nabla_{(x,y)} \pi_{0,1}((x^k, y^k)) = \alpha_k \nabla_{(x,y)} \pi((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\xi}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) - z^k,
$$

it follows that $\{\alpha_k \nabla \pi_{0,1}((x^k, y^k))\}$ is also bounded, say $\|\alpha_k \nabla_{(x,y)} \pi_{0,1}((x^k, y^k))\| \leq M$ for some constant $M > 0$. Now, by the feasibility of $(\hat{x}, \hat{y})$, we have $\pi_{0,1}((\hat{x}, \hat{y})) = 0$. Since $\{(x^k, y^k)\} \to (\hat{x}, \hat{y})$, we can assume w.l.o.g. that $(x^k, y^k) \in B_\delta((\hat{x}, \hat{y}))$ for all $k \in \mathbb{N}$, where δ denotes the constant from the GKL inequality. Then (6.3) implies that

$$
|\alpha_k \pi_{0,1}((x^k, y^k))| \leq \psi((x^k, y^k)) \left\|\alpha_k \nabla_{(x,y)} \pi_{0,1}((x^k, y^k))\right\| \leq \psi((x^k, y^k)) M.
$$

Since $\{\psi((x^k, y^k))\} \to 0$, we conclude that $\{\alpha_k \pi_{0,1}((x^k, y^k))\} \to 0$. Now consider an $i \in I_{\pm}(\hat{x})$. By the definition of $\pi_{0,1}$, we obviously have $0 \le \alpha_k (x_i^k y_i^k)^2 \le 2\alpha_k \pi_{0,1}((x^k, y^k))$ for each $k \in \mathbb{N}$. Letting $k \to \infty$, we then obtain $\{\alpha_k (x_i^k y_i^k)^2\} \to 0$. Now since $\hat{x}_i \neq 0$, we can assume w.l.o.g. that $x_i^k \neq 0$ for all $k \in \mathbb{N}$. Then

$$\lim_{k\to\infty} \alpha_k x_i^k (y_i^k)^2 = \lim_{k\to\infty} \frac{1}{x_i^k} \alpha_k (x_i^k y_i^k)^2 = \frac{1}{\hat{x}_i} \cdot 0 = 0.$$

By the feasibility of $(\hat{x}, \hat{y})$ we have $\{y_i^k\} \to \hat{y}_i = 0$. Since $\{\bar{\gamma}_i^k\}$ is bounded, we then obtain

$$\{\gamma_i^k y_i^k\} = \{(\alpha_k x_i^k y_i^k + \bar{\gamma}_i^k) y_i^k\} = \{\alpha_k x_i^k (y_i^k)^2\} + \{\bar{\gamma}_i^k y_i^k\} \to 0.$$

The remainder of the proof is then analogous to the case where $\{\alpha_k\}$ is bounded and we conclude that $\hat{x}$ is a CC-AM-stationary point. □

Similiarly to the relaxation method, combining Theorem 6.3 with Theorem 4.7 and [21, Prop. 2.3] immediately allows the following conclusions:

Corollary 6.4 *If, in addition to the assumptions in Theorem 6.3, $\hat{x}$ also satisfies CC-AM-regularity, then $\hat{x}$ is CC-M-stationary. Furthermore, there then exists a $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary.*

While the regularization method from Sect. 5 generates CC-AM-stationary limit points without any further assumptions, the augmented Lagrangian approach discussed here requires an additional condition (GKL in our case). A counterexample presented in [2] for standard NLPs explains that one has to expect such an additional assumption in the context of augmented Lagrangian methods. In this context, one should also take into account that the regularization method is a problem-tailored solution technique designed specifically for the solution of CC-problems, whereas the augmented Lagrangian method is a standard solver for NLPs. In general, these standard NLP-solvers are expected to have severe problems in solving CC-type problems due to the lack of constraint qualifications and due to the fact that KKT points may not even exist. Nonetheless, the results from this section show that the standard (safeguarded) augmented Lagrangian algorithm is at least a viable tool for the solution of optimization problems with cardinality constraints. We also refer to [16] for a related discussion for MPCCs.

