Optimality Conditions for Cardinality Constrained Optimization Problems

by

Zhuoyu Xiao
B.Sc., Jinan University, 2020

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER of SCIENCE

in the Department of Mathematics and Statistics

Optimality Conditions for Cardinality Constrained Optimization Problems

by

Zhuoyu Xiao
B.Sc., Jinan University, 2020

Supervisory Committee

Dr. Jane Ye, Supervisor
(Department of Mathematics and Statistics)

Dr. Yu-Ting Chen, Departmental Member
(Department of Mathematics and Statistics)

ABSTRACT

Cardinality constrained optimization problems (CCOP) are a new class of optimization problems with many applications. In this thesis, we propose a framework called mathematical programs with disjunctive subspaces constraints (MPDSC), a special case of mathematical programs with disjunctive constraints (MPDC), to investigate CCOP. Our method is different from the relaxed complementarity-type reformulation in the literature.

The first contribution of this thesis is that we study various stationarity conditions for MPDSC, and then apply them to CCOP. In particular, we recover disjunctive-type strong (S-) stationarity and Mordukhovich (M-) stationarity for CCOP, and then reveal the relationship between them and those from the relaxed complementarity-type reformulation.

The second contribution of this thesis is that we obtain some new results for MPDSC, which do not hold for MPDC in general. We show that many constraint qualifications like the relaxed constant positive linear dependence (RCPLD) coincide with their piecewise versions for MPDSC. Based on such result, we prove that RCPLD implies error bounds for MPDSC. These two results also hold for CCOP. All of these disjunctive-type constraint qualifications for CCOP derived from MPDSC are weaker than those from the relaxed complementarity-type reformulation in some sense.

Contents

Supervisory Committee **ii**

Abstract **iii**

Contents **iv**

List of Figures **vi**

List of Tables **vii**

Acknowledgements **viii**

Dedication **ix**

1 Introduction **1**

2 Preliminaries **6**

2.1 Constraint qualifications for NLP 6

2.2 Background on variational analysis 8

3 Optimality conditions for MPDSC **12**

3.1 Computations of various cones to set S 13

3.2 Stationarity conditions for MPDSC 17

3.3 Constraint qualifications for MPDSC 31

3.3.1 The generator method for MPDSC 31

3.3.2 Constraint qualifications for MPDSC from MPDC 34

3.3.3 RCPLD as a sufficient condition for error bounds 38

4 Optimality Conditions for CCOP from MPDSC **44**

4.1 Disjunctive subspaces reformulation for CCOP 44

4.2 Stationarity conditions for CCOP 45

4.2.1 Disjunctive-type S-stationarity and M-stationarity for CCOP . 45
4.2.2 Other stationarity conditions for CCOP 50
4.3 Disjunctive-type constraint qualifications for CCOP 58

5 Conclusions and Future Work **65**

Bibliography **66**

List of Figures

Figure 1.1 Two examples: (i) $\|x\|_0 \leq 1$ in $\mathbb{R}^2$; (ii) $\|x\|_0 \leq 2$ in $\mathbb{R}^3$. 2

Figure 2.1 Relations among constraint qualifications for NLP 8

Figure 3.1 Relations among constraint qualifications for MPDSC 43

Figure 4.1 Relations among disjunctive-type stationarity conditions for CCOP 58

Figure 4.2 Relations among disjunctive-type constraint qualifications for CCOP 64

List of Tables

Table 3.1 Various cones to set S in MPDSC 16

Table 4.1 Various cones to set S in CCOP 46

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my advisor, Dr. Jane Ye. Half of my master's studies were done remotely due to the global pandemic. I would like to thank Dr. Ye for patiently discussing the progress of research with me via Zoom during the first year of the pandemic.

During my second year, I made the decision to pursue my PhD in the United States at short notice. Dr. Ye gave me many helpful suggestions on choosing a future career in academia or industry. She also told me that the most important thing is to figure out what I really want, what kind of person I want to be, and what kind of life I want to pursue in the future. I still remember these words, and plan to practice them in the future, just as she decided to choose a math career after getting her MBA.

I would like to thank my friends at the Department of Mathematics and Statistics: Chenqian Long, Xiaoxiao Ma, Manting Wang, Zhengxiao Wei, Yichen Yan, Tingzhou Yu, Shangzhi Zeng and Guangzhao Zhu. I would also like to thank all faculty and staff at the Department of Mathematics and Statistics for their hard work. I would also like to thank my landlord Rita Su for giving me a lot of advice and help in life. Victoria is a very beautiful place, I am grateful for all the kind and lovely people I have met in this land so far from home.

Finally, I would like to express my gratitude to my family. Thank you all for their unconditional love, support and encouragement over the years.

DEDICATION

In memory of Professor Asen L. Dontchev (1948-2021)

1 Introduction

Cardinality constrained optimization problems (CCOP) occur in many applications; e.g., image processing, portfolio optimization, machine learning, and other related problems [38, Section 2]. In this thesis, we consider cardinality constrained optimization problems of the form:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in X,\ \|x\|_0 \le s, \end{aligned} \tag{1.1}$$

where $X \subseteq \mathbb{R}^n$ defines the standard constraint set

$$X := \{x \in \mathbb{R}^n \mid g(x) \le 0,\ h(x) = 0\}$$

and $\|x\|_0$ denotes the number of nonzero components of vector $x \in \mathbb{R}^n$. We assume $s < n$, otherwise the cardinality constraint would be superfluous. Two simple examples about the l_0-norm are given in Figure 1.1. Unless otherwise mentioned, we assume that $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are all smooth functions.

The optimization problem (1.1) is difficult to solve since the l_0-norm of x is a nonconvex and noncontinuous function. Even testing the feasibility of the problem (1.1) is known to be NP-complete [11]. We emphasize some notable approaches in the literature for solving CCOP closely related to our work. Some optimality conditions for CCOP are obtained by using various cones in variational analysis [5, 35]. In [6], Beck and Eldar introduced three optimality criteria designed for CCOP, and proposed some efficient numerical algorithms aimed at finding points satisfying each criterion, respectively. Although Beck and Eldar's work has shown significant performance in convergence analysis, the biggest limitation of their work is its underlying assumption

Figure 1.1: Two examples: (i) $\|x\|_0 \leq 1$ in $\mathbb{R}^2$; (ii) $\|x\|_0 \leq 2$ in $\mathbb{R}^3$.

that $X = \mathbb{R}^n$, which means that equality and inequality constraints do not apply. In order to deal with the general situation, Burdakov et al. [14] took ideas from the mixed-integer problem, and made the following relaxed complementarity-type reformulation of problem (1.1):

$$\begin{aligned} \min_{x,y} \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \\ & e^T y \geq n - s, \\ & x_i y_i = 0,\ i = 1, \ldots, n, \\ & 0 \leq y_i \leq 1,\ i = 1, \ldots, n. \end{aligned} \tag{1.2}$$

where $y \in \mathbb{R}^n$ is an auxiliary variable and $e := (1, 1, \ldots, 1)^T \in \mathbb{R}^n$. Based on this continuous reformulation (1.2), some constraint qualifications and stationarity conditions for CCOP were introduced in Červinka et al. [15]. Some second-order necessary and sufficient optimality conditions for CCOP based on such reformulation are also introduced in Kanzow et al. [13]. However, this kind of reformulation has drawback that the local minimizer x^* of the reformulated problem (1.2) coincides with that of the original problem (1.1) only when the condition $\|x^*\|_0 = s$ is satisfied [14, Theorem 3.6] (a counter example is given in [14, Example 3]), and the case $\|x^*\|_0 < s$ is a harder problem.

As we can see from two examples above, the l_0-norm constraint set is the union of finitely many subspaces. Motivated by this observation, we propose a framework called mathematical programs with disjunctive subspaces constraints (MPDSC) whose constraint sets are the union of finitely many subspaces to investigate CCOP. MPDSC is a special case of mathematical programs with disjunctive constraints (MPDC). MPDC include many prominent optimization problems such as mathematical programs with equilibrium constraints (MPEC) [26], mathematical programs with vanishing constraints (MPVC) [1], and mathematical programs with switching constraints (MPSC) [29]. We would like to point out here that the disjunctive reformulation of CCOP was first mentioned in Pan et al. [34], and some new optimality conditions for CCOP based on such reformulation like the linear independence constraint qualification (LICQ) were derived in Mehlitz's work [28, Section 5.3] although they did not consider the general subspaces framework at all.

The stationarity conditions such as strong (S-) stationarity and Mordukhovich (M-) stationarity tailored for MPDC were introduced by Flegel et al. [17]. However, S-stationarity always requires strong constraint qualifications and M-stationarity does not preclude the existence of feasible descent directions. To deal with these issues, some new stationarity conditions stronger than M-stationarity like extended M-stationarity were introduced by Gfrerer [18]. In addition, Benko and Gfrerer [9] also introduced $\mathcal{Q}$-stationarity and $\mathcal{Q}_M$-stationarity for MPDC to build a bridge between Bouligand (B-) stationarity and M-stationarity, and $\mathcal{Q}_M$-stationarity is also stronger than M-stationarity. Despite the comprehensive calculus available for limiting normal cones which are closely related to M-stationarity, it is sometimes difficult to calculate the limiting normal cone to the general set. In order to deal with this difficulty, Gfrerer [19] introduced the linearized M-stationarity for MPDC. Tempted by the merit of sequential necessary conditions that they hold even without any additional constraint qualifications, asymptotically Mordukhovich (AM-) stationarity for MPDC has been proposed by Mehlitz [30].

Also, several recent developments of constraint qualifications for MPDC are notable. In Xu and Ye's recent work [39], many constraint qualifications such as the relaxed constant positive linear dependence (RCPLD) for general disjunctive programs were introduced. Further, they also introduced the piecewise RCPLD under which error bounds hold provided that inequality constraints and sets are Clarke regular. Recently directional variational analysis has become a popular topic in optimization. Gfrerer et al. [18] introduced the linear independence constraint qualification in the direction d (LICQ (d)) for MPDC. Bai et al. [4] introduced the directional quasi/pseudo-normality as sufficient conditions for the metric subregularity, and these results were simplified by Benko at el. [7] when considering the so-called mathematical programs with ortho-disjunctive constraints (MPODC), a special case of MPDC.

The main contributions of this thesis are summarized as follows:

(i) Firstly, we propose a general framework called mathematical programs with disjunctive subspaces constraints (MPDSC) which include mathematical programs with switching constraints (MPSC) [29] and CCOP as special cases. Under such framework, we study various optimality conditions and then apply them to CCOP. In particular, we recover disjunctive-type S-stationarity and M-stationarity for CCOP derived by Pan et al. [34]. Our disjunctive-type S-stationarity corresponds to the B-KKT conditions, and our disjunctive-type M-stationarity corresponds to the M-KKT conditions [34, Definition 3.1]. Moreover, we will reveal the relationship between them and those obtained from the relaxed complementarity-type reformulation [14].

(ii) Inspired by the fact that RCPLD coincides with piecewise RCPLD for MPSC [39, Section 5.3], we generalize this result and show that many constraint qualifications coincide with their piecewise versions for MPDSC. Based on such result, we prove that for MPDSC, RCPLD is a sufficient condition for error bounds. Note that this result does not hold for MPDC in general. Since CCOP is a special case of MPDSC, the above results also hold for CCOP. Our disjunctive-type constraint qualifications for CCOP are weaker than those from the relaxed complementarity-type reformulation [15] in some sense.

The remainder of this thesis is organized as follows. In Chapter 2, we review some preliminary knowledge about nonlinear programming and variational analysis. In Chapter 3, we study various stationarity conditions and constraint qualifications for MPDSC. We also show that RCPLD is a sufficient condition for error bounds for MPDSC, an important result which does not hold for MPDC in general. In Chapter 4, we first reformulate CCOP as MPDSC, and then apply the results in Chapter 3 to CCOP. In particular, we recover disjunctive-type S-stationarity and M-stationarity in [34] and reveal the relationship between them and complementarity-type ones [14]. We also study constraint qualifications for CCOP, and then show that RCPLD implies error bounds for CCOP. Comparisons between these disjunctive-type constraint qualifications and those from the relaxed complementarity-type reformulation are also mentioned at the end of this chapter. In Chapter 5, we will briefly talk about our future research.

Notation Given a point $x^* \in \mathbb{R}^n$, the symbol $\mathbb{B}_\varepsilon(x^*)$ stands for the open ball of radius ε centered at x^*, while the symbol $\mathbb{B}$ simply stands for the open unit ball centered at the origin. We denote by $\nabla f(x^*)$ the gradient of a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ at x^* and by $d_\Omega(x)$ the distance between the point x and the set Ω. Unless otherwise mentioned, $\| \cdot \|$ denotes an arbitrary norm in $\mathbb{R}^n$ and the notation $\langle \cdot, \cdot \rangle$ denotes the inner product. For a given nonempty set $A \subseteq \mathbb{R}^n$, we use notations $\text{cl}A$, $\text{cone}A$, and $\text{span}A$ to represent the closure of A, the conic hull of A, and the span of A, that is, the smallest subspace of $\mathbb{R}^n$ comprising A. For any given set B with finite elements, we denote the number of elements in B by $|B|$.

2 Preliminaries

2.1 Constraint qualifications for NLP

Consider the standard nonlinear program (NLP)

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g(x) \le 0, h(x) = 0, \end{aligned} \tag{2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable. We denote $\mathcal{F}_{NLP}$ by the feasible region of nonlinear program (2.1) and $\mathcal{I}_g(x^*) := \{i \in \{1, \ldots, m\} \mid g_i(x^*) = 0\}$ by the active set for inequality constraints at x^*. Now we recall some well-known constraint qualifications for NLP as follows.

Definition 2.1. *Let $x^* \in \mathbb{R}^n$ be a feasible point of problem* (2.1). *We say that x^* satisfies*

1. the linear independence constraint qualification (LICQ), *if the family of gradients $\{\nabla g_i(x^*)\}_{i \in \mathcal{I}_g(x^*)} \cup \{\nabla h_i(x^*)\}_{i=1}^p$ is linearly independent;*

2. Mangasarian-Fromovitz constraint qualification (MFCQ) [27], *if the family of gradients $\{\nabla g_i(x^*)\}_{i \in \mathcal{I}_g(x^*)} \cup \{\nabla h_i(x^*)\}_{i=1}^p$ is positively linearly independent, i.e., the family of gradients $\{\nabla g_i(x^*)\}_{i \in \mathcal{I}_g(x^*)} \cup \{\nabla h_i(x^*)\}_{i=1}^p$ is linearly independent with non-negative scalars associated with the gradients of the active inequality constraints;*

3. the constant rank constraint qualification (CRCQ) [23], *if there exists $\varepsilon > 0$*

such that for every $I \subseteq \mathcal{I}_g(x^*)$ *and every* $J \subseteq \{1, \dots, p\}$*, the family of gradients*

$$\{\nabla g_i(x^*)\}_{i \in I} \cup \{\nabla h_i(x^*)\}_{i \in J}$$

and the family of gradients

$$\{\nabla g_i(x)\}_{i \in I} \cup \{\nabla h_i(x)\}_{i \in J}$$

has the same rank for every $x \in \mathbb{B}_\varepsilon(x^*)$*;*

4. the relaxed constant rank constraint qualification (RCRCQ) [31]*, if there exists* $\varepsilon > 0$ *such that for every* $I \subseteq \mathcal{I}_g(x^*)$*, the family of gradients*

$$\{\nabla g_i(x^*)\}_{i \in I} \cup \{\nabla h_i(x^*)\}_{i=1}^p$$

and the family of gradients

$$\{\nabla g_i(x)\}_{i \in I} \cup \{\nabla h_i(x)\}_{i=1}^p$$

has the same rank for every $x \in \mathbb{B}_\varepsilon(x^*)$*;*

5. the constant positive linear dependence constraint qualification (CPLD) [36]*, if there exists* $\varepsilon > 0$ *such that for every* $I \subseteq \mathcal{I}_g(x^*)$ *and every* $J \subseteq \{1, \dots, p\}$*, whenever the family of gradients* $\{\nabla g_i(x^*)\}_{i \in I} \cup \{\nabla h_i(x^*)\}_{i \in J}$ *is positive linearly dependent, then* $\{\nabla g_i(x)\}_{i \in I} \cup \{\nabla h_i(x)\}_{i \in J}$ *is linearly dependent for every* $x \in \mathbb{B}_\varepsilon(x^*)$*;*

6. the relaxed constant positive linear dependence constraint qualification (RCPLD) [3]*, if there exists* $\varepsilon > 0$ *such that:*
   - (i) *The vectors* $\{\nabla h_i(x)\}_{i=1}^p$ *has the same rank for all* x *in* $\mathbb{B}_\varepsilon(x^*)$*;*
   - (ii) *Let* $J \subseteq \{1, \dots, p\}$ *be such that* $\{\nabla h_i(x^*)\}_{i \in J}$ *is a basis for* $\mathrm{span}\{\nabla h_i(x^*)\}_{i=1}^p$*. For every* $I \subseteq \mathcal{I}_g(x^*)$*, if the family of gradients* $\{\nabla g_i(x^*)\}_{i \in I} \cup \{\nabla h_i(x^*)\}_{i \in J}$ *is positive linearly dependent, then* $\{\nabla g_i(x)\}_{i \in I} \cup \{\nabla h_i(x)\}_{i \in J}$ *is linearly dependent for every* $x \in \mathbb{B}_\varepsilon(x^*)$*.*

Definition 2.2. (Error bounds for NLP) *We say that a feasible point* $x^* \in \mathcal{F}_{NLP}$

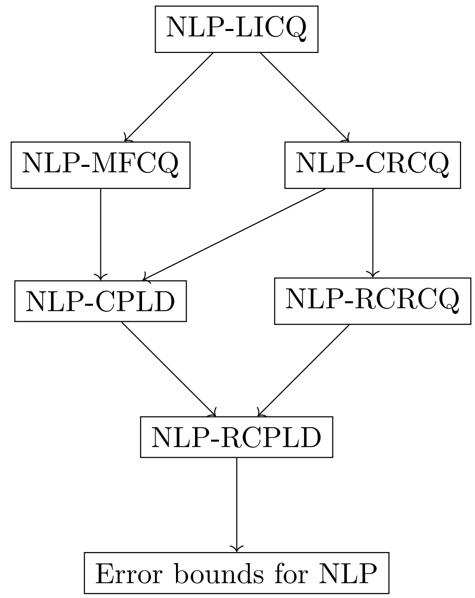

Figure 2.1: Relations among constraint qualifications for NLP

satisfies the error bound property if there exists $\alpha \geq 0$, $\varepsilon > 0$ *such that*

$$d_{\mathcal{F}_{NLP}}(x) \leq \alpha\Big(\sum_{i=1}^{m} \max\{g_i(x), 0\} + \sum_{i=1}^{p} |h_i(x)|\Big),\ \forall x \in \mathbb{B}_\varepsilon(x^*).$$

We conclude Section 2.1 with Figure 2.1 summarizing relations among various constraint qualifications for the standard nonlinear program (2.1).

2.2 Background on variational analysis

In this section, we recall some basic concepts from variational analysis which will be used frequently in this thesis. All of these subjects can be found in Clarke [16], Mordukhovich [32, 33], Bonnans and Shapiro [12], and Rockafellar and Wets [37].

Definition 2.3. (Cones and polar cones) [12, Section 2.1.4] *A nonempty subset* $K \subseteq \mathbb{R}^n$ *is said to be a cone if for any* $x \in K$ *and any* $t \geq 0$, *it follows that* $tx \in K$. *For a cone* $K \subseteq \mathbb{R}^n$ *its polar cone* K° *is defined as follows:*

$$K^\circ := \{y \in \mathbb{R}^n \mid \forall x \in K : \langle x, y\rangle \leq 0\}.$$

Remark 2.1. *Let us consider a special case. If* K *is a subspace in* $\mathbb{R}^n$, *then the polar cone* K° *coincides with the orthogonal complement* $K^\perp$ *in the sense of linear algebra.*

Now let us recall the concept of Painlevé-Kuratowski outer limits.

Definition 2.4. (Painlevé-Kuratowski outer limits) [33, p.2] *Consider a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with values $F(x) \subseteq \mathbb{R}^m$ in the collection of all the subsets of $\mathbb{R}^m$. The limiting construction*

$$\limsup_{x \to x^*} F(x) := \{y \in \mathbb{R}^m \mid \exists x_k \to x^*, y_k \to y \text{ with } y_k \in F(x_k),\ k \to \infty\}$$

is known as the Painlevé-Kuratowski outer limit of F at x^*.

The following concept of Bouligand tangent cones can be defined by Painlevé-Kuratowski outer limits.

Definition 2.5. (Bouligand tangent cones) [37, Definition 6.1] *Given a closed set $\Omega \subseteq \mathbb{R}^n$ and $x^* \in \Omega$, the Bouligand tangent cone to set Ω at x^* is defined by*

$$T_\Omega(x^*) := \limsup_{t \downarrow 0} \frac{\Omega - x^*}{t} = \{v \in \mathbb{R}^n \mid \exists x_k \overset{\Omega}{\to} x^*, t_k \downarrow 0 \text{ with } \frac{x_k - x^*}{t_k} \to v,\ k \to \infty\},$$

where $x_k \overset{\Omega}{\to} x^*$ means that $x_k \to x^*$ with $x_k \in \Omega$.

Using the definition of polar cones, we now recall the concepts of Fréchet/regular normal cones and Mordukhovich/limiting normal cones.

