

# Restricted Robinson Constraint Qualification and Optimality for Cardinality-Constrained Cone Programming

Lili Pan<sup>1,2</sup>  · Ziyang Luo<sup>3</sup> · Naihua Xiu<sup>1</sup>

Received: 9 January 2017 / Accepted: 4 September 2017 / Published online: 13 September 2017  
© Springer Science+Business Media, LLC 2017

**Abstract** In this paper, optimality conditions are presented and analyzed for the cardinality-constrained cone programming arising from finance, statistical regression, signal processing, etc. By introducing a restricted form of (strict) Robinson constraint qualification, the first-order optimality conditions for the cardinality-constrained cone programming are established based upon the properties of the normal cone. After characterizing further the second-order tangent set to the cardinality-constrained system, the second-order optimality conditions are also presented under some mild conditions. These proposed optimality conditions, to some extent, enrich the optimization theory for noncontinuous and nonconvex programming problems.

**Keywords** Cardinality constraint · Cone programming · Optimality condition · Constraint qualification

**Mathematics Subject Classification** 90C26 · 90C30 · 90C46

---

✉ Lili Pan  
panlili1979@163.com

Ziyang Luo  
zyluo@bjtu.edu.cn

Naihua Xiu  
nhxiu@bjtu.edu.cn

<sup>1</sup> Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, People's Republic of China

<sup>2</sup> Department of Mathematics, Shandong University of Technology, Zibo 255049, People's Republic of China

<sup>3</sup> State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, People's Republic of China

## 1 Introduction

The cardinality constraint, also known as the sparsity constraint, has attracted growing interest in various disciplines due to the “dimension curse” in the big data era. This paper treats a general cardinality-constrained optimization model called the cardinality-constrained cone programming (CCCP), which aims to minimize a general continuously differentiable function subject to the cardinality constraint together with some general nonlinear cone constraint. The CCCP problem provides a unified framework for a rich variety of cardinality-constrained problems arising from many applications including the subset selection problem in regression [1,2], the portfolio optimization problems [3–7], and the linear and nonlinear compressed sensing in signal processing [8–10].

The CCCP problem is generally NP-hard due to the cardinality function and so convex and nicely behaved nonconvex relaxation strategies are typically employed to handle this problem with various numerical algorithms (see, e.g., [11–17] and references therein). Most existing theoretical results are then addressed on relaxation exactness-guaranteed conditions such as the restricted isometry property (RIP) [10], while less results are explored on the optimality analysis for the cardinality-constrained problem directly. Such limited research results are mainly distributed in [18–26]. For example, with the single cardinality constraint, Beck and Eldar [19] presented three types of first-order necessary optimality conditions based upon the notions of stationarity and coordinatewise optimality. Extended analysis for the cardinality and the so-called symmetric set constraints were discussed in [20,25] in a similar manner. For a relatively more general case, where the feasible region is characterized by the cardinality constraint and nonlinear equality and inequality constraints, Lu and Zhang [24] established a first-order necessary optimality condition by employing the Robinson constraint qualification (RCQ) from variational analysis. Such a model was recently reformulated as a mixed-integer program [21] and complementarity constrained mathematical program [22], and optimality analysis was then built upon the corresponding reformulated models. Besides the tool of constraint qualifications (CQs) as employed in [21,22,24], the notions of tangent cones and normal cones were utilized to characterize optimality conditions for cardinality-constrained problems, such as the Bouligand tangent cone and the proximal and Mordukhovich normal cones to the cardinality-constrained set as discussed in [18], the Bouligand and Clarke tangent cones and corresponding normal cones to the intersection of the cardinality-constrained set and the nonnegative orthant in [26], and the Fréchet and Mordukhovich normal cones to the intersection of the cardinality-constrained set and any given convex polyhedron in [23].

Our purpose is to extend the first- and the second-order optimality conditions to the more general and unified CCCP model by proposing some tailored forms of the RCQ. These CQs are well designed to ensure decomposition properties of the tangent cone and the normal cone to the intersection of the cardinality-constrained set and the nonlinear cone constrained set, which will play a vital role in establishing the desired optimality conditions.

The remainder of the paper is organized as follows. The CCCP model and its related fundamental properties will be introduced in Sect. 2. The restricted forms of the RCQ

and the strict RCQ tailored for the CCCP problem will be proposed and discussed in Sect. 3. The first- and the second-order optimality conditions are then proposed in Sects. 4 and 5, respectively. Concluding remarks are drawn in Sect. 6.

## 2 Preliminaries

In this paper, we focus on the following CCCP problem:

$$\min f(x) \quad \text{s.t. } G(x) \in K, \|x\|_0 \leq s, \quad (1)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $G: \mathbb{R}^n \rightarrow \mathbb{R}^m$  are continuously differentiable functions,  $K$  is a nonempty, closed and convex cone in  $\mathbb{R}^m$ ,  $\|x\|_0$  denotes the cardinality of the vector  $x \in \mathbb{R}^n$ , i.e., the number of nonzero elements in  $x$ ,  $s \in ]0, n[$  is any given positive integer.

Several specific models of the form (1) arise from statistical regression, finance, signal processing and others; see, e.g., the subset selection in regression [27], the cardinality-constrained principal component analysis [28] and its relaxation [29], the cardinality-constrained semidefinite programming [30], the cardinality-constrained nonnegative matrix factorization [31], etc.

For convenience, we denote the feasible set of (1) as  $\Omega := \Phi \cap S$  with

$$\Phi := \{x \in \mathbb{R}^n: G(x) \in K\} \text{ and } S := \{x \in \mathbb{R}^n: \|x\|_0 \leq s\}. \quad (2)$$

Some related fundamental notation throughout the paper is listed as follows. For any given  $x \in \mathbb{R}^n$ , we denote  $\text{supp}(x) := \{i = 1, \dots, n: x_i \neq 0\}$ . For an index set  $J \subseteq \{1, 2, \dots, n\}$ ,  $|J|$  is the cardinality of  $J$  and  $\bar{J} = \{1, \dots, n\} \setminus J$ .  $\mathbb{R}_J^n := \text{span}\{e_i: i \in J\}$  denotes the subspace of  $\mathbb{R}^n$  spanned by  $\{e_i: i \in J\}$ , where  $e_i$  is the  $i$ -th unit vector in  $\mathbb{R}^n$ .  $\mathcal{J} := \{J \subseteq \{1, 2, \dots, n\}: |J| = s\}$ . For  $x^* \in S$ , we denote  $\Gamma^* := \text{supp}(x^*)$  and  $\mathcal{J}^* := \{J \in \mathcal{J}: J \supseteq \Gamma^*\}$ . For any given set  $\Omega \subseteq \mathbb{R}^n$ ,  $x' \xrightarrow{\Omega} x$  means  $x' \rightarrow x$  and  $x' \in \Omega$ . For a cone  $K \subseteq \mathbb{R}^n$ ,  $K^\circ := \{d \in \mathbb{R}^n: \langle d, x \rangle \leq 0, \forall x \in K\}$  is its (negative) polar cone. The distance from the point  $x \in \mathbb{R}^n$  to set  $\Omega \subseteq \mathbb{R}^n$  is denoted  $\text{dist}(x, \Omega) = \inf_{y \in \Omega} \|x - y\|$ . Denote  $G^{-1}: \mathbb{R}^m \rightarrow \mathbb{R}^n$  as the inverse mapping of  $G$ ,  $\text{DG}(x): \mathbb{R}^n \rightarrow \mathbb{R}^m$  as the derivative of  $G$  and  $\text{DG}(x)\mathbb{R}^n = \{\text{DG}(x)y: y \in \mathbb{R}^n\}$ . The support function of set  $\Omega$  is denoted as  $\sigma(x, \Omega) = \sup\{\langle x, y \rangle, y \in \Omega\}$ .  $\text{cl}(\Omega)$  and  $\text{int}(\Omega)$  stand for the closure and the interior of the set  $\Omega$ , respectively. Let  $x \in \Omega$ . The tangent (also called contingent or Bouligand) cone and Clarke (also called regular) tangent cone to  $\Omega$  at  $x$  are denoted as  $T_\Omega(x)$  and  $T_\Omega^C(x)$ , respectively (see, e.g., [32, Definition 6.1] and [32, Definition 6.25]). The normal cones in the sense of Fréchet (also called regular), Clarke, and Mordukhovich (also called limiting or basic) are denoted as  $\hat{N}_\Omega(x)$ ,  $N_\Omega^C(x)$  and  $N_\Omega(x)$ , respectively (see, e.g., [32, Definition 6.3] and [33]). For a nonempty and closed set  $\Omega \subseteq \mathbb{R}^n$  and a point  $x \in \Omega$ , the outer second-order tangent set to  $\Omega$  at  $x$  in the direction  $d$  is denoted as  $T_\Omega^2(x, d)$  (see, e.g., [32, Definition 13.11]).

