CONSTRAINT QUALIFICATIONS AND OPTIMALITY CONDITIONS FOR OPTIMIZATION PROBLEMS WITH CARDINALITY CONSTRAINTS[1]

Michal Červinka[2,3], Christian Kanzow[4], and Alexandra Schwartz[5]

Preprint 325   August 2014

[2]Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation
Pod Vodárenskou věží 4
CZ-182 08 Prague, Czech Republic
e-mail: cervinka@utia.cas.cz

[3]Charles University in Prague
Faculty of Social Sciences
Institute of Economic Studies
Opletalova 26
CZ-110 00 Prague, Czech Republic

[4]University of Würzburg
Institute of Mathematics
Emil-Fischer-Str. 30
97074 Würzburg, Germany
e-mail: kanzow@mathematik.uni-wuerzburg.de

[5]Technical University of Darmstadt
Graduate School Computational Engineering
Dolivostraße 15
64293 Darmstadt, Germany
e-mail: schwartz@gsc.tu-darmstadt.de

August 21, 2014

Abstract

This paper considers optimization problems with cardinality constraints. Based on a recently introduced reformulation of this problem as a nonlinear program with continuous variables, we first define some problem-tailored constraint qualifications and then show how these constraint qualifications can be used to obtain suitable optimality conditions for cardinality constrained problems. Here, the (KKT-like) optimality conditions hold under much weaker assumptions than the corresponding result that is known for the somewhat related class of mathematical programs with complementarity constraints.

Key Words: Cardinality constraints, constraint qualifications, optimality conditions, KKT conditions, strongly stationary points.

1 Introduction

Consider the *cardinality-constrained optimization problem*

$$\min_x \ f(x) \quad \text{s.t.} \quad x \in X, \|x\|_0 \leq \kappa \tag{1}$$

with a set $X \subseteq \mathbb{R}^n$ described by some standard constraints

$$X := \left\{ x \in \mathbb{R}^n \mid g_i(x) \leq 0 \ (i = 1, \ldots, m), \ h_i(x) = 0 \ (i = 1, \ldots, p) \right\} \tag{2}$$

and $\|x\|_0$ denoting the number of nonzero components of the vector x. Throughout the paper, we assume implicitly that $f, g_i, h_i : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable, and that the parameter κ is less than n (otherwise, there would be no sparsity constraint on the cardinality of the solution vector).

Problem (1) has many important applications including portfolio optimization problems with constraints on the number of assets [6], the subset selection problem in regression [15], or the compressed sensing technique applied in signal processing [8]. It is, however, very difficult to solve due to the presence of the cardinality constraint $\|x\|_0 \leq \kappa$. It is usually viewed as a mixed-integer problem since it can be reformulated in such a way by introducing suitable binary variables, see [6]. Methods for the solution of cardinality-constrained optimization problems therefore typically apply or adapt techniques from discrete optimization, see [5, 6, 9, 16, 23, 25, 28] and references therein for a couple of examples.

Here we focus on a different approach that follows an observation which is very close to the one made in [10] in the context of the related class of sparse optimization problems. This observation was slightly afterwards (and independently) also made by the authors of the accompanying paper [7] for cardinality-constrained problems: It states that a vector x^* is a solution (global minimum) of (1) with X as in (2) if and only if there exists a vector $y^* \in \mathbb{R}^n$ such that the pair (x^*, y^*) solves the continuous optimization problem

$$
\begin{aligned}
\min_{x,y} \ f(x) \quad \text{s.t.} \quad & g_i(x) \leq 0 \quad \forall i = 1, \ldots, m, \\
& h_i(x) = 0 \quad \forall i = 1, \ldots, p, \\
& e^T y \geq n - \kappa, \\
& 0 \leq y_i \leq 1 \quad \forall i = 1, \ldots, n, \\
& x_i y_i = 0 \quad \forall i = 1, \ldots, n,
\end{aligned} \tag{3}
$$

where $e := (1, \ldots, 1)^T \in \mathbb{R}^n$ denotes the all one vector. Furthermore, every local minimum of (1), where the feasible set is given by (2), yields a local minimum of (3), and the converse is also true under suitable conditions, see [7] for more details.

Hence (3) provides a reformulation of the difficult cardinality constrained optimization problem as a minimization problem in continuous variables. This reformulation has been used to obtain suitable methods for the solution of problem (1) in [7] which are, of course, not able to find a global minimum, but an appropriate stationary point in an efficient way, see also the extensive numerical results presented in [10] for sparse optimization problems. Here we want to exploit the relationship between the two problems (1) and (3) in order to obtain suitable optimality conditions for the original cardinality-constrained problem.

There is, however, still a difficulty since the reformulation (3) typically does not satisfy the standard constraint qualifications known for nonlinear programs, hence the usual KKT theory cannot be applied, at least not directly. In addition, the reformulation (3) looks similar to a mathematical program with complementarity constraints (MPCC) for which, in the meantime, there exists a rich theory, cf. [13, 19] for some background material on MPCCs.

However, direct application of most MPCC-tailored constraint qualifications is not possible in general. Furthermore, and most interestingly, it turns out that a direct inspection of the problem (3) yields results that are much stronger than the corresponding ones known for general MPCCs. Observations of this kind have already been given in [7] and will also arise in this paper, cf. Section 5 for a more detailed discussion.

The paper is organized in the following way: We first recall some basic definitions and a result on polar cones in Section 2. We then derive suitable constraint qualifications tailored to the cardinality-constrained problem (1) in Section 3 and provide sufficient conditions for these constraint qualifications. The main results are given in Section 4 where we use the previously introduced constraint qualifications to prove that a strong (KKT-like) stationarity result holds under very mild conditions (much less restrictive than for the corresponding result known for MPCCs). A detailed comparison of our results for the cardinality-constrained problem with the corresponding results known for MPCCs is given in Section 5.

Notation: e_i denotes the i-th unit vector in $\mathbb{R}^n$. We denote by $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{R}_- := (-\infty, 0]$ the set of nonnegative and nonpositive numbers, respectively. We say that a finite set of vectors a_i $(i \in I)$ and b_j $(j \in J)$ is *positively linearly independent* if there are no nonzero multipliers (λ, μ) with $\lambda_i \geq 0$ $(i \in I)$ such that

$$\sum_{i \in I} \lambda_i a_i + \sum_{j \in J} \mu_j b_j = 0;$$

otherwise these vectors are called *positively linearly dependent*. Note that, from the corresponding context, it should usually be clear which coefficients are sign-constrained and which coefficients are not.

2 Preliminaries

This section presents some preliminary material that will play an essential role in our subsequent analysis. To this end, we recall the definitions of several constraint qualifications for standard nonlinear programs and a result which simplifies the calculation of the cones that play a central role later on.

Let us begin by considering a nonlinear program in the standard form

$$\min_x \ f(x) \quad \text{s.t.} \quad g_i(x) \leq 0 \quad \forall i = 1, \dots, m, \\ h_i(x) = 0 \quad \forall i = 1, \dots, p, \tag{4}$$

Let X denote the feasible set of this program. Then the *(Bouligand) tangent cone* of X at a feasible point $x^* \in X$ is defined by

$$\mathcal{T}_X(x^*) := \big\{ d \in \mathbb{R}^n \ \big| \ \exists \{x^k\} \subseteq X, \exists \{t_k\} \downarrow 0 : x^k \to x^* \ \text{and} \ \lim_{k \to \infty} \frac{x^k - x^*}{t_k} = d \big\}.$$

On the other hand, the corresponding *linearization cone* of X at $x^* \in X$ is given by

$$\mathcal{L}_X(x^*) := \left\{ d \in \mathbb{R}^n \mid \nabla g_i(x^*)^T d \leq 0 \ (i : g_i(x^*) = 0), \nabla h_j(x^*)^T d = 0 \ (j = 1, \ldots, p) \right\}.$$

As part of our subsequent analysis, we will derive suitable expressions for the tangent and linearization cones corresponding to the specially structured nonlinear program from (3).

We also recall that, given an arbitrary cone $\mathcal{C} \subseteq \mathbb{R}^n$, its *dual cone* is defined by

$$\mathcal{C}^* := \{ v \in \mathbb{R}^n \mid v^T d \geq 0 \ \forall d \in \mathcal{C} \},$$

whereas

$$\mathcal{C}^\circ := -\mathcal{C}^* = \{ v \in \mathbb{R}^n \mid v^T d \leq 0 \ \forall d \in \mathcal{C} \}$$

denotes the corresponding *polar cone* of $\mathcal{C}$. In order to keep our notation as concise as possible, we will restrict ourselves to the use of the polar cone from now on. Note, however, that the polar cone is sometimes also denoted by $\mathcal{C}^*$ in the literature, see, for example, [22].

In the special case of a polyhedral convex cone, there is a simple duality relation known from convex analysis, see, e.g., [3, Proposition B.16].