Definition 2.6. (Fréchet and Mordukhovich normal cones) [37, Definition 6.3, Proposition 6.5] *Given a closed set $\Omega \subseteq \mathbb{R}^n$ and $x^* \in \Omega$, the Fréchet/regular normal cone to set Ω at x^* is defined by*

$$\hat{N}_\Omega(x^*) := (T_\Omega(x^*))^\circ = \{v \in \mathbb{R}^n \mid \langle v, x - x^* \rangle \leq o(\|x - x^*\|),\ \forall x \in \Omega\}$$

and the Mordukhovich/limiting normal cone to set Ω at x^* is defined by

$$N_\Omega(x^*) := \limsup_{x \to x^*} \hat{N}_\Omega(x) = \{v \in \mathbb{R}^n \mid \exists x^k \overset{\Omega}{\to} x^*,\ \exists v^k \to v,\ \text{s.t. } v^k \in \hat{N}_\Omega(x^k)\}.$$

From the definition, we can see that limiting normal cones are in general nonconvex while regular normal cones are always convex. If Ω is convex, we have $\hat{N}_\Omega(x^*) = N_\Omega(x^*)$. The following concept of directional limiting normal cones was introduced by Ginchev and Mordukhovich [21].

Definition 2.7. (Directional normal cones) [21, Definition 2.3] *Given a closed set $\Omega \subseteq \mathbb{R}^n$, $x^* \in \Omega$ and $d \in \mathbb{R}^n$. The limiting normal cone to Ω at x^* in direction d is defined by*

$$N_\Omega(x^*; d) := \{v \in \mathbb{R}^n \mid \exists t_k \downarrow 0, d^k \to d, v^k \to v, \text{ s.t. } v^k \in \hat{N}_\Omega(x^k + t_k d^k)\}.$$

It is easy to see that the limiting normal cone to Ω at x^* in direction $d = 0$ is equal to the limiting normal cone. Moreover, it is obvious that $N_\Omega(x^*; d) \subseteq N_\Omega(x^*)$ and $N_\Omega(x^*; d) = \emptyset$ if $d \notin T_\Omega(x^*)$. The following estimate for directional normal cones is useful in Section 3.

Theorem 2.1. [18, Lemma 2.1] *Let Ω be the union of finitely many closed convex sets. Then,*

$$N_\Omega(x^*; d) \subseteq \{v \in N_\Omega(x^*) \mid v^T d = 0\}.$$

Usually the Certesian product rule holds for normal cones without any assumption as shown in the following theorem. However, for tangent cones and directional normal cones, we need to impose on some additional condition.

Theorem 2.2. [37, Proposition 6.41] *Let $C = C_1 \times C_2 \times \cdots \times C_m$, where $C_i \subseteq \mathbb{R}^{n_i}$ is closed for $i = 1, 2, \ldots, m$ and $n = n_1 + n_2 + \cdots + n_m$. For any $x^* = (x_1^*, \ldots, x_m^*) \in C$ with $x_i^* \in C_i$ for $i = 1, 2, \ldots, m$, one has*

$$N_C(x^*) = N_{C_1}(x_1^*) \times \cdots \times N_{C_m}(x_m^*),$$
$$\hat{N}_C(x^*) = \hat{N}_{C_1}(x_1^*) \times \cdots \times \hat{N}_{C_m}(x_m^*).$$

Definition 2.8. (Directionally regularity) [40, Definitin 3.3] *Let $\Omega \subseteq \mathbb{R}^n$ be closed, $x^* \in \Omega$ and $d \in \mathbb{R}^n$. We say that set Ω is directionally regular at x^* if*

$$N_\Omega(x^*; d) = N_\Omega^i(x^*; d),$$

where $N_\Omega^i(x^*; d) := \{v \in \mathbb{R}^n \mid \forall t_k \downarrow 0, \exists d^k \to d, v^k \to v, \text{ s.t. } v^k \in \hat{N}_\Omega(x^k + t_k d^k)\}$. If Ω is directionally regular at any x, we say Ω is directionally regular.

By [40, Definitin 3.5], any closed convex set is directionally regular. Under such regularity condition, the following Certesian product rule for tangent cones and directional normal cones holds.

Theorem 2.3. [40, Proposition 3.3] *Let $C = C_1 \times C_2 \times \cdots \times C_m$, where $C_i \subseteq \mathbb{R}^{n_i}$ is closed for $i = 1, 2, \ldots, m$ and $n = n_1 + n_2 + \cdots + n_m$. Consider a point $x^* = (x_1^*, \ldots, x_m^*) \in C$ with $x_i^* \in C_i$ and a direction $d = (d_1, \ldots, d_m) \in \mathbb{R}^n$. Assume that all except at most one of C_i for $i = 1, \ldots, m$ are directionally regular at x_i^*, one has*

$$T_C(x^*) = T_{C_1}(x_1^*) \times \cdots \times T_{C_m}(x_m^*),$$
$$N_C(x^*; d) = N_{C_1}(x_1^*; d_1) \times \cdots \times N_{C_m}(x_m^*; d_m).$$

Finally, we recall the well-known metric subregularity property as well as its directional counterpart.

Definition 2.9. ((Directional) metric subregularity) [10, Definition 2.1] *Let $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $(x^*, y^*) \in gph\, M := \{(x, y) \mid y \in M(x)\}$. We say that M is metrically subregular at (x^*, y^*) provided there exists $\kappa > 0$ and $\rho > 0$ such that*

$$d_{M^{-1}(y^*)}(x) \leq \kappa d_{M(x)}(y^*),\ \forall x \in x^* + \rho\mathbb{B}.$$

Given $d \in \mathbb{R}^n$, we say M is metrically subregular in direction d at (x^*, y^*) if there exists $\rho > 0, \delta > 0$ with a directional neighborhood $V_{\rho,\delta}(d)$ of d such that the above estimate holds for all $x \in x^* + V_{\rho,\delta}(d)$, where $V_{\rho,\delta}(d)$ denotes the directional neighborhood at 0 in the direction d:

$$V_{\rho,\delta}(d) := \begin{cases} \{0\} \cup \left\{ z \in \rho\mathbb{B} \setminus \{0\} \;\middle|\; \left\| \frac{z}{\|z\|} - \frac{d}{\|d\|} \right\| \leq \delta \right\}, & \text{if } d \neq 0, \\ \rho\mathbb{B}, & \text{if } d = 0. \end{cases}$$

3 Optimality conditions for MPDSC

We first present mathematical programs with disjunctive constrains (MPDC):

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & F(x) \in \Lambda, \end{aligned} \tag{3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^d$ are continuously differentiable, and the constraint set $\Lambda \subseteq \mathbb{R}^d$ is a union of finitely many convex polyhedral sets.

In this thesis, we mainly consider a special case of MPDC called mathematical programs with disjunctive subspaces constraints (MPDSC):

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, \Phi_i(x) \in S,\ i = 1, \ldots, l, \end{aligned} \tag{3.2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$, $\Phi_i : \mathbb{R}^n \to \mathbb{R}^q$ are continuously differentiable, and constraint set $S := \bigcup_{r=1}^{R} S_r \subseteq \mathbb{R}^q$ with S_r being a subspace, $r = 1, \ldots, R$. Let $0 \neq c_j^r \in \mathbb{R}^n$ such that

$$S_r := \{x \in \mathbb{R}^n \mid \langle c_j^r, x \rangle = 0, j \in \mathcal{E}_r\}, \tag{3.3}$$

where $\mathcal{E}_r$ is a finite index set and vectors $\{c_j^r\}_{j \in \mathcal{E}_r}$ are linearly independent.

To see that problem (3.2) is a special case of MPDC, we can abbreviate $\Lambda := \mathbb{R}_-^m \times \{0\}^p \times S^l$ and $F(x) := (g(x), h(x), \Phi(x))$ where $\Phi(x) := (\Phi_1(x), \ldots, \Phi_l(x))$. Since each constraint set S is the union of R subspaces, set Λ is the union of $K := R^l$ convex polyhedral sets, hence Λ is a disjunctive set. Without loss of generality, we denote $\Lambda = \bigcup_{i=1}^{K} \Lambda_i$ where Λ_i is a polyhedral set and denote the feasible regions of problems (3.1) and (3.2) by $\mathcal{F} := \{x \in \mathbb{R}^n \mid F(x) \in \Lambda\}$, respectively. We will use the notation $I_F(x^*) := \{i \in \{1, \dots, K\} \mid F(x^*) \in \Lambda_i\}$ in the following context.

In Chapter 3, we study various optimality conditions for MPDSC. Before that, we compute various cones to set $S = \bigcup_{r=1}^{R} S_r$ in Section 3.1. In Section 3.2, we focus on various stationarity conditions for MPDSC. In Section 3.3, we focus on constraint qualifications for MPDSC. We show that for MPDSC many constraint qualifications like RCPLD coincide with corresponding piecewise constraint qualifications, respectively. Based on such result, we show that RCPLD implies error bounds for MPDSC, a result does not hold for MPDC in general.

3.1 Computations of various cones to set S

Let $x^* \in S$ and $d \in \mathbb{R}^n$. In this section, we compute various tangent and normal cones to set $S = \bigcup_{r=1}^{R} S_r$, such as the Bouligand tangent cone, the Mordukhovich/limiting normal cone, the Fréchet/regular normal cone, the directional normal cone, and the regular normal cone to tangent cone. We use the notations $I(x^*) := \{r \in \{1, \dots, R\} \mid x^* \in S_r\}$ for a given vector x in what follows.

Theorem 3.1. (Bouligand tangent cone) *Let $x^* \in S$. Then*

$$T_S(x^*) = \bigcup_{r \in I(x^*)} T_{S_r}(x^*) = \bigcup_{r \in I(x^*)} S_r. \tag{3.4}$$

Proof. The first equation in (3.4) is from [28, Lemma 2.2]. The second equation in (3.4) is obvious since each S_r is a subspace for $r = 1, \dots, R$. □

Theorem 3.2. (Fréchet/regular normal cone) *Let $x^* \in S$. Then*

$$\hat{N}_S(x^*) = \bigcap_{r \in I(x^*)} \hat{N}_{S_r}(x^*) = \bigcap_{r \in I(x^*)} S_r^{\perp}. \tag{3.5}$$

Proof. The proof is obvious by Theorem 3.1 and the fact $\hat{N}_S(x^*) = (T_S(x^*))^\circ$. □

Theorem 3.3. (Mordukhovich/limiting normal cone) *Let $x^* \in S$. Then*

$$N_S(x^*) = \bigcup_{r \in I(x^*)} N_{S_r}(x^*) = \bigcup_{r \in I(x^*)} S_r^{\perp}. \tag{3.6}$$

Proof. The second equation in (3.6) is obvious from the fact that each S_r is a subspace for $r = 1, \ldots, R$. We only need to prove the first equation.

The inclusion $N_S(x^*) \subseteq \bigcup_{r\in I(x^*)} N_{S_r}(x^*)$ is obvious from [28, Lemma 2.2]. Now we prove the converse inclusion. By noting that each S_r is a subspace thus convex for each $r \in I(x^*)$, we have

$$\bigcup_{r\in I(x^*)} N_{S_r}(x^*) = \bigcup_{r\in I(x^*)} \hat{N}_{S_r}(x^*). \tag{3.7}$$

Moreover, we can show that there exists $\delta > 0$ such that

$$\bigcup_{r\in I(x^*)} \hat{N}_{S_r}(x^*) \subseteq \bigcup_{x\in\mathbb{B}_\delta(x^*)} \hat{N}_S(x). \tag{3.8}$$

Indeed, for each $r \in I(x^*)$, we can pick a unit vector u_r such that $u_r \in S_r$ but $u_r \notin S_j$ for other $j \neq r$ since we only have finitely many subspaces here. Consider the point $x = x^* + tu_r$ such that $t < \delta$, we can see $x \in S_r$ but $x \notin S_j$ for other $j \neq r$. In this way, for every $i \in I(x^*)$ we have

$$\hat{N}_{S_r}(x^*) = (S_r)^\perp = \hat{N}_{S_r}(x) = \hat{N}_S(x) \subseteq \bigcup_{x\in\mathbb{B}_\delta(x^*)} \hat{N}_S(x),$$

where the equation $\hat{N}_{S_r}(x) = \hat{N}_S(x)$ is from the first equation in Theorem 3.2.

From (13) and (14) in Adam et al. [2], we further have the equality

$$\bigcup_{x\in\mathbb{B}_\delta(x^*)} \hat{N}_S(x) = N_S(x^*). \tag{3.9}$$

Combing (3.7), (3.8) and (3.9), we obtain the inclusion $\bigcup_{r\in I(x^*)} N_{S_r}(x^*) \subseteq N_S(x^*)$ as desired, which completes the proof. □

Theorem 3.4. (Directional normal cone) *Let $x^* \in S$ and $d \in T_S(x^*)$. Then*

$$N_S(x^*; d) = \{v \in N_S(x^*) \mid v^T d = 0\} = \bigcup_{r\in I(x^*)\cap I(d)} S_r^\perp. \tag{3.10}$$

Proof. Since S is a disjunctive set, by Theorem 2.1 we have

$$N_S(x^*; d) \subseteq \{v \in N_S(x^*) \mid v^T d = 0\}.$$

From Theorem 3.3, we have

$$\{v \in N_S(x^*) \mid v^T d = 0\} = \bigcup_{r \in I(x^*)} \{v \in N_{S_r}(x^*) \mid v^T d = 0\}. \tag{3.11}$$

Since each S_r is a subspace hence convex, we have

$$\bigcup_{r \in I(x^*)} \{v \in N_{S_r}(x^*) \mid v^T d = 0\} = \bigcup_{r \in I(x^*)} \{v \in \hat{N}_{S_r}(x^*) \mid v^T d = 0\}. \tag{3.12}$$

By [18, Lemma 2.1], it follows that

$$\bigcup_{r \in I(x^*)} \{v \in \hat{N}_{S_r}(x^*) \mid v^T d = 0\} \subseteq \bigcup_{r \in I(x^*)} \hat{N}_{T_{S_r}(x^*)}(d). \tag{3.13}$$

Further, since $T_{S_r}(x^*) = S_r$ for each $r \in I(x^*)$, we have

$$\bigcup_{r \in I(x^*)} \hat{N}_{T_{S_r}(x^*)}(d) = \bigcup_{r \in I(x^*)} \hat{N}_{S_r}(d) = \bigcup_{r \in I(x^*) \cap I(d)} S_r^{\perp}. \tag{3.14}$$

Combing (3.11), (3.12), (3.13) and (3.14), we establish the inclusion

$$\{v \in N_S(x^*) \mid v^T d = 0\} \subseteq \bigcup_{r \in I(x^*) \cap I(d)} S_r^{\perp}.$$

Finally, for any $v \in \bigcup_{r \in I(x^*) \cap I(d)} S_r^{\perp}$, there exists $r' \in I(x^*) \cap I(d)$ such that $v \in S_{r'}^{\perp}$. Since we only deal with finitely many subspaces, we can find a unit vector u' such that $u' \in S_{r'}$ but $u' \notin S_j$ for other $j \neq r'$. We define the sequences $\{d^k\}$ as

$$d^k := d + \frac{1}{k} u'.$$

Then $d^k \to d$ as $k \to \infty$.

Since $r' \in I(x^*) \cap I(d)$ we have $x^* \in S_{r'}$ and $d \in S_{r'}$, which implies that $d^k \in S_{r'}$ but $d^k \notin S_j$ for other $j \neq r'$. For any sequence $\{t_k\}$ such that $t_k \to 0^+$ as $k \to \infty$, it follows that $x^* + t_k d^k \in S_{r'}$ but $d^k \notin S_j$ for other $j \neq r'$. Therefore, we have

$$v \in S_{r'}^{\perp} = \hat{N}_{S_{r'}}(x^* + t_k d^k) = \hat{N}_S(x^* + t_k d^k),$$

where the last equation is from Theorem 3.2. From the definition of the directional normal cone (see Definition 2.7), we have

$$v \in N_S(x^*; d),$$

which implies

$$\bigcup_{r \in I(x^*) \cap I(d)} S_r^\perp \subseteq N_S(x^*; d)$$

as desired. □

Theorem 3.5. (Regular normal cone to tangent cone) *Let* $x^* \in S$ *and* $d \in \mathbb{R}^n$. *Then*

$$\hat{N}_{T_S(x^*)}(d) = \bigcap_{r \in I(x^*) \cap I(d)} S_r^\perp. \tag{3.15}$$

Proof. We denote $S(x^*) := \bigcup_{r \in I(x^*)} S_r$. Combing Theorem 3.1 with Theorem 3.2, we have

$$\hat{N}_{T_S(x^*)}(d) = \hat{N}_{S(x^*)}(d) = \bigcap_{r \in I(x^*) \cap I(d)} \hat{N}_{S_r}(d) = \bigcap_{r \in I(x^*) \cap I(d)} S_r^\perp,$$

which completes the proof. □

Table 3.1 summarizes various cones that were considered in this section.

Cones	Expressions	References
$T_S(x^*)$	$\bigcup_{r \in I(x^*)} S_r$	Theorem 3.1
$\hat{N}_S(x^*)$	$\bigcap_{r \in I(x^*)} S_r^\perp$	Theorem 3.2
$N_S(x^*)$	$\bigcup_{r \in I(x^*)} S_r^\perp$	Theorem 3.3
$N_S(x^*; d)$	$\bigcup_{r \in I(x^*) \cap I(d)} S_r^\perp$	Theorem 3.4
$\hat{N}_{T_S(x^*)}(d)$	$\bigcap_{r \in I(x^*) \cap I(d)} S_r^\perp$	Theorem 3.5

Table 3.1: Various cones to set S in MPDSC

3.2 Stationarity conditions for MPDSC

Before discussing various stationarity conditions for MPDSC, we first present two constraint qualifications for MPDSC that will be used frequently throughout this thesis. The following two definitions can be derived from [28, Definition 3.1] and [17, Definition 6], respectively.

Definition 3.1. (MPDSC-LICQ) *Let $x^* \in \mathcal{F}$ be a feasible point of problem* (3.2). *We say that x^* satisfies MPDSC-LICQ if there is no nonzero vector $\lambda := (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^{l} \mathbb{R}^q$ such that*

$$
\begin{aligned}
0 &= \sum_{i \in \mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{l} \nabla \Phi_i(x^*)^T \lambda_i^\Phi, \\
\lambda_i^\Phi &\in \sum_{r \in I(\Phi_i(x^*))} S_r^\perp, \; i = 1, \dots, l.
\end{aligned}
$$

Definition 3.2. (MPDSC-GGCQ) *Let $x^* \in \mathcal{F}$ be a feasible point of problem* (3.2). *We say that the generalized Guignard constraint qualifications* (GGCQ) *holds at x^* if*

$$\hat{N}_{\mathcal{F}}(x^*) = (L_{\mathcal{F}}^{lin}(x^*))^\circ,$$

where the linearization cone $L_{\mathcal{F}}^{lin}(x^*)$ of problem (3.2) *takes the following form:*

$$
L_{\mathcal{F}}^{lin}(x^*) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{cc} \nabla g_i(x^*) d \le 0, & i \in \mathcal{I}_g(x^*) \\ \nabla h_i(\bar{x}) d = 0, & i \in \{1, \dots, p\} \\ \nabla \Phi_i(x^*) d \in \bigcup\limits_{r \in I(\Phi_i(x^*))} S_r, & i \in \{1, \dots, l\} \end{array} \right\}.
$$

Remark 3.1. *Metric subregularity* (see Definition 2.9) *always implies GGCQ. Let $M(x) := F(x) - \Lambda$ and $(x^*, 0) \in gph\, M := \{(x, y) \mid y \in M(x)\}$. From* (13) *in* [17], *we know that GGCQ holds at x^* if M is metric subregular at $(x^*, 0)$.*

Utilizing two normal cones defined in Definition 2.6, some well-known stationarity conditions for general MPDC were defined.

Definition 3.3. ([17, Definition 1]) *Let $x^* \in \mathcal{F}$ be a feasible point of disjunctive problem* (3.1).

(i) *We say* x^* *is B-stationary if*

$$0 \in \nabla f(x^*) + \hat{N}_{\mathcal{F}}(x^*).$$

(ii) *We say* x^* *is S-stationary if*

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T \hat{N}_\Lambda(F(x^*)).$$

(iii) *We say* x^* *is M-stationary if*

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T N_\Lambda(F(x^*)).$$

We know that S-stationarity must be B-stationarity, since we always have the estimate $\nabla F(x^*)^T \hat{N}_\Lambda(F(x^*)) \subseteq \hat{N}_{\mathcal{F}}(x^*)$ [37, Theorem 6.14]. Now let us write down S-stationarity and M-stationarity for MPDSC (3.2) by the following analysis.

- (S-stationarity) For $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \hat{N}_\Lambda(F(x^*))$, from the product rule for normal cones (see Theorem 2.2), we know that $\lambda^g \in \hat{N}_{\mathbb{R}^m_-}(g(x^*)), \lambda^h \in \hat{N}_{\{0\}^p}(h(x^*))$ and $\lambda_i^\Phi \in \hat{N}_S(\Phi_i(x^*))$ for $i = 1, \ldots, l$.
  - (a) For $i \notin \mathcal{I}_g(x^*)$ we have $g_i(x^*) < 0$, which implies that $\hat{N}_{\mathbb{R}_-}(g_i(x^*)) = \{0\}$, hence $\lambda_i^g = 0$. For $i \in \mathcal{I}_g(x^*)$ we have $g_i(x^*) = 0$, which implies that $\hat{N}_{\mathbb{R}_-}(g_i(x^*)) = \mathbb{R}^+$, hence $\lambda_i^g \geq 0$.
  - (b) Since $h(x^*) = 0$ and $\{0\}^p$ is trivially convex, we have $\hat{N}_{\{0\}^p}(h(x^*)) = \mathbb{R}^p$, which means that $\lambda_i^h \in \hat{N}_{\{0\}}(h_i(x^*))$ can take any value for $i = 1, \ldots, p$.
  - (c) For any $\lambda_i^\Phi \in \hat{N}_S(\Phi_i(x^*))$, we have $\lambda_i^\Phi \in \bigcap_{r \in I(\Phi_i(x^*))} S_r^\perp$ from Theorem 3.2.
- (M-stationarity) For $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in N_\Lambda(F(x^*))$, most analysis are similar and the only difference is that we now have $\lambda_i^\Phi \in N_S(\Phi(x^*)) = \bigcup_{r \in I(\Phi_i(x^*))} S_r^\perp$ from Theorem 3.3.