By utilizing the observation  $S = \bigcup_{J \in \mathcal{J}} \mathbb{R}_J^n$ , we have the following properties on the tangent cones, normal cones and outer second-order tangent set to  $S$ .

**Lemma 2.1** [18, Theorems 3.9 and 3.15], [26, Theorems 2.1 and 2.2] *Let  $x^* \in S$ . We have  $N_S^C(x^*) = \mathbb{R}_{\Gamma^*}^n$ ,  $T_S^C(x^*) = \mathbb{R}_{\Gamma^*}^n$ , and*

$$\hat{N}_S(x^*) = \begin{cases} \mathbb{R}_{\Gamma^*}^n, & \text{if } |\Gamma^*| = s, \\ \{0\}, & \text{if } |\Gamma^*| < s, \end{cases} \quad T_S(x^*) = \begin{cases} \mathbb{R}_{\Gamma^*}^n, & \text{if } |\Gamma^*| = s, \\ \bigcup_{J \in \mathcal{J}^*} \mathbb{R}_J^n, & \text{if } |\Gamma^*| < s. \end{cases}$$

**Proposition 2.1** *Let  $x^* \in S$  and  $d \in T_S(x^*)$ . Then*

$$T_S^2(x^*, d) = T_S(x^*) \cap T_S(d).$$

*Proof* The desired result follows from

$$\begin{aligned} T_S^2(x^*, d) &= T_{\bigcup_{J \in \mathcal{J}} \mathbb{R}_J^n}^2(x^*, d) = \bigcup_{J \in \mathcal{J}} T_{\mathbb{R}_J^n}^2(x^*, d) = \bigcup_{J \in \mathcal{J}^*} T_{\mathbb{R}_J^n}^2(x^*, d) \\ &= \bigcup_{\substack{J \in \mathcal{J}^*, \\ d \in \mathbb{R}_J^n}} (T_{\mathbb{R}_J^n}(x^*) + \mathbb{R}d) = \bigcup_{\substack{J \in \mathcal{J}^*, \\ d \in \mathbb{R}_J^n}} (\mathbb{R}_J^n + \mathbb{R}d) = \bigcup_{\substack{J \in \mathcal{J}^*, \\ d \in \mathbb{R}_J^n}} \mathbb{R}_J^n = \bigcup_{\substack{|J|=s, J \supseteq \Gamma^*, \\ J \supseteq \text{supp}(d)}} \mathbb{R}_J^n \\ &= \left( \bigcup_{|J|=s, J \supseteq \Gamma^*} \mathbb{R}_J^n \right) \cap \left( \bigcup_{|J|=s, J \supseteq \text{supp}(d)} \mathbb{R}_J^n \right) = T_S(x^*) \cap T_S(d), \end{aligned}$$

where the second equality is from [34, Proposition 3.37], the third equality follows from the fact that for any  $\mathbb{R}_J^n$  satisfying  $x^* \notin \mathbb{R}_J^n$  (i.e.,  $J \notin \mathcal{J}^*$ ),  $T_{\mathbb{R}_J^n}^2(x^*, d) = \emptyset$  and the forth equality is from [32, Proposition 13.12]. This completes the proof.  $\square$

It is easy to verify that, for any  $d \in T_S(x^*)$ ,  $T_S^C(x^*) \subseteq T_S^2(x^*, d)$ .

### 3 Restricted Robinson Constraint Qualification

In order to get a decomposed form for the tangent cone and the normal cone to  $\Phi \cap S$ , we will introduce restricted variants of the classical RCQ and the strict RCQ tailored for the constraint system of (1). Recall that the RCQ for  $\Phi$  holds at  $x^* \in \Phi$  if  $0 \in \text{int}\{G(x^*) + DG(x^*)\mathbb{R}^n - K\}$ . According to [34, Corollary 2.98], the RCQ is equivalent to

$$DG(x^*)\mathbb{R}^n - T_K(G(x^*)) = \mathbb{R}^m. \tag{3}$$

By invoking [34, Lemma 3.27], this further leads to

$$\hat{N}_\Phi(x^*) = [T_\Phi(x^*)]^\circ = DG(x^*)^\top N_K(G(x^*)). \tag{4}$$

Some other essential properties on the RCQ are reviewed as follows.

**Lemma 3.1** [34, Proposition 2.90, Corollary 2.91] *Suppose that the RCQ holds at some  $x^* \in \Phi$ . Then  $T_\Phi(x^*) = \{d \in \mathbb{R}^n: DG(x^*)d \in T_K(G(x^*))\}$ , and  $\text{dist}(d, T_\Phi(x^*)) = o(\|d\|)$ ,  $\forall x^* + d \in \Phi$ .*

Recall that the strict RCQ for  $\Phi$  holds at  $(x^*, \lambda^*)$  with  $\lambda^* \in N_K(G(x^*))$  if  $0 \in \text{int}\{G(x^*) + DG(x^*)\mathbb{R}^n - K_0\}$  with  $K_0 := \{y \in K: \langle \lambda^*, y - G(x^*) \rangle = 0\}$ .

For  $\Phi \cap S$ , we introduce restricted forms of RCQ and strict RCQ as below.

**Definition 3.1** Let  $0 \neq x^* \in \Phi \cap S$ , and  $\lambda^* \in N_K(G(x^*))$ . We say that the restricted RCQ (RRCQ) with respect to  $\Phi \cap S$  holds at  $x^*$  if

$$0 \in \text{int}\{G(x^*) + DG(x^*)\mathbb{R}_{\Gamma^*}^n - K\}, \tag{5}$$

and the restricted strict RCQ (RSRCQ) to  $\Phi \cap S$  holds at  $(x^*, \lambda^*)$  if

$$0 \in \text{int}\{G(x^*) + DG(x^*)\mathbb{R}_{\Gamma^*}^n - K_0\}, \tag{6}$$

where  $K_0 = \{y \in K: \langle \lambda^*, y - G(x^*) \rangle = 0\}$ .