7 Final remarks

In this paper, we introduced CC-AM-stationarity and verified that this sequential optimality condition is satisfied at local minima of cardinality-constrained optimization problems without additional assumptions. Since CC-AM-stationarity is a weaker optimality condition than CC-M-stationarity, we also introduced CC-AM-regularity, a cone-continuity type condition, and showed that CC-AM-stationarity

and CC-M-stationary are equivalent under CC-AM-regularity. We illustrated that CC-AM-regularity is a rather weak assumption, which is satisfied under CC-CPLD and – contrary to NLPs – does not imply CC-ACQ or CC-GCQ.

As an application of the new optimality condition, we showed that the problem-tailored regularization method from [17] generates CC-AM-stationary points in both the exact and the inexact case and without the need for additional assumptions. As a direct consequence, we see that in the inexact case the regularization method still generates CC-M-stationary points under CC-AM-regularity. This is in contrast to MPCCs, where the convergence properties of this method deteriorate in the inexact setting. Additionally, we proved that limit points of a standard (safeguarded) augmented Lagrangian approach satisfy CC-AM-stationarity under an additional mild condition.

Besides the regularization method from [17], there exist a couple of other regularization techniques that can be applied or adapted to cardinality-constrained problems, e.g., the method by Scholtes [28] and the one by Steffensen and Ulbrich [29]. Regarding the local regularization method [29], we have a complete convergence theory similar to Sect. 5, but decided not to include the corresponding (lengthy and technical) results within this paper since, structurally, they are very similar to the corresponding theory presented here. On the other hand, we have to admit that, so far, we were not successful in translating our theory to the global regularization [28], and therefore leave this as part of our future research.

Equivalence of Sequential Constraint Qualifications

In [22], a sequential optimality condition called *AW-stationarity* was introduced. This condition was based the relaxed reformulation of (1.1) from [12], i.e.

$$\min_{x,y} f(x) \quad \text{s.t.} \quad x \in X,\ n - e^T y \le s,\ 0 \le y \le e,\ x \circ y = 0, \tag{A.1}$$

which is (2.2) with the additional constraint $y \ge 0$. To compare CC-AM-stationarity with AW-stationarity we first recall the definition of AW-stationarity from [22].

Definition A.1 Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be feasible for (A.1). Then $(\hat{x}, \hat{y})$ is called *approximately weakly stationary (AW-stationary)* for (A.1), if there exist sequences $\{(x^k, y^k)\} \subseteq \mathbb{R}^n \times \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, $\{\zeta_k\} \subseteq \mathbb{R}_+$, $\{\gamma^k\} \subseteq \mathbb{R}^n$, $\{\nu^k\} \subseteq \mathbb{R}^n$, and $\{\eta^k\} \subseteq \mathbb{R}^n_+$ such that

(a) $\{(x^k, y^k)\} \to (\hat{x}, \hat{y})$,
(b) $\{\nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k\} \to 0$,
(c) $\{-\zeta_k e - \nu^k + \eta^k\} \to 0$,
(d) $\forall i \in \{1, \dots, m\} : \left\{ \min\{-g_i(x^k), \lambda_i^k\} \right\} \to 0$,
(e) $\left\{ \min\{-(n - s - e^T y^k), \zeta_k\} \right\} \to 0$,
(f) $\forall i \in \{1, \dots, n\} : \left\{ \min\{|x_i^k|, |\gamma_i^k|\} \right\} \to 0$,
(g) $\forall i \in \{1, \dots, n\} : \left\{ \min\{y_i^k, |\nu_i^k|\} \right\} \to 0$,
(h) $\forall i \in \{1, \dots, n\} : \left\{ \min\{-(y_i^k - 1), \eta_i^k\} \right\} \to 0$.

Next we derive an equivalent formulation of CC-AM-stationarity.