Lemma 2.1 *Let the cones*

$$\mathcal{C}_1 := \{ d \in \mathbb{R}^n \mid a_i^T d \leq 0 \ (i = 1, \ldots, m), \ b_i^T d = 0 \ (i = 1, \ldots, p) \}$$

and

$$\mathcal{C}_2 := \left\{ v \in \mathbb{R}^n \ \middle| \ v = \sum_{i=1}^m \lambda_i a_i + \sum_{i=1}^p \mu_i b_i, \ \lambda_i \geq 0 \ (i = 1, \ldots, m) \right\}$$

be given. Then $\mathcal{C}_1 = \mathcal{C}_2^\circ$ and $\mathcal{C}_2 = \mathcal{C}_1^\circ$.

Based on the previously introduced notions, we are able to restate the definitions of several constraint qualifications known for standard nonlinear programs.

Definition 2.2 *Let x^* be a feasible point of the nonlinear program (4). Then we say that x^* satisfies the*

(a) linear independence constraint qualification (LICQ) *if the gradient vectors*

$$\nabla g_i(x^*) \ (i : g_i(x^*) = 0), \ \nabla h_i(x^*) \ (i = 1, \ldots, p)$$

are linearly independent;

(b) Mangasarian-Fromovitz constraint qualification (MFCQ) *if the gradient vectors $\nabla h_i(x^*) \ (i = 1, \ldots, p)$ are linearly independent and, in addition, there exists a vector $d \in \mathbb{R}^n$ such that $\nabla h_i(x^*)^T d = 0 \ (i = 1, \ldots, p)$ and $\nabla g_i(x^*)^T d < 0 \ (i : g_i(x^*) = 0)$ hold;*

(c) constant rank constraint qualification (CRCQ) *if for any subsets $I_1 \subseteq \{i : g_i(x^*) = 0\}$ and $I_2 \subseteq \{1, \ldots, p\}$ such that the gradient vectors*

$$\nabla g_i(x) \ (i \in I_1), \ \nabla h_i(x) \ (i \in I_2)$$

are linearly dependent in $x = x^*$, they remain linearly dependent for all x in a neighborhood of x^*;

(d) constant linear dependence condition (CPLD) *if for any subsets $I_1 \subseteq \{i : g_i(x^*) = 0\}$ and $I_2 \subseteq \{1, \ldots, p\}$ such that the gradient vectors*

$$\nabla g_i(x) \ (i \in I_1) \quad and \quad \nabla h_i(x) \ (i \in I_2)$$

are positive-linear dependent in $x = x^*$, they are linearly dependent for all x in a neighborhood of x^*;

(e) Abadie constraint qualification (ACQ) *if $\mathcal{T}_X(x^*) = \mathcal{L}_X(x^*)$ holds;*

(f) Guignard constraint qualification (GCQ) *if $\mathcal{T}_X(x^*)^\circ = \mathcal{L}_X(x^*)^\circ$ holds.*

Most of these constraint qualifications are well-known from the literature, see, e.g., [2, 17]. One exception might be the CPLD condition which was introduced in [21] and later shown to be a constraint qualification in [1]. The following implications hold between these conditions:
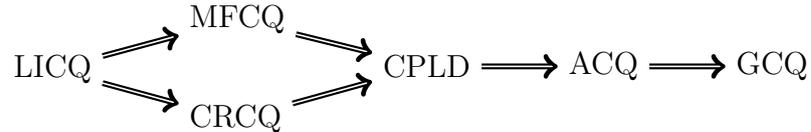
Almost all of these implications follow directly from the corresponding definitions. The only exception is that CPLD implies ACQ, but this observation can be derived from [1, 4]. It is clear from the above diagram that LICQ is the strongest and GCQ is the weakest constraint qualification among those considered here. In fact, it is possible to show that (in a certain sense), GCQ is the weakest possible constraint qualification which guarantees that a local minimum of the program (4) is also a stationary point, see [2] for more details.

3 Abadie- and Guignard-type Constraint Qualifications

Though being one of the weakest constraint qualifications for optimization problems, standard ACQ is usually violated at a feasible point of the cardinality-constrained optimization problem (1), see Example 3.6 for a counterexample. We therefore introduce a problem-tailored modification of the standard ACQ in this section which will be satisfied under much weaker assumptions than the usual ACQ condition. In a similar way, we also develop a variant of the standard GCQ condition. Quite surprisingly, however, this modified GCQ condition turns out to be equivalent to the usual GCQ assumption. We call these modified ACQ- and GCQ-conditions CC-ACQ and CC-GCQ (CC = cardinality constraints).

3.1 Derivation of Abadie- and Guignard-type Constraint Qualifications

Let Z be the feasible set of the continuous optimization problem (3), and let $(x^*, y^*) \in Z$ be any feasible point. Then recall that the tangent cone of Z at (x^*, y^*) is defined by

$$\mathcal{T}_Z(x^*, y^*) = \left\{ d = (d_x, d_y) \mid \exists \{(x^k, y^k)\} \subseteq Z, \exists \{t_k\} \downarrow 0 : \right.$$
$$\left. z^k := (x^k, y^k) \to (x^*, y^*) =: z^* \text{ and } \frac{z^k - z^*}{t_k} \to d \right\}.$$

Furthermore, let us introduce the following index sets:

$$\begin{aligned}
I_g(x^*) &:= \{i \in \{1, \ldots, m\} \mid g_i(x^*) = 0\}, \\
I_0(x^*) &:= \{i \in \{1, \ldots, n\} \mid x_i^* = 0\} \\
I_{\pm 0}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} \mid x_i^* \neq 0, y_i^* = 0\}, \\
I_{00}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} \mid x_i^* = 0, y_i^* = 0\}, \\
I_{0+}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} \mid x_i^* = 0, y_i^* \in (0, 1)\}, \\
I_{01}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} \mid x_i^* = 0, y_i^* = 1\}.
\end{aligned}$$

Note that the subscripts are used to indicate the signs of the variables x_i^* and y_i^*. The definitions of these index sets immediately show that we have the partitions

$$\{1, \ldots, n\} = I_0(x^*) \cup I_{\pm 0}(x^*, y^*)$$

and

$$I_0(x^*) = I_{00}(x^*, y^*) \cup I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*).$$

We further note that the index set $I_{\pm 0}(x^*, y^*)$ could alternatively be called $I_{\pm}(x^*)$ only, since $x_i^* \neq 0$ together with the assumed feasibility of (x^*, y^*) immediately yields $y_i^* = 0$. However, we prefer to use double indices also for this index set since this makes it easier to remember the sign distribution of the two components x_i^* and y_i^*.

Using these index sets, the linearization cone of Z at (x^*, y^*) is given by

$$\begin{aligned}
\mathcal{L}_Z(x^*, y^*) = \left\{ d = (d_x, d_y) \mid \right. & \nabla g_i(x^*)^T d_x \leq 0 \ \forall i \in I_g(x^*), \\
& \nabla h_i(x^*)^T d_x = 0 \ \forall i = 1, \ldots, p, \\
& e^T d_y \geq 0 \ \text{if } e^T y^* = n - \kappa, \\
& e_i^T d_y \geq 0 \ \forall i \in I_{\pm 0}(x^*, y^*) \cup I_{00}(x^*, y^*), \\
& e_i^T d_y \leq 0 \ \forall i \in I_{01}(x^*, y^*), \\
& \left. x_i^* e_i^T d_y + y_i^* e_i^T d_x = 0 \ \forall i = 1, \ldots, n \right\}
\end{aligned}$$

An alternative representation of the linearization cone, that is better suited for our purposes, is given in the following result whose proof is rather obvious and therefore omitted.

Lemma 3.1 *Let $(x^*, y^*) \in Z$ be a feasible point of the program (3). Then the linearization cone of Z at (x^*, y^*) is given by*

$$\mathcal{L}_Z(x^*, y^*) = \left\{ d = (d_x, d_y) \mid \nabla g_i(x^*)^T d_x \leq 0 \ \forall i \in I_g(x^*), \right.$$

5

$$\nabla h_i(x^*)^T d_x = 0 \;\; \forall i = 1, \ldots, p,$$
$$e^T d_y \geq 0 \;\; if \; e^T y^* = n - \kappa,$$
$$e_i^T d_y = 0 \;\; \forall i \in I_{\pm 0}(x^*, y^*),$$
$$e_i^T d_y \geq 0 \;\; \forall i \in I_{00}(x^*, y^*),$$
$$e_i^T d_y \leq 0 \;\; \forall i \in I_{01}(x^*, y^*),$$
$$e_i^T d_x = 0 \;\; \forall i \in I_{01}(x^*, y^*),$$
$$e_i^T d_x = 0 \;\; \forall i \in I_{0+}(x^*, y^*)\big\}.$$