Properties on the above proposed CQs are discussed as follows.

**Proposition 3.1** *If the RRCQ holds at some  $0 \neq x^* \in \Phi \cap S$ , then*

$$T_{\Phi \cap S}(x^*) = T_{\Phi}(x^*) \cap T_S(x^*). \tag{7}$$

*Proof* Analogous to the equivalence between RCQ and (3), we immediately get that RRCQ (5) is equivalent to  $DG(x^*)\mathbb{R}_{\Gamma^*}^n - T_K(G(x^*)) = \mathbb{R}^m$ , which yields that  $DG(x^*)\mathbb{R}_J^n - T_K(G(x^*)) = \mathbb{R}^m$ , for any  $J \in \mathcal{J}^*$  from  $\mathbb{R}_{\Gamma^*}^n \subseteq \mathbb{R}_J^n$ . By taking the polar to both sides, it follows from [32, Corollary 11.25] that  $[DG(x^*)^\top]^{-1}\mathbb{R}_J^n \cap N_K(G(x^*)) = \{0\}$ , and hence

$$\mathbb{R}_J^n \cap DG(x^*)^\top N_K(G(x^*)) = \{0\}. \tag{8}$$

Combining with (4) and the fact  $N_{\mathbb{R}_J^n}(x^*) = \mathbb{R}_J^n$ , it follows from [32, Theorem 6.42] that

$$T_{\Phi \cap \mathbb{R}_J^n}(x^*) = T_{\Phi}(x^*) \cap \mathbb{R}_J^n, \forall J \in \mathcal{J}^*. \tag{9}$$

The desired result then holds from

$$\begin{aligned} T_{\Phi \cap S}(x^*) &= T_{\bigcup_{J \in \mathcal{J}^*} \Phi \cap \mathbb{R}_J^n}(x^*) = \bigcup_{J \in \mathcal{J}^*} T_{\Phi \cap \mathbb{R}_J^n}(x^*) = \bigcup_{J \in \mathcal{J}^*} T_{\Phi}(x^*) \cap \mathbb{R}_J^n \\ &= T_{\Phi}(x^*) \cap \bigcup_{J \in \mathcal{J}^*} \mathbb{R}_J^n = T_{\Phi}(x^*) \cap T_S(x^*). \end{aligned} \tag{10}$$

□

**Lemma 3.2** *Let  $0 \neq x^* \in \Phi \cap S$ ,  $\lambda^* \in N_K(G(x^*))$ , and the RSRCQ holds at  $(x^*, \lambda^*)$ . Then for any  $\lambda \in N_K(G(x^*))$ ,  $(DG(x^*)^\top \lambda)_{\Gamma^*} = (DG(x^*)^\top \lambda^*)_{\Gamma^*}$  implies  $\lambda = \lambda^*$ .*

*Proof* Denote  $A := DG(x^*)$ ,  $r := |\Gamma^*|$ , and  $A_{\Gamma^*}$  as the submatrix consisting of columns of  $A$  indexed by  $\Gamma^*$ . For any given  $\lambda \in N_K(G(x^*))$  satisfying  $(A^\top \lambda)_{\Gamma^*} = (A^\top \lambda^*)_{\Gamma^*}$ , denote  $\mu := \lambda - \lambda^*$ . Then,  $A_{\Gamma^*}^\top \mu = 0$ . The RSRCQ implies that for any

$y \in \mathbb{R}^m$  and any sufficiently small  $\epsilon > 0$ , there exist  $h \in \mathbb{R}^r$  and  $k_0 \in K_0$  such that  $\epsilon y = G(x^*) + A_{\Gamma^*}h - k_0$ . Thus,

$$\langle \mu, G(x^*) - k_0 \rangle = \langle \mu, \epsilon y - A_{\Gamma^*}h \rangle = \epsilon \langle \mu, y \rangle - \langle A_{\Gamma^*}^\top \mu, h \rangle = \epsilon \langle \mu, y \rangle.$$

On the other hand, as  $k_0 \in K_0$ , we have

$$\langle \mu, G(x^*) - k_0 \rangle = \langle \lambda, G(x^*) - k_0 \rangle - \langle \lambda^*, G(x^*) - k_0 \rangle = \langle \lambda, G(x^*) - k_0 \rangle \geq 0.$$

The arbitrariness of  $y$  immediately implies  $\mu = 0$ . □

**Proposition 3.2** *Let  $0 \neq x^* \in \Phi \cap S$ ,  $\lambda^* \in N_K(G(x^*))$ , and the RSRCQ holds at  $(x^*, \lambda^*)$ . Then*

$$\hat{N}_{\Phi \cap S}(x^*) = \hat{N}_\Phi(x^*) + \hat{N}_S(x^*). \tag{11}$$

*Proof* It is known from (10) and Lemma 3.1 that

$$T_{\Phi \cap S}(x^*) = \bigcup_{J \in \mathcal{J}^*} T_\Phi(x^*) \cap \mathbb{R}_J^n = \bigcup_{J \in \mathcal{J}^*} \{d \in \mathbb{R}_J^n : DG(x^*)d \in T_K(G(x^*))\}. \tag{12}$$

Now we claim that

$$(T_\Phi(x^*) \cap \mathbb{R}_J^n)^\circ = \hat{N}_\Phi(x^*) + \mathbb{R}_J^n. \tag{13}$$

From [32, Corollary 11.25] and (4), we have

$$(T_\Phi(x^*) \cap \mathbb{R}_J^n)^\circ = cl[\hat{N}_\Phi(x^*) + \mathbb{R}_J^n] = cl[DG(x^*)^\top N_K(G(x^*)) + \mathbb{R}_J^n]. \tag{14}$$

Since  $DG(x^*)^\top N_K(G(x^*))$  is a nonempty, closed and convex cone under RCQ, and for any  $z_1 \in 0^+(DG(x^*)^\top N_K(G(x^*)))$  and  $z_2 \in 0^+(\mathbb{R}_J^n) = \mathbb{R}_J^n$  satisfying  $z_1 + z_2 = 0$ , we have  $z_1 \in DG(x^*)^\top N_K(G(x^*))$  and  $z_1 = -z_2 \in \mathbb{R}_J^n$ , which yields that  $z_1 \in DG(x^*)^\top N_K(G(x^*)) \cap \mathbb{R}_J^n$ . From (8), we obtain  $z_1 = z_2 = 0$ . Therefore,  $z_1$  and  $z_2$  belong to the linearity spaces of  $cl(DG(x^*)^\top N_K(G(x^*)))$  and  $cl(\mathbb{R}_J^n)$ . From [35, Corollary 9.1.1],

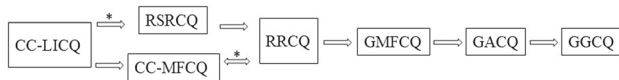
$$cl[DG(x^*)^\top N_K(G(x^*)) + \mathbb{R}_J^n] = DG(x^*)^\top N_K(G(x^*)) + \mathbb{R}_J^n.$$

Together with (14), we have proved the claim in (13).