Proposition A.2 *Let $\hat{x} \in \mathbb{R}^n$ be feasible for (1.1). Then $\hat{x}$ is CC-AM-stationary if and only if there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that*

(a) $\{x^k\} \to \hat{x}$,
(b) $\{\nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k\} \to 0$,
(c) $\forall i \in \{1, \ldots, m\} : \left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \to 0$,
(d) $\forall i \in \{1, \ldots, n\} : \left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \to 0$.

Proof "⇒": Assume first that $\hat{x}$ is CC-AM-stationary. We only need to prove that the corresponding sequences also satisfy conditions (*c*) and (*d*). For all $i \notin I_g(\hat{x})$ we have $g_i(x^k) < 0$ and $\lambda_i^k = 0$ for all k large. For all $i \in I_g(\hat{x})$ we have $\{g_i(x^k)\} \to 0$ and $\lambda_i^k \geq 0$ for all $k \in \mathbb{N}$. In bibliographystyle cases $\left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \to 0$ follows and hence assertion (c) holds. To verify part (d), note that for all $i \in I_{\pm}(\hat{x})$ we have $\gamma_i^k = 0$ for all $k \in \mathbb{N}$ and for all $i \in I_0(\hat{x})$ we know $x_i^k \to 0$. In both cases we obtain $\left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \to 0$.

"⇐": Suppose now that there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that conditions (*a*) – (*d*) hold. If we define $A^k := \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + \gamma^k$ for each $k \in \mathbb{N}$, then $\{A^k\} \to 0$. For all $i \notin I_g(\hat{x})$ we know $\{-g_i(x^k)\} \to -g_i(\hat{x}) > 0$ and $\left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \to 0$, which implies $\{\lambda_i^k\} \to 0$. For each $k \in \mathbb{N}$, define $\hat{\lambda}^k \in \mathbb{R}^m$ by

$$\hat{\lambda}_i^k := \begin{cases} 0 & \text{if } i \notin I_g(\hat{x}), \\ \lambda_i^k & \text{if } i \in I_g(\hat{x}). \end{cases}$$

Then $\{\lambda^k\} \subseteq \mathbb{R}^m_+$ implies $\{\hat{\lambda}^k\} \subseteq \mathbb{R}^m_+$ and by definition we have $\hat{\lambda}_i^k = 0$ for all $i \notin I_g(\hat{x})$ and all $k \in \mathbb{N}$. Next we define

$$B^k := A^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) = \nabla f(x^k) + \nabla g(x^k)\hat{\lambda}^k + \nabla h(x^k)\mu^k + \gamma^k.$$

Here $\{\lambda_i^k\} \to 0$ for all $i \notin I_g(\hat{x})$ implies $\{B^k\} \to 0$. For all $i \in I_{\pm}(\hat{x})$ we know $\{|x_i^k|\} \to |\hat{x}_i| > 0$ and $\left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \to 0$, which implies $\{\gamma_i^k\} \to 0$. For each $k \in \mathbb{N}$ define $\hat{\gamma}^k \in \mathbb{R}^n$ by

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

Then clearly we have $\hat{\gamma}_i^k = 0$ for all $i \in I_{\pm}(\hat{x})$ and all $k \in \mathbb{N}$. Now define for each $k \in \mathbb{N}$

$$C^k := B^k - \sum_{i \in I_{\pm}(\hat{x})} \gamma_i^k e_i = \nabla f(x^k) + \nabla g(x^k)\hat{\lambda}^k + \nabla h(x^k)\mu^k + \hat{\gamma}^k.$$

Here $\{\gamma_i^k\} \to 0$ for all $i \in I_{\pm}(\hat{x})$ implies $\{C^k\} \to 0$. Thus, we conclude that $\hat{x}$ is CC-AM-stationary with the corresponding sequences $\{x^k\}$, $\{\hat{\lambda}^k\}$, $\{\mu^k\}$, and $\{\hat{\gamma}^k\}$. □

Recall from [12] that feasibility of $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ for (A.1) implies feasibility of $\hat{x}$ for (1.1). An immediate consequence of Definition A.1 and Proposition A.2 is thus the following.