In order to derive suitable constraint qualifications which take into account the particular structure of the nonlinear program (3), we now introduce the *CC-linearization cone* (CC = cardinality constraints) by

$$
\begin{aligned}
\mathcal{L}_Z^{CC}(x^*, y^*) = \big\{ d = (d_x, d_y) \;\; | \;\; & \nabla g_i(x^*)^T d_x \leq 0 \;\; \forall i \in I_g(x^*), \\
& \nabla h_i(x^*)^T d_x = 0 \;\; \forall i = 1, \ldots, p, \\
& e^T d_y \geq 0 \;\; \text{if } e^T y^* = n - \kappa, \\
& e_i^T d_y = 0 \;\; \forall i \in I_{\pm 0}(x^*, y^*), \\
& e_i^T d_y \geq 0 \;\; \forall i \in I_{00}(x^*, y^*), \\
& e_i^T d_y \leq 0 \;\; \forall i \in I_{01}(x^*, y^*), \\
& e_i^T d_x = 0 \;\; \forall i \in I_{01}(x^*, y^*), \\
& e_i^T d_x = 0 \;\; \forall i \in I_{0+}(x^*, y^*), \\
& (e_i^T d_x)(e_i^T d_y) = 0 \;\; \forall i \in I_{00}(x^*, y^*) \big\}.
\end{aligned}
$$

Comparing this definition with the representation of the standard linearization cone from Lemma 3.1, it turns out that the only difference is that we included the last line into the CC-linearization cone. In particular, we therefore have

$$\mathcal{L}_Z^{CC}(x^*, y^*) \subseteq \mathcal{L}_Z(x^*, y^*). \tag{5}$$

While the linearization cone is, by definition, a polyhedral convex cone, simple examples such as Example 3.6 show that both the tangent cone $\mathcal{T}_Z(x^*, y^*)$ and the CC-linearization cone $\mathcal{L}_Z^{CC}(x^*, y^*)$ are nonconvex, specifically, they are the union of finitely many polyhedral convex cones.

This observation can be made precise by introducing certain subsets of the feasible set Z. Recall that $(x^*, y^*) \in Z$ still denotes a given (fixed) feasible point of the program (3). For an arbitrary subset $I \subseteq I_{00}(x^*, y^*)$, we then define the restricted feasible sets

$$
\begin{aligned}
Z_I := \big\{ (x, y) \;\; | \;\; & g_i(x) \leq 0 \;\; \forall i = 1, \ldots, m, \\
& h_i(x) = 0 \;\; \forall i = 1, \ldots, p, \\
& e^T y \geq n - \kappa, \\
& x_i = 0, y_i \in [0, 1] \;\; \forall i \in I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*) \cup I, \\
& y_i = 0 \;\; \forall i \in I_{\pm 0}(x^*, y^*) \cup \big(I_{00}(x^*, y^*) \setminus I \big) \big\},
\end{aligned}
$$

i.e., we split the bi-active index set $I_{00}(x^*, y^*)$ into the two sets I and $I_{00}(x^*, y^*) \setminus I$ and require that $x_i = 0$ on the first set and $y_i = 0$ on the second set. In particular,

we therefore have $Z_I \subseteq Z$ for all $I \subseteq I_{00}(x^*, y^*)$. Furthermore, we have the following result showing that the tangent cone is indeed the union of finitely many polyhedral convex cones.

Proposition 3.2 *Let $(x^*, y^*) \in Z$ be feasible for the program (3). Then the tangent cone and its polar satisfy the following equations:*

(a) $\mathcal{T}_Z(x^*, y^*) = \bigcup_{I \subseteq I_{00}(x^*, y^*)} \mathcal{T}_{Z_I}(x^*, y^*)$.

(b) $\mathcal{T}_Z(x^*, y^*)^\circ = \bigcap_{I \subseteq I_{00}(x^*, y^*)} \mathcal{T}_{Z_I}(x^*, y^*)^\circ$.

Proof: Statement (a) was given in [7]. Formally, that paper assumes that g_i and h_i are (affine-) linear. However, a simple inspection of that proof shows that possibly nonlinear functions g_i and h_i do not change anything. Furthermore, part (b) then follows from (a) and [2, Theorem 3.1.9]. $\qquad\square$

A similar representation can be shown for the CC-linearization cone.

Proposition 3.3 *Let $(x^*, y^*) \in Z$ be feasible for the program (3). Then the CC-linearization cone and its polar satisfy the following equations:*

(a) $\mathcal{L}_Z^{CC}(x^*, y^*) = \bigcup_{I \subseteq I_{00}(x^*, y^*)} \mathcal{L}_{Z_I}(x^*, y^*)$.

(b) $\mathcal{L}_Z^{CC}(x^*, y^*)^\circ = \bigcap_{I \subseteq I_{00}(x^*, y^*)} \mathcal{L}_{Z_I}(x^*, y^*)^\circ$.

Proof: Once again, statement (b) follows immediately from part (a) by applying [2, Theorem 3.1.9] to the nonempty cones $\mathcal{L}_{Z_I}(x^*, y^*)$.

In order to verify part (a), we first observe that the linearization cone of the set Z_I (for an arbitrary index set $I \subseteq I_{00}(x^*, y^*)$) is given by

$$
\begin{aligned}
\mathcal{L}_{Z_I}(x^*, y^*) = \big\{ d = (d_x, d_y) \quad | \quad & \nabla g_i(x^*)^T d_x \leq 0 \;\; \forall i \in I_g(x^*), \\
& \nabla h_i(x^*)^T d_x = 0 \;\; \forall i = 1, \ldots, p, \\
& e^T d_y \geq 0 \;\; \text{if } e^T y^* = n - \kappa, \\
& e_i^T d_x = 0 \;\; \forall i \in I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*) \cup I, \\
& e_i^T d_y \geq 0 \;\; \forall i \in I, \\
& e_i^T d_y \leq 0 \;\; \forall i \in I_{01}(x^*, y^*), \\
& e_i^T d_y = 0 \;\; \forall i \in I_{\pm 0}(x^*, y^*) \cup \big(I_{00}(x^*, y^*) \setminus I \big) \big\}.
\end{aligned}
$$

Now take an arbitrary $(d_x, d_y) \in \mathcal{L}_Z^{CC}(x^*, y^*)$. Then define the index set $I := \{i \in I_{00}(x^*, y^*) \mid e_i^T d_y > 0\}$. This implies $e_i^T d_y = 0$ for all $i \in I_{00}(x^*, y^*) \setminus I$ and $e_i^T d_x = 0$ for all $i \in I$. It therefore follows that $(d_x, d_y) \in \mathcal{L}_{Z_I}(x^*, y^*)$.

Conversely, suppose that (d_x, d_y) belongs to one of the sets $\mathcal{L}_{Z_I}(x^*, y^*)$ for some $I \subseteq I_{00}(x^*, y^*)$. Then the above representation of $\mathcal{L}_{Z_I}(x^*, y^*)$ shows that $(e_i^T d_x)(e_i^T d_y) = 0$ for all $i \in I_{00}(x^*, y^*)$ and $e_i^T d_y \geq 0$ holds for all $i \in I_{00}(x^*, y^*)$. But this immediately implies that $(d_x, d_y) \in \mathcal{L}_Z^{CC}(x^*, y^*)$. $\qquad\square$

As a consequence of the previous results, we now obtain a simple relationship between the Bouligand tangent cone and the CC-linearization cone.

Proposition 3.4 *The inclusions*

$$\mathcal{T}_Z(x^*, y^*) \subseteq \mathcal{L}_Z^{CC}(x^*, y^*) \subseteq \mathcal{L}_Z(x^*, y^*)$$

hold for any feasible point $(x^*, y^*) \in Z$.

Proof: The second inclusion was already noted in (5), so it remains to prove the first one. To this end, first note that the theory from standard nonlinear programs shows that the inclusion

$$\mathcal{T}_{Z_I}(x^*, y^*) \subseteq \mathcal{L}_{Z_I}(x^*, y^*)$$

always holds for an arbitrary index set $I \subseteq I_{00}(x^*, y^*)$. This implies

$$\bigcup_{I \subseteq I_{00}(x^*,y^*)} \mathcal{T}_{Z_I}(x^*, y^*) \subseteq \bigcup_{I \subseteq I_{00}(x^*,y^*)} \mathcal{L}_{Z_I}(x^*, y^*).$$

Together with Propositions 3.2 and 3.3, we therefore obtain

$$\mathcal{T}_Z(x^*, y^*) = \bigcup_{I \subseteq I_{00}(x^*,y^*)} \mathcal{T}_{Z_I}(x^*, y^*) \subseteq \bigcup_{I \subseteq I_{00}(x^*,y^*)} \mathcal{L}_{Z_I}(x^*, y^*) = \mathcal{L}_Z^{CC}(x^*, y^*), \qquad (6)$$

and this completes the proof. $\qquad\qquad\square$

Recall that the standard Abadie constraint qualification (ACQ) requires $\mathcal{T}_Z(x^*, y^*) = \mathcal{L}_Z(x^*, y^*)$. However, taking into account that the tangent cone is usually nonconvex as a finite union of polyhedral convex cones, whereas the linearization cone is polyhedral convex by its definition, it follows that the ACQ assumption typically does not hold in our context. On the other hand, the CC-linearization cone is also nonconvex, and Proposition 3.4 indeed motivates the following modifications of standard ACQ and standard GCQ.