Summarizing the analysis above, we have the following definition.

Definition 3.4. *Let* $x^* \in \mathcal{F}$ *be a feasible point of MPDSC* (3.2).

(i) *We say x^* is S-stationary if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ such that*

$$\nabla f(x^*) + \sum_{i \in \mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{l} \nabla \Phi_i(x^*)^T \lambda_i^\Phi = 0,$$
$$\lambda_i^g \geq 0,\ \forall i \in \mathcal{I}_g(x^*) \text{ and } \lambda_i^\Phi \in \bigcap_{r \in I(\Phi_i(x^*))} S_r^\perp,\ \forall i = 1, \ldots, l.$$

(ii) *We say x^* is M-stationary if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ such that*

$$\nabla f(x^*) + \sum_{i \in \mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{l} \nabla \Phi_i(x^*)^T \lambda_i^\Phi = 0,$$
$$\lambda_i^g \geq 0,\ \forall i \in \mathcal{I}_g(x^*) \text{ and } \lambda_i^\Phi \in \bigcup_{r \in I(\Phi_i(x^*))} S_r^\perp,\ \forall i = 1, \ldots, l.$$

The following theorem reveals the relationship between the local minimizer and stationarity conditions mentioned above.

Theorem 3.6. *Let x^* be a local minimizer of problem* (3.2)*. Then we have*

(i) *x^* is B-stationary.*

(ii) [28, Corollary 3.6] *Suppose MPDSC-LICQ holds at x^*, then x^* is S-stationary.*

(iii) [17, Theorem 7] *Suppose MPDSC-GGCQ holds at x^*, then x^* is M-stationary.*

Remark 3.2. *From* Remark 3.1 *and* Theorem 3.6*, we can see x^* is M-stationary if $M(x) = F(x) - \Lambda$ is metric subregular at* $(x^*, 0)$*.*

As we can see in the above theorem, S-stationarity always requires some strong constraint qualification conditions. Although M-stationarity holds under some weak conditions, it does not preclude the existence of feasible descent directions. Moreover, we can see from Definition 3.3 these well-known stationarity conditions are represented in terms of normal cones, hence a natural idea to discover new stationarity conditions stronger or weaker than the known ones is to make some new estimates of these cones.

In the recent paper [9], Benko and Gfrerer extended their initial work [8] in which the considered constraint set being an arbitrary closed set. They introduced $\mathcal{Q}$-stationarity and $\mathcal{Q}_M$-stationarity conditions for MPDC. When we consider MPDSC, we have the following statement:

"$\mathcal{Q}$-stationarity coincides with $\mathcal{Q}_M$-stationarity for MPDSC",

which does not hold for MPDC in general.

The key ideas behind $\mathcal{Q}$-stationarity and $\mathcal{Q}_M$-stationarity are as follows. Consider the disjunctive program (3.1), assume that we are given k convex cones Q_i such that $Q_i \subseteq T_\Lambda(F(x^*))$ for $i = 1\ldots,k$. Then we have

$$L_{\mathcal{F}}^{lin}(x^*) = \nabla F(x^*)^{-1}T_\Lambda(F(x^*)) \supseteq \nabla F(x^*)^{-1}Q_i,\ \forall i = 1,\ldots,k.$$

If we further assume GGCQ holds at x^* and

$$(F(x^*)^{-1}Q_i)^\circ = \nabla F(x^*)^T Q_i^\circ, \tag{3.16}$$

we have the following upper estimate of the regular normal cone

$$\hat{N}_{\mathcal{F}}(x^*) = (L_{\mathcal{F}}^{lin}(x^*))^\circ \ \subseteq (\nabla F(x^*)^{-1}Q_i)^\circ = \nabla F(x^*)^T Q_i^\circ,\ \forall i = 1,\ldots,k.$$

Condition (3.16) is fulfilled if for each $i = 1,\ldots,k$, $Q_i \subseteq T_\Lambda(F(x^*))$ is a convex polyhedral set, cf. [8, Proposition 1]. By [8, Lemma 1], we know that

$$(\nabla F(x^*)^T\Omega_1) \cap (\nabla F(x^*)^T\Omega_2) = \nabla F(x^*)^T(\Omega_1 \cap (\ker\nabla F(x^*)^T + \Omega_2))$$

holds for arbitrary sets $\Omega_1, \Omega_2 \subseteq \mathbb{R}^n$. It follows that

$$\begin{aligned}
\hat{N}_{\mathcal{F}}(x^*) &\subseteq \bigcap_{i=1}^{k} \nabla F(x^*)^T Q_i^\circ \\
&= \nabla F(x^*)^T Q_1^\circ \cap \nabla F(x^*)^T Q_2^\circ \cap \bigcap_{i=3}^{k} \nabla F(x^*)^T Q_i^\circ \\
&= \nabla F(x^*)^T (Q_1^\circ \cap (\ker\nabla F(x^*)^T + Q_2^\circ)) \cap \bigcap_{i=3}^{k} \nabla F(x^*)^T Q_i^\circ \\
&= (\nabla F(x^*)^T (Q_1^\circ \cap (\ker\nabla F(x^*)^T + Q_2^\circ))) \cap (\nabla F(x^*)^T Q_3^\circ) \cap \bigcap_{i=4}^{k} \nabla F(x^*)^T Q_i^\circ \\
&= \nabla F(x^*)^T (Q_1^\circ \cap (\ker\nabla F(x^*)^T + Q_2^\circ) \cap (\ker\nabla F(x^*)^T + Q_3^\circ)) \cap \bigcap_{i=4}^{k} \nabla F(x^*)^T Q_i^\circ \\
&= \cdots \\
&= \nabla F(x^*)^T (Q_1^\circ \cap \bigcap_{i=2}^{k} (\ker\nabla F(x^*)^T + Q_i^\circ)).
\end{aligned}$$

Therefore, we have the estimate

$$\hat{N}_{\mathcal{F}}(x^*) \subseteq \nabla F(x^*)^T (Q_1^\circ \cap \bigcap_{i=2}^{k} (\ker \nabla F(x^*)^T + Q_i^\circ)). \tag{3.17}$$

Further, if

$$\nabla F(x^*)^T (Q_1^\circ \cap \bigcap_{i=2}^{k} (\ker \nabla F(x^*)^T + Q_i^\circ)) \subseteq \nabla F(x^*)^T \hat{N}_\Lambda(F(x^*)), \tag{3.18}$$

then equality holds in (3.17) since by [37, Theorem 6.14] we always have the estimate

$$\nabla F(x^*)^T \hat{N}_\Lambda(F(x^*)) \subseteq \hat{N}_{\mathcal{F}}(x^*).$$

From the upper estimate (3.17), it is obvious that we want to choose $Q_i, i = 1, \ldots, k$ as large as possible in order that the inclusion is tight. It is reasonable that a good choice of $Q_1, Q_2, \ldots, Q_k$ should fulfill

$$\bigcap_{i=1}^{k} Q_i^\circ = \hat{N}_\Lambda(F(x^*)), \tag{3.19}$$

because under (3.19) equality holds in (3.18) either $\nabla F(x^*)$ has full rank or other weaker assumption are fulfilled (see [9, Theorem 3]).

Motivated by the upper estimate (3.17) of the regular normal cone $\hat{N}_{\mathcal{F}}(x^*)$, Benko and Gfrerer introduced $\mathcal{Q}$-stationarity and $\mathcal{Q}_M$-stationarity for MPDC. When considering MPDSC, we use the notation

$$I(\Phi(x^*)) := \prod_{j=1}^{l} I(\Phi_j(x^*)),$$

where $I(\Phi_j(x^*)) := \{r \in \{1, \ldots, R\} \mid \Phi_j(x^*) \in S_r\}$. Further, we denote $\mathcal{Q}(x^*)$ by the collection of all elements $(v^1, \ldots, v^k)$ with $v^i \in I(\Phi(x^*)),\ i = 1, \ldots, k$ such that

$$\{v_j^1, \ldots, v_j^k\} = I(\Phi_j(x^*)),\ j = 1, \ldots, l \tag{3.20}$$

and denote set $\Lambda(v^i)$ by

$$\Lambda(v^i) := \mathbb{R}^m_- \times \{0\}^p \times \prod_{j=1}^{l} S_{v^i_j},\ i = 1, \ldots, k.$$

Definition 3.5. *Let x^* be feasible for MPDSC* (3.2)*. We take convex cones $Q_i := T_{\Lambda(v^i)}(F(x^*))$ for $i = 1, \ldots, k$.*

(i) *We say that x^* is $\mathcal{Q}$-stationary for MPDSC* (3.2) *with respect to $Q_1, \ldots, Q_k$ if*

$$-\nabla f(x^*) \in \nabla F(x^*)^T (Q_1^\circ \cap \bigcap_{i=2}^{k} (\ker \nabla F(x^*)^T + Q_i^\circ))$$

(ii) *We say that x^* is $\mathcal{Q}_M$-stationary for MPDSC* (3.2) *with respect to $Q_1, \ldots, Q_k$ if*

$$-\nabla f(x^*) \in \nabla F(x^*)^T (N_\Lambda(F(x^*)) \cap Q_1^\circ \cap \bigcap_{i=2}^{k} (\ker \nabla F(x^*)^T + Q_i^\circ))$$

(iii) *We say that x^* is $\mathcal{Q}$-stationary or $\mathcal{Q}_M$-stationary for MPDSC* (3.2) *if x^* is $\mathcal{Q}$-stationary or $\mathcal{Q}_M$-stationary for MPDSC* (3.2) *with respect to some $Q_1, \ldots, Q_k$.*

Remark 3.3. *The original definitions of $\mathcal{Q}$-stationarity and $\mathcal{Q}_M$-stationarity for MPDC* (see [9, Definition 6]) *is more restrictive than the definitions we present here since they consider smaller subset such that* (26) *in* [9] *holds. We do not consider such case in this thesis for simplicity and such slight difference will disappear when considering MPDSC.*

The following theorem is immediate from the discussion before Definition 3.5. It holds not only for MPDSC but also for MPDC [9, Corollary 2].

Theorem 3.7. *Assume that MPDSC-GGCQ holds at a feasible point x^*. Given convex cones $Q_1, \ldots, Q_k \subseteq T_\Lambda(F(x^*))$ fulfilling* (3.16)*, x^* is $\mathcal{Q}$-stationary with respect to $Q_1, \ldots, Q_k$ if x^* is B-stationary. Conversely, if a feasible point x^* is $\mathcal{Q}$-stationary with respect to $Q_1, \ldots, Q_k$ and* (3.18) *is fulfilled, then x^* is S-stationary and consequently B-stationary.*

Now we prove $\mathcal{Q}$-stationarity coincides with $\mathcal{Q}_M$-stationarity for MPDSC, a fact does not hold for MPDC in general.

Theorem 3.8. *Let x^* be feasible for MPDSC* (3.2)*. Suppose x^* is $\mathcal{Q}$-stationary with respect to $Q_1, \ldots, Q_k$ defined as $Q_i = T_{\Lambda(v^i)}(F(x^*)),\ i = 1, \ldots, k$ in* Definition 3.5*, then x^* is also M-stationary* (see (ii) in Definition 3.4)*. In other words, $\mathcal{Q}$-stationarity coincides with $\mathcal{Q}_M$-stationarity for MPDSC.*

Proof. The proof is rather straightforward. Let x^* be $\mathcal{Q}$-stationary for MPDSC (3.2). Recalling the definition of $\mathcal{Q}$-stationarity with respect to $Q_1, \ldots, Q_k$, we have

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T (Q_1^\circ \cap \bigcap_{i=2}^{k} (\ker \nabla F(x^*)^T + Q_i^\circ)).$$

Assume that $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in Q_1^\circ \cap \bigcap_{i=2}^{k} (\ker \nabla F(x^*)^T + Q_i^\circ)$, it follows that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{l} \nabla \Phi_i(x^*)^T \lambda_i^\Phi = 0. \tag{3.21}$$

Since $Q_i = T_{\Lambda(v^i)}(F(x^*)),\ i = 1, \ldots, k$ we have

$$\begin{aligned} Q_i^\circ &= \hat{N}_{\Lambda(v^i)}(F(x^*)) \\ &= \hat{N}_{\mathbb{R}^m_-}(g(x^*)) \times \hat{N}_{\{0\}^p}(h(x^*)) \times \prod_{j=1}^{l} \hat{N}_{S_{v_j^i}}(\Phi_j(x^*)) \\ &= \hat{N}_{\mathbb{R}^m_-}(g(x^*)) \times \hat{N}_{\{0\}^p}(h(x^*)) \times \prod_{j=1}^{l} S^\perp_{v_j^i}. \end{aligned} \tag{3.22}$$

We can see the multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ must satisfy $\lambda \in Q_1^\circ$. The analysis of λ^g and λ^h are similar as the discussions before Definition 3.4, in what follows we focus on the discussion of λ^Φ. Since $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in Q_1^\circ$, we must have

$$\lambda_j^\Phi \in S^\perp_{v_j^1},\ j = 1, \ldots, l.$$

By (3.20), it follows that

$$\lambda_j^\Phi \in \bigcup_{r \in I(\Phi_j(x^*))} S_r^\perp,\ j = 1, \ldots, l,$$

which implies that x^* is M-stationary by (ii) in Definition 3.4. $\square$

We can see the upper estimate (3.17) plays a key role in the proposal of $\mathcal{Q}$-stationarity. Similarly, motivated by another estimate, Gfrerer [19] introduced the linearized M-stationarity by a repeated linearization procedure.

Theorem 3.9. [19, Theorem 2] *Assume that x^* is feasible for program* (3.1) *but with Λ being an arbitrary closed set and GGCQ is fulfilled at x^*. Then*

$$\hat{N}_{\mathcal{F}}(x^*) \subseteq \nabla F(x^*)^T N_{T_\Lambda(F(x^*))}(0) \subseteq F(x^*)^T N_\Lambda(F(x^*)). \tag{3.23}$$

In a word, the idea to estimate (3.23) is to utilize the limiting normal cone to the tangent cone to build a bridge between $\hat{N}_{\mathcal{F}}(x^*)$ and $F(x^*)^T N_\Lambda(F(x^*))$. We call this procedure linearization procedure. The linearization procedure would continue if $T_\Lambda(F(x^*))$ is not the union of finitely many convex polyhedral sets, until a series of tangent cones to tangent cones to set Λ is the union of finitely many convex polyhedral sets, which means the following estimate holds:

$$\hat{N}_{\mathcal{F}}(x^*) \subseteq \cdots \subseteq \nabla F(x^*)^T N_{T_{T_\Lambda(F(x^*))}(0)}(0) \subseteq \nabla F(x^*)^T N_{T_\Lambda(F(x^*))}(0).$$

The calculation of cone $N_{T_\Lambda(F(x^*))}(0)$ would be easier than that of Mordukhovich normal cone $N_\Lambda(F(x^*))$. Moreover, the linearized M-stationarity is sharper than M-stationarity since $N_{T_\Lambda(F(x^*))}(0) \subseteq N_\Lambda(F(x^*))$ always holds [37, Proposition 6.27]. But in our case, Λ is the union of finitely many convex polyhedral sets, hence we have $N_{T_\Lambda(F(x^*))}(0) = N_\Lambda(F(x^*))$ [22, p.59], which means that the linearized M-stationarity for MPDC should be

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T N_{T_\Lambda(F(x^*))}(0)$$

and linearized M-stationarity coincides with M-stationarity for MPDC. In particular, by the Certesian product rule (see Theorems 2.2 and 2.3), we can write down the linearized M-stationarity for MPDSC in the multipliers form, which is same as Definition 3.4 (ii).

From Theorem 3.6, we know even M-stationarity requires some mild constraint qualifications. The question is, whether it is possible to find a stationarity condition which holds for the local minimizer even in the absence of constraint qualifications. A potential candidate could be asymptotic version of M-stationarity (AM-stationarity) [30, Definition 3.1]. As the name suggests, the idea behind it is to ensure M-stationarity condition holds along a sequence of feasible points $\{x_k\}_{k\in\mathbb{N}}$ converging to local minimizer x^*. Based on this work, Liang and Ye defined the AM-stationarity for MPDC [25, Definition 3.4]. We can apply this definition to MPDSC directly.

Definition 3.6. *Let $x^* \in \mathcal{F}$ be feasible for MPDSC* (3.2)*. We say that x^* is AM-stationary if there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\varepsilon^k\} \subseteq \mathbb{R}^n$, $\{y^k\} := \{(y^{g,k}, y^{h,k}, y^{\Phi,k})\} \subseteq \mathbb{R}^d$ with $x^k \to x^*, \varepsilon^k \to 0$ and $y^k = (y^{g,k}, y^{h,k}, y^{\Phi,k}) \to 0$ such that $F(x^k) - y^k \in \Lambda$ and*

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^{p} \lambda_i^{h,k} \nabla h_i(x^k) + \sum_{i=1}^{n} \nabla \Phi_i(x^k)^T \lambda_i^{\Phi,k} = \varepsilon^k, \ \forall k, \tag{3.24}$$

the multipliers $\{(\lambda_i^{g,k}, \lambda_i^{h,k}, \lambda_i^{\Phi,k})\} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q$ should satisfy

$$\lambda_i^{g,k} = 0, \textit{if } g_i(x^k) - y_i^{g,k} < 0; \ \lambda_i^{g,k} \geq 0, \textit{if } g_i(x^k) - y_i^{g,k} = 0, \tag{3.25}$$

$$\lambda_i^{\Phi,k} \in N_S(\Phi_i(x^k) - y_i^{\Phi,k}) = \bigcup_{r \in I(\Phi_i(x^k) - y_i^{\Phi,k})} S_r^{\perp}. \tag{3.26}$$

It is easy to see an M-stationarity must be AM-stationary, now the question is under what conditions an AM-stationarity will be M-stationary. Now we review the condition called AM-regularity as follows. The following definition and theorem are immediately from [30] in which AM-regularity for MPDC was discussed.

Definition 3.7. *Let $x^* \in \mathcal{F}$ feasible for MPDSC* (3.2)*. Define a set-valued mapping $\mathcal{K} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by means of*

$$\mathcal{K}(x) := \left\{ \sum_{i=1}^{m} \lambda_i^g \nabla g_i(x) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x) + \sum_{i=1}^{l} \nabla \Phi_i(x)^T \lambda_i^\Phi \;\middle|\; \begin{array}{c} \lambda_i^g \geq 0, \ i \in \mathcal{I}_g(x^*), \\ \lambda_i^\Phi \in N_S(\Phi_i(x^*)) = \\ \displaystyle\bigcup_{r \in I(\Phi_i(x^*))} S_r^{\perp}, \ i = 1, \ldots, l. \end{array} \right\}.$$

We say that x^* is AM-regular if the following condition holds:

$$\limsup_{x \to x^*} \mathcal{K}(x) \subseteq \mathcal{K}(x^*),$$

where the notation $\limsup_{x \to x^*}$ denotes the Painlevé-Kuratowski outer limit at x^* defined in Definition 2.4.

Theorem 3.10. *Let x^* be a local minimizer of MPDSC* (3.2)*. Then, x^* is AM-stationary. Moreover, x^* is M-stationary if x^* is AM-regular.*

Recently many directional optimality conditions for MPDC keep emerging in the literature. We denote the critical cone by $\mathcal{C}(x^*) := \{d \in L^{lin}_{\mathcal{F}}(x^*) \mid \nabla f(x^*)d \leq 0\}$, a subset of the linearization cone which consists of all potential descent directions.

Definition 3.8. ([18, p.909 and Definition 3.4]) *Let x^* be feasible for MPDC* (3.1) *and $d \in \mathcal{C}(x^*)$.*

(i) *We say x^* is S-stationary* (d) *if*

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T \hat{N}_{T_\Lambda(F(x^*))}(\nabla F(x^*)d).$$

(ii) *We say x^* is M-stationary* (d) *if*

$$0 \in \nabla f(x^*) + \nabla F(x^*)^T N_\Lambda(F(x^*); \nabla F(x^*)d).$$

(iii) *We say x^* is extended M-stationary if x^* is M-stationary* (d) *for all $d \in \mathcal{C}(x^*)$.*

Now we apply S-stationarity in the direction d (S-stationarity (d)), M-stationarity in the direction d (M-stationarity (d)), and extended M-stationarity (ext. M-stationarity) to MPDSC. Using the following two arguments, we obtain S-stationarity (d) and M-stationarity (d) for MPDSC.

- (S-stationarity (d)) For any $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \hat{N}_{T_\Lambda(F(x^*))}(\nabla F(x^*)d)$, from the product rule, we have $\lambda^g \in \hat{N}_{T_{\mathbb{R}^m_-}(g(x^*))}(\nabla g(x^*)d)$, $\lambda^h \in \hat{N}_{T_{\{0\}^p}(h(x^*))}(\nabla h(x^*)d)$ and $\lambda^\Phi \in \hat{N}_{T_{S^l}(\Phi(x^*))}(\nabla \Phi(x^*)d)$.
  - (a) For any $i \notin \mathcal{I}_g(d) := \{i \in \mathcal{I}_g(x^*) \mid \nabla g_i(x^*)d = 0\}$, we have two cases:
    - (i) If $i \notin \mathcal{I}_g(x^*)$, we have $g_i(x^*) < 0$ which implies $T_{\mathbb{R}_-}(g_i(x^*)) = \mathbb{R}$. Then, it follows that $\lambda^g_i = 0$ because $\hat{N}_{\mathbb{R}}(\nabla g_i(x^*)d) = \{0\}$.
    - (ii) If $i \in \mathcal{I}_g(x^*)$ and $\nabla g_i(x^*)d < 0$, we have $g_i(x^*) = 0$ which implies $T_{\mathbb{R}_-}(g_i(x^*)) = \mathbb{R}_-$. It follows that $\lambda^g_i = 0$ since $\hat{N}_{\mathbb{R}_-}(\nabla g_i(x^*)d) = \{0\}$.

