Constant rank-type constraint qualifications and second-order optimality conditions

Thiago Parente da Silveira

Thesis submitted
to the
Institute of Mathematics and Statistics
of the
University of São Paulo
for the
acquisition of the academic degree
of
Doctor of Science

Program: Applied Mathematics
Advisor: Prof. Dr. Gabriel Haeser
Co-advisor: Prof. Dr. Héctor Ariel Ramírez Cabrera

Financial support by São Paulo Research Foundation (FAPESP) grants 2017/12187-9 and 2020/001305

São Paulo, December 2023

Constant rank-type constraint qualifications and second-order optimality conditions

This version of the thesis contains the corrections and changes suggested by the examination committee during the defense of the original version, that happened on 04/12/2023. A copy of the original version is available at the Institute of Mathematics and Statistics of the University of São Paulo.

Examination committee:

- Prof. Dr. Gabriel Haeser (orientador) - IME-USP
- Prof. Dr. Mituhiro Fukuda - UFABC
- Prof. Dr. Roberto Andreani - UNICAMP
- Prof. Dr. Roberto Cristóvão Mesquita Silva - UFAM
- Profa. Dra. Flávia Morgana de Oliveira Jacinto - UFAM

Acknowledgements

This part was not as easy as I expected. First of all, I would like to dedicate this thesis to my parents José and Elizabete. You are the reason of everything and my inspiration. You always supported me and I know that you give me the best of you every day. When I started college in 2011, I decided to be a Ph.D. in Optimization inspired by Prof Sandro Bitar. He was there in my first class and has been supporting me since then. I confess that sometimes I still do not believe that I got there. Thank you, Prof Sandro, for being such a good reference in my life. I really consider you as a father to me.

I can divide this phase into some blocks. The initial one, when I moved from Manaus to São Paulo I met important colleagues who helped me to adapt to "USP rhythm". As part of the life, I do not have contact with some of them anymore, and others became good and close friends. We shared good moments at Lab 118 and I will always remember the difficulties and joys that we had there. Among them, I would like to highlight my friend Diaulas, with whom I shared an apartment, good games, laughs, and some of the best ideas for my thesis. To all of you, thank you.

The second part was when I moved to Dresden, Germany. I was completely afraid of it, especially because it was during the pandemic. However, it was in Dresden that I could learn the most important lessons of my life. When I was alone, I could identify what is the most important thing in my life. I met amazing people there who helped me a lot and made my stay there incredible. Sometimes I really miss Dresden because of you. I would like to thank Mrs. Beate Müller, who welcomed me in a wonderful way that I did not even imagine was possible. She advised me of all the German bureaucracies and also invited me to stay with her and her family for Christmas. It was special to me. At TUD, Prof. Fischer and other colleagues told me how professional life works in Germany, and it was completely different from what I was expecting at the time. I was lucky to meet you all. Thank you!

The third and last one, was when I moved back to São Paulo. Initially, things were not in the way that I wanted, but God knows everything and He has the best for us. I could meet some friends again and, after some time, I found the most incredible woman on Earth. And, for my luck, she is going to be my wife in two months. Carol Moraes, thank you for your patience and for encouraging me at all times to complete this important stage in my life. I love you.

I would like to thank Prof Gabriel, especially for his patience. You gave me this opportunity some years ago and I will always be grateful for that. In addition, I would like to mention Prof Roberto Andreani (Nino) and Prof Héctor. During years we had meetings where I could learn a lot with all of you.

Lastly, I would like to thank the Financial support by São Paulo Research Foundation - FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, in Portuguese) grants 2017/12187-9 and 2020/001305. Without this financial support, I would not be able to do all those things.

Abstract

SILVEIRA, T. P. **Constant rank-type constraint qualifications and second-order optimality conditions**. Ph.D. thesis. Institute of Mathematics and Statistics of the University of São Paulo. Brazil, 2023.

The constant rank constraint qualification, introduced by Janin in [Math. Program. Study 21:110-126, 1984], has been shown very robust in diverse applications, such as global convergence of algorithms, second-order optimality conditions, computing the derivative of the value function, and stability analysis, but always in the nonlinear programming context. In this thesis, we propose different approaches to defining a constant rank-type constraint qualification for nonlinear second-order cone programming problems, that may be based either on the sequential optimality condition and then provide global convergence of an augmented Lagrangian algorithm, or a sequential approach based on the eigenvectors structure of the second-order cone and then get global convergence of algorithms based on an external penalty method, or a classical approach based on a constant rank theorem and then guarantees second-order necessary optimality condition based on the critical cone and holds for any Lagrange multiplier.

Keywords: second-order cone programming, constant rank, constraint qualification, second-order optimality conditions.

Resumo

SILVEIRA, T. P. **Condições de qualificações do tipo posto constante e condições de otimalidade de segunda ordem**. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo. Brasil, 2010.

A condição de qualificação de posto constante, introduzida por Janin em [Math. Program. Study 21:110-126, 1984], tem se mostrado muito robusta em diversas aplicações, tais como convergência global de algoritmos, condições de otimalidade de segunda ordem, cálculo da derivada da função valor, e análise de estabilidade, mas sempre no contexto de programação não linear. Nesta tese, nós propomos diferentes abordagens para definir uma condição de qualificação do tipo posto constante, que podem ser baseadas ou em condições sequenciais de otimalidade e então obter convergência global de um algoritmo tipo Lagrangiano aumentado, ou uma abordagem sequencial baseada na estrutura dos autovetores do cone de segunda ordem e então obter convergência global de algoritmos baseados em um método de penalidade externa, ou uma abordagem clássica baseada em um teorema de posto constante e então garantir condições necessárias de otimalidade de segunda ordem baseadas no cone crítico e que valem para qualquer multiplicador de Lagrange.

Palavras-chave: programação sob o cone de segunda ordem, posto constante, condição de qualificação, condições de otimalidade de segunda ordem.

Contents

List of Abbreviations ix

List of Figures xi

1 Introduction 1

2 Initial results of NSOCP 5

2.1 Nonlinear Second-Order Cone Programming Problem . . . 5

2.2 Revisiting Nonlinear Programming Problem . . . 7

2.3 First Approaches for CRCQ in NSOCP . . . 11

3 Sequential constraint qualifications for NSOCP 17

3.1 Revisiting nondegeneracy condition . . . 17

3.2 Eigenvectors and constraint qualifications . . . 21

4 Constant Rank-type Constraint Qualifications 29

4.1 Revisiting Abadie's and Guignard's constraint qualifications . . . 29

4.1.1 Guignard's Theorem . . . 30

4.2 Nondegeneracy and Abadie's CQ via Implicit Function Theorem . . . 32

4.3 Constant Rank Constraint Qualification for NSOCP . . . 38

4.4 A Constraint Qualification based on Curves and Second-Order Optimality Conditions 42

4.5 New proposals for second-order constraint qualifications for NLP . . . 48

5 Conclusion 55

5.1 Suggestion for future research . . . 57

Bibliography 59

Appendix A 63

Appendix B 69

Appendix C 93

Appendix D 131

Appendix E 159

List of Abbreviations

NSOCP	Nonlinear Second-Order Cone Programming
NLP	Nonlinear Programming
NSDP	Nonlinear Semidefinite Programming
LICQ	Linear Independence Constraint Qualification
CRCQ	Constant Rank Constraint Qualification
MFCQ	Mangasarian-Fromovitz Constraint Qualification
CPLD	Constant Positive Linear Dependence
RCRCQ	Relaxed Constant Rank Constraint Qualification
MFF	Modified Mangasarian-Fromovitz
RCPLD	Relaxed Constant Positive Linear Dependence
CRSC	Constant Rank of the Subspace Component
WCR	Weak Constant Rank
SSOC	Strong Second-Order Condition
WSOC	Weak Second-Order Condition
KKT	Karush-Kuhn-Tucker

List of Figures

2.1 Relation among the CQ's for (NSOCP). The boxes in blue are the well-known CQ's in nonlinear second-order cone programming and the green boxes are the naive proposals. 16

3.1 An arbitrary vector y and its eigenvectors. 18
3.2 Two LI eigenvectors converging to two LD eigenvectors. 26
3.3 Relation among the CQ's for (NSOCP). The boxes in blue are the well-known CQ's in nonlinear second-order cone programming and the green boxes are the new proposals including the weak versions. 27

4.1 Details about the proof of the Theorem 4.2.2 when $m = 2$. 35
4.2 Details about the proof of the Theorem 4.2.2 when $m > 2$. 36
4.3 Relation among the CQ's for (NSOCP). The boxes in blue are the well-known CQ's in nonlinear second-order cone programming and the green boxes are all the new proposals. 43
4.4 Relationship among the CQs mentioned for (NLP). CQ's in blue box are related with SSOC; CQ's in green box do not have such property. 54

Chapter 1

Introduction

The study of optimization is present throughout history even if implicitly. To solve problems in the best way possible walks side by side with human history. However, in the last century, the studies of optimization grew as fast as its importance, and it can be seen when we analyze the volume of research, publications, and applications of optimization nowadays.

One of the most studied classes of problems is Nonlinear Programming (NLP) with several applications. In addition, along with the development of computers with a bigger capacity for processing data, it was necessary the develop algorithms to solve more complex problems. But since the solution of an optimization problem is topological and to rewrite this to a computational language is not a trivial task, new optimality conditions are necessary. One of the most important optimality conditions are the so-called *Karush-Kuhn-Tucker* (KKT) conditions, that take into account a linear combination among the gradient of the objective function and the gradients of the constraints. However, in order to have the fulfillment of the KKT conditions, an additional requirement is necessary for a minimizer: the *constraint qualifications* (CQ).

In NLP the studies of constraint qualifications are well developed in different ways, among which we highlight the Linear Independence Constraint Qualification (LICQ) [NW99], Constant Rank Constraint Qualification (CRCQ) [Jan84], Mangasarian-Fromovitz Constraint Qualification (MFCQ) [MF67], Relaxed-CRCQ (RCRCQ) [MS11a], Constant Positive Linear Dependence (CPLD) [QW00], Relaxed-CPLD (RCPLD) [AHSS12a], Constant Rank of the Subspace Component (CRSC) [AHSS12b], Abadie's Constraint Qualification [Aba65] and Guignard's Constraint Qualification [Gui69].

It is important to notice that many of the constraint qualifications mentioned above are related to a constant rank condition, such as CRCQ, RCRCQ, and CRSC. The CRCQ was proposed by Janin in [Jan84] where he showed that under this condition it is possible to guarantee the existence of a Lagrange multiplier for a local minimizer, i.e., CRCQ is enough the fulfill the KKT conditions. In addition, Janin showed that under CRCQ is possible to compute the derivative of the value function, showed in a simple way that if the constraints are affine then every local minimizer of an NLP problem has a Lagrange multiplier, and, that CRCQ is independent of MFCQ and strictly weaker than LICQ. The CRCQ condition was used in many applications such as the study of stability [GM15, GO16] and global convergence of algorithm [BHR18]. Later, in [AES10] the authors showed that CRCQ also has second-order information, even if the set of Lagrange multipliers is neither unique (LICQ case) nor compact (MFCQ case).

When we pass to other classes of problems, the field of study of constraint qualifications was not as developed as it is in NLP. For example, we mention Nonlinear Second-Order Cone Programming (NSOCP), Nonlinear Semidefinite Programming (NSDP), and Nonlinear Cone Programming (NCP). The most known constraint qualifications in these classes of problems are the *Nondegeneracy Condition* (see [AG03] for NSOCP context, for example) and the Robinson's Constraint Qualification [Rob76], that can be seen as the generalizations of the LICQ and MFCQ, respectively. Under the nondegeneracy condition, the set of Lagrange multipliers is singleton and it makes all the analysis easier. When we consider Robinson's CQ, the set of Lagrange multipliers is nonempty and compact. However, when we analyze second-order conditions, that are desirable because they work as necessary and sufficient conditions for a local minimizer, Robinson's CQ is not so robust. See the discussion in Chapter 2 for more details. Thus, the development of constant rank-type constraint qualifications for NSOCP becomes more important.

To the best of our knowledge, just recently the first proposals of CRCQ, RCRCQ, and CRSC were made for NSOCP in [ZZ19]. However, analyzing their conditions we noticed that it was incorrect. Thus, these facts led us to the research that will be presented in this thesis and were published in some papers: [AFH+21, AHM+22a, AHM+22b, AHM+23].

In Chapter 2, we present the initial definitions and properties of an NSOCP problem. Later, we present the main constraint qualifications known in the NLP context, in order to show how well structured the field of constraint qualifications is in NLP. We also present the proposals made by Zhang and Zhang in [ZZ19] and our counterexample published in [AFH+21]. Such counterexample showed that defining a constant rank-type constraint qualification for NSOCP would not be an easy task, once such a condition would have to take into account that only vanish all the possible subsets of gradients and requires constant rank is not enough to guarantee the existence of Lagrange multipliers. See also the case when the constraints are linear in [ART02]. After that, we presented our first approaches in order to extend constant rank-type constraint qualifications for NSCOP. For such, the main idea was to "reduce" some second-order constraints to inequality constraints whenever it was possible and to use Robinson's CQ for the remaining constraints. In addition, we applied a powerful tool that is the sequential optimality conditions developed in [AFH+19]. These constraint qualifications were called *Naive-CRSC* and *Naive-RCPLD*. This approach was called "naive" because we could not deal with the pure second-order constraints, i.e., the ones that we could not reduce to NLP constraints and we used our expertise in constraint qualifications in NLP and we showed that we could mix both types of constraints and define a CQ. These results are based on [AHM+22a].

In Chapter 3, we continue our research in defining constant rank-type constraint qualifications, but now without avoiding taking into account the second-order structure. Actually, based on the ideas given by sequential optimality conditions, we analyzed the eigenvector structure of NSOCP and noticed that we could propose weaker versions of the nondegeneracy condition and Robinson's CQ called *weak-nondegeneracy* and *weak-Robinson*, respectively, where we showed that we do not have to analyze all the eigenvectors of the second-order cone in order to have a constraint qualification, just the ones that are limits of the eigenvectors of the constraints. With these weaker versions at hand, we could propose constraint qualifications called *weak-CRCQ* and *weak-CPLD*. This proposal has a "sequential" approach and it is enough to guarantee the global convergence of algorithms based on an external penalty method even if the set of Lagrange multipliers is not compact. The results of this chapter are based on [AHM+22b].

In Chapter 4, we finally proposed a constant rank-type constraint qualification in a similar vein that Janin did for NLP in [Jan84], that is, using a constant rank theorem. At first, we had to build a relation between the nondegeneracy condition and *Abadie's CQ*. However, the definition of Abadie's CQ was not so clear in the NSOCP context and it can be verified when we analyze the definition of Abadie's in [ZZ19, Theorem 3.1] and, in addition, recently Börgens et al. in [BKMW20, Definition 5.5] proposed Abadie's CQ for optimization problems in Banach spaces. In order to make everything clear, we recalled even the original Guignard's CQ [Gui69], and then we got the bridge between the nondegeneracy condition and the correct version of Abadie's CQ. After that, using the constant rank theorem used by Janin that was based on [Mal72], we introduced *Constant Rank Constraint Qualification* (CRCQ) for NSOCP. This proposal of CRCQ is strictly weaker than the nondegeneracy condition and independent of Robinson's CQ, as expected. It explains in a simple way the linear case and shows when we might have Lagrange multipliers for this class of problems. Moreover, under CRCQ we could also obtain second-order information in a similar way that Andreani et al. obtained in [AES10] for NLP, that is, a second-order condition based on the critical cone and holds for any Lagrange multiplier. This result is stronger than the one that can be obtained under Robison's CQ, even if the set of Lagrange multipliers is not compact. These results are based on [AHM+23]. Later in this chapter, we proposed a constraint qualification based on curves for NSOCP called *Ref-McCormick* inspired by the one made in [FSS22] in the NLP context. We showed that Ref-McCormick is weaker than CRCQ and stronger than Abadie's CQ, and keeps the second-order information that CRCQ has. Moreover, we showed that under Ref-McCormick the Hessian of the Lagrangian does not depend on the Lagrange multiplier, and this result is new even in NLP. Inspired by this discussion, we also proposed new constraint qualifications in the NLP context that imply the NLP-Ref-McCormick, and then are enough to guarantee second-order conditions based on the critical cone.

In the Appendices, we have the papers [AFH+21, AHM+22a, AHM+22b, AHM+23] and, moreover, an additional one in Appendix 5.1 (Appendix E), that was recently submitted where we show the difficulties on obtaining a practical algorithm that guarantees strong second-order conditions in NLP.

Chapter 2

Initial results of NSOCP

In this chapter, we will introduce the Nonlinear Second-Order Cone Programming (NSOCP) problem and the first approaches in trying to define a constant rank-type constraint qualification for this class of problems. The main results of this chapter are based on [AFH+21, AHM+22a].

2.1 Nonlinear Second-Order Cone Programming Problem

Let us consider the following problem

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & g_j(x) \in \mathbb{L}_{m_j}, \quad j = 1, \dots, q, \end{array} \tag{NSOCP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbb{L}_{m_j}$ is a second-order cone (or Lorentz cone), which is given by $\mathbb{L}_{m_j} := \{(z_0, \widehat{z}) \in \mathbb{R} \times \mathbb{R}^{m_j-1} \mid z_0 \geq \|\widehat{z}\|\}$ when $m_j > 1$ and $\mathbb{L}_1 := \{x \in \mathbb{R} \mid x \geq 0\}$. We will denote by $Dg_j(\overline{x})$ the first-order derivative of the function $g_j : \mathbb{R}^n \to \mathbb{R}^{m_j}$ at a given point $\overline{x} \in \mathbb{R}^n$, and by $Dg_j(\overline{x})^T$ the transpose of $Dg_j(\overline{x})$. In addition, $D^2 g_j(\overline{x})$ is the second-order derivative of g_j at $\overline{x}$ and $D^2 g_j(\overline{x})[d_1, d_2]$ denotes the operation of $D^2 g_j(\overline{x})$ on $d_1, d_2 \in \mathbb{R}^n$. We will assume that f, g_j, $j = 1, \dots, q$ are at least twice continuously differentiable.

The interior part of $\mathbb{L}_{m_j}$ is $\text{int}(\mathbb{L}_{m_j}) := \{(z_0, \widehat{z}) \in \mathbb{R} \times \mathbb{R}^{m_j-1} \mid z_0 > \|\widehat{z}\|\}$ and the nonzero boundary is $\text{bd}^+(\mathbb{L}_{m_j}) := \{(z_0, \widehat{z}) \in \mathbb{R} \times \mathbb{R}^{m_j-1} \mid z_0 = \|\widehat{z}\| > 0\}$. Let us denote the feasible set of (NSOCP) by Ω. Given a point $\overline{x} \in \Omega$, let us define the following index sets:

$$\begin{array}{rcl} I_{\text{int}}(\overline{x}) & := & \{j \in \{1, \dots, q\} \mid g_j(\overline{x}) \in \text{int}(\mathbb{L}_{m_j})\}, \\ I_B(\overline{x}) & := & \{j \in \{1, \dots, q\} \mid g_j(\overline{x}) \in \text{bd}^+(\mathbb{L}_{m_j})\}, \\ I_0(\overline{x}) & := & \{j \in \{1, \dots, q\} \mid g_j(\overline{x}) = 0\}, \end{array}$$

which consist of the indices $j \in \{1, \dots, q\}$ of the constraints that hit the interior, the nonzero boundary, and the vertex of their respective cones.

Two important cones in order to study optimality conditions at a feasible point $\overline{x}$ are the (Bouligand) *tangent* cone $\mathcal{T}_\Omega(\overline{x})$ and the *linearized* cone $\mathcal{L}_\Omega(\overline{x})$, which are given by

$$\mathcal{T}_\Omega(\overline{x}) := \{d \in \mathbb{R}^n \mid \exists\, t_k \to 0^+,\ \exists\, d^k \to d \text{ such that } \overline{x} + t_k d^k \in \Omega\}, \tag{2.1}$$

and

$$\begin{array}{rcl} \mathcal{L}_\Omega(\overline{x}) & := & \{d \in \mathbb{R}^n \mid Dg_j(\overline{x})d \in \mathcal{T}_{\mathbb{L}_{m_j}}(g_j(\overline{x})),\ j = 1, \dots, q\} \qquad (2.2) \\ & = & \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} Dg_j(\overline{x})d \in \mathbb{L}_{m_j}, & j \in I_0(\overline{x}); \\ \langle Dg_j(\overline{x})d, \Gamma_j g_j(\overline{x}) \rangle \geq 0, & j \in I_B(\overline{x}) \end{array} \right\}, \qquad (2.3) \end{array}$$

where Γ_j is an $m_j \times m_j$ diagonal matrix with 1 at its first entry and -1 at the others, and $\langle \cdot, \cdot \rangle$ denotes the usual inner product, and the equality was obtained in [BR05, Lemma 25]. It is known that $\mathcal{T}_\Omega(\overline{x}) \subseteq \mathcal{L}_\Omega(\overline{x})$, but the reverse it is not always satisfied.

Given a cone $K \subseteq \mathbb{R}^n$, the cone K° is called *polar cone* of K and it is defined as follows

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

The polar cone is convex and closed. In addition, we have that $(K^\circ)^\circ = \mathrm{cl}(K)$, that is, the closure of K. These relationships will be important when we analyze constraint qualifications.

One of the most well-known optimality conditions is called *first-order geometric necessary condition*. If $\overline{x}$ is a local minimizer of (NSOCP), then

$$-\nabla f(\overline{x}) \in \mathcal{T}_\Omega(\overline{x})^\circ. \tag{2.4}$$

Since the cone $\mathcal{T}_\Omega(\overline{x})$ is a geometrical object, computing its polar might not be an easy task. One could ask about the relationship between $-\nabla f(\overline{x})$ and $\mathcal{L}_\Omega(\overline{x})^\circ$, once we have an analytical description of $\mathcal{L}_\Omega(\overline{x})$, and then we could try to compute its polar. However, since the inclusion $\mathcal{T}_\Omega(\overline{x}) \subset \mathcal{L}_\Omega(\overline{x})$ might be strict, then we can obtain $\mathcal{L}_\Omega(\overline{x})^\circ \subsetneq \mathcal{T}_\Omega(\overline{x})^\circ$ and thus might there is a gap between the information obtained and the true optimality information of $\overline{x}$ analyzing only the polar of $\mathcal{L}_\Omega(\overline{x})$. These facts lead us to analyze other conditions in order to characterize local minimizers.

We say that the Karush-Kuhn-Tucker (KKT) conditions hold for problem (NSOCP) at a feasible point $\overline{x}$ if there exists $\mu_j \in \mathbb{L}_{m_j}$, $j = 1, \ldots, q$ such that

$$\nabla_x L(\overline{x}, \mu) = \nabla f(\overline{x}) - \sum_{j=1}^{q} Dg_j(\overline{x})^T \mu_j = 0, \tag{2.5}$$

$$\langle \mu_j, g_j(\overline{x})\rangle = 0, \quad j = 1, \ldots, q, \tag{2.6}$$

where $L(x, \mu) := f(x) - \sum_{j=1}^{q} \langle \mu_j, g_j(x)\rangle$ is the Lagrangian function for problem (NSOCP), $\nabla_x L(x, \mu)$ is the gradient of L at the point (x, μ) with respect to the variable x. The vectors μ_j that satisfy (2.5) and (2.6) are called *Lagrange multipliers*. The set of all Lagrange multipliers associated to a feasible point $\overline{x}$ will be denoted by $\Lambda(\overline{x})$.

Let us analyze with more detail the condition (2.6). Notice that if $j \in I_{\mathrm{int}}(\overline{x})$, then we must have $\mu_j = 0$. If $j \in I_0(\overline{x})$, then μ_j can be any vector in $\mathbb{L}_{m_j}$. In the last case, if $j \in I_B(\overline{x})$, then

$$\mu_j = \alpha_j \Gamma_j g_j(\overline{x}), \tag{2.7}$$

for some $\alpha_j \in \mathbb{R}_+$. See [AG03, Lemma 15] for more details. Substituting (2.7) in (2.5), we obtain that $\overline{x}$ is a KKT point if there are $\mu_j \in \mathbb{L}_{m_j}$, $j \in I_0(\overline{x})$ and $\alpha_j \geq 0$, $j \in I_B(\overline{x})$ such that

$$\nabla f(\overline{x}) - \sum_{j \in I_0(\overline{x})} Dg_j(\overline{x})^T \mu_j - \sum_{j \in I_B(\overline{x})} \alpha_j \nabla \phi_j(\overline{x}) = 0, \tag{2.8}$$

where

$$\phi_j(x) = [g_j(x)]_0 - \|\widehat{g}_j(x)\| \tag{2.9}$$

and

$$\nabla \phi_j(x) = \frac{1}{\|\widehat{g}_j(x)\|} Dg_j(x)^T \Gamma_j g_j(x), \tag{2.10}$$

$j \in I_B(\overline{x})$ and it is called *reduction mapping*. See more details about this in [BR05]. Notice that the conditions above are precisely the KKT conditions for the following reformulated problem

$$\begin{array}{lll} \text{Minimize} & f(x), & \\ \text{s.t.} & g_j(x) \in \mathbb{L}_{m_j}, & j \in I_0(\overline{x}), \\ & \phi_j(x) \geq 0, & j \in I_B(\overline{x}), \end{array}$$

where the original constraints $g_j(x) \in \mathbb{L}_{m_j}$ such that $j \in I_B(\overline{x})$ are replaced by the nonlinear constraints $\phi_j(x) \geq 0$ and the remaining ones, that is, $j \in I_{\mathrm{int}}(\overline{x})$, are omitted. These facts give us a powerful tool in order to study constraint qualifications for (NSOCP) problems, once constraint qualifications are well developed for nonlinear programming problems.

Despite describing a simple relationship between the gradient of the objective function and the first-order derivatives of the constraints, KKT conditions are not an optimality condition in the sense that it is satisfied by all local minimizers. In order to guarantee the existence of the Lagrange multipliers, it is necessary a *constraint qualification*.

One of the most studied constraint qualifications is the so-called *Nondegeneracy* condition. Using the ideas given by the reduction mapping, let us recall its definition. See [BS00, Equation 4.172] for more details.

Definition 2.1.1. ***(Nondegeneracy condition)*** *Let $\overline{x}$ be a feasible point of (NSOCP). Consider all the row vectors of the matrices $Dg_j(\overline{x})^T$, $j \in I_0(\overline{x})$, together with the row vectors $\nabla \phi_j(\overline{x})^T$, $j \in I_B(\overline{x})$. We say that* Nondegeneracy *condition holds at $\overline{x}$ when these vectors are linearly independent.*

The nondegeneracy condition is very similar to the *Linear Independence Constraint Qualification* for nonlinear programming problems [NW99]. Indeed, under nondegeneracy, it is possible to show that the set $\Lambda(\overline{x})$ is singleton. The reader can find more properties related to this condition in [BR05, Section 4].

Another well-known constraint qualification for (NSOCP) is the *Robinson's constraint qualification*. This condition was proposed in [Rob76] for a general conic context. However, since we are interested in second-order cone problems, let us restrict ourselves to this case and explore its properties in a deeper way. Following the ideas given in [BS00, Proposition 2.97, Corollary 2.98 and Lemma 2.99], we have the following:

Definition 2.1.2. ***(Robinson's CQ)*** *Let $\overline{x}$ be a feasible point of (NSOCP). We say that* Robinson's CQ *holds at $\overline{x}$ if*

$$\sum_{j=1}^{q} Dg_j(\overline{x})^T \mu_j = 0 \text{ and } \mu_j \in \mathbb{L}_{m_j},\ \langle \mu_j, g_j(\overline{x}) \rangle = 0 \implies \mu_j = 0,\ j = 1, \dots, q. \tag{2.11}$$

With the reduction mapping in mind, we can rewrite the condition (2.11) in the following way

$$\sum_{j \in I_0(\overline{x})} Dg_j(\overline{x})^T \mu_j + \sum_{j \in I_B(\overline{x})} \alpha_j \nabla \phi_j(\overline{x}) = 0,$$

where $\mu_j \in \mathbb{L}_{m_j}$, $j \in I_0(\overline{x})$; $\alpha_j \geq 0$, $j \in I_B(\overline{x})$ implies that $\mu_j = 0$, $j \in I_0(\overline{x})$ and $\alpha_j = 0$, $j \in I_B(\overline{x})$. Under Robinson's CQ it is possible to show that the set of Lagrange multipliers is compact and non-empty [BS00, Propositions 3.9 and 3.17]. Thus, Robinson's CQ can be seen as a natural generalization of the *Mangasarian-Fromovitz Constraint Qualification* for nonlinear programming problems [Rob82]. We will present the definition of LICQ and MFCQ properly in the following section.

Although these conditions are well established and have a counterpart in nonlinear programming problems, the field of study of constraint qualifications for (NSOCP) is not as well developed. To be more specific, to the best of our knowledge the first extension approach of constant rank-type constraint qualifications was in [ZZ19]. However, in [AFH+21] we showed that their proposals were incorrect. In order to have a better comprehension of the difficulties of extending constant rank-type constraint qualifications for (NSOCP), let us introduce them properly for nonlinear programming problems and, after that, we will analyze the proposals of [ZZ19].

2.2 Revisiting Nonlinear Programming Problem

Let us consider the standard nonlinear programming problem

$$\begin{array}{lll} \text{Minimize} & f(x), & \\ \text{s.t.} & g_j(x) \geq 0, & j = 1, \dots, m, \\ & h_i(x) = 0, & i = 1, \dots, p, \end{array} \tag{NLP}$$

where $f, g_j, h_i : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, m$, $i = 1, \dots, p$ are twice continuously differentiable. Given a feasible point $\overline{x}$ of (NLP), we define the set of active inequality constraints as $A(\overline{x}) := \{j \in \{1, \dots, m\} \mid g_j(\overline{x}) = 0\}$.

Given a solution $\overline{x}$ of (NLP), we are interested in using first-order conditions in order to characterize this point. In a nonlinear programming context, we say that the KKT conditions hold at a feasible point $\overline{x}$, if there exist $\lambda \in \mathbb{R}^p$ and $\mu \in \mathbb{R}^m_+$ such that

$$\nabla f(\overline{x}) + \sum_{i=1}^{p} \lambda_i \nabla h_i(\overline{x}) - \sum_{j \in A(\overline{x})} \mu_j \nabla g_j(\overline{x}) = 0. \tag{2.12}$$

The tangent cone and the linearized cone are defined in a similar way to (NSOCP). We just have to keep in mind that the feasible set Ω is given by (NLP). Thus,

$$\mathcal{T}_\Omega(\overline{x})_{\text{NLP}} := \{d \in \mathbb{R}^n \mid \exists\, t_k \to 0^+,\ \exists d^k \to d \text{ such that } \overline{x} + t_k d^k \in \Omega\}$$

and

$$\mathcal{L}_\Omega(\overline{x})_{\text{NLP}} := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla h_i(\overline{x})^T d = 0, & i = 1, 2, \dots, p \\ \nabla g_j(\overline{x})^T d \geq 0, & j \in A(\overline{x}) \end{array} \right\}.$$

In a nonlinear programming context, it is simple to compute the polar cone of $\mathcal{L}_\Omega(\overline{x})_{\text{NLP}}$. It is given by

$$\mathcal{L}_\Omega(\overline{x})^\circ_{\text{NLP}} = \left\{ v \in \mathbb{R}^n \mid v = \sum_{i=1}^{p} \lambda_i \nabla h_i(\overline{x}) - \sum_{j \in A(\overline{x})} \mu_j \nabla g_j(\overline{x}),\ \mu_j \geq 0 \right\}.$$

Note that the KKT conditions can be written as $-\nabla f(\overline{x}) \in \mathcal{L}_\Omega(\overline{x})^\circ_{\text{NLP}}$. Thus, since the geometric condition also holds in a nonlinear programming context, that is, if $\overline{x}$ is a local minimizer of (NLP), then $-\nabla f(\overline{x}) \in \mathcal{T}_\Omega(\overline{x})^\circ_{\text{NLP}}$, any condition which implies $\mathcal{L}_\Omega(\overline{x})^\circ_{\text{NLP}} \subseteq \mathcal{T}_\Omega(\overline{x})^\circ_{\text{NLP}}$ must be a constraint qualification in nonlinear programming context. Now, let us analyze the counterpart of the nondegeneracy condition and Robinson's CQ for (NLP):

Let $\overline{x}$ be a feasible point of (NLP). Then:

i) the Linear Independence Constraint Qualification (LICQ) holds at $\overline{x}$ if the set

$$\left\{ \{\nabla h_i(\overline{x})\}_{i=1}^{p}, \{\nabla g_j(\overline{x})\}_{j \in A(\overline{x})} \right\}$$

is linearly independent;

ii) the Mangasarian-Fromovitz Constraint Qualification [MF67] (MFCQ) holds at $\overline{x}$ if

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(\overline{x}) - \sum_{j \in A(\overline{x})} \mu_j \nabla g_j(\overline{x}) = 0, \quad \mu_j \geq 0$$

implies that $\lambda_i = 0$, $i = 1, 2, \dots, p$ and $\mu_j = 0$, $j \in A(\overline{x})$.

The Linear Independence CQ has some important properties. For example, under LICQ we have the existence and uniqueness of the Lagrange multiplier, that is, if $\overline{x}$ is a local minimizer of (NLP) and satisfies LICQ, then there is a unique Lagrange multiplier $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^m_+$ such that $(\overline{x}, \lambda, \mu)$ satisfies the KKT conditions for (NLP). See [NW99] for more details. The MFCQ condition can be seen as a positive linearly independence of the gradients of the active constraints, where the scalars associated with the gradients of the inequality constraints must be non-negative. In this way, it is possible to see that MFCQ is weaker than LICQ. The MFCQ condition implies that the set of Lagrange multipliers is a non-empty and compact set.

Despite having good properties as a constraint qualification, the MFCQ condition may fail in simple cases. For example, if we consider a nonlinear programming problem with two linear inequality constraints $g_1(x), g_2(x) \geq 0$ where $g_2(x) = -g_1(x)$, then MFCQ (and, consequently, LICQ) does not hold at any feasible point. A good way of dealing with linear constraints was presented by Janin in [Jan84] through the *Constant Rank Constraint Qualification* (CRCQ). Let us recall the definition.

Definition 2.2.1. ***(CRCQ for NLP)*** *Let $\overline{x}$ be a feasible point of (NLP). We say that the* Constant Rank Constraint Qualification (CRCQ) *holds at $\overline{x}$ of (NLP), if there exists a neighborhood V of $\overline{x}$, such that for every subsets $I \subseteq \{1, \ldots, p\}$ and $J \subseteq A(\overline{x})$, the rank of $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_j(x)\}_{j \in J}$ remains constant for all $x \in V$.*

In [Jan84], the author showed that CRCQ is strictly weaker than LICQ and independent of MFCQ. The CRCQ also explains in a simple way what happens at (NLP) where all the constraints are linear. In order to show that CRCQ is indeed a constraint qualification, Janin used a constant rank theorem to obtain the equality $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$, which is known as *Abadie's Constraint Qualification* in nonlinear programming problems.

At first, CRCQ seems no relation with LICQ, as CRCQ requires a piece of information in a neighborhood of $\overline{x}$ and is described for every subset of constraints. However, by requiring LICQ, we are requiring linear independence for all subsets of active constraints in a whole neighborhood of the point, in other words, LICQ can be equivalently described as for every subset $I \subseteq \{1, 2, \ldots, p\}$, $J \subseteq A(\overline{x})$, the set $\{\nabla h_i(\overline{x}), \nabla g_j(\overline{x}) \mid i \in I, j \in J\}$ is linearly independent. From this point of view, we can see that CRCQ is weaker than LICQ.

In addition to describing in a very simple way the existence of Lagrange multipliers in a problem with linear constraints, the CRCQ also has other important properties related to second-order optimality conditions. In order to have a better comprehension of this topic, let us define the following sets.

Let $\overline{x}$ be a feasible point of (NLP). The the *critical cone* $C(\overline{x})_{\text{NLP}}$, is defined as

$$C(\overline{x})_{\text{NLP}} := \mathcal{L}_\Omega(\overline{x})_{\text{NLP}} \cap \{\nabla f(\overline{x})\}^\perp,$$

where $\{\nabla f(\overline{x})\}^\perp$ denotes the set of vectors that are orthogonal to $\nabla f(\overline{x})$. In addition, when $\overline{x}$ admits a Lagrange multiplier pair (λ, μ) associated respectively with equalities and active inequalities, the critical cone can be written as

$$C(\overline{x})_{\text{NLP}} = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla h_i(\overline{x})^T d = 0, & i = 1, 2, \ldots, p \\ \nabla g_j(\overline{x})^T d \geq 0, & j \in A(\overline{x}),\ \mu_j = 0 \\ \nabla g_j(\overline{x})^T d = 0, & j \in A(\overline{x}),\ \mu_j > 0 \end{array} \right\}.$$

The critical cone represents true second-order information since sufficient optimality conditions are also based on the same critical cone, thus the necessary second-order conditions based on the critical cone are more desirable.

We say that the *Strong Second-Order Condition* (SSOC) holds at a KKT point $\overline{x}$ of (NLP) associated to the Lagrange multipliers $\lambda \in \mathbb{R}^p$, $\mu \in \mathbb{R}^m_+$, if for every d in the critical cone $C(\overline{x})_{\text{NLP}}$ we have that

$$d^T \left(\nabla^2 f(\overline{x}) + \sum_{i=1}^{p} \lambda_i \nabla^2 h_i(\overline{x}) - \sum_{j \in A(\overline{x})} \mu_j \nabla^2 g_j(\overline{x}) \right) d \geq 0. \tag{2.13}$$

If the inequality above is strict, that is, if the quadratic form in (2.13) is positive definite in the critical cone, then we have a sufficient condition for strict local optimality.

In contrast to theoretical optimality conditions, any known second-order practical algorithm is only guaranteed to satisfy a weaker necessary second-order condition, where the critical directions considered are those of the following critical subspace:

$$S(\overline{x})_{\text{NLP}} := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla h_i(\overline{x})^T d = 0, & i = 1, 2, \ldots, p \\ \nabla g_j(\overline{x})^T d = 0, & j \in A(\overline{x}) \end{array} \right\}, \tag{2.14}$$

which is the lineality space of $C(\overline{x})_{\text{NLP}}$. If in (2.13) we consider directions d in the critical subspace $S(\overline{x})$ instead of considering the direction in the critical cone, then we say that $\overline{x}$ satisfies the *Weak Second-Order Condition* (WSOC) for (NLP). At a first-order stationary point $\overline{x}$ satisfying strict complementarity, the critical cone is reduced to the critical subspace, but in general, these sets may be different. Note that even for linear constraints, at the vertices of a polytope, the critical subspace is empty, which causes the required second-order condition to be automatically satisfied regardless of the objective function, which indicates that this condition is too weak to attest optimality.

The study that makes the relation between constraint qualification and second-order optimality conditions is not so easy. For example, since LICQ implies the uniqueness of the Lagrange multipliers it is possible to show that a local minimizer $\overline{x}$ of (NLP) associated with (λ, μ) satisfies SSOC. On the other hand, even if MFCQ implies compactness of the Lagrange multiplier set, an example given by Arutyunov in [Aru98] shows that "min + MFCQ" does not imply even WSOC. Such a counter-example was rediscovered by Anitescu in [Ani00, Section 3]. This also implies that constraint qualifications that were proposed later that are weaker than MFCQ, also do not satisfy such second-order optimality conditions. With this in mind and recalling that CRCQ is independent of MFCQ, one can ask about the relationship between CRCQ and SSOC. This question is fully explained by Andreani et al. in [AES10].

Theorem 2.2.1. *([AES10, Theorem 3.1]) Suppose that $\overline{x} \in \Omega$ is a local minimizer of (NLP) such that CRCQ holds. Then, for any Lagrange multiplier $(\lambda, \mu) \in \Lambda(\overline{x})$, $(\overline{x}, \lambda, \mu)$ verifies the strong second-order condition.*

The theorem above has important implications. One of the most important is the fact that under CRCQ we may not have that the set of Lagrange multipliers is bounded. However, the Hessian of Lagrangian is positive semidefinite for all Lagrange multipliers. Again, since the condition based on the critical cone has the "non-gap" property, it is more desirable.

Later, some weaker versions of CRCQ were proposed. Let $\overline{x}$ be a feasible point of (NLP). Then:

i) the *Relaxed-CRCQ* (RCRCQ) [MS11a] holds at $\overline{x}$ if there exists a neighborhood V of $\overline{x}$ such that for any subset $J \subseteq A(\overline{x})$, the rank of the family $\{\nabla h_i(x), \nabla g_j(x) \mid i \in \{1, 2, \ldots, p\},\ j \in J\}$ remains constant for all $x \in V$.

ii) consider the following set

$$J^-_{\text{NLP}}(\overline{x}) := \{j \in A(\overline{x}) \mid -\nabla g_j(\overline{x}) \in \mathcal{L}_\Omega(\overline{x})^\circ_{\text{NLP}}\}. \tag{2.15}$$

The *Constant Rank of the Subspace Component* (CRSC) [AHSS12b] holds at $\overline{x}$, if there exists a neighborhood V of $\overline{x}$ such that the rank of $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla g_j(x)\}_{j \in J^-_{\text{NLP}}}$ remains constant for all $x \in V$.

iii) the *Weak Constant Rank* (WCR) [AMS07] holds at $\overline{x}$ if there exists a neighborhood V of $\overline{x}$ such that the rank of the family $\{\nabla h_i(x) \mid i \in \{1, 2, \ldots, p\}\} \cup \{\nabla g_j(x) \mid j \in A(\overline{x})\}$ remains constant for all $x \in V$.

Under RCRCQ it is possible to show that every local minimizer is a SSOC point for any Lagrange multiplier [MS11b, Theorem 6]. In [AHSS12b], the authors showed that CRSC is weaker than MFCQ. Thus, we have that SSOC does not hold under this condition. The difference between RCRCQ and CRSC is the subsets where constant rank is required around the feasible point. Under RCRCQ, we vary all the subsets of indexes of the inequality constraints that are active at $\overline{x}$. On the other hand, CRSC captures exactly the set in which constant rank is necessary in order to define a constraint qualification. In addition, even though WCR condition seems to have similar properties as CRSC, only WCR is not enough to guarantee even the existence of Lagrange multipliers, that is, it is not a constraint qualification. See [AMS07, Counterexample 5.1] for more details. This fact shows us the importance of identifying correctly the sets of gradients such that constant rank is required. Proceeding in the correct way can give us not only the existence of Lagrange multipliers but some second-order information as well.

2.3 First Approaches for CRCQ in NSOCP

In this section, we will show some approaches to defining a constant rank-type constraint qualification in nonlinear second-order cone programming. To the best of our knowledge, the first tentative was made by Zhang and Zhang in [ZZ19]. In that paper, the authors proposed not only an extension of CRCQ but also RCRCQ and CRSC. Let us rewrite the proposal given in [ZZ19]

Definition 2.3.1. *The Constant Rank Constraint Qualification (CRCQ) as defined in [ZZ19] holds at a feasible point $\overline{x}$ of (NSOCP) if there exists a neighborhood V of $\overline{x}$ such that for any index sets $J_1 \subseteq I_0(\overline{x})$ and $J_2 \subseteq I_B(\overline{x})$, the family of matrices whose rows are the union of $Dg_j(x)$, $j \in J_1$ and the vector rows $(Dg_j(x)\Gamma_j g_j(x))^T$, $j \in J_2$ has the same rank for all $x \in V$.*

Let us have a first look at this definition. Given a feasible point $\overline{x}$ of (NSOCP), notice that the vectors $(Dg_j(x)\Gamma_j g_j(x))^T$ can be seen as the gradients of the functions $\tilde{\phi}_j(x) := \frac{1}{2}([g_j(x)]_0^2 - \|\widehat{g}_j(x)\|^2)$ where $j \in I_B(\overline{x})$, which is a different reduction mapping for (NSOCP). We will see more properties about this topic later. Since we already know that the constraints at the boundary of a second-order cone have a behavior similar to inequality constraints in nonlinear programming problems, it is expected to require constant rank for all subsets of $I_B(\overline{x})$. However, the main difference relies on the constraints in which $j \in I_0(\overline{x})$.

In order to make the analysis simpler, let us consider only one single SOCP constraint in (NSOCP), that is, $q = 1$ and $m_1 > 1$ with $g(\overline{x}) = 0$. We have a "multi-dimensionally active" constraint and, in this case, the CRCQ proposal given in [ZZ19] consists on requiring constant rank of $Dg(x)^T$ for all x around $\overline{x}$, i.e., constant rank of the set $\{\nabla g_0(x), \dots, \nabla g_{m_1-1}(x)\}$ for $x \in V$, where V is a neighborhood of $\overline{x}$. This condition is similar to WCR mentioned previously, which is not a constraint qualification. In fact, the following example given in [AFH+21] shows that the proposals given by Zhang and Zhang in [ZZ19] were incorrect.

$$\begin{aligned} \text{Minimize} \quad & f(x) := -x, \\ \text{s.t.} \quad & g(x) \in \mathbb{L}_2, \end{aligned} \tag{2.16}$$

with

$$g(x) = \begin{pmatrix} g_0(x) \\ g_1(x) \end{pmatrix} := \begin{pmatrix} x \\ x + x^2 \end{pmatrix}.$$

The point $\overline{x} = 0$ is the unique feasible point of the problem. Since $g(\overline{x}) = 0$, the KKT conditions for this problem are given by the existence of $\mu = (\mu_0, \mu_1) \in \mathbb{L}_2$ such that $\nabla f(\overline{x}) - Dg(\overline{x})^T \mu = 0$, that is,

$$-1 - \mu_0 - \mu_1 = 0. \tag{2.17}$$

Once $\mu \in \mathbb{L}_2$, we have that $\mu_0 \geq |\mu_1|$. Thus, (2.17) does not have a solution. In addition, if we look deeper at this example, we notice more important facts. We have that $\nabla g_0(x) = 1$ and $\nabla g_1(x) = 1 + 2x$ for all x. It means that all subsets of gradients

$$\{\nabla g_0(x)\}, \{\nabla g_1(x)\}, \{\nabla g_0(x), \nabla g_1(x)\}$$

have constant rank equal to 1 for all x near $\overline{x}$. Therefore, requiring constant rank of all subsets is not enough for being a constraint qualification in (NSOCP).

The proposal made by Zhang and Zhang in [ZZ19] was based on an implicit function theorem [Zor82] (which is similar to the approach used by Minchenko and Stakhovski in [MS11a] with the Lyusternik's Theorem [IT74]) in order to prove the existence of a feasible curve satisfying some properties for each direction in the linearized cone. The difficulty with this approach is the fact that there are "more types" of direction in the linearized cone. Indeed, let us consider the problem (NSOCP) with $q = 1$ and a feasible point $\overline{x}$ such that $g(\overline{x}) = 0$. Given a direction $d \in \mathcal{L}_\Omega(\overline{x})$, we have that $Dg(\overline{x})d \in \mathbb{L}_m$. If $Dg(\overline{x})d \in \text{int}(\mathbb{L}_m)$, then this direction does not interfere locally with the feasibility of $g(\overline{x} + td)$, with $t > 0$ small enough. If $Dg(\overline{x})d = 0$, then we have a similar case to the nonlinear programming problem, and an implicit function theorem can handle this situation (see [AES10] for more details). However, if $Dg(\overline{x})d \in \text{bd}^+(\mathbb{L}_m)$, we do not have known tools for handling with this case. With this information in mind, we need to try another approach to solving this issue.

A different approach to defining constraint qualifications is through the so-called *sequential optimality conditions* [AHM11]. The *Approximate-KKT* (AKKT) condition was proposed initially for (NLP) problems.

Theorem 2.3.1. *Let $\overline{x}$ be a local minimizer of (NLP). Then, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\mu^k\} \subset \mathbb{R}^m_+$ such that $x^k \to \overline{x}$ and*

$$\nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^k \nabla h_i(x^k) - \sum_{j=1}^{m} \mu_j^k \nabla g_j(x^k) \to 0 \tag{2.18}$$

The point $\overline{x}$ is called an AKKT-point. It is important to emphasize that the AKKT condition is a pure optimality condition, that is, it is satisfied by every local minimizer even when any constraint qualification is not. In addition, it is a powerful tool for proving algorithm convergence (see [BHR18] and references therein). For the sake of defining CRCQ through the AKKT condition, we will need the Carathéodory's Lemma. Let us recall it as stated in [AHSS12a].

Lemma 2.3.1. *(Carathéodory's Lemma) Let $v_1, \ldots, v_{p+q} \in \mathbb{R}^n$ be such that $\{v_i\}_{i=1}^p$ are linearly independent. Let α_i, $i = 1, \ldots, p+q$ be real numbers and consider the vector $v := \sum_{i=1}^{p+q} \alpha_i v_i$. Then, there exist $J \subseteq \{p+1, \ldots, p+q\}$ and scalars $\tilde{\alpha}_i$, $i \in \{1, \ldots, p\} \cup J$, such that $\{v_i\}_{i \in \{1,\ldots,p\} \cup J}$ are linearly independent, $\alpha_i > 0$ implies $\tilde{\alpha}_i > 0$, for all $i \in J$, and*

$$v = \sum_{i \in \{i=1,\ldots,p\} \cup J} \tilde{\alpha}_i v_i.$$

This lemma plays an important role once it says that we can rewrite a linear combination using a subset of linear independent vectors and keeping the signal of some of the scalars. Remember that in (2.12) the multipliers associated with the gradients of the inequalities must be non-negative. In (NSOCP) context, the multipliers associated must be in their respective second-order cones. Unfortunately, it may not be possible as illustrated in the following example given in [AFH+21, Example 1].

Example 2.3.1. *Consider the vector $v := \alpha_0 v_0 + \alpha_1 v_1 + \alpha_2 v_2$ with $(\alpha_0, \alpha_1, \alpha_2) := (\sqrt{2}, 1, 1) \in \mathbb{L}_3$, $v_0 := (1,1)^T$, $v_1 := (1,0)^T$ and $v_2 := (1,0)^T$. If we put any scalar $\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2$ equal to zero and considering $(\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2) \in \mathbb{L}_3$, then we can not rewrite the vector v as a linear combination of these new scalars.*

In the first moment, it looks like we can not apply any constant-rank approach for (NSOCP) in order to get a constraint qualification. However, since we already have some constraint qualifications well-defined for this context (nondegeneracy condition and Robinson's CQ), we will combine these results with the ideas given by the nonlinear programming problems. For such, let us consider the following class of problems

$$\begin{array}{lll} \text{Minimize} & f(x), & \\ \text{s.t.} & g_j(x) \in \mathbb{L}_{m_j}, & j = 1, \dots, q, \\ & h_i(x) = 0, & i = 1, \dots, p, \end{array} \tag{2.19}$$

where $f, h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, p$ and $g_j : \mathbb{R}^n \to \mathbb{R}^{m_j}$, $j = 1, \dots, q$ are twice continuously differentiable functions. Notice that this problem is essentially the problem (NSOCP) with equality constraints, once we will use some ideas from (NLP). This addition is not necessary, because given a feasible point $\overline{x}$ of (NSOCP) we could consider just the second-order constraints at the positive boundary and deal with them as nonlinear inequality constraints, through reduction mapping as presented previously. If on the one hand, we will use nonlinear programming ideas, on the other hand, we will use the results coming from Robinson's CQ. But in order to develop a formulation for constraint qualification for the problem (2.19) and, in particular, for (NSOCP), we will need an extension of AKKT condition for this context. It was developed by Andreani et al. in [AFH+19].

Theorem 2.3.2. *Let $\overline{x}$ be a local minimizer of (2.19). Then, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\mu_j^k\} \subset \mathbb{L}_{m_j}$ with $j \in I_0(\overline{x})$ and $\{\alpha_j^k\} \subset \mathbb{R}_+$ with $j \in I_B(\overline{x})$, such that*

$$\nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^k \nabla h_i(x^k) - \sum_{j \in I_0(\overline{x})} Dg_j(x^k)^T \mu_j - \sum_{j \in I_B(\overline{x})} \alpha_j \nabla \phi_j(x^k) \to 0 \tag{2.20}$$

Now we have all the tools necessary to define our first proposal of constant rank-type condition for a second-order cone programming problem. This proposal is called *naive* in the sense that we use ideas coming from nonlinear programming problems in order to give some support for the studies of constraint qualifications in second-order cone programming problems. It was proposed by us in [AHM+22a].

Definition 2.3.2. ***(Naive-RCPLD)*** *Let $\overline{x}$ be a feasible point of (2.19) and let $I \subseteq \{1, \dots, p\}$ be such that $\{\nabla h_i(\overline{x})\}_{i \in I}$ is a basis of the linear space generated by vectors $\{\nabla h_i(\overline{x})\}_{i=1}^p$. We say that the* Relaxed Constant Positive Linear Dependence (Naive - RCPLD) *condition holds at $\overline{x}$ when, for all $J \subseteq I_B(\overline{x})$, there exists a neighborhood V of $\overline{x}$ such that:*

- *$\{\nabla h_i(x)\}_{i=1}^p$ has constant rank for all $x \in V$;*
- *if the system*

$$\sum_{i \in I} \lambda_i \nabla h_i(\overline{x}) - \sum_{j \in I_0(\overline{x})} Dg_j(\overline{x})^T \mu_j - \sum_{j \in I_B(\overline{x})} \alpha_j \nabla \phi_j(\overline{x}) = 0$$

  where $\lambda_i \in \mathbb{R}$, $i \in I$; $\mu_j \in \mathbb{L}_{m_j}$, $j \in I_0(\overline{x})$; $\alpha_j \geq 0$, $j \in I_B(\overline{x})$, has a not all zero solution $(\lambda_i)_{i \in I}$, $(\mu_j)_{j \in I_0(\overline{x})}$, $(\alpha_j)_{j \in I_B(\overline{x})}$, then the vectors $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla \phi_j(x)\}_{j \in J}$ are linearly dependent for all $x \in V$.

Here is important to notice something. In [AHM+22a] we proposed initially the condition above using the reduction mapping $\tilde{\phi}_j(x) := \frac{1}{2}([g_j(x)]_0^2 - \|\widehat{g}_j(x)\|^2)$ where $j \in I_B(\overline{x})$, and now we are presenting according to (2.9), that is, $\phi_j(x) = [g_j(x)]_0 - \|\widehat{g}_j(x)\|$. However, the conditions are equivalent because we have that

$$\nabla \phi_j(\overline{x}) = \frac{1}{[g_j(\overline{x})]_0} Dg_j(\overline{x})^T \Gamma_j g_j(\overline{x}) = \frac{1}{[g_j(\overline{x})]_0} \nabla \tilde{\phi}_j(\overline{x}). \tag{2.21}$$

Furthermore, note that if $I_B(\overline{x}) = \emptyset$ and there is no equality constraints, the condition above is Robinson's CQ. In particular, we have that *naive-RCPLD* is weaker than Robinson's CQ. The word *naive* comes from the fact that we employ Robinson's CQ (which is well-defined) for the second-order constraints in which we can not reduce them to nonlinear inequality constraints in order to apply the knowledge that we already have from (NLP). The condition presented in Definition 2.3.2 also shows that we can combine different types of constraints in one constraint qualification, which can be very useful for more classes of problems.

Before we prove that Naive-RCPLD is a constraint qualification, we will present a naive extension of the Constant Rank of the Subspace Component (CRSC). This condition plays an important role in the studies of constraint qualifications in nonlinear programming problems, especially because CRSC unifies MFCQ and CRCQ conditions, in the sense that it is implied by both of them.

Definition 2.3.3. ***(Naive-CRSC)*** *Let $\overline{x}$ be a feasible point of (2.19). Define $P(\overline{x}) := \{j \in I_0(\overline{x}) \mid m_j = 1\}$ and $\tilde{I}_0(\overline{x}) := I_0(\overline{x}) \setminus P(\overline{x})$ and consider $J_-(\overline{x}) \subset I_B(\overline{x}) \cup P(\overline{x})$ as*

$$J_-(\overline{x}) := \left\{ j_0 \in I_B(\overline{x}) \cup P(\overline{x}) \;\middle|\; -\nabla\phi_{j_0}(\overline{x}) = \sum_{i=1}^{p} \lambda_i \nabla h_i(\overline{x}) - \sum_{j \in I_B(\overline{x}) \cup P(\overline{x})} \alpha_j \nabla \phi_j(\overline{x});\ \lambda_i \in \mathbb{R},\ \alpha_j \geq 0 \right\}.$$

Set $J_+(\overline{x}) := I_B(\overline{x}) \cup P(\overline{x}) \setminus J_-(\overline{x})$. Let $I \subset \{1, \dots, p\}$ and $J \subset \{J_-(\overline{x})\}$ be subsets such that $\{\nabla h_i(\overline{x})\}_{i\in I} \cup \{\nabla\phi_j(\overline{x})\}_{j\in J}$ is a basis of the linear space generated by $\{\nabla h_i(\overline{x})\}_{i=1}^p \cup \{\nabla\phi_j(\overline{x})\}_{j\in J_-(\overline{x})}$. We say that the Constant Rank of the Subspace Component (Naive - CRSC) *condition holds at $\overline{x}$ when there exists a neighborhood V of $\overline{x}$ such that:*

- *$\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla\phi_j(x)\}_{j\in J_-(\overline{x})}$ has constant rank for all $x \in V$;*
- *if the system*

$$\sum_{i\in I} \lambda_i \nabla h_i(\overline{x}) - \sum_{j\in \tilde{I}_0(\overline{x})} Dg_j(\overline{x})^T \mu_j - \sum_{j \in J \cup J_+(\overline{x})} \alpha_j \nabla\phi_j(\overline{x}) = 0$$

  where $\lambda_i \in \mathbb{R}$, $i \in I$; $\mu_j \in \mathbb{L}_{m_j}$, $j \in \tilde{I}_0(\overline{x})$; $\alpha_j \geq 0$, $j \in J_+(\overline{x})$; $\alpha_j \in \mathbb{R}$, $j \in J$ has only the trivial solution.

The following theorem shows that Naive-CRSC (and, consequently, Naive-RCPLD) are constraint qualifications. The proof is similar to the one given in [AHM+22a, Theorem 5.1], where we proved the result for Naive-RCPLD. The difference between the proofs relies on the fact that when we take a vector in spam$\{\nabla h_i(\overline{x})\}_{i\in I} \cup \{\nabla\phi_j(\overline{x})\}_{j\in J}$ instead of spam$\{\nabla h_i(\overline{x})\}_{i\in I}$, we have to take care of the signals of the scalars of the vectors $\{\nabla\phi_j(\overline{x})\}_{j\in J}$, because they must be non-negative. To prove the result, we will use AKKT condition as mentioned before.

Theorem 2.3.3. *Let $\overline{x}$ be a feasible point of (2.19) satisfying the AKKT condition 2.20 and Naive-CRSC. Then, the KKT conditions hold at $\overline{x}$. In particular, the conditions Naive-CRSC and Naive-RCPLD are constraint qualifications.*

Proof. Consider the sets $J_-(\overline{x})$, $J_+(\overline{x})$, J and I according to Definition 2.3.3. Due to the fact that $\overline{x}$ is an AKKT point, from 2.20 we know that there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\mu_j^k\} \subset \mathbb{L}_{m_j}$ with $j \in I_0(\overline{x})$ and $\{\alpha_j^k\} \subset \mathbb{R}_+$ with $j \in I_B(\overline{x})$, such that

$$\nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^k \nabla h_i(x^k) - \sum_{j\in\tilde{I}_0(\overline{x})} Dg_j(\overline{x})^T \mu_j - \sum_{j\in J_+(\overline{x})\cup J_-(\overline{x})} \alpha_j \nabla\phi_j(x^k) \to 0.$$

Since the set $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla\phi_j(x)\}_{j\in J_-(\overline{x})}$ has constant rank for all x close enough to $\overline{x}$ and, in addition, the set $\{\nabla h_i(\overline{x})\}_{i\in I} \cup \{\nabla\phi_j(\overline{x})\}_{j\in J}$ is a basis of the linear space generated by $\{\nabla h_i(\overline{x})\}_{i=1}^p \cup \{\nabla\phi_j(\overline{x})\}_{j\in J_-(\overline{x})}$, then we have that $\{\nabla h_i(x^k)\}_{i\in I} \cup \{\nabla\phi_j(x^k)\}_{j\in J}$ is linearly independent for k large enough. Furthermore, for each k large enough, by Carathéodory's Lemma we can rewrite

$$\sum_{i=1}^{p} \lambda_i^k \nabla h_i(x^k) - \sum_{j\in J_-(\overline{x})} \alpha_j \nabla\phi_j(x^k) = \sum_{i\in I} \tilde{\lambda}_i^k \nabla h_i(x^k) - \sum_{j\in J^k} \tilde{\alpha}_j \nabla\phi_j(x^k),$$

where J^k is the set J at iteration k and $\tilde{\alpha}_k \geq 0$. We obtain new scalars $\tilde{\lambda}_i$, $i \in I$ and $\tilde{\alpha}_j$, with $j \in J^k$ such that

$$\nabla f(x^k) + \sum_{i\in I} \tilde{\lambda}_i^k \nabla h_i(x^k) - \sum_{j\in \tilde{I}_0(\overline{x})} Dg_j(\overline{x})^T \mu_j - \sum_{j\in J_+(\overline{x})} \alpha_j \nabla \phi_j(x^k) - \sum_{j\in J^k} \tilde{\alpha}_j \nabla \phi_j(x^k) \to 0, \quad (2.22)$$

where the vectors $\{\nabla h_i(x^k)\}_{i\in I} \cup \{\nabla \phi_j(x^k)\}_{j\in J^k}$ are linearly independent for k large enough. One may ask about the signals of $\tilde{\alpha}_k$. If there exists an $j_0 \in J_-(\overline{x})$ such that $\alpha_{j_0} < 0$, then we have that

$$\sum_{i=1}^p \lambda_i^k \nabla h_i(x^k) - \sum_{j\in J_+(\overline{x})\cup J_-(\overline{x})} \alpha_j \nabla \phi_j(x^k) = \sum_{i\in I} \tilde{\lambda}_i^k \nabla h_i(x^k) - \sum_{j\in J^k\setminus\{j_0\}} \tilde{\alpha}_j \nabla \phi_j(x^k) - \tilde{\alpha}_{j_0} \nabla \phi_{j_0}(x^k). \quad (2.23)$$

Since $j_0 \in J_-(\overline{x})$, from the definition of $J_-(\overline{x})$ we have that

$$-\nabla \phi_{j_0}(\overline{x}) = \sum_{i=1}^p \lambda_i \nabla h_i(\overline{x}) + \sum_{j\in I_B(\overline{x})\cup P(\overline{x})} \alpha_j \nabla \phi_j(\overline{x});\ \lambda_i \in \mathbb{R},\ \alpha_j \geq 0$$

which implies

$$\tilde{\alpha}_{j_0} \nabla \phi_{j_0}(x^k) = |\alpha_{j_0}| \left(\sum_{i=1}^p \lambda_i \nabla h_i(\overline{x}) - \sum_{j\in I_B(\overline{x})\cup P(\overline{x})} \alpha_j \nabla \phi_j(\overline{x}) \right)$$

with correct signals of α_j for $j \in I_B(\overline{x}) \cup P(\overline{x}) = J_+(\overline{x}) \cup J_-(\overline{x})$. Substituting the expression above in the right side of the equation 2.23, we can apply Carathéodory's Lemma for this new linear combination and get the result in 2.22.