Definition 3.5 *Let* $(x^*, y^*) \in Z$ *be feasible for the program* (3). *Then we say that*

(a) CC-ACQ *holds at* (x^*, y^*) *if* $\mathcal{T}_Z(x^*, y^*) = \mathcal{L}_Z^{CC}(x^*, y^*)$ *holds.*

(b) CC-GCQ *holds at* (x^*, y^*) *if* $\mathcal{T}_Z(x^*, y^*)^\circ = \mathcal{L}_Z^{CC}(x^*, y^*)^\circ$ *holds.*

Note that CC-ACQ implies CC-GCQ, whereas the converse is not true in general. Moreover, standard ACQ also implies CC-ACQ, whereas the following example shows that CC-ACQ might hold also in situations where standard ACQ is violated.

Example 3.6 Consider the simplest possible cardinality-constrained problem

$$\min f(x) \quad \text{s.t.} \quad \|x\|_0 \le 1$$

with $x \in \mathbb{R}^2$ and the corresponding relaxed problem

$$\min_{x,y} f(x) \quad \text{s.t.} \quad y_1 + y_2 \ge 1,$$
$$0 \le y_i \le 1 \quad \forall i = 1, 2,$$
$$x_i y_i = 0 \quad \forall i = 1, 2.$$

If we choose the feasible point (x^*, y^*) with $x^* = (0,0)^T$ and $y^* = (0,1)^T$ and denote the feasible set of the relaxed problem by Z, we obtain

$$
\begin{aligned}
\mathcal{T}_Z(x^*, y^*) &= \{(d_x, d_y) \mid (d_x)_2 = 0, d_y = 0\} \\
&\quad \cup \{(d_x, d_y) \mid d_x = 0, (d_y)_1 \geq 0, (d_y)_2 \leq 0, (d_y)_1 + (d_y)_2 \geq 0\}, \\
\mathcal{L}_Z(x^*, y^*) &= \{(d_x, d_y) \mid (d_x)_2 = 0, (d_y)_1 \geq 0, (d_y)_2 \leq 0, (d_y)_1 + (d_y)_2 \geq 0\},
\end{aligned}
$$

where the tangent cone $\mathcal{T}_Z(x^*, y^*)$ and the linearization cone $\mathcal{L}_Z(x^*, y^*)$ can be calculated, e.g., using Proposition 3.2 and Lemma 3.1, respectively. Here, $\mathcal{T}_Z(x^*, y^*)$ is nonconvex, more precisely, it is the union of two polyhedral convex cones, and consequently $\mathcal{T}_Z(x^*, y^*) \subsetneq \mathcal{L}_Z(x^*, y^*)$. Hence standard ACQ is violated in this example. On the other hand, a simple computation shows that the CC-linearization cone $\mathcal{L}_Z^{CC}(x^*, y^*)$ and the tangent cone $\mathcal{T}_Z(x^*, y^*)$ coincide, hence CC-ACQ holds. $\lozenge$

Example 3.6 shows that CC-ACQ is indeed weaker than standard ACQ. In view of its definition, it is also clear that standard GCQ implies CC-GCQ, and it is rather tempting to believe that CC-GCQ is also strictly weaker than standard GCQ. Surprisingly, however, the following result shows that CC-GCQ and standard GCQ are exactly the same.

Theorem 3.7 *Let $(x^*, y^*) \in Z$ be feasible for (3). Then CC-GCQ holds in (x^*, y^*) if and only if GCQ holds there.*

Proof: We only have to verify that CC-GCQ implies standard GCQ. Using Proposition 3.4, we obtain

$$
\mathcal{L}_Z(x^*, y^*)^\circ \subseteq \mathcal{L}_Z^{CC}(x^*, y^*)^\circ \subseteq \mathcal{T}_Z(x^*, y^*)^\circ.
$$

Since CC-GCQ is assumed to hold, standard GCQ is therefore satisfied if we can show that $\mathcal{L}_Z(x^*, y^*)^\circ = \mathcal{L}_Z^{CC}(x^*, y^*)^\circ$ or, equivalently, that $\mathcal{L}_Z^{CC}(x^*, y^*)^\circ \subseteq \mathcal{L}_Z(x^*, y^*)^\circ$ holds. To verify this inclusion, let us calculate the corresponding polar cones. Using the representation of $\mathcal{L}_Z(x^*, y^*)$ given in Lemma 3.1 and applying Lemma 2.1, we immediately obtain

$$
\begin{aligned}
\mathcal{L}_Z(x^*, y^*)^\circ = \quad &\{w = (w_x, w_y) \mid \\
&w_x = \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) + \sum_{i \in I_{01} \cup I_{0+}} \gamma_i e_i, \\
&w_y = -\delta e + \sum_{i \in I_{\pm 0} \cup I_{00} \cup I_{01}} \nu_i e_i, \\
&\lambda_i \geq 0 \text{ for all } i \in I_g(x^*), \\
&\delta \geq 0 \text{ if } e^T y^* = n - \kappa \text{ and otherwise } \delta = 0, \\
&\nu_i \geq 0 \text{ for all } i \in I_{01}(x^*, y^*), \\
&\nu_i \leq 0 \text{ for all } i \in I_{00}(x^*, y^*)\}.
\end{aligned}
$$

In order to calculate the polar cone of $\mathcal{L}_Z^{CC}(x^*, y^*)$, we use the formula from Proposition 3.3. Using the representation of the linearization cones $\mathcal{L}_{Z_I}(x^*, y^*)$ stated in

the proof of Proposition 3.3 and applying once again Lemma 2.1, we obtain for all $I \subseteq I_{00}(x^*, y^*)$

$$
\begin{aligned}
\mathcal{L}_{Z_I}(x^*, y^*)^\circ \;=\; & \{ w = (w_x, w_y) \mid \\
& w_x = \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i \in I_{01} \cup I_{0+} \cup I} \gamma_i e_i, \\
& w_y = -\delta e + \sum_{i \in I_{\pm 0} \cup I_{00} \cup I_{01}} \nu_i e_i, \\
& \lambda_i \geq 0 \text{ for all } i \in I_g(x^*), \\
& \delta \geq 0 \text{ if } e^T y^* = n - \kappa \text{ and otherwise } \delta = 0, \\
& \nu_i \geq 0 \text{ for all } i \in I_{01}(x^*, y^*), \\
& \nu_i \leq 0 \text{ for all } i \in I \}.
\end{aligned}
$$

Now, take an arbitrary element $w = (w_x, w_y) \in \mathcal{T}_Z(x^*, y^*)^\circ$. Then $w \in \mathcal{L}_{Z_I}(x^*, y^*)^\circ$ for all $I \subseteq I_{00}(x^*, y^*)$. In particular, taking $I = \emptyset$, we obtain suitable scalars $\hat{\lambda}_i, \hat{\mu}_i, \hat{\delta}, \hat{\gamma}_i, \hat{\nu}_i$ such that the above representation holds with $I = \emptyset$ so that, in particular, we have

$$
\hat{\gamma}_i = 0 \quad \forall i \in I_{00}(x^*, y^*).
$$

On the other hand, taking $I = I_{00}(x^*, y^*)$, we get possibly different coefficients $\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\delta}, \tilde{\gamma}_i, \tilde{\nu}_i$ such that the above representation holds with $I = I_{00}(x^*, y^*)$ which, in particular, yields

$$
\tilde{\nu}_i \leq 0 \quad \forall i \in I_{00}(x^*, y^*).
$$

Since none of the constraints occuring in Z_I depends on both x and y, the partial derivatives with respect to x and y are completely independent of each other. Therefore, defining $\lambda_i, \mu_i, \eta, \gamma_i, \nu_i$ by

$$
\begin{aligned}
\lambda_i &:= \hat{\lambda}_i \quad \forall i \in I_g(x^*), \\
\mu_i &:= \hat{\mu}_i \quad \forall i = 1, \ldots, p, \\
\gamma_i &:= \hat{\gamma}_i \quad \forall i \in I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*) \cup I_{00}(x^*, y^*),
\end{aligned}
$$

i.e., using the scalars of the first representation for the derivatives with respect to the x-vector, and

$$
\begin{aligned}
\delta &:= \tilde{\delta}, \\
\nu_i &:= \tilde{\nu}_i \quad \forall i \in I_{00}(x^*, y^*) \cup I_{01}(x^*, y^*) \cup I_{\pm 0}(x^*, y^*),
\end{aligned}
$$

i.e., using the coefficients of the second representation for the derivatives with respect to the y-vector, we altogether obtain

$$
\begin{aligned}
w_x &= \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i \in I_{0+} \cup I_{01} \cup I_{00}} \gamma_i e_i, \\
w_y &= -\delta e + \sum_{i \in I_{00} \cup I_{01} \cup I_{\pm 0}} \nu_i e_i
\end{aligned}
$$

10

with $\lambda_i \geq 0$ for all $i \in I_g(x^*)$, $\delta \geq 0$ and $\delta = 0$ if we have $e^T y^* > n - \kappa$, $\gamma_i = 0$ for all $i \in I_{00}(x^*, y^*)$, $\nu_i \leq 0$ for all $i \in I_{00}(x^*, y^*)$, and $\nu_i \geq 0$ for all $i \in I_{01}(x^*, y^*)$. This shows that $w = (w_x, w_y) \in \mathcal{L}_Z(x^*, y^*)^\circ$, and therefore completes the proof. $\qquad\square$

Following the terminology coined in [12], the condition $\mathcal{L}_Z(x^*, y^*)^\circ = \mathcal{L}_Z^{CC}(x^*, y^*)^\circ$ from the above proof could be called the *CC-intersection property*.