    For those $i \in \mathcal{I}_g(d)$, we have $g_i(x^*) = 0$ and $\nabla g_i(x^*)d = 0$, which implies that $\lambda^g_i \geq 0$ since $T_{\mathbb{R}_-}(0) = \mathbb{R}_-$ and $\hat{N}_{\mathbb{R}_-}(0) = \mathbb{R}_+$.
  - (b) For any $i = 1, \ldots, p$, we obtain $\hat{N}_{T_{\{0\}}(h_i(x^*))}(\nabla h_i(x^*)d) = \mathbb{R}$ since $h_i(x^*) = 0$ and $\nabla h_i(x^*)d = 0$, which implies that λ^h_i can take any value for $i = 1, \ldots, p$.

(c) For any $\lambda_i^\Phi \in \hat{N}_{T_S(\Phi_i(x^*))}(\nabla\Phi_i(x^*)d)$, from Theorem 3.5 we have

$$\lambda_i^\Phi \in \hat{N}_{T_S(\Phi_i(x^*))}(\nabla\Phi_i(x^*)d) = \bigcap_{r\in I(\Phi_i(x^*))\cap I(\nabla\Phi_i(x^*)d)} S_r^\perp.$$

- (M-stationarity (d)) For $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in N_\Lambda(F(x^*); \nabla F(x^*)d)$, we have $\lambda^g \in N_{\mathbb{R}^m_-}(g(x^*); \nabla g(x^*)d)$, $\lambda^h \in N_{\{0\}^p}(h(x^*); \nabla h(x^*)d)$ and $\lambda^\Phi \in N_{S^l}(\Phi(x^*); \nabla\Phi(x^*)d)$:

(a) For multipliers λ_i^g, we have $\lambda_i^g = 0$ for $i \notin \mathcal{I}_g(d)$ and $\lambda_i^g \geq 0$ for $i \in \mathcal{I}_g(d)$.

(b) For multipliers λ_i^h, we have $\lambda_i^h \in N_{\{0\}}(h_i(x^*); \nabla h_i(x^*)d) = N_{\{0\}}(0; 0) = N_{\{0\}}(0) = \mathbb{R}$, which means that λ_i^h can take any value.

(c) For any $\lambda_i^\Phi \in N_S(\Phi_i(x^*); \nabla\Phi_i(x^*)d)$, from Theorem 3.4 we have

$$\lambda_i^\Phi \in N_S(\Phi_i(x^*); \nabla\Phi_i(x^*)d) = \bigcup_{r\in I(\Phi_i(x^*))\cap I(\nabla\Phi_i(x^*)d)} S_r^\perp.$$

Summarizing the analysis above, we have the following definition.

Definition 3.9. *Let $x^* \in \mathcal{F}$ be a feasible point of MPDSC* (3.2) *and $d \in \mathcal{C}(x^*)$.*

(i) *We say x^* is S-stationary (d) if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ such that*

$$\nabla f(x^*) + \sum_{i\in\mathcal{I}_g(d)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^l \nabla\Phi_i(x^*)^T \lambda_i^\Phi = 0,$$

$$\lambda_i^g \geq 0,\ \forall i \in \mathcal{I}_g(d) \text{ and } \lambda_i^\Phi \in \bigcap_{r\in I(\Phi_i(x^*))\cap I(\nabla\Phi_i(x^*)d)} S_r^\perp,\ \forall i = 1, \ldots, l.$$

(ii) *We say x^* is M-stationary (d) if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ such that*

$$\nabla f(x^*) + \sum_{i\in\mathcal{I}_g(d)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^l \nabla\Phi_i(x^*)^T \lambda_i^\Phi = 0,$$

$$\lambda_i^g \geq 0,\ \forall i \in \mathcal{I}_g(d) \text{ and } \lambda_i^\Phi \in \bigcup_{r\in I(\Phi_i(x^*))\cap I(\nabla\Phi_i(x^*)d)} S_r^\perp,\ \forall i = 1, \ldots, l.$$

Similar to Remark 3.2 and Theorem 3.6, we have the corresponding theorems. Before presenting them, we give the definition of linear independence constraint qualification in the direction d (LICQ (d)) for MPDSC. The following definition is derived from (31) in [20].

Definition 3.10. (MPDSC-LICQ (d)) *Let $x^* \in \mathcal{F}$ be a feasible point of problem* (3.2)*. We say that x^* satisfies MPDSC-LICQ (d) if there is no nonzero vector $\lambda := (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^{l} \mathbb{R}^q$ such that*

$$0 = \sum_{i \in \mathcal{I}_g(d)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{l} \nabla \Phi_i(x^*)^T \lambda_i^\Phi,$$
$$\lambda_i^\Phi \in \operatorname{span} \bigcup_{r \in I(\Phi_i(x^*)) \cap I(\nabla \Phi_i(x^*) d)} S_r^\perp,\ i = 1, \ldots, l.$$

Theorem 3.11. *Let x^* be a local minimizer of program* (3.2) *and $d \in \mathcal{C}(x^*)$.*

(i) [20, Lemma 7] *Suppose LICQ (d) holds at x^*. Then, x^* is S-stationary (d).*

(ii) [18, Theorem 3.3] *Suppose $M(x) := F(x) - \Lambda$ is metrically subregualr at $(x^*, 0)$ in direction d. Then, x^* is M-stationary (d).*

(iii) [18, Theorem 3.9] *Suppose GGCQ is satisfied at the feasible point x^*. Then, x^* is ext. M-stationary.*

The following theorem illustrates the relationship between M-stationarity (d) and M-stationarity and the relationship between S-stationarity (d) and S-stationarity for MPDSC.

Theorem 3.12. *Let x^* be a feasible point for MPDSC* (3.2) *and $d \in \mathcal{C}(x^*)$. x^* is M-stationary if x^* is M-stationary (d). Moreover, suppose the condition $\mathcal{I}_g(d) = \mathcal{I}_g(x^*)$ holds, each S-stationary point x^* is S-stationary (d).*

Proof. The argument which says M-stationarity (d) implies M-stationarity is obvious. To see this, we can take $\lambda_i^g = 0$ for $i \in \mathcal{I}_g(x^*) \setminus \mathcal{I}_g(d)$. Since the condition

$$I(\Phi_i(x^*)) \cap I(\nabla \Phi_i(x^*) d) \subseteq I(\Phi_i(x^*)),\ \forall i = 1, \ldots, l \tag{3.27}$$

always holds, we have

$$\bigcup_{r \in I(\Phi_i(x^*)) \cap I(\nabla \Phi_i(x^*) d)} S_r^\perp \subseteq \bigcup_{r \in I(\Phi_i(x^*))} S_r^\perp,\ \forall i = 1, \ldots, l.$$

Let x^* be S-stationary for MPDSC. The condition $\lambda_i^g \geq 0, \forall i \in \mathcal{I}_g(d)$ in the definition of S-stationarity (d) holds since now we assume $\mathcal{I}_g(d) = \mathcal{I}_g(x^*)$. Moreover, from condition (3.27) we know that

$$\bigcap_{r \in I(\Phi_i(x^*))} S_r^{\perp} \subseteq \bigcap_{r \in I(\Phi_i(x^*)) \cap I(\nabla \Phi_i(x^*)d)} S_r^{\perp}, \ \forall i = 1, \ldots, l,$$

which means that x^* is S-stationary (d). Combining two cases discussed above, we complete the proof. □

At the end of this section, we summarize some second-order necessary and sufficient conditions for MPDSC (3.2). In what follows, we assume all functions are twice continuously differentiable and denote the Lagrangian function of problem (3.2) by

$$\mathcal{L}(x, \lambda^g, \lambda^h, \lambda^\Phi) := f(x) + \langle \lambda^g, g(x) \rangle + \langle \lambda^h, h(x) \rangle + \langle \lambda^\Phi, \Phi(x) \rangle.$$

We say that the quadratic growth condition is fulfilled at x^* if there are constants $\varepsilon > 0$ and $C > 0$ such that the following condition holds:

$$f(x) \geq f(x^*) + C\|x - x^*\|^2, \ \forall x \in \mathcal{F} \cap \mathbb{B}_\varepsilon(x^*). \tag{3.28}$$

In particular, we can see the quadratic growth condition (3.28) is sufficient for x^* be a strict local minimizer, that is,

$$f(x) > f(x^*), \ \forall x \in \mathcal{F} \cap \mathbb{B}_\varepsilon(x^*) \text{ such that } x \neq x^*.$$

Theorem 3.13. [28, Theorem 4.2 and Theorem 4.3] *Let x^* be a local minimizer of problem* (3.2)*. Assume that MPDSC-LICQ is valid at x^*. Then, we have*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^\Phi) d \geq 0, \ \forall d \in \mathcal{C}(x^*),$$

where $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ is the uniquely determined S-stationary multiplier associated with x^*.

Conversely, let x^* be an S-stationary point of problem (3.2)*. Assume there exists an S-stationary multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ associated with x^* such that*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^\Phi) d > 0, \ \forall d \in \mathcal{C}(x^*) \setminus \{0\}$$

holds. Then, there are constants $\varepsilon > 0$ and $C > 0$ such that the quadratic growth condition is fulfilled at x^*. In particular, x^* is a strict local minimizer.

The above theorem presents the second-order condition associated with S-stationarity. Now we present second-order conditions associated with M-stationarity (d), S-stationarity (d) as well as extended M-stationarity.

Theorem 3.14. [18, Theorem 3.3] *Let x^* be a local minimizer of problem* (3.2) *and $d \in \mathcal{C}(x^*)$. Assume $M(x) := F(x) - \Lambda$ is metrically subregular in direction d at $(x^*, 0)$. Then, there exists an M-stationary (d) multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ associated with x^* such that the second-order condition holds:*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^\Phi) d \geq 0.$$

Theorem 3.15. [18, Theorem 3.17] *Let x^* be a feasible point of problem* (3.2)*. Assume for every nonzero critical direction $0 \neq d \in \mathcal{C}(x^*)$ there exists an S-stationary (d) multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ associated with x^* such that*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^\Phi) d > 0.$$

Then, the quadratic growth condition is fulfilled at x^*. In particular, x^* is a strict local minimizer of problem (3.2)*.*

Theorem 3.16. [18, Theorem 3.21] *Let x^* be an extended M-stationary point of problem* (3.2)*. Assume that for every nonzero critical direction $0 \neq d \in \mathcal{C}(x^*)$ one has*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^\Phi) d > 0$$

for any M-stationary (d) multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T$ associated with x^*. Then, the quadratic growth condition is fulfilled at x^*. In particular, x^* is a strict local minimizer of problem (3.2)*.*

In addition to stationarity conditions we reviewed, there are other stationarity conditions in the literature, such as strong M-stationarity (str. M-stationarity) for MPEC [18, Definition 4.2] and strong M-stationarity in direction d (str. M-stationarity (d)) for MPSC [25, Definition 4.9]. Since these two stationarity conditions are only for special cases and MPDSC does not include MPEC, we do not review them here. In Section 4.2, we will discuss strong M-stationarity as well as strong M-stationarity (d) for CCOP in details.

3.3 Constraint qualifications for MPDSC

We first introduce some background on the generator method which plays an important role in this section. Then, we study constraint qualifications for MPDSC which can be derived from Xu and Ye's work [39] directly and show that RCPLD implies error bounds for MPDSC.

3.3.1 The generator method for MPDSC

Now let us review some background knowledge about the generator method which will be used frequently in the following context. More detailed discussions on it can be found in [39]. We emphasize here the work [39] mainly deals with MPDC (3.1), while we will make some simplifications when considering MPDSC (3.2).

Definition 3.11. *Let A be a set with finitely many linearly independent vectors and D be a subspace. We say A is the generator of D if*

$$D := \mathcal{G}(A) := \text{span}(A).$$

Now we describe the regular normal cone to the disjunctive set $S := \bigcup_{r=1}^{R} S_r$ described in (3.3). For $x^* \in S_r$, let $A_{S_r}(x^*) := \{c_j^r \mid j \in \mathcal{E}_r\}$ denote the generator set of S_r at x^*, that is,

$$N_{S_r}(x^*) = \hat{N}_{S_r}(x^*) = \mathcal{G}(A_{S_r}(x^*)).$$

Since $A_{S_r}(x^*) = \{c_j^r \mid j \in \mathcal{E}_r\}$ is independent of x^*, we have

$$A_{S_r}(x) = A_{S_r}(x^*),\ \forall x, x^* \in S_r.$$

By Theorem 3.2, the regular normal cone to set S at x^* of interest is the intersection of finitely many subspaces, hence it is still a subspace. It follows that the regular normal cone $\hat{N}_S(x^*)$ also can be generated by a set of linearly independent vectors. We denote such set by $\hat{A}_S(x^*)$. Then, we have

$$\hat{N}_S(x^*) = \mathcal{G}(\hat{A}_S(x^*)). \tag{3.29}$$

We call $\hat{A}_S(x^*)$ satisfying (3.29) the generator of the regular normal cone $\hat{N}_S(x^*)$.

However, the limiting normal cone $N_S(x^*)$ usually cannot be generated by a set of linearly independent vectors since the limiting normal cone $N_S(x^*)$ is usually not a subspace from Theorem 3.3. By (3.9) and (3.29) we have

$$N_S(x^*) = \bigcup_{x\in\mathbb{B}_\delta(x^*)} \hat{N}_S(x) = \bigcup_{x\in\mathbb{B}_\delta(x^*)} \mathcal{G}(\hat{A}_S(x)). \tag{3.30}$$

We also define the set

$$A_S(x^*) := \bigcup_{x\in\mathbb{B}_\delta(x^*)} \hat{A}_S(x), \tag{3.31}$$

where $\delta > 0$ is the constant satisfying condition (3.30).

By Theorem 3.3, we know that

$$N_S(x^*) = \bigcup_{r\in I(x^*)} N_{S_r}(x^*). \tag{3.32}$$

Therefore an interesting question is, what is the relationship between generator $A_S(x^*)$ and the union of all $A_{S_r}(x^*)$ where $r \in I(x^*)$. In the following lemma, we show that they are the same. This result is stronger than Lemma 2.1 in [39] where the authors consider the MPDC case.

Lemma 3.1. *Let* $S = \bigcup_{r=1}^{R} S_r \subseteq \mathbb{R}^q$ *where* S_r *is a subspace. Then for any* $x^* \in S$*, we have*

$$A_S(x^*) = \bigcup_{r\in I(x^*)} A_{S_r}(x^*). \tag{3.33}$$

Proof. We will show that

$$A_S(x^*) = \bigcup_{x\in\mathbb{B}_\delta(x^*)} \hat{A}_S(x) = \bigcup_{r\in I(x^*)} A_{S_r}(x^*).$$

The first equality comes from the definition of $A_S(x^*)$. By Theorem 3.2, we know that

$$\hat{N}_S(x) = \bigcap_{r\in I(x)} \hat{N}_{S_r}(x) \tag{3.34}$$

for any $x \in S$, which implies the inclusion

$$\hat{A}_S(x) \subseteq \bigcup_{r\in I(x)} A_{S_r}(x) \tag{3.35}$$

for any $x \in S$.

For any $v \in \bigcup_{x \in \mathbb{B}_\delta(x^*)} \hat{A}_S(x)$, there exists $x' \in \mathbb{B}_\delta(x^*)$ such that

$$v \in \hat{A}_S(x') \subseteq \bigcup_{r \in I(x')} A_{S_r}(x') \subseteq \bigcup_{r \in I(x^*)} A_{S_r}(x^*). \tag{3.36}$$

Indeed, the first inclusion in (3.36) is from (3.35) directly. It is not difficult to see that we can take $\delta > 0$ sufficiently small to guarantee

$$I(x') \subseteq I(x^*). \tag{3.37}$$

Further, by noting that $A_{S_r}(x^*)$ is independent of x^*, hence we have

$$A_{S_r}(x^*) = A_{S_r}(x'). \tag{3.38}$$

By combining (3.37) with (3.38), we justify the second inclusion in (3.36). Therefore,

$$\bigcup_{x \in \mathbb{B}_\delta(x^*)} \hat{A}_S(x) \subseteq \bigcup_{r \in I(x^*)} A_{S_r}(x^*). \tag{3.39}$$

Conversely, for any $v \in \bigcup_{r \in I(x^*)} A_{S_r}(x^*)$, there exists $r \in I(x^*)$ such that

$$v \in A_{S_r}(x^*).$$

Since set S is the union of finitely many subspaces, we can pick a unit vector u_r such that $u_r \in S_r$ but $u_r \notin S_j$ for other $j \neq r$. In this way, for any $0 < t < \delta$ we know that the point $x_r := x^* + tu_r \in \mathbb{B}_\delta(x^*)$ satisfies $x_r \in S_r$ but $x_r \notin S_j$ for other $j \neq r$. Therefore, it follows from (3.34) that

$$\hat{N}_S(x_r) = \hat{N}_{S_r}(x_r),$$

which implies that

$$\hat{A}_S(x_r) = A_{S_r}(x_r). \tag{3.40}$$

From (3.38) and (3.40), we can see

$$v \in A_{S_r}(x^*) = A_{S_r}(x_r) = \hat{A}_S(x_r) \subseteq \bigcup_{x \in \mathbb{B}_\delta(x^*)} \hat{A}_S(x).$$

Then, we have

$$\bigcup_{r\in I(x^*)} A_{S_r}(x^*) \subseteq \bigcup_{x\in\mathbb{B}_\delta(x^*)} \hat{A}_S(x). \tag{3.41}$$

By combing (3.39) with (3.41), we complete the proof. □

3.3.2 Constraint qualifications for MPDSC from MPDC

Now we apply some constraint qualifications tailored for MPDC (see [39, Definition 4.2] and [39, Definition 3.1]) to MPDSC. In the following definition, we also propose a new constraint qualification called the weak linear independence constraint qualification (WLICQ) for MPDSC, which builds a bridge between MPDSC-LICQ (see Definition 3.1) and MPDSC-GMFCQ.

Definition 3.12. *Let $x^* \in \mathcal{F}$ be a feasible point for MPDSC* (3.2). *We say that x^* satisfies*

1. MPDSC-WLICQ *if there is no nonzero vector $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^l \mathbb{R}^q$ such that*

$$\begin{aligned} 0 &= \sum_{i\in\mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^l \nabla\Phi_i(x^*)^T \lambda_i^\Phi, \\ \lambda_i^\Phi &\in N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp,\ i = 1,\ldots,l. \end{aligned}$$

2. MPDSC-GMFCQ *if there is no nonzero vector $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^l \mathbb{R}^q$ such that*

$$\begin{aligned} 0 &= \sum_{i\in\mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^l \nabla\Phi_i(x^*)^T \lambda_i^\Phi, \\ \lambda_i^g &\geq 0,\ i \in \mathcal{I}_g(x^*) \text{ and } \lambda_i^\Phi \in N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp,\ i = 1,\ldots,l. \end{aligned}$$

3. MPDSC-CRCQ *if for every index sets $I \subseteq \mathcal{I}_g(x^*)$, $J \subseteq \{1,\ldots,p\}$, $L \subseteq \{1,\ldots,l\}$ and $\lambda_i^\Phi \in N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp, i \in L$, then the set of vectors*

$$\{\nabla g_i(x^*)\}_{i\in I} \cup \{\nabla h_i(x^*)\}_{i\in J} \cup \bigcup_{\beta_i\in A_i, i\in L} \{\nabla\Phi_i(x^*)^T \beta_i\}$$

and the set of vectors

$$\{\nabla g_i(x^k)\}_{i\in I} \cup \{\nabla h_i(x^k)\}_{i\in J} \cup \bigcup_{\beta_i\in A_i, i\in L} \{\nabla \Phi_i(x^k)^T \beta_i\}$$

have the same rank for k *sufficiently large, for all sequences* $\{x^k\}$ *satisfying* $x^k \to x^*, x^k \neq x^*$ *as* $k \to \infty$ *and any set of linearly independent vectors* A_i *where*

$$A_i \subseteq A_S(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp$$

and $A_i = \emptyset$ *if* $\lambda_i^\Phi = 0$ *for* $i = 1, \ldots, l$.

4. MPDSC-RCRCQ *if the index set* J *is taken as* $\{1, \ldots, p\}$ *in MPDSC-CRCQ.*

5. MPDSC-CPLD *if there exists index sets* $I \subseteq \mathcal{I}_g(x^*)$, $J \subseteq \{1, \ldots, p\}$ *and* $L \subseteq \{1, \ldots, l\}$, *a nonzero vector* $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^l \mathbb{R}^q$ *with* $\lambda_i^g \geq 0, i \in I$ *and* $\lambda_i^\Phi \in N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp, i \in L$ *such that*

$$0 = \sum_{i\in I} \lambda_i^g \nabla g_i(x^*) + \sum_{i\in J} \lambda_i^h \nabla h_i(x^*) + \sum_{i\in L} \nabla \Phi_i(x^*)^T \lambda_i^\Phi,$$

then the set of vectors

$$\{\nabla g_i(x^k)\}_{i\in I} \cup \{\nabla h_i(x^k)\}_{i\in J} \cup \bigcup_{\beta_i\in A_i, i\in L} \{\nabla \Phi_i(x^k)^T \beta_i\}$$

is linearly dependent for k *sufficiently large, for all sequences* $\{x^k\}$ *satisfying* $x^k \to x^*, x^k \neq x^*$ *as* $k \to \infty$ *and any set of linearly independent vectors* A_i *where*

$$A_i \subseteq A_S(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp$$

and $A_i = \emptyset$ *if* $\lambda_i^\Phi = 0$ *for* $i = 1, \ldots, l$.