The set J^k may not be the same for all k. However, by the pigeonhole principle, we can consider subsequences where the sets J^{k_i} are the same for all k_i. In order to simplify notation, let us call $J := J^{k_i}$. From this point, the proof follows exactly the same steps of [AHM+22a, Theorem 5.1] □

Notice that in Definition 2.3.3 we could consider all the subsets of $E \subseteq \{1,\dots,p\}$ and $F \subseteq I_B(\overline{x}) \cup P(\overline{x})$ and request constant rank of $\{\nabla h_i(x)\}_{i\in E} \cup \{\nabla \phi_j(x)\}_{j\in F}$ for all x around $\overline{x}$ and then define Naive-CRCQ, or consider $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla \phi_j(x)\}_{j\in F}$ and then get a Naive-RCRCQ definition, but this is not the goal of this work. Later, we will define a constant rank-type condition without using the help provided by Robinson's CQ.

The following example given in [AHM+22a, Example 5.1] shows that Naive-RCPLD is strictly weaker than Robinson's CQ. Here, we just consider a minor modification in the reduction mapping.

Example 2.3.2. *Consider the second-order constraint $g : \mathbb{R} \to \mathbb{R}^2$ given by $g(x) := (g_0(x), \overline{g}(x)) = (x, x) \in \mathbb{L}_2$ and the feasible point $\overline{x} = 1$. We have that $I_0(\overline{x}) = \emptyset$ and $I_B(\overline{x}) = J_-(\overline{x}) \neq \emptyset$. In addition, considering the reduction mapping $\phi(x) = g_0(x) - |\overline{g}(x)|$ we get that $\nabla \phi(\overline{x}) = 0$. Thus, Robinson's CQ does not hold at $\overline{x}$ while Naive-RCPLD holds. Last, if we consider the space generated by $\nabla \phi(\overline{x})$, we have that $J = \emptyset$ is a basis for it. Therefore, Naive-CRSC also holds at $\overline{x}$.*

The naive CQ's, in addition to being shown to be a constraint qualification using AKKT condition, also show that it is possible to deal with two types of constraints at the same optimization problem, namely, second-order constraints and nonlinear constraints. This approach was made based on the fact that we could not deal with the constraints "purely conic", that is, $\tilde{I}_0(\overline{x}) \neq \emptyset$. Somehow, we are avoiding to deal with them. In the following chapters, we will face such constraints directly and propose new constraint qualifications without this skip.

We finish this chapter with the following figure that shows the relation among the well-known constraint qualifications for the second-order cone programming problem and the naive proposals.

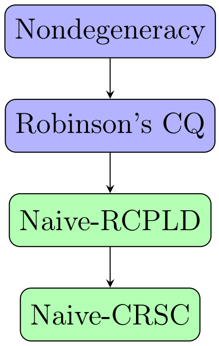

Figure 2.1: *Relation among the CQ's for (NSOCP). The boxes in blue are the well-known CQ's in nonlinear second-order cone programming and the green boxes are the naive proposals.*

Chapter 3

Sequential constraint qualifications for NSOCP

In this chapter, we will introduce new constraint qualifications for Nonlinear Second-Order Cone Programming (NSOCP). For such, we will revisit the nondegeneracy condition and propose a new point of view on it. With this in hand, we will be able to introduce weaker versions of Nondegeneracy and Robinson's CQ and, in addition, new constant rank-type constraint qualifications. The main results of this chapter are based on [AHM+22b].

3.1 Revisiting nondegeneracy condition

Let us consider the problem (NSOCP) presented in the previous chapter

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & g_j(x) \in \mathbb{L}_{m_j}, \quad j = 1, \ldots, q, \end{array}$$

and consider a feasible point $\overline{x}$. Also, let us consider for a while that $q = 1$ just to make some quick analysis. According to Definition 2.1.1, the nondegeneracy condition holds at $\overline{x}$ if the gradients of the coordinates of g at $\overline{x}$ are linearly independent, which is very similar to the LICQ for nonlinear programming problems and it is reasonable in order to define a regularity condition. However, the natural extension of a linearly independence condition is a constant rank condition. This leads one to think in a constant rank-type condition like Zhang and Zhang in [ZZ19], that is, the constant rank of the set which contains all the gradients of the coordinates of g around the point $\overline{x}$. Furthermore, even requiring a stronger condition (considering all possible subsets) we may not get a constraint qualification, which was shown in [AFH+21] and explained in the previous chapter. The situation is even harder when we take into account that Carathéodory's Lemma does not work in the second-order context (see [AFH+21, Example 1]).

All of the points mentioned in the previous paragraph rely on the fact that dealing with the second-order cone structure is not an easy task, especially if we try doing this just based on ideas that come from a nonlinear programming context. Thus, in order to avoid this issue, let us analyze the second-order cone structure in a deeper way.

Consider the m-dimensional second-order cone and let $y := (y_0, \widehat{y}) \in \mathbb{R} \times \mathbb{R}^{m-1}$ be any arbitrary vector. By definition, we have that $y_0 \geq \|\widehat{y}\|$. According to [AG03, Section 4], consider the following identity

$$y = \frac{1}{2}(y_0 - \|\widehat{y}\|) \begin{pmatrix} 1 \\ -\frac{\widehat{y}}{\|\widehat{y}\|} \end{pmatrix} + \frac{1}{2}(y_0 + \|\widehat{y}\|) \begin{pmatrix} 1 \\ \frac{\widehat{y}}{\|\widehat{y}\|} \end{pmatrix},$$

and define

$$\lambda_1(y) := y_0 - \|\widehat{y}\|,\ \lambda_2(y) := y_0 + \|\widehat{y}\| \quad \text{and} \quad u_1(y) := \frac{1}{2}\begin{pmatrix} 1 \\ -\frac{\widehat{y}}{\|\widehat{y}\|} \end{pmatrix},\ u_2(y) := \frac{1}{2}\begin{pmatrix} 1 \\ \frac{\widehat{y}}{\|\widehat{y}\|} \end{pmatrix} \tag{3.1}$$

if $\|\widehat{y}\| > 0$, and

$$u_1(y) := \frac{1}{2}\begin{pmatrix} 1 \\ -\widehat{w} \end{pmatrix},\ u_2(y) := \frac{1}{2}\begin{pmatrix} 1 \\ \widehat{w} \end{pmatrix},$$

if $\|\widehat{y}\| = 0$, where $\widehat{w} \in \mathbb{R}^{m-1}$ can be any arbitrary vector such that $\|\widehat{w}\| = 1$. The scalars $\lambda_1(y)$ and $\lambda_2(y)$ are called *eigenvalues* of y associated to the *eigenvectors* $u_1(y)$ and $u_2(y)$, respectively. Notice that we can analyze the belongingness to the second-order cone of a vector y through its eigenvalues. Indeed, we have that $y \in \mathbb{L}_m$ if, and only if, $\lambda_1(y), \lambda_2(y) \geq 0$. In addition, if both eigenvalues are strictly positive then we have that $y \in \operatorname{int}(\mathbb{L}_m)$, if $\lambda_1(y) = 0$ and $\lambda_2 > 0$, then we have that $y \in \operatorname{bd}^+(\mathbb{L}_m)$ and, lastly, if $\lambda_1(y) = \lambda_2(y) = 0$, then y is the vertex of the second-order cone. Based on this, we also can define the orthogonal projection of y onto $\mathbb{L}_m$, which is given by

$$\mathcal{P}_{\mathbb{L}_m}(y) := [\lambda_1(y)]_+ u_1(y) + [\lambda_2(y)]_+ u_2(y),$$

where $[\cdot]_+ := \max\{\cdot, 0\}$. The following figure shows the vector y and its eigenvectors related to the second-order cone.

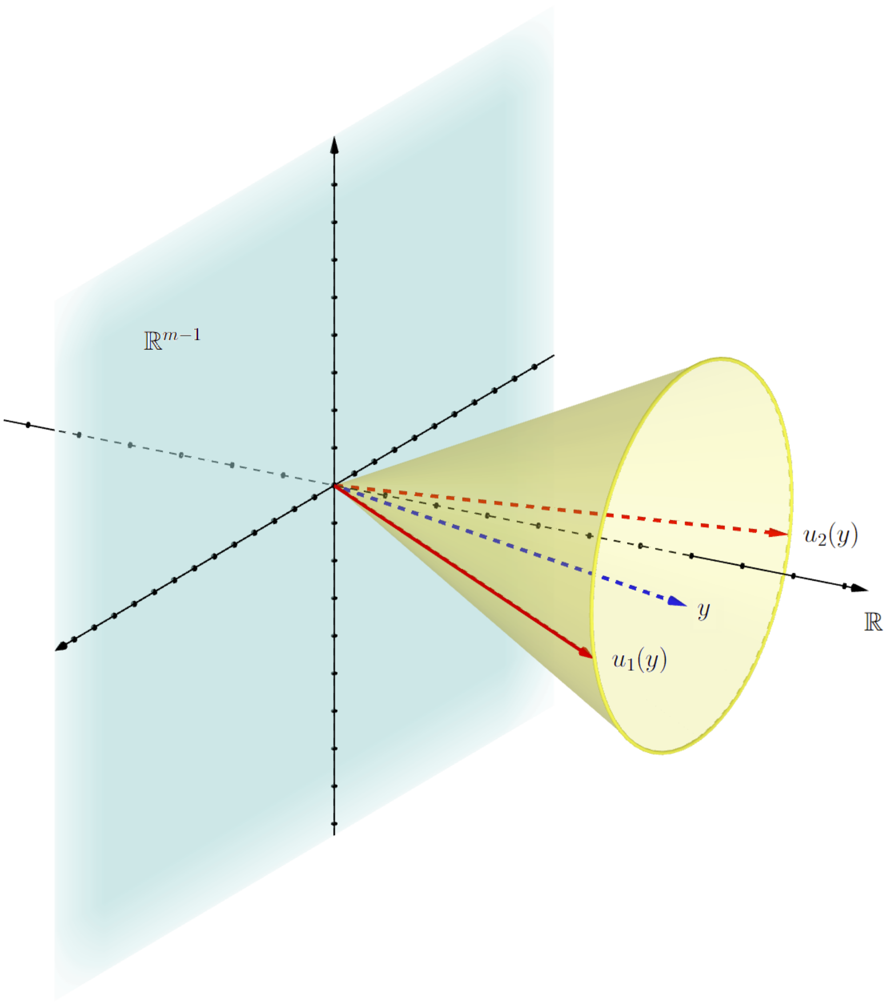

Figure 3.1: *An arbitrary vector y and its eigenvectors.*

With these ideas in mind, the goal now is rewriting the nondegeneracy condition in terms of the eigenvectors. Just to avoid overwriting, from now on we will assume that $m_j > 1$ for $j = 1, \dots, q$.

Definition 3.1.1. *Let $K \subseteq \mathbb{R}^n$ be a nonempty closed convex cone and consider a matrix $M \in \mathbb{R}^{n\times m}$. We say that M is* K-linearly independent *if given any $v \in K \setminus \{0\}$, we have that $Mv \neq 0$.*

The definition above is related to the concept of injectivity over K. We can rewrite the definition above in the following way: the matrix M is K-linearly independent, if $Mv = 0$ with $v \in K$, then $v = 0$. Notice that if we consider $K = \mathbb{R}^n$, then we get that M is injective.

The following lemma is a particular case of [AHM+22b, Lemma 2.1], which provides an equivalence in order to define $\mathbb{L}_m$-linearly independence. We will omit the proof.

Lemma 3.1.1. *Let $\mathbb{L}_m$ be the m-dimensional second-order cone. Consider the set $S := \{\widehat{w} \in \mathbb{R}^{m-1} \mid \|\widehat{w}\| = 1\}$ and, for each $\widehat{w} \in S$, consider the vectors $K_{\widehat{w}} := \{(1, -\widehat{w}), (1, \widehat{w})\}$. We have that*

$$\mathbb{L}_m = \bigcup_{w\in S} \operatorname{cone}(K_{\widehat{w}}), \tag{3.2}$$

where $\operatorname{cone}(K_{\widehat{w}})$ denotes the conic hull of $K_{\widehat{w}}$. We have that a matrix $M \in \mathbb{R}^{n\times m}$ is $\mathbb{L}_m$-linearly independent if, and only if, the vectors $\{(1, -\widehat{w}), (1, \widehat{w})\}$ are positively linearly independent, for every fixed $\widehat{w} \in S$. In addition, notice that

$$\mathbb{R}^m = \bigcup_{\substack{\widehat{w}\in\mathbb{R}^{m-1}\\ \|\widehat{w}\|=1}} \operatorname{span}(\{(1, -\widehat{w}), (1, \widehat{w})\}), \tag{3.3}$$

where $\operatorname{span}(\{(1, -\widehat{w}), (1, \widehat{w})\})$ denotes the linear span of the vectors $(1, -\widehat{w})$ and $(1, \widehat{w})$. We have that the matrix M is injective if, and only if, the vectors $\{(1, -\widehat{w}), (1, \widehat{w})\}$ are linearly independent, for every fixed $\widehat{w} \in S$.

The lemma above explains $\mathbb{L}_m$-linear independence (and the usual concept of injectivity) from a different point of view. Despite being initially defined only for one cone, we can consider the product of closed convex cones $\{\mathbb{L}_{m_j}\}_{j=1}^q$ and the family of matrices related to them $\{M_j\}_{j=1}^q$. For such, we just need to consider the cone $\mathbb{L} := \prod_{j=1}^q \mathbb{L}_{m_j}$ and the matrix M whose lines are the matrices M_j. This will be pivotal for what we will do in the second-order cone programming context. With this new tool at hand, we can rewrite nondegeneracy (and, consequently, Robinson's CQ) in terms of $\mathbb{L}$-linear independence. The reader can find more details in [BR05].

Definition 3.1.2. *Let $\overline{x}$ be a feasible point of (NSOCP). We say that*

- Nondegeneracy condition *holds at $\overline{x}$ if the family*

$$\left\{Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x})\right\}_{j\in I_B(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^T\right\}_{j\in I_0(\overline{x})} \tag{3.4}$$

  is $\mathbb{R}^{|I_B(\overline{x})|} \times \prod_{j\in I_0(\overline{x})} \mathbb{R}^{m_j}$-linearly independent;

- Robinson's CQ *holds at $\overline{x}$ if the family 3.4 is $\mathbb{R}_+^{|I_B(\overline{x})|} \times \prod_{j\in I_0(\overline{x})} \mathbb{L}_{m_j}$-linearly independent.*

Finally, we are able to write the nondegeneracy condition and Robinson's CQ in a way that takes into account the eigenvectors of the second-order cone. Here, we will present a different proof from [AHM+22b, Corollary 3.1] because we will be able to get a "hint" related to the eigenstructure of (NSOCP). For such, let us consider the following result that follows directly from Lemma 3.1.1 and Definition 3.1.2.

Proposition 3.1.1. *Let $\overline{x}$ be a feasible point of (NSOCP). We say that*

- Nondegeneracy condition *holds at $\overline{x}$ if, and only if,*

$$\left\{Dg_j(\overline{x})^T u_1(g_j(\overline{x}))\right\}_{j\in I_B(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^T(1,-\widehat{w}_j), Dg_j(\overline{x})^T(1,\widehat{w}_j)\right\}_{j\in I_0(\overline{x})} \tag{3.5}$$

  is linearly independent for every $\widehat{w}_j \in \mathbb{R}^{m_j-1}$ such that $\|\widehat{w}_j\| = 1$, $j \in I_0(\overline{x})$;

- Robinson's CQ *holds at $\overline{x}$ if, and only if, the family (3.5) is positively linearly independent for every $\widehat{w}_j \in \mathbb{R}^{m_j-1}$ such that $\|\widehat{w}_j\| = 1$, $j \in I_0(\overline{x})$.*

Proof. Since the proof of nondegeneracy condition and Robinson's CQ are quite similar in this approach and the first one is easier, we will prove nondegeneracy and give more details for better comprehension. Let $\overline{x}$ be a feasible point of (NSOCP) such that

$$\left\{Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x})\right\}_{j\in I_B(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^T\right\}_{j\in I_0(\overline{x})}$$

is $\mathbb{R}^{|I_B(\overline{x})|} \times \prod_{j\in I_0(\overline{x})} \mathbb{R}^{m_j}$-linearly independent. For each $j \in I_0(\overline{x})$ take a vector $\widehat{w}_j \in \mathbb{R}^{m_j-1}$ such that $\|\widehat{w}_j\| = 1$. Assume that there are scalars $\alpha_j, \beta_j, j \in I_0(\overline{x})$ and $\gamma_j, j \in I_B(\overline{x})$, not all of them simultaneously zero such that

$$\sum_{j\in I_0(\overline{x})} \alpha_j Dg_j(\overline{x})^\top(1,\widehat{w}_j) + \beta_j Dg_j(\overline{x})^\top(1,-\widehat{w}_j) + \sum_{j\in I_B(\overline{x})} \gamma_j Dg_j(\overline{x})^\top u_1(g_j(\overline{x})) = 0.$$

Rearranging the terms above and recalling the expression given in (3.1) to compute $u_1(g_j(\overline{x}))$ we get

$$\sum_{j\in I_0(\overline{x})} (\alpha_j+\beta_j)\nabla g_{j,0}(\overline{x}) + \sum_{j\in I_0(\overline{x})} (\beta_j-\alpha_j) D\widehat{g}_j(\overline{x})^\top \widehat{w}_j + \frac{1}{2}\sum_{j\in I_B(\overline{x})} \gamma_j Dg_j(\overline{x})^T \left(1, -\frac{\widehat{g}_j(\overline{x})}{\|\widehat{g}_j(\overline{x})\|}\right) = 0.$$

In addition, since $g_{j,0}(\overline{x}) = \|\widehat{g}_j(\overline{x})\|$ for $j \in I_B(\overline{x})$ and $\frac{1}{2}\left(1, -\frac{\widehat{g}_j(\overline{x})}{\|\widehat{g}_j(\overline{x})\|}\right) = \frac{1}{2\|\widehat{g}_j(\overline{x})\|}\Gamma_j g_j(\overline{x})$, it follows that

$$\sum_{j\in I_0(\overline{x})} (\alpha_j+\beta_j)\nabla g_{j,0}(\overline{x}) + \sum_{j\in I_0(\overline{x})} (\beta_j-\alpha_j) D\widehat{g}_j(\overline{x})^\top \widehat{w}_j + \frac{1}{2}\sum_{j\in I_B(\overline{x})} \frac{\gamma_j}{\|\widehat{g}_j(\overline{x})\|} Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x}) = 0.$$

Due to the fact that $\|\widehat{w}_j\| = 1$ and we have that (3.4) is linearly independent, we obtain $\alpha_j = \beta_j = \gamma_j = 0$ for all j.

Now assume that $\overline{x}$ is such that

$$\left\{Dg_j(\overline{x})^T u_1(g_j(\overline{x}))\right\}_{j\in I_B(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^T(1,-\widehat{w}_j), Dg_j(\overline{x})^T(1,\widehat{w}_j)\right\}_{j\in I_0(\overline{x})} \tag{3.6}$$

is linearly independent for every $\widehat{w}_j \in \mathbb{R}^{m_j-1}$ such that $\|\widehat{w}_j\| = 1$, $j \in I_0(\overline{x})$. Suppose, by absurd, that there are vectors $v_j \in \mathbb{R}^{m_j}, j \in I_0(\overline{x})$ and $\theta_j \in \mathbb{R}, j \in I_B(\overline{x})$ not all of them simultaneously zero such that

$$\sum_{j\in I_0(\overline{x})} Dg_j(\overline{x})^T v_j + \sum_{j\in I_B(\overline{x})} \theta_j Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x}) = 0.$$

Again, using the fact that

$$\begin{aligned} \theta_j Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x}) = \theta_j Dg_j(\overline{x})^T(g_{j,0}(\overline{x}), -\widehat{g}_j(\overline{x})) &= \|\widehat{g}_j(\overline{x})\|\theta_j Dg_j(\overline{x})^T\left(1, -\frac{\widehat{g}_j(\overline{x})}{\|\widehat{g}_j(\overline{x})\|}\right) \\ &= 2\|\widehat{g}_j(\overline{x})\|\theta_j Dg_j(\overline{x})^T u_1(g_j(\overline{x})) \end{aligned}$$

where $g_{j,l}$ the l-th coordinate of g_j, we obtain

$$\sum_{j\in I_0(\overline{x})} Dg_j(\overline{x})^T v_j + \sum_{j\in I_B(\overline{x})} 2\|\widehat{g}_j(\overline{x})\|\theta_j Dg_j(\overline{x})^T u_1(g_j(\overline{x})) = 0.$$

Using the fact that $\|\widehat{g}_j(\overline{x})\| \neq 0$ for all $j \in I_B(\overline{x})$ and (3.6) is linearly independent, then we obtain that there is at least one vector $v_j \neq 0$. Defining $\eta_j = \frac{1}{2}(v_{j,0} - \|\widehat{v}_j\|)$, $\vartheta = \frac{1}{2}(v_{j,0} + \|\widehat{v}_j\|)$ and $\widehat{w}_j = \frac{\widehat{v}_j}{\|\widehat{v}_j\|}$ if $\|\widehat{v}_j\| \neq 0$ or $\widehat{w}_j$ as being any unitary vector otherwise, we get

$$\sum_{j\in I_0(\overline{x})} Dg_j(\overline{x})^T v_j = \sum_{j\in I_0(\overline{x})} \eta_j Dg_j(\overline{x})^T(1, -\widehat{w}_j) + \sum_{j\in I_0(\overline{x})} \vartheta_j Dg_j(\overline{x})^T(1, \widehat{w}_j).$$

Finally, we obtain

$$\sum_{j\in I_0(\overline{x})} \eta_j Dg_j(\overline{x})^T(1, -\widehat{w}_j) + \sum_{j\in I_0(\overline{x})} \vartheta_j Dg_j(\overline{x})^T(1, \widehat{w}_j) + \sum_{j\in I_B(\overline{x})} 2\|\widehat{g}_j(\overline{x})\|\theta_j Dg_j(\overline{x})^T u_1(g_j(\overline{x})) = 0.$$

Since (3.6) is linearly independent, we must have $\eta_j = \vartheta_j = \theta_j = 0$ for every j, which implies that v_j must be equal to zero for every j due to the definition of η_j and ϑ_j, which is a contradiction. □

3.2 Eigenvectors and constraint qualifications

Based on the previous section, we will introduce new constraint qualifications using the results given in Proposition 3.1.1. In order to build these new constraint qualifications under the light of Proposition 3.1.1, consider the problem (NSOCP when $q = 1$ and $g(\overline{x}) = 0$ in. Assume that $\overline{x}$ is nondegenerate. According to Definition 2.1.1, we have that the columns of $Dg(\overline{x})^T$ are linearly independent, that is, given scalars $\alpha_i \in \mathbb{R}$, $i = 0, \ldots, m-1$, the equation

$$\sum_{i=0}^{m-1} \alpha_i \nabla g_i(\overline{x}) = 0$$

implies that $\alpha_0 = \ldots = \alpha_{m-1} = 0$, where $g = (g_0, \ldots, g_{m-1})$. In this case, we are analyzing the linear independence of m vectors, as usual. Under the light of Proposition 3.1.1, the point $\overline{x}$ is nondegenerate if, and only if, given $\alpha, \beta \in \mathbb{R}$ such that

$$\alpha Dg(\overline{x})^T(1, -\widehat{w}) + \beta Dg(\overline{x})^T(1, \widehat{w}) = 0, \tag{3.7}$$

we have that $\alpha = \beta = 0$ for every $\widehat{w} \in \mathbb{R}^{m-1}$ such that $\|\widehat{w}\| = 1$. In this second case, we are analyzing the linear independence of two vectors for each fixed unitary vector $\widehat{w} \in \mathbb{R}^{m-1}$. If we analyze the equation (3.7) in a deeper way, we obtain the following equation

$$(\alpha + \beta)\nabla g_0(\overline{x}) + (\beta - \alpha)D\widehat{g}(\overline{x})^T\widehat{w} = 0, \tag{3.8}$$

where $\widehat{g} = (g_1, \ldots, g_{m-1})$. Taking a look at equation (3.8), we can observe that the nondegeneracy condition relates the "linear independence between $\nabla g_0(\overline{x})$ and $D\widehat{g}(\overline{x})^T$", which is expected once we know that all the gradients are linearly independent. However, the information about the linear independence of the coordinates of $\widehat{g}$ is kind of hidden if we analyze the nondegeneracy condition only under the light of Proposition 3.1.1. Indeed, just when we vanish the vectors $\widehat{w} \in \mathbb{R}^{m-1}$ such that $\|\widehat{w}\| = 1$ we may obtain some information about the linear independence of the columns of $D\widehat{g}(\overline{x})^T$. For example, if for every unitary vector $\widehat{w} \in \mathbb{R}^{m-1}$ we have that $D\widehat{g}(\overline{x})^T\widehat{w} \neq 0$, it means that the columns of $D\widehat{g}(\overline{x})^T$ are linearly independent. On the other hand, if there exists a unitary vector $\widehat{w} \in \mathbb{R}^{m-1}$ such that $D\widehat{g}(\overline{x})^T\widehat{w} = 0$, then we can obtain a non-trivial solution of (3.8) and nondegeneracy fails at $\overline{x}$ but it does not necessarily impossibility the fullness of weak-nondegeneracy (see Example 3.2.1 for more details).

In order to finish this little discussion, we would like to point out that the second part of the proof of Proposition 3.1.1 showed us that despite choosing all unitary vectors $\overline{w}_j \in \mathbb{R}^{m_j-1}$, there are some vectors that are more important than others. Indeed, following similar ideas given in [AHMR23b], we can propose new constraint qualifications where only the limit points of sequences consisting of eigenvectors of $g(x^k)$ are already enough.

Definition 3.2.1 (Weak-nondegeneracy and weak-Robinson's CQ [AHM+22b])**.** *Let $\overline{x}$ be a feasible point of (NSOCP). We say that:*

- Weak-nondegeneracy *holds at $\overline{x}$ if, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there is an infinite subset $I \subseteq \mathbb{N}$ and convergent eigenvectors sequences $\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, -\widehat{w}_j)$ and $\{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, \widehat{w}_j)$, with $\widehat{w}_j \in \mathbb{R}^{m_j-1}$ and $\|\widehat{w}_j\| = 1$, for every $j \in I_0(\overline{x})$, such that*
$$\left\{Dg_j(\overline{x})^T u_1(g_j(\overline{x}))\right\}_{j\in I_B(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^T (1, -\widehat{w}_j), Dg_j(\overline{x})^T (1, \widehat{w}_j)\right\}_{j\in I_0(\overline{x})} \tag{3.9}$$
is linearly independent;
- Weak-Robinson's CQ *holds at $\overline{x}$ if, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there is an infinite subset $I \subseteq \mathbb{N}$ and convergent eigenvectors sequences $\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, -\widehat{w}_j)$ and $\{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, \widehat{w}_j)$, for every $j \in I_0(\overline{x})$, such that (3.9) is positively linearly independent.*

These conditions are proved to be constraint qualifications in [AHM+22b]. By definition we have that weak-nondegeneracy and weak-Robinson are weaker than nondegeneracy and Robinson's CQ, respectively. In order to establish more details among their relation, let us remember the brief discussion of some paragraphs above where we explored the problem (NSOCP) with only one second-order cone constraint at a feasible point $\overline{x}$ such that $g(\overline{x}) = 0$. As mentioned previously, in order to make nondegeneracy fail and keep the fullness of weak-nondegeneracy, we need to have a situation where (at least) $\nabla g_0(\overline{x})$ does not belong the the space generated by the columns of $D\widehat{g}(\overline{x})^T$ and require that $D\widehat{g}(\overline{x})^T$ does not have full rank.

Example 3.2.1. *Consider the following feasible set $\Omega = \{x = (x_1, x_2) \in \mathbb{R}^2 \mid g(x) \in \mathbb{L}_3\}$ where $g(x) := (\exp^{x_1} - 1, 3\sin(x_2), 4\sin(x_2))$ at the point $\overline{x} = (0, 0)$. We have that $g(\overline{x}) = (0, 0, 0)$. Take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$. Since x_2 converges to 0, for all k large enough we have that $\widehat{g}(x_2) = (0, 0)$ if, and only if, $x_2 = 0$.*

If there exists a subsequence $\{x^k\}_{k\in S}$ where S is an infinite subset of $\mathbb{N}$ such that $x_2^k \neq 0$, then the eigenvectors of $g(x^k)$ are uniquely determined. It follows that $\|\widehat{g}(x^k)\| = 5\sin(x_2^k)$ and

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -\frac{3}{5}, -\frac{4}{5}\right) \text{ and } u_2(g(x^k)) = \frac{1}{2}\left(1, \frac{3}{5}, \frac{4}{5}\right),$$

for all $k \in S$ (remember that $\sin(x_2^k) \neq 0$ if $x_2^k \neq 0$ and x_2^k is close enough to zero). Defining $\widehat{w} := \left(\frac{3}{5}, \frac{4}{5}\right)$ we get that

$$\lim_{k\in S} u_1(g(x^k)) = \frac{1}{2}(1, -\widehat{w}) \text{ and } \lim_{k\in S} u_2(g(x^k)) = \frac{1}{2}(1, \widehat{w}),$$

and, moreover, since

$$Dg(x)^T = \begin{bmatrix} \exp^{x_1} & 0 & 0 \\ 0 & 3\cos(x_2) & 4\cos(x_2) \end{bmatrix} \Rightarrow Dg(\overline{x})^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 4 \end{bmatrix}$$

we obtain that

$$Dg(\overline{x})^\top(1, -\widehat{w}) = \frac{1}{2}\begin{pmatrix} 1 \\ -5 \end{pmatrix} \quad \text{and} \quad Dg(\overline{x})^\top(1, \widehat{w}) = \frac{1}{2}\begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

are linearly independent.

On the other hand, if there exists a subsequence $\{x^k\}_{k\in S}$ such that $x_2^k = 0$ the eigenvectors of $g(x^k)$ are not uniquely determined. From the computation of $Dg(\overline{x})^T$ we obtain that nondegeneracy does not hold at $\overline{x}$. From the proof of Proposition 3.1.1 we know that must exists a vector $\widehat{w} \in \mathbb{R}^2$ with $\|\widehat{w}\| = 1$ such that $Dg(\overline{x})^T(1,-\widehat{w})$ and $Dg(\overline{x})^T(1,\widehat{w})$ are not linearly independent. Indeed, it is enough to take any unitary vector $\widehat{w} \in \ker(D\widehat{g}(\overline{x})^T)$, which exists due to the fact that $D\widehat{g}(\overline{x})^T$ does not have full rank. Thus, avoiding the eigenvectors that belong to the kernel of $D\widehat{g}(\overline{x})^T$ we will obtain limit points such that $Dg(\overline{x})^T(1,-\widehat{w})$ and $Dg(\overline{x})^T(1,\widehat{w})$ are linearly independent and then weak-nondegeneracy holds at $\overline{x}$.

The next point is to find "the difference" between nondegeneracy and weak-nondegeneracy, that is, what must be required additionally to weak-nondegeneracy in order to get nondegeneracy condition. From the previous discussion, we already know the answer for the case with only one second-order constraint: the surjectivity of $D\widehat{g}(\overline{x})^T$. Indeed, this follows directly from the relation established in 3.8 and the definition of the nondegeneracy condition. The remaining issue now is the multifold case, which is answered in the following proposition and the proof can be found in [AHM+22b, Proposition 3.1].

Proposition 3.2.1. *Let $\overline{x}$ be a feasible point of (NSOCP). The nondegeneracy condition holds at $\overline{x}$ if, and only if, weak-nondegeneracy holds at $\overline{x}$ and, moreover, the matrix*

$$M := \begin{bmatrix} \vdots \\ D\widehat{g}_j(\overline{x}) \\ \vdots \end{bmatrix}_{j\in I_0(\overline{x})}$$

is surjective.

In order to establish the relation between weak-nondegeneracy and Robinson's CQ (and weak-Robinson as well), we will analyze Proposition 3.1.1 when Robinson's CQ holds at a feasible point $\overline{x}$ in a second-order cone programming problem with only one constraint $g(x) \in \mathbb{L}_m$ where $g(\overline{x}) = 0$, in a similar vein that we did for nondegeneracy condition.

If Robinson's CQ holds at $\overline{x}$, given non-negative scalars α and β, we have that the vectors $Dg(\overline{x})^T(1,-\widehat{w})$ and $Dg(\overline{x})^T(1,\widehat{w})$ are positively linearly independent, that is, the following equation

$$\alpha Dg(\overline{x})^T(1,-\widehat{w}) + \beta Dg(\overline{x})^T(1,\widehat{w}) = 0,$$

which can be rewritten as

$$(\alpha+\beta)\nabla g_0(\overline{x}) + (\beta-\alpha)D\widehat{g}(\overline{x})^T\widehat{w} = 0,$$

only admits the solution $\alpha = \beta = 0$. In the first moment one could expect a similar behavior between $\nabla g_0(\overline{x})$ and $D\widehat{g}(\overline{x})^T\widehat{w}$ for a fixed unitary vector $\widehat{w} \in \mathbb{R}^{m-1}$, as we obtained in the weak-nondegeneracy case (the linear independence between them). The more natural result would be to get the positive linear independence between these vectors. However, the following example shows that it may not hold.

Example 3.2.2. *Consider the following constraint $g : \mathbb{R} \to \mathbb{R}^2$ given by $g(x) = (2x, -x)$ at the point $\overline{x} = 0$. The eigenvectors are $\widehat{w} = \pm 1$. Let us consider $\widehat{w} = 1$ and take scalars $\alpha, \beta \geq 0$. It follows that*

$$\alpha Dg(\overline{x})^T(1,-\widehat{w}) + \beta Dg(\overline{x})^T(1,\widehat{w}) = 0$$

give us

$$\begin{aligned} \alpha(2,-1)\begin{pmatrix} 1 \\ -1 \end{pmatrix} + \beta(2,-1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} &= 0 \\ 3\alpha + \beta &= 0, \end{aligned}$$

which only admits the solution $\alpha = \beta = 0$. *Thus, we obtain that the vectors* $Dg(\overline{x})^T(1, -\widehat{w})$ *and* $Dg(\overline{x})^T(1, \widehat{w})$ *are positively linearly independent. On the other hand,*

$$\alpha \nabla g_0(\overline{x}) + \beta D\widehat{g}(\overline{x})^T \widehat{w} \quad = \quad 0,$$

give us $2\alpha - \beta = 0$, *that is, the vectors* $\nabla g_0(\overline{x})$ *and* $D\widehat{g}(\overline{x})^T \widehat{w}$ *are not positively linearly independent.*

With this in mind, in order to find the relation between weak-nondegeneracy and Robinson's CQ (and weak-Robinson) we can explore the fact that $\nabla g_0(\overline{x})$ and $D\widehat{g}(\overline{x})^T \widehat{w}$ may be linearly dependent. Moreover, we will use another characterization of Robinson's CQ for this purpose. Consider the problem (NSOCP) and a feasible point $\overline{x}$. The Robinson's CQ holds at $\overline{x}$ if there exists $d \in \mathbb{R}^n$ such that $g_j(\overline{x}) + Dg_j(\overline{x})d \in \operatorname{int}(\mathbb{L}_{m_j})$ for every j. When we have only one constraint $g(x) \in \mathbb{L}_m$ at a feasible point $\overline{x}$ such that $g(\overline{x}) = 0$, the fulfillness of Robinson's CQ means the existence of a direction $d \in \mathbb{R}^n$ in which $Dg(\overline{x})d \in \operatorname{int}\mathbb{L}_m$, that is,

$$\langle \nabla g_0(\overline{x}), d\rangle > \|D\widehat{g}(\overline{x})d\|.$$

Somehow the inequality above shows that the achievement of Robinson's CQ is due to the fact the "magnitude" of $\nabla g_0(\overline{x})$ is strictly greater than the "magnitude" of $D\widehat{g}(\overline{x})$. With this information and hints at hand, we can finally show the relation between weak-nondegeneracy and Robinson's CQ (and consequently weak-Robinson).

Example 3.2.3. *Consider the constraint* $g : \mathbb{R} \to \mathbb{R}^3$ *given by* $g(x) = (\tan(x), \sin\left(\frac{x}{4}\right), \sin\left(\frac{x}{4}\right))$ *at the point* $\overline{x} = 0$. *We have that* $g(\overline{x}) = (0, 0, 0)$ *and*

$$\nabla g_0(x) = \sec^2(x), \;\; D\widehat{g}(x) = \left(\frac{1}{4}\cos\left(\frac{x}{4}\right), \frac{1}{4}\cos\left(\frac{x}{4}\right)\right).$$

In particular, when we consider the point $\overline{x}$ *we get* $\nabla g_0(\overline{x}) = 1$ *and* $D\widehat{g}(\overline{x}) = \left(\frac{1}{4}, \frac{1}{4}\right)$. *If we consider* $d = 1$ *we obtain* $\nabla g_0(\overline{x})d > \|D\widehat{g}(\overline{x})d\|$ *and then Robinson's CQ (and consequently weak-Robinson) holds at* $\overline{x}$.

Now take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ *in which* $x^k \neq 0$ *for all* k *large enough. Thus, we get that* $\widehat{g}(x^k) \neq (0, 0)$ *and then the eigenvectors of* $g(x^k)$ *are uniquely determined and they are given by*

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) \text{ and } u_2(g(x^k)) = \frac{1}{2}\left(1, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right).$$

Define $\widehat{w} := \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$. *It follows that*

$$\lim_{k\to\infty} u_1(g(x^k)) = \frac{1}{2}(1, -\widehat{w}) \text{ and } \lim_{k\to\infty} u_2(g(x^k)) = \frac{1}{2}(1, \widehat{w}),$$

and, in addition,

$$Dg(\overline{x})^\top(1, -\widehat{w}) = \frac{4-\sqrt{2}}{4} > 0 \quad \text{and} \quad Dg(\overline{x})^\top(1, \widehat{w}) = \frac{4+\sqrt{2}}{4} > 0$$

are linearly dependent. Therefore, weak-nondegeneracy does not hold at $\overline{x}$.

To finish this discussion about the relation among the CQ's, we need to establish what happens between Robinson's CQ and weak-Robinson. Before presenting the result that we have at hand, let us remember some points that differ from the nondegeneracy case from Robinson's CQ. To fix the ideas, consider the programming problem with only one second-order constraint $g(x) \in \mathbb{L}_m$ at a feasible point $\overline{x} \in \mathbb{R}^n$ such that $g(\overline{x}) = 0$. When we take nondegeneracy condition, we know that $\nabla g_0(\overline{x})$ does not belong to the image of $D\widehat{g}(\overline{x})^T$, which can also be provided by weak-nondegeneracy condition as discussed previously. Furthermore, from the nondegeneracy condition, we know that $D\widehat{g}(\overline{x})^T$ has full column rank, and this is pivotal in order to build the difference with weak-nondegeneracy, once we can have the fulfillness of weak-nondegeneracy without this condition as we showed in Example 3.2.1.

On the other hand, even if we have the positive linear independence between $Dg(\overline{x})^T(1,-\widehat{w})$ and $Dg(\overline{x})^T(1,\widehat{w})$, which is part of the condition of weak-Robinson, we may not have similar information about $\nabla g_0(\overline{x})$ and $D\widehat{g}(\overline{x})^T\widehat{w}$ as we showed in Example 3.2.2. Moreover, if Robinson's CQ holds at $\overline{x}$, we do not have any information about the rank of $D\widehat{g}(\overline{x})$. Indeed, we can have a case where $D\widehat{g}(\overline{x})^T$ has full column-rank like $g(x) = x$, or a case where $D\widehat{g}(\overline{x})^T$ is null as $g(x) = (x_1, 0, \ldots, 0)$. In both cases, Robinson's CQ holds at $\overline{x} = 0$ and the rank varies from zero to complete. With this information in mind, the best result that we could get is the following one and the proof can be found in [AHM[+]22b, Theorem 3.1].

Theorem 3.2.1. *Consider the problem (NSOCP). If $q = 1$, that is, if we have only one second-order constraint $g(x) \in \mathbb{L}_m$, then weak-Robinson CQ holds at $\overline{x}$ if, and only if, Robinson's CQ also holds. Moreover, if $q > 1$ and weak-Robinson CQ holds at $\overline{x}$, then for each constraint $g_j, j \in \{1, \ldots, q\}$ we have that Robinson's CQ holds at $\overline{x}$ separately.*

Since we already have a new formulation for the nondegeneracy condition based on the eigenvectors of the second-order cone, we can now introduce a constant rank-type constraint qualification for problem (NSOCP). For such, consider the following notation

$$\begin{aligned}\mathcal{J}_{J_B,J_0^-,J_0^+}(x,\widehat{w}) \quad &:= \quad \left\{Dg_j(x)^\top u_1(g_j(x))\right\}_{j\in J_B} \bigcup \left\{Dg_j(x)^\top(1,-\widehat{w}_j)\right\}_{j\in J_0^-} \\ & \qquad \bigcup \left\{Dg_j(x)^\top(1,\widehat{w}_j)\right\}_{j\in J_0^+} \end{aligned} \tag{3.10}$$

with $\widehat{w} = [\widehat{w}_j]_{j\in J_0^-\cup J_0^+}$, $J_B \subseteq I_B(\overline{x})$, and $J_+, J_- \subseteq I_0(\overline{x})$, where J_+ is related to $(1,\widehat{w})$ and J_- to $(1,-\widehat{w})$.

Definition 3.2.2 (weak-CRCQ and weak-CPLD)**.** *We say that a feasible point $\overline{x}$ of (NSOCP) satisfies:*

- Weak constant rank constraint qualification *(weak-CRCQ) if for every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $I \subseteq_\infty \mathbb{N}$, and convergent eigenvector sequences*

$$\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,-\widehat{w}_j) \quad \text{and} \quad \{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,\widehat{w}_j),$$

with $\|\widehat{w}_j\| = 1$, for all $j \in I_0(\overline{x})$, such that for all subsets $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$, we have that: if the family of vectors $\mathcal{J}_{J_B,J_-,J_+}(\overline{x},\widehat{w})$ is linearly dependent, then $\mathcal{J}_{J_B,J_-,J_+}(x^k,\widehat{w}^k)$ remains linearly dependent for all $k \in I$ large enough, where $\widehat{w} = [\widehat{w}_j]_{j\in J_-\cup J_+}$ and $\widehat{w}^k = [\widehat{w}^k_j]_{j\in J_-\cup J_+}$ satisfies

$$u_1(g_j(x^k)) = \frac{1}{2}(1,-\widehat{w}^k_j) \quad \text{and} \quad u_2(g_j(x^k)) = \frac{1}{2}(1,\widehat{w}^k_j) \tag{3.11}$$

for every $j \in J_- \cup J_+$.

- Weak constant positive linear dependence *(weak-CPLD) condition if for every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there is some $I \subseteq_\infty \mathbb{N}$, and convergent eigenvector sequences*

$$\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,-\widehat{w}_j) \quad \text{and} \quad \{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,\widehat{w}_j),$$

with $\|\widehat{w}_j\| = 1$, for all $j \in I_0(\overline{x})$, such that for all subsets $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$, we have that: if $\mathcal{J}_{J_B,J_-,J_+}(\overline{x},\widehat{w})$ is positively linearly dependent, then $\mathcal{J}_{J_B,J_-,J_+}(x^k,\widehat{w}^k)$ is linearly dependent for all $k \in I$ large enough, where $\widehat{w}$ and $\widehat{w}^k$ are as above.

First of all, notice that by definition we have that weak-CPLD is weaker than weak-CRCQ. Also, we have the impression that the definition above seems too demanding, once we require a statement related to all possible combinations of subsets of $I_B(\overline{x})$ and $I_0(\overline{x})$. However, when we look at weak-nondegeneracy condition (weak-Robinson), keep in mind that even if we not require explicitly any information about every subset $\mathcal{J}_{J_B,J_0^-,J_0^+}(x,\widehat{w})$, we require (positive) linear independence for the set that contains all index and all vectors of (3.9). This means that we have (positive) linear independence for every subset $\mathcal{J}_{J_B,J_0^-,J_0^+}(x,\widehat{w})$, that is, we have by definition that weak-nondegeneracy is stronger than weak-CRCQ and, consequently, than weak-CPLD, and also that weak-Robinson is stronger than weak-CPLD. Now let us establish the relation among weak-nondegeneracy, weak-Robinson, and the CQ's proposed above:

Example 3.2.4. *Consider the constraint $g:\mathbb{R}\to\mathbb{R}^3$ given by $g(x) := (\sin(x)-1, 6\sin(x), 8\sin(x))$ at the feasible point $\overline{x} := 0$. It follows that $g_0(x) = \cos(x)-1$, $\widehat{g}(x) = (6\sin(x), 8\sin(x))$, $\nabla g_0(\overline{x}) = -1$ and $D\widehat{g}(\overline{x})^T = (6,8)$*

We already know from previous discussions how to quickly identify if a single constraint at the vertex of the second-order cone satisfies or not weak-nondegeneracy and weak-Robinson. In this case, we have that $\nabla g_0(\overline{x})$ belongs to the image of $D\widehat{g}(\overline{x})^T$, and then weak-nondegeneracy does not hold at $\overline{x}$. Moreover, the "magnitude" of $D\widehat{g}(\overline{x})^T$ it is bigger than the "magnitude" of $\nabla g_0(\overline{x})$ when we look at the norms of their respective derivatives, and than weak-Robinson also does not hold at $\overline{x}$. However, let us prove this affirmation properly.

Take any sequence $\{x^k\}_{k\in\mathbb{N}}\to\overline{x}$ and assume without loss of generality that there exists a subsequence $\{x^k\}_{k\in S}$ in which $x^k\neq 0$ for all $k\in S$. It follows that the eigenvectors of $\widehat{g}(x^k)$ are uniquely determined and they are given by:

$$u_1(g(x^k)) = \frac{1}{2}\left(1,-\frac{3}{5},-\frac{4}{5}\right) \text{ and } u_2(g(x^k)) = \frac{1}{2}\left(1,\frac{3}{5},\frac{4}{5}\right).$$

Now, defining $\widehat{w}^k = \widehat{w} = \left(\frac{3}{5},\frac{4}{5}\right)$, it give us

$$Dg(\overline{x})^\top(1,-\widehat{w}) = \langle(-1,6,8),(1,-\tfrac{3}{5},-\tfrac{4}{5})\rangle = -11 < 0,$$
$$Dg(\overline{x})^\top(1,\widehat{w}) = \langle(-1,6,8),(1,\tfrac{3}{5},\tfrac{4}{5})\rangle = 9 > 0.$$

Since these vectors have different signals, then they are positively linearly dependent. Thus, neither weak-Robinson nor weak-nondegeneracy holds at $\overline{x}$, as expected. Moreover, when we consider any subset of $\{Dg(x^k)^\top(1,-\widehat{w}^k), Dg(x^k)^\top(1,\widehat{w}^k)\}$ we have that the rank remains the same, once we are considering $\widehat{w}^k = \widehat{w}$ for all $k\in S$. Therefore, weak-CRCQ (and, consequently, weak-CPLD) holds at $\overline{x}$.

In order to build an example such that Robinson's CQ does not imply weak-CRCQ, let us consider a case where $Dg(\overline{x})^T(1,-\widehat{w})$ and $Dg(\overline{x})^T(1,\widehat{w})$ are LD but they are limits of PLI eigenvectors.

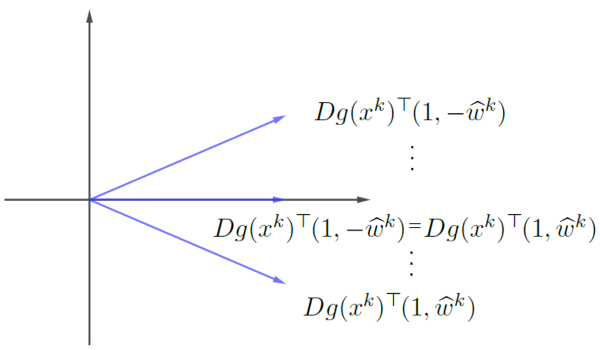

Figure 3.2: *Two LI eigenvectors converging to two LD eigenvectors.*

Moreover, such an example must have at least two variables. Indeed, otherwise, we would get that $Dg(\overline{x})^T(1,-\widehat{w}) \pm 0$ and, by the Theorem of the permanence of the signal we would have that $Dg(x^k)^\top(1, -\widehat{w}^k)$ has the same signal for k large enough, and the same result for the other eigenvector.

With these ideas in mind, we need to find a function $g : \mathbb{R}^2 \to \mathbb{R}^m$, $m \geq 2$ in which $\widehat{g}(\overline{x}) = 0$, $D\widehat{g}(\overline{x}) = 0$ but $D\widehat{g}(x) \neq 0$ for every x close to $\overline{x}$ ($\widehat{g}(x) := \|x\|^2$, for example). Also, remember that in order to make Robinson's CQ holds at the point $\overline{x}$, we need to consider a function $g_0(x)$ such that the norm of its derivative at $\overline{x}$ be greater than the norm of $D\widehat{g}(\overline{x})$. Finally, we can present the following example.

Example 3.2.5. *Consider the constraint $g : \mathbb{R}^2 \to \mathbb{R}^2$ given by*

$$g(x) := \big(5\exp(x_1) - 5, x_1^2 + x_2^2\big),$$

and the point $\overline{x} = (0, 0)$. It follows that

$$Dg(x) = \begin{bmatrix} 5\exp(x_1) & 0 \\ 2x_1 & 2x_2 \end{bmatrix} \text{ and } Dg(\overline{x}) = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}.$$

Considering the vector $d = (1, 0)^T$ we get that $Dg(\overline{x})d = (5, 0)^T \in \operatorname{int}(\mathbb{L}_2)$, that is, Robinson's CQ holds at $\overline{x}$.

Now, take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$. Without loss of generality, assume that $\|x^k\|^2 \neq 0$ for all k. We get that the eigenvectors of $g(x^k)$ are uniquely determined and given by

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -\frac{\widehat{g}(x^k)}{\|\widehat{g}(x^k)\|}\right) = \frac{1}{2}(1, -1) \text{ and } u_2(g(x^k)) = \frac{1}{2}\left(1, \frac{\widehat{g}(x^k)}{\|\widehat{g}(x^k)\|}\right) = \frac{1}{2}(1, 1),$$

that is associated to the vector $\widehat{w}^k = 1$ and, in addition, $\widehat{w} = w^k$. On the one hand, we have that

$$Dg(x^k)^T(1, -\widehat{w}^k) = \begin{pmatrix} 5\exp(x_1^k) - 2x_1^k \\ -2x_2^k \end{pmatrix} \text{ and } Dg(x^k)^T(1, \widehat{w}^k) = \begin{pmatrix} 5\exp(x_1^k) + 2x_1^k \\ 2x_2^k \end{pmatrix},$$

and then they are linearly independent. On the other hand, we have that $Dg(\overline{x})^T(1, -\widehat{w}) = Dg(\overline{x})^T(1, \widehat{w}) = (5, 0)^T$, that is, they are linearly dependent and it means that weak-CRCQ does not hold at $\overline{x}$.

Below we have a figure showing the relation among the constraint qualifications nondegeneracy and Robinson with the proposed ones in the two chapters until the moment.

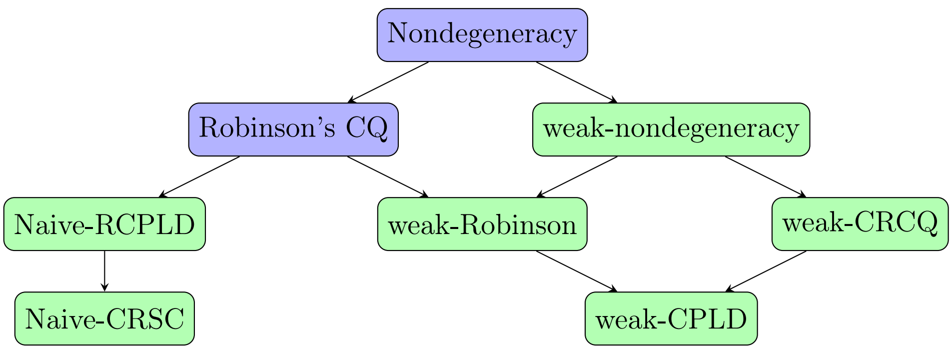

Figure 3.3: *Relation among the CQ's for (NSOCP). The boxes in blue are the well-known CQ's in nonlinear second-order cone programming and the green boxes are the new proposals including the weak versions.*

To finish this chapter and prove that the conditions proposed are indeed constraint qualifications, the strategy adopted was to use the external penalty method [AHM+22b, Theorem 4.1]. With this result at hand, given a local minimizer $\overline{x}$ of (NSOCP) we can build a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ where each x^k is a local minimizer of a regularized problem where the constraints are penalized. One implication of this result relies on the fact that we can give a formula for the Lagrange multipliers based on the penalty parameter and the unfeasibility of x^k. After that, with this tool, we are able to prove not only that weak-CPLD is a constraint qualification, but also that it is enough to guarantee convergence of an external penalty method. We will not show the proof of this result, but the reader can find it in detail in [AHM+22b, Theorem 4.2].

Theorem 3.2.2. *Consider the problem (NSOCP) and let $\overline{x}$ be a feasible point. Let $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ be a sequence of penalty parameters and take a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ in which*

$$\nabla_x L\left(x,\ \rho_k \mathcal{P}_{\mathbb{L}_{m_1}}(-g_1(x^k)), \ldots, \rho_k \mathcal{P}_{\mathbb{L}_{m_q}}(-g_q(x^k))\right) \to 0.$$

If weak-CPLD holds at $\overline{x}$, then $\overline{x}$ is a KKT point. In particular, weak-CPLD is a constraint qualification.

Chapter 4

Constant Rank-type Constraint Qualifications

In this chapter, we will present a new approach to defining constant rank-type constraint qualification for the problem (NSOCP). In Chapter 2, we showed the difficulty of such proposals, some authors tried before in [ZZ19] but made some mistakes as we showed in a counter-example in [AFH+21]. Also, we made a first proposal in [AHM+22a] where we used as much nonlinear programming structure as we could, once that in NLP the topic of constraint qualification is well-developed and we have the help of a powerful tool: the sequential optimality condition. In Chapter 3, we used the eigenvectors of the second-order cone in order to build new definitions of the nondegeneracy condition and Robinson's CQ, and then propose weaker constraint qualifications [AHM+22b]. However, we kind of avoided the pure conical structure of the problem through Proposition 3.1.1. In this chapter we will understand in a deeper way the essence of a constraint qualification of a constrained optimization problem in order to define CRCQ in a similar vein as Janin did in [Jan84] and, moreover, some second-order properties are held as proved in [AES10]. The results of this chapter are based on [AHM+23].

4.1 Revisiting Abadie's and Guignard's constraint qualifications

In nonlinear programming problems, it is known that if the constraints are linear, then every local minimizer has a Lagrange multiplier. A direct proof can be done using the constant rank condition presented by Janin in [Jan84]. In that paper, Janin showed that given a feasible point $\overline{x}$ of (NLP) for every direction $d \in \mathcal{L}_\Omega(\overline{x})_{\mathrm{NLP}}$, there exists a curve $\xi : [0, \varepsilon) \to \mathbb{R}^n$ such that $\xi(0) = \overline{x}$, $\xi(t) \in \Omega$ for all $t \geq 0$ small enough and $\xi'(0) = d$. In other words, Janin proved that

$$\mathcal{T}_\Omega(\overline{x})_{\mathrm{NLP}} = \mathcal{L}_\Omega(\overline{x})_{\mathrm{NLP}}, \tag{4.1}$$

which is called *Abadie's Constraint Qualification* in a nonlinear programming context. The equality between the tangent and linear cones implies that their polar are also equal, that is,

$$\mathcal{T}_\Omega(\overline{x})^\circ_{\mathrm{NLP}} = \mathcal{L}_\Omega(\overline{x})^\circ_{\mathrm{NLP}}, \tag{4.2}$$

which is called *Guignard's Constraint Qualification* in a nonlinear programming context. However, when we pass to second-order cone programming problems, these conditions are not enough to define a constraint qualification, as illustrated in the following example given by Andersen, Roos and Terlaky in [ART02, Subsection 2.1].

Example 4.1.1. *Consider the following problem*

$$\begin{array}{ll} \text{Minimize} & f(x) := -x_2, \\ \text{s.t.} & g(x) := (x_1, x_1, x_2) \in \mathbb{L}_3. \end{array} \tag{4.3}$$

The feasible set is given by $(x_1, x_2) \in \mathbb{R}^2$ *such that* $x_1 \geq \sqrt{x_1^2 + x_2^2}$. *It implies that* $x_2 = 0$ *and that* $\overline{x} := (0, 0)$ *is a local minimizer of* (4.3). *Take any direction* $d \in \mathcal{L}_\Omega(\overline{x})$, *that is,* $d = (d_1, d_2) \in \mathbb{R}^2$ *such that* $Dg(\overline{x})d \in \mathbb{L}_3$. *It follows that* $Dg(\overline{x})d = (d_1, d_1, d_2) = g(d)$ *and* $Dg(\overline{x})d \in \mathbb{L}_3$. *Define* $d^k := d$ *and* $t_k := \frac{1}{k}$. *Since* g *is linear, we get that*

$$g(\overline{x} + t_k d^k) = g(\overline{x}) + t_k g(d) \in \mathbb{L}_3,$$

because $g(\overline{x}), g(d) \in \mathbb{L}_3$ *and* $t_k \geq 0$. *Thus,* $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$ *and, in addition,* $\mathcal{T}_\Omega(\overline{x})^\circ = \mathcal{L}_\Omega(\overline{x})^\circ$. *However, as showed in [ART02, Subsection 2.1] there is no Lagrange multiplier for* $\overline{x}$.

The example above shows us some important points. The first one is the fact that the linearity of the constraints is not enough to guarantee the existence of Lagrange multipliers in (NSOCP). The second one, if there exists an Abadie-type constraint qualification for (NSOCP), then it is not equal to the one that we know in nonlinear programming problems. Moreover, notice that even a Guignard-type constraint qualification must be different from (NLP) case. The third one, Abadie's Constraint Qualification proposed by Zhang and Zhang in [ZZ19, Theorem 3.2] is also incorrect (besides the constant rank-type constraint qualifications proposed by them, as we showed in [AFH+21]).

With this information in mind, let us make a parallel path between nonlinear programming and second-order cone programming problems in order to find the moment when the two theories start to have different results. We know, for example, that the first-order geometric necessary condition (2.4) still holds for both classes of problems. Thus, the next step will be to analyze Guignard's constraint qualification for both problems, which we already know contains some differences between NLP and NSCOP. This is important because at least in the NLP context, Guignard's constraint qualification is the weakest CQ that we can request for a local minimizer in order to guarantee the existence of Lagrange multipliers [GT71]. Moreover, the proposition of Guinard's constraint qualification was made for a more general class of problems, namely, in Banach spaces [Gui69].

4.1.1 Guignard's Theorem

To fix the ideas, let $\mathbb{B}$ be a Banach Space (a vector space with a norm defined, namely, $\|\cdot\|_{\mathbb{B}}$, that is complete when we consider the metric induced by its norm, i.e., $d(x, y) := \|x - y\|_{\mathbb{B}}$). For the sake of completeness, let us consider that $\mathbb{B}$ also has an inner product defined $\langle \cdot, \cdot \rangle_{\mathbb{B}}$. In addition, let $\mathbb{K} \subseteq \mathbb{B}$ be a nonempty closed convex pointed cone (we will not go deep in the details about Banach Spaces because it is not the goal of this subsection. The reader can find more details in [Lim83, Bea11]). Consider the following problem

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & g(x) \in \mathbb{K}, \end{array} \tag{4.4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{B}$ are twice continuously differentiable. The tangent and linear cones for (4.4) are defined in a similar way as (2.1) and (2.2), respectively, just adapting their respective feasible sets. Now, consider the following pivotal set

$$\begin{aligned} H(\overline{x}) &:= Dg(\overline{x})^T N_{\mathbb{K}}(g(\overline{x})) \\ &= \left\{ Dg(\overline{x})^T y \mid y \in N_{\mathbb{K}}(g(\overline{x})) \right\}, \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} N_{\mathbb{K}}(g(\overline{x})) &:= \mathcal{T}_{\mathbb{K}}(g(\overline{x}))^\circ \\ &= \{ y \in \mathbb{K}^\circ \mid \langle g(\overline{x}), y \rangle_{\mathbb{B}} = 0 \} \end{aligned} \tag{4.6}$$

is the normal cone to $\mathbb{K}$ at $g(\overline{x})$. With this information at hand, we can present the main result of this subsection that was proved by Guignard in [Gui69].

Theorem 4.1.1. *(Theorem 2 of [Gui69]) Let $\overline{x}$ be feasible point of (4.4). Then*

i) $\mathcal{L}_\Omega(\overline{x})^\circ = \mathrm{cl}(H(\overline{x}))$ *or, equivalently,* $\mathcal{L}_\Omega(\overline{x}) = H(\overline{x})^\circ$*;*

ii) if $\overline{x}$ is a local minimizer and, in addition, $\mathcal{T}_\Omega(\overline{x})^\circ = \mathcal{L}_\Omega(\overline{x})^\circ$ and $H(\overline{x})$ is closed, than there exists some $\overline{\mu} \in \mathbb{K}^\circ$ such that

$$\nabla f(\overline{x}) + Dg(\overline{x})^T \overline{\mu} = 0 \quad \text{and} \quad \langle g(\overline{x}), \overline{\mu} \rangle = 0. \tag{4.7}$$

In the first moment, the theorem above seems too similar to what is known in the classical literature of nonlinear programming problems. Actually, the unique difference is the requesting of closedness of $H(\overline{x})$, which seems to be a little detail when we consider its structure: $H(\overline{x})$ is the linear image of a closed convex cone. The issue is that despite the first thought that $H(\overline{x})$ is always closed, this is not true. To be more specific, this is not a trivial problem and Pataki wrote a whole paper only about this question in [Pat07]. Moreover, when we take into account the structure of a nonlinear programming problem, we have that $N_{\mathbb{R}^m_+}(g(\overline{x}))$ is polyhedral and then its image by a linear application is always closed. In other words, in nonlinear programming problems the set $H(\overline{x})$ is automatically closed and this is the reason why Guignard's constraint qualification reduces to the equality $\mathcal{T}_\Omega(\overline{x})^\circ = \mathcal{L}_\Omega(\overline{x})^\circ$. However, when we pass to the second-order cone programming context, this issue comes back and needs to be dealt with in a proper way.