3.2 Sufficient Conditions for CC-ACQ

This section presents some conditions which imply that CC-ACQ (hence also CC-GCQ and thus standard GCQ) holds. A first and very simple result is contained in our next lemma.

Lemma 3.8 *Let $(x^*, y^*) \in Z$ be feasible for the program (3), and assume that each of the restricted feasible sets Z_I with $I \subseteq I_{00}(x^*, y^*)$ satisfies the standard ACQ. Then CC-ACQ (hence also GCQ) holds at (x^*, y^*).*

Proof: The statement follows immediately from (6), since the only inclusion in that formula is now an equation due to the assumed ACQ condition. $\qquad\square$

Since linear constraints automatically satisfy the standard ACQ condition, it follows that CC-ACQ also holds in the case of (affine-) linear functions g_i, h_i.

Corollary 3.9 *Let $(x^*, y^*) \in Z$ be feasible for the program (3), and assume that the functions g_i and h_i are all (affine-) linear. Then CC-ACQ (hence also GCQ) holds at (x^*, y^*).*

Our next aim is to show that the assertion of Corollary 3.9 also holds for possibly nonlinear functions g_i and h_i. To this end, we first recall a number of other tailored constrained qualifications that were introduced in [7] for cardinality-constrained optimization problems. To motivate these definitions, let (x^*, y^*) be a feasible point of the relaxed program (3), and define the corresponding tightened nonlinear program TNLP(x^*):

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

We then say that (x^*, y^*) satisfies a constraint qualification for the relaxed problem (3) if x^* satisfies the corresponding standard constraint qualification for TNLP(x^*). In this way, we obtain the following definitions.

Definition 3.10 *Let (x^*, y^*) be feasible for the relaxed problem (3). Then (x^*, y^*) satisfies*

(a) CC-LICQ *if the gradients*

$$\nabla g_i(x^*) \ (i \in I_g(x^*)), \ \nabla h_i(x^*) \ (i = 1, \dots, p), \ e_i \ (i \in I_0(x^*))$$

are linearly independent;

(b) CC-MFCQ *if the gradients*

$$\nabla g_i(x^*) \ (i \in I_g(x^*)), \quad and \quad \nabla h_i(x^*) \ (i = 1, \ldots, p), \ e_i \ (i \in I_0(x^*))$$

are positively linearly independent;

(c) CC-CRCQ *if for any subsets* $I_1 \subseteq I_g(x^*)$, $I_2 \subseteq \{1, \ldots, p\}$, *and* $I_3 \subseteq I_0(x^*)$ *such that the gradients*

$$\nabla g_i(x) \ (i \in I_1), \ \nabla h_i(x) \ (i \in I_2), \ e_i \ (i \in I_3)$$

are linearly dependent in $x = x^*$, *they remain linearly dependent in a neighborhood of* x^*;

(d) CC-CPLD *if for any subsets* $I_1 \subseteq I_g(x^*)$, $I_2 \subseteq \{1, \ldots, p\}$, *and* $I_3 \subseteq I_0(x^*)$ *such that the gradients*

$$\nabla g_i(x) \ (i \in I_1), \quad and \quad \nabla h_i(x) \ (i \in I_2), \ e_i \ (i \in I_3)$$

are positively linearly dependent in $x = x^*$, *they are linearly dependent in a neighborhood of* x^*.

Note that all these constraint qualifications depend on the vector x^* only, and not on the vector pair (x^*, y^*). Hence these conditions may be viewed as constraint qualifications for the original cardinality constrained optimization problem (1).

We claim that the following implications among all these constraint qualifications hold:

$$\text{CC-LICQ} \Rrightarrow \begin{matrix} \text{CC-MFCQ} \\ \\ \text{CC-CRCQ} \end{matrix} \Rrightarrow \text{CC-CPLD} \Longrightarrow \text{CC-ACQ} \Longrightarrow \text{CC-GCQ}$$

Note that these implications are the direct counterparts of those known for the corresponding standard constraint qualifications, cf. Section 2. Most of these implications (therefore) follow directly from the corresponding definitions. The only nontrivial part is that CC-CPLD implies CC-ACQ. To verify this statement, we begin with a preliminary result.

Lemma 3.11 *Let* $(x^*, y^*) \in Z$ *be feasible for the relaxed program* (3), *and suppose that CC-CPLD holds at* (x^*, y^*). *Then, for any* $I \subseteq I_{00}(x^*, y^*)$, *standard CPLD is satisfied for the restricted feasible set* Z_I.

Proof: Consider a fixed index set $I \subseteq I_{00}(x^*, y^*)$. The corresponding feasible set Z_I then has constraints that either depend on x or on y, but never on both. Consequently, it suffices to verify CPLD for the constraints depending on x and on y separately. Since all constraints depending on y are linear, they satisfy CRCQ and thus also CPLD. It therefore suffices to show that the standard CPLD condition holds for those constraints that depend on x only. We can restrict ourselves to the

gradient vectors that arise by taking the partial derivatives with respect to the x-variables only, since the partial derivatives with respect to the y-variables are in this case all zero. More precisely, this means that we have to show that for all subsets $I_1 \subseteq I_g(x^*), I_2 \subseteq \{1, \ldots, p\}$, and $I_3 \subseteq I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*) \cup I$ such that the gradient vectors

$$\nabla g_i(x^*) \ (i \in I_1), \quad \text{and} \quad \nabla h_i(x^*) \ (i \in I_2), \ e_i \ (i \in I_3)$$

are positively linearly dependent, they are linearly dependent in a whole neighborhood of x^*. However, since I_3 may, in particular, be viewed as a subset of $I_0(x^*)$, this statement follows immediately from the definition of CC-CPLD. $\qquad \square$

Similar to the previous result, one can also show that CC-CRCQ implies that a piecewise CRCQ condition holds, by which we mean that standard CRCQ holds for all sets $Z_I, I \subseteq I_{00}(x^*, y^*)$. On the other hand, CC-LICQ does not imply piecewise LICQ. In fact, piecewise LICQ would require that the following gradients (with respect to x and y) are linearly independent for all subsets $I \subseteq I_{00}(x^*, y^*)$:

$$\begin{pmatrix} \nabla g_i(x^*) \\ 0 \end{pmatrix} (i \in I_g(x^*)), \ \begin{pmatrix} \nabla h_i(x^*) \\ 0 \end{pmatrix} (i = 1, \ldots, p), \ \begin{pmatrix} 0 \\ -e \end{pmatrix} (\text{if } e^T y^* = n - \kappa),$$

$$\begin{pmatrix} e_i \\ 0 \end{pmatrix} (i \in I_{0+} \cup I_{01} \cup I), \ \begin{pmatrix} 0 \\ -e_i \end{pmatrix} (i \in I), \ \begin{pmatrix} 0 \\ e_i \end{pmatrix} (i \in I_{01}), \ \begin{pmatrix} 0 \\ e_i \end{pmatrix} (i \in I_{\pm 0} \cup (I_{00} \setminus I)).$$

While CC-LICQ implies that those gradients which have nonzero entries with respect to the x-part are linearly independent, it is clear that the other gradients are linearly dependent whenever $e^T y^* = n - \kappa$ holds and the set $I_{0+} := I_{0+}(x^*, y^*)$ is empty, a situation that typically holds at a solution of the cardinality-constrained optimization problem.

In a similar way, it is also possible to see that CC-MFCQ does, in general, not imply a piecewise MFCQ condition. We skip the corresponding details especially since this observation will not be used subsequently.

Lemma 3.11 allows us to state the following generalization of Corollary 3.9.

Theorem 3.12 *Let $(x^*, y^*) \in Z$ be feasible for the program (3), and assume that CC-CPLD holds. Then CC-ACQ (hence also GCQ) holds at (x^*, y^*).*

Proof: Since CC-CPLD holds at (x^*, y^*), it follows from Lemma 3.11 that standard CPLD holds for each of the feasible sets $Z_I, I \subseteq I_{00}(x^*, y^*)$. But standard CPLD implies that standard ACQ holds for each of the feasible sets Z_I. The statement therefore follows immediately from Lemma 3.8. $\qquad \square$

4 Stationarity Conditions

This section shows that, in every local minimum (x^*, y^*) of the relaxed program (3), in which a suitable CC-constraint qualification holds, certain KKT-type optimality conditions are satisfied. We distinguish two optimality conditions here, one is

called strong stationarity (and is equivalent to the standard KKT conditions), and the other one is called M-stationarity. We show that strong stationarity provides a necessary optimality condition under the CC-GCQ assumption. Under certain assumptions, strong stationarity is also a sufficient condition for a local minimum. The slightly weaker M-stationarity condition is therefore a necessary optimality condition under the CC-GCQ condition, too. This type of stationary points arises quite naturally as limit points within the algorithmic framework in [7].