Theorem A.3 *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a feasible point of (A.1). If $(\hat{x}, \hat{y})$ is AW-stationary, then $\hat{x}$ is CC-AM-stationary.*

The converse is also true as the following result shows.

Theorem A.4 *Let $\hat{x} \in \mathbb{R}^n$ be a feasible point of (1.1). If $\hat{x}$ is a CC-AM-stationary point, then for all $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for (A.1), it follows that $(\hat{x}, \hat{y})$ is AW-stationary.*

Proof Assume that $\hat{x}$ is CC-AM-stationary. Then there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that conditions (a) – (d) in Proposition A.2 hold. Now consider an arbitrary $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for (A.1) and define

$$y^k := \hat{y}, \quad \zeta_k := 0, \quad \nu^k := 0, \quad \eta^k := 0$$

for each $k \in \mathbb{N}$. Hence conditions (a) – (d) and (f) in Definition A.1 are trivially satisfied. Using the feasibility of $y^k = \hat{y}$, it is easy to see that the remaining conditions also hold. Consequently, $(\hat{x}, \hat{y})$ is AW-stationary. □

An obvious advantage of CC-AM-stationarity over AW-stationarity is that it does not depend on the artificial variable y. Hence, CC-AM-stationarity is a genuine optimality condition for the original problem (1.1). Indeed, one can even derive CC-AM-stationarity directly from (1.1) without referring to the relaxed reformulation by using the Fréchet normal cone of the set $\{x \in \mathbb{R}^n \mid \|x\|_0 \leq s\}$.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**(4), 1286–1309 (2007). https://doi.org/10.1137/060654797. (**ISSN 1052-6234**)
2. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**(6), 3533–3554 (2010). https://doi.org/10.1137/090777189. (**ISSN 1052-6234.**)
3. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011). https://doi.org/10.1080/02331930903578700. (**ISSN 0233-1934**)
4. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**(1), 96–110 (2016). https://doi.org/10.1137/15M1008488. (**ISSN 1052-6234**)
5. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: Strict constraint qualifications and sequential optimality conditions for constrained optimization. Math. Oper. Res. **43**(3), 693–717 (2018). https://doi.org/10.1287/moor.2017.0879. (**ISSN 0364-765X**)
6. Andreani, R., Haeser, G., Secchin, L.D., Silva, P.J.S.: New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. SIAM J. Optim. **29**(4), 3201–3230 (2019). https://doi.org/10.1137/18M121040X. (**ISSN 1052-6234**)
7. Beck, A., Eldar, Y.C.: Sparsity constrained nonlinear optimization: optimality conditions and algorithms. SIAM J. Optim. **23**(3), 1480–1509 (2013). https://doi.org/10.1137/120869778. (**ISSN 1052-6234**)
8. Bertsimas, D., Shioda, R.: Algorithm for cardinality-constrained quadratic optimization. Comput. Optim. Appl. **43**(1), 1–22 (2009). https://doi.org/10.1007/s10589-007-9126-9. (**ISSN 0926-6003**)
9. Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. Math. Program. **74**((2, Ser. A)), 121–140 (1996). https://doi.org/10.1016/0025-5610(96)00044-5. (**ISSN 0025-5610**)
10. Birgin, E. G., Martínez, J. M.: *Practical augmented Lagrangian methods for constrained optimization*, volume 10 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014. ISBN 978-1-611973-35-8. https://doi.org/10.1137/1.9781611973365
11. Branda, M., Bucher, M., Červinka, M., Schwartz, A.: 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). https://doi.org/10.1007/s10589-018-9985-2. (**ISSN 0926-6003**)
12. Burdakov, O.P., Kanzow, C., Schwartz, A.: Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. SIAM J. Optim. **26**(1), 397–425 (2016). https://doi.org/10.1137/140978077. (**ISSN 1052-6234**)
13. Di Lorenzo, D., Liuzzi, G., Rinaldi, F., Schoen, F., Sciandrone, M.: A concave optimization-based approach for sparse portfolio selection. Optim. Methods Softw. **27**(6), 983–1000 (2012). https://doi.org/10.1080/10556788.2011.577773. (**ISSN 1055-6788**)
14. Dong, H., Ahn, M., Pang, J.-S.: Structural properties of affine sparsity constraints. Math. Program. **176**((1–2, Ser. B)), 95–135 (2019). https://doi.org/10.1007/s10107-018-1283-3. (**ISSN 0025-5610**)
15. Feng, M., Mitchell, J.E., Pang, J.-S., Shen, X., Wächter, A.: Complementarity formulations of ℓ_0 -norm optimization problems. Pac. J. Optim. **14**(2), 273–305 (2018). (**ISSN 1348-9151**)
16. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: 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). https://doi.org/10.1137/120868359. (**ISSN 1052-6234**)
17. Kanzow, C., Schwartz, A.: A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. SIAM J. Optim. **23**(2), 770–798 (2013). https://doi.org/10.1137/100802487. (**ISSN 1052-6234**)
18. Kanzow, C., Schwartz, A.: The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited. Math. Oper. Res. **40**(2), 253–275 (2015). https://doi.org/10.1287/moor.2014.0667. (**ISSN 0364-765X**)