6. MPDSC-ERCPLD *if the following conditions hold.*

   (i) *The vectors* $\{\nabla h_i(x)\}_{i=1}^p$ *have the same rank for all* $x \in \mathbb{B}_\varepsilon(x^*)$;

(ii) *Let* $J \subseteq \{1,\ldots,p\}$ *be such that the set of vectors* $\{\nabla h_i(x^*)\}_{i\in J}$ *is a basis for* $\mathrm{span}\{\nabla h_i(x^*)\}_{i=1}^p$. *If there exist index sets* $I \subseteq \mathcal{I}_g(x^*)$ *and* $L \subseteq \{1,\ldots,l\}$, $\lambda_i^\Phi \in N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp, i \in L$ *such that the set of vectors*

$$\{\nabla g_i(x^*)\}_{i\in I} \cup \Big\{\{\nabla h_i(x^*)\}_{i\in J} \cup \bigcup_{\beta_i\in A_i, i\in L} \{\nabla\Phi_i(x^*)^T\beta_i\}\Big\}$$

is positive linearly dependent, then the set of vectors

$$\{\nabla g_i(x^k)\}_{i\in I} \cup \{\nabla h_i(x^k)\}_{i\in J} \cup \bigcup_{\beta_i\in A_i, i\in L} \{\nabla\Phi_i(x^k)^T\beta_i\}$$

is linearly dependent for k *sufficiently large, for all sequences* $\{x^k\}$ *satisfying* $x^k \to x^*, x^k \neq x^*$ *as* $k \to \infty$ *and any set of linearly independent vectors* A_i *where*

$$A_i \subseteq A_S(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp$$

and $A_i = \emptyset$ *if* $\lambda_i^\Phi = 0$ *for* $i = 1,\ldots,l$.

7. MPDSC-RCPLD *if the following conditions hold.*

(i) *The vectors* $\{\nabla h_i(x)\}_{i=1}^p$ *have the same rank for all* $x \in \mathbb{B}_\varepsilon(x^*)$;

(ii) *Let* $J \subseteq \{1,\ldots,p\}$ *be such that the set of vectors* $\{\nabla h_i(x^*)\}_{i\in J}$ *is a basis for* $\mathrm{span}\{\nabla h_i(x^*)\}_{i=1}^p$. *If there exist an index set* $I \subseteq \mathcal{I}_g(x^*)$, *a nonzero vector* $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^l \mathbb{R}^q$ *with* $\lambda_i^g \geq 0, i \in I$ *and* $\lambda_i^\Phi \in N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} S_r^\perp$, $i \in \{1,\ldots,l\}$ *such that*

$$0 = \sum_{i\in I} \lambda_i^g \nabla g_i(x^*) + \sum_{i\in J} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^l \nabla\Phi_i(x^*)^T\lambda_i^\Phi, \tag{3.42}$$

then the set of vectors

$$\{\nabla g_i(x^k)\}_{i\in I} \cup \{\nabla h_i(x^k)\}_{i\in J} \cup \bigcup_{\beta_i\in A_i, i\in\{1,\ldots,l\}} \{\nabla\Phi_i(x^k)^T\beta_i\} \tag{3.43}$$

is linearly dependent for k *sufficiently large, for all sequences* $\{x^k\}$ *satis-*

fying $x^k \to x^*, x^k \neq x^*$ as $k \to \infty$ and any set of linearly independent vectors A_i where

$$A_i \subseteq A_S(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq N_S(\Phi_i(x^*)) = \bigcup_{r \in I(\Phi_i(x^*))} S_r^\perp \quad (3.44)$$

and $A_i = \emptyset$ if $\lambda_i^\Phi = 0$ for $i = 1, \ldots, l$.

In fact, we can give a more accurate description of generator sets A_i, $i = 1, \ldots, l$ when we consider MPDSC instead of general MPDC. The following lemma is useful in what follows. We take RCPLD as an example.

Lemma 3.2. *Assume that MPDSC-RCPLD holds at x^*. Given $\lambda_i^\Phi \in N_S(\Phi_i(x^*))$, for any set of linearly independent vectors A_i, $i = 1, \ldots, l$ such that*

$$A_i \subseteq A_S(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq N_S(\Phi_i(x^*)),$$

there must exist $r \in I(\Phi_i(x^*))$ such that

$$A_i \subseteq A_{S_r}(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq \mathcal{G}(A_{S_r}(\Phi_i(x^*))). \quad (3.45)$$

Proof. Suppose on the contrary that there does not exist $r \in I(\Phi_i(x^*))$ satisfying (3.45). Without loss of generality, we assume there are $r_1, r_2 \in I(\Phi_i(x^*))$ such that

$$A_i \subseteq A_{S_{r_1}}(\Phi_i(x^*)) \cup A_{S_{r_2}}(\Phi_i(x^*)) \text{ but } A_i \not\subseteq A_{S_{r_1}}(\Phi_i(x^*)) \text{ and } A_i \not\subseteq A_{S_{r_2}}(\Phi_i(x^*)).$$

Since both S_{r_1} and S_{r_2} are subspaces, it follows that

$$\mathcal{G}(A_i) \subseteq \text{span}\{S_{r_1}^\perp \cup S_{r_2}^\perp\} \text{ but } \mathcal{G}(A_i) \not\subseteq S_{r_1}^\perp \text{ and } \mathcal{G}(A_i) \not\subseteq S_{r_2}^\perp,$$

which contradicts the fact that

$$\mathcal{G}(A_i) \subseteq N_S(\Phi_i(x^*)) = \bigcup_{r \in I(\Phi_i(x^*))} S_r^\perp.$$

In this way, we complete the proof. □

Like nonlinear programs (2.1), we also have the concept of error bounds for MPDSC. Note that we should describe the deviation of the generalized equations

$\Phi_i(x) \in S$ for $i = 1, \ldots, l$ in terms of distance functions. The following definition can be obtained from [39, Definition 2.2] directly.

Definition 3.13. (Error bounds for MPDSC) *We say that a feasible point* $x^* \in \mathcal{F}$ *of problem* (3.2) *satisfies the error bound property if there exists* $\alpha \geq 0$ *and* $\varepsilon > 0$ *such that*

$$d_{\mathcal{F}}(x) \leq \alpha\Big(\sum_{i=1}^{m} \max\{g_i(x), 0\} + \sum_{i=1}^{p} |h_i(x)| + \sum_{i=1}^{l} d_S(\Phi_i(x))\Big), \ \forall x \in \mathbb{B}_\varepsilon(x^*).$$

3.3.3 RCPLD as a sufficient condition for error bounds

Unlike nonlinear programs, we usually do not have the conclusion that RCPLD implies error bounds for MPDC. To deal with this issue, Xu and Ye introduced MPDC-piecewise RCPLD as a sufficient condition for error bounds [39, Theorem 4.2]. In this subsection, we will show that for MPDSC many constraint qualifications such as CRCQ, RCRCQ, CPLD, ERCPLD, and RCPLD coincide with their piecewise versions, respectively. In particular, since RCPLD coincides with piecewise RCPLD, it follows that RCPLD implies error bounds for MPDSC.

We emphasize that in this thesis we only focus on the discussions on RCPLD and piecewise RCPLD for simplicity, the discussions on other constraint qualifications as well as their piecewise versions are similar. Let x^* be feasible for MPDSC (3.2). Let sets $P_1, \ldots, P_R$ (sometimes some of them may be empty) be a partition of $\{1, \ldots, l\}$. We denote such partition by $P := \{P_1, \ldots, P_R\}$ and consider the subsystem for the partition P:

$$\begin{cases} g(x) \leq 0, \ h(x) = 0, \\ \Phi_i(x) \in S_1, \ i \in P_1, \\ \qquad \vdots \\ \Phi_i(x) \in S_R, \ i \in P_R. \end{cases} \tag{3.46}$$

We denote the feasible region of subsystem (3.46) by $\mathcal{F}_P$. Since each partition $P = \{P_1, \ldots, P_R\}$ is one of the possible partitions of $\{1, \ldots, l\}$, we have $\mathcal{F}_P \subseteq \mathcal{F}$. Conversely, for any $x^* \in \mathcal{F}$, there exists one partition P such that $x^* \in \mathcal{F}_P$.

Definition 3.14. (MPDSC-PRCPLD) [39, Definition 4.1] *We say that the piecewise RCPLD holds for MPDSC* (3.2) *at* $x^* \in \mathcal{F}$*, if MPDSC-RCPLD holds for subsystem*

(3.46) *for any partition* $P = \{P_1, \ldots, P_R\}$ *such that* $x^* \in \mathcal{F}_P$. *That is, the following conditions hold for any partition* $P = \{P_1, \ldots, P_R\}$ *such that* $x^* \in \mathcal{F}_P$.

(i) *The vectors* $\{\nabla h_i(x)\}_{i=1}^p$ *have the same rank for all* $x \in \mathbb{B}_\varepsilon(x^*)$;

(ii) *Let* $J \subseteq \{1, \ldots, p\}$ *be such that the set of vectors* $\{\nabla h_i(x^*)\}_{i \in J}$ *is a basis for* $\text{span}\{\nabla h_i(x^*)\}_{i=1}^p$. *If there exist index sets* $I \subseteq \mathcal{I}_g(x^*)$, *a nonzero vector* $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^l \mathbb{R}^q$ *with* $\lambda_i^g \geq 0, i \in I$ *and* $\lambda_i^\Phi \in N_{S_r}(\Phi_i(x^*)) = S_r^\perp$ *for* $i \in P_r$, $r = 1, \ldots, R$ *such that*

$$0 = \sum_{i \in I} \lambda_i^g \nabla g_i(x^*) + \sum_{i \in J} \lambda_i^h \nabla h_i(x^*) + \sum_{i \in P_1} \nabla \Phi_i(x^*)^T \lambda_i^\Phi + \cdots + \sum_{i \in P_R} \nabla \Phi_i(x^*)^T \lambda_i^\Phi, \tag{3.47}$$

then the set of vectors

$$\{\nabla g_i(x^k)\}_{i \in I} \cup \{\nabla h_i(x^k)\}_{i \in J} \cup \bigcup_{r=1,\ldots,R} \bigcup_{\beta_i^r \in A_i^r, i \in P_r} \{\nabla \Phi_i(x^k)^T \beta_i^r\} \tag{3.48}$$

is linearly dependent for k *sufficiently large, for all sequences* $\{x^k\}$ *satisfying* $x^k \to x^*, x^k \neq x^*$ *as* $k \to \infty$ *and any set of linearly independent vectors* A_i^r *where*

$$A_i^r \subseteq A_{S_r}(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i^r) \subseteq N_{S_r}(\Phi_i(x^*)) = S_r^\perp \tag{3.49}$$

and $A_i^r = \emptyset$ *if* $\lambda_i^\Phi = 0$ *for* $i \in P_r$, $r = 1, \ldots, R$.

The following example is useful for us to understand piecewise RCPLD.

Example 3.1. *Let us consider the following cardinality constrained system in* $\mathbb{R}^3$:

$$g(x) \leq 0, h(x) = 0, \|x\|_0 \leq 2. \tag{3.50}$$

From Figure 1.1 *we know that problem* (3.50) *is equivalent to*

$$g(x) \leq 0, h(x) = 0, \Phi(x) := x \in S,$$

where $S = S_1 \cup S_2 \cup S_3$ *such that* $S_1 = \{0\} \times \mathbb{R} \times \mathbb{R}$, $S_2 = \mathbb{R} \times \{0\} \times \mathbb{R}$ *and* $S_3 = \mathbb{R} \times \mathbb{R} \times \{0\}$.

Assume that point $x^* = (0, 0, 1)^T$ *is feasible for system* (3.50). *It is easy to see that* $l = 1$, $R = 3$, $x^* \in S_1$ *and* $x^* \in S_2$. *There are three partitions of* $\{1\}$ *into sets* $P = \{P_1, P_2, P_3\}$: (i) $P_1 = \{1\}, P_2 = \emptyset, P_3 = \emptyset$; (ii) $P_1 = \emptyset, P_2 = \{1\}, P_3 = \emptyset$;

(iii) $P_1 = \emptyset, P_2 = \emptyset, P_3 = \{1\}$. *Since* $x^* \notin S_3$, *we only have two possible subsystems. Therefore, piecewise RCPLD holds for system* (3.50) *if RCPLD holds for each of the following two subsystems:*

$$(\mathcal{P}_1)\begin{cases} g(x) \leq 0, \\ h(x) = 0, \\ x \in S_1. \end{cases} \qquad \text{and} \qquad (\mathcal{P}_2)\begin{cases} g(x) \leq 0, \\ h(x) = 0, \\ x \in S_2. \end{cases}$$

That is, we say that $x^* = (0,0,1)^T$ *satisfies piecewise RCPLD for system* (3.50) *if the following conditions hold:*

(i) *The vectors* $\{\nabla h_i(x)\}_{i=1}^p$ *have the same rank for all* $x \in \mathbb{B}_\varepsilon(x^*)$*;*

(ii) *Let* $J \subseteq \{1,\ldots,p\}$ *be such that the set of vectors* $\{\nabla h_i(x^*)\}_{i\in J}$ *is a basis for* $\text{span}\{\nabla h_i(x^*)\}_{i=1}^p$. *If there exist index sets* $I \subseteq \mathcal{I}_g(x^*)$, *a nonzero vector* $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}$ *with* $\lambda_i^g \geq 0, i \in I$ *such that*

$$0 = \sum_{i\in I} \lambda_i^g \nabla g_i(x^*) + \sum_{i\in J} \lambda_i^h \nabla h_i(x^*) + \lambda^I e,$$

then the set of vectors

$$\{\nabla g_i(x^k)\}_{i\in I} \cup \{\nabla h_i(x^k)\}_{i\in J} \cup \{e\}$$

is linearly dependent for k *sufficiently large, for all sequences* $\{x^k\}$ *satisfying* $x^k \to x^*, x^k \neq x^*$ *as* $k \to \infty$, *where the vector* e *can be taken as* $e_1 = (1,0,0)^T$ *or* $e_2 = (0,1,0)^T$.

If we write down RCPLD for system (3.50), we can see RCPLD and picewise RCPLD are the same. Now we show that MPDSC-RCPLD coincides with MPDSC-piecewise RCPLD. As we can see as follows, the following two equalities from Thereom 3.3 and Lemma 3.1 play key roles in the proof of Theorem 3.17:

$$N_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} N_{S_r}(\Phi_i(x^*)), \tag{3.51}$$

$$A_S(\Phi_i(x^*)) = \bigcup_{r\in I(\Phi_i(x^*))} A_{S_r}(\Phi_i(x^*)). \tag{3.52}$$

Theorem 3.17. *For mathematical programs with disjunctive subspaces constraints* (MPDSC)*, we have RCPLD coincides with piecewise RCPLD.*

Proof. If $R = 1$, set $S = \bigcup_{r=1}^{R} S_r$ is a subspace itself, hence the conclusion is obvious. Let us assume that $R \geq 2$.

Assume that MPDSC-RCPLD holds at x^*. Given any possible partition $P = \{P_1, \ldots, P_R\}$ of $\{1, \ldots, l\}$ such that $x^* \in \mathcal{F}_P$. We only need to prove that MPDSC-PRCPLD (ii) holds at x^*.

Let $J \subseteq \{1, \ldots, p\}$ be such that the set of vectors $\{\nabla h_i(x^*)\}_{i \in J}$ is a basis for $\text{span}\{\nabla h_i(x^*)\}_{i=1}^{p}$. Suppose there exists index sets $I \subseteq \mathcal{I}_g(x^*)$, $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^{l} \mathbb{R}^q$ with $\lambda_i^g \geq 0, i \in I$ and $\lambda_i^\Phi \in N_{S_r}(\Phi_i(x^*))$ for $i \in P_r$, $r = 1, \ldots, R$ such that (3.47) holds. We want to prove the linear dependence of (3.48) for k sufficiently large, for all sequences $\{x^k\}$ satisfying $x^k \to x^*$, $x^k \neq x^*$ as $k \to \infty$ and any A_i^r satisfying (3.49). By (3.51), we have

$$\lambda_i^\Phi \in N_{S_r}(\Phi_i(x^*)) \subseteq N_S(\Phi_i(x^*)),\ i = 1, \ldots, l. \tag{3.53}$$

In this way, we can see (3.47) implies (3.42). According to MPDSC-RCPLD (ii), the set of vectors of (3.43) is linearly dependent for k sufficiently large, for all sequences $\{x^k\}$ satisfying $x^k \to x^*, x^k \neq x^*$ and any A_i satisfying (3.44). We know from (3.52) that for $r \in I(\Phi_i(x^*))$, $A_{S_r}(\Phi_i(x^*))$ is a subset of $A_S(\Phi_i(x^*))$. Since $A_i \subseteq A_S(\Phi_i(x^*))$ is arbitrary in (3.44) and now we have $\lambda_i^\Phi \in N_{S_r}(\Phi_i(x^*))$, we can take A_i such that

$$A_i \subseteq A_{S_r}(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq N_{S_r}(\Phi_i(x^*)),\ i \in P_r. \tag{3.54}$$

We define $A_i^r := A_i$ for $i \in P_r$ where A_i satisfies (3.54), which means that the linear dependence of (3.43) implies that of (3.48). Therefore, MPDSC-RCPLD implies MPDSC-PRCPLD.

Conversely, we suppose MPDSC-PRCPLD holds at x^*. Now we are going to show that MPDSC-RCPLD (ii) holds at x^*.

Let $J \subseteq \{1, \ldots, p\}$ be such that the set of vectors $\{\nabla h_i(x^*)\}_{i \in J}$ is a basis for $\text{span}\{\nabla h_i(x^*)\}_{i=1}^{p}$. Suppose there exists an index set $I \subseteq \mathcal{I}_g(x^*)$, $\lambda = (\lambda^g, \lambda^h, \lambda^\Phi)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \prod_{i=1}^{l} \mathbb{R}^q$ with $\lambda_i^g \geq 0, i \in I$ and $\lambda_i^\Phi \in N_S(\Phi_i(x^*)), i \in \{1, \ldots, l\}$ such that (3.42) holds. We want to prove the linear dependence of (3.43) for all sequences $\{x^k\}$ satisfying $x^k \to x^*$, $x^k \neq x^*$ as $k \to \infty$ and any A_i satisfying (3.44). By Lemma 3.2, for any A_i satisfying (3.44) there must exist $r \in I(\Phi_i(x^*))$ such that

$$A_i \subseteq A_{S_r}(\Phi_i(x^*)), 0 \neq \lambda_i^\Phi \in \mathcal{G}(A_i) \subseteq \mathcal{G}(A_{S_r}(\Phi_i(x^*))). \tag{3.55}$$

From (3.55), we know that now we also have

$$\lambda_i^\Phi \in N_{S_r}(\Phi_i(x^*)). \tag{3.56}$$

For each $i \in \{1, \ldots, l\}$, we associate with a specific r such that (3.55) and (3.56) hold and define $A_i^r := A_i$. In this way, we construct a partition $P = \{P_1, \ldots, P_R\}$ of $\{1, \ldots, l\}$ and (3.42) implies (3.47).

According to MPDSC-PRCPLD (ii), the set of vectors of (3.48) is linearly dependent for k sufficiently large, for all sequences $\{x^k\}$ satisfying $x^k \to x^*, x^k \neq x^*$ and any set of linearly independent vectors A_i^r satisfying (3.49). Since $A_i^r = A_i$ and $P = \{P_1, \ldots, P_R\}$ is a partition of $\{1, \ldots, l\}$, the linear dependence of (3.48) implies that of (3.43). Therefore, MPDSC-RCPLD implies MPDSC-PRCPLD, which completes the proof. □

The discussions on other constraint qualifications are similar as that of RCPLD, hence we have the following corollary.

Corollary 3.1. *For mathematical programs with disjunctive subspaces constraints* (MPDSC)*, their constraint qualifications such as CRCQ, RCRCQ, CPLD, ERCPLD, and RCPLD coincide with their piecewise versions, respectively.*

Based on Theorem 3.17, now we show that for MPDSC the constraint qualification RCPLD implies the error bound property.

Lemma 3.3. [39, Theorem 4.2] *Suppose that MPDSC-piecewise RCPLD holds at x^* which is feasible for problem* (3.2)*. Then, the error bound property holds at x^*.*

Theorem 3.18. *Suppose that MPDSC-RCPLD holds at x^* which is feasible for problem* (3.2)*. Then, the error bound property holds at x^*.*

Proof. The proof is rather straightforward by combining Theorem 3.17 and Lemma 3.3. The constraint qualification MPDSC-RCPLD implies the error bound property since MPDSC-RCPLD coincides with MPDSC-piecewise RCPLD. □

We conclude Chapter 3 with Figure 3.1 which summarizes the relations among various constraint qualifications for MPDSC.