If  $\|x^*\|_0 = s$ , then  $T_{\Phi \cap S}(x^*) = T_\Phi(x^*) \cap \mathbb{R}_{\Gamma^*}^n$ . By (13), we have

$$\hat{N}_{\Phi \cap S}(x^*) = (T_{\Phi \cap S}(x^*))^\circ = \hat{N}_\Phi(x^*) + \mathbb{R}_{\Gamma^*}^n = \hat{N}_\Phi(x^*) + \hat{N}_S(x^*).$$

If  $0 < \|x^*\|_0 < s$ , from (10) and (13), we have



**Fig. 1** The relationship among the above CQs, where \* means the case of  $K = \mathbb{R}_+^l \times \{0\}^{m-l}$

$$\begin{aligned} \hat{N}_{\Phi \cap S}(x^*) &= (T_{\Phi \cap S}(x^*))^\circ = \left( \bigcup_{J \in \mathcal{J}^*} T_\Phi(x^*) \cap \mathbb{R}_J^n \right)^\circ \\ &= \bigcap_{J \in \mathcal{J}^*} (T_\Phi(x^*) \cap \mathbb{R}_J^n)^\circ = \bigcap_{J \in \mathcal{J}^*} (\hat{N}_\Phi(x^*) + \mathbb{R}_J^n). \end{aligned}$$

We next show that  $\bigcap_{J \in \mathcal{J}^*} (\hat{N}_\Phi(x^*) + \mathbb{R}_J^n) = \hat{N}_\Phi(x^*)$ . It suffices to show  $\bigcap_{J \in \mathcal{J}^*} (\hat{N}_\Phi(x^*) + \mathbb{R}_J^n) \subseteq \hat{N}_\Phi(x^*)$ . Let  $v \in \bigcap_{J \in \mathcal{J}^*} (\hat{N}_\Phi(x^*) + \mathbb{R}_J^n)$ . Pick any two index sets  $J', J'' \in \mathcal{J}^*$ , we have  $v = DG(x^*)^\top \lambda' + \xi' = DG(x^*)^\top \lambda'' + \xi''$ , with  $\lambda', \lambda'' \in N_K(G(x^*))$  and  $\xi' \in \mathbb{R}_{J'}^n, \xi'' \in \mathbb{R}_{J''}^n$ . Since  $\bigcap_{J \in \mathcal{J}^*} J = \Gamma^*$ , then

$$v_{\Gamma^*} = (DG(x^*)^\top \lambda')_{\Gamma^*} = (DG(x^*)^\top \lambda'')_{\Gamma^*}.$$

From the RSRCQ and Lemma 3.2, we have  $\lambda' = \lambda''$ .  $\bigcup_{J \in \mathcal{J}^*} J = \{1, \dots, m\}$ , then  $v \in DG(x^*)^\top N_K(G(x^*))$ . By invoking Lemma 2.1, we obtain (11).  $\square$

It is worth pointing out that when  $K = \mathbb{R}_+^l \times \{0\}^{m-l}$ , the RRCQ is exactly the CC-MFCQ as proposed in [22] and the RSRCQ can be implied by the CC-LICQ in [22]. Some weak CQs, which are generalized versions of Mangasarian–Fromovitz, Abadie and Guignard constraint qualifications (abbreviated as GMFCQ, GACQ and GGCQ, respectively) are proposed in [36] for a general minimization problem whose feasible region is  $\{x \in \mathbb{R}^n : F(x) \in \Lambda\}$  with some continuously differentiable function  $F$  and some nonempty and closed set  $\Lambda$ . Since the constrained system of (1) can be rewritten as  $(G(x), x) \in K \times \bigcup_{J \in \mathcal{J}} \mathbb{R}_J^n =: K(s)$ , where  $K(s)$  is evidently nonempty and closed, the corresponding GMFCQ can be easily shown to be weaker than RRCQ. The relationship among all the above CQs is illustrated as follows (Fig. 1).

All the involved functions in the constrained systems as discussed in the aforementioned CQs are continuously differentiable. Functions without continuous differentiability are also considered in [37,38] to define the weakest CQ for equality and inequality constrained systems. In [38], the involved functions are called  $\mathcal{C}$ -differentiable functions and hence the corresponding CQ is called the CCQ. If the problem with differentiable equality and inequality constraints, we have the following scheme:

$$\text{RCQ} \Leftrightarrow \text{MFCQ} \Leftrightarrow \text{GMFCQ} \Rightarrow \text{GACQ} \Rightarrow \text{GGCQ} \Leftrightarrow \text{CCQ}.$$

It will be a challenging work to extend this CCQ to conic constrained system with cardinality constraint or to consider CCCP problem where the involved functions are not continuously differentiable.

### 4 First-Order Optimality Conditions

Denote the Lagrangian function of problem (1) as

$$L(x, \lambda) := f(x) + \langle \lambda, G(x) \rangle, \quad (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m. \tag{15}$$

Lagrangian multiplier sets in the sense of Fréchet and Clarke are defined as

$$\Lambda^B(x^*) := \{ \lambda \in \mathbb{R}^m : -\nabla_x L(x^*, \lambda) \in \hat{N}_S(x^*), \lambda \in N_K(G(x^*)) \}, \tag{16}$$

$$\Lambda^C(x^*) := \{ \lambda \in \mathbb{R}^m : -\nabla_x L(x^*, \lambda) \in N_S^C(x^*), \lambda \in N_K(G(x^*)) \}, \tag{17}$$

respectively. Obviously,  $\Lambda^B(x^*) \subseteq \Lambda^C(x^*)$ .

**Theorem 4.1** (First-order necessary optimality condition) *Let  $x^* \neq 0$  be a locally optimal solution of (1), and  $\lambda^* \in N_K(G(x^*))$ .*

(i) *If the RRCQ holds at  $x^*$ , then  $d = 0$  is an optimal solution of*

$$\min_{d \in \mathbb{R}^n} \nabla f(x^*)^\top d \text{ s.t. } d \in T_S(x^*) \text{ and } DG(x^*)d \in T_K(G(x^*)). \tag{18}$$

- (ii) *If the RRCQ holds at  $x^*$ , then  $\Lambda^C(x^*)$  is a nonempty, convex and compact set.*
- (iii) *If the RSRCQ holds at  $(x^*, \lambda^*)$ , then  $\Lambda^B(x^*)$  is a singleton.*

*Proof* (i) By Proposition 3.1 and Lemma 3.1, the RRCQ implies

$$T_\Omega(x^*) = T_\Phi(x^*) \cap T_S(x^*) = \{ d \in T_S(x^*) : DG(x^*)d \in T_K(G(x^*)) \}.$$

Utilizing the property of tangent cone, we can easily get (i).

(ii) Consider the following convex problem

$$\min_{d \in \mathbb{R}^n} \nabla f(x^*)^\top d \text{ s.t. } d \in T_S^C(x^*) \text{ and } DG(x^*)d \in T_K(G(x^*)). \tag{19}$$

Obviously,  $d = 0$  is an optimal solution of (19) since  $T_S^C(x^*) \subseteq T_S(x^*)$ , and its RCQ holds by the hypothesis. Note that the dual of (19) takes the form

$$\max_{\lambda} \{ 0 : -\nabla_x L(x^*, \lambda) \in N_S^C(x^*), \lambda \in N_K(G(x^*)) \}.$$

The strong duality theory for convex programming implies (ii) immediately.