19. Kanzow, C., Steck, D., Wachsmuth, D.: An augmented Lagrangian method for optimization problems in Banach spaces. SIAM J. Control Optim. **56**(1), 272–291 (2018). https://doi.org/10.1137/16M1107103. (**ISSN 0363-0129**)
20. Kanzow, C., Mehlitz, P., Steck, D.: Relaxation schemes for mathematical programs with switching constraints. Optim. Method Softw. (2019). https://doi.org/10.1080/10556788.2019.1663425
21. Kanzow, C., Raharja, A. B., Schwartz, A.: An augmented Lagrangian method for cardinality-constrained optimization problems. Technical report, Institute of Mathematics, University of Würzburg, December (2020)
22. Krulikovski, E. H. M., Ribeiro, A. A., Sachine, M.: A sequential optimality condition for mathematical programs with cardinality constraints. *ArXiv e-prints*, (2020). ArXiv:2008.03158
23. Mehlitz, P.: Asymptotic stationarity and regularity for nonsmooth optimization problems. J. Nonsmooth Anal. Optim. (2020). https://doi.org/10.46298/jnsao-2020-6575
24. Murray, W., Shek, H.: A local relaxation method for the cardinality constrained portfolio optimization problem. Comput. Optim. Appl. **53**(3), 681–709 (2012). https://doi.org/10.1007/s10589-012-9471-1. (**ISSN 0926-6003**)
25. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**(4), 963–981 (2000). https://doi.org/10.1137/S1052623497326629. (**ISSN 1052-6234**)
26. Ramos, A.: Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications, and algorithmic consequences. Optim. Methods Softw. (2019). https://doi.org/10.1080/10556788.2019.1702661
27. Rockafellar, R.T., Wets, R. J.-B.: *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, (1998). ISBN 3-540-62772-3.https://doi.org/10.1007/978-3-642-02431-3
28. Scholtes, S.: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM J. Optim. **11**(4), 918–936 (2001)
29. Steffensen, S., Ulbrich, M.: A new relaxation scheme for mathematical programs with equilibrium constraints. SIAM J. Optim. **20**(5), 2504–2539 (2010)
30. Sun, X., Zheng, X., Li, D.: Recent advances in mathematical programming with semi-continuous variables and cardinality constraint. J. Op. Res. Soc. China **1**, 55–77 (2013)
31. Červinka, M., Kanzow, C., Schwartz, A.: Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. Math. Program. **160**((1–2, Ser. A)), 353–377 (2016). https://doi.org/10.1007/s10107-016-0986-6. (**ISSN 0025-5610**)
32. Zheng, X., Sun, X., Li, D., Sun, J.: Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear DC approach. Comput. Optim. Appl. **59**(1–2), 379–397 (2014). https://doi.org/10.1007/s10589-013-9582-3. (**ISSN 0926-6003**)