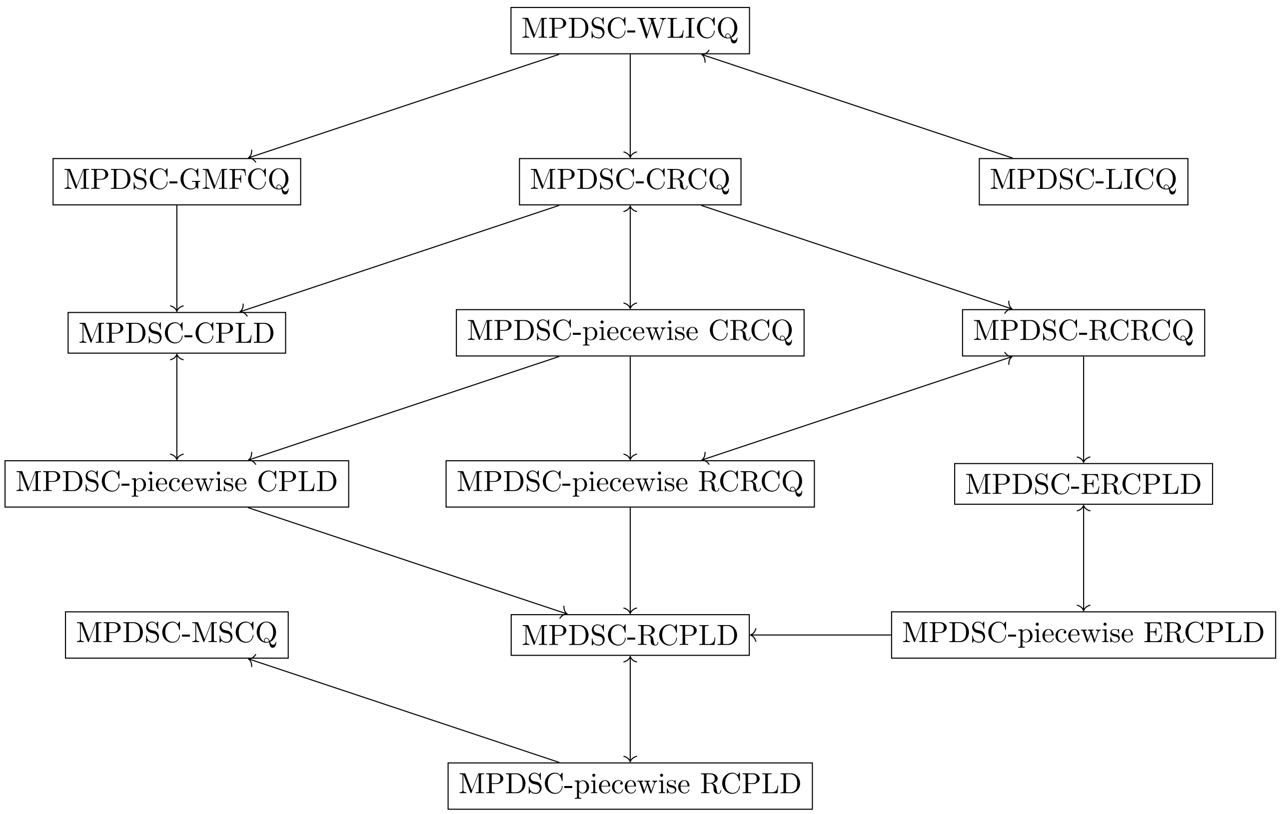

Figure 3.1: Relations among constraint qualifications for MPDSC

4 Optimality Conditions for CCOP from MPDSC

In Chapter 4, we will apply the results of MPDSC from Chapter 3 to CCOP. In Section 4.1, we give a disjunctive subspaces reformulation of CCOP, which is different from the relaxed complementarity-type reformulation (1.2). In Section 4.2, we survey various stationarity conditions for CCOP. In particular, we recover disjunctive-type S-stationarity and M-stationarity for CCOP in [34], and then make comparisons between them and those from the relaxed complementarity-type reformulation. In Section 4.3, in addition to showing that many constraint qualifications for CCOP coincide with their piecewise versions, we also make comparisons between these disjunctive-type constraint qualifications and those from the relaxed complementarity-type reformulation.

4.1 Disjunctive subspaces reformulation for CCOP

Motivated by the two simple examples given in the introduction (see Figure 1.1), we abbreviate $\mathcal{I}_s := \{I \in 2^{\{1,2,\ldots,n\}} \mid |I| = s\}$ and $\mathbb{R}_I := \text{span}\{e_i \mid i \in I\}$ for a given set I, where e_i is the unit vector with the ith component equals to one. Now we reformulate CCOP (1.1) as MPDSC as follows:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, x \in S, \end{aligned} \tag{4.1}$$

where $S := \bigcup_{I \in \mathcal{I}_s} \mathbb{R}_I$ is the union of finitely many subspaces which are orthogonal to each other. In fact, problem (4.1) is a special case of MPDSC (3.2) with $l = 1$ and $\Phi(x) = I(x)$ where $I(x)$ is the identity mapping.

We consider a special case of (4.1). For a feasible point $x^* \in \mathbb{R}^n$ such that $\|x^*\|_0 = s$, the point x^* only locates in one specific subspace $\mathbb{R}_I$ ($I \in \mathcal{I}_s$). Therefore, the following simpler reformulation without disjunctive subspaces structure is locally equivalent to CCOP:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0, h(x) = 0, x_i = 0 \ (i \in I_0(x^*)), \end{aligned} \tag{4.2}$$

where $I_0(x^*) := \{i \in \{1, \ldots, n\} \,|\, x_i^* = 0\}$. However, such reformulation does not apply in the case $\|x^*\|_0 < s$. We follow the convention in [15] and call problem (4.2) the tightened nonlinear program (TNLP) at x^* for CCOP.

4.2 Stationarity conditions for CCOP

4.2.1 Disjunctive-type S-stationarity and M-stationarity for CCOP

The following theorem is immediate from Theorems 3.1-3.5. In what follows, we use the notation $I_\pm(x) := \{i \in \{1, \ldots, n\} \mid x_i \neq 0\}$ for a given vector $x \in \mathbb{R}^n$.

Theorem 4.1. *Let x^* be a feasible point of program* (4.1) *and $d \in \mathbb{R}^n$. Then, we summarize various cones in two cases in* Table 4.1.

Combining Definition 3.4 and Theorem 4.1, now we discuss disjunctive-type S-stationarity and M-stationarity for CCOP (4.1). The multipliers λ^g and λ^h are same as those in Definition 3.4. Now we focus on the discussion about multiplier λ^I.

- (Disjunctive-type S-stationarity for CCOP) Let $\lambda^I \in \hat{N}_S(x^*)$. From Theorem 4.1 we have

  (a1) if $\|x^*\|_0 = s$, then $\lambda_i^I = 0$, $\forall i \in I_\pm(x^*)$;

  (a2) if $\|x^*\|_0 < s$, then $\lambda_i^I = 0$, $\forall i \in \{1, \ldots, n\}$.

Cones	$\lvert I_\pm(x^*)\rvert < s$ (or $\lvert I_\pm(x^*) \cup I_\pm(d)\rvert < s$)	$\lvert I_\pm(x^*)\rvert = s$ (or $\lvert I_\pm(x^*) \cup I_\pm(d)\rvert = s$)
$T_S(x^*)$	$\bigcup_{I_\pm(x^*)\subseteq\gamma\in\mathcal{I}_s} \mathbb{R}_\gamma$	$\mathbb{R}_{I_\pm(x^*)}$
$\hat{N}_S(x^*)$	$\{0\}$	$\mathbb{R}^\perp_{I_\pm(x^*)}$
$N_S(x^*)$	$\bigcup_{I_\pm(x^*)\subseteq\gamma\in\mathcal{I}_s} \mathbb{R}^\perp_\gamma$	$\mathbb{R}^\perp_{I_\pm(x^*)}$
$N_S(x^*; d)$	$\bigcup_{I_\pm(x^*)\cup I_\pm(d)\subseteq\gamma\in\mathcal{I}_s} \mathbb{R}^\perp_\gamma$	$\mathbb{R}^\perp_{I_\pm(x^*)\cup I_\pm(d)}$
$\hat{N}_{T_S(x^*)}(d)$	$\{0\}$	$\mathbb{R}^\perp_{I_\pm(x^*)\cup I_\pm(d)}$

Table 4.1: Various cones to set S in CCOP

- (Disjunctive-type M-stationarity for CCOP) Let $\lambda^I \in N_S(x^*)$. From Theorem 4.1 we have

  (b1) if $\|x^*\|_0 = s$, then $\lambda_i^I = 0,\ \forall i \in I_\pm(x^*)$;

  (b2) if $\|x^*\|_0 < s$, then $\lambda_i^I = 0,\ \forall i \in I_\pm(x^*)$ and $\|\lambda^I\|_0 \le n - s$.

  Therefore, for any $\lambda^I \in N_S(x^*)$ we have

$$\lambda_i^I = 0,\ \forall i \in I_\pm(x^*) \ \text{ and } \ \|\lambda^I\|_0 \le n - s.$$

By the discussion above, we have the following definitions.

Definition 4.1. (Disjunctive-type S-staionarity and M-stationarity) *Let $x^* \in \mathbb{R}^n$ be a feasible point of CCOP* (4.1)*.*

(i) [28, Section 5.3] *We say x^* is a S-stationary point if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ satisfying*

$$\begin{gathered}\nabla f(x^*) + \sum_{i\in\mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{n} \lambda_i^I e_i = 0,\\ \lambda_i^g \ge 0, \forall i \in \mathcal{I}_g(x^*),\end{gathered} \tag{4.3}$$

such that

(a) *if $\|x^*\|_0 = s$, then $\lambda_i^I = 0,\ \forall i \in I_\pm(x^*)$;*

(b) if $\|x^*\|_0 < s$, *then* $\lambda_i^I = 0,\ \forall i \in \{1, \dots, n\}$.

(ii) *We say* x^* *is a M-stationary point if there exists* $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ *satisfying* (4.3) *such that*

$$\lambda_i^I = 0,\ \forall i \in I_{\pm}(x^*) \text{ and } \|\lambda^I\|_0 \le n - s.$$

For the purpose of comparison, we present the definitions of S-stationarity and M-stationarity obtained from the relaxed complementarity-type reformulation (1.2) as follows.

Definition 4.2. (Complementarity-type S-staionarity and M-stationarity) [14, Defintion 4.6] *Let* (x^*, y^*) *be a feasible point of the problem* (1.2).

(i) *We say* (x^*, y^*) *is a S-stationary point if there exists* $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ *satisfying* (4.3) *such that*

$$\lambda_i^I = 0,\ \forall i \in I_0(y^*),$$

where $I_0(y^*) := \{i \in \{1, \dots, n\} \mid y_i^* = 0\}$.

(ii) *We say* (x^*, y^*) *is a M-stationary point if there exists* $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ *satisfying* (4.3) *such that*

$$\lambda_i^I = 0,\ \forall i \in I_{\pm}(x^*).$$

Remark 4.1. *From* Definition 4.2, *we can see that* (x^*, y^*) *feasible for CCOP is a complementarity-type M-stationary point if and only if* x^* *is a KKT point of the tightened nonlinear program* (4.2).

By comparing Definition 4.1 and Definition 4.2, we now reveal the relationship between our disjunctive-type stationarities and complementarity-type ones. We would like to point out here, Mehlitz was the first to show that S-staionarity from disjunctive reformulation is more restrictive in some sense [28, Section 5.3], while the discussion on different M-stationarities leave blank in his paper.

Theorem 4.2. *Let* x^* *be any feasible point for CCOP in the disjunctive form* (4.1). *We define* $y^* \in \mathbb{R}^n$ *as follows:*

$$y_i^* = \begin{cases} 0, & \text{if } i \in I_{\pm}(x^*); \\ 1, & \text{if } i \in I_0(x^*). \end{cases} \tag{4.4}$$

Then, (x^*, y^*) *is a feasible point of problem* (1.2)*. Further, if* x^* *is a disjunctive-type S-stationarity or M-stationarity of problem* (4.1)*, then* (x^*, y^*) *is a complementarity-type S-stationarity or M-stationarity of problem* (1.2).

Proof. We first show that (x^*, y^*) is feasible for problem (1.2). The complementary condition $x_i y_i = 0, i = 1, \ldots, n$ is obvious from (4.4). Now we verify the condition $e^T y \geq n - s$. Since $\|x^*\|_0 \leq s$ we have $|I_\pm(x^*)| \leq s$, which implies $|I_0(x^*)| \geq n - s$. From (4.4), we know $I_0(x^*) = I_\pm(y^*)$. Therefore, we have $|I_\pm(y^*)| \geq n - s$ hence $e^T y \geq n - s$, which shows that (x^*, y^*) is feasible.

The implication from disjunctive-type M-stationarity to complementarity-type M-stationarity is obvious. Now we focus on the implication of different S-stationarities. Assume that x^* is disjunctive-type S-stationary. If $\|x^*\|_0 = s$, disjunctive-type S-stationarity and complementarity-type S-stationarity are the same since we have $I_\pm(x^*) = I_0(y^*)$. If $\|x^*\|_0 < s$, disjunctive-type S-stationarity must be complementarity-type S-stationary since $\{1, \ldots, n\} \supseteq I_0(y^*)$ always holds. Combing two cases we discussed above, we complete the proof.

□

The reverse implication of above theorem do not hold when $\|x^*\|_0 < s$, that is, we cannot say that x^* is a disjunctive-type S-stationarity or M-stationarity of problem (4.1) if (x^*, y^*) is a complementarity-type S-stationarity or M-stationarity of problem (1.2). Let us look at two examples.

Example 4.1. (S-stationarity) *Consider the following cardinality constrained optimization problem in* $\mathbb{R}^2$*:*

$$\begin{aligned} \min_x \quad & f(x) = x_2 - x_1 \\ s.t. \quad & g(x) = x_1^2 + (x_2 - 1)^2 - 1 \leq 0, \ \|x\|_0 \leq 1. \end{aligned}$$

The unique global minimizer of this problem is $x^* = (0,0)^T$*. Take* $y^* = (1,0)^T$ *feasible for the relaxed complementarity-type reformulation* (1.2)*, now we show that* (x^*, y^*) *is a complementarity-type S-stationarity. It is easy to see* $\nabla f(x^*) = (-1,1)^T$*,* $\nabla g(x^*) = (0,-2)^T$*,* $e_1 = (1,0)^T$ *and* $e_2 = (0,1)^T$*. In this way,* (4.3) *reduces to*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda^g \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \lambda_1^I \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda_2^I \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Hence we choose $\lambda^g = \frac{1}{2}$, $\lambda_1^I = 1$ *and* $\lambda_2^I = 0$ *such that* Definition 4.2 (i) *holds at* (x^*, y^*). *Therefore,* (x^*, y^*) *is S-stationary in the sense of complementarity.*

However, x^* *is not a disjunctive-type S-stationarity. Since* $\|x^*\|_0 < 1$ *we have* $\lambda_1^I = 0$ *and* $\lambda_2^I = 0$ *from* Definition 4.1 (i). *We do not have suitable* λ^g *to satisfy* (4.3) *since the vectors* $\nabla f(x^*) = (-1, 1)^T$ *and* $\nabla g(x^*) = (0, -2)^T$ *are linearly independent, which shows the desired conclusion.*

Further, by Theorem 3.6 (ii), *LICQ is violated at* x^* *for this program. Now we verify this statement. By* [28, Section 5.3], *when we apply MPDSC-LICQ* (see Definition 3.1) *to CCOP* (1.1), *it is equivalent to say the following vectors*

$$\{\nabla g_i(x^*)\}_{i \in \mathcal{I}_g(x^*)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i \in I_0(x^*)} \tag{4.5}$$

are linearly independent. However, in our example, the vectors $\{\nabla g(x^*)\} \cup \{e_1\} \cup \{e_2\}$ *are linearly dependent, which means LICQ does not hold at* x^*.

Example 4.2. (M-stationarity) *Consider the following cardinality constrained optimization problem in* $\mathbb{R}^2$*:*

$$\begin{aligned} \min_x \quad & f(x) = x_2 - x_1 \\ s.t. \quad & g(x) = x_1^2 + x_2^2 \leq 0, \ \|x\|_0 \leq 1. \end{aligned}$$

The only feasible point of this problem is $x^* = (0, 0)^T$ *therefore the unique solution must be* x^*. *Take any feasible point* y^* *for the relaxed complementarity-type reformulation* (1.2), *we show that* (x^*, y^*) *is a complementarity-type M-stationarity. It is easy to see* $\nabla f(x^*) = (-1, 1)^T$, $\nabla g(x^*) = (0, 0)^T$, $e_1 = (1, 0)^T$ *and* $e_2 = (0, 1)^T$*.In this way,* (4.3) *reduces to*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_1^I \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda_2^I \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Hence we choose $\lambda^g = 1$, $\lambda_1^I = 1$ *and* $\lambda_2^I = -1$ *such that* Definition 4.2 (ii) *holds at* (x^*, y^*). *Therefore,* (x^*, y^*) *is M-stationary in the sense of complementarity.*

However, x^* *is not a disjunctive-type M-stationarity. Since* $\|x^*\|_0 < 1$ *we have* $\lambda_1^I = 0$ *or* $\lambda_2^I = 0$ *from* Definition 4.1 (ii). *We do not have suitable* λ^I *to satisfy* (4.3), *which shows the desired conclusion.*

Further, by Theorem 3.6 (iii), *GGCQ is violated at* x^* *for this program. Now we verify this statement. On the one hand, it is easy to see* $\mathcal{F} = \{0\}$ *and* $x^* = 0$, *hence*

$\hat{N}_{\mathcal{F}}(x^*) = \mathbb{R}^2$. *On the other hand, by* Definition 3.2 *we know the linearization cone of this example takes the form*

$$L^{lin}_{\mathcal{F}}(x^*) = (\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R}),$$

hence $(L^{lin}_{\mathcal{F}}(x^*))^\circ = \{0\}$. *Therefore, we conclude* $\hat{N}_{\mathcal{F}}(x^*) \neq (L^{lin}_{\mathcal{F}}(x^*))^\circ$ *and GGCQ fails to hold at* x^*.

4.2.2 Other stationarity conditions for CCOP

In this subsection, we will survey other stationarity conditions from MPDSC for CCOP. Moreover, as we said at the end of Section 3.2, we will discuss strong M-stationarity (str. M-stationarity) and strong M-stationarity in the direction d (str. M-stationarity (d)) for CCOP in details.

The following corollary which says that $\mathcal{Q}$-stationarity coincides with $\mathcal{Q}_M$-stationarity for CCOP is obvious from Theorem 3.8.

Corollary 4.1. *Let* x^* *be feasible for CCOP* (1.1). *Suppose* x^* *is Q-stationary, then* x^* *is also M-stationary. In other words,* $\mathcal{Q}$*-stationarity coincides with* $\mathcal{Q}_M$*-stationarity for CCOP.*

Applying Definition 3.6, we obtain AM-stationary condition for CCOP. Let $x^* \in \mathcal{F}$ feasible for problem (4.1), AM-stationary condition for CCOP holds at x^* if there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\varepsilon^k\} \subseteq \mathbb{R}^n$, $\{(y^{g,k}, y^{h,k}, y^{I,k})\} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n$ with $\varepsilon^k \to 0$, $x^k \to x^*$, $(y^{g,k}, y^{h,k}, y^{I,k}) \to 0$ such that $x^k - y^{I,k} \in S$ and

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^{p} \lambda_i^{h,k} \nabla h_i(x^k) + \sum_{i=1}^{n} \lambda_i^{I,k} e_i = \varepsilon^k, \tag{4.6}$$

the multipliers $\{(\lambda_i^{g,k}, \lambda_i^{h,k}, \lambda_i^{I,k})\} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n$ should satisfy

$$\lambda_i^{g,k} = 0, \text{if } g_i(x^k) - y_i^{g,k} < 0;\ \lambda_i^{g,k} \geq 0, \text{if } g_i(x^k) - y_i^{g,k} = 0, \tag{4.7}$$

$$\lambda_i^{I,k} = 0, \text{if } x_i^k - y_i^{I,k} \neq 0 \ \text{ and } \ \|\lambda^{I,k}\|_0 \leq n - s. \tag{4.8}$$

We now show that in fact the sequences $\{\varepsilon^k\}$ and $\{y^k\}$ in (4.6) to (4.8) can be removed.

(a) For $i \notin \mathcal{I}_g(x^*)$, we have $g_i(x^k) < 0$ for k sufficiently large, hence $g_i(x^k) - y_i^{g,k} < 0$ for k sufficiently large. By (4.7), we have $\lim_{k\to\infty} \min\{\lambda_i^{g,k}, -g_i(x^k)\} = 0$.

(b) For $i \in \mathcal{I}_g(x^*)$, since $\min\{\lambda_i^{g,k}, -g_i(x^k)\} \le -g_i(x^k)$ and $g_i(x^k) \to 0$, we also have $\lim_{k\to\infty} \min\{\lambda_i^{g,k}, -g_i(x^k)\} = 0$.

(c) For $i \in I_\pm(x^*)$, we have $x_i^* \neq 0$, which implies $x_i^k \neq 0$ for k sufficiently large. Then, it follows that $x_i^k - y_i^{I,k} \neq 0$ for k sufficiently large, hence we have $\lambda_i^{I,k} = 0$ and $\|\lambda^{I,k}\|_0 \le n - s$ from (4.8).

By the above discussion, we obtain the following AM-stationarity for CCOP.

Definition 4.3. *Let $x^* \in \mathcal{F}$ be feasible for CCOP* (4.1). *We say that x^* is AM-stationary for CCOP if and only if there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$ with $\lim_{k\to\infty} x^k = x^*$ and multipliers $\{(\lambda_i^{g,k}, \lambda_i^{h,k}, \lambda_i^{I,k})\} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n$ such that*

$$
\begin{aligned}
&\lim_{k\to\infty} \|\nabla f(x^k) + \sum_{i=1}^m \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^p \lambda_i^{h,k} \nabla h_i(x^k) + \sum_{i=1}^n \lambda_i^{I,k} e_i\| = 0,\\
&\lim_{k\to\infty} \min\{\lambda_i^{g,k}, -g_i(x^k)\} = 0,\\
&\lambda_i^{I,k} = 0,\ \forall i \in I_\pm(x^*) \ \text{ and } \ \|\lambda^{I,k}\|_0 \le n - s \ \ (k \to \infty).
\end{aligned}
$$

Remark 4.2. *The above AM-stationary condition for CCOP from disjunctive subspaces reformulation is sharper than that obtained from the relaxed complementarity-type reformulation* (see [24, Definition 3.1]) *like M-stationarity. The only difference between them is for the latter, the cardinality constraint $\|\lambda^{I,k}\|_0 \le n-s$ is not required.*

The following definition and theorem about AM-regularity for CCOP are obvious from Definition 3.7 and Theorem 3.10.