(iii) Under Proposition 3.2 and (4), we have

$$\hat{N}_\Omega(x^*) = \hat{N}_\Phi(x^*) + \hat{N}_S(x^*) = (DG(x^*))^\top N_K(G(x^*)) + \hat{N}_S(x^*).$$

Additionally, it is known from (i) that  $-\nabla f(x^*) \in \hat{N}_\Omega(x^*)$ . Thus, there always exists some  $\lambda \in N_K(G(x^*))$  such that

$$-\nabla_x L(x^*, \lambda) = -\nabla f(x^*) - (DG(x^*))^\top \lambda \in \hat{N}_S(x^*),$$



which indicates that  $\Lambda^B(x^*) \neq \emptyset$ . For any  $\lambda' \in \Lambda^B(x^*)$ , it follows from Lemma 2.1 and (16) that  $(DG(x^*)^\top \lambda)_{\Gamma^*} = (DG(x^*)^\top \lambda')_{\Gamma^*} = 0$ . Immediately we have  $\lambda' = \lambda$  by invoking Lemma 3.2. This completes the proof.  $\square$

The inverse implications of the conditions (ii) and (iii) in Theorem 4.1 do not hold as the following two examples illustrate.

*Example 4.1* Consider the problem  $\min_{x \in \mathbb{R}^2} \{x_2 \text{ s.t. } x_2 = 0, \|x\|_0 \leq 1\}$ . Obviously,  $x^* = (1, 0)^\top$  is an optimal solution, and the RRCQ does not hold at  $x^*$  since  $(0, 1)(\mathbb{R}, 0)^\top - \{0\} \neq \mathbb{R}$ . To prove that the inverse implication of (ii) in Theorem 4.1 does not hold,  $\Lambda^C(x^*)$  should be a nonempty, convex and compact set; however,  $\Lambda^C(x^*) = \mathbb{R}$  is not compact.

*Example 4.2* Consider the problem

$$\min_{x \in \mathbb{R}^3} \left\{ f(x) = x_1^2 \text{ s.t. } G(x) = (g_{i,j}(x)) \in \mathbb{S}_+^3, \|x\|_0 \leq 2 \right\}, \tag{20}$$

where  $g_{1,1}(x) = x_1$ ,  $g_{i,4-i}(x) = x_2$ ,  $i = 1, 2, 3$ ,  $g_{3,3}(x) = 1 - x_3$  and others are zero.  $\mathbb{S}_+^3$  is the cone consisting of all 3-by-3 symmetric positive semidefinite matrices. Evidently, the optimal solution set of (20) is  $\{(0, 0, x_3) : x_3 \leq 1\}$ . Let  $x^* = (0, 0, 1)^\top$ . Then  $\Gamma^* = \{3\}$  and  $\Lambda^B(x^*) = \{O_{3 \times 3}\}$ . For any given  $Z \in N_{\mathbb{S}_+^3}(G(x^*))$ , denote  $K_0 := \{Y \in \mathbb{S}_+^3 : \langle Z, Y - G(x^*) \rangle = 0\} = \mathbb{S}_+^3 \cap Z^\perp$ . Thus,

$$\mathcal{M} := G(x^*) + DG(x^*)\mathbb{R}_{\Gamma^*}^3 - K_0 = \left\{ te_3e_3^\top - Y : t \in \mathbb{R}, Y \in \mathbb{S}_+^3 \cap Z^\perp \right\}.$$

Note that for any  $U \in \mathcal{M}$ , the submatrix  $U_{[1:2,1:2]} \in -\mathbb{S}_+^2$ . This immediately implies that  $0 \notin \text{int}\mathcal{M}$ , which indicates that the RSRCQ does not hold at  $(x^*, \lambda^*)$  with  $\lambda^* \in N_K(G(x^*))$  for any  $Z \in N_K(G(x^*))$ .

Next, we proceed with establishing the first-order sufficient conditions. We define a sufficient optimality condition for the local minimizer.

**Definition 4.1** We say that the  $\gamma$ -growth ( $\gamma > 0$ ) condition holds for a function  $f$  at point  $x^*$ , if there exist a constant  $c > 0$  and a neighborhood  $\mathcal{N}(x^*)$  of  $x^*$  such that for all  $x \in \mathcal{N}(x^*)$ , the following inequality holds

$$f(x) \geq f(x^*) + c\|x - x^*\|^\gamma, \quad \forall x \in \Omega \cap \mathcal{N}(x^*).$$

For a feasible point  $x^* \in \Omega$  and  $\eta \geq 0$ , define the set

$$\mathcal{F}_\eta(x^*) := \{d \in T_S(x^*), \text{dist}[DG(x^*)d, T_K(G(x^*))] \leq \eta\|d\|\}. \tag{21}$$

$\mathcal{F}_\eta(x^*)$  is a closed cone. If  $\eta = 0$ ,  $\mathcal{F}_0(x^*) := \{d \in T_S(x^*), DG(x^*)d \in T_K(G(x^*))\}$  is the feasible set of problem (18).

**Theorem 4.2** (First-order sufficient optimality condition) *Let  $x^*$  be a feasible point of the problem (1). The following assertions hold.*

(i) If there exist constants  $\alpha > 0$  and  $\eta > 0$  such that

$$\nabla f(x^*)^\top d \geq \alpha \|d\|, \quad \forall d \in \mathcal{F}_\eta(x^*), \tag{22}$$

then the first-order growth condition holds at  $x^*$ .

(ii) Suppose that RRCQ is satisfied at  $x^*$ . Set  $\eta = 0$ . If there exists an  $\alpha > 0$  such that (22) holds, then the first-order growth condition holds at  $x^*$ .

*Proof* (i) Consider a feasible point  $x^* + d \in \Phi \cap S$  for a sufficiently small  $d \in \mathbb{R}^n$ . We first show  $d \in T_S(x^*)$ . Assume  $d_i \neq 0$ . Then  $(x^* + d)_i \neq 0$  holds if  $x_i^* = 0$  obviously and also holds if  $x_i^* \neq 0$  by letting  $\|d\| < \min\{|x_i^*|, i \in \Gamma^*\}$ . This means that  $\text{supp}(d) \subseteq \text{supp}(x^* + d)$ , or  $\|d\|_0 \leq \|x^* + d\|_0 \leq s$ . Furthermore, we assert that  $\Gamma^* \subseteq \text{supp}(d)$ . In fact, if there is an  $i_0 \in \Gamma^*$  with  $d_{i_0} = 0$ , then we can let  $s + 1 - |\Gamma^*|$  elements of  $d_{\Gamma^*}$  be nonzeros. In such case,  $\|d\|_0 = s$  but  $\|x^* + d\|_0 \geq s + 1$ , which is a contradiction with  $x^* + d \in S$ . Hence,  $d \in T_S(x^*)$ .