When we look at Example 4.3, it is proved that $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$. Thus, since we know that there is no Lagrange multiplier due to the proof of Andersen in [ART02, Subsection 2.1], the only explanation for this fact lies in the closedness of $H(\overline{x})$. Indeed, Pataki showed in [Pat07, Example 4.3] that the set $H(\overline{x})$ for the problem (4.3) is not closed and in [AHM$^+$23, Example 2.1] we showed in another way that $H(\overline{x})$ is not closed. Actually, if the constraints of (NSOCP) are linear, we always have $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$ and the existence of Lagrange multipliers only depends on the closedness of $H(\overline{x})$.

Corollary 4.1.1. *(Linear SOCP) Consider the following problem*

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & g_j(x) \in \mathbb{L}_{m_j}, \quad j = 1, \dots, q, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}^{m_j}$ are continuously differentiable and the constraints g_j, $j = 1, \dots, q$ are linear. Define $\mathbb{K} := \mathbb{L}_{m_1} \times \dots \times \mathbb{L}_{m_q}$ and $g(x) := (g_1(x), \dots, g_q(x))$ and let $\overline{x}$ be a local minimizer. If the set $H(\overline{x})$ defined in (4.5) is closed, then KKT conditions hold at $\overline{x}$.

Proof. The inclusion $\mathcal{T}_\Omega(\overline{x}) \subseteq \mathcal{L}_\Omega(\overline{x})$ is always satisfied. Conversely, assume without loss of generality that $g_j(\overline{x}) = 0$ for $j = 1, \dots, q$. Take a direction $d \in \mathcal{L}_\Omega(\overline{x})$, consider any sequence of non-negative scalars $t_k \to 0$, and define $d^k = d$ for all k. Also, due to the linearity of g_j's we have that

$$g_j(\overline{x} + t_k d^k) = g_j(\overline{x}) + t_k Dg_j(\overline{x}) d.$$

Since $d \in \mathcal{L}_\Omega(\overline{x})$, we have that $Dg_j(\overline{x}) d \in \mathbb{L}_{m_j}$ for all $j = 1, \dots, q$, and then $\overline{x} + t_k d^k$ is feasible for all k. □

To finish this little discussion about Guignard's Constraint Qualification, we take this opportunity to present below the correct definition for Abadie's Constraint Qualification. Unfortunately, we did not have access to the original publication of Abadie [Aba65] where it was presented his constraint qualification. However, recently Börgens et al. also recalled this definition for optimization problems in Banach Spaces in [BKMW20, Definition 5.5]. Just to make the future explanation about the closedness of $H(\overline{x})$ for the problem (NSOCP), from this point let us consider $\mathbb{K} := \mathbb{L}_{m_1} \times \dots \times \mathbb{L}_{m_q}$, where $g(x) := (g_1(x), \dots, g_q(x))$. We will need this for the following definition.

Definition 4.1.1. *Let $\overline{x}$ be a feasible point of (NSOCP). We say that Abadie's Constraint Qualification holds at $\overline{x}$ if $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$ and $H(\overline{x})$ is closed.*

Before we pass to the next session, let us just standardize our notation with the last results related to Guignard's CQ. Given an m-dimensional second-order cone $\mathbb{L}_m$, we have that it is a self-dual cone, that is, $\mathbb{L}_m^\circ = -\mathbb{L}_m$. The vectors in (4.6) will play the role of Lagrange multipliers, so in order to keep the consistency with the KKT conditions presented in Chapter 1, we will consider the Lagrange multipliers as being elements of $\mathbb{L}_m$ but with the negative signal in the equation as defined in (2.5). In order to define the set $H(\overline{x})$ as presented in (4.5) for the general problem (NSOCP), we will consider the cone $\mathbb{K} := \mathbb{L}_{m_1} \times \ldots \times \mathbb{L}_{m_q}$.

4.2 Nondegeneracy and Abadie's CQ via Implicit Function Theorem

In [Jan84], Janin proposed a constant rank-type constraint qualification showing that his proposal implies Abadie's CQ in a nonlinear programming problem. For such, he used an adaptation [Jan84, Proposition 2.2] of a constant rank theorem given by Malliavin in [Mal72, Subsection 5.3]. We know that the constant rank theorem is a generalization of the well-known implicit function theorem, which can be used to show that nondegeneracy implies Abadie's CQ, for example. This is the goal of this section, once we are interested in building CRCQ for (NSOCP), and the natural path for that is to build the bridge between nondegeneracy and Abadie's conditions.

In order to use the implicit function theorem, we will define some concepts that will be useful for us. The reader can find more details in [IT74]. Let X and Y be Banach spaces and let U be a neighborhood of a point $\overline{x} \in X$.

Definition 4.2.1. *Let $F : U \subseteq X \to Y$. We say that F is* Frechet differentiable *at the point $\overline{x}$, if there exists a continuous linear operator $\Lambda : X \to Y$ such that*

$$F(\overline{x} + h) = F(\overline{x}) + \Lambda h + r(h),$$

where

$$\lim_{\|h\|_X \to 0} \frac{\|r(h)\|_Y}{\|h\|_X} = 0.$$

In addition, we say that F is regular *at the point $\overline{x} \in X$ if it is Frechet differentiable at this point and*

$$\mathrm{Im} DF(\overline{x}) = Y,$$

where $DF(\overline{x})$ denotes the Frechet differential of F at $\overline{x}$.

The definition above is the natural generalization of the derivatives in $\mathbb{R}^n$. With this concept at hand, let us introduce the idea of tangent set and then we will be able to introduce the implicit function theorem.

Definition 4.2.2. *Let M be a subspace of X. We say that a vector $v \in X$ is* tangent *to the set M at the point $\overline{x}$, if there exist an $\varepsilon > 0$ and a function $r : [0, \varepsilon) \to X$ such that*

$$\overline{x} + tv + r(t) \in M, \ \forall t \in [0, \varepsilon),$$

where

$$\lim_{t \to 0} \frac{\|r(t)\|_X}{t} = 0.$$

The set of all of these tangent vectors is called tangent set *to M at $\overline{x}$ and will be denoted by $T_M(\overline{x})$.*

Theorem 4.2.1. *(Liusternik's Theorem [IT74]) Let X, Y be Banach spaces and consider a neighborhood U of a point $\overline{x} \in X$. Consider a function $F : U \to Y$ that is Frechet differentiable and* *suppose that F is regular at $\overline{x}$. Also, assume that its derivative is continuous at $\overline{x}$. Then, the tangent space to the set*

$$M := \{x \in U \mid F(x) = F(\overline{x})\}$$

at the point $\overline{x}$ is equal to the kernel of the operator $F'(\overline{x})$, i.e.,

$$T_M(\overline{x}) = \ker(DF(\overline{x})).$$

In addition, there exists a neighborhood $U_{\overline{x}} \subset U$ and a function $\phi : U_{\overline{x}} \to X$ such that

$$F(x + \phi(x)) = F(\overline{x})$$

for all $x \in U_{\overline{x}}$.

Now we will prove that the nondegeneracy condition implies Abadie's CQ. But before proving the multifold case, we will do this for the case with only one single second-order constraint $g : \mathbb{R}^n \to \mathbb{R}^m$ at the vertex of the second-order cone and after that at the boundary of the second-order cone, in order to understand the concepts of the proofs.

Theorem 4.2.2. *Consider the problem (NSOCP) with $q = 1$. Let $\overline{x}$ be a feasible point. Assume that $g(\overline{x}) = 0$. If nondegeneracy holds at $\overline{x}$, then Abadie's CQ also holds at $\overline{x}$.*

Proof. First, let us prove that if $\overline{x}$ is nondegenerate, then $H(\overline{x})$ is closed and then we will prove that $\mathcal{L}_\Omega(\overline{x}) \subseteq \mathcal{T}_\Omega(\overline{x})$. Take a convergent sequence $\{y^k\}_{k\in\mathbb{N}}$ in which $y^k \to \overline{y}$. For each $k \in \mathbb{N}$, there exists only one vector $v^k \in -\mathbb{L}_m$ such that $y^k = Dg(\overline{x})^T v^k$, due to the fact that $Dg(\overline{x})^T$ has full column rank. In addition, since $Dg(\overline{x})^T$ is a continuous application, we get that the sequence $\{v^k\}_{k\in\mathbb{N}}$ is convergent and it converges to an element $\overline{v} \in -\mathbb{L}_m$, because it is a closed set. Thus.

$$\overline{y} = \lim_{k\to\infty} y^k = \lim_{k\to\infty} Dg(\overline{x})^T v^k = Dg(\overline{x})^T \lim_{k\to\infty} v^k = Dg(\overline{x})^T \overline{v},$$

and then $H(\overline{x})$ is closed.

Now, let us prove that $\mathcal{L}_\Omega(\overline{x}) \subseteq \mathcal{T}_\Omega(\overline{x})$. Take a direction $d \in \mathcal{L}_\Omega(\overline{x})$. Since we are assuming that $g(\overline{x}) = 0$, then we have three possibilities for d:

i) $Dg(\overline{x})d \in \operatorname{int}(\mathbb{L}_m)$;

ii) $Dg(\overline{x})d = 0$;

iii) $Dg(\overline{x})d \in \operatorname{bd}^+(\mathbb{L}_m)$.

Let us consider each case separately. Assume that $Dg(\overline{x})d \in \operatorname{int}(\mathbb{L}_m)$. Then

$$\begin{aligned} g(\overline{x} + td) &= g(\overline{x}) + tDg(\overline{x})d + r(t), \quad \text{where } \lim_{t\to 0} \frac{\|r(t)\|}{t} = 0, \\ &= tDg(\overline{x})d + r(t) \in \operatorname{int}(\mathbb{L}_m) \end{aligned}$$

for all $t \geq 0$ small enough. Thus, defining $t_k \to 0$ and $d^k = d$ for all k, we have that $g(\overline{x}+t_k d^k) \in \mathbb{L}_m$ for k large enough and then $d \in \mathcal{T}_\Omega(\overline{x})$.

Now, assume that $Dg(\overline{x})d = 0$, that is, $d \in \ker(Dg(\overline{x}))$. Consider the set

$$M := \{x \in \mathbb{R}^n \mid g(x) = g(\overline{x}) = 0\}.$$

Since $\overline{x}$ is nondegenerate, by Theorem 4.2.1 we have that $T_M(\overline{x}) = \ker(Dg(\overline{x}))$. Thus, there exists an $\varepsilon > 0$ and a function $r : [0, \varepsilon) \to \mathbb{R}^n$ such that

$$\overline{x} + td + r(t) \in M, \; t \in [0, \varepsilon),$$

where

$$\lim_{t\to 0}\frac{\|r(t)\|}{t}=0.$$

In other words, we have that $g(\overline{x}+td+r(t))=0$ where $t\in[0,\varepsilon)$. Defining $t_k:=\frac{\varepsilon}{k+1}$ and $d^k:=d+\frac{r(t_k)}{t_k}$, we get that $g(\overline{x}+t_kd^k)\in\mathbb{L}_m$ for all k.

In the last case, let us assume that $Dg(\overline{x})d\in\mathrm{bd}^+(\mathbb{L}_m)$. Before, we just like to emphasize that this is the hardest case and this is the reason why we left it as being the last one. The main difficulty relies on the fact that the direction d neither is orthogonal to all gradients of the constraints at the same time nor points to the interior of the second-order cone. Let us deal with this case carefully.

Notice that since $Dg(\overline{x})d\in\mathrm{bd}^+(\mathbb{L}_m)$, we have that $\nabla g_0(\overline{x})^Td=\|D\widehat{g}(\overline{x})d\|>0$. Moreover, the vector $\Gamma Dg(\overline{x})d$ is orthogonal to $Dg(\overline{x})d$ and both belong to the second-order cone. If $m>2$, let $\{b_1,\ldots,b_{m-2}\}$ be an orthonormal basis of the linear subspace

$$\mathrm{span}\{Dg(\overline{x})d,\Gamma Dg(\overline{x})d\}^{\perp},$$

where $\perp$ denotes the orthogonal complement. Now, define the matrix $B=[b_i]_{i=1,\ldots,m-2}$ whose i-th column is the vector b_i. Now, define the following matrix $A=\Gamma Dg(\overline{x})d$ if $m=2$, otherwise

$$A:=[\Gamma Dg(\overline{x})d,B] \tag{4.8}$$

and consider the function $F:\mathbb{R}^n\to\mathbb{R}^m$ given by $F(x):=A^Tg(x)$. We have that

$$DF(\overline{x})=A^TDg(\overline{x}).$$

It follows that F is regular at $\overline{x}$ because $Dg(\overline{x})$ has full rank and A has orthonormal columns. Define the set

$$M:=\{x\in\mathbb{R}^n\mid F(x)=F(\overline{x})=0\}.$$

Since we are assuming that $Dg(\overline{x})d\in\mathrm{bd}^+(\mathbb{L}_m)$, by definition of A we have that $d\in\ker(DF(\overline{x}))$. By Theorem 4.2.1, we have that $T_M(\overline{x})=\ker(DF(\overline{x}))$. Thus, there exists an $\varepsilon>0$ and a function $r:[0,\varepsilon)\to\mathbb{R}^n$ such that

$$\overline{x}+td+r(t)\in M,\ \ t\in[0,\varepsilon),$$

where

$$\lim_{t\to 0}\frac{\|r(t)\|}{t}=0.$$

In other words, we obtain $F(\overline{x}+td+r(t))=0$. By definition of F and of the matrix A, we get that

$$\begin{aligned}\langle g(\overline{x}+td+r(t)),b_i\rangle &= 0,\ i=1,\ldots,m-2,\\ \langle g(\overline{x}+td+r(t)),\Gamma Dg(\overline{x})d\rangle &= 0,\ t\in[0,\varepsilon).\end{aligned}$$

Thus, since we have eliminated $m-1$ dimensions, we obtain $g(\overline{x}+td+r(t))\in\mathrm{span}\{Dg(\overline{x})d\}$. However, this is not enough to guarantee the feasibility of $\overline{x}+td+r(t)$. In order to get the feasibility, remember that $\|Dg(\overline{x})d\|>0$. Also, consider the following computing

$$\begin{aligned}\langle g(\overline{x}+td+r(t)),Dg(\overline{x})d\rangle &= \langle g(\overline{x})+tDg(\overline{x})d,Dg(\overline{x})d\rangle+r(t)\\ &= t\|Dg(\overline{x})d\|+r(t)>0,\end{aligned}$$

for $t>0$ small enough. In other words, we have that the cosine between the vectors $g(\overline{x}+td+r(t))$ and $Dg(\overline{x})d$ is positive. In addition, since $g(\overline{x}+td+r(t))\in\mathrm{span}\{Dg(\overline{x})d\}$ we get that for all $t>0$ small enough, there exists an $\alpha_t>0$ such that

$$g(\overline{x}+td+r(t))=\alpha_tDg(\overline{x})d\in\mathrm{bd}^+(\mathbb{L}_m). \tag{4.9}$$

Defining $t_k:=\frac{\varepsilon}{k+1}$ and $d^k:=d+\frac{r(t_k)}{t_k}$, we get that $g(\overline{x}+t_kd^k)\in\mathbb{L}_m$ for all k and, i.e., $d\in\mathcal{T}_\Omega(\overline{x})$. □

Before we deal with the case $g(\overline{x}) \in \text{bd}^+(\mathbb{L}_m)$, let us analyze some details of the proof above in order to have a better comprehension. The first point is that the proof that $H(\overline{x})$ is closed is easily extensible for the multifold case. It is possible to see that the nondegeneracy assumption plays an important role in obtaining this result. The second point is that when we consider the cases when $Dg(\overline{x})d \in \text{int}(\mathbb{L}_m)$ and $Dg(\overline{x})d = 0$, the proof follows the same ideas as presented in [Jan84, AES10]. The main point then relies on the case $Dg(\overline{x})d \in \text{bd}^+(\mathbb{L}_m)$, because we can not apply the ideas used previously in order to solve this case, and this is the main contribution of this proof. It is not easy to understand how to get the feasibility of the points in the curve given by the Theorem 4.2.1. Let us understand it step by step. For example, when $m = 2$ it is easier to understand the feasibility because the two pieces of information that we have, namely,

$$\langle g(\overline{x} + td + r(t)), \Gamma Dg(\overline{x})d \rangle = 0 \quad \text{and} \quad \langle g(\overline{x} + td + r(t)), Dg(\overline{x})d \rangle > 0.$$

are enough to guarantee the feasibility of $g(\overline{x}+td+r(t)) = \alpha_t Dg(\overline{x})d$, as illustrated in the following figure.

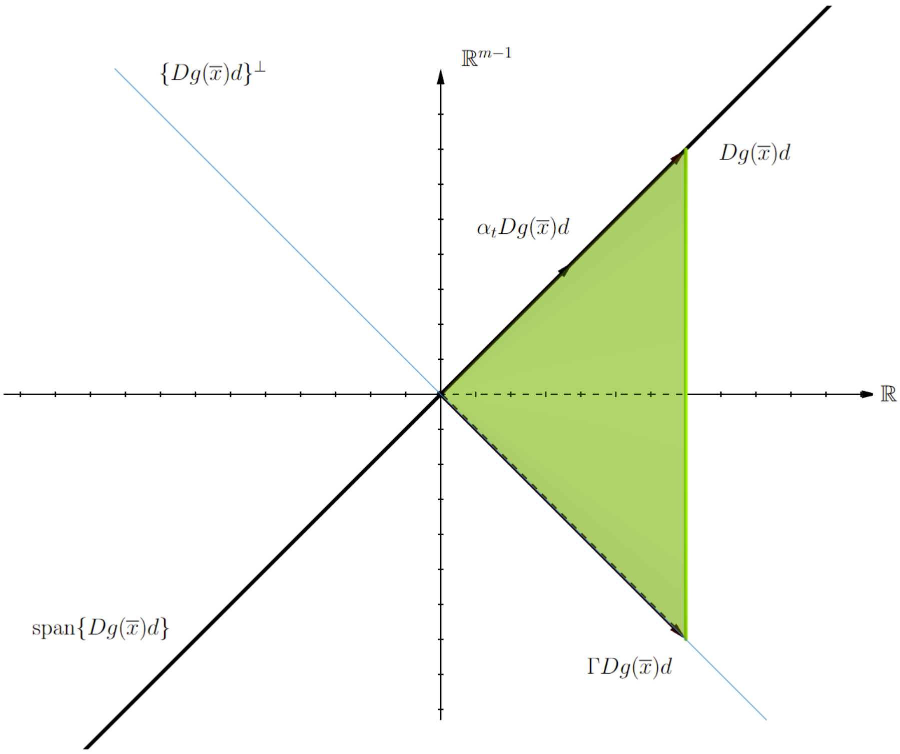

Figure 4.1: *Details about the proof of the Theorem 4.2.2 when $m = 2$.*

However, when we consider the case when $m > 2$, only these two inner products are not enough to guarantee feasibility. This is why we need to build the matrix A with the columns constructed by the columns of B. The following figure illustrates the feasibility of $g(\overline{x} + td + r(t))$ when $m > 2$ and $Dg(\overline{x})d \in \text{bd}^+(\mathbb{L}_m)$.

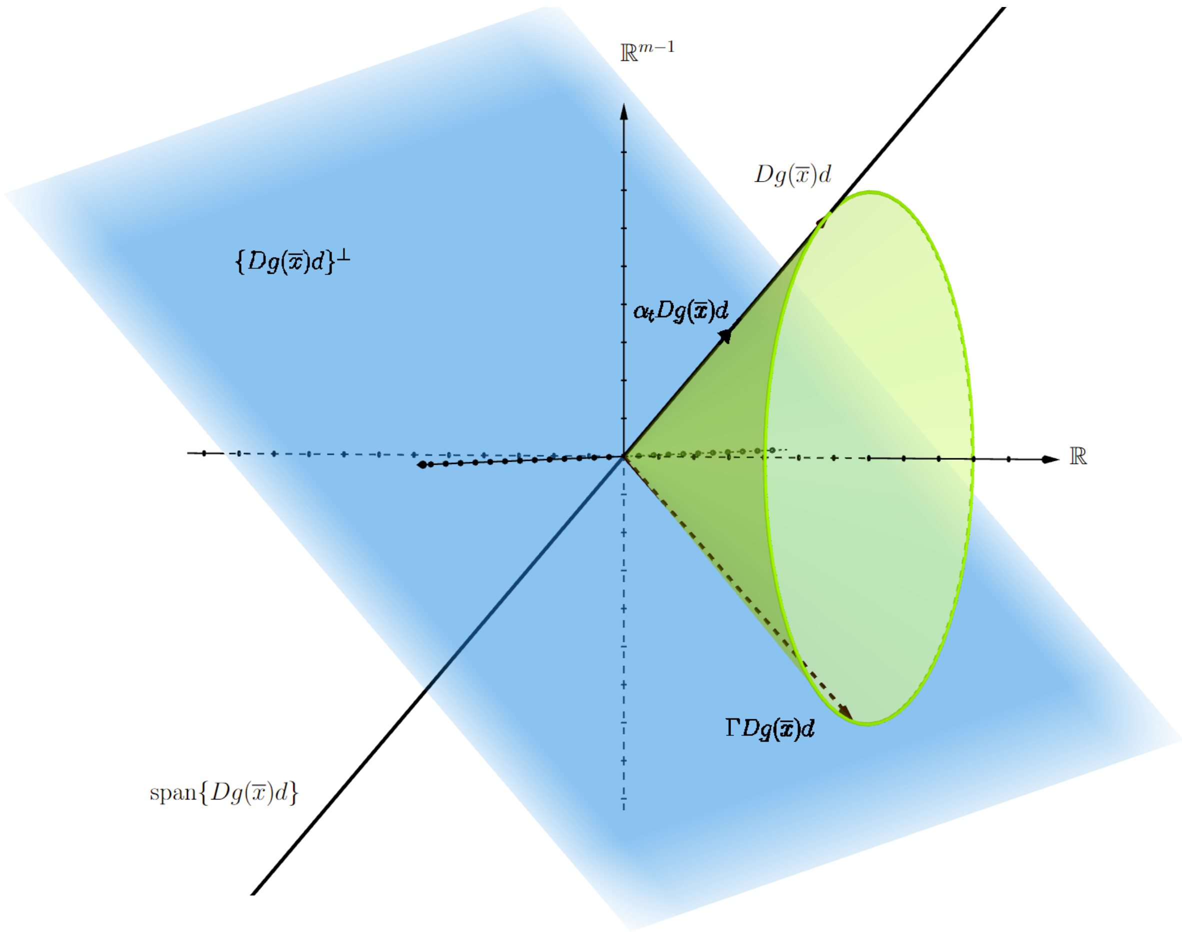

Figure 4.2: *Details about the proof of the Theorem 4.2.2 when $m > 2$.*

Now let us analyze the case $g(\overline{x}) \in \text{bd}^+(\mathbb{L}_m)$ but still with only one single second-order constraint.

Theorem 4.2.3. *Consider the problem (NSOCP) with $q = 1$. Let $\overline{x}$ be a feasible point. Assume that $g(\overline{x}) \in \text{bd}^+(\mathbb{L}_m)$. If nondegeneracy holds at $\overline{x}$, then Abadie's CQ also holds at $\overline{x}$.*

Proof. First, let us prove that $H(\overline{x})$ is closed. Remember that we are considering that $g(\overline{x}) \in \text{bd}^+(\mathbb{L}_m)$ and then the normal cone is unitary and given by $N_{\mathbb{L}}(g(\overline{x})) = \{-\Gamma g(\overline{x})\}$. Thus, the set $H(\overline{x})$ is also closed because it is unitary.

Now, let us prove that $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$. Take any direction $d \in \mathcal{L}_\Omega(\overline{x})$. From (2.2) we have that

$$\langle Dg(\overline{x})d, \Gamma g(\overline{x})\rangle \geq 0.$$

This is equivalent to $\langle \nabla\tilde{\phi}(\overline{x}), d\rangle \geq 0$, where

$$\tilde{\phi}(x) := \frac{1}{2}((g_0(x))^2 - \|\widehat{g}(x)\|^2)$$

is a reduction mapping as presented in Chapter 1, whose gradient is

$$\nabla\tilde{\phi}(\overline{x}) = Dg(\overline{x})^T\Gamma g(\overline{x}).$$

See equation (2.21) for more details. This situation is equivalent to a nonlinear programming problem with the constraint $\tilde{\phi}(x) \geq 0$.

Consider the case $\langle \nabla\tilde{\phi}(\overline{x}), d\rangle > 0$. It follows that

$$\begin{aligned}\tilde{\phi}(\overline{x}+td) &= \tilde{\phi}(\overline{x}) + t\nabla\tilde{\phi}(\overline{x})^T d + o(t), \quad \text{where} \quad \lim_{t\to 0}\frac{o(t)}{t} = 0,\\ &= t\nabla\tilde{\phi}(\overline{x})^T d + o(t) \geq 0,\end{aligned}$$

for all $t > 0$ small enough. It implies that $g_0(\overline{x}+td) \geq \|\widehat{g}(\overline{x}+td)\|$ and then $d \in \mathcal{L}_\Omega(\overline{x})$.

Now let us consider the case $\langle \nabla\tilde{\phi}(\overline{x}), d\rangle = 0$, that is, $d \in \ker(\nabla\tilde{\phi}(\overline{x}))$. Remember that $\nabla\tilde{\phi}(\overline{x}) \neq 0$, because $g(\overline{x}) \neq 0$ and $\overline{x}$ is nondegenerate. It means that $\tilde{\phi}$ is regular at $\overline{x}$. Consider the following set

$$M := \{x \in \mathbb{R}^n \mid \tilde{\phi}(x) = \tilde{\phi}(\overline{x}) = 0\}.$$

By Theorem 4.2.1, we have that $T_M(\overline{x}) = \ker(\nabla\tilde{\phi}(\overline{x}))$. Then there exists an $\varepsilon > 0$ and a function $r(t) : [0, \varepsilon) \to \mathbb{R}^n$ such that

$$\overline{x} + td + r(t) \in M, \quad \text{where} \quad \lim_{t\to 0^+}\frac{r(t)}{t} = 0,$$

that is, $\tilde{\phi}(\overline{x}+td+r(t)) = \tilde{\phi}(\overline{x}) = 0$. It means that $|g_0(\overline{x}+td+r(t))| = \|\widehat{g}(\overline{x}+td+r(t))\|$. In addition,

$$\lim_{t\to 0^+} g_0(\overline{x}+td+r(t)) = g_0(\overline{x}) > 0.$$

Therefore, $g_0(\overline{x}+td+r(t)) \geq 0$ for all $t \geq 0$ small enough and $g(\overline{x}+td+r(t)) \in \mathrm{bd}^+(\mathbb{L}_m)$, i.e., $d \in \mathcal{T}_\Omega(\overline{x})$. □

The Theorems 4.2.2 and 4.2.3 show the relation between the nondegeneracy condition and Abadie's CQ for (NSOCP) when we have only one second-order constraint. The case $g(\overline{x}) \in \mathrm{bd}^+(\mathbb{L}_m)$ is similar to a nonlinear programming problem with inequality constraint, and it is easier to see this equivalence when we consider a reduction mapping. The main difference between the nondegeneracy condition in (NSOCP) and LICQ in (NLP) relies on in the case $g(\overline{x}) = 0$, when we have that all coordinates of g are active at $\overline{x}$ and, in addition, we have one more case to analyze when we take a direction $d \in \mathcal{L}_\Omega(\overline{x})$, namely, $Dg(\overline{x})d \in \mathrm{bd}^+(\mathbb{L}_m)$.

We finish this section by presenting the result for the multifold case. We will not present the whole proof because it is similar to the ones presented previously, but we will present the main parts of the proof.

Theorem 4.2.4. *Consider the problem (NSOCP). Let $\overline{x}$ be a feasible point. If $\overline{x}$ is nondegenerate, then Abadie's CQ holds at $\overline{x}$.*

Proof. The fact that $H(\overline{x})$ is closed is similar to the proof of the Theorems 4.2.2 and 4.2.3. Now, take a direction $d \in \mathcal{L}_\Omega(\overline{x})$ and consider the following sets:

$$\begin{aligned} D_B^{int}(\overline{x}) &:= \{j \in \{1,2,\ldots,q\} \mid g_j(\overline{x}) \in \mathrm{bd}^+(\mathbb{L}_{m_j}), \nabla\tilde{\phi}_j(\overline{x})^T d > 0\}\\ D_B^0(\overline{x}) &:= \{j \in \{1,2,\ldots,q\} \mid g_j(\overline{x}) \in \mathrm{bd}^+(\mathbb{L}_{m_j}), \nabla\tilde{\phi}_j(\overline{x})^T d = 0\}\\ D_0^{int}(\overline{x}) &:= \{j \in \{1,2,\ldots,q\} \mid g_j(\overline{x}) = 0, Dg_j(\overline{x})d \in \mathrm{int}(\mathbb{L}_{m_j})\}\\ D_0^0(\overline{x}) &:= \{j \in \{1,2,\ldots,q\} \mid g_j(\overline{x}) = 0, Dg_j(\overline{x})d = 0\}\\ D_0^B(\overline{x}) &:= \{j \in \{1,2,\ldots,q\} \mid g_j(\overline{x}) = 0, Dg_j(\overline{x})d \in \mathrm{bd}^+(\mathbb{L}_{m_j})\},\end{aligned} \tag{4.10}$$

The constraints g_j where we have $Dg_j(\overline{x})d \in \mathrm{int}(\mathbb{L}(m_j))$ and $\nabla\tilde{\phi}_j(\overline{x})^T d > 0$ are not relevant for the proof, since d is a feasible direction for such constraints. For each index $j \in D_0^B(\overline{x})$, let $\{b_1^j, \ldots, m_{m_j-2}^j\}$ be an orthonormal basis for the linear subspace

$$\mathrm{span}\{Dg_j(\overline{x})d,\ \Gamma_j Dg_j(\overline{x})d\}^\perp,$$

and define the matrix $A_j := \left[\Gamma_j Dg_j(\overline{x})d,\ b_1^j, \ldots,\ b_{m_j-2}^j\right]$. Now, consider the function

$$F(x) := \begin{cases} A_j^T g_j(x), & \text{if } j \in D_0^B(\overline{x}), \\ g_j(x), & \text{if } j \in D_0^0(\overline{x}), \\ \tilde{\phi}_j(x), & \text{if } j \in D_B^0(\overline{x}), \end{cases}$$

and the following set

$$M := \{x \in \mathbb{R}^n \mid F(x) = F(\overline{x}) = 0\}.$$

By construction, we have that $d \in \ker(DF(\overline{x}))$ and, in addition, since $\overline{x}$ is nondegenerate we have that F is regular at $\overline{x}$. Thus, by Theorem 4.2.1 we have that $T_M(\overline{x}) = \ker(DF(\overline{x}))$, i.e., there exists an $\varepsilon > 0$ and a function $r(t) : [0, \varepsilon) \to \mathbb{R}^n$, such that

$$\overline{x} + td + r(t) \in M, \quad \text{where} \quad \lim_{t \to 0^+} \frac{r(t)}{t} = 0,$$

that is, $F(\overline{x}+td+r(t)) = 0$. Using similar arguments to the proofs of the Theorems 4.2.2 and 4.2.3, we obtain that $d \in \mathcal{T}_\Omega(\overline{x})$. □

4.3 Constant Rank Constraint Qualification for NSOCP

In this section, we will present a constant rank-type constraint qualification for (NSOCP) based on the approach presented in the previous section. Once we have already built a bridge between the nondegeneracy condition and Abadie's CQ through an implicit function theorem, namely, Theorem 4.2.1, the next natural step is to replace the Theorem 4.2.1 by a constant rank-type theorem, which is the natural generalization of an implicit function theorem. For such, we will now revisit the results used by Janin in [Jan84, Proposition 2.2], and Andreani, Echagüe and Schuverdt in [AES10, Proposition 3.1]. The version presented by Janin [Jan84] was based on [Mal72, Subsection 5.3], where Janin considered nonlinear programming problems in a parametric form. Later, Andreani, Echagüe and Schuverdt [AES10] presented a reread of Janin's version in a simpler way, also in a nonlinear programming context.

Despite the following theorem being presented in a context similar to the nonlinear programming problems, we can apply it for second-order cone programming problems, due to the fact that the constraints in (NSOCP) are functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, and the original proposal made by Malliavin in [Mal72] was made in a more general context, namely topological varieties.

Theorem 4.3.1. *Let F be a function of class C^k, that is, k times continuously differentiable, where $F : \mathbb{R}^n \to \mathbb{R}^p$. Let $\overline{x} \in \mathbb{R}^n$ be a point such that there exists a neighborhood V of $\overline{x}$, where the matrix $DF(x)$ has constant rank for all $x \in V$. Let $v \in \ker(DF(\overline{x}))$ be an arbitrary vector. Then, there are neighborhoods V_1 and V_2 of $\overline{x}$ and a diffeomorphism $\Phi : V_1 \to V_2$ of class C^k in which:*

i) $\Phi(\overline{x}) = \overline{x}$;

ii) the Jacobian matrix of Φ at $\overline{x}$ is the identity matrix, that is, $D\Phi(\overline{x}) = \mathbb{I}_n$;

iii) the function $F(\Phi^{-1}(\overline{x} + v))$ has constant value for all $v \in \ker(DF(\overline{x})) \cap (V_2 - \overline{x})$, that is,

$$F(\Phi^{-1}(\overline{x} + v)) = F(\Phi^{-1}(\overline{x})).$$

Notice that there are some similarities between the Theorem 4.2.1 and Theorem 4.3.1. Indeed, we rewrote the theorem presented by Janin in [Jan84, Proposition 2.2], and Andreani, Echagüe and Schuverdt in [AES10, Proposition 3.1] in order to keep a consistency with the results that were obtained in the previous section and show that these theorems are similar. Now, let us present our constant rank condition for (NSOCP).

Definition 4.3.1. *Consider the problem* (NSOCP) *and let* $\overline{x}$ *be a feasible point. We say that the* facial constant rank property *holds at* $\overline{x}$*, if there is a neighborhood* V *of* $\overline{x}$ *such that for all subsets* $J_1, J_2 \subseteq I_0(\overline{x})$*,* $J_3 \subseteq I_B(\overline{x})$*, with* $J_1 \cap J_2 = \emptyset$*, and all matrices* $A_j \in \mathbb{R}^{m_j \times m_j - 1}$ *of full column rank with* $j \in J_1$*, the rank of*

$$\bigcup_{j \in J_1} \{Dg_j(x)^T A_j\} \bigcup_{j \in J_2} \{Dg_j(x)\} \bigcup_{j \in J_3} \{\nabla \tilde{\phi}_j(x)\}.$$

remains constant for all $x \in V$.

First of all, notice that if $j \in I_B(\overline{x})$ for all $j = 1, \ldots, q$, we recover exactly the definition of CRCQ for nonlinear programming problems. This result also can be obtained if we have that $m_1 = \ldots = m_q = 1$. Second, we did not call this condition as a constraint qualification, because it is not a CQ due to the fact that only the facial constant rank property is not enough to imply directly that $H(\overline{x})$ is closed. To see this, we can consider the Example 4.1.1 where the constraint is linear and satisfies the facial constant rank property, but there is no Lagrange multiplier for that problem as we already know. Third, the division of the indexes $j \in I_0(\overline{x})$ in two subsets is because given a direction $d \in \mathcal{L}_\Omega(\overline{x})$, we do not know in advance what constraints belong to either D_0^0 or D_0^B. Moreover, the requiring of matrices A_j is motivated by (4.8) in the proof that nondegeneracy implies Abadie's CQ. Moreover, it is possible to see that nondegeneracy implies the facial constant rank condition due to the fact that the matrices A_j have full column rank. Lastly, the condition presented in Definition 4.3.1 is equivalent to the condition that we presented in [AHM+23, Definition 4.1], but here we built it in a different way without using directly the concept of faces. The approach proposed here follows the ideas presented at the beginning of this chapter.

With the Definition 4.3.1 at hand, we can show that the facial constant rank property implies $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$.

Theorem 4.3.2. *Consider the problem* (NSOCP) *and let* $\overline{x}$ *be a feasible point. If the facial constant rank property holds at* $\overline{x}$*, then* $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$.

Proof. Take a direction $d \in \mathcal{L}_\Omega(\overline{x})$ and consider the sets defined in (4.10). For each index $j \in D_0^B(\overline{x})$, let $\{b_1^j, \ldots, m_{m_j-2}^j\}$ be an orthonormal basis for the linear subspace

$$\operatorname{span}\{Dg_j(\overline{x})d,\ \Gamma_j Dg_j(\overline{x})d\}^\perp,$$

and define the matrix $A_j := \left[\Gamma_j Dg_j(\overline{x})d,\ b_1^j, \ldots,\ b_{m_j-2}^j\right]$. Notice that A_j has full column rank and define $J_1 := D_0^B(\overline{x})$, $J_2 := D_0^0(\overline{x})$ and $J_3 := D_B^0(\overline{x})$.

Now, consider the function

$$F(x) := \begin{cases} A_j^T g_j(x), & \text{if } j \in D_0^B(\overline{x}), \\ g_j(x), & \text{if } j \in D_0^0(\overline{x}), \\ \tilde{\phi}_j(x), & \text{if } j \in D_B^0(\overline{x}). \end{cases} \tag{4.11}$$

Since the facial constant rank property holds at $\overline{x}$, then there exists a neighborhood V of $\overline{x}$ such that the rank of $DF(x)$ is constant for all $x \in V$. Also, by the definition of F, we have that $d \in \ker(DF(\overline{x}))$. Thus, by Theorem 4.3.1 there is a diffeomorphism $\Phi : V_1 \to V_2$ and an $\varepsilon > 0$ such that $\overline{x} + td \in V_2$, $t \in [0, \varepsilon)$ and, moreover,

$$F(\Phi^{-1}(\overline{x} + td)) = F(\Phi^{-1}(\overline{x})) = 0. \tag{4.12}$$

It means that

$$\begin{cases} A_j^T g_j(\Phi^{-1}(\overline{x} + td)) = 0, & \text{if } j \in D_0^B(\overline{x}), \\ g_j(\Phi^{-1}(\overline{x} + td)) = 0, & \text{if } j \in D_0^0(\overline{x}), \\ \tilde{\phi}_j(\Phi^{-1}(\overline{x} + td)) = 0, & \text{if } j \in D_B^0(\overline{x}), \end{cases}$$

and $g_j(\Phi^{-1}(\overline{x}+td)) \in \mathbb{L}_{m_j}$, if $D_0^{int}(\overline{x})$ and $\tilde{\phi}_j(\Phi^{-1}(\overline{x}+td)) \geq 0$, if $D_B^{int}(\overline{x})$. Last, defining $t_k \to 0^+$ and

$$d^k := \frac{\Phi^{-1}(\overline{x}+t_k d) - \overline{x}}{t_k},$$

give us that $d^k \to d$ (using i) and ii) from Theorem 4.3.1) and we obtain that $g_j(\overline{x}+t_k d^k) = g_j(\Phi^{-1}(\overline{x}+t_k d)) \in \mathbb{L}_{m_j}$ for all $j = 1, \ldots, q$, i.e., $d \in \mathcal{T}_\Omega(\overline{x})$ as we wanted to show. □

The theorem above is a different way to show that a linear second-order cone programming problem always satisfies the equality $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$, because if the constraints are linear then the facial constant rank property holds automatically. However, as we already explained previously the facial constant rank property is not a constraint qualification, because it is not enough to imply that $H(\overline{x})$ is closed. Moreover, the unique sufficient condition based on rank that implies the closedness of $H(\overline{x})$ is the fulfillment of the nondegeneracy condition. Also, remember that from Definition 4.1.1, Abadie's constraint qualification requires the closedness of $H(\overline{x})$ separately from the equality $\mathcal{T}_\Omega(\overline{x}) = \mathcal{L}_\Omega(\overline{x})$. With these ideas in mind, we introduce the following definition.

Definition 4.3.2. *Consider the problem* (NSOCP) *and let $\overline{x}$ be a feasible point. We say that the* constant rank constraint qualification (CRCQ) *holds at $\overline{x}$ if the facial constant rank property holds at $\overline{x}$ and, in addition, if the set $H(\overline{x})$ is closed.*

With this new proposal of constant rank condition at hand together with the Theorem 4.3.2, we obtain the following result:

Theorem 4.3.3. *The CRCQ condition according to the Definition 4.3.2 implies Abadie's CQ. In particular, the CRCQ condition is a constraint qualification for the problem* (NSOCP).

Once we have a constant rank-type constraint qualification at hand, let us establish the relation between CRCQ and the other constraint qualifications mentioned previously. Since we know some short paths in order to make an example where Robinson's CQ and weak-nondegeneracy hold, let us solve both relations with only one example. Keep in mind the following three hints:

- in order to make an example where Robinson's CQ does not hold at a feasible point $\overline{x}$, we need a constraint $g(x) = (g_0(x), \widehat{g}(x))$ be such that $\|\nabla g_0(\overline{x})\| \leq \|D\widehat{g}(\overline{x})\|$;
- in order to make weak-nondegeneracy does not hold at a feasible point $\overline{x}$, it is enough to show that $\nabla g_0(\overline{x}) \in \mathrm{Im} D\widehat{g}(\overline{x})$;
- linear constraints satisfy CRCQ if $H(\overline{x})$ is closed. Moreover, according to Pataki in [Pat07], if $g : \mathbb{R}^n \to \mathbb{R}^2$ then we have that $H(\overline{x})$ is closed because $\mathbb{L}_2$ is polyhedral.

Example 4.3.1. *Consider the following constraint $g(x) := (x, 2x)$ and the feasible point $\overline{x} = 0$. We have that $g(\overline{x}) = 0$ and $Dg(\overline{x}) = (1,2)^T$. Given any direction $d \in \mathbb{R}$, we have that $Dg(\overline{x})d = (d, 2d)^T$ and, in addition, $Dg(\overline{x})d \notin \mathrm{int}(\mathbb{L}_2)$, that is, Robinson's CQ does not hold at $\overline{x}$. In particular, notice that nondegeneracy does not hold at $\overline{x}$.*

Take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$. Without loss of generality, assume that $x^k \neq 0$ for all k (the other case is analogous). Then we have that the eigenvectors of $g(x^k)$ are uniquely determined, namely, $u_1(g(x^k)) = \frac{1}{2}(1,-1)$ and $u_1(g(x^k)) = \frac{1}{2}(1,1)$. Define $\widehat{w} := 1$. It follows that

$$\lim_{k\to\infty} u_1(g(x^k)) = \frac{1}{2}(1, -\widehat{w}) \text{ and } \lim_{k\to\infty} u_2(g(x^k)) = \frac{1}{2}(1, \widehat{w}),$$

and, in addition,

$$Dg(\overline{x})^\top(1, -\widehat{w}) = -1 \quad \text{and} \quad Dg(\overline{x})^\top(1, \widehat{w}) = 3$$

are linearly dependent. Thus, weak-nondegeneracy does not hold at $\overline{x}$.

Lastly, CRCQ holds at $\overline{x}$ because the constraint is linear, and, in addition, $H(\overline{x})$ is closed because $\mathbb{L}_2$ is polyhedral.

If the interested reader wants to build an example where CRCQ holds without using the fact that $\mathbb{L}_2$ is polyhedral, it is possible to consider any constraint in the following form: $g(x) = (ax, bx, 0)$, where $a, b \in \mathbb{R}$ are such that $b > a > 0$. The example above shows that CRCQ neither implies Robinson's CQ nor weak-nondegeneracy. Next, we will show the other implications among these constraint qualifications. For such, we will take the "opposite" hints from the ones considered before, that is, we need a constraint $g(x) = (g_0(x), \widehat{g}(x)) \in \mathbb{L}_m$ such that

- $\|\nabla g_0(\overline{x})\| > \|D\widehat{g}(\overline{x})\|$;
- $\nabla g_0(\overline{x}) \notin \mathrm{Im} D\widehat{g}(\overline{x})$;

Example 4.3.2. *Consider the constraint $g : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $g(x) := (5x_1, x_2, x_3^2)$. Let $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ be any sequence such that $x \neq 0$ for all k (the other case is analogous). Thus, the eigenvectors of $(g(x^k)$ are uniquely determined and they are given by*

$$u_i(g(x^k)) = \frac{1}{2}\left(1, (-1)^i \frac{x_2^k}{\sqrt{(x_2^k)^2 + (x_3^k)^4}}, (-1)^i \frac{(x_3^k)^2}{\sqrt{(x_2^k)^2 + (x_3^k)^4}}\right), \text{ with } i = 1, 2.$$

It follows that

$$\lim_{k\to\infty} u_i(g(x^k)) = \frac{1}{2}\left(1, (-1)^i \widehat{w}\right),$$

where $\widehat{w} = (\widehat{w}_1, \widehat{w}_2) \in \mathbb{R}^2$ is any vector such that $\|\widehat{w}\| = 1$ and $\widehat{w}_1 \neq 0$, because $(x_3^k)^4$ goes to zero faster than $(x_2^k)^2$. We have that

$$Dg(x)^T = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2x_3 \end{bmatrix} \tag{4.13}$$

and then we get that the vectors

$$Dg(\overline{x})^\top(1, -\widehat{w}) = \begin{pmatrix} 5 \\ -\widehat{w}_1 \\ 0 \end{pmatrix} \quad \text{and} \quad Dg(\overline{x})^\top(1, \widehat{w}) = \begin{pmatrix} 5 \\ \widehat{w}_1 \\ 0 \end{pmatrix}$$

are linearly dependent, that is, weak-nondegeneracy holds at $\overline{x}$.

In order to see that Robinson's CQ holds at $\overline{x}$, take the direction $d = (1, 0, 0)$. We get that $Dg(\overline{x})d = (5, 0, 0) \in \mathrm{int}(\mathbb{L}_3)$.

Lastly, let us show that CRCQ does not hold at $\overline{x}$. Consider the set $J_1 = I_0(\overline{x})$ and the matrix

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

which has full column rank. Using (4.13), we obtain

$$Dg(x)^T A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 2x_3 \end{bmatrix}$$

and it has rank equals to 1 at $\overline{x}$ and rank equals to 2 for x in a neighborhood of $\overline{x}$ in which $x_3 \neq 0$.

In the examples above we showed that CRCQ is strictly weaker than the nondegeneracy condition it is independent of Robinson's CQ (in a similar way to what happens in nonlinear programming problems) and of weak-nondegeneracy. Moreover, it explains the gap that exists when we consider linear constraints and we showed that it implies the equality between the tangent and linear cones, as Janin did in [Jan84] when it was presented the CRCQ for nonlinear programming problems.

4.4 A Constraint Qualification based on Curves and Second-Order Optimality Conditions

In this section, we will present a new constraint qualification based on curves for (NSOCP) that is related to the CRCQ presented in the previous section. The main motivation for this comes from [AES10, Proposition 3.2], [FSS22, Definition 2.1] and [McC67]. In [McC67], McCormick proposed some constraint qualifications based on curves for (NLP). Such proposals considered: i) the existence of a feasible curve whose tangent is a direction in the linearized cone $\mathcal{L}_{\Omega}(\overline{x})_{\mathrm{NLP}}$ (McCormick First-Order Constraint Qualification); ii) the existence of a feasible curve that is twice differentiable and whose tangent is a vector in the critical subspace (2.14) (McCormick Second-Order Constraint Qualification).

More recently, Fazzion, Sánchez and Schuverdt in [FSS22] proposed a reformulation of the constraints given by McCormick in order to get second-order optimality conditions that are related to the critical cone, instead of the critical subspace. We recall their reformulation in the following definition:

Definition 4.4.1. *([FSS22, Definition 2.1]) Consider the problem* (NLP) *and let $\overline{x}$ be a feasible point. Given a nonzero direction $d \in \mathcal{L}_{\Omega}(\overline{x})_{\mathrm{NLP}}$, we say that the* Reformulation of the McCormick for (NLP) *(NLP-Ref-McCormick) holds at $\overline{x}$ if there exists a twice differentiable curve $\xi : [0, \varepsilon] \to \mathbb{R}^n$ such that $\xi(0) = \overline{x}$ and $\xi'(0) = d$, and, in addition, for all $t \in (0, \varepsilon]$ we have that*

$$\begin{array}{rcll} h_i(\xi(t)) & = & 0, & i = 1, \ldots, p, \\ g_j(\xi(t)) & = & 0, & j \in A(\overline{x}) \;\; \text{such that } \nabla g_j(\overline{x})^T d = 0, \\ g_j(\xi(t)) & > & 0, & j \in A(\overline{x}) \;\; \text{such that } \nabla g_j(\overline{x})^T d > 0. \end{array} \tag{4.14}$$

In [FSS22, Theorem 2.1] the authors showed that NLP-Ref-McCormick is a constraint qualification and also that it satisfies the Strong Second-Order Condition for nonlinear programming problems (2.13). Nevertheless, this condition was implicitly mentioned previously in [AES10, Proposition 3.2]. The thesis of the proposition is precisely the NLP-Ref-McCormick condition, but without using the result as a constraint qualification itself, but as a natural result of the constant rank constraint qualification for nonlinear programming problems. In order to obtain the NLP-Ref-McCormick condition, the authors in [AES10] used the constant rank theorem (Theorem 4.3.1) in a nonlinear programming problem context. With these ideas in mind, we will propose an extension of Ref-McCormick for (NSOCP) and get some second-order optimality conditions, in a similar vein of [AES10, FSS22].

Definition 4.4.2. *Consider the problem* (NSOCP) *and let $\overline{x}$ be a feasible point. Given a direction $d \in \mathcal{L}_{\Omega}(\overline{x})$, consider the sets defined in* (4.10)*. We say that the* Reformulation of the McCormick *(Ref-McCormick) for* (NSOCP) *holds at $\overline{x}$ if the set $H(\overline{x})$ is closed and if there exists a twice differentiable curve $\xi : [0, \varepsilon] \to \mathbb{R}^n$ such that $\xi(0) = \overline{x}$ and $\xi'(0) = d$, and, in addition, for all $t \in (0, \varepsilon]$ we have that*

$$g_j(\xi(t)) \in \left\{ \begin{array}{ll} \mathrm{bd}^+(\mathbb{L}_{m_j}), & j \in D_0^B(\overline{x}) \cup D_B^0(\overline{x}), \\ \mathrm{int}\,(\mathbb{L}_{m_j}), & j \in D_0^{int}(\overline{x}) \cup D_B^{int}(\overline{x}), \\ \{0\}, & j \in D_0^0(\overline{x}) \end{array} \right. \tag{4.15}$$

Notice that this definition coincides with the definition proposed by the authors in [FSS22] when the problem (NLP) just has inequality constraints. The case when the problem (NSOCP) has equality constraints can be done in the same way and we will omit it here because it is not the goal of this section. Before we prove that the condition above is indeed a constraint qualification for (NSOCP), let us establish the relation with the constant rank constraint qualification proposed in the previous section. For such, we will use the Theorem 4.3.1.

Theorem 4.4.1. *Consider the problem* (NSOCP) *and let $\overline{x}$ be a feasible point. If CRCQ holds at $\overline{x}$, then Ref-McCormick also holds.*

Proof. Since we are assuming that CRCQ holds at $\overline{x}$, then given a direction $d \in \mathcal{L}_\Omega(\overline{x})$ we can build a function F as (4.11) and get a C^2 diffeomorphism $\Phi : [0,\varepsilon] \to \mathbb{R}^n$ such that (4.12) holds. Thus, define the curve

$$\xi(t) := \Phi^{-1}(\overline{x} + td). \tag{4.16}$$

By items i) and ii) of Theorem 4.3.1, we obtain that

$$\lim_{t\to 0^+} \xi(t) = \overline{x} \quad \text{and} \quad \lim_{t\to 0^+} \frac{\xi(t) - \overline{x}}{t} = d. \tag{4.17}$$

From (4.12), we get that $g_j(\xi(t)) \in \mathrm{bd}^+(\mathbb{L}_{m_j})$ if $j \in D_0^B(\overline{x}) \cup D_B^0(\overline{x})$ and $g_j(\xi(t)) = 0$ if $j \in D_0^0(\overline{x})$. If $d \in D_0^{int}(\overline{x})$, then

$$\begin{aligned} g_j(\xi(t)) &= g_j(\xi(0)) + tDg_j(\xi(0))\xi'(0) + r(t), \quad \text{where } \lim_{t\to 0^+} \frac{r(t)}{t} = 0 \\ &= tDg_j(\overline{x})d + r(t) \in \mathrm{int}(\mathbb{L}_{m_j}), \end{aligned}$$

for all $t > 0$ small enough. Lastly, if $d \in D_B^{int}(\overline{x})$, we have that

$$\begin{aligned} \tilde{\phi}_j(\xi(t)) &= \tilde{\phi}_j(\xi(0)) + t\nabla\tilde{\phi}_j(\xi(0))\xi'(0) + r(t), \quad \text{where } \lim_{t\to 0^+} \frac{r(t)}{t} = 0 \\ &= t\nabla\tilde{\phi}_j(\overline{x})d + r(t) > 0, \end{aligned}$$

for all $t > 0$ small enough, that is, we obtain that $g_j(\xi(t)) \in \mathrm{int}(\mathbb{L}_{m_j})$. Therefore, Ref-McCormick holds at $\overline{x}$ as we wanted to show. □

Now, let us prove that Ref-McCormick is a constraint qualification.

Theorem 4.4.2. *Consider the problem (NSOCP) and let $\overline{x}$ be a feasible point such that Ref-McCormick holds. Then, Abadie's CQ holds at $\overline{x}$.*

Proof. It is enough to show that $\mathcal{L}_\Omega(\overline{x}) \subseteq \mathcal{T}_\Omega(\overline{x})$. Take a direction $d \in \mathcal{L}_\Omega(\overline{x})$ and a sequence $\{t_k\} \to 0^+$. Define

$$d^k := \frac{\xi(t_k) - \overline{x}}{t_k},$$

where ξ is given by (4.16). We have that $g_j(\overline{x} + t_k d^k) = g_j(\xi(t_k)) \in \mathbb{L}_{m_j}$ for all k large enough, $j = 1, \ldots, q$. Therefore, we have that $d \in \mathcal{T}_\Omega(\overline{x})$. □

Next, we present a figure summarizing the relation among the CQ's introduced in this chapter.

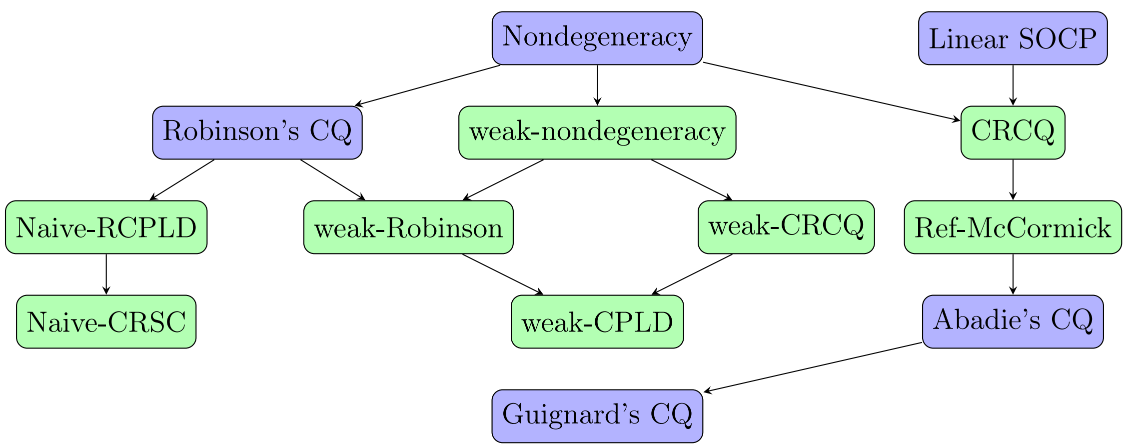

Figure 4.3: *Relation among the CQ's for (NSOCP). The boxes in blue are the well-known CQ's in nonlinear second-order cone programming and the green boxes are all the new proposals.*

To finish this section, let us analyze the second-order optimality conditions at a feasible point $\overline{x}$ of (NSOCP) such that Ref-McCormick holds. In Section 2.2 we presented some second-order optimality conditions for (NLP), among which we highlight the so-called Strong Second-Order Condition (SSOC) (2.13) that is based on the critical cone $C(\overline{x})_{\mathrm{NLP}}$ and it is more desirable because under SSOC it is possible to define a sufficient condition in order to have a local minimizer.

In the nonlinear programming context, it is known that a local minimizer $\overline{x}$ of (NLP) implies SSOC under LICQ [NW99, Theorem 12.5], under CRCQ [AES10, Theorem 3.1] and under RCRCQ [MS11b, Theorem 6]. Moreover, in [RS23] the authors reiterated that RCRCQ is the weakest constraint qualification that ensures the fulfillment of SSOC.

In order to analyze second-order optimality conditions as SSOC for (NSOCP), let us remember some important concepts of second-order information in a conic context. Let us start with the critical cone in the second-order cone programming context. Given a feasible point $\overline{x}$ of (NSOCP), the critical cone $C(\overline{x})$ as defined as

$$C(\overline{x}) := \mathcal{L}_{\Omega}(\overline{x}) \cap \{\nabla f(\overline{x})\}^{\perp}, \tag{4.18}$$

in a similar vein as we presented in the nonlinear programming context. In addition, if $\overline{x}$ is a KKT point of (NSOCP), that is, there exists $\mu_j \in \mathbb{L}_{m_j}$, $j = 1, \ldots, q$, such that (2.5) and (2.6) hold, from [BR05, Corollary 26] we can rewrite the critical cone in terms of the Lagrange multipliers. For such, consider the following indices sets

$$\begin{array}{rcl}
M_{int}^0 & := & \{j \mid g_j(\overline{x}) \in \mathrm{int}(\mathbb{L}_{m_j}),\ \mu_j = 0\} \\
M_B^0 & := & \{j \mid g_j(\overline{x}) \in \mathrm{bd}^+(\mathbb{L}_{m_j}),\ \mu_j = 0\} \\
M_B^B & := & \{j \mid g_j(\overline{x}) \in \mathrm{bd}^+(\mathbb{L}_{m_j}),\ \mu_j \in \mathrm{bd}^+(\mathbb{L}_{m_j})\} \\
M_0^0 & := & \{j \mid g_j(\overline{x}) = 0,\ \mu_j = 0\} \\
M_0^{int} & := & \{j \mid g_j(\overline{x}) = 0,\ \mu_j \in \mathrm{int}(\mathbb{L}_{m_j})\} \\
M_0^B & := & \{j \mid g_j(\overline{x}) = 0,\ \mu_j \in \mathrm{bd}^+(\mathbb{L}_{m_j})\}.
\end{array}$$

With these sets at hand, from [BR05, Corollary 26] we obtain

$$\begin{array}{rcl}
C(\overline{x}) & = & \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} Dg_j(\overline{x})d \in \mathcal{T}_{\mathbb{L}_{m_j}}(g_j(\overline{x})), & \text{if } \mu_j = 0 \\ Dg_j(\overline{x})d = 0, & \text{if } \mu_j \in \mathrm{int}(\mathbb{L}_{m_j}) \\ \langle Dg_j(\overline{x})d, \mu_j\rangle = 0, & \text{if } \mu_j \in \mathrm{bd}^+(\mathbb{L}_{m_j}),\ j \in J_B(\overline{x}) \\ Dg_j(\overline{x})d \in \mathrm{cone}(\Gamma_j \mu_j), & \text{if } \mu_j \in \mathrm{bd}^+(\mathbb{L}_{m_j}),\ j \in J_0(\overline{x}) \end{array} \right\} \\
& = & \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \langle Dg_j(\overline{x})d, \Gamma_j g_j(\overline{x})\rangle \geq 0, & j \in M_B^0 \\ \langle Dg_j(\overline{x})d, \mu_j\rangle = 0, & j \in M_B^B \\ Dg_j(\overline{x})d \in \mathbb{L}_{m_j}, & j \in M_0^0 \\ Dg_j(\overline{x})d = 0, & j \in M_0^{int} \\ Dg_j(\overline{x})d \in \mathrm{cone}(\Gamma_j \mu_j), & j \in M_0^B \end{array} \right\}.
\end{array}$$

Since we have an explicit form for the critical cone in the second-order cone programming context, let us analyze the quadratic form that will be evaluated on $C(\overline{x})$ to get SSOC. The natural answer for this topic would be the Hessian of the Lagrangian function

$$L(x, \mu_1, \ldots, \mu_q) := f(x) - \sum_{j=1}^{q} \langle \mu_j, g_j(x)\rangle$$

with respect to variable x. However, when we consider a second-order cone programming context, there is a difference when we compare to SSOC for (NLP), as we presented in (2.13). Based on this little discussion, let us introduce SSOC for (NSOCP). See [BR05] for more details.