Definition 4.4. *Let $x^* \in \mathcal{F}$. Define a set-valued mapping $\mathcal{K} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by means of*

$$
\mathcal{K}(x) := \left\{ \sum_{i=1}^m \lambda_i^g \nabla g_i(x) + \sum_{i=1}^p \lambda_i^h \nabla h_i(x) + \sum_{i=1}^n \lambda_i^I e_i \;\middle|\; \begin{array}{c} \lambda_i^g \ge 0,\ i \in \mathcal{I}_g(x^*), \\ \lambda_i^I = 0,\ i \in I_\pm(x^*) \\ \text{and } \|\lambda^I\|_0 \le n - s. \end{array} \right\}.
$$

We say that x^* is AM-regular if the following condition holds:

$$
\limsup_{x\to x^*} \mathcal{K}(x) \subseteq \mathcal{K}(x^*).
$$

Theorem 4.3. *Let x^* be a local minimizer of CCOP, then x^* is AM-stationary. Moreover, suppose that x^* is AM-regular. Then, x^* is M-stationary.*

We now apply Definition 3.9 to CCOP (4.1) to discuss S-stationarity (d) and M-stationarity (d) where $d \in \mathcal{C}(x^*)$. The multipliers λ^g and λ^h are same as those in Definition 3.9. Now we focus on the discussion about multiplier λ^I.

- (S-stationarity (d) for CCOP) Let $\lambda^I \in \hat{N}_{T_S(x^*)}(d)$, from Theorem 4.1 we have
  (a1) if $|I_\pm(x^*) \cup I_\pm(d)| = s$, then $\lambda_i^I = 0$, $\forall i \in I_\pm(x^*) \cup I_\pm(d)$;
  (a2) if $|I_\pm(x^*) \cup I_\pm(d)| < s$, then $\lambda_i^I = 0$, $\forall i \in \{1, \ldots, n\}$.
- (M-stationarity (d) for CCOP) Let $\lambda^I \in N_S(x^*; d)$, from Theorem 4.1 we have
  (b1) if $|I_\pm(x^*) \cup I_\pm(d)| = s$, then $\lambda_i^I = 0$, $\forall i \in I_\pm(x^*) \cup I_\pm(d)$;
  (b2) if $|I_\pm(x^*) \cup I_\pm(d)| < s$, then $\lambda_i^I = 0, \forall i \in I_\pm(x^*) \cup I_\pm(d)$ and $\|\lambda^I\|_0 \le n - s$.

  Therefore, for $\lambda^I \in N_S(x^*; d)$ we have

$$\lambda_i^I = 0,\ \forall i \in I_\pm(x^*) \cup I_\pm(d) \text{ and } \|\lambda^I\|_0 \le n - s.$$

By the discussion above, we have the following definitions.

Definition 4.5. *Let $x^* \in \mathbb{R}^n$ be a feasible point of CCOP in the disjunctive form* (4.1) *and $d \in \mathcal{C}(x^*)$.*

(i) *We say x^* is S-stationary (d) if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ satisfying*

$$\begin{gathered}\nabla f(x^*) + \sum_{i \in \mathcal{I}_g(d)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{n} \lambda_i^I e_i = 0, \\ \lambda_i^g \ge 0, \forall i \in \mathcal{I}_g(d),\end{gathered} \tag{4.9}$$

such that

(a) *if $|I_\pm(x^*) \cup I_\pm(d)| = s$, then $\lambda_i^I = 0$, $\forall i \in I_\pm(x^*) \cup I_\pm(d)$.*
(b) *if $|I_\pm(x^*) \cup I_\pm(d)| < s$, then $\lambda_i^I = 0$, $\forall i \in \{1, \ldots, n\}$.*

(ii) *We say x^* is M-stationary (d) if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ satisfying* (4.9) *such that*

$$\lambda_i^I = 0,\ \forall i \in I_\pm(x^*) \cup I_\pm(d) \text{ and } \|\lambda^I\|_0 \le n - s.$$

The following theorem follows from Theorem 3.12 directly.

Theorem 4.4. *Let x^* be a feasible point for CCOP (4.1) and $d \in \mathcal{C}(x^*)$. x^* is M-stationary if x^* is M-stationary (d). Moreover, suppose the condition $\mathcal{I}_g(d) = \mathcal{I}_g(x^*)$ holds, each S-stationary point x^* is S-stationary (d).*

The following theorems summarizing second-order necessary and sufficient conditions for CCOP follows from Theorems 3.13-3.16 directly. In what follows, we assume all functions are twice continuously differentiable and denote the Lagrangian function of CCOP in the disjunctive subspaces form (4.1) by

$$\mathcal{L}(x, \lambda^g, \lambda^h, \lambda^I) := f(x) + \langle \lambda^g, g(x) \rangle + \langle \lambda^h, h(x) \rangle + \langle \lambda^I, x \rangle.$$

Theorem 4.5. *Let x^* be a local minimizer of problem (4.1). Assume that CCOP-LICQ is valid at x^*. Then, we have*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^I) d \geq 0, \ \forall d \in \mathcal{C}(x^*),$$

where $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ is the uniquely determined S-stationary multiplier associated with x^*.

Conversely, let x^* be an S-stationary point of problem (4.1). Assume there exists an S-stationary multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ associated with x^* such that

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^I) d > 0, \ \forall d \in \mathcal{C}(x^*) \setminus \{0\}$$

holds. Then, there are constants $\varepsilon > 0$ and $C > 0$ such that the quadratic growth condition is fulfilled at x^*. In particular, x^* is a strict local minimizer.

Theorem 4.6. *Let x^* be a local minimizer of problem (4.1) and $d \in \mathcal{C}(x^*)$. Assume $M(x) := F(x) - \Lambda$ is metrically subregular in direction d at $(x^*, 0)$, where $F(x) := (g(x), h(x), I(x))$ and $\Lambda := \mathbb{R}^m_- \times \{0\}^p \times S$. Then, there exists an M-stationary (d) multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ associated with x^* such that the following second-order condition holds:*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^I) d \geq 0.$$

Theorem 4.7. *Let x^* be a feasible point of problem (4.1). Assume for every nonzero critical direction $0 \neq d \in \mathcal{C}(x^*)$ there exists an S-stationary (d) multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ associated with x^* such that*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^I) d > 0.$$

Then, the quadratic growth condition is fulfilled at x^*. In particular, x^* is a strict local minimizer of problem (4.1).

Theorem 4.8. *Let x^* be an extended M-stationary point of problem* (4.1)*. Assume that for every nonzero critical direction $0 \neq d \in \mathcal{C}(x^*)$ one has*

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^g, \lambda^h, \lambda^I) d > 0$$

for any M-stationary (d) *multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ associated with x^*. Then, the quadratic growth condition is fulfilled at x^*. In particular, x^* is a strict local minimizer of problem* (4.1).

In Gfrerer's work [18], the author introduced strong M-stationarity (str. M-stationarity) for MPEC to build a bridge between M-stationarity and S-stationarity. In Liang and Ye's work [25], they introduced strong M-stationarity in the direction d (str. M-stationarity (d)) for MPSC to build a bridge between M-stationarity (d) and S-stationarity (d). Similar to their work, we also propose strong M-stationarity and strong M-stationarity (d) for CCOP. Moreover, we will discuss the relations among these two new concepts and extended M-stationarity which is usually difficult to verify in practice. We denote by $r(x^*)$ the rank of the family of gradients

$$\{\nabla g_i(x^*)\}_{i \in \mathcal{I}_g(x^*)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i \in I_0(x^*)}.$$

Now we give the definition of strong M-stationarity for CCOP.

Definition 4.6. *Assume that $J_g \subseteq \mathcal{I}_g(x^*)$ and $J_I \subseteq I_0(x^*)$, a pair of index sets (J_g, J_I) is called a CCOP working set if*

$$|J_g| + p + |J_I| = r(x^*)$$

and the family of gradients

$$\{\nabla g_i(x^*)\}_{i \in J_g} \cup \{\nabla h_i(x^*\}_{i=1}^p \cup \{e_i\}_{i \in J_I}$$

is linearly independent.

We say x^* is strongly M-stationary for CCOP, if there is a CCOP working set

(J_g, J_I) together with a M-stationary multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ such that

$$\lambda_i^g = 0,\ \forall i \in \{1, \dots, m\} \setminus J_g, \tag{4.10}$$

$$\lambda_i^I = 0,\ \forall i \in \{1, \dots, n\} \setminus J_I. \tag{4.11}$$

From the definition above, it follows immediately that a strong M-stationarity is M-stationary. A natural question is under what conditions a S-stationarity is strongly M-stationary. In the case of MPEC, the author showed that the implication holds under the constraint qualification condition MPEC-LICQ [18, Theorem 4.4]. In the case of CCOP, we also have the similar result.

Theorem 4.9. *Let x^* be S-stationary for CCOP in the disjunctive form* (4.1) *and assume that the constraint qualification CCOP-LICQ is fulfilled at x^*. Then x^* is strongly M-stationary.*

Proof. By (4.5), we can see that under the constraint qualification CCOP-LICQ, the only working set for CCOP is $(J_g, J_I) = (\mathcal{I}_g(x^*), I_0(x^*))$.

It follows immediately that x^* is M-stationary since it is S-stationary. Assume $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ is the corresponding M-stationary multiplier. Since the multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ satisfies $\lambda_i^g \geq 0,\ \forall i \in \mathcal{I}_g(x^*)$ and $\lambda_i^I = 0,\ \forall i \in I_\pm(x^*)$ and $\|\lambda^I\|_0 \leq n - s$, conditions (4.10) and (4.11) are fulfilled automatically. □

Now we discuss the relationship between strong M-stationarity and extended M-stationarity (see (iii) in Definition 3.8). In the MPEC case, Gfrerer showed that extended M-stationarity implies strong M-stationarity for MPEC if there exists some MPEC working set [18, Theorem 4.3].

Let us consider the CCOP case, assume x^* is extended M-stationary, hence M-stationary (take the direction $d = 0$) and there exists some CCOP working set (J_g, J_I) such that $|J_I| \leq n - s$. From the definition of M-stationarity, it follows that $-\nabla f(x^*)$ can be represented as a linear combination of the following family of vectors

$$\{\nabla g_i(x^*)\}_{i \in \mathcal{I}_g(x^*)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i \in I_0(x^*)}.$$

For every CCOP working set (J_g, J_I), we know from the definition of CCOP working set that $-\nabla f(x^*)$ can be represented as a unique linear combination of the following family of vectors

$$\{\nabla g_i(x^*)\}_{i \in J_g} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i \in J_I}.$$

We denote this unique multiplier by $\lambda(J) := (\lambda^g, \lambda^h, \lambda^I)^T$. It is obvious that conditions (4.10) and (4.11) are fulfilled. We have $\|\lambda^I\|_0 \leq n - s$ from $|J_I| \leq n - s$, which implies that $\lambda(J) = (\lambda^g, \lambda^h, \lambda^I)^T$ is a M-stationary multiplier. Therefore, x^* is strongly M-stationary. In this way, we obtain the following theorem.

Theorem 4.10. *Let x^* be extended M-stationary for CCOP in the disjunctive form* (4.1)*. Suppose that there exists some CCOP working set (J_g, J_I) such that $|J_I| \leq n-s$. Then x^* is strongly M-stationary.*

We also propose the concept of CCOP working set in the direction d as follows. We denote by $r(x^*; d)$ the rank of the following family of gradients

$$\{\nabla g_i(x^*)\}_{i \in \mathcal{I}_g(d)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i \in I_0(x^*) \cap I_0(d)}.$$

Now we introduce the definition of strong M-stationarity in the direction d (str. M-stationarity (d)) for CCOP and present corresponding theorems.

Definition 4.7. *Assume that $J_g(d) \subseteq \mathcal{I}_g(d)$ and $J_I(d) \subseteq I_0(x^*) \cap I_0(d)$, a pair of index sets $(J_g(d), J_I(d))$ is called a CCOP working set in the direction d if*

$$|J_g(d)| + p + |J_I(d)| = r(x^*; d)$$

and the family of gradients

$$\{\nabla g_i(x^*)\}_{i \in J_g(d)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i \in J_I(d)}$$

is linearly independent.

We say x^* is strongly M-stationary in the direction d for CCOP, if there is a CCOP working set in the direction d, that is $(J_g(d), J_I(d))$ together with a M-stationary (d) multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T$ such that

$$\lambda_i^g = 0, \ \forall i \in \{1, \ldots, m\} \setminus J_g(d), \tag{4.12}$$

$$\lambda_i^I = 0, \ \forall i \in \{1, \ldots, n\} \setminus J_I(d). \tag{4.13}$$

Before presenting corresponding theorems, let us introduce CCOP-LICQ (d) as follows. Let x^* be a feasible point, we say CCOP-LICQ (d) holds at x^* if the following family of vectors

$$\{\nabla g_i(x^*)\}_{i\in\mathcal{I}_g(d)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i\in I_0(x^*)\cap I_0(d)} \tag{4.14}$$

is linearly independent.

Theorem 4.11. *Let x^* be S-stationary (d) for CCOP in the disjunctive form (4.1) and assume that the constraint qualification CCOP-LICQ (d) is fulfilled at x^*. Then x^* is strongly M-stationary (d).*

Proof. The proof is very similar to that of Theorem 4.9. By (4.14), we can see under CCOP-LICQ (d), the only working set in the direction d for CCOP is $(J_g(d), J_I(d)) = (\mathcal{I}_g(d), I_0(x^*) \cap I_0(d))$.

It follows immediately that x^* is M-stationary (d) since it is S-stationary (d). Assume $\lambda = (\lambda^g, \lambda^h, \lambda^I)$ is the corresponding M-stationary (d) multiplier. Since the multiplier $\lambda = (\lambda^g, \lambda^h, \lambda^I)$ satisfies $\lambda_i^g \geq 0$, $\forall i \in \mathcal{I}_g(d)$ and $\lambda_i^I = 0$, $\forall i \in I_\pm(x^*) \cup I_\pm(d)$ and $\|\lambda^I\|_0 \leq n - s$, conditions (4.12) and (4.13) are fulfilled automatically. □

Theorem 4.12. *Let x^* be extended M-stationary for CCOP in the disjunctive form (4.1). Suppose that there exists some CCOP working set in the direction d $(J_g(d), J_I(d))$ such that $|J_I(d)| \leq n - s$. Then x^* is strongly M-stationary (d).*

Proof. The proof is very similar to that of Theorem 4.10. The feasible point x^* is M-stationary (d) since it is extended M-stationary. From the definition of M-stationarity (d), it follows that $-\nabla f(x^*)$ can be represented as a linear combination of the following family of vectors

$$\{\nabla g_i(x^*)\}_{i\in\mathcal{I}_g(d)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i\in I_0(x^*)\cap I_0(d)}.$$

For every CCOP working set in the direction d, that is $(J_g(d), J_I(d))$, we know from the definition that $-\nabla f(x^*)$ can be represented as a unique linear combination of the following family of vectors

$$\{\nabla g_i(x^*)\}_{i\in J_g(d)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i\in J_I(d)}.$$

We denote this unique multiplier by $\lambda(J;d) := (\lambda^g, \lambda^h, \lambda^I)$. It is obvious that conditions (4.12) and (4.13) are fulfilled. We have $\|\lambda^I\|_0 \leq n - s$ by noting that $|J_I(d)| \leq n - s$, which implies that $\lambda(J;d) = (\lambda^g, \lambda^h, \lambda^I)$ is a M-stationary (d) multiplier. Therefore, x^* is strongly M-stationary (d). □

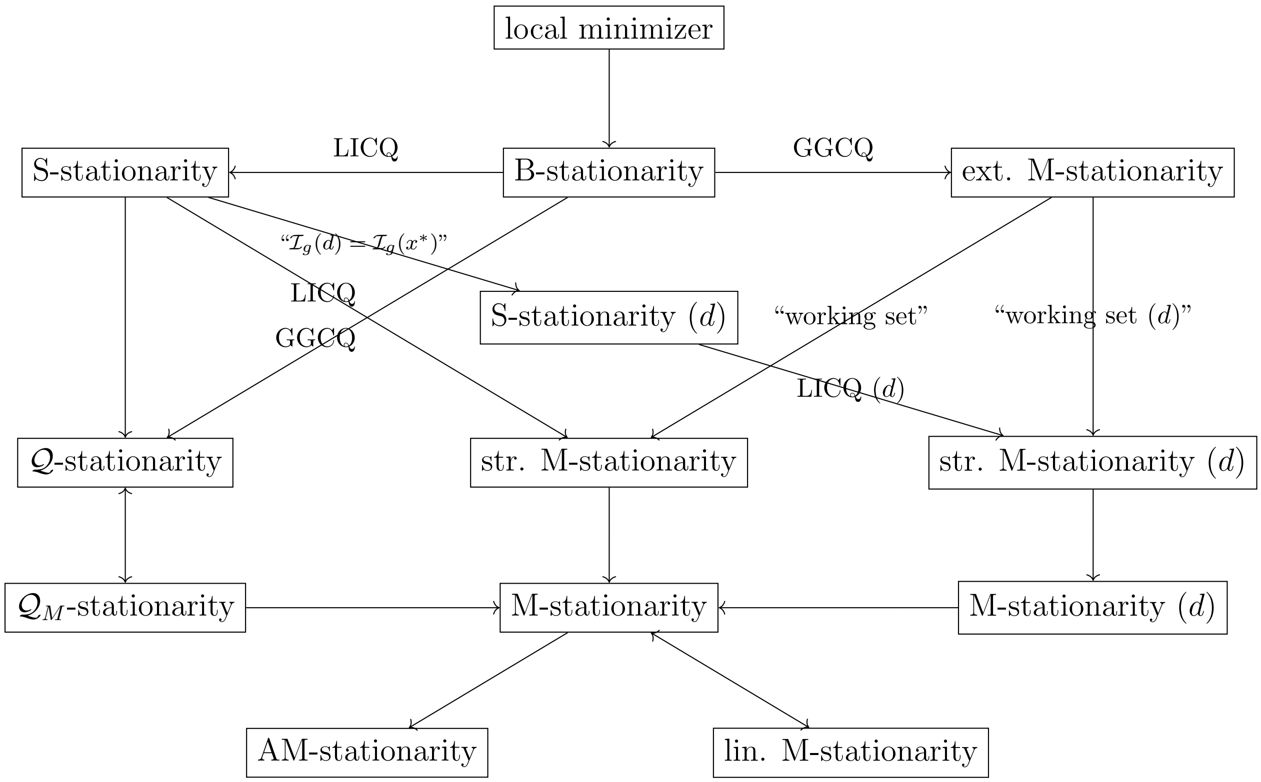

Figure 4.1: Relations among disjunctive-type stationarity conditions for CCOP

We conclude Section 4.2 with Figure 4.1 summarizing the relations among disjunctive-type stationarity conditions for CCOP we discussed above.

4.3 Disjunctive-type constraint qualifications for CCOP

In this section, we will show our disjunctive-type constraint qualifications for CCOP are weaker than complementarity-type ones. The following two corollaries are obvious from Corollary 3.1 since MPDSC includes MPSC and CCOP as special cases.

Corollary 4.2. *For mathematical programs with switching constraints* (MPSC)*, their constraint qualifications CRCQ, RCRCQ, CPLD, ERCPLD, and RCPLD coincide with their piecewise versions, respectively.*

Corollary 4.3. *For cardinality constrained optimization problems* (CCOP)*, their constraint qualifications CRCQ, RCRCQ, CPLD, ERCPLD, and RCPLD coincide with their piecewise versions, respectively.*

Now let us write down disjunctive-type constraint qualifications for CCOP in more specific forms. Let x^* be feasible for CCOP (1.1). Take CCOP-RCPLD for example, by Lemma 3.2 for $\lambda^I \in N_S(x^*)$ and the generator A, there exists $I \in \mathcal{I}_s$ such that

$$A \subseteq A_{\mathbb{R}_I}(x^*), 0 \neq \lambda^I \in \mathcal{G}(A) \subseteq \mathcal{G}(A_{\mathbb{R}_I}(x^*)).$$

Since $N_{\mathbb{R}_I}(x^*) = \mathbb{R}_I^\perp$ and $I_\pm(x^*) \subseteq I \in \mathcal{I}_s$, we have

$$A_{\mathbb{R}_I}(x^*) = \{e_i \mid i \in K \text{ such that } K \subseteq I_0(x^*) \text{ and } |K| \leq n-s\}.$$

By noting the condition $0 \neq \lambda^I \in \mathcal{G}(A)$, we can take generator A as

$$A = \{e_i \mid i \in K \text{ such that } I_\pm(\lambda^I) \subseteq K \subseteq I_0(x^*) \text{ and } |K| \leq n-s\}.$$

If $\lambda^I = 0$, we have $A = \emptyset$ hence $K = \emptyset$.

We need to emphasize here, we will obtain the condition $L \subseteq \{1\}$ if we apply constraint qualifications tailored for MPDSC (see Definition 3.12) to CCOP in the disjunctive form (4.1). In fact, the case $L = \emptyset$ is the same as the case $\lambda^I = 0$, at this time we have $A = \emptyset$ and constraint qualfications for CCOP will deduce to those for the standard nonlinear program (2.1). Now let us look at the following definition. Note that we have discussed the constraint qualification CCOP-LICQ in (4.5).

Definition 4.8. (Disjunctive-type constraint qualifications for CCOP) *Let x^* be feasible for CCOP* (1.1). *We say that x^* satisfies*

1. CCOP-LICQ *if the following vectors*

$$\{\nabla g_i(x^*)\}_{i\in\mathcal{I}_g(x^*)} \cup \{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i\in I_0(x^*)}$$

   is linearly independent.