We begin by stating the two stationary concepts that will be used in our analysis. To shorten the notation, we sometimes abbreviate index sets such as $I_{00} := I_{00}(x^*, y^*)$ when the reference point (x^*, y^*) is clear from the context.

Definition 4.1 *Let (x^*, y^*) be feasible for the relaxed program* (3). *Then (x^*, y^*) is called*

(a) S-stationary *(S = strong) if there exist multipliers $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that the following conditions hold:*

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i \in I_{01} \cup I_{0+}} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*).$$

(b) M-stationary *(M = Mordukhovich) if there exist multipliers $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that the following conditions hold:*

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i \in I_{01} \cup I_{0+} \cup I_{00}} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*).$$

The notions of S- and M-stationarity are motivated by a similar terminology used in the context of mathematical programs with complementarity constraints (MPCCs), see Section 5 for more details. Note that S-stationarity requires $\gamma_i = 0$ for all indices i such that $y_i^* = 0$, whereas M-stationarity says that this has to hold only for those indices i where $x_i^* \neq 0$; since the feasibility of the vector pair (x^*, y^*) then implies that $y_i^* = 0$, it follows that M-stationarity is a weaker stationarity concept than S-stationarity since it does not require anything for the multipliers γ_i for the bi-active indices, where $x_i^* = 0$ and $y_i^* = 0$. We further note that S-stationarity was noted to be equivalent to the usual KKT conditions of the relaxed program (3) in [7].

As a consequence of the previous section, we can show that a local minimum satisfying CC-GCQ is a strongly stationary point of the relaxed program (3).

Theorem 4.2 *Let (x^*, y^*) be a local minimum of* (3) *such that CC-GCQ holds at (x^*, y^*). Then (x^*, y^*) is an S-stationary point.*

Proof: Since CC-GCQ holds at (x^*, y^*) by assumption, it follows from Theorem 3.7 that standard GCQ holds at (x^*, y^*). Under standard GCQ, however, the usual KKT conditions are necessary optimality conditions at the local minimum (x^*, y^*), see, e.g., [2]. On the other hand, these KKT conditions are shown to be equivalent

to strong stationarity in [7]. Hence the assertion follows. $\qquad\square$

Since, by Corollary 3.9, CC-GCQ holds if both g_i and h_i are linear, we re-obtain the following result from [7] as a special case of our theory.

Corollary 4.3 *Assume that all functions g_i and h_i are linear, and let (x^*, y^*) be a local minimum of the corresponding relaxed program (3). Then (x^*, y^*) is an S-stationary point.*

For standard nonlinear programs it is well-known that any KKT point yields a *global* minimum provided that we have a convex program. The cardinality-constrained program is, of course, nonconvex, therefore we cannot expect a result of this kind. However, under a convexity-type condition, the following result shows that every strongly stationary point yields a *local* minimum of the relaxed program (3). This observation is similar to one in the MPCC-setting, see [27].

Theorem 4.4 *Assume that f and each g_i are convex and each h_i is linear. Let (x^*, y^*) be an S-stationary point of the relaxed program (3). Then (x^*, y^*) is a local minimum of this program.*

Proof: Let (x, y) be an arbitrary feasible point of the relaxed program (3). We then obtain

$$
\begin{aligned}
f(x) &\geq f(x^*) + \nabla f(x^*)^T (x - x^*) \\
&= f(x^*) - \sum_{i \in I_g(x^*)} \underbrace{\lambda_i}_{\geq 0} \underbrace{\nabla g_i(x^*)^T (x - x^*)}_{\leq g_i(x) - g_i(x^*) = g_i(x) \leq 0} \\
&\quad - \sum_{i=1}^{p} \mu_i \underbrace{\nabla h_i(x^*)^T (x - x^*)}_{= h_i(x) - h_i(x^*) = 0} - \sum_{i \in I_{01} \cup I_{0+}} \gamma_i e_i^T (x - x^*) \\
&\geq f(x^*) - \sum_{i \in I_{01} \cup I_{0+}} \gamma_i (x_i - x_i^*).
\end{aligned}
$$

Since we take the last sum only over all indices $i \in I_{01}(x^*, y^*) \cup I_{0+}(x^*, y^*)$, it follows that, in a sufficiently small neighborhood of (x^*, y^*), we still have $y_i \neq 0$ for all $i \in I_{01}(x^*, y^*) \cup I_{0+}(x^*, y^*)$, hence the feasibility of the pair (x, y) yields $x_i = 0$. Consequently, we have $f(x) \geq f(x^*)$ for all (x, y) in a sufficiently small neighborhood of (x^*, y^*). $\qquad\square$

Note that the previous proof shows that the S-stationary point (x^*, y^*) is actually a *global* minimum of the relaxed program (3) under the convexity-type assumption provided that $\gamma_i = 0$ for all indices i such that $y_i^* \neq 0$. Of course, this assumption is usually not satisfied.

5 Comparison with MPCCs

This section gives a detailed comparison between (a special class of) cardinality-constrained optimization problems on the one hand and MPCCs on the other hand.

Despite several similarities, it turns out that both problems have substantially different properties. In particular, this justifies to treat cardinality-constrained problems separately.

A mathematical program with complementarity constraints (MPCC) is an optimization problem of the form

$$\min_x f(x) \quad \text{s.t.} \quad g_i(x) \leq 0 \quad \forall i = 1, \ldots, m,$$
$$h_i(x) = 0 \quad \forall i = 1, \ldots, p, \tag{7}$$
$$G_i(x) \geq 0, H_i(x) \geq 0, G_i(x)H_i(x) = 0 \quad \forall i = 1, \ldots, q,$$

see [13, 19] for more information on this problem class.

Consequently, in the case where the inequality constraints $g(x) \leq 0$ contain nonnegativity constraints $x \geq 0$, the cardinality-constrained problem (3) is (after a redefinition of g) of the form

$$\min_{x,y} f(x) \quad \text{s.t.} \quad g_i(x) \leq 0 \quad \forall i = 1, \ldots, m,$$
$$h_i(x) = 0 \quad \forall i = 1, \ldots, p,$$
$$e^T y \geq n - \kappa, \tag{8}$$
$$y_i \leq 1 \quad \forall i = 1, \ldots, n,$$
$$x_i \geq 0, y_i \geq 0, x_i y_i = 0 \quad \forall i = 1, \ldots, n,$$

i.e. it is an MPCC. The situation $x \geq 0$ occurs, for example, in portfolio optimization, see [6].

In order to compare the results obtained in this paper for cardinality constrained problems with those known for MPCCs, let us state the corresponding definitions for MPCCs, see [18, 24, 26] for some discussion and a derivation of these stationarity concepts.

Definition 5.1 *Let x^* be feasible for* (7). *Then x^* is called*

(a) W-stationary (W = weakly), *if there are multipliers $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p, \gamma \in \mathbb{R}^q$ and $\nu \in \mathbb{R}^q$ such that the following conditions hold:*

$$\nabla f(x^*) + \sum_{i:g_i(x^*)=0} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) - \sum_{i:G_i(x^*)=0} \gamma_i \nabla G_i(x^*) - \sum_{i:H_i(x^*)=0} \nu_i \nabla H_i(x^*) = 0,$$
$$\lambda_i \geq 0 \quad \forall i : g_i(x^*) = 0;$$

(b) C-stationary (C = Clarke), *if it is W-stationary and, in addition, it holds that $\gamma_i \nu_i \geq 0$ for all i such that $G_i(x^*) = 0$ and $H_i(x^*) = 0$;*

(c) M-stationary (M = Mordukhovich), *if it is W-stationary and, in addition, for all i such that $G_i(x^*) = 0$ and $H_i(x^*) = 0$, we either have $\gamma_i, \nu_i \geq 0$ or $\gamma_i \nu_i = 0$;*

(d) S-stationary (S = strongly), *if it is W-stationary and, in addition, it holds that $\gamma_i, \nu_i \geq 0$ for all i such that $G_i(x^*) = 0$ and $H_i(x^*) = 0$.*

Note that S-stationarity implies M-stationarity, M-stationarity implies C-stationarity, and C-stationarity implies W-stationarity. Furthermore, there exist examples which show that each of these implications is strict, i.e. none of these concepts coincides for general MPCCs.

Now, there are two different ways to look at the cardinality-constrained problem (8) with nonnegativity constraints on the variables x: One way is to view this as a special cardinality-constrained problem with the nonnegativity constraints $x \geq 0$ as additional inequality constraints, and the other way is to view this problem as an MPCC, with the nonnegativity constraints being a part of the complementarity conditions. Taking the first point of view, we write down the S- and M-stationarity conditions (in the sense of Definition 4.1) in the following result.