Definition 4.4.3. *Consider the problem* (NSOCP) *and let* $\overline{x}$ *be a KKT point associated to a Lagrange multiplier* $(\mu_1, \ldots, \mu_q)$. *We say that the* Strong Second-Order Condition (SSOC) *holds at* $(\overline{x}, \mu_1, \ldots, \mu_q)$ *if*

$$d^T \nabla^2_{xx} L(\overline{x}, \mu_1, \ldots, \mu_q) d + d^T \mathcal{H}(\overline{x}, \mu_1, \ldots, \mu_q) d \geq 0,$$

for all $d \in C(\overline{x})$, *where* $\mathcal{H}(\overline{x}, \mu_1, \ldots, \mu_q) = \sum_{j=1}^{q} \mathcal{H}_j(\overline{x}, \mu_j)$ *with*

$$\mathcal{H}_j(\overline{x}, \mu_j) := \begin{cases} -\dfrac{[\mu_j]_0}{[g_j(\overline{x})]_0} Dg_j(\overline{x})^T \Gamma_j Dg_j(\overline{x}), & \text{if } g_j(\overline{x}) \in \operatorname{bd}^+(\mathbb{L}_{m_j}), \\ 0, & \text{otherwise.} \end{cases}$$

Notice that we have $m_1 = \ldots = m_q = 1$, the condition above reduces to SSOC as (2.13). The difference in the definition above relies on the fact that we have to take into account the term

$$\begin{aligned} d^T \mathcal{H}(\overline{x}, \mu_1, \ldots, \mu_q) d &= d^T \left(\sum_{j \in I_B(\overline{x})} -\frac{[\mu_j]_0}{[g_j(\overline{x})]_0} Dg_j(\overline{x})^T \Gamma_j Dg_j(\overline{x}) \right) d \\ &= \sum_{j \in M_B^B} -\frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \left(Dg_j(\overline{x}) d\right)^T \Gamma_j Dg_j(\overline{x}) d \qquad (4.19) \\ &= \sum_{j \in M_B^B} -\frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \langle Dg_j(\overline{x}) d, \Gamma_j Dg_j(\overline{x}) d \rangle, \end{aligned}$$

that is known as the "sigma term". It represents the curvature of the second-order cone $\mathbb{L}_{m_j}$ along $Dg_j(\overline{x})d$ where $d \in C(\overline{x})$.

In order to prove that Ref-McCormick is a second-order constraint qualification that ensures the fulfillment of SSOC according to the Definition 4.4.3, we will need the following auxiliary result, which is kind of a "complementarity function".

Lemma 4.4.1. *Let* $\overline{x}$ *be a local minimizer of* (NSOCP) *such that Ref-McCormick holds associated to the Lagrange multiplier* $(\mu_1, \ldots, \mu_q)$. *Take a direction* $d \in \mathcal{L}_\Omega(\overline{x})$. *Consider the function* $R : \mathbb{R}_+ \to \mathbb{R}$ *given by*

$$R(t) := \sum_{j \in M_0^{int} \cup M_0^B} \langle g_j(\xi(t)), \mu_j \rangle + \sum_{j \in M_B^B} \frac{[\mu_j]_0}{2[g_j(\overline{x})]_0} g_j(\xi(t))^T \Gamma_j g_j(\xi(t)), \qquad (4.20)$$

where ξ *comes from Definition 4.4.2 is such that* $\xi(0) = \overline{x}$ *and* $\xi'(0) = d$. *Then, there exists an* $\varepsilon > 0$ *such that* $R(t) = 0$ *for all* $t \in [0, \varepsilon)$ *and, in particular,* $R'(0) = R''(0) = 0$.

Proof. Take an index j such that $j \in M_B^B$. If $\mu_j = 0$ we obtain that

$$\frac{[\mu_j]_0}{2[g_j(\overline{x})]_0} g_j(\xi(t))^T \Gamma_j g_j(\xi(t)) = 0. \qquad (4.21)$$

Otherwise, if $\mu_j \neq 0$ we obtain that $\mu_j = \Gamma_j Dg_j(\overline{x}) d$. In addition, using the complementarity condition (2.6), we have that $\langle g_j(\overline{x}), \mu_j \rangle = 0$ and then we obtain that $Dg_j(\overline{x}) d \in \operatorname{bd}^+(\mathbb{L}_{m_j})$. From Theorem 4.4.1, it follows that for all $t \geq 0$ small enough,

$$g_j(\xi(t)) \in \operatorname{bd}^+(\mathbb{L}_{m_j}),$$

that is, (4.21) holds.

Now, take any index j such that $j \in M_0^{int}$. From Theorem 4.4.1, we have that $g_j(\xi(t)) = 0$ for all $t \geq 0$ small enough.

Last, take an index $j \in M_0^B$. It follows from the definition of M_0^B that $Dg_j(\overline{x}) d \in \operatorname{bd}^+(\mathbb{L}_{m_j})$. Thus, from Theorem 4.4.1 we have that for each $t \geq 0$ small enough there exists an $\alpha_t^j \geq 0$ such that $g_j(\xi(t)) = \alpha_t^j Dg_j(\overline{x}) d$, which means that

$$\langle g_j(\xi(t)), \mu_j\rangle = \langle \alpha_t^j Dg_j(\overline{x})d, \mu_j\rangle = 0,$$

because $Dg_j(\overline{x})d \in \mathrm{cone}(\Gamma_j\mu_j)$. Therefore, exists an $\varepsilon > 0$ such that $R(t) = 0$ for all $t \in [0, \varepsilon)$. For the sake of completeness, let us compute $R'(0)$ and $R''(0)$.

Differentiating $R(t)$ once and taking $t = 0$, give us

$$R'(t) = \sum_{j\in M_0^{int}\cup M_0^B} \langle Dg_j(\xi(t))\xi'(t), \mu_j\rangle + \sum_{j\in M_B^B} \frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \langle Dg_j(\xi(t))^T\Gamma_j g_j(\xi(t)), \xi'(t)\rangle$$

and then

$$\begin{aligned} R'(0) &= \sum_{j\in M_0^{int}\cup M_0^B} \langle Dg_j(\overline{x}))d, \mu_j\rangle + \sum_{j\in M_B^B} \frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \langle Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x}), d\rangle \\ &= \sum_{j\in M_0^{int}\cup M_0^B} \langle Dg_j(\overline{x}))d, \mu_j\rangle + \sum_{j\in M_B^B} \frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \langle \nabla\tilde{\phi}_j(\overline{x}), d\rangle = 0. \end{aligned}$$

Now, differentiating $R(t)$ twice and taking $t = 0$, we obtain

$$\begin{aligned} R''(0) &= \sum_{j\in M_0^{int}\cup M_0^B} \langle D^2 g_j(\overline{x})[d,d], \mu_j\rangle + \sum_{j\in M_0^{int}\cup M_0^B} \langle Dg_j(\overline{x})^T\mu_j, \xi''(0)\rangle \\ &+ \sum_{j\in M_B^B} \frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \langle D^2 g_j(\overline{x})[d,d], \Gamma_j g_j(\overline{x})\rangle + \sum_{j\in M_B^B} \frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \langle Dg_j(\overline{x})d, \Gamma_j Dg_j(\overline{x})d\rangle \\ &+ \sum_{j\in M_B^B} \frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \left(Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x})\right)^T \xi''(0) = 0. \end{aligned}$$

Rearranging the terms and recalling that $\nabla\tilde{\phi}_j(\overline{x}) = Dg_j(\overline{x})^T\Gamma_j g_j(\overline{x})$ when $j \in M_B^B$, and

$$\mu_j = \frac{[\mu_j]_0}{[g_j(\overline{x})]_0}\Gamma_j g_j(\overline{x}),\ j \in M_B^B,$$

give us

$$\begin{aligned} R''(0) &= \sum_{j\in M_B^B\cup M_0^{int}\cup M_0^B} \langle D^2 g_j(\overline{x})[d,d], \mu_j\rangle + \sum_{j\in M_B^B\cup M_0^{int}\cup I_0^B} \langle Dg_j(\overline{x})^T\mu_j, \xi''(0)\rangle \\ &+ \sum_{j\in M_B^B} \frac{[\mu_j]_0}{[g_j(\overline{x})]_0} \langle Dg_j(\overline{x})d, \Gamma_j Dg_j(\overline{x})d\rangle = 0. \end{aligned} \tag{4.22}$$

□

With this result at hand, we will be able to prove that under Ref-McCormick, a local minimizer $\overline{x}$ of (NSOCP) satisfies SSOC according to Definition 4.4.3 for any Lagrange multiplier associated.

Theorem 4.4.3. *Let $\overline{x}$ be a local minimizer of the problem (NSOCP) such that Ref-McCormick holds. Then, for any Lagrange multiplier $(\mu_1, \ldots, \mu_q)$, we have that $(\overline{x}, \mu_1, \ldots, \mu_q)$ satisfies SSOC.*

Proof. Just for commodity, define the set of indices $M := M_B^B \cup M_0^{int} \cup M_0^B$. Since $\overline{x}$ satisfies the KKT conditions, from (2.5) we have that

$$\nabla f(\overline{x}) - \sum_{j=1}^{q} Dg_j(\overline{x})^T\mu_j = 0$$

that can be rewritten as

$$\nabla f(\overline{x}) - \sum_{j\in M} Dg_j(\overline{x})^T\mu_j = 0$$

Let $d \in C(\overline{x})$ be any direction. We have that

$$\left\langle \nabla f(\overline{x}) - \sum_{j\in M} Dg_j(\overline{x})^T\mu_j, d \right\rangle = 0,$$

which implies that $\nabla f(\overline{x})^T d = 0$. In fact, we have that

$$\langle Dg_j(\overline{x})d, \mu_j\rangle = 0,\ j \in M_B^B; \quad Dg_j(\overline{x})d = 0,\ j \in M_0^{int}$$

and, in addition,

$$Dg_j(\overline{x})d = \alpha_j\Gamma_j\mu_j, \text{ for some } \alpha_j \geq 0 \text{ and } j \in M_0^B$$

that implies

$$\langle Dg_j(\overline{x})^T\mu_j, d\rangle = \langle \mu_j, Dg_j(\overline{x})d\rangle = 0.$$

Since Ref-McCormick holds at $\overline{x}$, there exists $\xi : [0, \varepsilon) \to \mathbb{R}^n$ twice continuously differentiable such that $\xi(0) = \overline{x}$ and $\xi'(0) = d$. Let $\varphi(t) := f(\xi(t))$. Since $\overline{x}$ is a local minimizer f in Ω, then $t = 0$ is a local minimizer of φ, that is,

$$\varphi'(0) = \nabla f(\overline{x})^T d = 0,$$

and, moreover

$$\varphi''(0) = d^T\nabla^2 f(\overline{x})d + \nabla f(\overline{x})^T\xi''(0) \geq 0. \tag{4.23}$$

In order to compute $\nabla f(\overline{x})^T\xi''(0)$, consider the function R as defined in (4.20). By Lemma 4.4.1, we have that $R'(0) = R''(0) = 0$. Subtracting (4.22) from (4.23), we obtain

$$\begin{aligned} d^T\nabla^2 f(\overline{x})d \quad &- \sum_{j\in M}\langle D^2g_j(\overline{x})[d,d], \mu_j\rangle - \sum_{j\in M_B^B}\frac{[\mu_j]_0}{[g_j(\overline{x})]_0}\langle Dg_j(\overline{x})d, \Gamma_j Dg_j(\overline{x})d\rangle \\ &+ \left\langle \nabla f(\overline{x}) - \sum_{j\in M} Dg_j(\overline{x})^T\mu_j, \xi''(0)\right\rangle \geq 0. \end{aligned}$$

and, from (4.19), we obtain

$$d^T\nabla^2_{xx}L(\overline{x}, \mu_1, \ldots, \mu_q)d + d^T\mathcal{H}(\overline{x}, \mu_1, \ldots, \mu_q)d \geq 0,$$

for all $d \in C(\overline{x})$. Therefore, SSOC holds at $\overline{x}$ as we wanted to prove. □

The importance of the result above relies on some facts. First, to the best of our knowledge, this is the weakest constraint qualification in second-order cone programming context that ensures SSOC, once it is known that under Robinson's CQ SSOC does not hold and, in addition, Ref-McCormick is strictly weaker than the nondegeneracy condition. Second, under Ref-McCormick we do not necessarily have that the set of Lagrange multipliers is compact, but SSOC still holds for any Lagrange multiplier.

Next, we will present a result that is new even in the nonlinear programming context, which explains why SSOC holds under Ref-McCormick for any Lagrange multiplier.

Theorem 4.4.4. *Let $\overline{x}$ be a local minimizer of* (NSOCP) *such that Ref-McCormick holds. The quadratic form*

$$d^T\nabla^2_{xx}L(\overline{x}, \mu_1, \ldots, \mu_q)d + d^T\mathcal{H}(\overline{x}, \mu_1, \ldots, \mu_q)d \tag{4.24}$$

for $d \in C(\overline{x})$, does not depend on $(\mu_1, \ldots, \mu_q) \in \Lambda(\overline{x})$.

Proof. Let $(\mu_1, \ldots, \mu_q)$ be any Lagrange multiplier associated to $\overline{x}$. It is enough to show that

$$-\sum_{j\in M}\langle D^2 g_j(\overline{x})[d,d], \mu_j\rangle - \sum_{j\in M_B^B}\frac{[\mu_j]_0}{[g_j(\overline{x})]_0}\langle Dg_j(\overline{x})d, \Gamma_j Dg_j(\overline{x})d\rangle$$

does not depend on $(\mu_1, \ldots, \mu_q) \in \Lambda(\overline{x})$. From (4.22), we have that

$$\sum_{j\in M}\langle Dg_j(\overline{x})^T\mu_j, \xi''(0)\rangle = -\sum_{j\in M}\langle D^2 g_j(\overline{x})[d,d], \mu_j\rangle - \sum_{j\in M_B^B}\frac{[\mu_j]_0}{[g_j(\overline{x})]_0}\langle Dg_j(\overline{x})d, \Gamma_j Dg_j(\overline{x})d\rangle, \quad (4.25)$$

where $M := M_B^B \cup M_0^{int} \cup M_0^B$. Since $\overline{x}$ is a KKT point, we have that (2.5) holds, that is,

$$\nabla f(\overline{x}) = \sum_{j\in M} Dg_j(\overline{x})^T\mu_j. \quad (4.26)$$

Substituting (4.26) in (4.25), we get that

$$-\sum_{j\in M}\langle D^2 g_j(\overline{x})[d,d], \mu_j\rangle - \sum_{j\in M_B^B}\frac{[\mu_j]_0}{[g_j(\overline{x})]_0}\langle Dg_j(\overline{x})d, \Gamma_j Dg_j(\overline{x})d\rangle = \langle\nabla f(\overline{x}), \xi''(0)\rangle,$$

whose the right part of the equation above does not depend on the Lagrange multiplier. □

The theorem above shows that the quadratic form (4.24) is constant when we vanish all the Lagrange multipliers associated to a local minimizer in which Ref-McCormick holds. This result is stronger than the result obtained in [BHRV18, Theorem 3.3] for nonlinear programming problems, where the authors showed that under WCR the Hessian of the Lagrangian does not depend on the Lagrange multiplier when we consider the direction on the critical subspace.

It is important to mention that in the proofs of the Lemma 4.4.1, Theorem 4.4.3 and Theorem 4.4.4, we did not use in an explicit way the fact that $H(\overline{x})$ is closed, just for guarantee the existence of a Lagrange multiplier. It means that the second-order results obtained are related to the form of how we define the facial constant rank property (Definition 4.3.1) and the properties of the curve required by the Ref-McCormick condition in (4.15).

4.5 New proposals for second-order constraint qualifications for NLP

In this section we will consider the problem (NLP) and explore more the fact that the Ref-McCormick proposed in [FSS22] for nonlinear programming problems is a strong second-order constraint qualification, that is, it is enough to guarantee SSOC for any Lagrange multiplier associated to a local minimizer $\overline{x}$ of (NLP). First of all, let us recall some results obtained in the second-order cone programming problem for nonlinear programming problem.

Proposition 4.5.1. *([AES10, Theorem 3.1]) Let $\overline{x}$ be a local minimizer of (NLP) that satisfies NLP-Ref-McCormick. Take a direction $d \in \mathcal{L}_\Omega(\overline{x})_{\text{NLP}}$ and consider the function*

$$R(t) = \sum_{i=1}^{p}\lambda_i h_i(\xi(t)) + \sum_{j\in A(\overline{x})}\mu_j g_j(\xi(t)), \quad (4.27)$$

where ξ comes from Definition 4.4.1 is such that $\xi(0) = \overline{x}$ and $\xi'(0) = d$. Then, there exists an $\varepsilon > 0$ such that $R(t) = 0$ for all $t \in [0, \varepsilon)$ and, in particular, $R'(0) = R''(0) = 0$.

The proof of the proposition above is contained in the proof of the [AES10, Theorem 3.1], where the authors showed that CRCQ is enough to guarantee SSOC for any Lagrange multiplier in a nonlinear programming context. They also obtained the following expression for $R''(0)$

$$\begin{aligned} R''(0) &= d^T\left(\sum_{i=1}^{p}\lambda_i\nabla^2 h_i(\overline{x})) - \sum_{j\in A(\overline{x})}\mu_j\nabla^2 g_j(\overline{x})\right)d + \\ &+ \left(\sum_{i=1}^{p}\lambda_i\nabla h_i(\overline{x}) - \sum_{j\in A(\overline{x})}\mu_j\nabla g_j(\overline{x})\right)^T\xi''(0) = 0. \end{aligned} \tag{4.28}$$

With this result at hand, we can reproduce the result obtained in Theorem 4.4.4 for (NLP).

Theorem 4.5.1. *Let $\overline{x}$ be a KKT point of (NLP) such that NLP-Ref-McCormick holds at $\overline{x}$. The quadratic form*

$$d^T\left(\nabla^2 f(\overline{x}) + \sum_{i=1}^{p}\lambda_i\nabla^2 h_i(\overline{x}) - \sum_{j\in A(\overline{x})}\mu_j\nabla^2 g_j(\overline{x})\right)d,\ \ d\in C(\overline{x})_{\text{NLP}} \tag{4.29}$$

does not depend on $(\lambda,\mu)\in\Lambda(\overline{x})$.

Proof. Let $(\tilde{\lambda},\tilde{\mu})$, $(\overline{\lambda},\overline{\mu})\in\Lambda(\overline{x})$ be different Lagrange multipliers and take any direction $d\in C(\overline{x})_{\text{NLP}}$. Computing $d^T\left(\nabla^2_{xx}L(\overline{x},\tilde{\lambda},\tilde{\mu}) - \nabla^2_{xx}L(\overline{x},\overline{\lambda},\overline{\mu})\right)d$, give to us

$$d^T\left[\sum_{i=1}^{p}\tilde{\lambda}_i\nabla^2 h_i(\overline{x}) - \sum_{j\in A(\overline{x})}\tilde{\mu}_j\nabla^2 g_j(\overline{x}) - \left(\sum_{i=1}^{p}\overline{\lambda}_i\nabla^2 h_i(\overline{x}) - \sum_{j\in A(\overline{x})}\overline{\mu}_j\nabla^2 g_j(\overline{x})\right)\right]d.$$

Since $\overline{x}$ is a KKT point, we have that

$$\begin{aligned} \nabla f(\overline{x})^T\xi''(0) &= -\left(\sum_{i=1}^{p}\tilde{\lambda}_i\nabla h_i(\overline{x}) - \sum_{j\in A(\overline{x})}\tilde{\mu}_j\nabla g_j(\overline{x})\right)^T\xi''(0) \\ &= -\left(\sum_{i=1}^{p}\overline{\lambda}_i\nabla h_i(\overline{x}) - \sum_{j\in A(\overline{x})}\overline{\mu}_j\nabla g_j(\overline{x})\right)^T\xi''(0) \end{aligned}$$

Using the relation above and the equation (4.28), we obtain

$$d^T\left(\sum_{i=1}^{p}\tilde{\lambda}_i\nabla^2 h_i(\overline{x})) - \sum_{j\in A(\overline{x})}\tilde{\mu}_j\nabla^2 g_j(\overline{x})\right)d = d^T\left(\sum_{i=1}^{p}\overline{\lambda}_i\nabla^2 h_i(\overline{x})) - \sum_{j\in A(\overline{x})}\overline{\mu}_j\nabla^2 g_j(\overline{x})\right)d.$$

Therefore $d^T\left(\nabla^2_{xx}L(\overline{x},\tilde{\lambda},\tilde{\mu}) - \nabla^2_{xx}L(\overline{x},\overline{\lambda},\overline{\mu})\right)d = 0$, as we wanted to show. □

Notice that when we take a direction $d\in\mathcal{L}_\Omega(\overline{x})_{\text{NLP}}$ and we build the curve ξ according to Definition 4.4.1, there are some constraints $j\in A(\overline{x})$ that are not so relevant in order to evaluate feasibility of $g_j(\xi(t))$ for $t\geq 0$ small enough. Indeed, remember that from Definition 4.4.1, we may have cases where $\nabla g_j(\overline{x})^T d > 0$ and then we get

$$g_j(\xi(t)) > 0,$$

that is, at least for this constraint we get feasibility "for free". Moreover, if there exists a nonzero direction $d\in\mathcal{L}_\Omega(\overline{x})_{\text{NLP}}$ such that $\nabla g_j(\overline{x})^T d > 0$ for all $j\in A(\overline{x})$, this direction is not worrisome for any active constraint when we analyze feasibility through this direction. On the other hand, the existence of such direction is pivotal when we take into account the fulfillment of MFCQ, which can be rewritten as

i) the set $\{\nabla h_i(\overline{x})\}_{i=1}^p$ is linearly independent;

ii) there exists d such that $\nabla g_j(\overline{x})^T d > 0$ for all $j \in A(\overline{x})$ and $\nabla h_i(\overline{x})^T d = 0$ for all $i = 1, \ldots, p$.

It is known that only MFCQ is not enough to guarantee even WSOC, that is, the quadratic form (4.29) is not necessarily positive semidefinite when we vanish all directions in the critical subspace $S(\overline{x})_{\text{NLP}}$ defined in (2.14). In [AMS07], the authors showed that MFCQ+WCR implies WSOC. In addition, another condition can be added to MFCQ in order to guarantee WSOC.

Proposition 4.5.2. *[AMS07, Hae17, Mas19] Let $\overline{x}$ be a local minimizer of (NLP) such that MFCQ holds. Define the matrix $M(x)$ whose columns are $(\nabla h_i(x), \nabla g_j(x))_{i=1,\ldots,p;\ j\in A(\overline{x})}$. If there exists a neighborhood V of $\overline{x}$ such that*

$$\operatorname{rank}(M(x)) \leq \operatorname{rank}(M(x)) + 1 \tag{4.30}$$

for all $x \in V$, then there exists $(\lambda, \mu) \in \Lambda(\overline{x})$ such that WSOC holds.

The condition above was called in [ABHS17, Definition 4.1] as *Modified Mangasarian-Fromovitz (MMF)*. With these results at hand, let us resume the discussion about the nonzero directions $d \in \mathcal{L}_\Omega(\overline{x})_{\text{NLP}}$ such that there exists $j \in A(\overline{x})$ satisfying $\nabla g_j(\overline{x})^T d = 0$. Given a direction $d \in \mathcal{L}_\Omega(\overline{x})_{NLP}$ such that $d \neq 0$, let us define the following set:

$$J_d(\overline{x}) := \{j \in A(\overline{x}) \mid \nabla g_j(\overline{x})^T d = 0\}. \tag{4.31}$$

This set leads us to define the following

$$J_0(\overline{x}) := \bigcup_{\substack{d\in\mathcal{L}_\Omega(\overline{x})_{\text{NLP}} \\ d\neq 0}} J_d(\overline{x}) = \{j \in A(\overline{x}) \mid \exists\, d \in \mathcal{L}_\Omega(\overline{x})_{\text{NLP}},\ d \neq 0 :\ \nabla g_j(\overline{x})^T d = 0\}. \tag{4.32}$$

Notice that $J_0(\overline{x}) \subseteq A(\overline{x})$ and this inclusion might be strict and that this set captures exactly which constraints can be violated when walking towards direction d. This means that we do not need to consider all the indexes $j \in A(\overline{x})$. Indeed, if there exists $j_0 \in A(\overline{x})$ such that $\nabla g_{j_0}(\overline{x})^T d > 0$ for all $d \in \mathcal{L}_\Omega(\overline{x})_{\text{NLP}}$ such that $d \neq 0$, then we have that $g_{j_0}(\overline{x} + td) \geq 0$ for all $t \geq 0$ small enough. With this set in mind, let us define the following assumption:

Assumption 1. *Consider the problem (NLP) and let $\overline{x}$ be a feasible point. We say that Assumption 1 (A1) holds at $\overline{x} \in \Omega$, if for every index $j \in J_0(\overline{x})$, we have that there exists a neighborhood V of $\overline{x}$ such that*

$$\nabla g_j(x) \in \operatorname{span}\left(\{\nabla h_i(x)\}_{i=1}^p\right) \tag{4.33}$$

for all $x \in V$.

Notice that A1 is strictly stronger than (4.30). On the one hand, by definition we have that A1 implies (4.30). On the other hand, if the set $\{\nabla h_i(\overline{x})\}_{i=1}^p \cup \{\nabla g_j(\overline{x})\}_{j\in A(\overline{x})}$ is linearly independent, then we have that the hypothesis of Proposition 4.5.2 is satisfied while A1 does not. Later we will show that MFCQ+A1 implies NLP-Ref-McCormick. For such, we will need the following result from [AMS05].

Lemma 4.5.1. *([AMS05, Lemma 2]) Let the functions $f, f_1, \ldots, f_q : D \to \mathbb{R}^n$ be twice continuously differentiable; let $x \in D$, with D an open set. Assume that the gradients $\nabla f_1(x), \ldots, \nabla f_q(x)$ are linearly independent and that $\nabla f(y)$ is a linear combination of $\nabla f_1(y), \ldots, \nabla f_q(y)$ for all $y \in D$. In particular,*

$$\nabla f(x) = \sum_{i=1}^q \alpha_i \nabla f_i(x). \tag{4.34}$$

Then, there exists $D_2 \subset \mathbb{R}^q$*, an open neighborhood of* $(f_1(x), \ldots, f_q(x))$*, and a function* $\varphi : D_2 \to \mathbb{R}$*,* $\varphi \in C^2(D_2)$*, such that, for all* $y \in D_1$*, we have that* $(f_1(y), \ldots, f_q(y)) \in D_2$ *and*

$$f(y) = \varphi(f_1(y), \ldots, f_q(y)).$$

Moreover, for all $i = 1, \ldots, q$*,*

$$\alpha_i = \frac{\partial \varphi}{\partial u_i}(f_1(x), \ldots, f_q(x)). \tag{4.35}$$

In the following theorem we will show that a weaker hypothesis than MFCQ+A1 is enough to guarantee the fulfillment of NLP-Ref-McCormick.

Theorem 4.5.2. *Let* $\overline{x}$ *be a feasible point of (NLP). If there exists a neighborhood* V *of* $\overline{x}$ *such that* $\{\nabla h_i(x)\}_{i=1}^p$ *has constant rank for all* $x \in V$ *and, in addition, A1 holds at* $\overline{x}$*, then NLP-Ref-McCormick also holds at* $\overline{x}$*.*

Proof. Take a nonzero direction $d \in \mathcal{L}_\Omega(\overline{x})$ and consider the set J_d previously defined. Let $I \subset \{1, 2, \ldots, p\}$ be a subset such that $\{\nabla h_i(\overline{x})\}_{i \in I}$ is a basis for $\text{span}\{\nabla h_i(\overline{x})\}_{i=1}^p$. Define the function $H(x) := (h_i(x))_{i \in I}$ and consider the set

$$M := \{x \in \mathbb{R}^n \mid H(x) = H(\overline{x}) = 0\}.$$

By Lyusternik Theorem (Theorem 4.2.1), it follows that

$$\mathcal{T}_M(\overline{x}) = \text{Ker}(H(\overline{x})),$$

where $\mathcal{T}_M(\overline{x})$ is the tangent cone to the set M at the point $\overline{x}$. Since $d \in \ker(H(\overline{x}))$ (because we are taking $d \in \mathcal{L}_\Omega(\overline{x})_{\text{NLP}}$ and H is twice continuously differentiable, then there exists a twice continuously differentiable arc $r(t)$ such that

$$\overline{x} + td + r(t) \in M, \quad \lim_{t \to 0^+} \frac{r(t)}{t} = 0,$$

that is,

$$H(\overline{x} + td + r(t)) = H(\overline{x}) = 0, \tag{4.36}$$

or, in other words,

$$h_i(\overline{x} + td + r(t)) = h_i(\overline{x}) = 0, \ i \in I.$$

Now, take any index $i_0 \in \{1, 2, \ldots, p\} \backslash I$. Since $\{\nabla h_i(\overline{x})\}_{i \in I}$ is linearly independent, so $\{\nabla h_i(x)\}_{i \in I}$ also is for all x close enough to $\overline{x}$. Moreover, using the fact that the rank of $\{\nabla h_i(x)\}_{i=1}^p$ remains constant for x close to $\overline{x}$, then we obtain that $\nabla h_{i_0}(x) \in \text{span}\{\nabla h_i(x)\}_{i \in I}$. Thus, by Lemma (4.5.1), there is a C^2 function φ_{i_0} such that $h_{i_0}(x) = \varphi_{i_0}(\{h_i(x)\}_{i \in I})$, for all x near to $\overline{x}$. Thus, we obtain

$$\varphi_{i_0}(\{h_i(\overline{x})\}_{i \in I}) = h_{i_0}(\overline{x}) = 0.$$

In addition,

$$\begin{aligned} h_{i_0}(\overline{x} + td + r(t)) &= \varphi_{i_0}(\{h_i(\overline{x} + td + r(t))\}_{i \in I}) \\ &= \varphi_{i_0}(\{h_i(\overline{x})\}_{i \in I}) \\ &= h_{i_0}(\overline{x}) = 0. \end{aligned}$$

It follows that $h_i(\overline{x} + td + r(t)) = 0$ for all $t \geq 0$ small enough and $i = 1, 2, \ldots, p$.

Proceeding in a similar way, now consider the indexes $j \in J_d$. By hypothesis, we have that

$$\nabla g_j(x) \in \text{span}\{\nabla h_i(x)\}_{i \in I}.$$

Since $\{\nabla h_i(x)\}_{i \in I}$ is linearly independent, by Lemma (4.5.1), there is a C^2 function φ_j such that

$g_j(x) = \varphi_j(\{h_i(x)\}_{i\in I})$, for all x near $\overline{x}$. Thus, we obtain

$$\varphi_j(\{h_i(\overline{x})\}_{i\in I}) = g_j(\overline{x}) = 0.$$

In addition,

$$\begin{aligned} g_j(\overline{x}+td+r(t)) &= \varphi_j(\{h_i(\overline{x}+td+r(t))\}_{i\in I}) \\ &= \varphi_j(\{h_i(\overline{x})\}_{i\in I}) \\ &= g_j(\overline{x}) = 0. \end{aligned}$$

It follows that $g_j(\overline{x}+td+r(t)) = 0$ for all $t \geq 0$ small enough with $j \in J_d$, that is, for $j \in A(\overline{x})$ such that $\nabla g_j(\overline{x})^T d = 0$.

Lastly, for $j \in \{1, 2, \ldots, m\} \setminus J_d$, we obtain that

$$g_j(\overline{x}+td+r(t)) = g_j(\overline{x}) + t\nabla g_j(\overline{x})^T d + o(t) \geq 0,$$

for all $t \geq 0$ small enough. Therefore, define $\xi(t) := \overline{x}+td+r(t)$ and it follows that NLP-Ref-McCormick holds at $\overline{x}$. □

Corollary 4.5.1. *If $\{\nabla h_i(\overline{x})\}_{i=1}^p$ is linearly independent and, moreover, A1 holds at $\overline{x}$, then NLP-Ref-McCormick holds at $\overline{x}$.*

Corollary 4.5.2. *If MFCQ + A1 holds at $\overline{x}$, then NLP-Ref-McCormick also holds at $\overline{x}$.*

The results above explain another point of view on the relation between constant rank-type conditions and NLP-Ref-McCormick. In addition, Corollary 4.5.2 showed the additional assumption that can be required to MFCQ in order to have SSOC. Since in the Theorem 4.5.2 we considered just the constant rank of $\{\nabla h_i(x)\}_{i=1}^p$ in a neighborhood of a feasible point $\overline{x}$, we may consider then other constant rank-type constraint qualifications weaker than MFCQ such that, when added to the Assumption A1, guarantees the fulfillment of SSOC for (NLP). This leads us to the following constraint qualification:

Definition 4.5.1. *[AHSS12a] Consider the problem (NLP) and let $\overline{x}$ be a feasible point. We say that the* Relaxed Constant Positive Linear Dependence *constraint qualification (RCPLD) holds at $\overline{x}$ if fixed a set $B \subseteq \{1, \ldots, p\}$ such that $\{\nabla h_i(\overline{x})\}_{i\in B}$ is a basis for* $\operatorname{span}\{\nabla h_i(\overline{x}) \mid i = 1, \ldots, p\}$*, the following statements hold:*

i) $\{\nabla h_i(x)\}_{i=1}^p$ has constant rank around $\overline{x}$;

ii) for every $J \subseteq A(\overline{x})$, if $\{\nabla h_i(\overline{x}), \nabla g_j(\overline{x}) \mid i \in B,\ j \in J\}$ is positive linearly dependent, then $\{\nabla h_i(x), \nabla g_j(x) \mid i \in B,\ j \in J\}$ is positive linearly dependent for all x around $\overline{x}$.

Notice that in the definition above, we have the requirement of the constant rank of $\{\nabla h_i(x)\}_{i=1}^p$ around $\overline{x}$. In [AHSS12b, Theorem 4.3] the authors showed that RCPLD implies CRSC and in [AHSS12a] the authors showed that RCPLD is weaker than MFCQ and than RCRCQ, it means that RCPLD does not imply WSOC. However, using Assumption A1 and Theorem 4.5.2, give us the following result:

Corollary 4.5.3. *Consider the problem (NLP) and let $\overline{x}$ be a feasible point. If RCPLD and Assumption A1 hold at $\overline{x}$, then NLP-Ref-McCormick holds.*

In order to summarize the results present in this section, let us introduce the following constraint qualifications for nonlinear programming problems.

Definition 4.5.2. *Let $\overline{x}$ be a feasible point of (NLP). We say that:*

i) the span-regularity CQ *holds at $\overline{x}$ if $\{\nabla h_i(\overline{x})\}_{i=1}^p$ is linearly independent and Assumption A1 also holds at $\overline{x}$;*

ii) the span-constant rank CQ *holds* $\overline{x}$ *if there exists a neighborhood* V *of* $\overline{x}$ *such that the rank of* $\{\nabla h_i(y)\}_{i=1}^p$ *remains constant for all* $y \in V$ *and Assumption A1 also holds at* $\overline{x}$*;*

By construction, we have that span-regularity CQ implies span-constant rank CQ. And from Theorem 4.5.2 both definitions implies Ref-McCormick. Thus, they are second-order constraint qualifications. Moreover, these constraint qualifications have strong second-order properties, that is, if $\overline{x}$ is a local minimizer such that span-constant rank CQ holds, then SSOC holds for any Lagrange multiplier associated to $\overline{x}$. we will establish the relation among these CQ's and LICQ.

Example 4.5.1. *Consider the constraints* $g_j : \mathbb{R}^n \to \mathbb{R}$ *given by* $g_j(x) := x_j$ *for* $j = 1, \ldots, n-1$ *and the constraint* $h(x) := x_n$*. It follows that the gradients of the constraints are* $\{e_i\}_{i=1}^n$*, where* e_i *denotes the i-th canonical vector of* $\mathbb{R}^n$*. Thus, we get that LICQ holds and that Span-Regularity fails at the point* $\overline{x} = 0$*.*

On the other hand, consider the constraints $g_j, h_i : \mathbb{R}^n \to \mathbb{R}$ *given by* $g_j(x) = h_i(x) := x_j$ *for* $j = 1, \ldots, n$ *at the feasible point* $\overline{x} = 0$*. We have that the set* $\{\nabla h_i(\overline{x})\}_{i=1}^p$ *is linearly independent and A1 is trivially satisfied, that is, Span-Regularity holds at* $\overline{x}$*. Moreover, it is possible to see that LICQ fails at* $\overline{x}$*.*

The other constant rank-type constraint qualification that is well-known in nonlinear programming problems is the so called CRSC that was introduced by Andreani et al. in [AHSS12b] and we recalled in Section 2.2. The authors showed that this condition is implied by RCRCQ and MFCQ. Due to the second fact, then we get that CRSC does not imply SSOC, even if it has good properties inherited by RCRCQ. Thus, the natural question that arises is the following: is there any additional assumption that can be added to CRSC in order to get SSOC and keep it weaker than RCRCQ? The answer is yes and we will show it now. For such, let us consider two approaches.

The first one was presented by Andreani et al. in [AHMR23a] in a conic context (which encompasses the problem (NSOCP)), where the authors noticed that from CRSC some inequality constraints have the behavior of equality constraints. Thus, it is not necessary to vanish all the subsets of inequality constraints as it is required in RCRCQ, just the ones that are not in the subspace component, that is, in the set J_{NLP}^- defined in (2.15). We present here nonlinear programming problem version.

Definition 4.5.3. *([AHMR23a, Definition 4.1]) Let* $\overline{x}$ *be a feasible point of (NLP). We say that the* Strong-CRSC *condition holds at* $\overline{x}$ *if there is a neighborhood* V *of* $\overline{x}$ *such that for every subset* $J \subseteq A(\overline{x}) \setminus J_{\text{NLP}}^-$*, the rank of* $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla g_j(x)\}_{j \in J_{\text{NLP}}^-} \cup \{\nabla g_j(x)\}_{j \in J}$ *remains constant for all* $x \in V$*.*

By definition it is possible to notice that Strong-CRSC is weaker than RCRCQ and we will shot that it implies NLP-Ref-McCormick later. In addition, notice that this condition takes into account the sets $A(\overline{x})$ and J_{NLP}^-, but not consider the analyses of $J_0(\overline{x})$ defined in (4.32). It motivates the following definition.

Definition 4.5.4. *Let* $\overline{x}$ *be a feasible point of (NLP). We say that the* constraint qualification 1 *(CQ1) holds at* $\overline{x}$*, if there exists a neighborhood* V *of* $\overline{x}$ *such that for every subset* $J \subseteq J_0(\overline{x})$*, the rank of* $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla g_j(x)\}_{j \in J}$ *remains constant for all* $x \in V$*.*

The difference between the definition above and RCRCQ is that we consider all the subsets of $J_0(\overline{x})$ instead of $A(\overline{x})$. We will show that CQ1 is indeed a constraint qualification later. Despite this, the definition above does not take into account the good properties of the set J_{NLP}^-. In order to summarize the discussion about CRSC and what we need to require additionally to CRCS in order to obtain a second-order constraint qualification, consider the following definition.

Definition 4.5.5. *Let* $\overline{x}$ *be a feasible point of (NLP). We say that the* Constraint Qualification 2 *(CQ2) holds at* $\overline{x}$ *if there is a neighborhood* V *of* $\overline{x}$ *such that for every subset* $J \subseteq J_0(\overline{x}) \setminus J_{\text{NLP}}^-$*, the rank of* $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla g_j(x)\}_{j \in J_{\text{NLP}}^-} \cup \{\nabla g_j(x)\}_{j \in J}$ *remains constant for all* $x \in V$*.*

This condition is weaker than Strong-CRSC and CQ1 (and, consequently, weaker than RCRCQ) and it is independent of MFCQ. Now, let us prove that the conditions proposed imply NLP-Ref-McCormick.

Theorem 4.5.3. *Let $\overline{x}$ be a feasible point of (NLP). If CQ2 holds at $\overline{x}$, then NLP-Ref-McCormick also hods at $\overline{x}$.*

Proof. Take a direction $d \in \mathcal{L}_\Omega(\overline{x})$ such that $d \neq 0$. Define $J = J_d(\overline{x}) \setminus J^-_{\text{NLP}}$ where $J_d(\overline{x}) \subset J_0(\overline{x})$ was defined in (4.31). Consider the function

$$F(x) := \begin{cases} h_i(x), & i = 1, \dots, p, \\ g_j(x), & \text{if } j \in J^-_{\text{NLP}}, \\ g_k(x), & \text{if } k \in J. \end{cases} \tag{4.37}$$

By construction, we have that $d \in \ker(DF(\overline{x}))$ and, by hypothesis, we have that $DF(x)$ has constant rank around $\overline{x}$. Thus, by Theorem 4.3.1 there exists a diffeomorphism $\Phi : V_1 \to V_2$ of class C^2 and an $\varepsilon > 0$ such that $\overline{x} + td \in V_2$ and, moreover,

$$F(\Phi^{-1}(\overline{x} + td)) = F(\Phi^{-1}(\overline{x})) = F(\overline{x}) = 0.$$

It means that

$$F(\Phi^{-1}(\overline{x} + td)) = \begin{cases} h_i(\Phi^{-1}(\overline{x} + td)) = 0, & i = 1, \dots, p, \\ g_j(\Phi^{-1}(\overline{x} + td)) = 0, & \text{if } j \in J^-_{\text{NLP}}, \\ g_k(\Phi^{-1}(\overline{x} + td)) = 0, & \text{if } k \in J. \end{cases}$$

For the remaining constraints, that is, $j \notin J^-_{\text{NLP}} \cup J_d(\overline{x})$, we have that $\nabla g_j(\overline{x})^T d > 0$. To finish this proof, define $\xi(t) := \Phi^{-1}(\overline{x} + td)$ and then we get that NLP-Ref-McCormick holds at $\overline{x}$. ☐

To finish this section, next we present a figure that shows the relationship among the CQ's mentioned before.

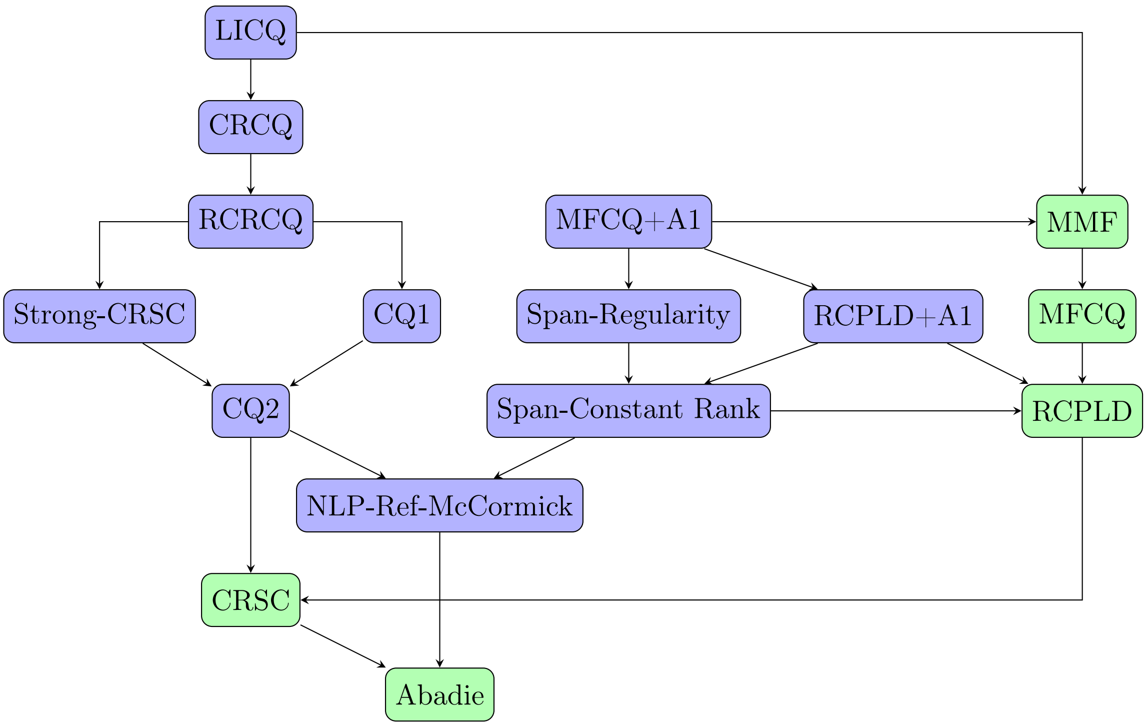

Figure 4.4: *Relationship among the CQs mentioned for (NLP). CQ's in blue box are related with SSOC; CQ's in green box do not have such property.*

Chapter 5

Conclusion

The study of constraint qualifications is well-developed in nonlinear programming (NLP). In particular, we highlight the constant rank-type constraint qualifications such as CRCQ [Jan84], CPLD [QW00], RCRCQ [MS11a], RCPLD [AHSS12a] and CRSC [AHSS12b]. However, in the nonlinear second-order cone programming (NSOCP) context the situation was different. Indeed, the most well-known constraint qualifications for (NSOCP) are the nondegeneracy condition and Robinson's CQ, that are the generalization of LICQ and MFCQ for (NSOCP), respectively.

This difficulty in defining constant rank-type constraint qualifications in (NSOCP) may be due the following facts: first, the conical structure of the second-order cone. Initially it does not seems to be an issue, once we also have a cone defining the constraints in nonlinear programming, namely, the non-negative orthant. But, the second-order cone structure is indeed harder to deal when we compare to the non-negative orthant (or, maybe, the non-negative orthant has a well behavior that makes our lives easier), and this difference appeared when we revisited classical constraint qualifications for constrained optimization, such as Guignard's CQ [Gui69] and Abadie's CQ [Aba65]. The second point that can be seen as a difficult in order to define constant rank-type constraint qualifications for (NSOCP), may be the fact that only linearity is not enough to guarantee the existence of Lagrange multiplier, as Andersen showed in [ART02], which is not the case when we analyze only linear constraints in the non-negative orthant. Actually, this field of study is so important with good properties and algorithms that there is a line research to study only linear programming problems (LP). Last, is that in (NLP) we have a powerful tool that is the so-called Sequential Optimality Conditions. On the one hand, in the nonlinear programming context the sequential optimality conditions was introduced in [AHM11] and it has several applications, for both convergence of classes of algorithms and study of new constraint qualifications. See [BHR18] and references therein. On the other hand, in the second-order cone programming context, the sequential optimality conditions were developed in [AFH+19] that analyzed the structure given by the eigenvectors of the second-order cone, which almost coincides with the first proposal of constant rank-type constraint qualifications for (NSOCP) made in [ZZ19]. In addition, when we analyzed the conditions proposed in [ZZ19], we noticed that it was incorrect. All of these facts leaded us in the following timeline: the counter-example to [ZZ19] presented in [AFH+21]; the first approaches that we presented in [AHM+22a] using the sequential optimality conditions; the consolidation of our constant rank-type conditions presented in [AHM+22b] using the eigenstructure of the second-order cone; and, finally, the cherry on the cake, the approach using a constant rank theorem that was presented in [AHM+23], that followed a similar vein of the original proposal of Janin for (NLP) in [Jan84] and has a similar second-order properties that was presented by Andreani et al. in [AES10].

In order to overcome the difficulties listed above, the most natural idea was to bring the constraints of a second-order cone programming problem to our knowledge in constraint qualifications in (NLP). The is the reason why our first approach [AHM+22a] was called "naive", because we could not deal with all of the pure conic constraints in a proper way and then we dealt with such constraints using the requirements of Robinson's CQ. By "pure conic constraints", we mean constraints that we could not apply a reduction mapping in order to get an inequality constraint. For the remaining constraints, that is, the ones that we could apply a reduction mapping, we brought them to a nonlinear programming context through a reduction mapping and then we applied the expertise of constraint qualifications for (NLP) that was already developed, such as RCPLD and CRSC, for example. However, in some cases, the naive CQ's coincided with Robinson's CQ justly our approach was to use it when we could not reduce to a (NLP) context. In this approach we could prove that Naive-RCPLD and Naive-CRSC are indeed constraint qualifications for (NSOCP), weaker than Robinson's CQ and we also provided that these conditions are enough to prove global convergence to a KKT point for algorithms that generates AKKT sequences.

In [AHM+22b] the approach was more solid, in the sense that we did not avoid the pure conic structure of the second-order cone. On the contrary, we explored the structure of the eigenvectors of the second-order cone developed in [AG03] following similar ideas to the ones that were developed in [AHMR23b] for nonlinear semidefinite programming (NSDP). With this tool at hand, we noticed that we could establish weaker versions of nondegeneracy and Robinson's CQ just avoiding vectors in the second-order cone that somehow are not related to the eigenvectors of the constraints. With these new CQ's that we called by weak-nondegeneracy and weak-Robinson, we could then establish proposals of CRCQ and CPLD, that were called weak-CRCQ and weak-CPLD, respectively, that coincides with their counterpart in (NLP) and, in addition, are enough to show global convergence of algorithms related to an external penalty method, using the results provided in [AFH+19]. In this thesis, we presented an approach a little bit different of our paper but keeping the same essence of the results and, in addition, we showed a way to build examples to establish relation among the CQ's proposed and the classical ones, that is, the nondegeneracy condition and Robinson's CQ.

Later, in [AHM+23] we encompassed the pure conic structure of the problem and the attached the constant rank historical approach, that is, to apply the constant rank theorem initially used by Janin in [Jan84] to show that CRCQ implies Abadie's CQ, as a natural generalization of an Implicit Function Theorem, like as [IT74]. But, before we apply this approach, we needed before to build a relation between the nondegeneracy condition and Abadie's CQ in (NSOCP). In this middle time, we faced some difficulties once the Abadie's CQ that we were considering was incorrect, namely, only the equality between the tangent cone and linearized cone (this condition was also presented in [ZZ19]), which was probably caused by a bias of (NLP) researches and, in addition, due to the fact that we do not have access to the physical publication of Abadie [Aba65]. Thus, revisiting Guignard's CQ [Gui69] we could rebuild Abadie's CQ for (NSOCP). In order to understand this issue, just in [BKMW20] we found the correct definition of Abadie's CQ, almost at the same time that we were developing our research. Finally, with all the tools and experience at hand, we could introduce our definition of CRCQ for (NSOCP) using a constant rank theorem in a similar way that Janin did and, moreover, with second-order optimality conditions as was made in [AES10].

Last, in this thesis we added a constraint qualification based on curves for (NSOCP), that is weaker than CRCQ and stronger than Abadie's CQ. This condition was inspired by a recent research [FSS22] for (NLP). In our proposal, we showed not only that Ref-McCormick was a constraint qualification, but we could also show that this condition also implies SSOC for (NSOCP) for any Lagrange multiplier and, in addition, we showed that this result is related to the fact the the Hessian of the Lagragian does not depend on the Lagrange multiplier when we vanish all the directions in the critical cone. To the best of our knowledge, this is new even for (NLP) and it is the weakest condition that ensures this result. Inspired by Ref-McCormick for (NLP) and this second-order results, in the last section of this thesis we proposed new constraint qualifications in (NLP) based on constant rank property, that implies NLP-Ref-McCormick and then inherit all the properties mentioned before. Moreover, in the Appendix E of this thesis, contains a preprint showing the difficulty of obtaining points that satisfies SSOC in (NLP) by a practical algorithm when do not have hypothesis based on constant rank.

5.1 Suggestion for future research

In Figure 4.3 we showed our contribution for the developing of the study of constant rank-type constraint qualifications for (NSOCP). Since a new "bridge" was built using techniques such as sequential optimality conditions and also the constant rank theorem, the most natural step are

i) to find the natural generalizations of the CQ's that are already well established in (NLP) for the (NSOCP) context;

ii) to get results related to the CQ's proposed by us that are not proved yet, such as the study of stability analysis;

iii) the computing of the derivative of the value function using CRCQ;

iv) convergence of algorithms under CRCQ.

In addition, another possibility is to study in a deeper way the constant rank theorem in [Mal72] that was used by Janin. It is important to notice that the constant rank theorem was proved as being an equivalence, as it has been being used just as an implication, in the sense of: *if a family of functions has constant rank, then some properties hold*. However, the original theorem has the structure of: *a family of functions has constant rank if, and only if, some properties hold*. It means that there exists a family of functions that captures exactly the requirement in order to have an implication to Abadie's CQ but keeping second-order properties. For example, the difference between CRCQ and RCRCQ relies on what set of constraints we require constant rank. While in the CRCQ proposed by Janin we vanish all the subsets of the equality constraints, in RCRCQ we just consider the subsets that contains all the equality constraints. In CRSC, it is required just one set to have constant rank, but since it is weaker than MFCQ then we lost SSOC. Recently, in [AHMR23a] the authors proposed a CQ between RCRCQ and CRSC that has such properties and it can be a candidate of being the weakest constant rank-type CQ that ensures SSOC.

Bibliography

[Aba65] J Abadie. *Problèmes d'optimisation*. Institut Blaise Pascal, 1965. 1, 31, 55, 56

[ABHS17] R. Andreani, R. Behling, G. Haeser e P. J. S. Silva. On second-order optimality conditions in nonlinear optimization. *Optimization Methods and Software*, 32(1):22–38, 2017. 50

[AES10] R. Andreani, C. E. Echagüe e M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization theory and Applications*, 146(2):255–266, 2010. 1, 2, 10, 12, 29, 35, 38, 42, 44, 48, 55, 56

[AFH+19] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos e L. D. Secchin. Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. *Optimization online*, 2019. 2, 13, 55, 56

[AFH+21] R. Andreani, E. H. Fukuda, G. Haeser, H. Ramírez, D. O. Santos, P. J. S. Silva e T. P. Silveira. Erratum to: New constraint qualifications and optimality conditions for second order cone programs. *Set-Valued and Variational Analysis*, páginas 1–5, 2021. 2, 3, 5, 7, 11, 12, 17, 29, 30, 55

[AG03] F. Alizadeh e D. Goldfarb. Second-order cone programming. *Mathematical Programming Series B*, 95:3–51, 2003. 1, 6, 17, 56

[AHM11] R. Andreani, G. Haeser e J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011. 12, 55

[AHM+22a] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez, D. O. Santos e T. P. Silveira. Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. *Optimization Letters*, páginas 1–22, 2022. 2, 3, 5, 13, 14, 15, 29, 55

[AHM+22b] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez e T. P. Silveira. Global convergence of algorithms under constant rank conditions for nonlinear second-order cone programming. *Journal of Optimization Theory and Applications*, 195(1):42–78, 2022. 2, 3, 17, 19, 22, 23, 25, 27, 28, 29, 55, 56

[AHM+23] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez e T. P. Silveira. First-and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. *Mathematical Programming*, páginas 1–41, 2023. 2, 3, 29, 31, 39, 55, 56

[AHMR23a] R. Andreani, G. Haeser, L. M. Mito e H. Ramírez. A minimal face constant rank constraint qualification for reducible conic programming. *arXiv preprint arXiv:2304.13881*, 2023. 53, 57

[AHMR23b] R. Andreani, G. Haeser, L. M. Mito e H. Ramírez. Weak notions of nondegeneracy in nonlinear semidefinite programming. *Mathematical Programming*, páginas 1–32, 2023. 22, 56

[AHSS12a] R. Andreani, G. Haeser, M. L. Schuverdt e P. J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1-2):255–273, 2012. 1, 12, 52, 55

[AHSS12b] R. Andreani, G. Haeser, M. L. Schuverdt e P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012. 1, 10, 52, 53, 55

[AMS05] R. Andreani, J. M. Martínez e M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of optimization theory and applications*, 125(2):473–483, 2005. 50

[AMS07] R. Andreani, J. M. Martínez e M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56(5-6):529–542, 2007. 10, 11, 50

[Ani00] M. Anitescu. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, 10(4):1116–1135, 2000. 10

[ART02] E. D. Andersen, C. Roos e T. Terlaky. Notes on duality in second order and p-order cone optimization. 2002. 2, 29, 30, 31, 55

[Aru98] A. Arutyunov. Second-order conditions in extremal problems. the abnormal points. *Transactions of the American Mathematical Society*, 350(11):4341–4365, 1998. 10

[Bea11] B. Beauzamy. *Introduction to Banach spaces and their geometry*. Elsevier, 2011. 30

[BHR18] E. Birgin, G. Haeser e A. Ramos. Augmented lagrangians with constrained subproblems and convergence to second-order stationary points. *Computational Optimization and Applications*, 69(1):51–75, 2018. 1, 12, 55

[BHRV18] R. Behling, G. Haeser, A. Ramos e D. S. Viana. On a conjecture in second-order optimality conditions. *Journal of Optimization Theory and Applications*, 176:625–633, 2018. 48

[BKMW20] E. Börgens, C. Kanzow, P. Mehlitz e G. Wachsmuth. New constraint qualifications for optimization problems in banach spaces based on asymptotic kkt conditions. *SIAM Journal on Optimization*, 30(4):2956–2982, 2020. 2, 31, 56

[BR05] J. F. Bonnans e H. Ramírez. Perturbation analysis of second-order cone programming problems. *Mathematical Programming, Series B*, 104:205–227, 2005. 5, 6, 7, 19, 44

[BS00] J. F. Bonnans e A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media, 2000. 7

[FSS22] N. S. Fazzion, M.D. Sánchez e M. L. Schuverdt. A note on the mccormick second-order constraint qualification. *Trends in Computational and Applied Mathematics*, 23:769–781, 2022. 3, 42, 48, 56

[GM15] H. Gfrerer e B. S. Mordukhovich. Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions. *SIAM Journal on Optimization*, 25(4):2081–2119, 2015. 1

[GO16] H. Gfrerer e J. V. Outrata. On computation of generalized derivatives of the normal-cone mapping and their applications. *Mathematics of Operations Research*, 41(4):1535–1556, 2016. 1

[GT71] F. J. Gould e J. W. Tolle. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20(2):164–172, 1971. 30

[Gui69] M. Guignard. Generalized kuhn–tucker conditions for mathematical programming problems in a banach space. *SIAM Journal on Control*, 7(2):232–241, 1969. 1, 2, 30, 31, 55, 56

[Hae17] G. Haeser. An extension of yuanâs lemma and its applications in optimization. *Journal of Optimization Theory and Applications*, 174(3):641–649, 2017. 50

[IT74] A. D. Ioffe e V. M. Tihomirov. *Theory of extremal problems*. North-Holland, Amsterdam, 1974. 12, 32, 56

[Jan84] R. Janin. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, 21:110–126, 1984. 1, 2, 9, 29, 32, 35, 38, 41, 55, 56

[Lim83] E. L. Lima. *Espaços métricos*, volume 4. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1983. 30

[Mal72] M. P. Malliavin. Géométrie différentielle intrinseque. *Herman, Paris*, 1972. 2, 32, 38, 57

[Mas19] W. F. Mascarenhas. A simple canonical form for nonlinear programming problems and its use. *Journal of Optimization Theory and Applications*, 181(2):456–469, 2019. 50

[McC67] G. P. McCormick. Second order conditions for constrained minima. *SIAM Journal on Applied Mathematics*, 15:641–652, 1967. 42

[MF67] O. L. Mangasarian e S. Fromovitz. The fritz-john necessary conditions in presence of equality and inequality constraints. *J. Math. Anal. Appl*, 17:37–47, 1967. 1, 8

[MS11a] L. Minchenko e S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011. 1, 10, 12, 55

[MS11b] L. Minchenko e S. Stakhovski. Parametric nonlinear programming problems under the relaxed constant rank condition. *SIAM Journal on Optimization*, 21(1):314–332, 2011. 10, 44

[NW99] J. Nocedal e S. J. Wright. *Numerical optimization*. Springer, 1999. 1, 7, 8, 44

[Pat07] G. Pataki. On the closedness of the linear image of a closed convex cone. *Mathematics of Operations Research*, 32(2):395–412, 2007. 31, 40

[QW00] L. Qi e Z. Wei. On the constant positive linear dependence condition and its application to sqp methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000. 1, 55

[Rob76] S. M. Robinson. Stability theory for systems of inequalities, part ii: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13(4):497–513, 1976. 1, 7

[Rob82] S. M. Robinson. *Generalized equations and their solutions, Part II: Applications to nonlinear programming*. Springer, 1982. 7

[RS23] A. A. Ribeiro e M. Sachine. On strong second-order necessary optimality conditions under relaxed constant rank constraint qualification. *Optimization*, páginas 1–13, 2023. 44

[Zor82] V. A. Zorich. *Mathematical Analysis*. Moscow, Nauka, 1982. 12

[ZZ19] Y. Zhang e L. Zhang. New constraint qualifications and optimality conditions for second order cone programs. *Set-Valued and Variational Analysis*, 27:693–712, 2019. 2, 7, 11, 12, 17, 29, 30, 55, 56

Appendix A

Article: Erratum to: New Constraint Qualifications and Optimality Conditions for Second Order Cone Programs. doi.org/10.1007/s11228-021-00573-5.

Set-Valued and Variational Analysis (2022) 30:329–333
https://doi.org/10.1007/s11228-021-00573-5

Erratum to: New Constraint Qualifications and Optimality Conditions for Second Order Cone Programs

R. Andreani[1] · E. H. Fukuda[2] · G. Haeser[3] · H. Ramírez[4] · D. O. Santos[5] · P. J. S. Silva[1] · T. P. Silveira[3]

Received: 8 September 2020 / Accepted: 13 January 2021 / Published online: 7 April 2021

Abstract

In this note we show with a counter-example that all conditions proposed in Zhang and Zhang (Set-Valued Var. Anal **27**:693–712 2019) are not constraint qualifications for second-order cone programming.

Keywords Constraint qualifications · Optimality conditions · Second-order cone programming

This comment refers to the article available online at https://doi.org/10.1007/s11228-018-0487-2.

✉ G. Haeser
ghaeser@ime.usp.br

R. Andreani
andreani@ime.unicamp.br

E. H. Fukuda
ellen@i.kyoto-u.ac.jp

H. Ramírez
hramirez@dim.uchile.cl

D. O. Santos
daiana@ime.usp.br

P. J. S. Silva
pjssilva@ime.unicamp.br

T. P. Silveira
thiagops@ime.usp.br

1 Department of Applied Mathematics, University of Campinas, Campinas-SP, Brazil

2 Graduate School of Informatics, Kyoto University, Kyoto, Japan

3 Department of Applied Mathematics, University of São Paulo, São Paulo-SP, Brazil

4 Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Santiago, Chile

5 Institute of Science and Technology, Federal University of São Paulo, São José dos Campos-SP, Brazil

Mathematics Subject Classification (2010) 90C22 · 90C46 · 90C30

We consider the (nonlinear) second-order cone programming problem

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & g_j(x) \in K_{m_j}, \; j = 1, \ldots, \ell, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_j : \mathbb{R}^n \to \mathbb{R}^{m_j}$, $j = 1, \ldots, \ell$ are continuously differentiable and the second-order cone K_m is defined as $K_m := \{z := (z_0, \overline{z}) \in \mathbb{R} \times \mathbb{R}^{m-1} \mid z_0 \geq \|\overline{z}\|\}$ if $m > 1$ and $K_1 := \{z \in \mathbb{R} \mid z \geq 0\}$. Here $\|\cdot\|$ is the Euclidean norm.