2. CCOP-WLICQ *if there is no nonzero vector $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ with $\lambda_i^I = 0, \forall i \in I_\pm(x^*)$ and $\|\lambda^I\|_0 \leq n-s$ such that*

$$0 = \sum_{i\in\mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^n \lambda_i^I e_i.$$

3. CCOP-GMFCQ *if there is no nonzero vector $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$*

such that

$$0 = \sum_{i \in \mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{n} \lambda_i^I e_i,$$
$$\lambda_i^g \geq 0,\ i \in \mathcal{I}_g(x^*),\ \ \lambda_i^I = 0, \forall i \in I_\pm(x^*) \text{ and } \|\lambda^I\|_0 \leq n - s.$$

4. CCOP-CRCQ *if for every index sets* $I \subseteq \mathcal{I}_g(x^*)$, $J \subseteq \{1, \ldots, p\}$ *and* $\lambda^I \neq 0$ *satisfying* $\lambda_i^I = 0, \forall i \in I_\pm(x^*)$ *and* $\|\lambda^I\|_0 \leq n - s$, *then the set of vectors*

$$\{\nabla g(x^*)\}_{i \in I} \cup \{\nabla h(x^*)\}_{i \in J} \cup \{e_i\}_{i \in K}$$

and the set of vectors

$$\{\nabla g(x^k)\}_{i \in I} \cup \{\nabla h(x^k)\}_{i \in J} \cup \{e_i\}_{i \in K}$$

have the same rank for all sequences $\{x^k\}$ *satisfying* $x^k \to x^*$, $x^k \neq x^*$ *as* $k \to \infty$, *where* $I_\pm(\lambda^I) \subseteq K \subseteq I_0(x^*)$ *such that* $|K| \leq n - s$. *In the case* $\lambda^I = 0$, K *is taken as an empty set above.*

5. CCOP-RCRCQ *if the index set* J *is taken as* $\{1, \ldots, p\}$ *in CCOP-CRCQ.*

6. CCOP-CPLD *if there exists index sets* $I \subseteq \mathcal{I}_g(x^*)$, $J \subseteq \{1, \ldots, p\}$, *a nonzero vector* $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ *with* $\lambda_i^g \geq 0, i \in I$ *and* $\lambda^I \neq 0$ *satisfying* $\lambda_i^I = 0, \forall i \in I_\pm(x^*)$ *and* $\|\lambda^I\|_0 \leq n - s$ *such that*

$$0 = \sum_{i \in I} \lambda_i^g \nabla g_i(x^*) + \sum_{i \in J} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^{n} \lambda_i^I e_i,$$

then the set of vectors

$$\{\nabla g(x^k)\}_{i \in I} \cup \{\nabla h(x^k)\}_{i \in J} \cup \{e_i\}_{i \in K}$$

is linearly dependent for all sequences $\{x^k\}$ *satisfying* $x^k \to x^*$, $x^k \neq x^*$ *as* $k \to \infty$, *where* $I_\pm(\lambda^I) \subseteq K \subseteq I_0(x^*)$ *such that* $|K| \leq n - s$. *In the case* $\lambda^I = 0$, K *is taken as an empty set above.*

7. CCOP-ERCPLD *if the following conditions hold.*

   (i) *The vectors* $\{\nabla h_i(x)\}_{i=1}^p$ *have the same rank for all* $x \in \mathbb{B}_\varepsilon(x^*)$;

(ii) *Let* $J \subseteq \{1, \ldots, p\}$ *be such that the set of vectors* $\{\nabla h_i(x^*)\}_{i \in J}$ *is a basis for* $\text{span}\{\nabla h_i(x^*)\}_{i=1}^p$. *If there exists an index set* $I \subseteq \mathcal{I}_g(x^*)$ *and* $\lambda^I \neq 0$ *satisfying* $\lambda_i^I = 0, \forall i \in I_\pm(x^*)$ *and* $\|\lambda^I\|_0 \leq n-s$ *such that the set of vectors*

$$\{\nabla g_i(x^*)\}_{i \in I} \cup \Big\{\{\nabla h_i(x^*)\}_{i \in J} \cup \{e_i\}_{i \in K}\Big\}$$

is positive linearly dependent, then the set of vectors

$$\{\nabla g_i(x^k)\}_{i \in I} \cup \{\nabla h_i(x^k)\}_{i \in J} \cup \{e_i\}_{i \in K}$$

is linearly dependent for all sequences $\{x^k\}$ *satisfying* $x^k \to x^*$, $x^k \neq x^*$ *as* $k \to \infty$, *where* $I_\pm(\lambda^I) \subseteq K \subseteq I_0(x^*)$ *such that* $|K| \leq n - s$. *In the case* $\lambda^I = 0$, K *is taken as an empty set above.*

8. CCOP-RCPLD *if the following conditions hold.*

(i) *The vectors* $\{\nabla h_i(x)\}_{i=1}^p$ *have the same rank for all* $x \in \mathbb{B}_\varepsilon(x^*)$*;*

(ii) *Let* $J \subseteq \{1, \ldots, p\}$ *be such that the set of vectors* $\{\nabla h_i(x^*)\}_{i \in J}$ *is a basis for* $\text{span}\{\nabla h_i(x^*)\}_{i=1}^p$. *If there exist an index set* $I \subseteq \mathcal{I}_g(x^*)$, *a nonzero vector* $\lambda = (\lambda^g, \lambda^h, \lambda^I)^T \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ *with* $\lambda_i^g \geq 0, i \in I$ *and* $\lambda^I \neq 0$ *satisfying* $\lambda_i^I = 0, \forall i \in I_\pm(x^*)$ *and* $\|\lambda^I\|_0 \leq n - s$ *such that*

$$0 = \sum_{i \in I} \lambda_i^g \nabla g_i(x^*) + \sum_{i \in J} \lambda_i^h \nabla h_i(x^*) + \sum_{i=1}^n \lambda_i^I e_i,$$

then the set of vectors

$$\{\nabla g_i(x^k)\}_{i \in I} \cup \{\nabla h_i(x^k)\}_{i \in J} \cup \{e_i\}_{i \in K}$$

is linearly dependent for all sequences $\{x^k\}$ *satisfying* $x^k \to x^*$, $x^k \neq x^*$ *as* $k \to \infty$, *where* $I_\pm(\lambda^I) \subseteq K \subseteq I_0(x^*)$ *such that* $|K| \leq n - s$. *In the case* $\lambda^I = 0$, K *is taken as an empty set above.*

In Červinka et al. [15], the authors say that a feasible point (x^*, y^*) for the relaxed complementarity-type reformulation (1.2) satisfies a constraint qualification if x^* satisfies the corresponding constraint qualification for the tightened nonlinear problem (4.2), which implies that constraint qualifications only depends on x^* not on the pair (x^*, y^*). Therefore, we have the following definition.

Definition 4.9. (Relaxed constraint qualifications for CCOP) *Let x^* be feasible for CCOP* (1.1). *We say that x^* satisfies*

1. relaxed-MFCQ [15, Definition 3.11] *if the set of vectors*

$$\{\nabla g_i(x^*)\}_{i\in\mathcal{I}_g(x^*)} \cup \left\{\{\nabla h_i(x^*)\}_{i=1}^p \cup \{e_i\}_{i\in I_0(x^*)}\right\}$$

is positive linearly independent;

2. relaxed-CRCQ [15, Definition 3.11] *if for every index sets $I \subseteq \mathcal{I}_g(x^*)$, $J \subseteq \{1,\ldots,p\}$ and $K \subseteq I_0(x^*)$ such that the set of vectors*

$$\{\nabla g_i(x^*)\}_{i\in I} \cup \{\nabla h_i(x^*)\}_{i\in J} \cup \{e_i\}_{i\in K}$$

and the set of vectors

$$\{\nabla g_i(x^k)\}_{i\in I} \cup \{\nabla h_i(x^k)\}_{i\in J} \cup \{e_i\}_{i\in K}$$

have the same rank for all sequences $\{x^k\}$ satisfying $x^k \to x^*, x^k \neq x^*$ as $k \to \infty$;

3. relaxed-RCRCQ *if the index set J is taken as $\{1,\ldots,p\}$ in relaxed-CRCQ;*

4. relaxed-CPLD [15, Definition 3.11] *if there exists index sets $I \subseteq \mathcal{I}_g(x^*)$, $J \subseteq \{1,\ldots,p\}$ and $K \subseteq I_0(x^*)$ such that the set of vectors*

$$\{\nabla g_i(x^*)\}_{i\in I} \cup \left\{\{\nabla h_i(x^*)\}_{i\in J} \cup \{e_i\}_{i\in K}\right\}$$

is positive linearly dependent, then the set of vectors

$$\{\nabla g_i(x^k)\}_{i\in I} \cup \{\nabla h_i(x^k)\}_{i\in J} \cup \{e_i\}_{i\in K}$$

is linearly dependent for all sequences $\{x^k\}$ satisfying $x^k \to x^*, x^k \neq x^*$ as $k \to \infty$;

5. relaxed-RCPLD *if the following conditions hold.*

   (i) *The vectors $\{\nabla h_i(x)\}_{i=1}^p$ have the same rank for all $x \in \mathbb{B}_\varepsilon(x^*)$;*

(ii) *Let* $J \subseteq \{1, \ldots, p\}$ *be such that the set of vectors* $\{\nabla h_i(x^*)\}_{i \in J}$ *is a basis for* $\text{span}\{\nabla h_i(x^*)\}_{i=1}^p$. *If there exists index sets* $I \subseteq \mathcal{I}_g(x^*)$, $J \subseteq \{1, \ldots, p\}$ *and* $K \subseteq I_0(x^*)$ *such that the set of vectors*

$$\{\nabla g_i(x^*)\}_{i \in I} \cup \left\{ \{\nabla h_i(x^*)\}_{i \in J} \cup \{e_i\}_{i \in K} \right\}$$

is positive linearly dependent, then the set of vectors

$$\{\nabla g_i(x^k)\}_{i \in I} \cup \{\nabla h_i(x^k)\}_{i \in J} \cup \{e_i\}_{i \in K}$$

is linearly dependent for all sequences $\{x^k\}$ *satisfying* $x^k \to x^*, x^k \neq x^*$ *as* $k \to \infty$*;*

By noting the difference of index sets K in the disjunctive-type constraint qualifications and the relaxed ones, we have the following theorem.

Theorem 4.13. *Let* x^* *be any feasible point for CCOP* (1.1)*. Then,*

(i) *If* $\|x^*\|_0 = s$, *disjunctive-type constraint qualifications such as CRCQ, RCRCQ, CPLD and RCPLD from the disjunctive subspaces reformulation* (4.1) *coincide with the relaxed ones in* Definition 4.9*.*

(ii) *If* $\|x^*\|_0 < s$, *disjunctive-type constraint qualifications such as CRCQ, RCRCQ, CPLD and RCPLD from the disjunctive subspaces reformulation* (4.1) *are weaker than the relaxed ones in* Definition 4.9*.*

We also conclude this section with Figure 4.2 summarizing the relations among disjunctive-type constraint qualifications for CCOP we discussed above. Again, we can see from the diagram many constraint qualifications such as CRCQ, RCRCQ, CPLD, ERCPLD, and RCPLD coincide with their piecewise versions for CCOP, respectively.

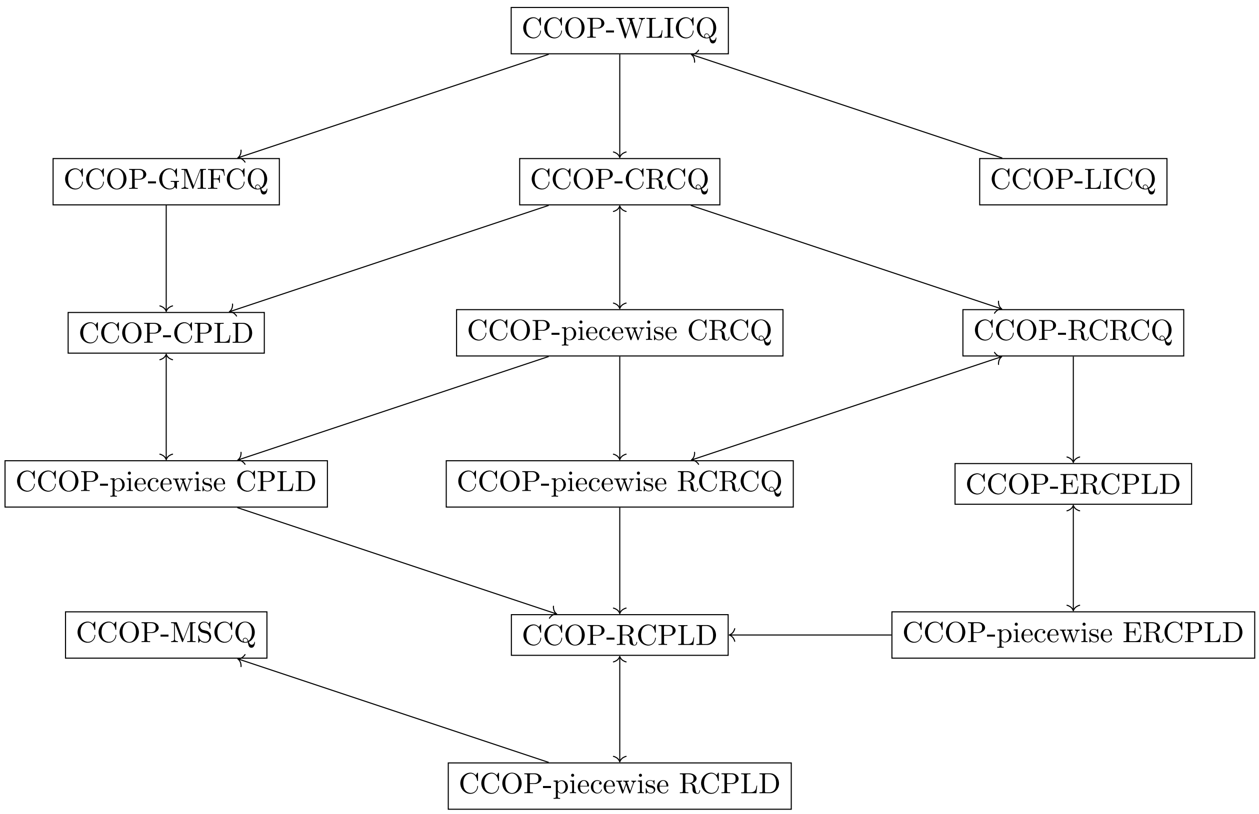

Figure 4.2: Relations among disjunctive-type constraint qualifications for CCOP

5 Conclusions and Future Work

In this thesis, we survey disjunctive-type optimality conditions for cardinality constrained optimization problems by reformulating them as mathematical programs with disjunctive subspaces constraints. In particular, we recover disjunctive-type S-stationarity and M-stationarity for CCOP, and then reveal the relationship between them and complementarity-type ones. We also show that CRCQ, RCRCQ, CPLD, ERCPLD, and RCPLD coincide with their piecewise versions for MPDSC, respectively. Based on this result, we prove that RCPLD is a sufficient condition for error bounds for MPDSC, which also holds for CCOP.

At the end of this thesis, we briefly talk about our future research. Directional variational analysis provide an intriguing area in optimization, and many directional optimality conditions keep emerging in recent papers. In our future work, we will try to use the tools from directional variational analysis to derive directional constraint qualifications for MPDC.

Bibliography

[1] W. Achtziger and C. Kanzow. *Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications.* Math. Program., **114**, (2008), 69-99.

[2] L. Adam, M. Červinka, and M. Pištěk. *Normally admissible stratifications and calculation of normal cones to a finite union of polyhedral sets.* Set-Valued Var. Anal., **24**, (2016), 207-229.

[3] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. Silva. *A relaxed constant positive linear dependence constraint qualification and applications.* Math. Program., **135**, (2012), 255-273.

[4] K. Bai, J. J. Ye, and J. Zhang. *Directional quasi-/pseudo-normality as sufficient conditions for metric subregularity.* SIAM J. Optim., **29**, (2019), 2625-2649.

[5] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. *Restricted normal cones and sparsity optimization with affine constraints.* Found. Comput. Math., **14**, (2014), 63-83.

[6] A. Beck and Y. C. Eldar. *Sparsity constrained nonlinear optimization: Optimality conditions and algorithms.* SIAM J. Optim., **23**, (2013), 1480-1509.

[7] M. Benko, M. Červinka, and T. Hoheisel. *Sufficient conditions for metric subregularity of constraint systems with applications to disjunctive and ortho-disjunctive programs.* Set-Valued Var. Anal., **30**, (2022), 143-177.

[8] M. Benko and H. Gfrerer. *On estimating the regular normal cone to constraint systems and stationarity conditions.* Optim., **66**, (2017), 61-92.

[9] M. Benko and H. Gfrerer. *New verifiable stationarity concepts for a class of mathematical programs with disjunctive constraints.* Optim., **67**, (2018), 1-23.

[10] M. Benko, H. Gfrerer, and J. Outrata. *Calculus for directional limiting normal cones and subdifferentials.* Set-Valued Var. Anal., **27**, (2019), 713-745.

[11] D. Bertsimas and R. Shioda. *Algorithm for cardinality-constrained quadratic optimization.* Comput. Optim. Appl., **43**, (2009), 1-22.

[12] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer, New York, 2000.

[13] M. Bucher and A. Schwartz. *Second-order optimality conditions and improved convergence results for regularization methods for cardinality-constrained optimization problems.* J. Optim. Theory Appl., **178**, (2018), 383-410.

[14] O. P. Burdakov, C. Kanzow, and A. Schwartz. *Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method.* SIAM J. Optim., **26**, (2016), 397-425.

[15] M. Červinka, C. Kanzow, and A. Schwartz. *Constraint qualifications and optimality conditions for optimization problems with cardinality constraints.* Math. Program., **160**, (2016), 353-377.

[16] F. H. Clarke. *Optimization and Nonsmooth Analysis.* Classics Appl. Math. 5, SIAM, Philadelphia, PA, 1990.

[17] M. L. Flegel, C. Kanzow, and J. Outrata. *Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints.* Set-Valued Var. Anal., **15**, (2007), 139-162.

[18] H. Gfrerer. *Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints.* SIAM J. Optim., **24**, (2014), 898-931.

[19] H. Gfrerer. *Linearized M-stationarity conditions for general optimization problems.* Set-Valued Var. Anal., **27**, (2019), 819-840.

[20] H. Gfrerer, J. J. Ye, and J. Zhou. *Second-order optimality conditions for nonconvex set-constrained optimization problems.* Math. Oper. Res., (2022), 1211, doi.org/10.1287/moor.2021.1211.

[21] I. Ginchev and B. S. Mordukhovich. *On directionally dependent subdifferentials.* C.R. Bulg. Acad. Sci., **64**, (2011), 497–508.

[22] R. Henrion and J. Outrata. *On calculating the normal cone to a finite union of convex polyhedra.* Optim., **57**, (2008), 57-78.

[23] R. Janin. *Directional derivative of the marginal function in nonlinear programming.* Math. Program. Stud., **21**, (1984), 110-126.

[24] C. Kanzow, A. B. Raharja, and A. Schwartz. *Sequential optimality conditions for cardinality-constrained optimization problems with applications.* Comput. Optim. Appl., **80**, (2021), 185-211.

[25] Y.-C. Liang and J. J. Ye. *Optimality conditions and exact penalty for mathematical programs with switching constraints.* J. Optim. Theory Appl., **190**, (2021), 1-31.

[26] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, Cambridge, 1996.

[27] O. L. Mangasarian and S. Fromovitz. *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints.* J. Math. Anal. Appl., **17**, (1967), 37-47.

[28] P. Mehlitz. *On the linear independence constraint qualification in disjunctive programming.* Optim., **69**, (2020), 2241-2277.

[29] P. Mehlitz. *Stationarity conditions and constraint qualifications for mathematical programs with switching constraints.* Math. Program., **181**, (2020), 149-186.

[30] P. Mehlitz. *Asymptotic stationarity and regularity for nonsmooth optimization problems.* J. Nonsmooth Anal. Optim., **1**, (2020), 6575.

[31] L. Minchenko and S. Stakhovski. *On relaxed constant rank regularity condition in mathematical programming.* Optim., **60**, (2011), 429-440.

[32] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation, Vol. 1: Basic Theory, Vol. 2: Applications.* Springer, Berlin, 2006.

[33] B. S. Mordukhovich. *Variational Analysis and Applications.* Monographs in Mathematics, Springer, Cham, Switzerland, 2018.

[34] L. L. Pan, N. H. Xiu, and J. Fan. *Optimality conditions for sparse nonlinear programming.* Sci. China Math., **60**, (2017), 759-776.

[35] L.-L. Pan, N.-H. Xiu, and S.-L. Zhou. *On solutions of sparsity constrained optimization.* J. Oper. Res. Soc. China, **3**, (2015), 421-439.

[36] L. Qi and Z. Wei. *On the constant positive linear dependence condition and its application to SQP methods.* SIAM J. Optim., **10**, (2000), 963-981.

[37] R. T. Rockafellar and R. J. Wets. *Variational Analysis.* Springer, Berlin, 1998.

[38] A. M. Tillmann, D. Bienstock, A. Lodi, and A. Schwartz. *Cardinality minimization, constraints, and regularization: a survey.* arXiv:2106.09606.

[39] M. Xu and J. J. Ye. *Relaxed constant positive linear dependence constraint qualification for disjunctive programs.* arXiv:2204.09869.

[40] J. J. Ye and J. Zhou. *Verifiable sufficient conditions for the error bound property of second-order cone complementarity problems.* Math. Program., **171**, (2018), 361-395.