It follows from the Taylor expansion  $G(x^* + d) = G(x^*) + DG(x^*)d + o(\|d\|)$  and  $G(x^* + d) \in K \subseteq G(x^*) + T_K(G(x^*))$  that  $DG(x^*)d + o(\|d\|) \in T_K(G(x^*))$ . Then it yields that

$$\text{dist}(DG(x^*)d, T_K(G(x^*))) \leq \|DG(x^*)d - DG(x^*)d - o(\|d\|)\| = o(\|d\|). \tag{23}$$

So  $d \in \mathcal{F}_\eta(x^*)$  for sufficiently small  $d$ . By (22), we obtain

$$f(x^* + d) = f(x^*) + \nabla f(x^*)^\top d + o(\|d\|) \geq f(x^*) + \alpha \|d\| + o(\|d\|),$$

which proves (i).

(ii) Using the similar proof as (i), for a feasible point  $x^* + d \in \Omega$  with a sufficiently small  $d \in \mathbb{R}^n$ , we have  $d \in T_S(x^*)$ . Suppose  $d \in \mathbb{R}_J^n$  for some  $J \in \mathcal{J}^*$ . From  $G(x^* + d) \in K$ , we also have (23).

Under (5), we have  $DG(x^*)\mathbb{R}_J^n - T_K(G(x^*)) = \mathbb{R}^m$ . Lemma 3.1 yields that  $\text{dist}(d, (DG(x^*))^{-1}T_K(G(x^*))) \leq c \cdot \text{dist}(DG(x^*)d, T_K(G(x^*))) = o(\|d\|)$ . Then there is a  $d' \in (DG(x^*))^{-1}T_K(G(x^*))$  such that  $\|d - d'\| = o(\|d\|)$ . From  $\|d - d'\| = o(\|d\|)$ , we have  $\text{supp}(d') \subseteq \text{supp}(d)$ . Otherwise, if  $d'_i \neq 0$  but  $d_i = 0$ , then  $\|d - d'\| \geq |d'_i|$ . This is contradicted with  $\|d - d'\| = o(\|d\|)$ . Then  $d' \in \mathbb{R}_J^n$  from  $d \in \mathbb{R}_J^n$ . Therefore, we obtain that there exists a  $d' \in \mathbb{R}_J^n \subseteq T_S(x^*)$  satisfying  $DG(x^*)d' \in T_K(G(x^*))$  and  $\|d - d'\| = o(\|d\|)$ . By (22), we have  $\nabla f(x^*)^\top d' \geq \alpha \|d'\|$ .

Using the first-order Taylor expansion of  $f$ ,

$$\begin{aligned} f(x^* + d) &= f(x^*) + \nabla f(x^*)^\top d + o(\|d\|) \\ &= f(x^*) + \nabla f(x^*)^\top d' + \nabla f(x^*)^\top (d - d') + o(\|d\|) \\ &\geq f(x^*) + \alpha \|d'\| + \nabla f(x^*)^\top (d - d') + o(\|d\|) \\ &\geq f(x^*) + \alpha (\|d\| - \|d - d'\|) + \nabla f(x^*)^\top (d - d') + o(\|d\|) \\ &= f(x^*) + \alpha \|d\| + o(\|d\|). \end{aligned}$$

Then the first-order growth condition holds. □

### 5 Second-Order Optimality Conditions

In this section, we assume that functions  $f(\cdot)$  and  $G(\cdot)$  in problem (1) are twice continuously differentiable on  $\Omega$ . The linearized problem of (1) is defined by (18) and the critical cone is

$$C(x^*) := \{d \in T_S(x^*) : DG(x^*)d \in T_K(G(x^*)), \nabla f(x^*)^\top d = 0\}. \tag{24}$$

For some  $d, w \in \mathbb{R}^n$ , denote a path  $x(\cdot) : \mathbb{R}_+ \mapsto S$  with the form:

$$x(t) = x^* + td + \frac{1}{2}t^2w + r(t)$$

such that  $r(t) = o(t^2)$ . Then by using the second-order Taylor expansion of  $G$  at  $x^*$ , we obtain

$$\begin{aligned} G(x(t)) &= G\left(x^* + td + \frac{1}{2}t^2w + o(t^2)\right) \\ &= G(x^*) + tDG(x^*)d + \frac{1}{2}t^2[DG(x^*)w + D^2G(x^*)(d, d)] + o(t^2), \end{aligned} \tag{25}$$

where  $D^2G(x^*)(d, d) = [D^2G(x^*)d]d$  is the quadratic form corresponding to the second-order derivative  $D^2G(x)$  of  $G$  at  $x^*$ . It follows from the definition of the outer second-order tangent set that for some  $t_n \downarrow 0$ ,  $\text{dist}(G(x(t_n)), K) = o(t_n^2)$  if and only if

$$DG(x^*)w + D^2G(x^*)(d, d) \in T_K^2(G(x^*), DG(x^*)d). \tag{26}$$

**Lemma 5.1** *Suppose that  $x^*$  is a locally optimal solution of problem (1) and RRCQ holds at  $x^*$ . Then for any  $d \in C(x^*)$  and  $w \in T_S^2(x^*, d)$  satisfying (26), it follows that*

$$\nabla f(x^*)^\top w + \langle d, \nabla^2 f(x^*)d \rangle \geq 0. \tag{27}$$

*Proof* Consider  $d \in C(x^*)$  and  $w \in T_S^2(x^*, d) = T_S(x^*) \cap T_S(d)$  satisfying (26). It yields that there exists a sequence  $t_n \downarrow 0$  such that for some  $J \in \mathcal{J}^*$ ,  $\bar{x}(t_n) := x^* + t_nd + \frac{1}{2}t_n^2w \in \mathbb{R}_J^n$  and  $\text{dist}(G(\bar{x}(t_n)), K) = o(t_n^2)$ . Therefore, we obtain  $\text{dist}(\bar{x}(t_n), G^{-1}(K)) \leq \text{dist}(G(\bar{x}(t_n)), K) = o(t_n^2)$  from RRCQ and Lemma 3.1. By restricting  $o(t_n^2)$  on  $\mathbb{R}_J^n$ , we have

$x(t_n) = \bar{x}(t_n) + o(t_n^2) = x^* + t_nd + \frac{1}{2}t_n^2w + o(t_n^2) \in G^{-1}(K) \cap \mathbb{R}_J^n$ , i.e., the point  $x(t_n)$  are feasible.

By the second-order Taylor expansion of  $f$  at  $x^*$ , we have

$$\begin{aligned} f(x(t_n)) &= f\left(x^* + t_nd + \frac{1}{2}t_n^2w + o(t_n^2)\right) \\ &= f(x^*) + \frac{1}{2}t_n^2[\nabla f(x^*)^\top w + \langle d, \nabla^2 f(x^*)d \rangle] + o(t_n^2). \end{aligned}$$

Combining with  $f(x(t_n)) \geq f(x^*)$ , we get (27). □

Immediately, for any  $d \in C(x^*)$ , the optimal value of the following problem is nonnegative

$$\begin{aligned} \min_w \quad & \nabla f(x^*)^\top w + \langle d, \nabla^2 f(x^*)d \rangle \\ \text{s.t.} \quad & \text{DG}(x^*)w + D^2G(x^*)(d, d) \in T_K^2(G(x^*), \text{DG}(x^*)d), \\ & w \in T_S^2(x^*, d). \end{aligned} \tag{28}$$

**Theorem 5.1** (Second-order necessary optimality condition) *Suppose  $x^* \neq 0$  is a locally optimal solution of (1) and the RRCQ holds at  $x^*$ . Then for any  $d \in C(x^*)$  and any convex set  $T(d) \subseteq T_K^2(G(x^*), \text{DG}(x^*)d)$ , the following inequality holds:*

$$\sup_{\lambda \in \Lambda^C(x^*)} \{ \langle d, \nabla_{xx}^2 L((x^*, \lambda)d) \rangle - \sigma(\lambda, T(d)) \} \geq 0. \tag{29}$$

*Proof* It follows from  $T_S^2(x^*, d) = T_S(x^*) \cap T_S(d)$  for any  $d \in T_S(x^*)$  that  $\mathbb{R}_{\Gamma^*}^n \subseteq T_S^2(x^*, d)$ . Let  $T(d) := \text{cl}\{T(d) + T_K(G(x^*))\}$ , which is the topological closures of the sum of two convex sets. Hence  $T(d)$  is convex.