Given a feasible point x^*, we denote by $I_0(x^*) := \{j \in \{1, \ldots, \ell\} \mid g_j(x^*) = 0\}$ the index set of constraints at the vertex of the corresponding second-order cone and by $I_B(x^*) := \{j \in \{1, \ldots, \ell\} \mid [g_j(x^*)]_0 = \|\overline{g_j(x^*)}\| > 0\}$ the index set of constraints at the non-zero boundary of the corresponding second-order cone. For $j \in I_B(x^*)$ we define $\phi_j(x) := \frac{1}{2}([g_j(x)]_0^2 - \|\overline{g_j(x)}\|^2)$, with $\nabla \phi_j(x) = J_{g_j}(x)^T R_{m_j} g_j(x)$, where $J_{g_j}(x)^T$ is the $n \times m_j$ transposed Jacobian of g_j and R_m is the $m \times m$ diagonal matriz with 1 at the first position and -1 at the remaining positions.

In [11], the authors present an extension of the classical constant rank constraint qualification (CRCQ, [9]) for the second-order cone programming problem (1). It reads as follows:

Definition 1 The Constant Rank Constraint Qualification (CRCQ) as defined in [11] holds at a feasible point x^* of (1) if there exists a neighborhood V of x^* such that for any index sets $J_1 \subseteq I_0(x^*)$ and $J_2 \subseteq I_B(x^*)$, the family of matrices whose rows are the union of $J_{g_j}(x)$, $j \in J_1$ and the vector rows $\nabla \phi_j(x)^T$, $j \in J_2$ has the same rank for all $x \in V$.

When $j \in I_B(x^*)$, the conic constraint $g_j(x) \in K_{m_j}$ can be locally replaced by the nonlinear constraint $\phi_j(x) \geq 0$, which is active at x^* (see e.g. [7, Section 4] for more details). Note also that for $j \in I_0(x^*)$ such that K_{m_j} is one-dimensional, the constraint $g_j(x) \in K_{m_j}$ is also a standard nonlinear constraint. Hence, the particularity of a second-order cone lies on the fact that one may have a "multi-dimensionally active" constraint $g_j(x^*) = 0$, which must be treated accordingly since these are tipically the constraints that are hard to tackle. The first impression one has when reading Definition 1 is that there is no special treatment for these active constraints. In particular, one would expect some regularity to be assumed for each constraint $g_j(x) \in K_{m_j}$ when $j \in I_0(x^*)$. To emphasize this last point, let us consider problem (1) with a single second-order cone, that is, $\ell = 1$, with constraint $g(x) \in K_{m_1}$. Let x^* be a feasible point such that $g(x^*) = 0$. According to Definition (1), CRCQ holds at x^* when the set of vectors given by all rows of $J_g(x)$ has constant rank, i.e., the full set of gradients $\{\nabla g_0(x), \ldots, \nabla g_{m_1 - 1}(x)\}$ has constant rank, and no subset of these vectors is considered. However, it is well known that the classical CRCQ for nonlinear programming requires that all subsets of active constraints possesses the constant rank property.

Despite these considerations, the example given below shows that even a strengthen definition of CRCQ, that takes all these subsets into account, is not a constraint qualification. This thus invalidates all the results proved in [11]. Therein, the authors also propose a definition for the relaxed-CRCQ (RCRCQ, [10]) and for the Constant Rank of the Subspace Component (CRSC, [6]), which, being weaker than their definition of CRCQ, are not constraint qualifications either. In particular, the definition of RCRCQ is done in such a way that only the full set of *all* gradients in $I_0(x^*)$ is considered, while every subset $J_2 \subseteq I_B(x^*)$ is considered (namely, J_1 is taken to be fixed and equal to $I_0(x^*)$ in Definition 1). However, it is easy to see that this is not a constraint qualification, since when one considers only one-dimensional cones, and consequently (1) reduces to a nonlinear programming problem, RCRCQ reads identical to the so-called Weak Constant Rank property from [1], which is not a constraint qualification. Our counter-example is discussed in the sequel.

Consider the following problem of one-dimensional variable:

$$\begin{aligned} \text{Minimize} \quad & f(x) := -x, \\ \text{s.t.} \quad & g(x) \in K_2, \end{aligned} \tag{2}$$

with

$$g(x) = \begin{pmatrix} g_0(x) \\ g_1(x) \end{pmatrix} := \begin{pmatrix} x \\ x + x^2 \end{pmatrix}.$$

The unique feasible point is $x^* = 0$, thus, it is a global solution. Since $g(x^*) = 0$, the Karush-Kuhn-Tucker conditions for this problem are given by the existence of $\mu \in K_2$ such that $\nabla f(x^*) - J_g(x^*)^T \mu = 0$, that is

$$-1 - \mu_0 - \mu_1 = 0, \tag{3}$$

with $\mu = (\mu_0, \mu_1)^T \in K_2$, or, equivalently, $\mu_0 \geq |\mu_1|$. Thus, (3) can not hold and the Karush-Kuhn-Tucker conditions fail. On the other hand $J_g(x) = \begin{pmatrix} 1 \\ 1 + 2x \end{pmatrix}$ for all x. In particular, $\nabla g_0(x) = 1$ and $\nabla g_1(x) = 1 + 2x$ for all x. Thus, all subsets of gradients

$$\{\nabla g_0(x)\}, \{\nabla g_1(x)\}, \{\nabla g_0(x), \nabla g_1(x)\}$$

have constant rank equal to 1 for all x near x^*. This shows that the definition of CRCQ from [11] is not a constraint qualification, as this property is characterized by the fact that the Karush-Kuhn-Tucker conditions hold at any local minimizer.

We next briefly point out the possible mistake in the approach followed in [11]. It is based on the proof of RCRCQ from [10], which is also similar to [1]. It is shown therein that $\mathscr{L}(x^*) \subseteq \mathscr{T}(x^*)$, for apropriate definitions of the linearized cone $\mathscr{L}(x^*)$ and tangent cone $\mathscr{T}(x^*)$ for second-order cone programming, by means of applying an implicit function-type theorem (Lyusternik's theorem [8]). This theorem allows constructing a suitable tangent curve and can be applied provided the constant rank assumption holds true. However, in the nonlinear programming context, when constraint $g_j(x^*) = 0$ is analyzed, direction $d \in \mathscr{L}(x^*)$ must be orthogonal to the gradient $\nabla g_j(x^*)$ in order to ensure the existence of a tangent curve to $\{x \mid g_j(x) = 0\}$ along the direction d. This seems to be ignored in [11].

Instead of applying the implicit function approach, constant rank constraint qualifications may be defined using the approach of sequential optimality conditions [2]. See, for instance, [4–6]. For this, one would need a proper extension of the so-called Carathéodory Lemma (see, e.g., [5]), which permits rewriting a linear combination $y := \sum_{i=1}^m \lambda_i v_i$ with $\lambda_i \in \mathbb{R}$ and $v_i \in \mathbb{R}^n$ for all i in the following way: $y = \sum_{i \in I} \tilde{\lambda}_i v_i$ with $I \subseteq \{1, \ldots, m\}$, $\{v_i\}_{i \in I}$ linearly independent, and $\tilde{\lambda}_i$ with the same sign of λ_i for each i. In the case of second-order cones, for which the vector of scalars $(\alpha_i)_{i=1}^m$ belongs to the second-order cone K_m, one would want to rewrite the same vector y by only using a linearly independent subset of $\{v_i\}_{i=1}^m$ and such that the new scalars still belong to the cone. However, this is not possible in general as the following examples show.

Example 1 Take $y := \beta_0 v_0 + \beta_1 v_1 + \beta_2 v_2$, with $(\beta_0, \beta_1, \beta_2) := (\sqrt{2}, 1, 1) \in K_3$, $v_0 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $v_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $v_2 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. There is no way of rewriting y using new scalars $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) \in K_3$ such that $\hat{\beta}_i = 0$ for some $i = 0, 1, 2$.

In the case of more than one block of constraints ($\ell > 1$), even assuming more regularity for each block, a conic variant of Carathéodory's Lemma seems not possible to obtain.

Example 2 Take $y := \beta_0 v_0 + \beta_1 v_1 + \gamma_0 w_0 + \gamma_1 w_1$ with $(\beta_0, \beta_1) := (1, 1) \in K_2$, $(\gamma_0, \gamma_1) := (1, 1) \in K_2$, and vectors

$$v_0 := \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, v_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, w_0 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{and } w_1 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

It is not possible to rewrite y with new scalars $(\hat{\beta}_0, \hat{\beta}_1) \in K_2$, $(\hat{\gamma}_0, \hat{\gamma}_1) \in K_2$ in such a way that at least one component vanishes. Note that both $\{v_0, v_1\}$ and $\{w_0, w_1\}$ are linearly independent sets, but the necessity of dealing with the product of two second-order cones makes it impossible to fulfill the desired property.

We end this erratum with the following observation. Since it is well-known that linear second-order cone programs may possess duality gap, a definition of CRCQ could not be automatically satisfied by linear problems at the vertex. In [3], a naive proposition of CRCQ is presented where the "multi-dimensionally" active constraints are treated similarly to Robinson's CQ while the remaining constraints are treated similarly to CRCQ for nonlinear programming.

Acknowledgements This work was supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2018/24293-0, 2017/18308-2, and 2017/12187-9), CNPq (grants 301888/2017-5, 303427/2018-3, and 404656/2018-8), Grant-in-Aid for scientific research (C)(19K11840) from Japan Society for the Promotion of Science, and FONDECYT (grant 1201982) and Basal Program CMM-AFB 170001, both from ANID (Chile).

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. Andreani, R., Echagüe, C.E., Schuverdt, M.L.: Constant-rank condition and second-order constraint qualification. J. Optim. Theory Appl. **146**(2), 255–266 (2010)
2. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011)
3. Andreani, R., Haeser, G., Mito, L.M., Ramírez, H., Santos, D.O., Silveira, T.P.: Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. Optimization online (2020)
4. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A second-order sequential optimality condition associated to the convergence of algorithms. IMA J. Numer. Anal. **37**(4), 1902–1929 (2017)
5. Andreani, R., Haeser, G., Schuverdt, M., Silva, P.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**(1-2), 255–273 (2012)
6. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**(3), 1109–1135 (2012)

7. Bonnans, J.F., Ramírez, H.: Perturbation analysis of second-order cone programming problems. Math. Program. **104**(2), 205–227 (2005)
8. Ioffe, A.D., Tihomirov, V.M.: Theory of Extremal Problems. North-Holland, Amsterdam (1974)
9. Janin, R.: Direction derivate of the marginal function in nonlinear programming. Math. Program. Study **21**, 110–126 (1984)
10. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)
11. Zhang, Y., Zhang, L.: New constraint qualifications and optimality conditions for second order cone programs. Set-Valued Var. Anal. **27**, 693–712 (2019)

Appendix B

Article: Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. doi.org/10.1007/s11590-021-01737-w

Optimization Letters (2022) 16:589–610
https://doi.org/10.1007/s11590-021-01737-w

ORIGINAL PAPER

Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming

R. Andreani[1] · G. Haeser[2] · L. M. Mito[2] · H. Ramírez[3] · D. O. Santos[4] · T. P. Silveira[2]

Received: 12 September 2020 / Accepted: 7 April 2021 / Published online: 13 April 2021

Abstract
The constant rank constraint qualification, introduced by Janin in 1984 for nonlinear programming, has been extensively used for sensitivity analysis, global convergence of first- and second-order algorithms, and for computing the directional derivative of the value function. In this paper we discuss naive extensions of constant rank-type constraint qualifications to second-order cone programming and semidefinite programming, which are based on the Approximate-Karush–Kuhn–Tucker necessary optimality condition and on the application of the reduction approach. Our definitions are strictly weaker than Robinson's constraint qualification, and an application to the global convergence of an augmented Lagrangian algorithm is obtained.

Keywords Constraint qualifications · Optimality conditions · Second-order cone programming · Semidefinite programming · Global convergence

1 Introduction

In this paper we investigate constraint qualifications (CQs) for second-order cone programming and semidefinite programming. In particular, we are interested in constant rank CQs as defined first in [15] and later extended in [7,8,19,21] in the context of nonlinear programming. In particular, the definition in [15] gained some notoriety for its ability to compute the directional derivative of the value function, a result known to hold at the time only under Mangasarian-Fromovitz CQ [24]. Also, the definition

We would like to thank Ellen H. Fukuda (Kyoto University) and Paulo J.S. Silva (University of Campinas) for initial discussions on this topic. This work was supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2018/24293-0, 2017/18308-2, 2017/17840-2, and 2017/12187-9), CNPq (grants 301888/2017-5, 303427/2018-3, and 404656/2018-8), PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016), and FONDECYT grant 1201982 and Basal Program CMM-AFB 170001, both from ANID (Chile).

Extended author information available on the last page of the article

from [15] includes naturally the case of linear constraints, which does not follow under Mangasarian-Fromovitz CQ. The ability to handle redundant constraints (in particular, linear ones) in the case of nonlinear programming is a powerful modeling tool that frees the model builder from the apprehension of including them without preprocessing. Actually, the effort of finding which constraints are redundant may be equivalent to the effort of solving the problem.

For conic programming, it is well known that linearity of the constraints is not a CQ [2,22] and this somehow stresses the difficulties in extending these ideas to the conic context. In particular, a previous tentative extension to second-order cones [28] has been shown to be incorrect [3].

In this paper, we make use of the reduction approach in order to propose new constant rank-type CQs for second-order cone programming and semidefinite programming that are strictly weaker than Robinson's CQ. In our approach, we separate the constraints into two sets: one consisting of the constraints that can be completely characterized by standard equality and inequality nonlinear programming constraints, and other with the irreducible conic constraints. For second-order cone programming, the second block consists of constraints that are active at the vertex of a multi-dimensional second-order cone, while for semidefinite programming these correspond to semidefinite blocks where the zero eigenvalue is non-simple.

We consider our conditions to be naive extensions of the corresponding nonlinear programming CQ in the sense that if the problem only has irreducible constraints then all our conditions coincide with Robinson's CQ; however we show some interesting examples where our condition holds while Robinson's CQ fails. Extending these ideas to consider also the irreducible constraints is an ongoing topic of research.

Despite our inability of dealing with the irreducible conic constraints, the Approximate-Karush–Kuhn–Tucker (AKKT) [5] necessary optimality condition, recently extended to second-order cones [4] and semidefinite programming [9], can easily be used to handle the remaining constraints by means of the reduction approach. This allows obtaining CQs analogous to those defined in [7,8,15,19,21]. Analogous definitions of [15,19] are independent of Robinson's CQ, while analogues of [7,8,21] are strictly weaker than Robinson's CQ.

Since several algorithms are expected to generate AKKT sequences (this is the case, for instance, of the augmented Lagrangian algorithms of [4,9]), a relevant corollary of our analysis is that all CQs introduced in this paper can be used for proving global convergence of these algorithms to a KKT point.

This paper is organized as follows. In Sect. 2, we briefly introduce constant rank CQs for nonlinear programming. In Sect. 3, we revisit constraint qualifications for second-order cone programming. Section 4 is devoted to the AKKT approach, while in Sect. 5 we introduce and explain our new CQs for second-order cones. In Sect. 6 we extend these ideas to semidefinite programming. Finally, our conclusions are presented in Sect. 7.

Notation: For a continuously differentiable function $g\colon \mathbb{R}^n \to \mathbb{R}^m$, we denote $J_g(x)$ the $m \times n$ Jacobian matrix of g at x, for which the j-th row is given by the transposed gradient $\nabla g_j(x)^T$ of the j-th component function $g_j\colon \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, m$. Any finite-dimensional space $\mathbb{R}^m$ is equipped with its standard Euclidean inner product $\langle x, y\rangle := x^T y = \sum_{j=1}^{m} x_j y_j$. Then, given a closed convex cone $K \subseteq$

$\mathbb{R}^m$, we denote its polar by $K^\circ := \{v \in \mathbb{R}^m \mid \langle v, y \rangle \leq 0, \forall y \in K\}$. Finally, we adopt the following standard conventions on the empty set $\emptyset$: the sum over an empty index set is null (i.e., $\sum_\emptyset = 0$) and $\emptyset$ is linearly independent (considered as the basis of the trivial linear space $\{0\}$).

2 Constant rank-type CQ conditions in nonlinear programming

Consider the following nonlinear programming problem (NLP):

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t. } h_i(x) = 0, \quad & i = 1, \ldots, p, \\ g_j(x) \leq 0, \quad & j = 1, \ldots, q, \end{aligned} \tag{1}$$

where $f, h_i, g_j \colon \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions. We denote by $A(x^*) := \{j \in \{1, \ldots, q\} \mid g_j(x^*) = 0\}$, the set of indices of active inequality constraints at a feasible point x^*.

It is well known that at a local minimizer x^*, it holds that $-\nabla f(x^*) \in \mathscr{T}(x^*)^\circ$, where $\mathscr{T}(x^*)$ denotes the (Bouligand) tangent cone to the feasible set at x^* (see, e.g., [20, Theorem 12.8]). However, since the tangent cone is a geometric object, this necessary optimality condition is not always easy to manipulate. For this reason, one considers the linearized cone, which is defined as follows:

$$\mathscr{L}(x^*) := \left\{ d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0, i = 1, \ldots, p; \ \nabla g_j(x^*)^T d \leq 0, \ j \in A(x^*) \right\}.$$

Its polar may be computed via Farkas' Lemma, obtaining:

$$\mathscr{L}(x^*)^\circ = \left\{ v \in \mathbb{R}^n \;\middle|\; v = \sum_{i=1}^p \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*), \mu_j \geq 0, \ j \in A(x^*) \right\}.$$

Hence, when $\mathscr{T}(x^*)^\circ = \mathscr{L}(x^*)^\circ$, this geometric optimality condition takes the form of the usual, much more tractable, Karush–Kuhn–Tucker conditions. Vectors (λ_i, μ_j) above are called Lagrange multipliers associated with x^*, and the set of all these vectors is denoted by $\Lambda(x^*)$ in this manuscript.

A constraint qualification (CQ) is a condition that ensures the equality $\mathscr{T}(x^*)^\circ = \mathscr{L}(x^*)^\circ$. One of the most used CQ in the NLP literature is the well-known Linear Independence Constraint Qualification (LICQ), which states the linear independence of the set of gradients $\{\nabla h_i(x^*)\}_{i=1}^p \cup \{\nabla g_j(x^*)\}_{j \in A(x^*)}$. LICQ ensures not only the existence, but also the uniqueness of the Lagrange multiplier (see, e.g., [20, Section 12.3]). Several weaker CQs have been defined for NLP. In this paper, we are interested in constant rank-type ones as first introduced by Janin in [15]. Recall that in the NLP setting, we say that the Constant Rank Constraint Qualification (CRCQ) holds at a feasible point x^* if there exists a neighborhood V of x^*, such that for every subsets $I \subseteq \{1, \ldots, p\}$ and $J \subseteq A(x^*)$, the rank of $\{\nabla h_i(x), \nabla g_j(x); i \in I, j \in J\}$ remains constant for all $x \in V$. CRCQ is clearly weaker than LICQ.

Note that requiring only constant rank of the full set of gradients $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla g_j(x)\}_{j\in A(x^*)}$ (which is known as the Weak Constant Rank (WCR) property) is not a CQ, as shown in [10]. The necessity of considering every subset of this set of gradients may be seen from the definition of the linearized cone. Indeed, given $d \in \mathscr{L}(x^*)$, the relevant index set of inequality constraints gradients is given by $J = J_d := \{j \in A(x^*) \mid \nabla g_j(x^*)^T d = 0\}$, which cannot be chosen in advance if one only considers the point x^*. However, this suggests that there is no need to consider subsets of indices for the equality constraints, that is, it is enough to fix $I = \{1, \ldots, p\}$. This condition, called Relaxed-CRCQ (RCRCQ), has been shown to be a CQ in [18]. This condition reads as follows: RCRCQ holds at a feasible point x^* if there exists a neighborhood V of x^*, such that for every subset $J \subseteq A(x^*)$, the rank of $\{\nabla h_i(x), \nabla g_j(x); i \in \{1, \ldots, p\}, j \in J\}$ remains constant for all $x \in V$.

These conditions can be seen as *constant linear dependence* conditions and thus it is natural to weaken these definitions by considering only *constant positive linear dependence*, providing conditions CPLD [21] and its relaxed variant RCPLD [7], both strictly weaker than Mangasarian-Formovitz CQ. This will be the most natural formulation for the CQs we propose in this paper. We refer the reader to [7].

It turns out that the idea behind the construction of RCRCQ can be also extended to inequality constraints, providing an even weaker CQ. One seeks at characterizing a single index set J which is relevant of having the constant rank property. This set consists of the indices of gradients defining the subspace component of $\mathscr{L}(x^*)^\circ$, which is given by its lineality space. More precisely, the lineality space of $\mathscr{L}(x^*)^\circ$, defined as the largest linear space contained in $\mathscr{L}(x^*)^\circ$, is in this case given by $\mathscr{L}(x^*)^\circ \cap -\mathscr{L}(x^*)^\circ$. So, a gradient $\nabla g_j(x^*)$ belongs to $\mathscr{L}(x^*)^\circ \cap -\mathscr{L}(x^*)^\circ$ if, and only if, $-\nabla g_j(x^*) \in \mathscr{L}(x^*)^\circ$. Thus, for $J = J_-(x^*) := \{j \in A(x^*) \mid -\nabla g_j(x^*) \in \mathscr{L}(x^*)^\circ\}$, we say that the Constant Rank of the Subspace Component (CRSC) CQ holds at a feasible point x^* if there exists a neighborhood V of x^*, such that the rank of $\{\nabla h_i(x), \nabla g_j(x); i \in \{1, \ldots, p\}, j \in J_-(x^*)\}$ remains constant for all $x \in V$. It was proved in [8] that CRSC is sufficient for the existence of Lagrange multipliers at a local minimizer, and this is the weakest of the CQs we have discussed.

CQ conditions discussed above in the NLP context have multiple applications. For instance, RCRCQ was used to compute the directional derivative of the value function in [19], as well as to prove the convergence of a second-order augmented Lagrangian algorithm to second-order stationary points in [6]. RCPLD and CRSC were shown to be sufficient for proving first-order global convergence of several algorithms while also implying the validity of an error bound property (cf. [8]). Noteworthy, under CRSC, all inequality constraints in the set $J_-(x^*)$ behave locally as equality constraints, in the sense that they are active at any feasible point in a neighborhood of x^*. Therefore, we strongly believe that the extension of these notions to a conic framework may have a major impact in stability and algorithmic theory for conic programming.

3 Constraint qualifications conditions in second-order cone programming

Let us consider the second-order cone programming (SOCP) problem as follows:

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & h_i(x) = 0, \quad i = 1, \ldots, p, \\ & g_j(x) \in K_{m_j}, \quad j = 1, \ldots, \ell, \end{aligned} \tag{2}$$

where the functions are continuously differentiable and the second-order cones are denoted by $K_{m_j} := \{(z_0, \overline{z}) \in \mathbb{R} \times \mathbb{R}^{m_j - 1} \mid z_0 \geq \|\overline{z}\|\}$ when $m_j > 1$, and $K_{m_j} := \mathbb{R}_+$ (non-negative reals) otherwise.

We say that the Karush–Kuhn–Tucker (KKT) conditions hold for problem (2) at a feasible point x^* if there exists $\lambda \in \mathbb{R}^p$, $\mu_j \in K_{m_j}$, $j = 1, \ldots, \ell$, such that

$$\nabla_x L(x^*, \lambda, \mu) = \nabla f(x^*) + J_h(x^*)^T \lambda - \sum_{j=1}^{\ell} J_{g_j}(x^*)^T \mu_j = 0, \tag{3}$$

$$\langle \mu_j, g_j(x^*) \rangle = 0, \quad j = 1, \ldots, \ell. \tag{4}$$

Here, $L(x, \lambda, \mu) := f(x) + \langle \lambda, h(x) \rangle - \sum_{j=1}^{\ell} \langle \mu_j, g_j(x) \rangle$ is the standard Lagrangian function for problem (2), and $\nabla_x L(x, \lambda, \mu)$ denotes the gradient of L at (x, λ, μ) with respect to x. As usual, the set of all Lagrange multipliers (λ, μ) associated with the feasible point x^*, such that (3)–(4) are fulfilled, is denoted by $\Lambda(x^*)$.

As in NLP, one needs to assume a suitable CQ in order to ensure the existence of Lagrange multipliers associated with a local minimizer. In what follows, we recall the elements needed to define these CQs in the SOCP context.

The topological interior of K_{m_j}, denoted by $\text{int}(K_{m_j})$, and the non-zero boundary, denoted by $\text{bd}^+(K_{m_j})$, are respectively defined by

$$\begin{aligned} \text{int}(K_{m_j}) &:= \{(z_0, \overline{z}) \in \mathbb{R} \times \mathbb{R}^{m_j - 1} \mid z_0 > \|\overline{z}\|\}, \\ \text{bd}^+(K_{m_j}) &:= \{(z_0, \overline{z}) \in \mathbb{R} \times \mathbb{R}^{m_j - 1} \mid z_0 = \|\overline{z}\| > 0\}. \end{aligned}$$

Thus, given a feasible point x^*, we introduce the index sets:

$$\begin{aligned} I_{int}(x^*) &:= \{j \in \{1, \ldots, \ell\} \mid g_j(x^*) \in \text{int}(K_{m_j})\}, \\ I_B(x^*) &:= \{j \in \{1, \ldots, \ell\} \mid g_j(x^*) \in \text{bd}^+(K_{m_j})\}, \\ I_0(x^*) &:= \{j \in \{1, \ldots, \ell\} \mid g_j(x^*) = 0\}. \end{aligned}$$

Moreover, the complementarity condition (4) can be equivalently written as

$$\mu_j \circ g_j(x^*) = 0, \quad j = 1, \ldots, \ell, \tag{5}$$

where the operation $\circ$ is defined for any couple of vectors $y := (y_0, \bar{y})$ and $s := (s_0, \bar{s})$, with the same dimension, as follows:

$$y \circ s := \begin{pmatrix} \langle y, s \rangle \\ y_0 \bar{s} + s_0 \bar{y} \end{pmatrix}.$$

For more details about this operation, its algebraic properties and its relation with Jordan algebras, see [1, Section 4] and references therein.

From (5), it is easy to check that complementarity condition is equivalently written in terms of the above-mentioned index sets as follows:

$$\mu_j = 0 \text{ if } j \in I_{int}(x^*), \quad \mu_j = \alpha_j R_{m_j} g_j(x^*), \text{ for some } \alpha_j \geq 0, \text{ if } j \in I_B(x^*), \tag{6}$$

and no condition on μ_j can be inferred when $j \in I_0(x^*)$. Here, R_m is an $m \times m$ diagonal matrix whose first entry is 1 and the remaining ones are -1. Consequently, KKT conditions at x^* can be characterized as the existence of $\lambda \in \mathbb{R}^p$, $\mu_j \in K_{m_j}$, $j \in I_0(x^*)$, and $\alpha_j \geq 0$, $j \in I_B(x^*)$, such that

$$\nabla f(x^*) + J_h(x^*)^T \lambda - \sum_{j \in I_0(x^*)} J_{g_j}(x^*)^T \mu_j - \sum_{j \in I_B(x^*)} \alpha_j \nabla \phi_j(x^*) = 0, \tag{7}$$

where

$$\phi_j(x) := \frac{1}{2}([g_j(x)]_0^2 - \|\overline{g_j(x)}\|^2) \quad \text{for all } j \in I_B(x^*).$$

Indeed, it is straightforward to check that $\nabla \phi_j(x) = J_{g_j}(x)^T R_{m_j} g_j(x)$ and multipliers μ_j for all $j \notin I_0(x^*)$ are recovered from (6).

The use of mappings ϕ_j is a consequence of applying the reduction approach to problem (2). Actually, condition (7) is simply KKT conditions at point x^* for a locally equivalent version of problem (2) for which constraints $g_j(x) \in K_{m_j}$ are replaced by $\phi_j(x) \geq 0$ when $j \in I_B(x^*)$, and are omitted when $j \in I_{int}(x^*)$. For the sake of completeness, this reduced equivalent problem is explicitly stated here below:

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & h_i(x) = 0, \quad i = 1, \ldots, p, \\ & g_j(x) \in K_{m_j}, \; j \in I_0(x^*), \\ & \phi_j(x) \geq 0, \quad j \in I_B(x^*). \end{aligned} \tag{8}$$

Despite its apparent simplicity in the SOCP setting, the reduction approach is a key tool in conic programming. It permits obtaining first- and second-order optimality conditions, to simplify some well-known CQs, among other crucial properties. See [13, Section 3.4.4] and [12, Section 4] for more details. Throughout this article we will use KKT condition (7) and problem (8) to adapt CQ conditions from NLP to the SOCP setting (2).

One of the most used (and strong) conditions to guarantee the existence of a Lagrange multiplier at a local minimizer x^* is the nondegeneracy condition. Thanks to the reduction approach (cf. [13, Equation 4.172]), this condition can be equivalently defined as follows:

Definition 1 Let x^* be a feasible point of (2). Consider all the row vectors of the matrices $J_h(x^*)$ and $J_{g_j}(x^*)$, $j \in I_0(x^*)$ together with the row vectors $\nabla\phi_j(x^*)^T$, $j \in I_B(x^*)$. We say that *nondegeneracy* holds at x^* when these vectors are linearly independent.

The nondegeneracy condition implies the existence and uniqueness of a Lagrange multiplier at a local minimizer x^*, and the reciprocal is true provided that (x^*, λ, μ) (with $(\lambda, \mu) \in \Lambda(x^*)$) is strictly complementary, that is, $g_j(x^*) + \mu_j \in \operatorname{int}(K_{m_j})$ for all $j = 1, \ldots, \ell$; see [13, Proposition 4.75]. Thus, nondegeneracy is the analogue of LICQ from nonlinear programming. Note that there are other definitions of nondegeneracy e.g. [1, Definition 18] and [12, Definition 16]. However, all these definitions coincide in the case of SOCP problem (2). We address the reader to [12, Section 4] for more details about nondegeneracy in the context of SOCP.

As LICQ in NLP, nondegeneracy condition is often considered too strong. For this reason, one typically assumes a weaker condition, called Robinson's CQ, which was originally defined in [23] for a general conic setting. In our SOCP setting, we can use characterizations given in [13, Proposition 2.97, Corollary 2.98 and Lemma 2.99] to obtain the following equivalent definition:

Definition 2 Let x^* be a feasible point of (2). We say that *Robinson's CQ* holds at x^* if

$$J_h(x^*)^T\lambda + \sum_{j=1}^{\ell} J_{g_j}(x^*)^T\mu_j = 0 \text{ and } \lambda \in \mathbb{R}^m,\ \mu_j \in K_{m_j},\ \langle \mu_j, g_j(x^*)\rangle = 0,\ j = 1, \ldots, \ell \quad (9)$$
$$\Rightarrow\ \lambda = 0 \text{ and } \mu_j = 0,\ j = 1, \ldots, \ell.$$

As in NLP, when x^* is assumed to be a local solution of (2), Robinson's CQ (9) is equivalent to saying that the set of Lagrange multipliers $\Lambda(x^*)$ is nonempty and compact (cf. [13, Props. 3.9 and 3.17]). In this sense, condition (9) can be seen as an extension of Mangasarian-Fromovitz CQ in NLP to the SOCP setting (2), written in a dual form.

Thanks to (6), condition (9) can be rewritten as follows:

$$\begin{aligned} &J_h(x^*)^T\lambda + \sum_{j\in I_0(x^*)} J_{g_j}(x^*)^T\mu_j + \sum_{j\in I_B(x^*)} \alpha_j\nabla\phi_j(x^*) = 0,\\ &\lambda \in \mathbb{R}^m, \mu_j \in K_{m_j}, j \in I_0(x^*);\ \alpha_j \geq 0, j \in I_B(x^*)\\ &\Rightarrow\ \lambda = 0, \mu_j = 0, j \in I_0(x^*);\ \alpha_j = 0, j \in I_B(x^*). \end{aligned} \quad (10)$$

As we will see in the forthcoming sections, condition (10) best fits our analysis.

Note that (10) can be interpreted as a conic linear independence of the (transposed) Jacobians and gradients involved in its definition. Indeed, given some finite number of convex and closed cones C_j and denoting by $\prod_j C_j$ the cartesian product of these sets, we say that a correspondent set of matrices V_j of appropriate dimensions is $\prod_j C_j$-linearly independent if

$$\sum_j V_j s_j = 0 \text{ and } -s_j \in C_j^\circ \text{ for all } j \;\; \Rightarrow \;\; s_j = 0 \text{ for all } j.$$

Then, (10) coincides with the $\{0_p\} \times \prod_{j \in I_0(x^*)} K_{m_j} \times \mathbb{R}_+^{|I_B(x^*)|}$-linear independence of matrices: $J_h(x^*)^T$, $J_{g_i}(x^*)^T$ with $j \in I_0(x^*)$, and $\nabla \phi_j(x^*)$ with $j \in I_B(x^*)$. Here, 0_p denotes the null vector in $\mathbb{R}^p$. Moreover, when $C_j = \mathbb{R}_+$ for all j in the definition above (and consequently, each matrix V_j is simply a column vector), $\prod_j C_j$-linear independence coincides with the well-known positive linear independence. Then, condition (10) reminds the characterization of Mangasarian-Fromovitz CQ condition given by the positive linear independence of the gradients of active constraints (after replacing each equality constraint $h_i(x) = 0$ by two inequalities $h_i(x) \geq 0$ and $h_i(x) \leq 0$). It is also interesting to note that $\{0_p\} \times \prod_{j=1,\ldots,\ell} K_{m_j}$-linear independence of matrices $J_h(x^*)^T$ and $J_{g_i}(x^*)^T$ with $j = 1, \ldots, \ell$, is strictly stronger than Robinson's CQ (9). This again shows how useful is the reduction approach for our analysis. Given the analyzed above, when Robinson's CQ fails, we say that the corresponding matrices in (10) are conic linearly dependent.

4 The Approximate-KKT approach

For the nonlinear programming problem (1), the following *Approximate-KKT* (AKKT) necessary optimality condition [5] is well known:

Theorem 1 *Let x^* be a local minimizer of (1). Then, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\mu^k\} \subset \mathbb{R}^q_+$ such that $x^k \to x^*$ and*

$$\nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \to 0. \tag{11}$$

We define $\mu_j^k \to 0$ (or, equivalently, $\mu_j^k = 0$) for $j \notin A(x^*)$. Note that this does not require any constraint qualification at all and the sequence of approximate Lagrange multipliers $\{(\lambda^k, \mu^k)\}$ may be unbounded. If the sequence has a bounded subsequence, one may take a convergent subsequence such that the KKT conditions hold. In the unbounded case, one may define $M^k := \max\{|\lambda_i^k|, i = 1, \ldots, p; \mu_j^k, j \in A(x^*)\} \to +\infty$ and divide the expression in (11) by M^k. Thus, one may take an appropriate subsequence such that

$$\frac{\lambda^k}{M^k} \to \lambda \in \mathbb{R}^p \quad \text{and} \quad \frac{\mu_j^k}{M^k} \to \mu_j \geq 0, \;\; j \in A(x^*),$$

obtaining the existence of scalars λ_i, $i = 1, \dots, p$; $\mu_j \geq 0$, $j \in A(x^*)$, not all equal to zero, satisfying

$$\sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0.$$

That is, the gradients of equality constraints and active inequality constraints are positive linearly dependent. This provides a simple proof for the existence of Lagrange multipliers under the Mangasarian-Fromovitz CQ (MFCQ). A very similar argument shows that the set of Lagrange multipliers at x^* is bounded if, and only if, MFCQ holds.

In order to go beyond MFCQ in nonlinear programming, one relies on the well-known *Carathéodory's Lemma*, as stated in [7]:

Lemma 1 *Let $v_1, \dots, v_{p+q} \in \mathbb{R}^n$ be such that $\{v_i\}_{i=1}^p$ are linearly independent. Consider scalars $\beta_i, i = 1, \dots, p+q$, and denote $y := \sum_{i=1}^{p+q} \beta_i v_i$. Then, there exist $J \subseteq \{p+1, \dots, p+q\}$ and scalars $\hat{\beta}_i$, $i \in \{1, \dots, p\} \cup J$, such that $\{v_i\}_{i \in \{1,\dots,p\} \cup J}$ are linearly independent, $\beta_i > 0$ implies $\hat{\beta}_i > 0$, for all $i \in J$, and $y = \sum_{i \in \{1,\dots,p\} \cup J} \hat{\beta}_i v_i$.*

Thus, in order to prove that CRCQ (and its weaker variants) is a CQ for the nonlinear programming problem (1), we apply Carathéodory's Lemma to (11). This yields

$$\nabla f(x^k) + \sum_{i \in I^k} \tilde{\lambda}_i^k \nabla h_i(x^k) + \sum_{j \in J^k} \tilde{\mu}_j^k \nabla g_j(x^k) \to 0,$$

with $I^k \subseteq \{1, \dots, p\}$, $J^k \subseteq A(x^*)$, $\tilde{\mu}_j^k \geq 0$, $j \in J^k$, and such that the vectors of the set $\{\nabla h_i(x^k)\}_{i \in I^k} \cup \{\nabla g_j(x^k)\}_{j \in J^k}$ are linearly independent for all k. Here, by the infinite pigeonhole principle and passing to a subsequence if necessary, index subsets I^k and J^k can be taken as fixed and not depending on k. Then, the AKKT approach described above is similarly followed. It is worth to emphasize here that the application of Carathéodory's Lemma preserves the sign of the candidate to multipliers, that is, $\tilde{\mu}_j^k$ has the same sign than μ_j^k. This is a crucial step which is not clearly extended to the conic case (see [3]). Note that if $\{\nabla h_i(x^k)\}_{i=1}^p$ is linearly independent for all k, we may take $I_k = \{1, \dots, p\}$, which will be relevant in our analysis.

In the sequel, we will use the extension of the AKKT necessary optimality condition for second-order cone programming (2), as presented in [4]:

Theorem 2 *Let x^* be a local minimizer of (2). Then, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\mu_j^k\} \subset K_{m_j}$, $j \in I_0(x^*)$, $\{\alpha_j^k\} \subset \mathbb{R}_+$, $j \in I_B(x^*)$ such that $x^k \to x^*$ and*

$$\nabla f(x^k) + J_h(x^k)^T \lambda^k - \sum_{j \in I_0(x^*)} J_{g_j}(x^k)^T \mu_j^k - \sum_{j \in I_B(x^*)} \alpha_j^k \nabla \phi_j(x^k) \to 0. \quad (12)$$

5 A proposal of constraint qualifications for second-order cones

Following the previous discussion, we present a "naive" formulation of constant rank constraint qualifications for the second-order cone programming problem (2).

Definition 3 Let x^* be a feasible point of problem (2) and $I \subseteq \{1, \ldots, p\}$ be such that $\{\nabla h_i(x^*)\}_{i \in I}$ is a basis of the linear space generated by vectors $\{\nabla h_i(x^*)\}_{i=1}^p$. We say that the *Relaxed Constant Positive Linear Dependence (RCPLD)* condition holds at x^* when, for all $J \subseteq I_B(x^*)$, there exists a neighborhood V of x^* such that:

- $\{\nabla h_i(x)\}_{i=1}^p$ has constant rank for all x in V;
- if the system

$$\sum_{i \in I} \lambda_i \nabla h_i(x^*) + \sum_{j \in I_0(x^*)} J_{g_j}(x^*)^T \mu_j + \sum_{j \in J} \alpha_j \nabla \phi_j(x^*) = 0,$$
$$\lambda_i \in \mathbb{R}, i \in I; \quad \mu_j \in K_{m_j}, j \in I_0(x^*); \quad \alpha_j \geq 0, j \in J,$$

  has a not all zero solution $(\lambda_i)_{i \in I}, (\mu_j)_{j \in I_0(x^*)}, (\alpha_j)_{j \in I_B(x^*)}$, then vectors $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla \phi_j(x)\}_{j \in J}$ are linearly dependent for all x in V.

Note that Robinson's CQ implies RCPLD since it states the conic linear independence of the corresponding sets (and thus, for all its subsets) while RCPLD allows its conic linear dependence, as long as the linearly dependence is maintained for a reduced subset in a neighborhood.

The definition above takes into account our inability to relax Robinson's CQ for cones K_{m_j} with $j \in I_0(x^*)$, as the linear dependence for x near x^* is required only for equalities and for constraints at the boundary. Indeed, note that in the case when $I_B(x^*) = \emptyset$ and no equalities are considered (i.e., $p = 0$), RCPLD coincides with Robinson's CQ (9). This is an immediate consequence of the adopted convention that states that the empty set is always a linear independent set. On the other hand, we are aware that Definition 3 is unnecessarily strong when $m_j = 1$ for an index $j \in I_0(x^*)$. Indeed, in such case, the associated inequality $g_j(x) \in K_{m_j}$ corresponds to an inequality constraint of the form $g_j(x) \geq 0$, which is active at x^*. Hence, RCPLD definition can be slightly modified to take this situation into account as follows: define $A(x^*) := \{j \in I_0(x^*) \mid m_j = 1\}$, and remove those indices from $I_0(x^*)$, that is, define $\tilde{I}_0(x^*) := I_0(x^*) \backslash A(x^*)$. Indices in $A(x^*)$ can thus be treated similarly to those in $I_B(x^*)$. So, by defining $\phi_j(x) := g_j(x)$ when $j \in A(x^*)$, a slightly weaker version of RCPLD can be obtained by replacing $I_0(x^*)$ by $\tilde{I}_0(x^*)$ and $I_B(x^*)$ by $I_B(x^*) \cup A(x^*)$ in Definition 3. Since this modification has no consequence in the proof of Theorem 3, we do not include it in its statement.

The point raised in the last paragraph explains why Definition 3 is considered a "naive" extension of a constant rank-type condition. Before proving that RCPLD is a CQ for problem (2), we make further observations related to this point.

Remark 1 (a) When we choose $J = \emptyset$ in Definition 3, we necessarily obtain that there is no non-zero solution (λ_i, μ_j), with $i \in I$ and $j \in I_0(x^*)$, to the system:

$$\sum_{i\in I} \lambda_i \nabla h_i(x^*) + \sum_{j\in I_0(x^*)} J_{g_j}(x^*)^T \mu_j = 0 \quad \text{and} \quad \lambda_i \in \mathbb{R}, i \in I; \quad \mu_j \in K_{m_j}, j \in I_0(x^*).$$

This is equivalent to saying that Robinson's CQ holds at x^* for the constrained set $\Gamma_0 := \{x \mid h_i(x) = 0,\ i \in I,\ g_j(x) \in K_{m_j},\ j \in I_0(x^*)\}$. So, RCPLD ensures that Robinson's CQ is fulfilled at x^* for the active set Γ_0. Actually, by using the slight modification discussed above, we can exclude standard nonlinear constraints from $I_0(x^*)$, and conclude that it only implies the weaker condition: Robinson's CQ holds at x^* for the constrained set $\tilde{\Gamma}_0 := \{x \mid h_i(x) = 0,\ i \in I,\ g_j(x) \in K_{m_j},\ j \in I_0(x^*),\ m_j > 1\}$.

(b) Consider the case when problem (2) reduces to NLP (1), that is, $\tilde{I}_0(x^*) = \emptyset$ and $I_B(x^*) = \emptyset$. Then, RCPLD in Definition 3 reduces to the respective definition for nonlinear programming [7]. In particular, by enlarging the system to include $\alpha_j \in \mathbb{R}$, $j \in J$, instead of only considering $\alpha_j \geq 0$, $j \in J$, the definition reduces to an equivalent characterization (see [7]) of RCRCQ: $\{\nabla h_i(x)\}_{i=1}^p$ has constant rank for x around x^* and for all $J \subseteq A(x^*)$, if the set $\{\nabla h_i(x^*)\}_{i\in I} \cup \{\nabla \phi_j(x^*)\}_{j\in J}$ is linearly dependent, then $\{\nabla h_i(x)\}_{i\in I} \cup \{\nabla \phi_j(x)\}_{j\in J}$ must remain linearly dependent for all x in a neighborhood of x^* (here, the set I is fixed as in Definition 3). The latter also explains why RCPLD, given in Definition 3, is considered a constant rank-type condition for problem (2).

(c) Differently from the definition of nondegeneracy and Robinson's CQ, the choice of the reduction function $\phi(\cdot)$ gives rise to different constant rank conditions. For instance, one could formulate a similar, but different, condition by considering the alternative reduction function $\tilde{\phi}_j(x) := [g_j(x)]_0 - \|\overline{g_j(x)}\|$ for $j \in I_B(x^*)$. This is a well-known fact for nonlinear programming, which establishes that when a constraint set satisfies CRCQ, it can be rewritten in such a way that it fulfills Robinson's CQ [16]. See also [17] where the result is proved under a weaker CQ.

Theorem 3 *Let x^* be a feasible point of problem (2) satisfying the AKKT condition (12) and RCPLD. Then, the KKT conditions hold at x^*. In particular, RCPLD is a constraint qualification.*

Proof AKKT condition (12) ensures the existence of sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\mu_j^k\} \subset K_{m_j}$, $j \in I_0(x^*)$, $\{\alpha_j^k\} \subset \mathbb{R}_+$, $j \in I_B(x^*)$, such that $x^k \to x^*$ and

$$\nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^k \nabla h_i(x^k) - \sum_{j\in I_0(x^*)} J_{g_j}(x^k)^T \mu_j^k - \sum_{j\in I_B(x^*)} \alpha_j^k \nabla \phi_j(x^k) \to 0.$$

By the constant rank assumption on the equality constraints, and the definition of I, we may rewrite $\sum_{i=1}^p \lambda_i^k \nabla h_i(x^k) = \sum_{i\in I} \tilde{\lambda}_i^k \nabla h_i(x^k)$ for new scalars $\tilde{\lambda}_i^k \in \mathbb{R}$, $i \in I$, such that vectors $\{\nabla h_i(x^k)\}_{i\in I}$ are linearly independent. Applying Carathéodory's Lemma, for each k, we get $J^k \subseteq I_B(x^*)$ and new scalars $\hat{\lambda}_i^k \in \mathbb{R}$, $i \in I$, $\hat{\alpha}_j^k \geq 0$, $j \in J^k$, such that

$$\nabla f(x^k) + \sum_{i \in I} \hat{\lambda}_i^k \nabla h_i(x^k) - \sum_{j \in I_0(x^*)} J_{g_j}(x^k)^T \mu_j^k - \sum_{j \in J^k} \hat{\alpha}_j^k \nabla \phi_j(x^k) \to 0, \quad (13)$$

and vectors $\{\nabla h_i(x^k)\}_{i \in I} \cup \{\nabla \phi_j(x^k)\}_{j \in J^k}$ are linearly independent. By the infinite pigeonhole principle, without loss of generality we can consider subsequences, which are renamed as the original ones, for which sets J^k are the same for all k. This set is denoted by J.

Define $M^k := \max\{|\hat{\lambda}_i^k|, i \in I; \|\mu_i^k\|, i \in I_0(x^*); \hat{\alpha}_j, j \in J\}$. If $\{M^k\}$ is bounded, any accumulation point of $\{\hat{\lambda}_i^k, i \in I; \mu_i^k, i \in I_0(x^*); \hat{\alpha}_j, j \in J\}$ (after replacing by 0 the values for indices that are neither in I, nor in J) satisfies (7). Hence, x^* is a KKT point of (2). Otherwise, we may take a subsequence such that $M^k \to +\infty$, and divide the expression in (13) by M^k, considering convergent subsequences such that

$$-\frac{\hat{\lambda}_i^k}{M^k} \to \lambda_i \in \mathbb{R}, \ i \in I; \qquad \frac{\mu_j^k}{M^k} \to \mu_j \in K_{m_j}, \ j \in I_0(x^*);$$
$$\frac{\hat{\alpha}_j^k}{M^k} \to \alpha_j \geq 0, \ j \in J, \qquad \text{with } (\lambda_i, \mu_j, \alpha_j) \neq 0,$$

and obtaining

$$\sum_{i \in I} \lambda_i \nabla h_i(x^*) + \sum_{j \in I_0(x^*)} J_{g_j}(x^*)^T \mu_j + \sum_{j \in J} \alpha_j \nabla \phi_j(x^*) = 0.$$

Then, since vectors $\{\nabla h_i(x^k)\}_{i \in I} \cup \{\nabla \phi_j(x^k)\}_{j \in J}$ are linearly independent, this contradicts the definition of RCPLD. □

Exact definition of RCPLD in nonlinear programming can be consulted in [7]. The definition of CRCQ [15], RCRCQ [19], and CPLD [21] may be analogously extended. They are omitted. We only introduce the extension of CRSC [8] for this SOCP setting, since its definition is more involving and differs from its nonlinear programming counterpart. For the sake of completeness, the definition of CRSC considers sets $\tilde{I}_0(x^*)$ and $A(x^*)$. To prove that CRSC is a CQ is enough to follow the proof of Theorem 3, so it is omitted.

Definition 4 Let x^* be a feasible point of (2) and $J_-(x^*) \subseteq I_B(x^*) \cup A(x^*)$ be defined as

$$J_-(x^*) := \Bigg\{ j_0 \in I_B(x^*) \cup A(x^*) \Bigg| -\nabla \phi_{j_0}(x^*) = \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j \in I_B(x^*) \cup A(x^*)} \alpha_j \nabla \phi_j(x^*),$$
$$\text{for some } \lambda_i \in \mathbb{R}, \alpha_j \geq 0 \Bigg\}.$$

Set $J_+(x^*) := I_B(x^*) \cup A(x^*) \backslash J_-(x^*)$. We also define $I \subseteq \{1, \ldots, p\}$ and $J \subseteq J_-(x^*)$ such that $\{\nabla h_i(x^*)\}_{i \in I} \cup \{\nabla \phi_j(x^*)\}_{j \in J}$ is a basis of the linear space generated by $\{\nabla h_i(x^*)\}_{i=1}^{p} \cup \{\nabla \phi_j(x^*)\}_{j \in J_-(x^*)}$. We say that the *Constant Rank of the Subspace Component* (CRSC) condition holds at x^* when there exists a neighborhood V of x^* such that:

– $\{\nabla h_i(x)\}_{i=1}^{p} \cup \{\nabla \phi_j(x)\}_{j\in J_-(x^*)}$ has constant rank for all x in V;
– the system

$$\sum_{i\in I} \nabla h_i(x^*)\lambda_i + \sum_{j\in \tilde{I}_0(x^*)} J_{g_j}(x^*)\mu_j + \sum_{j\in J\cup J_+(x^*)} \nabla\phi_j(x^*)\alpha_j = 0,$$

$$\lambda_i \in \mathbb{R}, i\in I; \quad \mu_j \in K_{m_j}, j\in \tilde{I}_0(x^*); \quad \alpha_j \in \mathbb{R}, j\in J; \quad \alpha_j \geq 0, j\in J_+(x^*),$$

has only the trivial solution.

Note that when $\tilde{I}_0(x^*) = \emptyset$, the second requirement in the definition of CRSC always holds [8].

As said above, both definitions, RCPLD and CRSC, are "naive" in the sense that they do not improve on Robinson's CQ regarding multi-dimensional cones at zero. That is, when all constraint indices belong to $\tilde{I}_0(x^*)$, both definitions coincide with Robinson's CQ (9). However, the example below shows that RCPLD and CRSC are strictly weaker than Robinson's CQ:

Example 1 Consider the constraint set defined by

$$g(x) := (g_0(x), g_1(x)) := (x, x) \in K_2,$$

where x is one-dimensional. Clearly, $x^* = 1$ is feasible and the single constraint is in the boundary, i.e. $I_B(x^*)$ is the only nonempty index set. Reduced constraint is such that $\phi(x) := \frac{1}{2}(g_0(x)^2 - g_1(x)^2) = 0$ for all x. Then, it follows that $\nabla\phi(x^*) = 0$ and consequently, Robinson's CQ fails. However, $\nabla\phi(x) = 0$ for all x, which implies that RCPLD holds. CRSC also holds by noting that the reduced constraint belongs to the index set $J_-(x^*)$, whose gradient has constant rank, and $\tilde{I}_0(x^*) = \emptyset$, which is sufficient for ensuring the second condition. Indeed, $J = \emptyset$ is a basis for the linear space generated by the constraint gradient in $J_-(x^*)$ and the result follows by the linear independence of the empty set.

6 Extension to semidefinite programming

Consider the semidefinite programming (SDP) problem with multiple constraints:

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & h(x) = 0, \\ & g_j(x) \in \mathbb{S}_+^{m_j}, \ j = 1, \ldots, \ell, \end{aligned} \tag{14}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^p$, and $g_j : \mathbb{R}^n \to \mathbb{S}^{m_j}$ are continuously differentiable functions, $\mathbb{S}^{m_j}$ is the linear space of $m_j \times m_j$ real symmetric matrices equipped with the inner product $A \cdot B := \text{trace}(AB)$, where $\text{trace}(AB)$ denotes the sum of the elements of the diagonal of AB for all matrices $A, B \in \mathbb{S}^{m_j}$, and

$$\mathbb{S}_+^{m_j} := \{M \in \mathbb{S}^{m_j} \mid z^T M z \geq 0, \forall z \in \mathbb{R}^{m_j}\}$$

is the closed convex cone of all positive semidefinite elements of $\mathbb{S}^{m_j}$, for all $j = 1, \ldots, \ell$. We denote by $\preceq_j$ the partial order relation induced by $\mathbb{S}^{m_j}_+$, that is, $A \preceq_j B$ if, and only if, $B - A \in \mathbb{S}^{m_j}_+$. For the sake of notation, the index j is omitted throughout the paper and this relation order is simply denoted by $\preceq$. The order relations $\succeq$, $\succ$, and $\prec$ are similarly defined.

We end this subsection by recalling the Karush–Kuhn–Tucker conditions in the SDP framework. We say that KKT conditions hold at a feasible point x^* of problem (14) when there exist Lagrange multipliers $\lambda \in \mathbb{R}^p$ and $\mu_j \in \mathbb{S}^{m_j}$, $j = 1, \ldots, \ell$ such that

$$\nabla f(x^*) + J_h(x^*)^T \lambda - \sum_{j=1}^{\ell} J_{g_j}(x^*)^T \mu_j, \tag{15a}$$

$$g_j(x^*) \cdot \mu_j = 0, \;\; j = 1, \ldots, \ell, \tag{15b}$$

with

$$J_{g_j}(x^*)^T z := (\partial_1 g_j(x^*) \cdot z, \ldots, \partial_n g_j(x^*) \cdot z)^T, \quad \forall z \in \mathbb{S}^{m_j},$$

where $\partial_i g_j(x^*)$ is the partial derivative of g_j with respect to the variable x_i, at x^*, for each $i = 1, \ldots, n$. In fact, $J_{g_j}(x^*)^T$ is the adjoint of the linear mapping $J_{g_j}(x^*)$, defined by

$$J_{g_j}(x^*) d := \sum_{i=1}^{n} d_i \partial_i g_j(x^*),$$

for all $d = (d_1, ..., d_n)^T \in \mathbb{R}^n$, $j = 1, \ldots, \ell$.

6.1 Revisiting constraint qualifications for multifold SDP

Constraint qualification conditions recalled in Sect. 3 for SOCP have been also well established for SDP problem (14). In this section, we start by quickly recalling Robinson's CQ, before proceeding with the study of nondegeneracy condition, which needs more attention for our purposes.

As in the SOCP setting, Robinson's CQ [23] can be equivalently characterized via the properties established in [13, Proposition 2.97, Corollary 2.98 and Lemma 2.99] in its dual form:

Definition 5 We say that *Robinson's CQ* holds at a feasible point x^* of problem (14) when

$$\left.\begin{aligned} J_h(x^*)^T \lambda + \sum_{j=1}^{\ell} J_{g_j}(x^*)^T \mu_j = 0, \\ g_j(x^*) \cdot \mu_j = 0, \; \forall j = 1, \ldots, \ell, \\ \mu_j \in \mathbb{S}^{m_j}_+, \; \forall j = 1, \ldots, \ell, \end{aligned}\right\} \;\; \Rightarrow \;\; \mu_j = 0, \; \forall j = 1, \ldots, \ell. \tag{16}$$

As in SOCP, Robinson's CQ is considered as the natural extension of Mangasarian-Fromovitz CQ from NLP to the SDP setting. Actually, when x^* is assumed to be a local solution of (2), Robinson's CQ (16) is equivalent to saying that the set of Lagrange multipliers $\Lambda(x^*)$ is nonempty and compact (cf. [13, Props. 3.9 and 3.17]).

Let us now recall nondegeneracy condition in the SDP context. The notion of nondegeneracy (called transversality therein) was introduced by Shapiro and Fan in [26, Section 2] by means of tangent spaces in the context of eigenvalue optimization. An equivalent form is proven in [13, Equation (4.172)] for reducible cones. This is adopted as a formal definition in our multifold SDP setting:

Definition 6 We say that a feasible point x^* of problem (14) is *nondegenerate* when the following relation is satisfied

$$\operatorname{Im} \mathscr{A}(x^*) + \{0\} \times \prod_{j=1}^{\ell} \operatorname{lin}(T_{\mathbb{S}_+^{m_j}}(g_j(x^*))) = \mathbb{R}^p \times \prod_{j=1}^{\ell} \mathbb{S}^{m_j}, \tag{17}$$

where

$$\mathscr{A}(x^*) := \begin{pmatrix} J_h(x^*) \\ J_{g_j}(x^*);\ j = 1, ..., \ell \end{pmatrix}$$

is a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^p \times \prod_{j=1}^{\ell} \mathbb{S}^{m_j}$.

As it happens in SOCP, the nondegeneracy condition is considered to be a natural analogue of LICQ from NLP to SDP. Actually, nondegeneracy condition (17) implies the existence and uniqueness of a Lagrange multiplier at a local minimizer x^*, and the reciprocal is true provided that (x^*, λ, μ) (with $(\lambda, \mu) \in \Lambda(x^*)$) is strictly complementary, that is, $g_j(x^*) + \mu_j \succ 0$ for all $j = 1, \ldots, \ell$; see [13, Proposition 4.75]. However, this analogy only makes sense when matrix blocks $g_j(x^*)$ are chosen in a "minimal" way, in the sense of avoiding zeros in the off diagonal entries. In particular, an NLP problem with ℓ inequality constraints should be modeled as an instance of (14) with $m_1 = \ldots = m_\ell = 1$. Only in that case, nondegeneracy coincides LICQ. To stress the point above, we recall here below some results from [11, Section 5].

Consider the NLP problem of minimizing $f(x)$ under two constraints: $g_1(x) \geq 0$ and $g_2(x) \geq 0$, where f, g_1, and g_2 are smooth real-valued functions. Let x^* be a local mimimun for which $g_1(x^*) = g_2(x^*) = 0$ and LICQ holds (i.e., vectors $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$ are linearly independent). Denote by $\bar{\mu}_1$ and $\bar{\mu}_2$ the unique associated Lagrange multipliers, and assume that strict complementarity holds: $\bar{\mu}_i > 0$ for $i = 1, 2$. If this NLP problem is written as the following SDP problem

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & \begin{bmatrix} g_1(x) & 0 \\ 0 & g_2(x) \end{bmatrix} \in \mathbb{S}_+^2, \end{aligned} \tag{18}$$

then nondegeneracy condition (17) never holds. Indeed, the Lagrange multiplier associated with x^* for the reformulated problem (18) is never unique. It is enough to note that the matrix

$$\bar{\mu} := \begin{bmatrix} \bar{\mu}_1 & 0 \\ 0 & \bar{\mu}_2 \end{bmatrix}$$

is an associated Lagrange multiplier as well as

$$\bar{\mu} + t \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

for any $t \in \mathbb{R}$ such that $t^2 \leq \bar{\mu}_1 \bar{\mu}_2$. Of course, this apparent inconsistency occurs not only for diagonal matrices but also for any SDP problem with a diagonal structure (see e.g. [11, Lemma 5.1]), and it is due to an inappropriate modeling decision regarding the sparse structure of the studied SDP problem.

On the other hand, this phenomenon does not occur with Robinson's CQ, which is always preserved independently of the block structure of the SDP constraint set. This may be one of the reasons why multifold SDP is not often taken into consideration in the literature, along with the fact that interior-point methods are knowingly capable of exploiting block-diagonal structure (see Gondzio's review [14] and references therein for details). It is not expected, though, that every constraint qualification will be preserved between multifold and block-diagonal representations. In particular, the constraint qualifications we define in the next section are defined by means of exploiting the multifold structure. In this context, they are strictly weaker than Robinson's CQ, while if one considers a single block-diagonal representation our condition would resume to Robinson's CQ. Furthermore, since our analysis is related to AKKT sequences, which describe the output of many practical algorithms, our results provide a stronger convergence theory for them when applied to SDP problems under multifold representation.

For more details about the nondegeneracy condition in the semidefinite programming context, see e.g. [11,25]. In particular, Nondegeneracy condition for multifold SDP given in Definition 6 and the discussion above are inspired from [11, Section 5].

In the next section we propose a naive RCPLD condition similar to Definition 3 for multifold SDP, as in (14). We note that CPLD has already been used in the context of SDP problems in [27], however, they consider the application of an augmented Lagrangian method for a mixed problem with SDP constraints and NLP constraints, where the NLP constraints are not penalized and are carried out to the subproblems. Hence, the usual CPLD is assumed for the NLP constrained subproblems, in the context of feasibility results, while Robinson's CQ is assumed for the full problem in the context of optimality results. In particular, no CPLD-type CQ is introduced for the full problem.

6.2 A constant rank condition for SDP

Denote the smallest eigenvalue of a matrix A by $\sigma_{\min}(A)$ and its associated unitary eigenvectors by $\nu_{\min}(A)$ and $-\nu_{\min}(A)$. It is known that $\sigma_{\min}$ is continuously differentiable at A when $\sigma_{\min}(A)$ is simple, i.e., when it has algebraic multiplicity equal to one, and that $J_{\sigma_{\min}}(A) = \nu_{\min}(A)\nu_{\min}(A)^T$ in this case (see, e.g., [26]). So, given a local minimizer x^*, the composition $\sigma_{\min} \circ g_j$ is a reduction mapping for the block j

when $\sigma_{\min}(g_j(x^*))$ is simple, playing a similar role to $\phi_j(x)$ for problem (8). Also, in this scenario,

$$\nabla(\sigma_{\min}(g_j(x)) = J_{g_j}(x)^T J_{\sigma_{\min}}(g_j(x)) \tag{19}$$

when x is close enough to x^*. This motivates us to define an analogue of problem (8) for SDP as follows:

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & h(x) = 0, \\ & g_j(x) \in \mathbb{S}_+^{m_j}, \quad j \in I_N(x^*), \\ & \sigma_{\min}(g_j(x)) \geq 0, \ j \in I_R(x^*), \end{aligned} \tag{20}$$

where

$$I_R(x^*) := \{j \in \{1, \dots, \ell\} \mid 0 = \sigma_{\min}(g_j(x^*)) \text{ is simple}\}$$

and

$$I_N(x^*) := \{j \in \{1, \dots, \ell\} \mid 0 = \sigma_{\min}(g_j(x^*)) \text{ is not simple}\}.$$

Note that (20) is locally equivalent to (14) and that we have removed for simplicity all the constraints such that $g_j(x^*) \succ 0$, i.e., the "inactive" ones, in the reformulated problem. However, in problem (20), we have not applied the reduction approach to blocks $j \in I_N(x^*)$. Roughly speaking, our approach consists of defining a constraint qualification that relaxes Robinson's CQ to a constant rank-type condition, but only at the constraints indexed by $I_R(x^*)$, which are the ones that are well-behaved enough to be fully replaceable by a single real-valued constraint. As in the SOCP case, our strategy for proving that this is indeed a constraint qualification is based on sequential optimality conditions.