Lemma 5.2 *Let (x^*, y^*) be feasible for (8). Then (x^*, y^*) is*

(a) *S-stationary if and only if there exist suitable multipliers such that the following conditions hold:*

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) + \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*),$$
$$\gamma_i \leq 0 \quad \forall i \in I_{00}(x^*, y^*).$$

(b) *M-stationary if and only if there exist suitable multipliers such that the following conditions hold:*

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) + \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*).$$

Proof: (a) Applying the S-stationarity conditions from Definition 4.1 to (8) gives

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) - \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \lambda_i^+ e_i + \sum_{i=1}^p \mu_i \nabla h_i(x^*) + \sum_{i \in I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*),$$
$$\lambda_i^+ \geq 0 \quad \forall i \in I_{00}(x^*, y^*) \cup I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*).$$

Fusing λ^+ and γ to one multiplier yields the desired statement.

(b) Writing down the M-stationarity conditions from Definition 4.1 to (8) yields

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) - \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \lambda_i^+ e_i + \sum_{i=1}^p \mu_i \nabla h_i(x^*) + \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*),$$
$$\lambda_i^+ \geq 0 \quad \forall i \in I_{00}(x^*, y^*) \cup I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*).$$

Replacing $\gamma_i - \lambda_i^+$ by a new γ_i for each index $i \in I_{00}(x^*, y^*) \cup I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*)$ gives the desired representation of M-stationarity. $\qquad\qquad\Box$

Note the close relation between the S- and M-stationarity conditions for the cardinality-constrained problem from (3) and the corresponding stationarity conditions for the specially structured problem (8) involving nonnegativity constraints: S-stationarity differs only in the last sum where now also the bi-active index set I_{00} is included (with some sign constraints on the corresponding multipliers). As for M-stationarity, there is absolutely no difference though the problem itself is different!

Our definition of M- and S-stationarity for cardinality-constrained problems was motivated by the corresponding concepts for MPCCs, and indeed in the special case (8), the definitions turn out to be identical.

Lemma 5.3 *Let (x^*, y^*) be feasible for (8). Then (x^*, y^*) is S-stationary (M-stationary) in the sense of Definition 4.1 if and only if it is S-stationary (M-stationary) in the sense of Definition 5.1.*

Proof: (a) We first verify the statement for S-stationary points. Using the same index sets as before, the S-stationarity conditions from Definition 5.1 read

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) - \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$

$$-\delta e + \sum_{i \in I_{01}} \nu_i e_i - \sum_{i \in I_{\pm 0} \cup I_{00}} \nu_i e_i = 0,$$

$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*),$$

$$\gamma_i \geq 0, \ \nu_i \geq 0 \quad \forall i \in I_{00}(x^*, y^*),$$

$$\delta \geq 0 \quad \text{if} \quad e^T y^* = n - \kappa, \quad \text{else} \quad \delta = 0,$$

$$\nu_i \geq 0 \quad \forall i \in I_{01}(x^*, y^*).$$

Since the conditions on δ and ν are can always be satisfied by choosing $\delta = 0$, $\nu = 0$, the S-stationarity conditions from Definition 5.1 hold if and only if the following conditions hold:

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) - \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$

$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*),$$

$$\gamma_i \geq 0 \quad \forall i \in I_{00}(x^*, y^*).$$

In view of Lemma 5.2 and replacing γ_i by $-\gamma_i$ everywhere, these are precisely the S-stationarity conditions from Definition 4.1.

(b) We next verify the corresponding statement for M-stationary points. The proof is completely analogous to the one for S-stationary points, but it is stated here since it yields an interesting observation that is stated formally in the subsequent Remark 5.4.

Let us first write down the M-stationarity conditions for problem (8) in the sense of Definition 5.1:

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) - \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$

$$-\delta e + \sum_{i \in I_{01}} \nu_i e_i - \sum_{i \in I_{\pm 0} \cup I_{00}} \nu_i e_i = 0,$$

$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*),$$

$$\gamma_i \geq 0, \nu_i \geq 0 \quad \text{or} \quad \gamma_i \nu_i = 0 \quad \forall i \in I_{00}(x^*, y^*),$$

$$\delta \geq 0 \quad \text{if} \quad e^T y^* = n - \kappa, \quad \text{else} \quad \delta = 0,$$

$$\nu_i \geq 0 \quad \forall i \in I_{01}(x^*, y^*).$$

Using once again that the conditions on δ and ν can always be satisfied by choosing $\delta = 0$, $\nu = 0$, it follows that there exist multipliers satisfying the M-stationarity conditions from Definition 5.1 if and only if there exist multipliers such that the following simplified conditions hold:

$$\nabla f(x^*) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) - \sum_{i \in I_{00} \cup I_{0+} \cup I_{01}} \gamma_i e_i = 0, \tag{9}$$

$$\lambda_i \geq 0 \quad \forall i \in I_g(x^*).$$

But these are precisely the M-stationarity conditions given in Lemma 5.2 (note that we can change the sign of the γ_i without loss of generality). $\qquad \square$

A simple inspection of part (b) in the previous proof shows that the M-stationarity conditions of problem (8) in the sense of Definition 5.1 are satisfied by some multipliers if and only if the C-stationarity conditions hold for this problem, and this, in turn, is also equivalent to the satisfaction of the W-stationarity conditions. Hence we have the following observation.

Remark 5.4 The M-, C-, and W-stationarity points in the sense of Definition 5.1 are the same for the particular MPCC (8) that arises from our cardinality-constrained problem in the case where all variables are assumed to be nonnegative. In view of Lemma 5.3, this means that, in this particular situation, the M-stationary points in the sense of Definition 4.1 are also the same as the W- and C-stationary points in the sense of Definition 5.1.

On the other hand, S- and M-stationarity are different concepts. This is shown by the following example.

Example 5.5 Consider the cardinality-constrained problem (1), (2) with $n = 2, \kappa = 1$, objective function $f(x) := x_2 - x_1$ and feasible set $X := \{x \mid x \geq 0, \ x_1^2 + (x_2 - 1)^2 \leq 1\}$, see Figure 1. The unique solution of this problem is $x^* := (0, 0)$. There exist different corresponding optimal y-parts. For example, taking $y^* := (0, 1)$, it is easy to see that (x^*, y^*) is not an S-stationary point; in particular, using Theorem 4.2, it follows that this problem does not satisfy GCQ, hence CC-GCQ is also violated. However, (x^*, y^*) is M-stationary, for example, one may take $\lambda = 0, \gamma_1 = 1, \gamma_2 = -1$ to see that the conditions from Lemma 5.2 (b) hold.

On the other hand, x^* together with $y^* := (1, 0)$ is also optimal, and this pair turns out to be S-stationary. Indeed, a simple calculation shows that, for example, the multipliers $\gamma_1 := 1, \gamma_2 := -1$, and $\lambda := 0$ satisfy the S-stationarity conditions from Lemma 5.2 (a). $\qquad \diamondsuit$
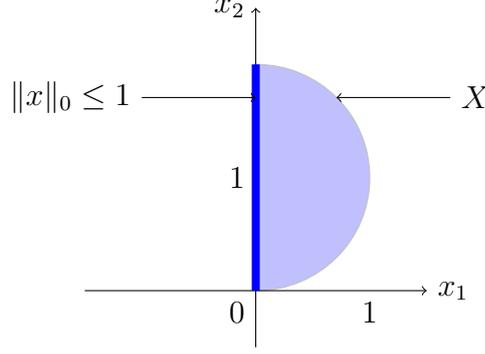
Figure 1: Illustration of Example 5.5

We next investigate the relation between our CC-linearized cone and the MPCC-linearized cone, which for a point x^* feasible for the MPCC (7) is defined by

$$
\begin{aligned}
\mathcal{L}_Z^{MPCC}(x^*) = \big\{ d \ \mid \ & \nabla g_i(x^*)^T d \le 0 \text{ if } g_i(x^*) = 0, \\
& \nabla h_i(x^*)^T d = 0 \ \forall i = 1, \dots, p, \\
& \nabla G_i(x^*)^T d = 0 \text{ if } G_i(x^*) = 0, H_i(x^*) > 0, \\
& \nabla H_i(x^*)^T d = 0 \text{ if } G_i(x^*) > 0, H_i(x^*) = 0, \\
& \nabla G_i(x^*)^T d \ge 0 \text{ if } G_i(x^*) = 0, H_i(x^*) = 0, \\
& \nabla H_i(x^*)^T d \ge 0 \text{ if } G_i(x^*) = 0, H_i(x^*) = 0, \\
& (\nabla G_i(x^*)^T d)(\nabla H_i(x^*)^T d) = 0 \text{ if } G_i(x^*) = 0, H_i(x^*) = 0 \big\},
\end{aligned}
$$

see, e.g., [11, 20].