Since  $T(d) \subseteq T_K^2(G(x^*), \text{DG}(x^*)d)$  and  $\text{DG}(x^*)d \in T_K(G(x^*))$ , by [34, Proposition 3.34] and the closedness of  $T_K^2(G(x^*), \text{DG}(x^*)d)$ , we have

$$T(d) \subset T_K^2(G(x^*), \text{DG}(x^*)d).$$

Clearly, if we replace  $T_S^2(x^*, d)$  and  $T_K^2(G(x^*), \text{DG}(x^*)d)$  in (28) by its subset  $\mathbb{R}_{\Gamma^*}^n$  and  $T(d)$ , respectively, the optimal value of the obtained optimization problem will be greater than or equal to the optimal value of (28), and hence the optimal value of the problem

$$\min_w \left\{ \nabla f(x^*)^\top w + \langle d, \nabla^2 f(x^*)d \rangle : \text{DG}(x^*)w + D^2G(x^*)(d, d) \in T(d), w \in \mathbb{R}_{\Gamma^*}^n \right\} \tag{30}$$

is nonnegative as well. Equation (30) is linear, and its dual is

$$\max_{\lambda \in \Lambda^C(x^*)} \{ \langle d, \nabla_{xx}^2 L((x^*, \lambda)d) \rangle - \sigma(\lambda, T(d)) \}. \tag{31}$$

Indeed, the Lagrangian function of (30) is

$$\begin{aligned} \mathcal{L}(w, \mu, \lambda) &= \nabla f(x^*)^\top w + \langle d, \nabla^2 f(x^*)d \rangle + \langle \lambda, \text{DG}(x^*)w \rangle + \langle \lambda, D^2G(x^*)(d, d) \rangle + \langle \mu, w \rangle \\ &= \nabla f(x^*)^\top w + \langle d, \nabla^2 f(x^*)d \rangle + \langle \lambda, \text{DG}(x^*)w \rangle + \langle d, [D^2G(x^*)d]^\top \lambda \rangle + \langle \mu, w \rangle \\ &= \langle \nabla_x L(x^*, \lambda) + \mu, w \rangle + \langle d, \nabla_{xx}^2 L((x^*, \lambda)d) \rangle. \end{aligned}$$

Thus,  $\inf_w \mathcal{L}(w, \mu, \lambda) = \langle d, \nabla_{xx}^2 L((x^*, \lambda)d) \rangle, \nabla_x L(x^*, \lambda) + \mu = 0$ .

It follows that  $\sigma((\mu, \lambda), \mathbb{R}_{\Gamma^*}^n \times T(d)) = \sigma(\mu, \mathbb{R}_{\Gamma^*}^n) + \sigma(\lambda, T(d)) = \infty$ , for any  $\mu \notin \mathbb{R}_{\Gamma^*}^n = N_S^C(x^*)$  or  $\lambda \notin (T_K(G(x^*)))^\circ = N_K(G(x^*))$ . Therefore, the effective domain of the parametric dual of (30) is contained in  $\Lambda^C(x^*)$ . The duality then follows.

For any  $z = (z_1, z_2) \in \mathbb{R}_{\Gamma^*}^n \times T(d) = cl\{\mathbb{R}_{\Gamma^*}^n \times T(d) + \mathbb{R}_{\Gamma^*}^n \times T_K(G(x^*))\}$ , we have  $z + \mathbb{R}_{\Gamma^*}^n \times T_K(G(x^*)) \subset \mathbb{R}_{\Gamma^*}^n \times T(d)$ . Then  $z_2 + T_K(G(x^*)) \subseteq T(d)$ . Moreover, the RRCQ implies that  $DG(x^*)\mathbb{R}_{\Gamma^*}^n - T_K(G(x^*)) = \mathbb{R}^m$ . Thus,  $z_2 + DG(x^*)\mathbb{R}_{\Gamma^*}^n - T(d) = \mathbb{R}^m$ , which means  $DG(x^*)\mathbb{R}_{\Gamma^*}^n - T(d) = \mathbb{R}^m$ . Hence there is  $w \in \mathbb{R}_{\Gamma^*}^n$  such that  $DG(x^*)w + D^2G(x^*)(d, d) \in T(d)$ . Therefore, Eq. (30) has a feasible solution. Moreover, from RRCQ, we have

$$0 \in \text{int}\{DG(x^*)\mathbb{R}_{\Gamma^*}^n - T_K(G(x^*))\} \subseteq \text{int}\{DG(x^*)\mathbb{R}_{\Gamma^*}^n - T(d)\},$$

where the last inclusion is due to  $T_K(G(x^*)) \subseteq T(d)$ . From the feasibility of  $w^*$ , it holds that  $DG(x^*)w^* + D^2G(x^*)(d, d) \in T(d)$ . Then

$$\begin{aligned} & \text{int}\{DG(x^*)\mathbb{R}_{\Gamma^*}^n - T(d)\} \\ &= \text{int}\{DG(x^*)\mathbb{R}_{\Gamma^*}^n + DG(x^*)w^* + D^2G(x^*)(d, d) - T(d)\} \\ &= \text{int}\{DG(x^*)\mathbb{R}_{\Gamma^*}^n + D^2G(x^*)(d, d) - T(d)\}, \end{aligned}$$

where the last equality follows from the feasibility of  $w^*$  and  $\mathbb{R}_{\Gamma^*}^n + w^* = \mathbb{R}_{\Gamma^*}^n$ . Therefore, the RCQ for problem (30) holds. Consequently, from [34, Theorem 2.165], there is no duality gap between (30) and its dual (31). We obtain that the optimal value of (31) is nonnegative. Since  $\mathcal{T}(d) \subset T(d)$ , we have that  $\sigma(\lambda, \mathcal{T}(d)) \leq \sigma(\lambda, T(d))$ , and hence (29) follows. □

Let  $\mathcal{A}_{K,M}(d) := \mathcal{A}_{K,M}(G(x^*), DG(x^*)d)$  with  $M := DG(x^*)$  be the upper second-order approximation set for  $K$  at the point  $G(x^*)$  in the direction  $DG(x^*)d \in T_K(G(x^*))$  (see [34, Definition 3.82]). We have the following second-order sufficient condition for the CCCP.