In [9], the AKKT condition was extended for SDP. Next, we present an adapted version of it for problems with mixed NLP and SDP constraints, like (20):

Theorem 4 *Let x^* be a local minimizer of (20). Then, there exist AKKT sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\alpha_j^k\} \subset \mathbb{R}_+$, and $\{\mu_j^k\} \subset \mathbb{S}_+^{m_j}$ such that $x^k \to x^*$ and*

$$\begin{aligned} & \nabla f(x^k) + J_h(x^k)^T \lambda^k - \sum_{j \in I_N(x^*)} J_{g_j}(x^k)^T \mu_j^k \\ & \quad - \sum_{j \in I_R(x^*)} \alpha_j^k \nabla \sigma_{\min}(g_j(x^k)) \to 0, \end{aligned} \tag{21}$$

$$\sigma_i(g_j(x^*)) > 0 \Rightarrow \sigma_i(\mu_j^k) \to 0, \quad i = 1, \dots, m_j, \quad \forall j \in I_N(x^*), \tag{22}$$

where $\sigma_i(\mu_j^k)$ *and* $\sigma_i(g_j(x^*))$ *denote corresponding eigenvalues of* μ_j^k *and* $g_j(x^*)$*, respectively, regarding ordered orthonormal eigenbasis* $\{v_i(\mu_j^k)\}_{i=1}^{m_j}$ *and* $\{v_i(g_j(x^*))\}_{i=1}^{m_j}$ *such that* $v_i(\mu_j^k) \to v_i(g_j(x^*))$ *for all* $i = 1, \ldots, m_j$ *and all* $j \in I_N(x^*)$.

With this result at hand, we proceed in a similar manner to Definition 3 in order to extend the *Relaxed Constant Positive Linear Dependence* (RCPLD) condition to SDP via problem (20).

Definition 7 Let x^* be feasible for problem (14) and let $I \subseteq \{1, \ldots, p\}$ be such that $\{\nabla h_i(x^*)\}_{i \in I}$ is a basis for the space spanned by $\{\nabla h_i(x^*)\}_{i=1}^p$. We say that *Relaxed Constant Positive Linear Dependence* holds at x^* when, for every $J \subseteq I_R(x^*)$, there exists a neighborhood V of x^* such that:

- $\{\nabla h_i(x)\}_{i=1}^p$ has constant rank for all $x \in V$;
- If the system

$$J_h(x^*)^T \lambda + \sum_{j \in I_N(x^*)} J_{g_j}(x^*)^T \mu_j + \sum_{j \in J} \alpha_j \nabla \sigma_{\min}(g_j(x^*)) = 0,$$
$$\lambda \in \mathbb{R}^p, \qquad \mu_j \succeq 0, \ \forall j \in I_N(x^*), \qquad \alpha_j \geq 0, \ \forall j \in J$$

  has a nontrivial solution, then $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla \sigma_{\min}(g_j(x))\}_{j \in J}$ is linearly dependent for every $x \in V$.

Next, we show that RCPLD is a constraint qualification using AKKT sequences (Theorem 4).

Theorem 5 *Let* x^* *be a feasible point of problem* (14) *satisfying the AKKT condition* (21) *and RCPLD stated in Definition* 7*. Then, the KKT conditions* (15) *hold at* x^**. In particular, RCPLD is a constraint qualification.*

Proof Let $\{x^k\} \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^p$, $\{\alpha_j^k\} \subset \mathbb{R}_+$, and $\{\mu_j^k\} \subset \mathbb{S}_+^{m_j}$ be sequences such that (21) and (22) hold. By the constant rank assumption and the definition of I, the set $\{\nabla h_i(x^k)\}_{i \in I}$ is a basis for the space spanned by $\{\nabla h_i(x^k)\}_{i=1}^p$ when k is large enough. Hence, for all such k, there are new scalars $\tilde{\lambda}^k \in \mathbb{R}^{|I|}$ such that

$$\sum_{i=1}^p \lambda_i^k \nabla h_i(x^k) = \sum_{i \in I} \tilde{\lambda}_i^k \nabla h_i(x^k),$$

for all k. Set $\tilde{\lambda}_i^k = 0$ for all $i \notin I$. So, $J_h(x^k)^T \lambda^k = J_h(x^k)^T \tilde{\lambda}^k$ for all k.

Also, thanks to Carathéodory's Lemma (Lemma 1) in (21), for every fixed k there is a nonempty subset $J^k \subset I_R(x^*)$ such that $\{\nabla h_i(x^k)\}_{i \in I} \bigcup \{\nabla \sigma_{\min}(g_j(x^k))\}_{j \in J^k}$ is linearly independent and, consequently, (21) can be rewritten as follows

$$\nabla f(x^k) + J_h(x^k)^T \tilde{\lambda}^k - \sum_{j \in I_N(x^*)} J_{g_j}(x^k)^T \mu_j^k - \sum_{j \in J^k} \tilde{\alpha}_j^k \nabla \sigma_{\min}(g_j(x^k)) \to 0, \tag{23}$$

for some $\tilde{\alpha}_j^k \geq 0$, where $j \in J^k$. Note that in this process the scalars $\tilde{\lambda}_i^k$, $i \in I$, also changes, but we abuse the notation by still denoting them by $\tilde{\lambda}_i^k$. Now, by the infinite pigeonhole principle, we can assume, without loss of generality, that $J^k = J$, for all $k \in \mathbb{N}$. That is, we can take a subsequence if necessary such that J^k does not vary with k.

Now, we claim that the sequences $\{\tilde{\lambda}^k\}$, $\{\mu_j^k\}$, $j \in I_N(x^*)$, and $\{\tilde{\alpha}_j^k\}$, $j \in J$ are bounded. Indeed, set

$$M_k := \max\{\tilde{\alpha}_j^k,\ j \in J;\ \|\mu_j^k\|,\ j \in I_N(x^*);\ \|\tilde{\lambda}^k\|\}$$

and suppose that $\{M_k\}$ is unbounded. This implies, by passing to a subsequence if necessary, that

$$-\frac{\tilde{\lambda}_i^k}{M^k} \to \lambda_i \in \mathbb{R},\ i \in I; \qquad \frac{\mu_j^k}{M^k} \to \mu_j \in K_{m_j},\ j \in I_N(x^*);$$
$$\frac{\tilde{\alpha}_j^k}{M^k} \to \alpha_j \geq 0,\ j \in J, \qquad \text{with } (\lambda_i, \mu_j, \alpha_j) \neq 0.$$

Then, by dividing (21) by M_k and passing to the limit, we contradict RCPLD.

Finally, let $\bar{\mu}_j \in \mathbb{S}_+^{m_j}$ $(j \in I_N(x^*))$, $\bar{\alpha}_j \geq 0$ $(j \in I_R(x^*))$, and $\bar{\lambda}$, be limit points of the sequences $\{\mu_j^k\}$ $(j \in I_N(x^*))$, $\{\tilde{\alpha}_j^k\}$ $(j \in I_R(x^*))$, and $\{\tilde{\lambda}^k\}$, respectively. Note that these limit points are Lagrange multipliers associated with x^*. Indeed, by definition of $I_R(x^*)$, we always have $\sigma_{\min}(g_j(x^*))\bar{\alpha}_j = 0$, for all $j \in I_R(x^*)$. So, for each $j \in I_R(x^*)$ the matrix $\bar{\mu}_j := \bar{\alpha}_j v_{\min}(g_j(x^*))v_{\min}(g_j(x^*))^T$ is positive semidefinite and satisfies that $J_{g_j}(x^*)^T\bar{\mu}_j = \bar{\alpha}_j^k \nabla\sigma_{\min}(g_j(x^k))$ (cf. (19)). Additionally, set $\bar{\mu}_j := 0$ when j is such that $g_j(x^*) \succ 0$. Then, it follows from (21) that

$$\nabla f(x^*) + J_h(x^*)^T\bar{\lambda} - \sum_{j=1}^{\ell} J_{g_j}(x^*)^T\bar{\mu}_j = 0,$$

which together with (22) implies that $g_j(x^*) \cdot \bar{\mu}_j = 0$ for every j. The desired result follows.

□

The CRSC condition can also be extended in a very similar manner. That is, we treat the conic constraints that "look like equality constraints" near the feasible point x^*, as equality constraints, which means it is not necessary to consider the rank-type structure of every subset of their gradients, but only of one fixed set. To formalize our analyses, we define the set

$$J_-(x^*) := \Bigg\{ j_0 \in I_R(x^*) \Bigg| -\nabla\sigma_{min}(g_{j_0}(x^*)) = \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j \in I_R(x^*)} \alpha_j \nabla\sigma_{\min}(g_j(x^*)), \quad \text{for some } \lambda_i \in \mathbb{R}, \alpha_j \geq 0 \Bigg\}, \tag{24}$$

and the set $J_+(x^*) := I_R(x^*)\backslash J_-(x^*)$. Now, the *Constant Rank of the Subspace Component* (CRSC) constraint qualification for SDP is defined as follows:

Definition 8 Let x^* be a feasible point of (2) and $J_-(x^*) \subseteq I_R(x^*)$ be defined as in (24). We also take $I \subseteq \{1, \dots, p\}$ and $J \subseteq J_-(x^*)$ such that $\{\nabla h_i(x^*)\}_{i\in I} \cup \{\nabla \sigma_{\min}(g_j(x^*))\}_{j\in J}$ is a basis of the space spanned by the set $\{\nabla h_i(x^*)\}_{i=1}^p \cup \{\nabla \sigma_{\min}(g_j(x^*))\}_{j\in J_-(x^*)}$. We say that *Constant Rank of the Subspace Component* (CRSC) condition holds at x^* when there exists a neighborhood V of x^* such that:

– $\{\nabla h_i(x)\}_{i=1}^p \cup \{\nabla \sigma_{\min}(g_j(x))\}_{j\in J_-(x^*)}$ has constant rank for all x in V;
– the system

$$\sum_{i\in I} \lambda_i \nabla h_i(x^*) + \sum_{j\in I_N(x^*)} J_{g_j}(x^*)^T \mu_j + \sum_{j\in J\cup J_+(x^*)} \alpha_j \nabla \sigma_{\min}(g_j(x^*)) = 0,$$

$$\lambda_i \in \mathbb{R}, i \in I; \quad \mu_j \in \mathbb{S}_+^{m_j}, j \in I_N(x^*); \quad \alpha_j \in \mathbb{R}, j \in J; \quad \alpha_j \geq 0, j \in J_+(x^*),$$

has only the trivial solution.

It is possible to prove that CRSC is indeed a constraint qualification, but since the proof follows from the same arguments provided in the proof of Theorem 5, it is omitted. The next counterexample, analogous to Example 1, shows that CRSC and RCPLD are strictly weaker than Robinson's CQ.

Example 2 Consider the following pair of constraints:

$$g_1(x) := \frac{1}{2}\begin{bmatrix} x+1 & x-1 \\ x-1 & x+1 \end{bmatrix} \in \mathbb{S}_+^2, \quad g_2(x) := \frac{1}{2}\begin{bmatrix} 1-x & -x-1 \\ -x-1 & 1-x \end{bmatrix} \in \mathbb{S}_+^2$$

and the point $x^* = 0$, which is the unique feasible point. The eigenvalues of $g_1(x)$ are $\sigma_{\min}(g_1(x)) = x$ and $\sigma_{\max}(g_1(x)) = 1$, with corresponding eigenvectors $v_{\min}(g_1(x)) = (1, 1)^T$ and $v_{\max}(g_1(x)) = (1, -1)^T$, respectively, for all x close to x^*. With the same eigenvectors, the eigenvalues of $g_2(x)$ are $\sigma_{\min}(g_2(x)) = -x$ and $\sigma_{\max}(g_2(x)) = 1$, when x is close to x^*.

Also, note that $\sigma_{\min}(g_1(x^*))$ and $\sigma_{\min}(g_2(x^*))$ are both simple, which means the reformulation of the problem as in (20) is simply an NLP problem. Moreover, we have that $\nabla\sigma_{\min}(g_1(x)) = 1$, $\nabla\sigma_{\min}(g_2(x)) = -1$, for all x close enough to $x^* = 0$. Then, RCPLD and CRSC (with $J_-(x^*) = \{1, 2\}$ and, consequently, $J_+(x^*) = \emptyset$ and J equals either $\{1\}$ or $\{2\}$) hold. However, Robinson's CQ does not hold. Thus, RCPLD and CRSC are strictly implied by Robinson's CQ.

7 Conclusion

We have presented naive definitions of constant rank-type CQs for second-order cone programming and semidefinite programming. The definition is naive in the sense that no improvement is made with respect to irreducible constraints, where our definitions resume to Robinson's CQ. However, in general, our definitions are strictly weaker than Robinson's CQ. In order to present a definition that takes into account the true conic constraints, we expect that a much more involving implicit function approach or Approximate-KKT approach would be needed, which is a subject of current research. Note that, since augmented Lagrangian algorithms described in [4] and [9] generate an AKKT sequence for SOCP (2) and SDP (14) problems, respectively, CQs introduced in these notes are sufficient for showing global convergence to a KKT point without assuming Robinson's CQ.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**(1), 3–51 (2003)
2. Andersen, E.D., Roos, C., Terlaky, T.: Notes on duality in second order and p-order cone optimization. Optimization **4**(51), 627–643 (2002)
3. Andreani, R., Fukuda, E.H., Haeser, G., Ramírez, H., Santos, D.O., Silva, P.J.S., Silveira, T.P.: Erratum to: New constraint qualifications and optimality conditions for second order cone programs. submitted to Set-Valued and Variational Analysis (2020)
4. Andreani, R., Fukuda, E.H., Haeser, G., Santos, D.O., Secchin, L.D.: Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. Optimization Online (2019)
5. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011)
6. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A second-order sequential optimality condition associated to the convergence of algorithms. IMA J. Numer. Anal. **37**(4), 1902–1929 (2017)
7. Andreani, R., Haeser, G., Schuverdt, M., Silva, P.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**(1–2), 255–273 (2012)
8. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**(3), 1109–1135 (2012)
9. Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. Math. Program. **180**(1), 203–235 (2020)
10. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On second-order optimality conditions for nonlinear programming. Optimization **56**, 529–542 (2007)
11. Bonnans, J.F., Ramírez, H.: Strong regularity of semidefinite programs. Technical report DIM-CMM B-05-06-137, (2005)
12. Bonnans, J.F., Ramírez, H.: Perturbation analysis of second-order cone programming problems. Math. Program. **104**(2), 205–227 (2005)
13. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Verlag, New York (2000)
14. Gondzio, J.: Interior point methods 25 years later. European Journal of Operational Research **216**(3), 587–601 (2012)
15. Janin, R.: Direction derivate of the marginal function in nonlinear programming. Math. Program. Study **21**, 110–126 (1984)
16. Lu, S.: Implications of the constant rank constraint qualification. Math. Program. **126**(2), 365–392 (2011)
17. Minchenko, L.: Note on Mangasarian-Fromovitz-like constraint qualifications. J. Optim. Theory Appl. **182**, 1199–1204 (2019)
18. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)
19. Minchenko, L., Stakhovski, S.: Parametric nonlinear programming problems under the relaxed constant rank condition. SIAM J. Optim. **1**, 314–332 (2011)

20. Nocedal, J., Wright, S.: Numerical Optimization. Springer Science & Business Media, Berlin (2006) (2006)
21. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Optim. **10**(4), 963–981 (2000)
22. Ramana, M.V., Tunçel, L., Wolkowicz, H.: Strong duality for semidefinite programming. SIAM J. Optim. **3**(7), 641–662 (1997)
23. Robinson, S.M.: Stability theorems for systems of inequalities, Part II: differentiable nonlinear systems. SIAM J. Numer. Anal. **13**, 497–513 (1976)
24. Robinson, S.M.: Generalized equations and their solutions, Part II: applications to nonlinear programming. In: Optimality and stability in mathematical programming, pp. 200–221. Springer (1982)
25. Shapiro, A.: First and second-order analysis of nonlinear semidefinite programs. Math. Program. **77**(2), 301–320 (1997)
26. Shapiro, A., Fan, M.K.H.: On eigenvalue optimization. SIAM J. Optim. **5**, 552–569 (1995)
27. Wu, H., Luo, H., Ding, X., Chen, G.: Global convergence of modified augmented Lagrangian methods for nonlinear semidefinite programmings. Comput. Optim. Appl. **56**(3), 531–558 (2013)
28. Zhang, Y., Zhang, L.: New constraint qualifications and optimality conditions for second order cone programs. Set-Valued Var. Anal. **27**, 693–712 (2019)

Authors and Affiliations

R. Andreani[1] · G. Haeser[2] · L. M. Mito[2] · H. Ramírez[3] · D. O. Santos[4] · T. P. Silveira[2]

G. Haeser
ghaeser@ime.usp.br

R. Andreani
andreani@ime.unicamp.br

L. M. Mito
leokoto@ime.usp.br

H. Ramírez
hramirez@dim.uchile.cl

D. O. Santos
daiana@ime.usp.br

T. P. Silveira
thiagops@ime.usp.br

[1] Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil

[2] Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil

[3] Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Santiago, Chile

[4] Institute of Science and Technology, Federal University of São Paulo, São José dos Campos, SP, Brazil

Appendix C

Article: Global Convergence of Algorithms Under Constant Rank Conditions for Nonlinear Second-Order Cone Programming. doi.org/10.1007/s10957-022-02056-5

Journal of Optimization Theory and Applications (2022) 195:42–78
https://doi.org/10.1007/s10957-022-02056-5

Global Convergence of Algorithms Under Constant Rank Conditions for Nonlinear Second-Order Cone Programming

Roberto Andreani[1] · Gabriel Haeser[2] · Leonardo M. Mito[2] · C. Héctor Ramírez[3] · Thiago P. Silveira[2]

Received: 22 October 2021 / Accepted: 20 May 2022 / Published online: 27 June 2022

Abstract

In Andreani et al. (Weak notions of nondegeneracy in nonlinear semidefinite programming, 2020), the classical notion of nondegeneracy (or transversality) and Robinson's constraint qualification have been revisited in the context of nonlinear semidefinite programming exploiting the structure of the problem, namely its eigendecomposition. This allows formulating the conditions equivalently in terms of (positive) linear independence of significantly smaller sets of vectors. In this paper, we extend these ideas to the context of nonlinear second-order cone programming. For instance, for an m-dimensional second-order cone, instead of stating nondegeneracy at the vertex as the linear independence of m derivative vectors, we do it in terms of several statements of linear independence of 2 derivative vectors. This allows embedding the structure of the second-order cone into the formulation of nondegeneracy and, by extension, Robinson's constraint qualification as well. This point of view is shown to be crucial in defining significantly weaker constraint qualifications such as the constant rank

Communicated by Sándor Zoltán Németh.

Leonardo M. Mito
leokoto@ime.usp.br

Roberto Andreani
andreani@unicamp.br

Gabriel Haeser
ghaeser@ime.usp.br

C. Héctor Ramírez
hramirez@dim.uchile.cl

Thiago P. Silveira
thiagops@ime.usp.br

[1] Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil

[2] Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil

[3] Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Universidad de Chile, Santiago, Chile

constraint qualification and the constant positive linear dependence condition. Also, these conditions are shown to be sufficient for guaranteeing global convergence of several algorithms, while still implying metric subregularity and without requiring boundedness of the set of Lagrange multipliers.

Keywords Second-order cone programming · Constraint qualifications · Algorithms · Global convergence · Constant rank.

Mathematics Subject Classification 90C46 · 90C30

1 Introduction

The well-known *constant rank constraint qualification* (CRCQ) was introduced by Janin [29], for *nonlinear programming* (NLP), with the purpose of obtaining a formula for the Hadamard directional derivative of the value function. Prior to his work, similar results were known under the *Mangasarian–Fromovitz constraint qualification* (MFCQ) and the *linear independence constraint qualification* (LICQ).

Janin also showed that CRCQ neither implies nor is implied by MFCQ and, moreover, that CRCQ is strictly weaker than LICQ. After that, CRCQ has been widely employed in the NLP literature for instance in the study of stability , strong second-order necessary optimality conditions [5], global convergence of algorithms , among other applications. We remark that CRCQ explains in a very simple way the existence of Lagrange multipliers associated with affine constraints, such as in linear programming.

More recently, Qi and Wei [42] presented a condition called *constant positive linear dependence* (CPLD), which is strictly weaker than both MFCQ and CRCQ, and showed its application on the convergence of a general *sequential quadratic programming* (SQP) method for NLP. However, they did not prove that CPLD was a constraint qualification at the time. This issue was settled in a later article by Andreani et al. [16], where they proved that CPLD implies the *quasinormality* constraint qualification condition. Later, in [4], the convergence of an *augmented Lagrangian* method was also proved under CPLD. Other uses of constant rank-type constraint qualifications in NLP are discussed, for instance, in [14, 15, 29, 34, 35] and their references. In particular, the appropriate way of incorporating equality constraints in the definitions of CRCQ and CPLD is discussed, respectively, in [34] and [14].

Although constraint qualifications with applications toward convergence of algorithms are largely studied in NLP, the situation is quite different in *nonlinear second-order cone programming* (NSOCP), despite its many relevant applications—for example, in structural optimization and machine learning , hydroacoustic classification of fishes , and others . In NSOCP, this role is almost always covered by the so-called *nondegeneracy* condition (c.f. [18, Equation 25]) and *Robinson's constraint qualification* (Robinson's CQ) (c.f. [18, Equation 29]), which can be seen as natural generalizations of LICQ and MFCQ, respectively. The first work that attempted to extend CRCQ and its variants to the context of NSOCP is due to Zhang and Zhang [47], but their condition was invalidated by a counterexample given by Andreani et al. in [6].

Later, a "naive approach" to extend some constant rank-type constraint qualifications for NSOCP was presented by Andreani et al. in [11]; the adjective "naive" refers to the fact that some of the conic constraints were locally rewritten as NLP constraints whenever possible, yielding a new reformulated problem with mixed constraints, and then a hybrid condition between the NLP versions of CRCQ/CPLD and nondegeneracy/Robinson's CQ was presented. The major contribution of [11] is to show an effective way of dealing with those two distinct types of constraints via sequences of approximate stationary points.

Recently, we proposed in [12] a new geometrical characterization of CRCQ for NLP using the faces of the nonnegative orthant, which was naturally extended to the context of NSOCP as well as *nonlinear semidefinite programming* (NSDP). This has led us to an alternative constant rank-type constraint qualification that allowed us to derive strong second-order optimality conditions for NSDP and NSOCP without assuming compactness of the Lagrange multiplier set, similarly to what is known in NLP . However, no application toward algorithms was provided or suggested in [12]. Since the sequential approach from [11] seems more suitable for algorithms, we developed it even further for NSDP problems [9, 10] by directly exploiting the eigenvector structure of the problem, overcoming the limitations of the naive approach.

This paper introduces new constraint qualifications for NSOCP problems following similar ideas to those used in [9] and [10], but taking into account the structure of the second-order cone. For such, we will first introduce weak variants of the nondegeneracy condition and Robinson's CQ—here called *weak-nondegeneracy* and *weak-Robinson's CQ*—which are weaker than their original versions but that still reduce to LICQ and MFCQ, respectively, when an NLP problem is modeled as an NSOCP problem with several one-dimensional constraints. Moreover, we show that weak-nondegeneracy is strictly weaker than nondegeneracy, and we also clarify some relations between weak-nondegeneracy (weak-Robinson's CQ) and standard nondegeneracy (Robinson's CQ), which were only partially addressed in [9]. In particular, we show a new characterization of nondegeneracy in terms of the validity of weak-nondegeneracy plus the linear independence of a partial Jacobian of the constraints. The relationship of weak-Robinson's CQ and Robinson's CQ is also partially settled in our Theorem 3.1, which was left as an open problem for NSDP in [10]. With these new constraint qualifications at hand, we introduce new extensions of CRCQ and CPLD for NSOCP, which also recover their counterparts in NLP. We also discuss a mild adaptation of these new conditions that can be adopted with the purpose of proving global convergence results for algorithms that keep track of Lagrange multiplier estimates.

The structure of this paper is as follows: In Sect. 2, we present some notation and technical results. Sections 3 and 4 present weak constraint qualifications for NSOCP: weak-nondegeneracy condition, weak-Robinson's CQ, and two weak constant rank conditions. Also, we present some of their properties and a detailed comparison with other constraint qualifications from the literature, and among themselves. In Sect. 5, we introduce perturbed versions of the constant rank conditions of Sect. 4, and we present some algorithms related to them. We state the relationship between these perturbed variants and the so-called *metric subregularity CQ*. Finally, in Sect. 6, we summarize our results and discuss some ideas for future research.

2 Preliminaries

In this section, we will present our notation and some linear algebra and convex analysis tools needed for deriving the results of this paper.

2.1 Basic Results and Some Notation

For a given differentiable function $F\colon \mathbb{R}^n \to \mathbb{R}^m$, we denote the *Jacobian* matrix of F at a point $x \in \mathbb{R}^n$ by $DF(x)$; and the jth column of its *transpose*, $DF(x)^\top$, will be denoted by $\nabla F_j(x)$. We also adopt the usual inner product in $\mathbb{R}^m$, given by $\langle y, z\rangle := \sum_{j=1}^m y_j z_j$, along with the *Euclidean norm* $\|y\| := \sqrt{\langle y, y\rangle}$, for every $y, z \in \mathbb{R}^m$. The open ball (respective to the Euclidean norm) that has center at y and radius $\delta \geq 0$ will be denoted by $B(y, \delta)$, and its closure, by $\mathrm{cl}(B(y, \delta))$.

The orthogonal projection of a given $y \in \mathbb{R}^m$ onto a nonempty closed convex set $C \subseteq \mathbb{R}^m$ with respect to the Euclidean norm is defined as

$$\mathcal{P}_C(y) := \operatorname*{argmin}_{z \in C} \|z - y\|.$$

It is valid to mention that $\mathcal{P}_C(y)$ is well defined as a continuous function of y, since C is closed and convex. Also, when C is given by the Cartesian product of other nonempty closed convex sets $C_1, \ldots, C_q$, where $C_j \subseteq \mathbb{R}^{m_j}$ for every $j \in \{1, \ldots, q\}$, then for any $y := (y_1, \ldots, y_q) \in \mathbb{R}^{m_1+\cdots+m_q}$, we have

$$\mathcal{P}_C(y) = \left(\mathcal{P}_{C_1}(y_1), \ldots, \mathcal{P}_{C_q}(y_q)\right).$$

To relate our results to the classical ones from the literature, we will make use of a notion of *conic linear independence*, defined as follows:

Definition 2.1 Let $C \subseteq \mathbb{R}^m$ be a nonempty closed convex cone. A matrix $M \in \mathbb{R}^{n\times m}$ is said to be *C-linearly independent* if there is no nonzero $v \in C$ such that $Mv = 0$.

Roughly speaking, Definition 2.1 describes "injectivity over C." In particular, if C is the nonnegative orthant

$$\mathbb{R}^m_+ := \{y \in \mathbb{R}^m : \forall i \in \{1, \ldots, m\},\ y_i \geq 0\},$$

then Definition 2.1 reduces to a concept known in NLP as *positive linear independence* of the columns of M. Now, let us show a simple characterization of conic linear independence in terms of all finitely generated conical slices of the cone.

Lemma 2.1 *Let $C \subseteq \mathbb{R}^m$ be a closed convex cone such that there exists a (possibly infinite) index set S and, for each $w \in S$, a finite subset $\mathcal{E}_w \subseteq C$ whose elements are linearly independent, such that*

$$C = \bigcup_{w \in S} \mathrm{cone}(\mathcal{E}_w), \tag{1}$$

where cone$(\mathcal{E}_w)$ *denotes the conic hull of* $\mathcal{E}_w$*. Then, a matrix* $M \in \mathbb{R}^{n\times m}$ *is* C*-linearly independent if, and only if, the family* $\{Mv\}_{v\in\mathcal{E}_w}$ *is positively linearly independent, for every fixed* $w \in S$*.*

Proof Suppose that M is C-linearly independent, let $w \in S$ be arbitrary, and let $a_v \in \mathbb{R}_+$, $v \in \mathcal{E}_w$, be scalars such that

$$\sum_{v\in\mathcal{E}_w} a_v M v = M\left[\sum_{v\in\mathcal{E}_w} a_v v\right] = 0. \tag{2}$$

Since C is a convex cone, it follows that $\tilde{v} := \sum_{v\in\mathcal{E}_w} a_v v$ belongs to C, so $\tilde{v} = 0$ by hypothesis; and from the linear independence of $\mathcal{E}_w$ we have that $a_v = 0$ for every $v \in \mathcal{E}_w$. Thus, $\{Mv\}_{v\in\mathcal{E}_w}$ is positively linearly independent.

Conversely, assume that $\{Mv\}_{v\in\mathcal{E}_w}$ is positively linearly independent, and let $\tilde{v} \in C$ be such that $M\tilde{v} = 0$. Then, there is some $w \in S$ such that $\tilde{v} \in \text{cone}(\mathcal{E}_w)$; that is, there exist some scalars $a_v \geq 0$, $v \in \mathcal{E}_w$, such that $\tilde{v} = \sum_{v\in\mathcal{E}_w} a_v v$ and hence (2) holds, implying that $a_v = 0$ for all $v \in \mathcal{E}_w$; thus, $\tilde{v} = 0$. □

Remark 2.1 Considering $C = \mathbb{R}^m$ in the statement of the Lemma and replacing the conic hull by the linear span in (1), we arrive similarly at a characterization of the linear independence of the columns of M in terms of the linear independence of the family $\{Mv\}_{v\in\mathcal{E}_w}$, for every fixed $w \in S$.

A simple example to fix ideas on how to use Lemma 2.1 is to take the parametric representation of $\mathbb{R}^2$:

$$\begin{aligned} \mathbb{R}^2 &= \{(r\cos(w), r\sin(w))\colon w \in [0, 2\pi], r \geq 0\} \\ &= \bigcup_{w\in[0,2\pi]} \text{cone}((\cos(w), \sin(w))) \end{aligned} \tag{3}$$

so we have $C = \mathbb{R}^2$, $S = [0, 2\pi]$, and $\mathcal{E}_w = \{(\cos(w), \sin(w))\}$, $w \in S$. In this case, Lemma 2.1 simply states the trivial fact that a matrix $M \in \mathbb{R}^{n\times 2}$ is injective if, and only if, $M(\cos(w), \sin(w)) \neq 0$ for every $w \in [0, 2\pi]$. Moreover, the main object of our study, the *second-order cone* (or *Lorentz cone*):

$$\mathbb{L}_m := \begin{cases} \{y := (y_0, \widehat{y}) \in \mathbb{R}\times\mathbb{R}^{m-1}\colon y_0 \geq \|\widehat{y}\|\}, & \text{if } m > 1, \\ \mathbb{R}_+, & \text{if } m = 1, \end{cases}$$

may benefit from Lemma 2.1 as well, since it can be written as

$$\mathbb{L}_m = \bigcup_{\substack{w\in\mathbb{R}^{m-1} \\ \|w\|=1}} \text{cone}(\{(1, -w), (1, w)\}),$$

which corresponds to $S = \{w \in \mathbb{R}^{m-1}\colon \|w\| = 1\}$ and $\mathcal{E}_w = \{(1, -w), (1, w)\}$. In this case Lemma 2.1 states that a matrix $M \in \mathbb{R}^{n\times m}$ is $\mathbb{L}_m$-linearly independent if, and only if, the vectors

$$M(1, -w) \quad \text{and} \quad M(1, w) \tag{4}$$

are positively linearly independent for every $w \in \mathbb{R}^{m-1}$ such that $\|w\| = 1$. Furthermore, the standard notion of linear independence in $\mathbb{R}^m$ can also be stated in terms of the conical slices of $\mathbb{L}_m$, since it is a full-dimensional cone; indeed, observe that

$$\mathbb{R}^m = \bigcup_{\substack{w \in \mathbb{R}^{m-1} \\ \|w\|=1}} \operatorname{span}(\{(1, -w), (1, w)\}),$$

where span$(\{(1, -w), (1, w)\})$ denotes the *linear span* of the vectors $(1, -w)$ and $(1, w)$; then, the matrix M is $\mathbb{R}^m$-linearly independent (i.e., injective) if, and only if, the vectors (4) are linearly independent for every $w \in \mathbb{R}^{m-1}$ such that $\|w\| = 1$. Thus, we have replaced the linear independence of the m columns of M by a series of linear independence requirements of only two parameterized vectors in (4), independently of the size of m. With this point of view, we will be able to exploit the structure of the second-order cone, which will turn out to be essential in our analysis.

Furthermore, observe that Lemma 2.1 can be applied to products of closed convex cones $C = \prod_{j \in J} C_j$, where J is an index set, in order to describe C-linear independence of a family of matrices $\{M_j\}_{j \in J}$ mounted into a conveniently indexed block matrix

$$M := \begin{bmatrix} \vdots \\ M_j \\ \vdots \end{bmatrix}_{j \in J} \tag{5}$$

therefore, we will abuse the terminology to define the C-linear independence of the family $\{M_j\}_{j \in J}$ in terms of the above M throughout the paper.

To close this subsection, let us briefly recall the celebrated *Carathéodory's Lemma* [17, Exercise B.1.7] from convex analysis:

Lemma 2.2 *(Carathéodory's Lemma) Let $y_1, \ldots, y_p \in \mathbb{R}^n$, and let $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$ be arbitrary. Then, there exist some $J \subseteq \{1, \ldots, p\}$ and some scalars $\tilde{\alpha}_j$ with $j \in J$, such that $\{y_j\}_{j \in J}$ is linearly independent,*

$$\sum_{j=1}^{p} \alpha_j y_j = \sum_{j \in J} \tilde{\alpha}_j y_j,$$

and $\alpha_j \tilde{\alpha}_j > 0$, for all $j \in J$.

2.2 The Nonlinear Second-Order Cone Programming Problem

A (multifold) nonlinear second-order cone programming problem is usually stated in the form:

$$\begin{array}{lll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x), & \\ \text{subject to} & g_j(x) \in \mathbb{L}_{m_j}, \ \forall j \in \{1, \ldots, q\}, \end{array} \tag{NSOCP}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g_j\colon \mathbb{R}^n \to \mathbb{R}^{m_j}$ are continuously differentiable functions, for all $j \in \{1, \ldots, q\}$, and $\mathbb{L}_{m_j}$ is a second-order cone in $\mathbb{R}^{m_j}$. As usual, for a point $x \in \mathbb{R}^n$ we denote $g_j(x) = (g_{j,0}(x), \widehat{g}_j(x)) \in \mathbb{R} \times \mathbb{R}^{m_j-1}$. The feasible set of (NSOCP) will be denoted by $\mathcal{F}$. Also, we denote the *interior* and the *boundary excluding the origin* of $\mathbb{L}_{m_j}$ by $\operatorname{int}\mathbb{L}_{m_j}$ and $\operatorname{bd}_+\mathbb{L}_{m_j}$, respectively; and as usual in the study of NSOCP, for any $x \in \mathcal{F}$ we partition $\{1, \ldots, q\}$ as follows:

$$\begin{aligned} I_0(x) &:= \{j \in \{1, \ldots, q\}\colon g_j(x) = 0\}, \\ I_B(x) &:= \{j \in \{1, \ldots, q\}\colon g_j(x) \in \operatorname{bd}_+\mathbb{L}_{m_j}\}, \\ I_{\text{int}}(x) &:= \{j \in \{1, \ldots, q\}\colon g_j(x) \in \operatorname{int}\mathbb{L}_{m_j}\}. \end{aligned} \tag{6}$$

Following [2, Section 4], we recall that if $m_j > 1$, then every $y \in \mathbb{R}^{m_j}$ has a *spectral decomposition* with respect to $\mathbb{L}_{m_j}$, in the form

$$y = \lambda_1(y)u_1(y) + \lambda_2(y)u_2(y),$$

where

$$\lambda_i(y) := y_0 + (-1)^i\|\widehat{y}\| \quad \text{and} \quad u_i(y) := \begin{cases} \frac{1}{2}\left(1, (-1)^i \frac{\widehat{y}}{\|\widehat{y}\|}\right), & \text{if } \widehat{y} \neq 0, \\ \frac{1}{2}\left(1, (-1)^i w\right), & \text{otherwise}, \end{cases} \tag{7}$$

and $w \in \mathbb{R}^{m_j-1}$ can be any unitary vector, with $i \in \{1, 2\}$. In this setting, $\lambda_i(y)$ is said to be an eigenvalue of y associated with the eigenvector $u_i(y)$, $i \in \{1, 2\}$. By definition, we see that $y \in \mathbb{L}_{m_j}$ if, and only if, $\lambda_1(y) \geq 0$, $\lambda_2(y) \geq 0$, whence follows that the orthogonal projection of y onto $\mathbb{L}_{m_j}$ can be characterized as

$$\mathcal{P}_{\mathbb{L}_{m_j}}(y) = [\lambda_1(y)]_+ u_1(y) + [\lambda_2(y)]_+ u_2(y), \tag{8}$$

where $[\,\cdot\,]_+ := \max\{\,\cdot\,, 0\}$.

Remark 2.2 From this point onwards, we will assume that $m_j > 1$ for every $j \in \{1, \ldots, q\}$. The reason to do this is that if $m_j = 1$, then $g_j \in \mathbb{L}_{m_j}$ is a standard NLP constraint, which should be treated separately in our approach, together with equality constraints; we should remark that our approach is very friendly to this kind of mixed constraints, since it is based on [11]. In particular, inclusion of equality constraints can be done in the way suggested in [34] and [14]. Therefore, to avoid cumbersome notation, we will omit both types of NLP constraints in this paper. Alternatively, the spectral decomposition of $y \in \mathbb{L}_1$ could be interpreted as $y = \lambda_1(y)u_1(y)$, with $u_1(y) = 1$ and $\lambda_1(y) = y$. From this point of view, the definitions and theorems of this paper can be adjusted to fit the case $m_j = 1$ by simply disregarding all expressions involving $\lambda_2(y)$ and $u_2(y)$.

Let $\overline{x} \in \mathcal{F}$. The well-known *Karush–Kuhn–Tucker* (KKT) conditions for $\overline{x}$ consist of the existence of *Lagrange multipliers* $\overline{\mu}_j \in \mathbb{L}_{m_j}$, $j \in \{1, \ldots, q\}$, such that

$$\begin{aligned} \nabla_x L(\overline{x}, \overline{\mu}_1, \ldots, \overline{\mu}_q) = 0, \\ \langle \overline{\mu}_j, g_j(\overline{x}) \rangle = 0, \quad \forall j \in \{1, \ldots, q\}, \end{aligned} \tag{9}$$

where

$$L(x, \mu_1, \ldots, \mu_q) := f(x) - \sum_{j=1}^{q} \langle \mu_j, g_j(x) \rangle.$$

It is known that not every local minimizer satisfies the KKT conditions, unless a constraint qualification is present. The most prominent constraint qualifications in the literature are the nondegeneracy CQ and Robinson's CQ, which we recall next as characterized[1] in the work of Bonnans and Ramírez [18].

Definition 2.2 A point $\overline{x} \in \mathcal{F}$ satisfies

– *Nondegeneracy* if the family

$$\left\{ Dg_j(\overline{x})^\top \Gamma_j g_j(\overline{x}) \right\}_{j \in I_B(\overline{x})} \bigcup \left\{ Dg_j(\overline{x})^\top \right\}_{j \in I_0(\overline{x})} \tag{10}$$

is $\mathbb{R}^{|I_B(\overline{x})|} \times \prod_{j \in I_0(\overline{x})} \mathbb{R}^{m_j}$-linearly independent;

– *Robinson's CQ* if the family (10) is $\mathbb{R}_+^{|I_B(\overline{x})|} \times \prod_{j \in I_0(\overline{x})} \mathbb{L}_{m_j}$-linearly independent; where

$$\Gamma_j := \begin{bmatrix} 1 & 0 \\ 0 & -\mathbb{I}_{m_j - 1} \end{bmatrix} \tag{11}$$

and $\mathbb{I}_{m_j-1}$ is the identity matrix of dimension $m_j - 1$.

As mentioned in the Introduction, the nondegeneracy condition reduces to LICQ from NLP when it is seen as an instance of (NSOCP) with $m_1 = \cdots = m_q = 1$, while Robinson's CQ reduces to MFCQ in the same process.

3 Weak Constraint Qualifications for NSOCP

From the practical point of view, one of the standard strategies for proving first-order global convergence of iterative algorithms is proving that every feasible limit point $\overline{x}$ of the sequence $\{x^k\}_{k \in \mathbb{N}}$ of its iterates fulfills the KKT conditions whenever a given

[1] See [18, Proposition 19] for the characterization of nondegeneracy. The characterization of Robinson's CQ follows from [19, Proposition 2.97 and Corollary 2.98] using the fact $\langle y_j, g_j(\overline{x}) \rangle = 0$ with $j \in I_B(\overline{x})$ if, and only if, $y_j = \alpha \Gamma_j g_j(\overline{x})$ for some $\alpha \geq 0$; and similarly, $\langle y_j, g_j(\overline{x}) \rangle = 0$ with $j \in I_{\text{int}}(\overline{x})$ if, and only if, $y_j = 0$ [2, Lemma 15].

CQ holds. Roughly speaking, this means that the algorithm surely avoids all non-optimal points that satisfy the CQ but violate KKT; hence, building this reasoning under a more general (weaker) CQ means to narrow down the range of convergence of the method without removing optimal candidates from it. Moreover, it is well known that the existence of Lagrange multipliers is a relevant issue beyond algorithms—for example, in situations where they have some practical interpretation, such as in the electricity pricing context—meaning that there is also a theory-driven motivation for pursuing weaker constraint qualifications.

In this section, we will present weaker variants of nondegeneracy and Robinson's CQ, discuss some of their properties, and exemplify their usage with an external penalty method. Besides, these conditions shall pave the way for a more radical relaxation in terms of local constant rank, which will be discussed in the next section. A similar approach has been conducted in [9, 10] for NSDP problems, but although NSOCP can be seen as a particular case of NSDP via an arrowhead matrix transformation

$$(y_0, \widehat{y}) \mapsto \operatorname{Arw}(y_0, \widehat{y}) := \left[\begin{array}{ccc|c} y_0 & & & \\ & \ddots & & \widehat{y} \\ & & y_0 & \\ \hline & \widehat{y}^\top & & y_0 \end{array}\right],$$

it should be noted that constraint qualifications are not necessarily carried over with the transformation; that is, when dealing with weak constraint qualifications, one generally loses information when the problem is equivalently rewritten differently (a noticeable exception is Robinson's CQ, which turns out the be quite robust in this sense). For instance, the nondegeneracy condition for NSDP is never satisfied by a constraint in the form

$$\operatorname{Arw}(g_0(x), \widehat{g}(x)) \in \mathbb{S}^m_+ := \{M \in \mathbb{R}^{m\times m} : M = M^\top;\ \forall d \in \mathbb{R}^m, d^\top M d \geq 0\}$$

when $m > 2$, regardless of the fulfillment of nondegeneracy for NSOCP applied to the constraint $(g_0(x), \widehat{g}(x)) \in \mathbb{L}_m$. As it can be easily verified, the same conclusion holds for the constraint qualification called "weak-nondegeneracy" for NSDP that was introduced in [10]. Thus, a specialized analysis is required to obtain results similar to [9, 10], for NSOCP. In fact, the analysis we present in this section regarding those weak conditions is, in a sense, more refined than the one presented in [10] since there are some important questions that were left open in [10], which we are able to answer here.

3.1 Parametric Bases and Weak-Nondegeneracy for NSOCP

We open our studies by characterizing nondegeneracy and Robinson's CQ in terms of the eigenvectors of the constraint functions (as in (7)). To motivate it, let $g(x) := (g_0(x), \widehat{g}(x))$ and $\overline{x} \in \mathbb{R}^n$ be such that $g(\overline{x}) = 0$. Using Bonnans and Ramírez' characterization (Definition 2.2), we see that $\overline{x}$ is *nondegenerate* (that is, it satisfies nondegeneracy CQ) when the matrix $Dg(\overline{x})$ is surjective. This is clearly a representation of nondegeneracy in the canonical basis $e_1, \ldots, e_m$ of $\mathbb{R}^m$, where e_i has 1 in its ith position and zeros elsewhere. Other representations of $\mathbb{R}^m$ may lead to different characterizations of these constraint qualifications; and this simple fact leads us a natural way of imbuing the structure of the cone into the conditions.

For instance, the discussion after Lemma 2.1 allows us to represent nondegeneracy and Robinson's CQ in terms of each slice of $\mathbb{L}_m$, as long as we consider all of them. More precisely:

Corollary 3.1 *Let $\overline{x}$ be a feasible point of (NSOCP). Then:*

1. *Nondegeneracy holds at $\overline{x}$ if, and only if, the family of vectors*

$$\left\{Dg_j(\overline{x})^\top u_1(g_j(\overline{x}))\right\}_{j\in I_B(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^\top(1,-\overline{w}_j),\ Dg_j(\overline{x})^\top(1,\overline{w}_j)\right\}_{j\in I_0(\overline{x})} \tag{12}$$

   is linearly independent for every $\overline{w}_j \in \mathbb{R}^{m_j-1}$ such that $\|\overline{w}_j\| = 1$, $j \in I_0(\overline{x})$;
2. *Robinson's CQ holds at $\overline{x}$ if, and only if, the family (12) is positively linearly independent for every $\overline{w}_j$ such that $\|\overline{w}_j\| = 1$, $j \in I_0(\overline{x})$.*

Proof For item 2, it suffices to apply Lemma 2.1 considering the product $C = \prod_{j\in J} C_j$, $J := I_B(\overline{x}) \cup I_0(\overline{x})$, where

$$C_j := \begin{cases} \mathbb{R}_+, & \text{if } j \in I_B(\overline{x}),\\ \mathbb{L}_{m_j}, & \text{if } j \in I_0(\overline{x}),\end{cases}$$

to the matrix $M = [M_j]_{j\in J}$ arranged as in (5), whose blocks are given by

$$M_j := \begin{cases} Dg_j(\overline{x})^\top u_1(g_j(\overline{x})), & \text{if } j \in I_B(\overline{x}),\\ Dg_j(\overline{x})^\top, & \text{if } j \in I_0(\overline{x}).\end{cases}$$

To see why C fits the description of Lemma 2.1, define $S_j := \{1\}$ for every $j \in I_B(\overline{x})$, $S_j := \{\overline{w}_j \in \mathbb{R}^{m_j-1} \colon \|\overline{w}_j\| = 1\}$ for every $j \in I_0(\overline{x})$; then, let $S := \prod_{j\in J} S_j$ and for each $\overline{w} := (\overline{w}_j)_{j\in J} \in S$, with $\overline{w}_j \in S_j$ for $j \in J$, define $\mathcal{E}_{\overline{w}} := \prod_{j\in J} \mathcal{E}_{\overline{w}_j}$, where

$$\mathcal{E}_{\overline{w}_j} := \begin{cases} 1, & \text{if } j \in I_B(\overline{x}),\\ \{(1,-\overline{w}_j),(1,\overline{w}_j)\}, & \text{if } j \in I_0(\overline{x}),\end{cases}$$

for every $j \in J$. Observe that $C = \bigcup_{\overline{w}\in S} \operatorname{cone}(\mathcal{E}_{\overline{w}})$ and the proof of item 2 is over. The proof for item 1 is similar, considering Remark 2.1. □

For a better understanding of the meaning of Corollary 3.1, let us resume the short discussion after Lemma 2.1. Note that LICQ for a pair of constraints $g_1(x) \geq 0$ and $g_2(x) \geq 0$ at a point $\overline{x}$ such that $g_1(\overline{x}) = g_2(\overline{x}) = 0$, when seen through Corollary 3.1, becomes equivalent to $Dg(\overline{x})^\top \begin{pmatrix}\cos(w)\\ \sin(w)\end{pmatrix}$ being nonzero, for every $w \in [0, 2\pi]$, where $g := (g_1, g_2)$. On the one hand, this is obvious; but on the other hand, note that the process of checking linear independence of a couple of n-dimensional vectors is reduced to checking whether one n-dimensional vector is zero or not, for each fixed real parameter w. Of course, this reasoning can be extended to arbitrary dimensions and arbitrary parameterizations, and Corollary 3.1 is simply one of these extensions where the parametrization is given in terms of the second-order cone. This will turn out to be relevant in our analysis as we will be able to identify that some of the linear independence requirements will be superfluous for a constraint qualification to be defined. This kind of reasoning can also be applied to the cone of symmetric positive semidefinite matrices, leading to a different, in fact simpler, proof of [10, Proposition 3.2], which is the analog of Corollary 3.1 in the context of NSDP, hence providing some intuition for a result that was originally presented as a mere technical tool in [10].

With the characterization of Corollary 3.1 at hand, we can take a close look at a simple example that shall motivate our next steps:

Example 3.1 Let $g_0, g_1 : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable functions, define $g := (g_0, g_1)$, and let $\overline{x}$ be a point such that:

- $g(\overline{x}) = 0$;
- $\nabla g_0(\overline{x})$ and $\nabla g_1(\overline{x})$ are linearly independent.

Observe that nondegeneracy holds for the constraint $g(x) \in \mathbb{L}_2$ at $\overline{x}$ since $Dg(\overline{x})^\top$ is $\mathbb{R}^2$-linearly independent. Now consider the equivalent NSOCP constraint

$$\tilde{g}(x) := (g_0(x), g_1(x), 0, \dots, 0) \in \mathbb{L}_m$$

and observe that the KKT conditions for it are the same as for the constraint $g(x) \in \mathbb{L}_2$. However, by Corollary 3.1, nondegeneracy for the reformulated problem is equivalent to the linear independence of the vectors

$$D\tilde{g}(\overline{x})^\top(1, -\overline{w}) = \nabla g_0(\overline{x}) - \overline{w}_1 \nabla g_1(\overline{x})$$

and

$$D\tilde{g}(\overline{x})^\top(1, \overline{w}) = \nabla g_0(\overline{x}) + \overline{w}_1 \nabla g_1(\overline{x})$$

for every $\overline{w} = (\overline{w}_1, \dots, \overline{w}_{m-1})$ such that $\|\overline{w}\| = 1$, which is violated when $\overline{w}_1 = 0$.

On the other hand, note that for every x such that $g_1(x) \neq 0$ the eigenvectors of $\tilde{g}(x)$ are uniquely determined by

$$u_1(\tilde{g}(x)) = \frac{1}{2}\left(1, -\frac{g_1(x)}{|g_1(x)|}, 0, \dots, 0\right)$$

and

$$u_2(\tilde{g}(x)) = \frac{1}{2}\left(1, \frac{g_1(x)}{|g_1(x)|}, 0, \dots, 0\right).$$

This suggests that although $\tilde{g}(\overline{x})$ admits multiple eigenvector decompositions $\frac{1}{2}(1, -\overline{w})$ and $\frac{1}{2}(1, \overline{w})$ with $\|\overline{w}\| = 1$, the only relevant ones are$\overline{w} = (\pm 1, 0, \ldots, 0)$. That is, in light of our previous work in NSDP , we can infer that the problematic choices of $\frac{1}{2}(1, -\overline{w})$ and $\frac{1}{2}(1, \overline{w})$ such that $\overline{w}_1 = 0$ may be disregarded when defining a constraint qualification. In fact, we may consider all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and we have that when $g_1(x^k) \neq 0$ for every $k \in \mathbb{N}$, the sequences $\{u_1(\tilde{g}(x^k))\}_{k\in\mathbb{N}}$ and $\{u_2(\tilde{g}(x^k))\}_{k\in\mathbb{N}}$ of eigenvectors of $\tilde{g}(x^k)$ are uniquely defined and $\frac{1}{2}(1, -\overline{w})$ and $\frac{1}{2}(1, \overline{w})$ with $\overline{w}_1 = 0$ are not among their limit points. Similarly, when $g_1(x^k) = 0$ for some indexes $k \in \mathbb{N}$ one may also choose the eigendecompositions of $\tilde{g}(x^k)$ that avoids having $\frac{1}{2}(1, -\overline{w})$ and $\frac{1}{2}(1, \overline{w})$ with $\overline{w}_1 = 0$ as limit points.

Conversely, note that for any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, the choice $\overline{w} = (\pm 1, 0, \ldots, 0)$ does not present the same issue, and in this case we get that the vectors

$$D\tilde{g}(\overline{x})^\top (1, -\overline{w}) = \nabla g_0(\overline{x}) \mp \nabla g_1(\overline{x}) \quad \text{and} \quad D\tilde{g}(\overline{x})^\top (1, \overline{w}) = \nabla g_0(\overline{x}) \pm \nabla g_1(\overline{x})$$

are linearly independent.

Example 3.1 suggests that demanding linear independence of (12) for all $\overline{w}_j$ may be unnecessarily strong for a constraint qualification. In fact, it also suggests that only the limit points of sequences consisting of eigenvectors of $g(x^k)$, for each $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, are needed. This observation leads to two new constraint qualifications for NSOCP:

Definition 3.1 (*Weak-nondegeneracy and weak-Robinson's CQ*) Let $\overline{x} \in \mathcal{F}$. We say that $\overline{x}$ satisfies:

- *Weak-nondegeneracy* if, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $I \subseteq_\infty \mathbb{N}$ and convergent eigenvectors sequences $\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, -\overline{w}_j)$ and $\{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, \overline{w}_j)$, with $\|\overline{w}_j\| = 1$, for every $j \in I_0(\overline{x})$, such that (12) is linearly independent;
- *Weak-Robinson's CQ* if, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $I \subseteq_\infty \mathbb{N}$ and convergent eigenvectors sequences $\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, -\overline{w}_j)$ and $\{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, \overline{w}_j)$, for every $j \in I_0(\overline{x})$, such that (12) is positively linearly independent;

where the notation $I \subseteq_\infty \mathbb{N}$ means that I is an infinite subset of $\mathbb{N}$.

Both conditions presented in Definition 3.1 will be proved to be CQs later on; let us first discuss their properties and relations with other CQs. From Definition 3.1, it is clear that weak-nondegeneracy is implied by nondegeneracy, but the converse is not necessarily true, as illustrated by Example 3.1. Notice also that both conditions from Definition 3.1 are maintained under the addition of structural zeros as in Example 3.1, which somehow shows the robustness of the conditions we define. Similarly, for NSDPs, in [10], it is shown that the analogous conditions from Definition 3.1 are maintained when stacking several semidefinite constraints into a single block diagonal semidefinite constraint. The next example shows, however, that weak-nondegeneracy may hold when nondegeneracy fails even when the problem does not have structural zeros:

Example 3.2 (Weak-nondegeneracy is weaker than nondegeneracy) Consider the constraint

$$g(x) := (x_1, x_2, x_2) \in \mathbb{L}_3$$

at the point $\overline{x} := (0, 0)$, which does not satisfy nondegeneracy. Now, take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$. There are three possible cases to consider:

1. There exists some infinite subset $I \subseteq_\infty \mathbb{N}$ such that $x_2^k > 0$ for all $k \in I$;
2. Case 1 fails to hold, but there exists some infinite subset $I \subseteq_\infty \mathbb{N}$ such that $x_2^k < 0$ for all $k \in I$;
3. Cases 1 and 2 both fail, implying $x_2 = 0$ for all k large enough;

In Case 1, the eigenvectors $u_1(g(x^k))$ and $u_2(g(x^k))$ are uniquely determined by

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \quad \text{and} \quad u_2(g(x^k)) = \frac{1}{2}\left(1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right),$$

for all $k \in I$. Define $\overline{w} := \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and note that

$$\lim_{k\in I} u_1(g(x^k)) = \frac{1}{2}(1, -\overline{w}) \quad \text{and} \quad \lim_{k\in I} u_2(g(x^k)) = \frac{1}{2}(1, \overline{w}).$$

In addition,

$$Dg(\overline{x})^\top(1, -\overline{w}) = \frac{1}{2}\begin{pmatrix} 1 \\ -\sqrt{2} \end{pmatrix} \quad \text{and} \quad Dg(\overline{x})^\top(1, \overline{w}) = \frac{1}{2}\begin{pmatrix} 1 \\ \sqrt{2} \end{pmatrix}$$

are linearly independent. Case 2 is analogous. In Case 3, we have that the eigenvectors of $g(x^k)$ are not uniquely defined in (7); thus, in checking Definition 3.1 we may choose an appropriate eigendecomposition of each $g(x^k)$. In particular, we may pick the same decomposition analyzed previously to conclude that weak-nondegeneracy holds at $\overline{x}$. Notice that since nondegeneracy fails, by Corollary 3.1 there must exist some $\overline{w}$, $\|w\| = 1$, such that $Dg(\overline{x})^\top(1, -\overline{w})$ and $Dg(\overline{x})^\top(1, \overline{w})$ are linearly dependent. This is the case of $\overline{w} := \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$ or $\overline{w} := \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$; however, since weak-nondegeneracy holds, these limit points can be avoided considering the eigendecompositions of $\{g(x^k)\}_{k\in\mathbb{N}}$ for any sequence $x^k \to \overline{x}$.

At this point, we acknowledge that weak-nondegeneracy may be hard to check. However, besides its robustness in terms of structural zeros as discussed in Example 3.1, let us prove that there is a deeper connection between nondegeneracy and weak-nondegeneracy, in the sense that we may characterize nondegeneracy by the validity of weak-nondegeneracy plus a simple linear independence requirement of a partial family of derivative vectors in $I_0(\overline{x})$, namely, by removing from consideration in the family (10) that defines nondegeneracy all gradients of first component entries, that is, $\nabla g_{j,0}(\overline{x})$, $j \in I_0(\overline{x})$ together with the vectors indexed by $I_B(\overline{x})$. In fact, in Example 3.2, this family of vectors reduces to the rows of $D\widehat{g}(\overline{x})$, where $\widehat{g}(x) := (x_2, x_2)$, which are linearly dependent. Loosely speaking, weak-nondegeneracy may be thought as an appropriate form of nondegeneracy but without requiring linear independence of this partial family of vectors.

(Difference between weak-nondegeneracy and Nondegeneracy) Let $\overline{x}$ be a feasible point of (NSOCP). We have that nondegeneracy holds at $\overline{x}$ if, and only if, weak-nondegeneracy holds at $\overline{x}$ and, in addition, the matrix

$$M := \begin{bmatrix} \vdots \\ D\widehat{g}_j(\overline{x}) \\ \vdots \end{bmatrix}_{j \in I_0(\overline{x})}$$

is surjective.

Proof From Definition 3.1, it is clear that if nondegeneracy holds at $\overline{x}$, then weak-nondegeneracy also holds at $\overline{x}$. Moreover, from (10) we obtain that M is surjective. Conversely, suppose that nondegeneracy does not hold at $\overline{x}$. By Corollary 3.1, there are unitary vectors $\overline{w}_j \in \mathbb{R}^{m_j-1}$, $j \in I_0(\overline{x})$, such that (12) is linearly dependent.

Let us define $\overline{w} = (\overline{w}_j)_{j \in I_0(\overline{x})}$. By the surjectivity of M, there exists a nonzero vector $d \in \mathbb{R}^n$ such that $\overline{w} = Md$. That is, we have that $D\widehat{g}_j(\overline{x})d = \overline{w}_j$ for all $j \in I_0(\overline{x})$. Now, take any positive sequence $\{t_k\}_{k\in\mathbb{N}} \to 0^+$ and let

$$x^k := \overline{x} + t_k d, \ \forall k \in \mathbb{N}.$$

We have that $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and when we consider $j \in I_0(\overline{x})$ and the Taylor expansion of $\widehat{g}_j(x^k)$ around $\overline{x}$, we obtain that

$$\widehat{g}_j(x^k) = t_k \overline{w}_j + o(t_k) \neq 0$$

for all $k \in \mathbb{N}$ large enough, since $\overline{w}_j \neq 0$. Moreover, for the indices $j \in I_B(\overline{x})$ we also have that $\widehat{g}_j(x^k) \neq 0$ for all k large enough, because $\widehat{g}_j(\overline{x}) \neq 0$. This means that the eigenvectors of $\widehat{g}_j(x^k)$ are uniquely determined from (7) for all $j \in I_0(\overline{x}) \cup I_B(\overline{x})$ and all $k \in \mathbb{N}$. In particular, for $j \in I_0(\overline{x})$ we have that

$$\frac{\widehat{g}_j(x^k)}{\|\widehat{g}_j(x^k)\|} = \frac{D\widehat{g}_j(\overline{x})d + o(t_k)/t_k}{\|D\widehat{g}_j(\overline{x})d + o(t_k)/t_k\|} \to \overline{w}_j.$$

As a consequence, since $\overline{w}_j \in \mathbb{R}^{m_j-1}$, $j \in I_0(\overline{x})$, is such that (12) is linearly dependent, we conclude that weak-nondegeneracy does not hold at $\overline{x}$. □

The following example shows that although weak-nondegeneracy implies weak-Robinson's CQ, the converse is not true:

Example 3.3 (Weak-Robinson is weaker than weak-nondegeneracy) Consider the constraint

$$g(x) := (4x, 2x, x) \in \mathbb{L}_3$$

and the point $\overline{x} := 0$. Clearly, it satisfies Robinson's CQ, hence it also satisfies weak-Robinson's CQ. However, observe that taking any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that $x^k > 0$ for all $k \in \mathbb{N}$, we have

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right) \quad \text{and} \quad u_2(g(x^k)) = \frac{1}{2}\left(1, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right),$$

hence we have $u_1(g(x^k)) \to \frac{1}{2}(1, -\overline{w})$ and $u_2(g(x^k)) \to \frac{1}{2}(1, \overline{w})$ where $\overline{w} = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)$. Then,

$$Dg(\overline{x})^\top(1, -\overline{w}) = \frac{4\sqrt{5}-5}{2\sqrt{5}} > 0 \quad \text{and} \quad Dg(\overline{x})^\top(1, \overline{w}) = \frac{4\sqrt{5}+5}{2\sqrt{5}} > 0$$

are linearly dependent, although positively linearly independent, implying that weak-nondegeneracy does not hold at $\overline{x}$.