Lemma 5.6 *Let (x^*, y^*) be feasible for (8). Then $\mathcal{L}_Z^{MPCC}(x^*, y^*) = \mathcal{L}_Z^{CC}(x^*, y^*)$.*

Proof: For (x^*, y^*) feasible for (8), the MPCC-linearized cone is of the form

$$
\begin{aligned}
\mathcal{L}_Z^{MPCC}(x^*, y^*) = \big\{ d = (d_x, d_y) \ \mid \ & \nabla g_i(x^*)^T d_x \le 0 \ \forall i \in I_g(x^*), \\
& \nabla h_i(x^*)^T d_x = 0 \ \forall i = 1, \dots, p, \\
& e^T d_y \ge 0 \text{ if } e^T y^* = n - \kappa, \\
& e_i^T d_y \le 0 \ \forall i \in I_{01}(x^*, y^*), \\
& e_i^T d_x = 0 \ \forall i \in I_{0+}(x^*, y^*) \cup I_{01}(x^*, y^*), \\
& e_i^T d_y = 0 \ \forall i \in I_{\pm 0}(x^*, y^*), \\
& e_i^T d_x \ge 0 \ \forall i \in I_{00}(x^*, y^*), \\
& e_i^T d_y \ge 0 \ \forall i \in I_{00}(x^*, y^*), \\
& (e_i^T d_x)(e_i^T d_y) = 0 \ \forall i \in I_{00}(x^*, y^*) \big\},
\end{aligned}
$$

which is exactly the same as the CC-linearized cone $\mathcal{L}_Z^{CC}(x^*, y^*)$. $\qquad \square$

Consequently, for points (x^*, y^*) feasible for (8), the CC-constraint qualifications CC-ACQ and CC-GCQ coincide with their MPCC counterparts MPCC-ACQ and MPCC-GCQ.

However, even though there are these close connections between cardinality - constrained problems and MPCCs in case $x \geq 0$, the results we have proven in this paper are quite different from what can be shown for general MPCCs. We summarize some of the major differences between general MPCCs and cardinality-constrained optimization problems in the following remark.

Remark 5.7 (a) Standard GCQ holds for MPCCs under the MPCC-LICQ assumption, but already under the MPCC-MFCQ condition it can be violated, see, e.g., [24]. The different behavior we observe for cardinality constrained problems is due to the equality $\mathcal{L}_Z(x^*, y^*)^\circ = \mathcal{L}_Z^{CC}(x^*, y^*)^\circ$ established in the proof of Theorem 3.7. This so-called intersection property, which implies CC-GCQ = GCQ, is not satisfied for general MPCCs.

(b) As a consequence of the previous remark, it follows that S-stationarity is a necessary optimality condition under the fairly strong MPCC-LICQ condition, but neither under the MPCC-MFCQ nor under any weaker MPCC constraint qualification. Recall that this is very much in contrast to the situation for cardinality-constrained problems.

(c) Even if all functions involved in the MPCC are linear, a local minimum is, in general, only an M-stationary point. A counterexample from [24] shows that S-stationarity cannot be obtained without further assumptions.

(d) While W-, C-, M-, and S-stationarity are four different stationarity concepts that arise in different contexts for general MPCCs, it turns out that these four stationarity conditions reduce to only two when applied to the particular MPCC (8) that results from the cardinality-constrained problem in case where this includes nonnegativity constraints.

(e) For MPCCs, it is known that MPCC-LICQ implies a piecewise LICQ condition, whereas this is not true in our setting, cf. the corresponding discussion after Lemma 3.11.

(f) Finally, we would like to stress that the popular MPCC-LICQ condition is likely to be violated for the problem (8). To this end, let (x^*, y^*) be a solution satisfying $y_i^* \in \{0, 1\}$ for all $i = 1, \ldots, n$ and such that the cardinality constraint $e^T y \geq n - \kappa$ is active; note that this situation is very likely to hold at a solution. Then the gradient of the cardinality constraint is obviously linearly dependent from the gradients which one obtains from the activity of the constraints $y_i \geq 0$ and $y_i \leq 1$. Hence MPCC-LICQ is violated in this situation; for the same reason, also MPCC-MFCQ does not hold.

6 Final Remarks

In this paper, we exploited the relation between the cardinality-constrained optimization problem and a suitable nonlinear program to define some problem-tailored constraint qualifications which were then used to prove a KKT-type optimality conditions under fairly mild conditions. Like for standard nonlinear programs, these results depend on the feasible set, but not directly on the objective function. There are some recent contributions to MPCCs where optimality conditions are derived under certain assumptions which involve the particular objective function, and it

might be an interesting future research topic to see whether one can also obtain similar results for cardinality-constrained problems, possibly under weaker assumptions than in the MPCC-setting by taking into account the particular structure of our reformulated cardinality-constrained problem.

References

[1] R. Andreani, J.M. Martínez, and M.L. Schuverdt: *The CPLD condition of Qi and Wei implies the quasinormality qualification.* Journal of Optimization Theory and Applications 125, 2005, pp. 473–485.

[2] M.S. Bazaraa and C.M. Shetty: *Foundations of Optimization.* Lecture Notes in Economics and Mathematical Systems, Springer, 1976.

[3] D.P. Bertsekas: *Nonlinear Programming.* Athena Scientific, Belmont, MA, second edition 1999.

[4] D.P. Bertsekas and A.E. Ozdaglar: *Pseudonormality and a Lagrange multiplier theory for constrained optimization.* Journal of Optimization Theory and Applications 114, 2002, pp. 287–343.

[5] D. Bertsimas and R. Shioda: *Algorithm for cardinality-constrained quadratic optimization.* Computational Optimization and Applications 43, 2009, pp. 1–22.

[6] D. Bienstock: *Computational study of a family of mixed-integer quadratic programming problems.* Mathematical Programming 74, 1996, pp. 121–140.

[7] O.P. Burdakov, C. Kanzow, and A. Schwartz: *Mathematical programs with cardinality constraints: Reformulation by complementarity-type conditions and a regularization method.* Technical Report, Institute of Mathematics, University of Würzburg, Würzburg, Germany, July 2014.

[8] E.J. Candès and M.B. Wakin: *An introduction to compressive sampling.* IEEE Signal Processing Magazine 25, 2008, pp. 21–30.

[9] D. Di Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen, and M. Sciandrone: *A concave optimization-based approach for sparse portfolio selection.* Optimization Methods and Software 27, 2012, pp. 983–1000.

[10] M. Feng, J.E. Mitchell, J.-S. Pang, X. Shen, and A. Wächter: *Complementarity formulation of ℓ_0-norm optimization problems.* Technical Report, Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA, September 2013.

[11] M.L. Flegel and C. Kanzow: *Abadie-type constraint qualification for mathematical programs with equilibrium constraints.* Journal of Optimization Theory and Applications 124, 2005, pp. 595–614.

[12] M.L. Flegel, C. Kanzow and J.V. Outrata: *Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints.* Set-Valued Analysis 15, 2007, pp. 139–162.

[13] Z.-Q. Luo, J.-S. Pang, and D. Ralph: *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, Cambridge, New York, Melbourne, 1996.

[14] O.L. Mangasarian: *Nonlinear Programming.* McGraw-Hill, New York, NY, 1969 (reprinted by SIAM, Philadelphia, PA, 1994).

[15] A. Miller: *Subset Selection in Regression.* Chapman & Hall/CRC, Boca Raton, second edition 2002.

[16] W. Murray and H. Shek: *A local relaxation method for the cardinality-constrained portfolio optimization problem.* Computational Optimization and Applications 53, 2012, pp. 681–709.

[17] J. Nocedal and S.J. Wright: *Numerical Optimization.* Springer Series in Operations Research, Springer, New York, 1999.

[18] J.V. Outrata: *Optimality conditions for a class of mathematical programs with equilibrium constraints.* Mathematics of Operations Research 24, 1999, pp. 627–644.

[19] J.V. Outrata, M. Kočvara, and J. Zowe: *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[20] J.-S. Pang and M. Fukushima: *Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints.* Computational Optimization and Applications 13, 1999, pp. 111–136.

[21] L. Qi and Z. Wei: *On the constant positive linear dependence constraint qualification and its application to SQP methods.* SIAM Journal on Optimization 10, 2000, pp. 963–981.

[22] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, 1998.

[23] R. Ruiz-Torrubiano, S. García-Moratilla, and A. Suárez: *Optimization problems with cardinality constraints.* In Y. Tenne and C.-K. Goh (eds.): *Computational Intelligence in Optimization.* Springer, 2010, pp. 105–130.

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

[25] X. Sun, X. Zheng, and D. Li: *Recent advances in mathematical programming with semi-continuous variables and cardinality constraint.* Journal of the Operations Research Society of China 1, 2013, pp. 55–77.

[26] J.J. Ye: *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints.* SIAM Journal on Optimization 10, 2000, pp. 943–962.

[27] J.J. Ye: *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints.* Journal of Mathematical Analysis and Applications 307, 2005, pp. 350–369.

[28] X. Zheng, X. Sun, D. Li, and J. Sun: *Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear DC approach.* Computational Optimization and Applications, to appear (DOI: 10.1007/s10589-013-9582-3).