**Theorem 5.2** (Second-order sufficient optimality condition) *Suppose that  $x^* \neq 0$  is a feasible solution with  $\Lambda^B(x^*) \neq \emptyset$ , and the RRCQ holds at  $x^*$ . Then the second-order growth condition holds at  $x^*$  if*

$$\sup_{\lambda \in \Lambda^B(x^*)} \{ \langle d, \nabla_{xx}^2 L((x^*, \lambda)d) \rangle - \sigma(\lambda, \mathcal{A}_{K,M}(d)) \} > 0, \quad \forall d \in C(x^*) \setminus \{0\}. \quad (32)$$

*Proof* By invoking Proposition 3.1, the desired result can be obtained by mimicking the proof of [34, Theorem 3.83]. □

It is worth mentioning that for the above two second-order optimality conditions as stated in Theorems 5.1 and 5.2, one may observe that besides the change from the inequality in (29) to the strict inequality in (32), the Lagrangian multiplier set for taking the supremum over becomes smaller since  $\Lambda^B(x^*) \subseteq \Lambda^C(x^*)$ . This makes that (32) implies (29).

## 6 Conclusions

In this paper, we studied the first- and second-order optimality conditions for the CCCP problem which provides a unified framework for most existing cardinality-constrained optimization problems. By introducing the RRCQ and RSRCQ, we decomposed the tangent cone and the normal cone to the involved feasible region and consequently presented the first-order optimality conditions. The second-order optimality conditions were also proposed by exploiting the outer second-order tangent set of the cardinality-constrained set. These proposed results heavily relied on the tailored constraint qualifications and the special structure of the cardinality-constrained set. It might be an interesting future research topic to see whether one can obtain the optimality conditions when the involved objective function and the closed convex cone enjoy special structures in the general CCCP model.

**Acknowledgements** The work was supported in part by the National Natural Science Foundation of China (Grants 11431002, 11771038, 11771255) and Shandong Province Natural Science Foundation (Grant ZR2016AM07). The authors thank the editor and referees for their useful suggestions and comments that improved the quality of the paper.

## References

1. Liu, J., Chen, J., Ye, J.: Large-scale sparse logistic regression. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 547–556. ACM (2009)
2. Koh, K., Kim, S.J., Boyd, S.P.: An interior-point method for large-scale  $l_1$ -regularized logistic regression. *J. Mach. Learn. Res.* **8**(8), 1519–1555 (2007)
3. Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. *Math. Program.* **74**(2), 121–140 (1996)
4. Gao, J., Li, D.: Optimal cardinality constrained portfolio selection. *Oper. Res.* **61**(3), 745–761 (2013)
5. Gao, J., Li, D.: A polynomial case of the cardinality-constrained quadratic optimization problem. *J. Glob. Optim.* **56**(4), 1441–1455 (2013)
6. Li, D., Sun, X., Wang, J.: Optimal lot solution to cardinality constrained mean–variance formulation for portfolio selection. *Math. Finance* **16**(1), 83–101 (2006)
7. Xu, F., Lu, Z., Xu, Z.: An efficient optimization approach for a cardinality-constrained index tracking problem. *Optim. Method Softw.* **31**(2), 258–271 (2015)
8. Blumensath, T.: Compressed sensing with nonlinear observations and related nonlinear optimization problems. *IEEE Trans. Inf. Theory* **59**(6), 3466–3474 (2013)
9. Candès, E.J., Tao, T.: Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**(12), 4203–4215 (2005)
10. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006)
11. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Proc. Lett.* **14**(10), 707–710 (2007)
12. Chen, S., Donoho, D.: Basis pursuit. In: 1994 Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers, 1994, vol. 1, pp. 41–44. IEEE (1994)
13. Chen, X., Ge, D., Wang, Z., Ye, Y.: Complexity of unconstrained  $L_2 - L_p$  minimization. *Math. Program.* **143**(1–2), 371–383 (2014)
14. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **96**(456), 1348–1360 (2001)
15. Kong, L., Sun, J., Tao, J., Xiu, N.: Sparse recovery on Euclidean Jordan algebras. *Linear Algebra Appl.* **465**, 65–87 (2015)
16. Zhang, C.H.: Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **38**(2), 894–942 (2010)

17. Zhao, Y.B.: RSP-based analysis for sparsest and least  $l_1$ -norm solutions to underdetermined linear systems. *IEEE Trans. Signal Proc.* **61**(22), 5777–5788 (2013)
18. Bauschke, H.H., Luke, D.R., Phan, H.M., Wang, X.: Restricted normal cones and sparsity optimization with affine constraints. *Found. Comput. Math.* **14**(1), 63–83 (2014)
19. Beck, A., Eldar, Y.C.: Sparsity constrained nonlinear optimization: optimality conditions and algorithms. *SIAM J. Optim.* **23**(3), 1480–1509 (2013)
20. Beck, A., Hallak, N.: On the minimization over sparse symmetric sets: projections, optimality conditions, and algorithms. *Math. Oper. Res.* **41**(1), 196–223 (2016)
21. Burdakov, O., Kanzow, C., Schwartz, A.: Mathematical programs with cardinality constraints: reformulation by complementarity-type constraints and a regularization method. *SIAM J. Optim.* **26**(1), 397–425 (2016)
22. Červinka, M., Kanzow, C., Schwartz, A.: Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. *Math. Program.* **160**, 353–377 (2016)
23. Li, X., Song, W.: The first-order necessary conditions for sparsity constrained optimization. *J. Oper. Res. Soc. China* **3**(4), 521–535 (2015)
24. Lu, Z., Zhang, Y.: Sparse approximation via penalty decomposition methods. *SIAM J. Optim.* **23**(4), 2448–2478 (2013)
25. Lu, Z.: Optimization over sparse symmetric sets via a nonmonotone projected gradient method. [arXiv:1509.08581v3](https://arxiv.org/abs/1509.08581v3) (2015)
26. Pan, L.L., Xiu, N.H., Zhou, S.L.: On solutions of sparsity constrained optimization. *J. Oper. Res. Soc. China* **3**(4), 421–439 (2015)
27. Bertsimas, D., Shioda, R.: Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.* **43**(1), 1–22 (2009)
28. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. *J. Comput. Graph. Stat.* **15**(2), 265–286 (2006)
29. d’Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R.: A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49**(3), 434–448 (2007)
30. Fukuda, M., Kojima, M., Murota, K., Nakata, K.: Exploiting sparsity in semidefinite programming via matrix completion I: general framework. *SIAM J. Optim.* **11**(3), 647–674 (2001)
31. Heiler, M., Schnörr, C.: Learning sparse representations by non-negative matrix factorization and sequential cone programming. *J. Mach. Learn. Res.* **7**, 1385–1407 (2006)
32. Rockafellar, R.T., Wets, R.J.: *Variational Analysis*. Springer, Berlin (1998)
33. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. SIAM, Philadelphia, PA (1990)
34. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer, Berlin (2000)
35. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton, NJ (2015)
36. Flegel, M.L., Kanzow, C., Outrata, J.V.: Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints. *Set-Valued Anal.* **15**(2), 139–162 (2007)
37. Moldovan, A., Pellegrini, L.: On regularity for constrained extremum problems. Part 1: sufficient optimality conditions. *J. Optim. Theory Appl.* **142**(1), 147–163 (2009)
38. Moldovan, A., Pellegrini, L.: On regularity for constrained extremum problems. Part 2: necessary optimality conditions. *J. Optim. Theory Appl.* **142**(1), 165–183 (2009)