To discuss in detail the relation between weak-Robinson's CQ and Robinson's CQ for (NSOCP), we rely on a simple lemma:

Lemma 3.1 *Let $\overline{x}$ be a feasible point of (NSOCP). If (weak-Robinson's CQ) weak-nondegeneracy holds at $\overline{x}$, then the family of vectors*

$$\left\{\nabla g_{j,0}(\overline{x})\right\}_{j\in I_0(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^\top u_1(g_j(\overline{x}))\right\}_{j\in I_B(\overline{x})} \tag{13}$$

is (positively) linearly independent.

Proof Assume that weak-Robinson's CQ holds at $\overline{x}$, so there exists some vectors $\overline{w}_j \in \mathbb{R}^{m_j-1}$, $\|\overline{w}_j\| = 1$, $j \in I_0(\overline{x})$, such that (12) is positively linearly independent; and, by contradiction, suppose that (13) is positively linearly dependent. Then, there are some $\eta_j \geq 0$, $j \in I_B(\overline{x}) \cup I_0(\overline{x})$, not all zero, such that

$$\sum_{j\in I_0(\overline{x})} \eta_j \nabla g_{j,0}(\overline{x}) + \sum_{j\in I_B(\overline{x})} \eta_j Dg_j(\overline{x})^\top u_1(g_j(\overline{x})) = 0. \tag{14}$$

Now set

$$\alpha_j = \beta_j = \frac{\eta_j}{2}$$

for every $j \in I_0(\overline{x})$ and (14) can be rewritten as

$$\sum_{j\in I_0(\overline{x})} \alpha_j Dg_j(\overline{x})^\top(1, -\overline{w}_j) + \sum_{j\in I_0(\overline{x})} \beta_j Dg_j(\overline{x})^\top(1, \overline{w}_j) + \sum_{j\in I_B(\overline{x})} \eta_j Dg_j(x)^\top u_1(g_j(\overline{x})) = 0,$$

which implies (12) is positively linearly dependent, contradicting weak-Robinson's CQ. The statement regarding weak-nondegeneracy follows analogously. □

Recall that Robinson's CQ can be evaluated separately for each of the constraints $g_j(x) \in \mathbb{L}_{m_j}$, $j \in \{1, \dots, q\}$, and that this is weaker than Robinson's CQ when such system is regarded as a whole (however, not being a CQ). In fact, for any given $\overline{x} \in \mathcal{F}$, the former can be characterized by the existence of some vectors $d_j \in \mathbb{R}^n$, $j \in \{1, \dots, q\}$, such that $g_j(\overline{x}) + Dg_j(\overline{x})d_j \in \operatorname{int}\mathbb{L}_{m_j}$, whereas the latter requires in addition $d_1 = d_2 = \cdots = d_q$ to hold. With this in mind, we prove next that weak-Robinson's CQ is somewhat in-between these two forms of Robinson's CQ.

Theorem 3.1 *Consider Problem (NSOCP) and let $\overline{x} \in \mathcal{F}$. If weak-Robinson's CQ holds at $\overline{x}$, then for each index $j \in \{1, \dots, q\}$ the point $\overline{x}$ satisfies Robinson's CQ for the isolated constraint $g_j(x) \in \mathbb{L}_{m_j}$.*

Proof Let $\overline{x} \in \mathcal{F}$ be a point such that weak-Robinson's CQ holds and assume that there exists an index $\ell \in \{1, \dots, q\}$ such that Robinson's CQ does not hold. Then, it follows by Lemma 3.1 that $g_\ell(\overline{x}) = 0$. So there exists some $\overline{w}_\ell \in \mathbb{R}^{m_\ell - 1}$ such that $\|\overline{w}_\ell\| = 1$ and the vectors $Dg_\ell(\overline{x})^\top(1, -\overline{w}_\ell)$ and $Dg_\ell(\overline{x})^\top(1, \overline{w}_\ell)$ are positively linearly dependent, that is, there exist scalars $\alpha \geq 0$, $\beta \geq 0$, at least one of them nonzero, such that

$$\alpha Dg_\ell(\overline{x})^\top(1, -\overline{w}_\ell) + \beta Dg_\ell(\overline{x})^\top(1, \overline{w}_\ell) = 0.$$

Defining $\tilde{w} := \left(\frac{\beta-\alpha}{\alpha+\beta}\right)\overline{w}_\ell$, it follows that

$$\nabla g_{\ell,0}(\overline{x}) = -D\widehat{g}_\ell(\overline{x})^\top \tilde{w}. \tag{15}$$

Note that $\|\tilde{w}\| \leq 1$, and that $\tilde{w} \notin \operatorname{Ker} D\widehat{g}_\ell(\overline{x})^\top$; otherwise, $\nabla g_{\ell,0}(\overline{x}) = 0$ and according to Lemma 3.1 weak-Robinson's CQ fails.

Since $\operatorname{Ker} D\widehat{g}_\ell(\overline{x})^\top + \operatorname{Im} D\widehat{g}_\ell(\overline{x}) = \mathbb{R}^{m_\ell - 1}$, there exist some $v \in \operatorname{Ker} D\widehat{g}_\ell(\overline{x})^\top$ and some $d \in \mathbb{R}^n$ such that $\tilde{w} = v + D\widehat{g}_\ell(\overline{x})d$. Note that $D\widehat{g}_\ell(\overline{x})d \neq 0$, otherwise we would have that $\tilde{w} \in \operatorname{Ker} D\widehat{g}_\ell(\overline{x})^\top$. In addition, $0 \neq \tilde{w} - v = \mathcal{P}_{\operatorname{Im} D\widehat{g}_\ell(\overline{x})}(\tilde{w})$ and by the non-expansiveness of the projection, we obtain $0 < \|\tilde{w} - v\| \leq \|\tilde{w}\| \leq 1$.

Now, proceeding similarly to the proof of Proposition 3.1, consider the sequence $\{x^k\}_{k\in\mathbb{N}}$ given by $x^k := \overline{x} + t_k d$, for any positive scalars sequence $\{t_k\}_{k\in\mathbb{N}} \to 0^+$, and consider the Taylor expansion of $\widehat{g}_\ell(x^k)$ around $\overline{x}$:

$$\widehat{g}_\ell(x^k) = t_k D\widehat{g}_\ell(\overline{x})d + o(t_k).$$

Since $D\widehat{g}_\ell(\overline{x})d \neq 0$, it follows that there exists some $k_0 \in \mathbb{N}$ such that $\widehat{g}_\ell(x^k) \neq 0$ for every $k > k_0$, which implies that its eigenvectors, and $u_2(g_\ell(x^k))$ $u_1(g_\ell(x^k))$ and $u_2(g_\ell(x^k))$, are uniquely determined from (7) for every $k > k_0$. Then, we obtain that

$$\frac{\widehat{g}_\ell(x^k)}{\|\widehat{g}_\ell(x^k)\|} = \frac{D\widehat{g}_\ell(\overline{x})d + o(t_k)/t_k}{\|D\widehat{g}_\ell(\overline{x})d + o(t_k)/t_k\|} \to \frac{\tilde{w} - v}{\|\tilde{w} - v\|}.$$

It follows that

$$\lim_{k\to\infty} u_1(g_\ell(x^k)) = \frac{1}{2}\left(1, -\frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)$$

and

$$\lim_{k\to\infty} u_2(g_\ell(x^k)) = \frac{1}{2}\left(1, \frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)$$

and, by weak-Robinson's CQ, the vectors and $Dg_\ell(\bar{x})^\top\left(1, \frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)$ $Dg_\ell(\bar{x})^\top \left(1, -\frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)$ and $Dg_\ell(\bar{x})^\top\left(1, \frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)$ are positively linearly independent. However, the following system in the variables a and b:

$$\begin{aligned}
0 &= aDg_\ell(\bar{x})^\top\left(1, \frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right) + bDg_\ell(\bar{x})^\top\left(1, -\frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)\\
&= a\nabla g_{\ell,0}(\bar{x}) + \frac{a}{\|\tilde{w}-v\|} D\widehat{g}_\ell(\bar{x})^\top\tilde{w} + b\nabla g_{\ell,0}(\bar{x}) - \frac{b}{\|\tilde{w}-v\|} D\widehat{g}_\ell(\bar{x})^\top\tilde{w}\\
&= \left[a\left(\frac{1}{\|\tilde{w}-v\|} - 1\right) - b\left(\frac{1}{\|\tilde{w}-v\|} + 1\right)\right] D\widehat{g}_\ell(\bar{x})^\top\tilde{w}
\end{aligned}$$

has a nontrivial solution $a = 1/\|\tilde{w}-v\| + 1 > 0$ and $b = 1/\|\tilde{w}-v\| - 1 \geq 0$, which is a contradiction. In the second equality of the above chain, we used $D\widehat{g}_\ell(\bar{x})^\top v = 0$; and in the last equality, we used (15). □

Remark 3.1 The same strategy of the previous proof actually allows proving a slightly stronger result: If a feasible point $\bar{x}$ satisfies weak-Robinson's CQ, then for each index $j \in I_0(\bar{x})$ the constraint

$$g_\ell(x) \in \mathbb{L}_{m_\ell}, \ \forall \ell \in I_B(\bar{x}) \cup \{j\}$$

satisfies Robinson's CQ at $\bar{x}$. In particular, if $I_0(\bar{x})$ is a singleton, then weak-Robinson's CQ and Robinson's CQ are equivalent, which is somewhat remarkable and highlights the "robustness" of Robinson's CQ. The situation where $I_0(\bar{x})$ is a singleton has been previously considered, for instance, in [36, 40]. In the general case, we were not able to prove nor provide a counterexample for the equivalence between Robinson's CQ and weak-Robinson's CQ.

4 Constant Rank Conditions for NSOCP

Let us consider an NLP problem for a moment; that is, (NSOCP) with $m_1 = \cdots = m_q = 1$, whose constraints take the form $g_1(x) \geq 0, \ldots, g_q(x) \geq 0$, and let $\bar{x} \in \mathcal{F}$. We recall that the nondegeneracy condition in this case is equivalent to LICQ, which holds when the family of vectors

$$\left\{\nabla g_j(\overline{x})\right\}_{j\in I_0(\overline{x})} \tag{16}$$

has full rank. The constant rank constraint qualification (CRCQ) condition can be considered a relaxation of LICQ, since it allows the rank of (16) to be incomplete, as long as the rank of the family

$$\left\{\nabla g_j(x)\right\}_{j\in J_0} \tag{17}$$

remains constant in a neighborhood of $\overline{x}$, for every subset $J_0 \subseteq I_0(\overline{x})$. Qi and Wei [42] described CRCQ in a slightly different but equivalent way: CRCQ holds at $\overline{x}$ if, for every $J_0 \subseteq I_0(\overline{x})$, if (17) is linearly dependent at $\overline{x}$, then it must also remain linearly dependent for every x in a neighborhood of $\overline{x}$. Similarly, Robinson's CQ is equivalent to the positive linear independence of (16), and the relaxation of it in the same style as CRCQ characterizes the constraint qualification known as constant positive linear dependence (CPLD) [16]. That is, CPLD holds at $\overline{x}$ if, for every subset $J_0 \subseteq I_0(\overline{x})$, if (17) is positively linearly dependent at $\overline{x}$, then it must remain linearly dependent for every x in a neighborhood of $\overline{x}$.

Extending such constant rank-type constraint qualifications to the context of NSOCP with an arbitrary dimension is not trivial. For instance, it is known that linear second-order cone programming problems may present a positive or infinite duality gap even when the primal problem is bounded, feasible and its solution is attained. This means that "constraint linearity" is not a constraint qualification in NSOCP, contrary to NLP. However, note that any kind of constant rank condition that depends solely on the derivatives of the constraint functions will always be satisfied for every linear problem, implying it cannot be a constraint qualification—see, for instance, [6]. See also [12, Section 2.1] for a detailed discussion on this issue regarding linear problems.

In a previous work, we noticed that weak-nondegeneracy imbues the cone structure into the constraint functions, allowing us to properly define a constant rank-type condition that is not retained by the linearity bottleneck. In this section, we shall follow a similar approach, making the necessary adaptations to overcome the difficulties that arise from the particularities of the second-order cone along the way.

4.1 Weak Constant Rank Conditions

With the definitions of weak-nondegeneracy and weak-Robinson's CQ for NSOCP at hand, we can present new extensions of CRCQ and CPLD for NSOCP by means of a simple relaxation of Definition 3.1, in the same lines as in NLP. Basically, the idea is to demand every subfamily of (12) to locally retain its (positive) linear dependence. So let us define, for any sets $J_B, J_-, J_+ \subseteq \{1, \ldots, q\}$ such that $\widehat{g}_j(x) \neq 0$ for every

$j \in J_B$, the family of vectors

$$\mathcal{D}_{J_B,J_-,J_+}(x,w) := \left\{Dg_j(x)^\top u_1(g_j(x))\right\}_{j\in J_B} \bigcup \left\{Dg_j(x)^\top(1,-w_j)\right\}_{j\in J_-} \bigcup \left\{Dg_j(x)^\top(1,w_j)\right\}_{j\in J_+} \tag{18}$$

where $w = [w_j]_{j\in J_-\cup J_+}$. Above, the index set J_B refers to an arbitrary subset of $I_B(\overline{x})$, and the indices J_- and J_+ both refer to $I_0(\overline{x})$, but with distinct eigenvectors; see (12).

Definition 4.1 (weak-CRCQ and weak-CPLD) We say that a feasible point $\overline{x}$ of (NSOCP) satisfies the:

- *Weak constant rank constraint qualification* (weak-CRCQ) if the following holds: For every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $I \subseteq_\infty \mathbb{N}$, and convergent eigenvector sequences

$$\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,-\overline{w}_j) \quad \text{and} \quad \{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,\overline{w}_j),$$

  with $\|\overline{w}_j\| = 1$, for all $j \in I_0(\overline{x})$, such that for all subsets $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$, we have that if the family of vectors $\mathcal{D}_{J_B,J_-,J_+}(\overline{x},\overline{w})$ is linearly dependent, then $\mathcal{D}_{J_B,J_-,J_+}(x^k,w^k)$ remains linearly dependent for all $k \in I$ large enough, where $\overline{w} = [\overline{w}_j]_{j\in J_-\cup J_+}$ and $w^k = [w^k_j]_{j\in J_-\cup J_+}$ satisfies

$$u_1(g_j(x^k)) = \frac{1}{2}(1,-w^k_j) \quad \text{and} \quad u_2(g_j(x^k)) = \frac{1}{2}(1,w^k_j) \tag{19}$$

  for each $j \in J_- \cup J_+$.
- *Weak constant positive linear dependence* (weak-CPLD) condition if the following holds: For every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there is some $I \subseteq_\infty \mathbb{N}$, and convergent eigenvector sequences

$$\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,-\overline{w}_j) \quad \text{and} \quad \{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,\overline{w}_j),$$

  with $\|\overline{w}_j\| = 1$, for all $j \in I_0(\overline{x})$, such that for all subsets $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$, we have that, if $\mathcal{D}_{J_B,J_-,J_+}(\overline{x},\overline{w})$ is positively linearly dependent, then $\mathcal{D}_{J_B,J_-,J_+}(x^k,w^k)$ is linearly dependent for all $k \in I$ large enough, where $\overline{w}$ and w^k are as in the previous item.

There are some features about Definition 4.1 that should be highlighted for a better understanding of it. First, weak-CRCQ fully recovers CRCQ when we set $m_j = 1$ for every $j \in \{1,\ldots,q\}$—see also Remark 2.2 for a clarification about the case $m_j = 1$. Similarly, note that weak-CPLD recovers CPLD in the same setting. Second, in view of Corollary 3.1, we see that weak-CRCQ is implied by (weak-)nondegeneracy as in Definition 3.1, and weak-CPLD is implied by both (weak-)Robinson's CQ and weak-CRCQ. However, due to such equivalence in NLP, those implications in the conic setting are strict (see Example 4.2 and [16, Counterexample 4.2], respectively). Third, we point out that weak-CRCQ is not comparable with (weak-)Robinson's CQ (see, for instance, [29, Examples 2.1 and 2.2]).

Remark 4.1 To fix ideas, let us consider a single conic constraint $g(x) \in \mathbb{L}_m$ at the point $\overline{x} \in \mathcal{F}$. First, suppose that $g(\overline{x}) = 0$ and take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$. We consider a partition of $\mathbb{N}$ as follows:

- $\mathcal{N}_0 := \{k \in \mathbb{N} : \widehat{g}(x^k) = 0\}$. For $k \in \mathcal{N}_0$, we can choose

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -w^k\right) \quad \text{and} \quad u_2(g(x^k)) = \frac{1}{2}\left(1, w^k\right),$$

  for any w^k such that $\|w^k\| = 1$. When $\mathcal{N}_0$ is infinite, weak-CRCQ demands, in particular, the existence of a choice of $\{w_k\}_{k\in\mathcal{N}_0}$ with some convergent subsequence $\{w^k\}_{k\in I} \to \overline{w}$, $I \subseteq_\infty \mathcal{N}_0$, such that

$$Dg(\overline{x})^\top (1, (-1)^i \overline{w}) = 0$$

  only if

$$Dg(x^k)^\top \left(1, (-1)^i w^k\right) = 0$$

  for all large $k \in I$, $i \in \{1, 2\}$; and, in addition, if $Dg(\overline{x})^\top(1, -\overline{w})$ and $Dg(\overline{x})^\top(1, \overline{w})$ are linearly dependent, then $Dg(x^k)^\top\left(1, -w^k\right)$ and $Dg(x^k)^\top\left(1, w^k\right)$ must also be linearly dependent, for every sufficiently large $k \in I$.
- $\mathcal{N}_1 := \{k \in \mathbb{N} : \widehat{g}(x^k) \neq 0\}$. This case is similar to the previous one, except that there is no freedom in the choice of w^k, as it is uniquely determined by $w^k = \widehat{g}(x^k)/\|\widehat{g}(x^k)\|$, for every $k \in \mathcal{N}_1$.

The reason why both eigenvectors are taken into consideration is that both eigenvalues of $g(\overline{x})$ are zero, in this case. Naturally, in case $g(\overline{x}) \in \mathrm{bd}_+\mathbb{L}_m$, we have only one zero eigenvalue, which is $\lambda_1(g(\overline{x}))$, then weak-CRCQ simply demands the vector

$$Dg(x)^\top u_1(g(x)) = \frac{1}{2} Dg(x)^\top \left(1, -\frac{\widehat{g}(x)}{\|\widehat{g}(x)\|}\right)$$

to be either nonzero at $\overline{x}$ or equal to zero in a whole neighborhood of $\overline{x}$. Note that this coincides with the naive approach [11], obtained by reducing the problem to an NLP. This observation remains true for more than one conic constraint as long as $I_0(\overline{x}) = \emptyset$. See also Remark 4.2.

Now, let us check how Definition 4.1 behaves when it is applied to example [6, Equation 2], which was used to refute the CRCQ proposal of [47].

Example 4.1 (Equation 2 from [6]) Consider the problem

$$\begin{aligned} \underset{x\in\mathbb{R}}{\text{Minimize}} \quad & -x, \\ \text{subject to} \quad & g(x) := (x, x + x^2) \in \mathbb{L}_2. \end{aligned} \tag{20}$$

and its unique feasible point $\overline{x} := 0$, which does not satisfy the KKT conditions. Our aim is to show that Definition 4.1 is not satisfied at $\overline{x}$. To do so, it suffices to take any sequence $\{x^k\}_{k\in\mathbb{N}} \to 0$ such that $x^k > 0$ for all $k \in \mathbb{N}$. In this case, for each $k \in \mathbb{N}$, the eigenvectors of $g(x^k)$ are uniquely determined by

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -\frac{x^k + (x^k)^2}{|x^k + (x^k)^2|}\right) = \frac{1}{2}(1, -1)$$

and

$$u_2(g(x^k)) = \frac{1}{2}\left(1, \frac{x^k + (x^k)^2}{|x^k + (x^k)^2|}\right) = \frac{1}{2}(1, 1),$$

so there is only one trivial limit point for each eigenvector sequence; also, $w^k = \overline{w} = 1$ for every $k \in \mathbb{N}$. However, note that

$$Dg(\overline{x})^\top(1, -\overline{w}) = 0 \quad \text{but} \quad Dg(x^k)^\top(1, -w^k) = -2x^k,$$

so for $J_B := I_B(\overline{x}) = \emptyset$, $J_- := \{1\}$, and $J_+ := \emptyset$, we have $\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k) = \{-2x^k\}$ is linearly independent for every $k \in \mathbb{N}$ whereas $\mathcal{D}_{J_B,J_-,J_+}(\overline{x}, \overline{w}) = \{0\}$ is (positively) linearly dependent. Thus, neither weak-CRCQ nor weak-CPLD is satisfied at $\overline{x}$.

As mentioned before, weak-nondegeneracy and weak-Robinson's CQ are strictly stronger than weak-CRCQ and weak-CPLD, respectively. It is clear that the former implies the latter, so let us prove the "strict" statement:

Example 4.2 (Weak-CRCQ is weaker than weak-nondegeneracy and does not imply weak-Robinson) Consider the constraint

$$g(x) := (-x, x, x) \in \mathbb{L}_3,$$

and its unique feasible point $\overline{x} := 0$. To prove that weak-CPLD holds at $\overline{x}$, let $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ be any sequence. Just as in Example 3.2, there are three cases to be considered, but it suffices to analyze one of them, since the other cases follow analogously. Then, for simplicity, we assume that there is some $I \subseteq_\infty \mathbb{N}$ such that $x^k > 0$ for every $k \in I$, and in this case the eigenvectors of $g(x^k)$ are uniquely determined by

$$u_1(g(x^k)) = \frac{1}{2}\left(1, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \quad \text{and} \quad u_2(g(x^k)) = \frac{1}{2}\left(1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right),$$

leading to $w^k = \overline{w} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. Then,

$$Dg(x^k)^\top(1, (-1)^i w^k) = Dg(\overline{x})^\top(1, -\overline{w}) = \left(-1 + (-1)^i \frac{2}{\sqrt{2}}\right)$$

for each $i \in \{1, 2\}$. Then, the family (12) will have opposite signs, making it positively linearly dependent, so weak-Robinson's CQ and weak-nondegeneracy both fail at $\overline{x}$, without violating the weak-CRCQ and weak-CPLD requirements since in this example

$$\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k) = \mathcal{D}_{J_B,J_-,J_+}(\overline{x}, \overline{w})$$

for every $k \in I$ regardless of J_B, J_-, and J_+.

Example 4.2 can also be used to verify that weak-CRCQ does not imply Robinson's CQ. In fact, Robinson's CQ does not imply weak-CRCQ either, making them independent. Let us show this with another example:

Example 4.3 (Weak-Robinson does not imply weak-CRCQ) Consider the constraint

$$g(x) := (2x_1, x_2^2) \in \mathbb{L}_2$$

at $\overline{x} := 0$. To see that $\overline{x}$ violates weak-CRCQ, it is enough to take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that $x_2^k \neq 0$ for every $k \in \mathbb{N}$. Then, the eigenvectors of $g(x^k)$ must be

$$u_1(g(x^k)) = \frac{1}{2}(1, -1) \quad \text{and} \quad u_2(g(x^k)) = \frac{1}{2}(1, 1),$$

which are defined by $w^k = \overline{w} = 1$ for all $k \in \mathbb{N}$. This implies that the vectors $Dg(x^k)^\top(1, -w^k) = (1, -2x_2^k)$ and $Dg(x^k)^\top(1, w^k) = (1, 2x_2^k)$ are linearly independent for all k, whereas the vectors $Dg(\overline{x})^\top(1, -\overline{w}) = (1, 0)$ and $Dg(\overline{x})^\top(1, \overline{w}) = (1, 0)$ are linearly dependent, violating weak-CRCQ.

On the other hand, in view of Corollary 3.1, it is easy to check that Robinson's CQ holds at $\overline{x}$, since $Dg(\overline{x})^\top(1, -\overline{w}) = (1, 0)$ and $Dg(\overline{x})^\top(1, \overline{w}) = (1, 0)$ are positively linearly independent for every $\overline{w} \in \mathbb{R}$ with $|\overline{w}| = 1$.

Finally, we shall prove that weak-CPLD (and by consequence weak-CRCQ, weak-nondegeneracy, and weak-Robinson's CQ) is a constraint qualification for (NSOCP) employing a result from [7], regarding the output sequences of an external penalty method:

Theorem 4.1 *Let $\overline{x}$ be a local minimizer of (NSOCP), and let $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$. Then, there exists some sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, such that for each $k \in \mathbb{N}$, x^k is a local minimizer of the regularized penalized function*

$$f(x) + \frac{1}{2}\|x - \overline{x}\|_2^2 + \frac{\rho_k}{2}\left(\sum_{j=1}^{q} \|\mathcal{P}_{\mathbb{L}_{m_j}}(-g_j(x))\|^2\right). \tag{21}$$

Proof The proof of this theorem is contained in the proof of [7, Theorem 3.1]. □

Observe that the gradient of (21) can be computed as

$$\nabla_x L\left(x, \rho_k \mathcal{P}_{\mathbb{L}_{m_1}}(-g_1(x)), \ldots, \rho_k \mathcal{P}_{\mathbb{L}_{m_q}}(-g_q(x))\right) + (x - \overline{x}),$$

for each $k \in \mathbb{N}$, which vanish at $x := x^k$. So defining $\mu_j^k := \rho_k \mathcal{P}_{\mathbb{L}_{m_j}}(-g_j(x^k))$, for all $j \in \{1, \ldots, q\}$, induces approximate Lagrange multiplier sequences associated with $\{x^k\}_{k\in\mathbb{N}}$—see also [7]. Then, to prove that weak-CPLD is a CQ, it suffices to construct bounded approximate multiplier sequences out of $\{\mu_j^k\}_{k\in\mathbb{N}}$. For convenience, we will prove a slightly more general result that also encompasses the convergence theory of an external penalty method under weak-CPLD; see [7] for details.

Theorem 4.2 (Weak-Robinson, weak-CRCQ and weak-CPLD are constraint qualifications) *Let* $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ *and* $\{x^k\}_{k\in\mathbb{N}} \to \overline{x} \in \mathcal{F}$ *be such that*

$$\nabla_x L\left(x^k, \rho_k \mathcal{P}_{\mathbb{L}_{m_1}}(-g_1(x^k)), \ldots, \rho_k \mathcal{P}_{\mathbb{L}_{m_q}}(-g_q(x^k))\right) \to 0,$$

and suppose that weak-CPLD holds at $\overline{x}$. *Then,* $\overline{x}$ *satisfies the KKT conditions. Moreover, any local minimizer of (NSOCP) that satisfies weak-CPLD is a KKT point.*

Proof For each $k \in \mathbb{N}$ and $j \in \{1, \ldots, q\}$, define $\mu_j^k := \rho_k \mathcal{P}_{\mathbb{L}_{m_j}}(-g_j(x^k))$. Then, we have

$$\nabla f(x^k) - \sum_{j=1}^{q} Dg_j(x^k)^\top \mu_j^k \to 0. \tag{22}$$

Let us consider an arbitrary spectral decomposition of μ_j^k:

$$\mu_j^k = \alpha_j^k u_1(g_j(x^k)) + \beta_j^k u_2(g_j(x^k)),$$

where $\alpha_j^k = [-\rho_k \lambda_1(g_j(x^k))]_+ \geq 0$ and $\beta_j^k = [-\rho_k \lambda_2(g_j(x^k))]_+ \geq 0$. See (8). Define

$$\begin{aligned} \Psi^k := & \sum_{j\in I_B(\overline{x})\cup I_0(\overline{x})} \alpha_j^k Dg_j(x^k)^\top u_1(g_j(x^k)) + \\ & + \sum_{j\in I_0(\overline{x})} \beta_j^k Dg_j(x^k)^\top u_2(g_j(x^k)) \end{aligned} \tag{23}$$

and note that (22) can be equivalently stated as $\nabla f(x^k) - \Psi^k \to 0$.

By Carathéodory's Lemma 2.2, for each $k \in \mathbb{N}$, there exists some $J_B^k \subseteq I_B(\overline{x})$ and $J_-^k, J_+^k \subseteq I_0(\overline{x})$ such that

$$\left\{Dg_j(x^k)^\top u_1(g_j(x^k))\right\}_{j\in J_B^k\cup J_-^k} \bigcup \left\{Dg_j(x^k)^\top u_2(g_j(x^k))\right\}_{j\in J_+^k} \tag{24}$$

is linearly independent and

$$\Psi^k = \sum_{j\in J_B^k\cup J_-^k} \tilde{\alpha}_j^k Dg_j(x^k)^\top u_1(g_j(x^k)) + \sum_{j\in J_+^k} \tilde{\beta}_j^k Dg_j(x^k)^\top u_2(g_j(x^k)),$$

for some new scalars $\tilde{\alpha}_j^k \geq 0$, $j \in J_B^k \cup J_-^k$, and $\tilde{\beta}_j^k \geq 0$, $j \in J_+^k$. By the infinite pigeonhole principle, we can take a subsequence if necessary such that J_B^k, J_-^k, and J_+^k do not depend on k; that is, we can assume, without loss of generality, that $J_B^k = J_B$, $J_-^k = J_-$, and $J_+^k = J_+$, for every $k \in \mathbb{N}$.

We claim that the sequences $\{\tilde{\alpha}_j^k\}_{k\in\mathbb{N}}$ are bounded for every $j \in J_B \cup J_-$, as well as $\{\tilde{\beta}_j^k\}_{k\in\mathbb{N}}$ for every $j \in J_+$. Indeed, by contradiction, suppose that the sequence $\{m^k\}_{k\in\mathbb{N}}$, given by

$$m^k := \max\{\max\{\tilde{\alpha}_j^k \colon j \in J_B \cup J_-\},\ \max\{\tilde{\beta}_j^k \colon j \in J_+\}\},$$

diverges. Dividing (22) by m^k, we obtain

$$\sum_{j\in J_B\cup J_-} \frac{\tilde{\alpha}_j^k}{m^k} Dg_j(x^k)^\top u_1(g_j(x^k)) + \sum_{j\in J_+} \frac{\tilde{\beta}_j^k}{m^k} Dg_j(x^k)^\top u_2(g_j(x^k)) \to 0$$

and since the sequences $\{\tilde{\alpha}_j^k/m^k\}_{k\in\mathbb{N}}$ are bounded, we can assume, without loss of generality, that they converge to, say, $\overline{\alpha}_j \geq 0$, for all $j \in J_B \cup J_-$; and, similarly, we can also assume that the sequences $\{\tilde{\beta}_j^k/m^k\}_{k\in\mathbb{N}}$ converge to some $\overline{\beta}_j \geq 0$, for all $j \in J_+$. Note that at least one element of $\{\overline{\alpha}_j\}_{j\in J_B\cup J_-} \cup \{\overline{\beta}_j\}_{j\in J_+}$ is nonzero, which makes the correspondent set $\mathcal{D}_{J_B,J_-,J_+}(\overline{x},\overline{w})$ as in Definition 4.1 linearly dependent for any limit point $\overline{w}$ of any subsequence of $\{w^k\}_{k\in\mathbb{N}}$, contradicting weak-CPLD since $\mathcal{D}_{J_B,J_-,J_+}(x^k,w^k)$, which coincides with (24) with w^k defined as in (19), is linearly independent for every $k \in \mathbb{N}$.

Since $\{\tilde{\alpha}_j^k\}_{k\in\mathbb{N}}$ and $\{\tilde{\beta}_j^k\}_{k\in\mathbb{N}}$ are bounded, the sequence $\{(\tilde{\mu}_1^k,\ldots,\tilde{\mu}_q^k)\}_{k\in\mathbb{N}} \subseteq \mathbb{L}_{m_1} \times \cdots \times \mathbb{L}_{m_q}$ defined by

$$\tilde{\mu}_j^k := \begin{cases} \tilde{\alpha}_j^k u_1(g_j(x^k)) + \tilde{\beta}_j^k u_2(g_j(x^k)), & \text{if } j \in J_- \cap J_+, \\ \tilde{\alpha}_j^k u_1(g_j(x^k)), & \text{if } j \in J_B \cup (J_- \setminus J_+), \\ \tilde{\alpha}_j^k u_2(g_j(x^k)), & \text{if } j \in J_+ \setminus J_-, \\ 0, & \text{if } j \in I_{\text{int}}(\overline{x}) \text{ or } j \notin (J_B \cup J_- \cup J_+) \end{cases}$$

is also bounded. Finally, note that all limit points of $\{(\tilde{\mu}_1^k,\ldots,\tilde{\mu}_q^k)\}_{k\in\mathbb{N}}$ are Lagrange multipliers associated with $\overline{x}$, which completes the first part of the proof. The second part follows directly from Theorem 4.1. □

Remark 4.2 In [11, Section 5], we proposed the so-called naive extensions of CRCQ (and CPLD) to NSOCP, which were obtained by replacing the conic constraints of (NSOCP) that satisfy $g_j(\overline{x}) \in \mathrm{bd}_+\mathbb{L}_{m_j}$ with standard NLP constraints, via a reduction function

$$\Phi_j(x) := g_{j,0}(x)^2 - \|\widehat{g}_j(x)\|^2,$$

and then applying the NLP definition of CRCQ (respectively, CPLD) to those reduced constraints. However, in order to compare it with the conditions we presented, we use another reduction function,

$$\tilde{\Phi}_j(x) := g_{j,0}(x) - \|\widehat{g}_j(x)\|,$$

instead of $\Phi_j(x)$, since $\nabla\tilde{\Phi}_j(x) = 2Dg_j(x)^\top u_1(g_j(x))$ for all x close enough to $\overline{x}$ and $j \in I_B(\overline{x})$. As mentioned in [11, Remark 5.1-c], using Φ_j or $\tilde{\Phi}_j$ characterizes different approaches. Assuming the second type of naive approach, we recall that naive-CRCQ (respectively, naive-CPLD) is satisfied at $\overline{x} \in \mathcal{F}$ when there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every $J_B \subseteq I_B(\overline{x})$, the following holds: If the family (10) is $\mathbb{R}^{|I_B(\overline{x})|} \times \prod_{j\in I_0(\overline{x})} \mathbb{R}^{m_j}$-linearly dependent (respectively, $\mathbb{R}_+^{|I_B(\overline{x})|} \times \prod_{j\in I_0(\overline{x})} \mathbb{L}_{m_j}$-linearly dependent), then the family $\{Dg_j(x)^\top u_1(g_j(x))\}_{j\in J_B}$ remains linearly dependent for all x in $\mathcal{V}$. Note that this definition coincides with nondegeneracy (respectively, Robinson's CQ) when no constraints are reducible—that is, when $I_B(\overline{x}) = \emptyset$—because $\emptyset$ is linearly independent. On the other hand, when all constraints are reducible, Definition 4.1 coincides with naive-CRCQ/CPLD. Thus, in the general case, both CQs of Definition 4.1 are strictly weaker than their "naive" counterparts.

5 Stronger Constant Rank Conditions With Applications

As we already mentioned, our study of constraint qualifications is driven toward global convergence of algorithms for solving (NSOCP). In particular, we presented in the previous section a global convergence proof for the external penalty method under weak-CPLD; to extend this result for a broader class of iterative methods, we now introduce more robust adaptations of weak-CPLD and weak-CRCQ. This is similar to what we did in [9] for NSDP problems. We start this section with an analog of [9, Definition 4.2] in NSOCP, which characterizes a perturbed version of weak-CRCQ and weak-CPLD.

Definition 5.1 (seq-CRCQ and seq-CPLD) We say that $\overline{x} \in \mathcal{F}$ satisfies the:

- *Sequential CRCQ condition for NSOCP* (seq-CRCQ) if for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\Delta_j^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{m_j}$, $j \in I_0(\overline{x}) \cup I_B(\overline{x})$, such that $\Delta_j^k \to 0$ for every j, there exists some $I \subseteq_\infty \mathbb{N}$, and convergent eigenvector sequences $\{u_1(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to \frac{1}{2}(1, -\overline{w}_j)$ and $\{u_2(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to \frac{1}{2}(1, \overline{w}_j)$, with $\|\overline{w}_j\| = 1$, for all $j \in I_0(\overline{x})$, such that for all subsets $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$, we have that, if the family of vectors $\mathcal{D}_{J_B,J_-,J_+}(\overline{x}, \overline{w})$ is linearly dependent, then $\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k)$ remains linearly dependent for all $k \in I$ large enough, where $\overline{w} = [\overline{w}_j]_{j\in J_-\cup J_+}$ and $w^k = [w_j^k]_{j\in J_-\cup J_+}$ with

$$u_1(g_j(x^k) + \Delta_j^k) = \frac{1}{2}(1, -w_j^k) \quad \text{and} \quad u_2(g_j(x^k) + \Delta_j^k) = \frac{1}{2}(1, w_j^k) \qquad (25)$$

for each $j \in J_- \cup J_+$. Recall that $\mathcal{D}_{J_B,J_-,J_+}(x, w)$ was defined in (18).

– *Sequential CPLD condition for NSOCP* (seq-CPLD) if for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Delta_j^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{m_j}$, $j \in I_0(\bar{x}) \cup I_B(\bar{x})$, such that $\Delta_j^k \to 0$ for every j, there exists some $I \subseteq_\infty \mathbb{N}$, and convergent eigenvector sequences $\{u_1(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to \frac{1}{2}(1, -\bar{w}_j)$ and $\{u_2(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to \frac{1}{2}(1, \bar{w}_j)$, with $\|\bar{w}_j\| = 1$, for all $j \in I_0(\bar{x})$, such that for all subsets $J_B \subseteq I_B(\bar{x})$ and $J_-, J_+ \subseteq I_0(\bar{x})$, we have that, if $\mathcal{D}_{J_B,J_-,J_+}(\bar{x}, \bar{w})$ is positively linearly dependent, then $\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k)$ remains linearly dependent for all $k \in I$ large enough, where $\bar{w}$ and w^k are as the previous item.

Note that the nondegeneracy condition (as in Proposition 2.1) implies seq-CRCQ, whereas Robinson's CQ implies seq-CPLD. Moreover, these implications are strict, as it is shown in the next counterexample:

Example 5.1 (Nondegeneracy and Robinson's CQ are strictly stronger than seq-CRCQ and seq-CPLD, respectively) Consider the constraint

$$g(x) := (-x, x) \in \mathbb{L}_2$$

at the point $\bar{x} := 0$, which is the only feasible point of the problem. In order to verify that $\bar{x}$ satisfies seq-CPLD and seq-CRCQ, let $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Delta^k\}_{k\in\mathbb{N}} \to 0$ be arbitrary sequences. We will assume that there is some $I \subseteq_\infty \mathbb{N}$ such that $\widehat{g}(x^k) + \widehat{\Delta}^k > 0$ for all $k \in I$, where $\Delta^k := (\Delta_0^k, \widehat{\Delta}^k) \in \mathbb{R}^2$, since the other cases (as in Example 3.2) follow analogously. Then, we have

$$u_1(g(x^k) + \Delta^k) = \frac{1}{2}(1, -1) \quad \text{and} \quad u_2(g(x^k) + \Delta^k) = \frac{1}{2}(1, 1),$$

which implies that $w^k = \bar{w} = 1$ for all $k \in I$. Hence, the vectors $Dg(\bar{x})^\top(1, -\bar{w}) = -2$ and $Dg(x^k)^\top(1, w^k) = 0$ are (positively) linearly dependent, but since $Dg(x^k)^\top(1, -w^k) = -2$ and $Dg(x^k)^\top(1, w^k) = 0$ are also linearly dependent for every $k \in I$, we see that seq-CPLD and seq-CRCQ both hold, while Robinson's CQ and nondegeneracy do not.

Example 5.1 shows that seq-CRCQ does not imply Robinson's CQ, and the converse is also false; otherwise, Robinson's CQ would imply weak-CRCQ, contradicting Example 4.3. Further, note that Definition 5.1 is basically Definition 4.1 with the addition of some perturbation sequences $\{\Delta_j^k\}_{k\in\mathbb{N}}$. Then, seq-CPLD implies weak-CPLD and seq-CRCQ implies weak-CRCQ, implying *a fortiori* that seq-CPLD and seq-CRCQ are constraint qualifications. However, the next example shows that these implications are both strict.

Example 5.2 (Seq-CRCQ and seq-CPLD are stronger than weak-CRCQ and weak-CPLD, respectively) Consider the constraint

$$g(x) := (x^2, x, 0) \in \mathbb{L}_3$$

at $\bar{x} := 0$. Let us begin by showing that $\bar{x}$ satisfies both weak-CRCQ and weak-CPLD, so let $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ be an arbitrary sequence. Again, as in Example 3.2, we will assume, without loss of generality, that there exists some $I \subseteq_\infty \mathbb{N}$ such that $x^k > 0$ for every $k \in I$. In this case, we must have

$$u_1(g(x^k)) = \frac{1}{2}(1, -1, 0) \quad \text{and} \quad u_2(g(x)) = \frac{1}{2}(1, 1, 0),$$

which yields $w^k = \overline{w} = (1, 0)$ for every $k \in I$. Then, $Dg(\overline{x})^\top(1, -\overline{w}) = -1$ and $Dg(\overline{x})^\top(1, \overline{w}) = 1$ are (positively) linearly dependent, but since $Dg(x^k)^\top(1, -w^k) = 2x^k - 1$ and $Dg(x)^\top u_2(g(x)) = 2x^k + 1$ are also linearly dependent for all $k \in I$ large enough so that $x^k \in (-\frac{1}{2}, \frac{1}{2})$, it means that weak-CRCQ and weak-CPLD both hold at $\overline{x}$.

However, taking any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that $x^k > 0$ for every $k \in \mathbb{N}$, and the perturbation vector

$$\Delta^k := (-(x^k)^2, -x^k, x^k) \to 0,$$

we have that $g(x^k) + \Delta^k := (0, 0, x^k)$, so its eigenvectors are uniquely determined by

$$u_1(g(x^k) + \Delta^k) = \frac{1}{2}(1, 0, -1) \quad \text{and} \quad u_2(g(x^k) + \Delta^k) = \frac{1}{2}(1, 0, 1),$$

implying $Dg(x^k)^\top u_1(g(x^k) + \Delta^k) = 2x^k > 0$ and $Dg(x^k)^\top u_2(g(x^k) + \Delta^k) = 2x^k > 0$ are positively linearly independent for every $k \in \mathbb{N}$. But since $Dg(\overline{x})^\top(1, 0, -1) = Dg(\overline{x})^\top(1, 0, 1) = 0$ we conclude that seq-CPLD and, by extension, seq-CRCQ, both fail at $\overline{x}$.

Furthermore, conditions seq-CRCQ and seq-CPLD can also be characterized in terms of a neighborhood, without sequences, just as the original CRCQ and CPLD conditions from NLP. Let us prove this:

Proposition 5.1 *Let $\overline{x} \in \mathcal{F}$. Condition seq-CRCQ (respectively, seq-CPLD) holds at $\overline{x}$ if, and only if, for every $\overline{w} := [\overline{w}_j]_{j\in I_0(\overline{x})}$ with $\|\overline{w}_j\| = 1$, $j \in I_0(\overline{x})$, there exists a neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{w})$ such that for every $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$, if $\mathcal{D}_{J_B,J_-,J_+}(\overline{x}, \overline{w})$ is (positively) linearly dependent, then $\mathcal{D}_{J_B,J_-,J_+}(x, w)$ remains linearly dependent for every $(x, w) \in \mathcal{V}$ with $w := [w_j]_{j\in I_0(\overline{x})}$ and $\|w_j\| = 1$ for every $j \in J_- \cup J_+$. Here, $\mathcal{D}_{J_B,J_-,J_+}(x, w)$ is as defined in (18).*

Proof Suppose that there exist some subsets $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$, and some $\overline{w} = [\overline{w}_j]_{j\in J_-\cup J_+}$ such that $\mathcal{D}_{J_B,J_-,J_+}(\overline{x}, \overline{w})$ is (positively) linearly dependent, but there is a sequence $\{(x^k, w^k)\}_{k\in\mathbb{N}} \to (\overline{x}, \overline{w})$ with $w^k := [w^k_j]_{j\in J_-\cup J_+}$ and $\|w^k_j\| = 1$, such that $\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k)$ is linearly independent for all $k \in \mathbb{N}$. Define, for each $k \in \mathbb{N}$ and $j \in J_B \cup I_- \cup I_+$, the perturbation vector

$$\Delta^k_j := \begin{cases} \frac{1}{k}\left(1, w^k_j\right) - g_j(x^k), & \text{if } j \in J_- \cup J_+ \\ g_{j,0}(\overline{x})\left(1, \frac{\widehat{g}_j(x^k)}{\|\widehat{g}_j(x^k)\|}\right) - g_j(x^k), & \text{if } j \in J_B, \end{cases} \tag{26}$$

which implies that $g_j(x^k) + \Delta_j^k \in \mathrm{bd}_+\mathbb{L}_{m_j}$ and hence its eigenvectors are uniquely determined for every such j and k. This contradicts Definition 5.1.

Conversely, pick any sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\Delta_j^k\}_{k\in\mathbb{N}} \to 0$, $j \in I_0(\overline{x}) \cup I_B(\overline{x})$, and any subsets $J_B \subseteq I_B(\overline{x})$ and $J_-, J_+ \subseteq I_0(\overline{x})$. Then, define $\{w^k\}_{k\in\mathbb{N}}$ as in Definition 5.1 and let $\overline{w} = [\overline{w}_j]_{j\in J_-\cup J_+}$ be such that $\|\overline{w}_j\| = 1$ for every $j \in J_- \cup J_+$ and $\lim_{k\in I} u_1(g_j(x^k) + \Delta_j^k) = \frac{1}{2}(1, -\overline{w}_j)$ and $\lim_{k\in I} u_2(g_j(x^k) + \Delta_j^k) = \frac{1}{2}(1, \overline{w}_j)$, for some $I \subseteq_\infty \mathbb{N}$. Note that $\lim_{k\in I} w^k = \overline{w}$, so if $\mathcal{D}_{J_B,J_-,J_+}(\overline{x}, \overline{w})$ is (positively) linearly dependent, then $\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k)$ remains linearly dependent for every k large enough. □

Remark 5.1 Note that Proposition 5.1 reveals that Definition 5.1 characterizes a "constant rank condition, or constant (positive) linear dependence, by conical slices." For example, consider a single constraint $g(x) \in \mathbb{L}_m$ at a point $\overline{x}$ such that $g(\overline{x}) \in \mathbb{L}_m$; then, seq-CRCQ holds at $\overline{x}$ if, and only if, for each conical slice of $\mathbb{L}_m$, which can be of two types:

1. $C_{\overline{w}}^1 = \mathrm{cone}(\{(1, \overline{w})\})$, for some $\overline{w} \in \mathbb{R}^{m-1}$ such that $\|\overline{w}\| = 1$;

2. $C_{\overline{w}}^2 = \mathrm{cone}(\{(1, -\overline{w}), (1, \overline{w})\})$, for some $\overline{w} \in \mathbb{R}^{m-1}$ such that $\|\overline{w}\| = 1$;

the dimension of

$$Dg(x)^\top \mathrm{span}(C_w^i) = \begin{cases} \mathrm{span}(\{Dg(x)^\top(1, w)\}), & \text{if } i = 1, \\ \mathrm{span}(\{Dg(x)^\top(1, -w), Dg(x)^\top(1, w)\}), & \text{if } i = 2, \end{cases}$$

remains constant for every (x, w) close enough to $(\overline{x}, \overline{w})$. The seq-CPLD condition admits a similar phrasing. That is, the local constant rank property must hold for every perturbation of $\overline{x}$ and every perturbation of the slice as well, roughly speaking, and the existence of two types of conical slices describes, intuitively, why should one consider every subset of $\{Dg(x)^\top(1, -w), Dg(x)^\top(1, w)\}$.

5.1 Global Convergence of Algorithms With Some Examples

Here, we show that the condition seq-CPLD can be used to prove global convergence of an abstract class of iterative algorithms, namely the ones that generate sequences of approximate solutions $\{x^k\}_{k\in\mathbb{N}}$, which we will assume to be convergent to some $\overline{x}$, and approximate Lagrange multipliers $\{\mu_j^k\}_{k\in\mathbb{N}} \subseteq \mathbb{L}_{m_j}$, $j \in \{1, \ldots, q\}$, in the sense that

$$\nabla_x L(x^k, \mu_1^k, \ldots, \mu_q^k) \to 0 \tag{27}$$

and for every $k \in \mathbb{N}$,

$$g_j(x^k) + \Delta_j^k \in \mathbb{L}_{m_j} \quad \text{and} \quad \langle g_j(x^k) + \Delta_j^k, \mu_j^k \rangle = 0 \tag{28}$$

for some sequences $\Delta_j^k \to 0$, $j \in \{1, \ldots, q\}$. Later in this section, we will discuss some details about some popular algorithms that generate this kind of sequence. But first, let us prove our unified global convergence result:

Theorem 5.1 (Global convergence under seq-CPLD) *Let* $\{x^k\}_{k\in\mathbb{N}}$ *and* $\{\mu_j^k\}_{k\in\mathbb{N}} \subseteq \mathbb{L}_{m_j}$, $j \in \{1, \ldots, q\}$ *satisfy* (27) *and* (28)*, and let* $\overline{x}$ *be a feasible limit point of* $\{x^k\}_{k\in\mathbb{N}}$ *that satisfies seq-CPLD. Then,* $\overline{x}$ *satisfies the KKT conditions.*

Proof For simplicity, let us assume that $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$. From (27), we obtain that

$$\nabla f(x^k) - \sum_{j=1}^{q} Dg_j(x^k)^\top \mu_j^k \to 0. \tag{29}$$

Now, by (28) we obtain

$$\mu_j^k = \begin{cases} 0, & \text{if } g_j(x^k) + \Delta_j^k \in \operatorname{int} \mathbb{L}_{m_j}, \\ \frac{\mu_{j,0}^k}{g_{j,0}(x^k) + \Delta_{j,0}^k} \Gamma_j(g_j(x^k) + \Delta_j^k), & \text{if } g_j(x^k) + \Delta_j^k \in \operatorname{bd}^+ \mathbb{L}_{m_j}, \end{cases}$$

where Γ_j is defined in (11), and μ_j^k can be any point of $\mathbb{L}_{m_j}$ if $g_j(x^k) + \Delta_j^k = 0$. Thus, there exists a spectral decomposition of

$$\mu_j^k := \alpha_j^k u_1(\mu_j^k) + \beta_j^k u_2(\mu_j^k),$$

such that $u_1(\mu_j^k)$ and $u_2(\mu_j^k)$ are also eigenvectors of $g_j(x^k) + \Delta_j^k$ for every $k \in \mathbb{N}$. Moreover, note that (28) implies that $\alpha_j^k \lambda_1(g_j(x_j^k) + \Delta_j^k) = 0$ and $\beta_j^k \lambda_2(g_j(x_j^k) + \Delta_j^k) = 0$ for every $k \in \mathbb{N}$ and every $j \in \{1, \ldots, q\}$. Then, $\beta_j^k = 0$ for all k large enough and for every $j \in I_B(\overline{x}) \cup I_{\text{int}}(\overline{x})$, because $\lambda_2(g_j(x_j^k) + \Delta_j^k) > 0$ for all large k in these cases. Therefore, we can rewrite (29) as

$$\nabla f(x^k) - \sum_{j\in I_0(\overline{x})} \left(\alpha_j^k Dg_j(x^k)^\top u_1(\mu_j^k) + \beta_j^k Dg_j(x^k)^\top u_2(\mu_j^k) \right) - \sum_{j\in I_B(\overline{x})} \alpha_j^k Dg_j(x^k)^\top u_1(\mu_j^k) \to 0.$$

The rest of the proof is similar to the proof of Theorem 4.2, which consists of using Carathéodory's lemma in the above relation, assuming that the new scalars are unbounded, and then directly applying Definition 5.1 to reach a contradiction, hence it shall be omitted. □

The sequences satisfying (27) and (28) are known as *Approximate-KKT* (AKKT) sequences, which define a sequential optimality condition introduced by Andreani et al. in [7] for NSOCP problems. Also, we must mention that several algorithms generate AKKT sequences; one recurrent example (see [7, Algorithm 5.1]) is the classical Hestenes–Powell–Rockafellar augmented Lagrangian method, which is based on the perturbed penalty function

$$L_{\rho,\tilde{\mu}_1,\dots,\tilde{\mu}_q}(x) := f(x) + \frac{\rho}{2}\left[\sum_{j=1}^{q}\left\|\mathcal{P}_{\mathbb{L}_{m_j}}\left(-g_j(x) - \frac{\tilde{\mu}_j}{\rho}\right)\right\|^2 - \left\|\frac{\tilde{\mu}_j}{\rho}\right\|^2\right],$$

where $\rho \in \mathbb{R}_+$ and $\tilde{\mu}_j \in \mathbb{L}_{m_j}$, $j \in \{1,\dots,q\}$, are given parameters. The sequence $\{x^k\}_{k\in\mathbb{N}}$ is computed as approximate stationary points of $L_{\rho_k,\tilde{\mu}_1^k,\dots,\tilde{\mu}_q^k}(x)$ and their associate approximate Lagrange multipliers are given by

$$\mu_j^k := \mathcal{P}_{\mathbb{L}_{m_j}}\left(-\rho_k g_j(x^k) - \tilde{\mu}_j^k\right),$$

where $\{\rho_k\}_{k\in\mathbb{N}}$ is the penalty parameter and $\{\tilde{\mu}_j^k\}_{k\in\mathbb{N}} \subseteq \mathbb{L}_{m_j}$ are given sequences and $\Delta_j^k := \frac{\mu_j^k - \tilde{\mu}_j^k}{\rho_k}$ for every $j \in \{1,\dots,q\}$. In particular, note that $\nabla L_{\rho_k,\tilde{\mu}_1^k,\dots,\tilde{\mu}_q^k}(x^k) = \nabla_x L(x^k,\mu_1^k,\dots,\mu_q^k)$ for every $k \in \mathbb{N}$. See also [8] for a more detailed discussion on this topic.

Besides the augmented Lagrangian and its variants, the *sequential quadratic programming* (SQP) algorithm of Kato and Fukushima [30, Algorithm 1] can also be proved to generate output sequences that satisfy (27) and (28). For completeness, we state their algorithm below:

In [30], Kato and Fukushima proved the global convergence of Algorithm 1 under the following assumptions:

A1. Step 1 is well defined for every $k \in \mathbb{N}$;
A2. The output sequence $\{x^k\}_{k\in\mathbb{N}}$ of Algorithm 1 is bounded;
A3. The multiplier sequences $\{\mu_j^k\}_{k\in\mathbb{N}}$, $j \in \{1,\dots,q\}$ computed by the method are all bounded.

Observe that these assumptions, although somewhat standard, are demands over the behavior of the algorithm itself instead of the problem, and a convergence theory that makes strong assumptions over the behavior of the method is, to say the best, fragile. Even so, A1 and A2 can be considered a "necessary evil" since their violation means that the execution of the method has terminated in failure. Assumption A3, on the other hand, is not plausible since it basically guides the method toward convergence. Instead of A3, an assumption over the problem (and not the method), for instance the fulfillment of a constraint qualification at every limit point of $\{x^k\}_{k\in\mathbb{N}}$, would be more reasonable for illustrating its strength. Of course, Robinson's CQ is well suited for this role since it implies A3, but an improvement can be made with the weaker constraint qualification seq-CPLD; that is, under the following assumption:

A4. All limit points of $\{x^k\}_{k\in\mathbb{N}}$ satisfy seq-CPLD.

Algorithm 1 Sequential quadratic programming algorithm of [30].

Input: An initial point $x^0 \in \mathbb{R}^n$ and some parameters $\alpha_0 > 0$, $\sigma \in (0,1)$, $\gamma_1 > 0$, $\gamma_2 > 0$, and $\tau > 0$.

Set $k := 0$. Then:

Step 1: Choose a symmetric positive definite matrix $M^k \in \mathbb{R}^{n\times n}$ such that $\gamma_1\|z\|^2 \leq z^\top M^k z \leq \gamma_2\|z\|^2$ for every $z \in \mathbb{R}^n$, and find a solution d^k if possible of the problem:

$$\begin{aligned} &\underset{d\in\mathbb{R}^n}{\text{Minimize}} && \nabla f(x^k)^\top d + \frac{1}{2}d^\top M^k d, \\ &\text{subject to} && g_j(x^k) + Dg_j(x^k)d \in \mathbb{L}_{m_j},\ \forall j \in \{1,\dots,q\} \end{aligned} \tag{QP}$$

together with its Lagrange multipliers $\mu_j^k \in \mathbb{L}_{m_j}$, $j \in \{1,\dots,q\}$; if $d^k = 0$, then **stop**;

Step 2: Set the penalty parameter as follows: If $\alpha^k \geq \max\{|\mu_{j,0}^k| : j \in \{1,\dots,q\}\}$, then $\alpha^{k+1} := \alpha^k$; otherwise, $\alpha^{k+1} := \max\{\alpha^k, |\mu_{j,0}^k| : j \in \{1,\dots,q\}\} + \tau$;

Step 3: Compute some scalar $t^k \in (0,1]$ satisfying

$$\Phi_{\alpha^{k+1}}(x^k) - \Phi_{\alpha^{k+1}}(x^k + t^k d^k) \leq \sigma t^k (d^k)^\top M^k d^k; \tag{30}$$

where

$$\Phi_\alpha(x) := f(x) + \alpha \sum_{j=1}^{q} \max\{0, -g_{j,0}(x) - \|\widehat{g}_j(x)\|\}$$

is a penalty function;

Step 4: Set $x^{k+1} := x^k + t^k d^k$ and $k := k+1$, and go to Step 1.

Then, we can easily rephrase an excerpt from the proof of [30, Theorem 1] and apply Theorem 5.1 to obtain the same convergence result of [30] under A1, A2, and A4, instead of A3 or Robinson's CQ. However, it should be noticed that A4 may hold even when the approximate Lagrange multiplier sequences are unbounded.

Proposition 5.2 *Under A1, the output sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\mu_j^k\}_{k\in\mathbb{N}}$, $j \in \{1,\dots,q\}$, of Algorithm 1 satisfy (27) and (28).*

Proof For each $k \in \mathbb{N}$, assumption A1 tells us that x^k and $\mu_j^k \in \mathbb{L}_{m_j}$, $j \in \{1,\dots,q\}$ satisfy the following:

$$\begin{aligned} \nabla f(x^k) + M^k d^k - \textstyle\sum_{j=1}^q Dg_j(x^k)^\top \mu_j^k = 0, \\ \langle \mu_j^k, g_j(x^k) + Dg_j(x^k)d^k\rangle = 0, \forall j \in \{1,\dots,q\}, \\ g_j(x^k) + Dg_j(x^k)d^k \in \mathbb{L}_{m_j}, \forall j \in \{1,\dots,q\}. \end{aligned}$$

Since by construction $\{M^k\}_{k\in\mathbb{N}}$ is bounded and by [30, Theorem 1] we have $\{d^k\}_{k\in\mathbb{N}} \to 0$, the conclusion follows by taking $\Delta_j^k := Dg_j(x^k)d^k$ for every $k \in \mathbb{N}$ and every $j \in \{1,\dots,q\}$. □

For the sake of completeness, we present a formal statement of the convergence result of Algorithm 1 under seq-CPLD, which follows immediately from the previous proposition.

Corollary 5.1 *Assume A1, A2 and A4. Every limit point of the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *generated by Algorithm 1 satisfies the KKT conditions.*

5.2 On Error Bounds and Robustness

Another interesting implication of CRCQ and CPLD from the literature concerns error bounds. To address it to NSOCP, let us recall the definition of the so-called *metric subregularity CQ* for (NSOCP) problems.

Definition 5.2 (*MSCQ*) Let $\overline{x}$ be a feasible point of (NSOCP) and let $g(x) := (g_1(x), \ldots, g_q(x))$. We say that $\overline{x}$ satisfies the metric subregularity CQ (MSCQ) when there exists some $\gamma > 0$ and a neighborhood $\mathcal{V}$ of $\overline{x}$ such that

$$\operatorname{dist}(x, \mathcal{F}) \leq \gamma \operatorname{dist}(g(x), \Pi_{j=1}^{q} \mathbb{L}_{m_j})$$

for every $x \in \mathcal{V}$, where $\mathcal{F}$ is the feasible set of (NSOCP).

The following result shows a sufficient condition in order to obtain MSCQ. This result is an adaptation from Minchenko and Stakhovski [34, Theorem 2] for nonlinear programming problems. Also, an extension for semidefinite programming was made in [9, Proposition 5.1] and hence its proof will be omitted.

Proposition 5.3 *Let* $\overline{x} \in \mathcal{F}$ *and assume that* g_j *are twice differentiable around* $\overline{x}$*, with* $j \in \{1, \ldots, q\}$*. Given* $x \in \mathbb{R}^n$*, let* $\Lambda_x(y)$ *denote the set of Lagrange multipliers associated with any given solution* y *of the problem of minimizing* $\|z - x\|$ *subject to* $g_j(z) \in \mathbb{L}_{m_j}$*,* $j \in \{1, \ldots, q\}$*,* $z \in \mathbb{R}^n$*. If there exist numbers* $\tau > 0$ *and* $\delta > 0$ *such that* $\Lambda_x(y) \cap \operatorname{cl}(B(0, \tau)) \neq \emptyset$ *for every* $x \in B(\overline{x}, \delta)$*, then* $\overline{x}$ *satisfies MSCQ.*

Then, we shall prove that seq-CPLD and seq-CRCQ are robust, and this, together with Proposition 5.3, is enough to show that they imply MSCQ.

Theorem 5.2 (Robustness of seq-CPLD (and seq-CRCQ)) *If* $\overline{x} \in \mathcal{F}$ *satisfies seq-CPLD (or seq-CRCQ), then:*

1. *There is a neighborhood* $\mathcal{V}$ *of* $\overline{x}$*, such that every* $x \in \mathcal{V} \cap \mathcal{F}$ *also satisfies seq-CPLD (respectively, seq-CRCQ);*
2. *MSCQ holds at* $\overline{x}$*.*

Proof We will only exhibit the proof for seq-CPLD, since the proof for seq-CRCQ is analogous. Suppose that item 1 is false, then there is a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that seq-CPLD fails at x^k, for all $k \in \mathbb{N}$. That is, for each $k \in \mathbb{N}$ there is some $w^k := [w_j^k]_{j\in I_0(x^k)}$ with $\|w_k^k\| = 1$ for every $j \in I_0(x^k)$, some sequences $\{x_\ell^k\}_{\ell\in\mathbb{N}} \to x^k$ and $\{w_\ell^k\}_{\ell\in\mathbb{N}} \to w^k$, and subsets $J_B^k \subseteq I_B(x^k)$ and $J_-^k, J_+^k \subseteq I_0(x^k)$ such that $\mathcal{D}_{J_B^k, J_-^k, J_+^k}(x^k, w^k)$ is positively linearly dependent, but $\mathcal{D}_{J_B^k, J_-^k, J_+^k}(x_\ell^k, w_\ell^k)$

is linearly independent for every $\ell \in \mathbb{N}$. By the infinite pigeonhole principle, we can assume that $I_0 = I_0(x^k)$ and $I_B = I_B(x^k)$ are the same for every $k \in \mathbb{N}$, and also that $J_B = J_B^k$, $J_- = J_-^k$, and $J_+ = J_+^k$ for every $k \in \mathbb{N}$, passing to a subsequence if necessary. Moreover, note that we can also assume that $I_0 \subseteq I_0(\overline{x})$ and $I_B \subseteq I_0(\overline{x}) \cup I_B(\overline{x})$. Now consider the following sets:

$$\tilde{J}_B := J_B \cap I_B(\overline{x}), \quad \tilde{J}_- := J_- \cup (J_B \cap I_0(\overline{x})), \quad \text{and} \quad \tilde{J}_+ := J_+.$$

By construction, note that $\mathcal{D}_{\tilde{J}_B, \tilde{J}_-, \tilde{J}_+}(x_\ell^k, w_\ell^k)$ is linearly independent for every $k, \ell \in \mathbb{N}$. For each k, let $\ell(k)$ be such that $\|w^k - w_{\ell(k)}^k\| < \frac{1}{k}$, and let $\overline{w}$ be any limit point of $\{w^k\}_{k \in \mathbb{N}}$. Without loss of generality, we will assume that $w^k \to \overline{w}$, which also implies that $w_{\ell(k)}^k \to \overline{w}$.

Analogously to (26), we can construct some $\Delta_j^k \in \mathbb{R}^{m_j}$ for every $j \in I_0(\overline{x}) \cup I_B(\overline{x})$, such that $g_j(x_{\ell(k)}^k) + \Delta_j^k \in \text{bd}_+ \mathbb{L}_{m_j}$ and hence its eigenvectors are uniquely determined by

$$u_1(g_j(x_{\ell(k)}^k) + \Delta_j^k) = \frac{1}{2}\left(1, \frac{\widehat{g}_j(x_{\ell(k)}^k)}{\|\widehat{g}_j(x_{\ell(k)}^k)\|}\right), \ \forall j \in \tilde{J}_B,$$

and

$$u_1(g_j(x_{\ell(k)}^k) + \Delta_j^k) = \frac{1}{2}\left(1, -w_{\ell(k)}^k\right)$$

and

$$u_2(g_j(x_{\ell(k)}^k) + \Delta_j^k) = \frac{1}{2}\left(1, w_{\ell(k)}^k\right), \ \forall j \in \tilde{J}_- \cup \tilde{J}_+.$$

With this in mind, on the one hand, we have that $\mathcal{D}_{\tilde{J}_B, \tilde{J}_-, \tilde{J}_+}(\overline{x}, \overline{w})$ is linearly dependent, because the family $\mathcal{D}_{\tilde{J}_B, \tilde{J}_-, \tilde{J}_+}(x^k, w^k)$ is linearly dependent for every $k \in \mathbb{N}$. But on the other hand, $\mathcal{D}_{\tilde{J}_B, \tilde{J}_-, \tilde{J}_+}(x_{\ell(k)}^k, w_{\ell(k)}^k)$ is linearly independent for every $k \in \mathbb{N}$, and the fact that the eigenvectors of $g_j(x_{\ell(k)}^k) + \Delta_j^k$ are uniquely determined for all $j \in \tilde{J}_B \cup \tilde{J}_- \cup \tilde{J}_+$, together with $w_{\ell(k)}^k \to \overline{w}$, contradicts seq-CPLD at $\overline{x}$.

The proof of item 2 follows analogously to the proof of [9, Theorem 5.1], which is essentially a corollary of item 1 and Proposition 5.3; hence it will be omitted. □

For a better exposition, what follows is a diagram that represents the relationship of some existing constraint qualifications and the ones that we present in this paper (Fig. 1).

6 Conclusion

In our previous work, we studied two ways of incorporating some structural features of the semidefinite cone into the nondegeneracy condition of Shapiro and Fan [45];

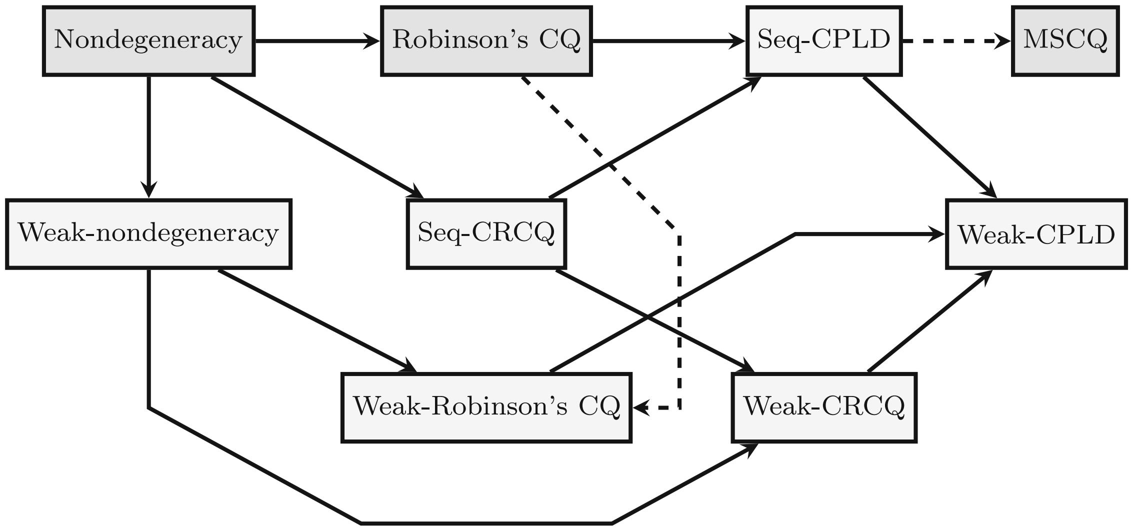

Fig. 1 Constraint qualifications for NSOCP. Strict implications are represented by solid arrows. Possibly two-sided implications are represented by dashed arrows

among them was the eigendecomposition, which has always been widely exploited in the design of algorithms for NSDP—for instance, see [31]. Quite surprisingly, after incorporating eigendecompositions into the nondegeneracy condition (and also Robinson's CQ) we obtained a strictly weaker constraint qualification by means of considering only converging sequences of eigenvectors associated with a given point of interest, which was called weak-nondegeneracy (respectively, weak-Robinson's CQ). Moreover, this "sequential approach" allowed us to bypass the main difficulty in generalizing the celebrated constant rank constraint qualification of NLP, to NSDP [9], which is the presence of a potentially nonzero duality gap even in feasible linear problems (see also [12] for a more detailed discussion on this topic). In this paper we bring those concepts to the context of NSOCP where several improvements with respect to the NSDP approach were made.

It is well known (see, for instance, the seminal work of Alizadeh and Goldfarb [2]) that although NSOCP problems can be reformulated as particular instances of NSDP problems, solving them via such a reformulation is generally not a good practice for a handful of reasons. Likewise, extensions of the sequential-type constraint qualifications of [9, 10] to NSOCP demand a specialized analysis to be properly conducted. In fact, the second-order cone induces a distinguished eigendecomposition that is easily computable, contrary to NSDP, which allows a deeper analysis to be made. For instance, besides extending the weak variants of the nondegeneracy condition and Robinson's CQ from NSDP to NSOCP, this paper also presents a full comparison between these weak conditions and their standard versions, which is an issue we could not properly address in [10]. Some technical results from [10] could also be explained in a somewhat natural way in this paper. Moreover, besides extending the constant rank conditions from [9], we also gave them a geometrical interpretation in terms of the conical slices of the second-order cone (Remark 5.1).

Very recently, we have been extending the notions of constant rank-type constraint qualifications to the contexts of NSDP and NSOCP. While [12] follows an implicit function approach pioneered by Janin [29] and giving rise to a definition of CRCQ that enjoys strong second-order properties, in this paper we exploit a sequential approach [7], which allows even weaker conditions to be defined, such as the CPLD condition, while enjoying global convergence properties of several algorithms without assuming boundedness of the set of Lagrange multipliers but still allowing computation of error bounds. Not surprisingly, when extending NLP concepts to the conic context, different points of view may give rise to different possible extensions, each one extending different applications of the concept. Some relevant topics in conic programming that we expect the conditions we define in this paper will be particularly relevant are: in the global convergence analysis of other classes of algorithms, including second-order algorithms; the study of the boundedness of Lagrange multipliers estimates and the use of scaled stopping criteria ; stability analysis of parametric optimization problems ; and necessary optimality conditions for some extended classes of bilevel optimization problems with conic constraints .

Acknowledgements The authors received financial support from FAPESP (Grants 2017/18308-2, 2017/17840-2,2017/12187-9, 2018/24293-0 and 2020/00130-5), CEPID-CeMEAI (Granted by FAPESP 2013/07375-0), CNPq (Grants 301888/2017-5, 303427/2018-3, 404656/2018-8 and 306988/2021-6), PRONEX—CNPq/FAPERJ (Grant E-26/010.001247/2016), and ANID (FONDECYT Grant 1201982, Program ANID ACE210010 and Basal Program CMM ANID AFB210005).

References

1. Adam, L., Henrion, R., Outrata, J.: On M-stationarity conditions in MPECs and the associated qualification conditions. Math. Program. **168**, 229–259 (2018)
2. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. Ser. B **95**, 3–51 (2003)
3. Alvarez, F., López, J., Ramírez, C., H.: Interior proximal algorithm with variable metric for second-order cone programming: applications to structural optimization and support vector machines. Optim. Methods Softw. **25**(6), 859–881 (2010)
4. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagragrian methods under the constant positive linear dependence constraint qualification. Math. Program. **111**(1–2), 5–32 (2008)
5. Andreani, R., Echagüe, C.E., Schuverdt, M.L.: Constant-rank condition and second-order constraint qualification. J. Optim. Theory Appl. **146**, 255–266 (2010)
6. Andreani, R., Fukuda, E.H., Haeser, G., Ramírez C., H., Santos, D.O., Silva, P.J.S.: Silveira Erratum to, Silva, P.J.S., Silveira, T.P.: Erratum to: New constraint qualifications and optimality conditions for second order cone programs (2021) (To appear in Set-Valued and Variational Analysis)
7. Andreani, R., Fukuda, E.H., Haeser, G., Santos, D.O., Secchin, L.D.: Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. Technical report (2019)
8. Andreani, R., Gómez, W., Haeser, G., Mito, L.M. , Ramos, A.: On optimality conditions for nonlinear conic programming (2021) (To appear in Mathematics of Operations Research)
9. Andreani, R., Haeser, G., Mito, L.M., Ramírez C., H.: Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. Technical report (2020). arXiv:2106.00775
10. Andreani, R., Haeser, G., Mito, L.M., Ramírez C., H.: Weak notions of nondegeneracy in nonlinear semidefinite programming. Technical report (2020). arXiv:2012.14810
11. Andreani, R., Haeser, G., Mito, L.M. , Ramírez C., H. Santos, D.O. , Silveira, T.P.: Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming (2021) (To appear in Optimization Letters)
12. Andreani, R., Haeser, G., Mito, L.M., Ramírez C., H., Silveira, T.P.: First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. Technical report (2021). arXiv:2107.04693
13. Andreani, R., Haeser,G., Schuverdt, M.L. , Secchin, L.D., Silva,P.J.S.: On scaled stopping criteria for a safeguarded augmented Lagrangian method with theoretical guarantees. (2021) (To appear in Mathematical Programming Computation)

14. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. Ser. A **135**(1–2), 255–273 (2012)
15. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**(3), 1109–1135 (2012)
16. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On the relation between constant positive linear dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. **125**(2), 473–485 (2005)
17. Bertsekas, D.P.: Nonlinear Programming. Athenas Scientific, Belmont, Mass (1999)
18. Bonnans, J.F., Ramírez C., H.: Perturbation analysis of second-order cone programming problems. Math. Program. Ser. B **104**, 205–227 (2005)
19. Bonnans, J.F., Shapiro, A.: Pertubation Analysis of Optimization Problems. Springer-Verlag, Berlin (2000)
20. Bosch, P., López, J., Ramírez, C., H., Robotham, H.: Support vector machine under uncertainty: an application for hydroacoustic classification of fish-schools in Chile. Expert Syst. Appl. **40**(10), 4029–4034 (2013)
21. Chen, J.-S., Ye, J.J., Zhang, J., Zhou, J.: Exact formula for the second-order tangent set of the second-order cone complementarity set. SIAM J. Optim. **29**(4), 2986–3011 (2019)
22. Chi, X., Wan, Z., Hao, Z.: Second order sufficient conditions for a class of bilevel programs with lower level second-order cone programming problem. J. Indus. Manage. Optim. **11**(4), 1111–1125 (2015)
23. Fukuda, E.H., Haeser, G., Mito, L.M.: On the weak second-order optimality condition for nonlinear semidefinite and second-order cone programming. Technical report (2020) (available at Optmization online)
24. Gauvin, J., Dubeau, F.: Differential properties of the marginal function in mathematical programming. Math. Program. Study **19**, 101–119 (1982)
25. Gfrerer, H., Mordukhovich, B.S.: Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions. SIAM J. Optim. **25**(4), 2081–2119 (2015)
26. Gfrerer, H., Outrata, J.V.: On computation of generalized derivatives of the normal-cone mapping and their applications. Math. Oper. Res. **41**(4), 1535–1556 (2016)
27. Henrion, R., Kruger, A.Y., Outrata, J.V.: Some remarks on stability of generalized equations. J. Optim. Theory Appl. **159**, 681–697 (2013)
28. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
29. Janin, R.: Directional derivative of the marginal function in nonlinear programming. Math. Program. Study **21**, 127–138 (1984)
30. Kato, H., Fukushima, M.: An SQP-type algorithm for nonlinear second-order cone programs. Optim. Lett. **1**(2), 129–144 (2007)
31. Kočvara, M., Stingl, M.: PENNON: A code for convex nonlinear and semidefinite programming. Optim. Methods Softw. **18**(3), 317–333 (2003)
32. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Applications of second-order cone programming. Linear Algebra Appl. **284**, 193–228 (1998)
33. Luna, J.P., Sagastizábal, C., Silva, P.J.S.: A discussion on electricity prices, or the two sides of the coin. Phil. Trans. R. Soc. A **379**, 20190428 (2021)
34. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)
35. Minchenko, L., Stakhovski, S.: Parametric nonlinear programming problems under the relaxed constant rank condition. SIAM J. Optim. **21**, 314–332 (2011)
36. Mordukhovich, B.S., Hang, N.T.V., Sarabi, M.E.: Augmented Lagrangian method for second-order conic programs under second-order sufficiency (2021) (To appear in Journal of Global Optimization)
37. Mordukhovich, B.S., Outrata, J.V., Ramírez, C., H.: Graphical derivatives and stability analysis for parameterized equilibria with conic constraints. Set-Valued Var. Anal. **23**, 687–704 (2015)
38. Mordukhovich, B.S., Outrata, J.V., Ramírez, C., H.: Second-order variational analysis in conic programming with applications to optimality and stability. SIAM J. Optim. **25**(1), 76–101 (2015)
39. Mordukhovich, B.S., Outrata, J.V., Sarabi, M.E.: Full stability of locally optimal solutions in second-order cone programs. SIAM J. Optim. **24**(4), 1581–1613 (2014)
40. Outrata , J.V. , Ramírez C., H.: On the Aubin property of critical points to perturbed second-order cone programs. SIAM J. Optim. **21**(3), 798–823 (2011)
41. Powell, M.J.D.: A Method for Nonlinear Constraints in Minimization Problems. Academic Press, New York (1982)

42. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
43. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control. Optim. **12**, 268–285 (1974)
44. Rockafellar, R.T.: Directional differentiability of the optimal value function in a nonlinear programming problem. Math. Program. Study **21**, 213–226 (1984)
45. Shapiro, A., Fan, M.K.H.: On eigenvalue optimization. SIAM J. Optim. **5**(3), 552–569 (1995)
46. Ye, J.J., Zhu, D.L.: New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches. SIAM J. Optim. **20**, 1885–1905 (2010)
47. Zhang, Y., Zhang, L.: New constraint qualifications and optimality conditions for second order cone programs. Set-Valued Var. Anal. **27**, 693–712 (2019)

Appendix D

Article: First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. doi.org/10.1007/s10107-023-01942-8

Mathematical Programming (2023) 202:473–513
https://doi.org/10.1007/s10107-023-01942-8

FULL LENGTH PAPER

Series A

First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition

Roberto Andreani[1] · Gabriel Haeser[2] · Leonardo M. Mito[2] · Héctor Ramírez[3] · Thiago P. Silveira[2]

Received: 5 December 2021 / Accepted: 21 February 2023 / Published online: 23 March 2023

Abstract

The well known constant rank constraint qualification [Math. Program. Study 21:110–126, 1984] introduced by Janin for nonlinear programming has been recently extended to a conic context by exploiting the eigenvector structure of the problem. In this paper we propose a more general and geometric approach for defining a new extension of this condition to the conic context. The main advantage of our approach is that we are able to recast the strong second-order properties of the constant rank condition in a conic context. In particular, we obtain a second-order necessary optimality condition that is stronger than the classical one obtained under Robinson's constraint qualification, in the sense that it holds for every Lagrange multiplier, even though our condition is independent of Robinson's condition.

Keywords Constraint qualifications · Constant rank · Second-order optimality conditions · Second-order cone programming · Semidefinite programming

Gabriel Haeser
ghaeser@ime.usp.br

Roberto Andreani
andreani@unicamp.br

Leonardo M. Mito
leokoto@ime.usp.br

Héctor Ramírez
hramirez@dim.uchile.cl

Thiago P. Silveira
thiagops@ime.usp.br

1 Department of Applied Mathematics, State University of Campinas, São Paulo, Brazil

2 Department of Applied Mathematics, University of São Paulo, São Paulo, Brazil

3 Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS IRL 2807), Universidad de Chile, Santiago, Chile

Mathematics Subject Classification 90C46 · 90C30 · 90C22

1 Introduction

In the classical *nonlinear programming* (NLP) context, the so-called *constant rank constraint qualification* (CRCQ) [36] was first presented as a tool for stability analysis, which stood out for being independent of the usual *Mangasarian-Fromovitz constraint qualification* (MFCQ) and strictly weaker than the *linear independence constraint qualification* (LICQ). For instance, it has been applied with this purpose in NLP [29, 36, 46, 47, 49], *mathematical programs with equilibrium constraints* (MPEC) [33], *generalized equations* [34], and *bilevel optimization* [44, 59]. Also, it is the origin of several other constant rank-type conditions, such as the *constant positive linear dependence* [10, 12, 51] and the *constant rank of the subspace component* [11], which have been successfully applied in the convergence analysis of iterative algorithms. To name a few algorithms whose convergence theory relies on CRCQ and its variants, we point out: an augmented Lagrangian method [3, 13], a regularized interior point method [52], sequential quadratic programming methods for NLP [41, 51, 58] and MPEC [38], and some relaxation schemes for MPEC [35, 57]. In fact, a particularly interesting aspect of CRCQ that makes it suitable for supporting practical algorithms is the fact it can be roughly interpreted as a relaxation of LICQ that is able to separate the core information of the problem, ignoring redundant constraints. Moreover, all linear programming problems satisfy CRCQ, in contrast with LICQ and MFCQ.

Besides convergence of algorithms and stability analysis, CRCQ was used in several contexts, such as NLP [4, 13, 45], MPEC [32], vector optimization [43], and continuous-time NLP [48], for studying second-order necessary optimality conditions. One of the main goals of this paper is to bring such results to more general conic programming contexts, namely *nonlinear second-order cone programming* (NSOCP) and *nonlinear semidefinite programming* (NSDP). In the seminal paper by Bonnans, Cominetti, and Shapiro [21], the authors derived no-gap second-order optimality conditions for problems over *second-order regular cones* [21, Definition 3], such as NSDP and NSOCP, under the well-known *Robinson's CQ* (see (7) on page 8, or [53]), which is the natural extension of MFCQ to conic programming. In particular, their second-order necessary condition states that every local solution that satisfies Robinson's CQ must also satisfy the following: for every critical direction, there exists a Lagrange multiplier (possibly depending on this direction), such that a certain quadratic form is nonnegative with respect to such direction and multiplier. However, the second-order condition that is obtained under CRCQ in NLP replaces "there exists a Lagrange multiplier" with "for every Lagrange multiplier," which is stronger than the one of [21]. Although this stronger condition can be obtained from [21] after assuming that the Lagrange multiplier is unique, which is ensured by stronger constraint qualifications such as the *nondegeneracy* condition (see (8) on page 8), this assumption is often regarded as too stringent. To the best of our knowledge, no second-order result concerning every Lagrange multiplier, without assuming its uniqueness, has been presented so far in the literature of nonlinear conic programming. Moreover, no extension of CRCQ has been proposed for nonlinear conic programming until very recently.

In 2019, Zhang and Zhang [60] proposed an extension of CRCQ and its relaxed version [46] for NSOCP, but it was later discovered that their results were incorrect [5]. This event has motivated us to investigate other possible extensions of CRCQ to conic problems, and their properties. The first step in this direction was made in [6], for NSOCP and NSDP problems with multiple constraints. The idea of [6] is to rewrite some of the conic constraints as locally equivalent NLP constraints, whenever possible, and then jointly applying nondegeneracy and the NLP version of CRCQ to the resulting problem. Later, in [8], based on the ideas from [7], we improved this strategy by exploiting the eigenvector structure of the semidefinite cone to deal with the conic constraints that could not be rewritten as NLP constraints. This approach was also extended to NSOCP problems in [9]. In simple terms, the condition of [8, 9] demands the rank of some families of functions to remain constant along every sequence converging to the point of interest – roughly speaking, a constant rank "by paths" – therefore, this extension is highly specialized to deal with sequences generated by iterative algorithms, but since this rank may vary between paths, it is likely unsuitable for other purposes. Indeed, the focus of [8, 9] was the global convergence of a large class of algorithms to first-order stationary points, and no second-order results were provided in it. Nevertheless, it is reasonable to expect that CRCQ may have multiple independent and correct extensions, each one of them generalizing at least one important aspect of it, but perhaps not all of them.

A common feature of all previous attempts of extending CRCQ to a conic context is an approach based on re-characterizing the conic program and the nondegeneracy condition, trying to make them as similar to NLP and LICQ as possible, so the extension of CRCQ would come out straightforwardly. This is somehow understandable because, even in NLP, the CRCQ condition has never received a geometrical interpretation before. In this paper, we present a new geometrical characterization of CRCQ for NLP in terms of the faces of the nonnegative orthant, which suggests a natural extension of it to NSOCP and NSDP. A point that we should stress is that contrary to our previously mentioned works, the definition of CRCQ that we present here is very simple. We prove that this extension is a constraint qualification strictly weaker than nondegeneracy and independent of Robinson's CQ, as it should be, and we also compare it with the condition of [8, 9]. Then, as an application, we show that every local solution of the problem satisfies the strong second order optimality condition, provided our extension of CRCQ holds. Moreover, just as it happens in NLP, our result does not demand *a priori* any specific condition over the Lagrange multiplier set, besides nonemptiness.

The structure of this paper is as follows: Sect. 2 consists of a nonlinear conic programming review emphasizing some aspects of the theory that are not commonly discussed in the literature; in Sect. 3, we analyze CRCQ for NLP and we show how it can be interpreted in terms of the faces of the nonnegative orthant. In Sects. 4 and 5, we propose extensions of CRCQ for NSOCP and NSDP, respectively, and we prove some of its properties. Finally, in Sect. 6, we conclude this paper with a short discussion and some ideas of prospective work.

We end this section by introducing some of our basic notation: throughout this paper, $\mathbb{E}$ will denote a finite-dimensional linear space equipped with the inner product $\langle \cdot, \cdot \rangle$; and for a given set $S \subseteq \mathbb{E}$, we will denote the *polar* of S by

$$S^\circ := \{z \in \mathbb{E} \mid \langle z, y \rangle \le 0, \ \forall y \in S\}$$

and the *orthogonal complement* of S will be denoted by $S^\perp$. The notations $\mathrm{cl}(S)$, $\mathrm{int}(S)$, $\mathrm{bd}(S)$, and $\mathrm{bd}^+(S)$ stand for the topological closure, interior, boundary, and boundary excluding the origin of S in $\mathbb{E}$, respectively. The *smallest cone* that contains S will be denoted by $\mathrm{cone}(S)$, and the *smallest linear space* that contains S will be denoted by $\mathrm{span}(S)$. Finally, for a twice continuously differentiable function $g : \mathbb{R}^n \to \mathbb{E}$ and a given point $x \in \mathbb{R}^n$, we denote by $Dg(x)$ and $D^2 g(x)$ the first- and second-order derivative of g at x, respectively. As usual, $Dg(x)^T$ stands for the adjoint of $Dg(x)$, which by definition satisfies $\langle Dg(x)d, z \rangle = \langle d, Dg(x)^T z \rangle$ for all $d \in \mathbb{R}^n$ and $z \in \mathbb{E}$, and the action of $D^2 g(x)$ over $d_1, d_2 \in \mathbb{R}^n$ will be denoted by $D^2 g(x)[d_1, d_2]$.

2 Common framework: nonlinear conic programming

In this section, we will review some classical results of convex analysis, and first- and second-order optimality conditions and constraint qualifications for NSOCP and NSDP. These problems are the cornerstones of two independent research fields, but they can also be seen as particular cases of a *nonlinear conic programming* (NCP) problem, given by

$$\begin{aligned} &\text{Minimize} && f(x), \\ &\text{s.t.} && g(x) \in \mathcal{K}, \end{aligned} \tag{NCP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{E}$ are twice continuously differentiable, and $\mathcal{K} \subseteq \mathbb{E}$ is a closed convex pointed cone that is assumed to be nonempty. We will use (NCP) as a framework to discuss the common traits of NSOCP and NSDP simultaneously, before moving to specific traits. Throughout the whole paper, we will denote the feasible set of (NCP) by $\Omega := \{x \in \mathbb{R}^n \mid g(x) \in \mathcal{K}\}$.

Let us begin with two key ideas that underlie all the results of this paper: *reducibility* and *faces*. Recall from [24, Definition 3.135] that for any given linear spaces $\mathbb{E}$ and $\mathbb{F}$, a cone $\mathcal{K} \subseteq \mathbb{E}$ is said to be *reducible* (more precisely, C^2-*reducible*) at a point $y \in \mathcal{K}$, to a closed convex pointed cone $\mathcal{C} \subseteq \mathbb{F}$, if there exists a neighborhood $\mathcal{N}$ of y and a twice continuously differentiable reduction function $\Xi : \mathcal{N} \to \mathbb{F}$ (possibly depending on y) such that $\Xi(y) = 0$, $D\Xi(y)$ is surjective, and

$$\mathcal{K} \cap \mathcal{N} = \{z \in \mathcal{N} \mid \Xi(z) \in \mathcal{C}\}.$$

In general, reductions are meant to be used as a simplification tool that allows one to interpret any point of $\mathcal{K}$ as a vertex of some other cone $\mathcal{C}$, and then extend the results obtained at $\mathcal{C}$ to $\mathcal{K}$ in a smooth way. In this work, we are also interested in the geometrical properties of the reduced cone $\mathcal{C}$ as well; in particular, in its faces.

To make a brief revision, we recall that F is a *face* of $\mathcal{C}$ if every open line segment that contains a point of F also has its extrema in F; that is, if for every $y \in F$ and every $z, w \in \mathcal{C}$ such that $y = \alpha z + (1 - \alpha) w$ for some $\alpha \in (0, 1)$, we have that $z, w \in F$.

Further, when there exists some $\eta \in \mathcal{C}^\circ$ such that

$$F = \mathcal{C} \cap \{\eta\}^\perp,$$

that is, when F is the intersection between $\mathcal{C}$ and one of its supporting hyperplanes, we say that F is an *exposed face* of $\mathcal{C}$. Some cones, like the nonnegative orthant, the semidefinite cone, and the second-order cone, are *facially exposed*, meaning all of their faces are exposed. We use the notation $F \trianglelefteq \mathcal{C}$ to say that F is a face of $\mathcal{C}$.

Now, to contextualize our results, we will revisit the classical theory of NCP in the next section, with a special emphasis in the work of Guignard [31], and Bonnans, Cominetti, and Shapiro [21]. In particular, we stress some aspects of the NCP theory that are often disregarded in the literature.

2.1 Review of first-order optimality conditions

For any set $S \subseteq \mathbb{E}$ and any $z \in S$, recall the (Bouligand) *tangent cone* to S at z, defined as

$$\mathcal{T}_S(z) := \left\{ y \in \mathbb{E} \;\middle|\; \begin{array}{l} \exists \{t_k\}_{k\in\mathbb{N}} \to 0^+, \exists \{y^k\}_{k\in\mathbb{N}} \to y \text{ such that} \\ z + t_k y^k \in S \text{ for all } k \in \mathbb{N} \end{array} \right\}.$$

Our review of first-order constraint qualifications for (NCP) revolves around two particular cones: the tangent cone $\mathcal{T}_\Omega(\bar{x})$ to Ω at a feasible point $\bar{x} \in \Omega$, and the *linearized tangent cone*

$$\mathcal{L}_\Omega(\bar{x}) := \left\{ d \in \mathbb{R}^n \mid Dg(\bar{x})d \in \mathcal{T}_\mathcal{K}(g(\bar{x})) \right\},$$

where $\mathcal{T}_\mathcal{K}(g(\bar{x}))$ is the tangent cone to $\mathcal{K}$ at $g(\bar{x})$. The importance of these cones for our analyses lies on the necessary optimality conditions for (NCP) associated with them. Namely, given any local minimizer $\bar{x} \in \Omega$ of (NCP), it is easy to see that $\langle \nabla f(\bar{x}), d\rangle \geq 0$ for all $d \in \mathcal{T}_\Omega(\bar{x})$; that is,

$$-\nabla f(\bar{x}) \in \mathcal{T}_\Omega(\bar{x})^\circ. \tag{1}$$

This is one of the simplest necessary optimality conditions, sometimes called the *first-order geometric necessary condition* for the optimality of $\bar{x}$. However, it may be difficult to use (1) when Ω does not admit an explicit characterization because $\mathcal{T}_\Omega(\bar{x})^\circ$ may not be easily computable in this case. The polar of $\mathcal{L}_\Omega(\bar{x})$, on the other hand, admits a practical description, as it is shown in the following lemma, extracted from the proof of [31, Theorem 2] by Guignard:

Lemma 1 *Let* $\bar{x} \in \Omega$. *Then,* $\mathcal{L}_\Omega(\bar{x})^\circ = \mathrm{cl}(H(\bar{x}))$, *where*

$$H(\bar{x}) := Dg(\bar{x})^T \mathcal{N}_\mathcal{K}(g(\bar{x})) = \left\{ Dg(\bar{x})^T z \mid z \in \mathcal{N}_\mathcal{K}(g(\bar{x})) \right\}, \tag{2}$$

and $\mathcal{N}_\mathcal{K}(g(\bar{x})) := \mathcal{T}_\mathcal{K}(g(\bar{x}))^\circ$ *is the* normal cone *to* $\mathcal{K}$ *at* $g(\bar{x})$.

Proof By the bipolar theorem (see e.g. [24, Proposition 2.40]), it suffices to prove that $\mathcal{L}_\Omega(\bar{x}) = H(\bar{x})^\circ$. Take any direction $d \in \mathcal{L}_\Omega(\bar{x})$ and let $z \in \mathcal{T}_\mathcal{K}(g(\bar{x}))^\circ$. By definition, $Dg(\bar{x})d \in \mathcal{T}_\mathcal{K}(g(\bar{x}))$ and then

$$0 \geq \langle Dg(\bar{x})d, z\rangle = \langle d, Dg(\bar{x})^T z\rangle.$$

Thus, since z is arbitrary, we obtain that $d \in H(\bar{x})^\circ$; and since d is also arbitrary, it follows that $\mathcal{L}_\Omega(\bar{x}) \subseteq (H(\bar{x}))^\circ$. Conversely, assume that there exists a vector $v \in H(\bar{x})^\circ$ such that $v \notin \mathcal{L}_\Omega(\bar{x})$, that is, $Dg(\bar{x})v \notin \mathcal{T}_\mathcal{K}(g(\bar{x}))$. By the strong separation theorem (see e.g. [24, Theorem 2.14]), there exists a vector y such that $\langle y, Dg(\bar{x})v\rangle > 0$ and $\langle y, z\rangle < 0$, for all $z \in \mathcal{T}_\mathcal{K}(g(\bar{x}))$, that is, $y \in \mathcal{N}_\mathcal{K}(g(\bar{x}))$. Therefore, $Dg(\bar{x})^T y \in H(\bar{x})$, which is a contradiction with $\langle Dg(\bar{x})^T y, v\rangle > 0$, because $v \in H(\bar{x})^\circ$. □

Recall that because $\mathcal{K}$ is a closed convex cone, we have

$$\mathcal{N}_\mathcal{K}(g(\bar{x})) = \left\{z \in \mathcal{K}^\circ \mid \langle g(\bar{x}), z\rangle = 0\right\}.$$

Then, combining the first-order geometric necessary condition and Lemma 1 yields the following theorem, also by Guignard:

Theorem 1 (Theorem 2 of [31]) *Let $\bar{x} \in \Omega$ be a local minimizer of (NCP). If $\mathcal{T}_\Omega(\bar{x})^\circ = \mathcal{L}_\Omega(\bar{x})^\circ$ and $H(\bar{x})$ is closed, then there exists some $\bar{\lambda} \in \mathcal{K}^\circ$ such that*

$$\nabla f(\bar{x}) + Dg(\bar{x})^T\bar{\lambda} = 0 \quad \text{and} \quad \langle g(\bar{x}), \bar{\lambda}\rangle = 0. \tag{3}$$

Theorem 1 can be seen as the "dual form" of the first-order geometric condition (1), and any vector $\bar{\lambda} \in \mathcal{K}^\circ$ that satisfies the *Karush-Kuhn-Tucker conditions* (3) is called a *Lagrange multiplier* associated with $\bar{x}$. Moreover, the collection of all Lagrange multipliers associated with $\bar{x}$ will be denoted by $\Lambda(\bar{x})$, and when $\Lambda(\bar{x}) \neq \emptyset$ we say that $\bar{x}$ is a *KKT point* of (NCP).

The hypothesis of Theorem 1,

$$\mathcal{T}_\Omega(\bar{x})^\circ = \mathcal{L}_\Omega(\bar{x})^\circ \quad \text{and} \quad H(\bar{x}) \text{ is closed}, \tag{4}$$

is known in the literature as *Guignard's CQ*, and it is the weakest assumption that makes the KKT conditions necessary for the local optimality of $\bar{x}$, in the sense of: if $\Lambda(\bar{x}) \neq \emptyset$ for every continuously differentiable function f that has a local minimizer constrained to Ω at $\bar{x}$, then Guignard's CQ must also hold at $\bar{x}$ [30, Corollary 3.4]. Börgens et al. [25, Definition 5.11] defined Guignard's CQ for optimization problems in Banach spaces as a single equality

$$\mathcal{T}_\Omega(\bar{x})^\circ = H(\bar{x}),$$

which is equivalent to (4) due to Lemma 1. In NLP, Guignard's CQ is usually stated in the form $\mathcal{T}_\Omega(\bar{x})^\circ = \mathcal{L}_\Omega(\bar{x})^\circ$, because the closedness of $H(\bar{x})$ follows from the polyhedricity of $\mathbb{R}^m_+$. However, as it can be seen in the following example, the equality $\mathcal{T}_\Omega(\bar{x})^\circ = \mathcal{L}_\Omega(\bar{x})^\circ$ on its own may not ensure that $\Lambda(\bar{x}) \neq \emptyset$ when $H(\bar{x})$ is not closed.

Example 1 Consider the following problem, presented in [2, Subsection 2.1]:

$$\begin{array}{ll} \text{Minimize} & f(x) := -x_2, \\ \text{s.t.} & g(x) := (x_1, x_1, x_2) \in K_3, \end{array}$$

where K_3 is the three-dimensional second-order cone, given by

$$K_3 = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \geq \sqrt{x_2^2 + x_3^2} \right\}.$$

Note that its feasible set is given by $\Omega = \{x \in \mathbb{R}^2 \mid x_1 \geq 0 \text{ and } x_2 = 0\}$, and that the point $\bar{x} = (0, 0) \in \mathbb{R}^2$ is a local minimizer of it. Any Lagrange multiplier $\lambda := (\lambda_1, \lambda_2, \lambda_3) \in K_3^\circ$ associated with $\bar{x}$ must satisfy

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{5}$$

which implies that $\lambda_3 = 1$ and $\lambda_1 = -\lambda_2$. But because $\lambda \in K_3^\circ = -K_3$ this vector must also satisfy $-\lambda_1 \geq \sqrt{\lambda_1^2 + 1}$, which does not have a solution with $\lambda_3 = 1$ and $\lambda_1 = -\lambda_2$. Therefore, $\bar{x}$ does not satisfy the KKT conditions. However, note that $\mathcal{T}_\Omega(\bar{x}) = \Omega = \mathcal{L}_\Omega(\bar{x})$ and consequently, $\mathcal{T}_\Omega(\bar{x})^\circ = \mathcal{L}_\Omega(\bar{x})^\circ$. Additionally, note that

$$H(\bar{x}) = \{(y_1 + y_2, y_3) \in \mathbb{R}^2 \mid (y_1, y_2, y_3) \in K_3^\circ\}$$

is not closed, because the sequence $\left\{\left(-\frac{1}{k}, -1\right)\right\}_{k \in \mathbb{N}}$ is contained in $H(\bar{x})$ since $\left(-\frac{1}{k} - k, k, -1\right) \in K_3^\circ$, $\forall k \in \mathbb{N}$, but its limit point $(0, -1)$ does not belong to $H(\bar{x})$.

The condition

$$\mathcal{T}_\Omega(\bar{x}) = \mathcal{L}_\Omega(\bar{x}) \text{ and } H(\bar{x}) \text{ is closed}, \tag{6}$$

which implies Guignard's CQ, is known as *Abadie's CQ* (see also Börgens et al. [25, Definition 5.5]), and Example 1 tells us that the closedness of $H(\bar{x})$ cannot be omitted in this case, either. The reason why we emphasize this point is that, as far as we know, it appears that Abadie's CQ and Guignard's CQ are rarely seen in the literature of finite-dimensional conic programming problems other than NLP, and the closedness of $H(\bar{x})$ is rarely regarded in the study of constraint qualifications. In contrast, $H(\bar{x})$ plays an important role in our results.

In finite-dimensional conic contexts, the focus is usually on constraint qualifications that already imply $H(\bar{x})$ is closed without requiring it explicitly, such as Robinson's CQ, that holds at a given point $\bar{x} \in \Omega$ when

$$0 \in \operatorname{int}(\operatorname{Im}(Dg(\bar{x})) - \mathcal{K} + g(\bar{x})). \tag{7}$$

In particular, if $\mathcal{K}$ has nonempty interior, then Robinson's CQ holds at $\bar{x}$ if, and only if, there exists some $d \in \mathbb{R}^n$ such that

$$g(\bar{x}) + Dg(\bar{x})d \in \text{int}(\mathcal{K}).$$

Robinson's CQ is stronger than Abadie's CQ, and it implies that $\Lambda(\bar{x})$, besides being closed and convex, is also nonempty and bounded [24, Theorem 3.9] when $\bar{x}$ is a local minimizer of (NCP). Actually, in this finite-dimensional context, nonempty and boundedness are also sufficient conditions to ensure Robinson's CQ [24, Proposition 3.17]. For this reason, Robinson's CQ is considered the natural analogue of MFCQ in NCP. Moreover, when $\mathcal{K}$ is reducible at the point $g(\bar{x})$ to a cone $\mathcal{C}$ by the reduction function Ξ, Robinson's CQ holds at $\bar{x}$ for the original constraint if, and only if, it holds at the same point for the reduced equivalent constraint $\mathcal{G}(x) \in \mathcal{C}$, with $\mathcal{G} := \Xi \circ g$.

Another well-known constraint qualification in the context of conic programming is the *nondegeneracy* condition, which holds at $\bar{x}$ when

$$\text{Im}(Dg(\bar{x})) + \text{lin}(\mathcal{T}_{\mathcal{K}}(g(\bar{x}))) = \mathbb{E}, \tag{8}$$

where $\text{lin}(\mathcal{T}_{\mathcal{K}}(g(\bar{x}))) = \mathcal{T}_{\mathcal{K}}(g(\bar{x})) \cap -\mathcal{T}_{\mathcal{K}}(g(\bar{x}))$ denotes the largest linear space contained in $\mathcal{T}_{\mathcal{K}}(g(\bar{x}))$; that is, its *lineality space*. This CQ has first appeared in Shapiro and Fan's article [56] for NSDP, by the name *transversality*, and then it was generalized to NCP by Shapiro, in [55]. Nondegeneracy is strictly stronger than Robinson's CQ and it is known that if $\bar{x}$ is a local minimizer of (NCP) that satisfies nondegeneracy, then $\Lambda(\bar{x})$ is a singleton (see, for instance, [24, Proposition 4.75]). Moreover, if $\mathcal{K}$ is reducible, nondegeneracy is equivalent to the surjectivity of $D\mathcal{G}(\bar{x})$, as it can be easily deduced from the equality $\text{lin}(\mathcal{T}_{\mathcal{K}}(g(\bar{x}))) = \text{Ker}(D\Xi(g(\bar{x})))$; see [24, Section 4.6.1].

Due to their implications over the Lagrange multiplier set, nondegeneracy and Robinson's CQ are currently the most important CQs in the study of second-order optimality conditions for (NCP), which will be reviewed in the next subsection.

2.2 Second-order optimality conditions

Before starting, recall that the (inner) second-order tangent set to a nonempty set $S \subseteq \mathbb{E}$, at a point $z \in S$, in a direction $y \in \mathcal{T}_S(z)$, is defined by

$$\mathcal{T}_S^2(z, y) := \left\{ w \in \mathbb{E} \;\middle|\; z + ty + \frac{t^2}{2} w + o(t^2) \in S, \; \forall t > 0 \right\}, \tag{9}$$

which is closed for all such z, y, and S. In addition, if S is convex, then $\mathcal{T}_S^2(z, y)$ is also convex [24, Page 163]; and if S is second-order regular, as it is the case of the semidefinite cone and the second-order cone, then $\mathcal{T}_S^2(z, y)$ is nonempty [24, Page 202].

The role of second-order necessary optimality conditions is to provide additional information when first-order conditions are not meaningful enough; that is, along the directions in the cone

$$C(\bar{x}) := \left\{d \in \mathbb{R}^n \mid d \in \mathcal{T}_\Omega(\bar{x}),\ \langle \nabla f(\bar{x}), d\rangle = 0\right\},$$

which is often called the *cone of critical directions*, or simply, the *critical cone* of (NCP) at $\bar{x}$. Ben-Tal and Zowe [19] presented a *geometric second-order necessary optimality condition* for (NCP), stating that if $\bar{x}$ is a local minimizer of the problem, then

$$\langle \nabla f(\bar{x}), s\rangle + \langle \nabla^2 f(\bar{x})d, d\rangle \geq 0 \tag{10}$$

for every $d \in C(\bar{x})$ and every $s \in \mathcal{T}^2_\Omega(\bar{x}, d)$. Then, Kawasaki [40, Theorem 5.1] made the first advances to derive a “dual form” of (10) under Robinson’s CQ assuming that $\mathcal{K}$ is a closed convex cone with nonempty interior. This result was later generalized and refined by Cominetti [26, Theorem 4.2] to the case where $\mathcal{K}$ is assumed to be a closed convex set. An important improvement was made afterwards by Bonnans, Cominetti, and Shapiro [21], who clarified several key points of the previous works, and obtained no-gap[1] second-order conditions, in particular, for second-order regular cones [21, Section 4]. Let us recall Bonnans, Cominetti, and Shapiro’s necessary condition in the context of second-order regular cones:

Theorem 2 (Theorem 3.1 of [21]) *Let $\bar{x} \in \Omega$ be a local minimizer of (NCP) that satisfies Robinson’s CQ. Then, for every direction $d \in C(\bar{x})$, there exists some $\bar{\lambda}_d \in \Lambda(\bar{x})$, such that*

$$d^T \nabla^2 f(\bar{x})d + \langle D^2 g(\bar{x})[d, d], \bar{\lambda}_d\rangle - \sigma(d, \bar{x}, \bar{\lambda}_d) \geq 0, \tag{11}$$

where

$$\sigma(d, \bar{x}, \bar{\lambda}_d) := \sup\left\{\langle w, \bar{\lambda}_d\rangle \mid w \in \mathcal{T}^2_\mathcal{K}(g(\bar{x}), Dg(\bar{x})d)\right\} \tag{12}$$

is the support function of $\mathcal{T}^2_\mathcal{K}(g(\bar{x}), Dg(\bar{x})d)$ with respect to $\bar{\lambda}_d$.

The term $\sigma(d, \bar{x}, \bar{\lambda}_d)$ characterizes a possible curvature of the set $\mathcal{K}$ at $g(\bar{x})$ along $Dg(\bar{x})d$, and it is often called the “sigma-term” in the classical literature (for instance, in the book [24]). Because $\bar{\lambda}_d \in \Lambda(\bar{x})$ and $\mathcal{K}$ is convex, $\sigma(d, \bar{x}, \bar{\lambda}_d)$ is always non-negative; and if $\mathcal{K}$ is polyhedral, as in NLP, then the sigma-term is zero everywhere. See also the discussion on polyhedricity and extended polyhedricity in [24, Section 3.2.3]. It is also worth mentioning that the second-order optimality condition of Theorem 2 can be derived without constraint qualifications, using Fritz John (generalized) multipliers [24, Theorem 3.50].

[1] The term “zero gap,” or “no gap,” is often used in NLP to refer to a second-order condition that does not require constraint qualifications to be necessary (using Fritz John/generalized Lagrange multipliers), and that becomes sufficient after replacing an inequality by a strict inequality. However, in this paper, we say that a condition has zero gap when it satisfies the latter, possibly subject to a constraint qualification, in the same way as [21].

Although the condition of Theorem 2 is generally considered very natural and useful in the conic programming context and in NLP, a stronger condition where the Lagrange multiplier $\bar{\lambda}$ does not depend on d has several potential uses, in view of the NLP literature. This motivates the following definition:

Definition 1 Let $\bar{x} \in \Omega$ be a KKT point and let $\bar{\lambda} \in \Lambda(\bar{x})$ be given. We say that the pair $(\bar{x}, \bar{\lambda})$ satisfies the *second-order condition* (SOC) when

$$d^T \nabla^2 f(\bar{x})d + \langle D^2 g(\bar{x})[d, d], \bar{\lambda}\rangle - \sigma(d, \bar{x}, \bar{\lambda}) \geq 0, \tag{13}$$

for every $d \in C(\bar{x})$.

In NLP, the existence of some $\bar{\lambda} \in \Lambda(\bar{x})$ such that SOC holds for the pair $(\bar{x}, \bar{\lambda})$ is known as the *semi-strong second-order necessary optimality condition* [20]. Moreover, when SOC holds for every $\bar{\lambda} \in \Lambda(\bar{x})$, then we obtain what is known as the *strong second-order necessary optimality condition* [4]. However, while the condition of Theorem 2 is necessary for optimality under Robinson's CQ, this is not true, in general, for the strong and semi-strong conditions. In fact, there is a counterexample published by Baccari [17, Section 3] (see also Anitescu [14] and Arutyunov [15]), that shows that Robinson's CQ does not guarantee the existence of a $\bar{\lambda} \in \Lambda(\bar{x})$ such that the pair $(\bar{x}, \bar{\lambda})$ satisfies SOC (see also the extended version of [18] for details). Under nondegeneracy, the set $\Lambda(\bar{x})$ is a singleton and, in this case, the semi-strong and the strong second-order conditions both coincide with the condition of Theorem 2.

As far as we know, there is no result concerning the semi-strong and strong second-order conditions without assuming uniqueness of Lagrange multipliers in the literature of conic programming, except for NLP. In NLP, this has been addressed by means of constant rank-type constraint qualifications, which is also the path we will follow in this paper.

3 Revisiting constant rank CQs in NLP

In this section we will revisit some constant rank-type conditions for NLP from a geometrical point of view, in order to extend it to a more general conic context later on. Consider the standard NLP problem

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & g_j(x) \geq 0, \ j = 1, \dots, m, \\ & g_j(x) = 0, \ j = m+1, \dots, m+p, \end{array} \tag{NLP}$$

which is a particular case of (NCP) with $\mathbb{E} = \mathbb{R}^{m+p}$, $\mathcal{K} = \mathbb{R}^m_+ \times \{0\}^p$, and $g(x) := (g_1(x), \dots, g_{m+p}(x))$. As usual in NLP, given a feasible point $\bar{x}$ of (NLP), we will denote the set of active inequality constraints at $\bar{x}$ as $\mathcal{A}(\bar{x}) := \{j \in \{1, \dots, m\} \mid g_j(\bar{x}) = 0\}$.

Now, let us recall Janin's constant rank constraint qualification as it was first presented in [36].

Definition 2 (CRCQ [36]) Let $\bar{x}$ be a feasible point of (NLP). We say that the constant rank constraint qualification for NLP (CRCQ) holds at $\bar{x}$ if there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that, for every subset $J \subseteq \mathcal{A}(\bar{x}) \cup \{m+1, \ldots, m+p\}$, the rank of the family $\{\nabla g_j(x)\}_{j \in J}$ remains constant for all $x \in \mathcal{V}$.

To prove that CRCQ is a constraint qualification, Janin proved that it implies $\mathcal{L}_\Omega(\bar{x}) \subseteq \mathcal{T}_\Omega(\bar{x})$, which in turn implies Abadie's CQ in NLP. His proof is what motivates the requirement to consider every subset J of $\mathcal{A}(\bar{x}) \cup \{m+1, \ldots, m+p\}$ in Definition 2; indeed, after picking a direction

$$d \in \mathcal{L}_\Omega(\bar{x}) = \left\{ d \in \mathbb{R}^n \,\middle|\, \begin{array}{l} \nabla g_j(\bar{x})^T d \geq 0, \;\; j \in \mathcal{A}(\bar{x}), \\ \nabla g_j(\bar{x})^T d = 0, \;\; j \in \{m+1, \ldots, m+p\} \end{array} \right\},$$

in order to prove that $d \in \mathcal{T}_\Omega(\bar{x})$, it is sufficient to have the constant rank assumption for the constraints that correspond to the indices $j \in \mathcal{A}(\bar{x})$ such that $\nabla g_j(\bar{x})^T d = 0$. Since those indices depend on d, and they are not determined *a priori*, one considers all possibilities. However, as it was noted years later by Minchenko and Stakhovski [46], taking subsets of the equality constraints is quite superfluous. This enhanced definition of CRCQ that ignores proper subsets of indices of equality constraints was presented in [46] as follows:

Definition 3 (RCRCQ [46]) Let $\bar{x}$ be a feasible point of (NLP). We say that relaxed constant rank constraint qualification for NLP (RCRCQ) holds at $\bar{x}$ if there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that, for every subset $J \subseteq \mathcal{A}(\bar{x})$, the rank of the family $\{\nabla g_j(x)\}_{j \in J \cup \{m+1, \ldots, m+p\}}$ remains constant for all $x \in \mathcal{V}$.

In order to bring these CQs to the conic setting, our approach in this manuscript consists first in generalizing two key ideas of NLP: the notion of "active constraints" and the notion of "subsets of indices of active constraints." The former can be interpreted in the general context as a consequence of reducibility. Indeed, for any given $\bar{x} \in \Omega$, let $s := |\mathcal{A}(\bar{x})|$ and note that $\mathbb{R}^m_+ \times \{0\}^p$ is reducible at $g(\bar{x})$ to the cone

$$\mathcal{C} := \mathbb{R}^s_+ \times \{0\}^p$$

in a neighborhood $\mathcal{N}$ of $g(\bar{x})$ by the mapping $\Xi : \mathcal{N} \to \mathbb{R}^{s+p}$ such that

$$\Xi(y) := (y_j)_{j \in \mathcal{A}(\bar{x}) \cup \{m+1, \ldots, m+p\}}$$

for every $y \in \mathcal{N}$, and in this case the reduced constraint function of (NLP) at $\bar{x}$ takes the form

$$\mathcal{G}(x) := \Xi(g(x)) = (g_j(x))_{j \in \mathcal{A}(\bar{x}) \cup \{m+1, \ldots, m+p\}}. \tag{14}$$

Therefore, in NLP, reducing the problem is essentially the same as simply disregarding inactive constraints around the point $\bar{x}$. The notion of "subsets of indices of the active constraints," on the other hand, can be interpreted in terms of faces.

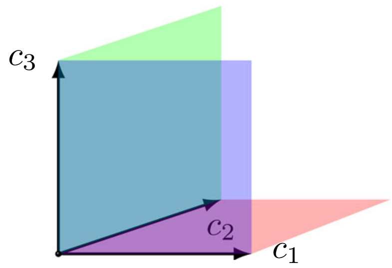

Fig. 1 Faces of $\mathbb{R}^3_+$

It is easy to see that every face of $\mathbb{R}^s_+$ can be written in terms of a unique subset of the canonical vectors of $\mathbb{R}^s$, which we will denote by $c_1, \ldots, c_s$. That is, $F \trianglelefteq \mathbb{R}^s_+$ if, and only if, there exists some $J \subseteq \{1, \ldots, s\}$ such that

$$F = \mathbb{R}^s_+ \bigcap_{j \in J} \{c_i\}^\perp, \tag{15}$$

where F and J are clearly in a one-to-one correspondence.

For example, in Fig. 1, the vertex of $\mathbb{R}^3_+$ corresponds to $J = \{1, 2, 3\}$; the one-dimensional faces cone(c_1), cone(c_2), and cone(c_3) correspond to $J = \{2, 3\}$, $J = \{1, 3\}$, and $J = \{1, 2\}$, respectively; the left, front, and bottom two-dimensional faces correspond to $J = \{1\}$, $J = \{2\}$, and $J = \{3\}$, respectively; and $\mathbb{R}^3_+$ itself corresponds to $J = \emptyset$.

Thus, considering all subsets of active constraints at $\bar{x}$ is the same as considering all faces of the reduced cone $\mathcal{C} = \mathbb{R}^s_+ \times \{0\}^p$. This discussion suggests a natural characterization of RCRCQ in terms of the faces of the reduced cone, as follows:

Proposition 1 *Let $\bar{x}$ be a feasible point of (NLP). Then, RCRCQ holds at $\bar{x}$ if, and only if, there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that, for each $F \trianglelefteq \mathbb{R}^{|\mathcal{A}(\bar{x})|}_+ \times \{0\}^p$, the dimension of*

$$D\mathcal{G}(x)^T[F^\perp]$$

remains constant for every $x \in \mathcal{V}$, where $\mathcal{G}$ is as defined in (14).

Proof Let $s := |\mathcal{A}(\bar{x})|$ and, without loss of generality, let us assume that $\mathcal{A}(\bar{x}) = \{1, \ldots, s\}$. Moreover, let $c_1, \ldots, c_{s+p}$ be the canonical basis of $\mathbb{R}^{s+p}$, and let $F \trianglelefteq \mathbb{R}^s_+ \times \{0\}^p$. Note that $F = R \times \{0\}^p$, where $R \trianglelefteq \mathbb{R}^s_+$. Then, there exists some $J \subseteq \{1, \ldots, s\}$ such that

$$F = \left(\mathbb{R}^s_+ \bigcap_{j \in J} \{c_i\}^\perp \right) \times \{0\}^p,$$

which implies

$$F^\perp = R^\perp \times \mathbb{R}^p = \operatorname{span}\left(\{c_j \mid j \in J \cup \{s+1, \ldots, s+p\}\} \right),$$

so

$$
\begin{aligned}
D\mathcal{G}(x)^T[F^{\perp}] &= \operatorname{span}(\{D\mathcal{G}(x)^T c_j\}_{j\in J\cup\{s+1,\dots,s+p\}}) \\
&= \operatorname{span}(\{\nabla g_j(x)\}_{j\in J\cup\{m+1,\dots,m+p\}}).
\end{aligned} \tag{16}
$$

Consequently,

$$
\dim(Dg(x)^T[F^{\perp}]) = \operatorname{rank}(\{\nabla g_j(x)\}_{j\in J\cup\{m+1,\dots,m+p\}}).
$$

The conclusion follows from the one-to-one correspondence between F and J. □

The equivalent form of RCRCQ presented in Proposition 1 allows us to visualize what it actually describes, geometrically. Indeed, recall that $\mathbb{R}^n = D\mathcal{G}(x)^{-1}(\operatorname{span}(F)) + (D\mathcal{G}(x)^{-1}(\operatorname{span}(F)))^{\perp}$ and it is elementary to see that

$$
(D\mathcal{G}(x)^{-1}(\operatorname{span}(F)))^{\perp} = D\mathcal{G}(x)^T[F^{\perp}].
$$

This implies the following relation:

$$
\dim(D\mathcal{G}(x)^{-1}(\operatorname{span}(F))) + \dim(D\mathcal{G}(x)^T[F^{\perp}]) = n.
$$

Thus, RCRCQ can be equivalently stated as the constant dimension of $D\mathcal{G}(x)^{-1}$ $(\operatorname{span}(F))$ for every $x \in \mathcal{V}$ at each $F \trianglelefteq \mathcal{C} = \mathbb{R}_+^{|\mathcal{A}(\bar{x})|} \times \{0\}^p$. The set $D\mathcal{G}(x)^{-1}(\operatorname{span}(F))$, on the other hand, can be regarded as a "linear approximation" of $\mathcal{G}^{-1}(\mathcal{C})$ around $\bar{x}$. Indeed, $D\mathcal{G}(x)$ is the best linear approximation of $\mathcal{G}$ at $x \in \mathcal{V}$ and, similarly, the faces of $\mathcal{C}$ can also be seen as "linear approximations" of it at $\mathcal{G}(\bar{x})$. In fact, each face induces a potentially different linear approximation of $\mathcal{G}^{-1}(\mathcal{C})$, which in turn coincides with Ω around $\bar{x}$. So roughly speaking: RCRCQ holds at $\bar{x}$ when the dimension of every linear approximation of the feasible set Ω at $\bar{x}$ is invariant to small perturbations. In particular, defining $g_J(x) := (g_j(x))_{j\in J\cup\{m+1,\dots,m+p\}}$ for every $J \subseteq \mathcal{A}(\bar{x})$, this characterization is equivalent to the constant dimension of $\operatorname{Ker}(Dg_J(x))$ for all x in a neighborhood of $\bar{x}$ at every $J \subseteq \mathcal{A}(\bar{x})$, which can also be trivially seen from the original definition of RCRCQ.

Note that the characterization of RCRCQ from Proposition 1 and the discussion above do not appear to be limited to the context of NLP, contrary to its original definition. In the next two sections, we will prove that the same idea can be applied to NSOCP and NSDP, respectively, giving rise to new constraint qualifications.

Remark 1 It is possible to obtain a characterization of CRCQ in the same style of Proposition 1. To do this, it suffices to reformulate the equality constraints $g_j(x) = 0$ as a pair of inequality constraints $g_j(x) \geq 0$ and $-g_j(x) \geq 0$, for $j \in \{m+1,\dots,m+p\}$. That is, consider $\mathcal{K} := \mathbb{R}_+^m \times \mathbb{R}_+^p \times \mathbb{R}_+^p$ and $g(x) := (g_1(x),\dots,g_{m+p}(x), -g_{m+1}(x),\dots,-g_{m+p}(x))$ in Proposition 1.

In view of Remark 1, we see that there are multiple ways of dealing with equality constraints in our approach, and they are not all equivalent. The suitability of each approach may depend on the application, but we highlight that our approach is able to deal with equality constraints regardless of how they are modelled. For simplicity, equality constraints are omitted in our exposition. See also Remarks 3 and 7. In the following two sections, we extend the ideas of this section to NSOCP and NSDP.

4 Nonlinear second-order cone programming

In this section, we consider the following problem:

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & g_j(x) \in K_{m_j},\ j = 1, \dots, q, \end{array} \tag{NSOCP}$$

where $K_{m_j} := \{(z_0, \widehat{z}) \in \mathbb{R} \times \mathbb{R}^{m_j - 1} \mid z_0 \geq \|\widehat{z}\|\}$ when $m_j > 1$ and $K_1 = \{x \in \mathbb{R} \mid x \geq 0\}$. Since K_{m_j} is self-dual, we have that $z \in K^{\circ}_{m_j}$ if, and only if, $-z \in K_{m_j}$, for any $j = 1, \dots, q$. Also, note that (NSOCP) can be seen as a particular case of (NCP) with

$$\mathcal{K} := K_{m_1} \times \dots \times K_{m_q} \quad \text{and} \quad g(x) := (g_1(x), \dots, g_q(x)).$$

Given a feasible point $\bar{x} \in \Omega$, let us define the following index sets:

$$\begin{aligned} I_{\text{int}}(\bar{x}) &:= \{j \in \{1, \dots, q\} \mid g_j(\bar{x}) \in \text{int}(K_{m_j})\}, \\ I_B(\bar{x}) &:= \{j \in \{1, \dots, q\} \mid g_j(\bar{x}) \in \text{bd}^+(K_{m_j})\}, \\ I_0(\bar{x}) &:= \{j \in \{1, \dots, q\} \mid g_j(\bar{x}) = 0\}, \end{aligned}$$

which consist of the indices of the constraints that hit the interior, the boundary excluding zero, and the vertex of their respective cones. For simplicity, we will omit equality constraints; we should mention, nevertheless, that our results can be easily adapted to deal with equality constraints — see Remark 3 for details. As another measure to avoid cumbersome notation, we will assume that $I_B(\bar{x}) = \{1, \dots, |I_B(\bar{x})|\}$; this assumption will often be recalled throughout this section.

Following Bonnans and Ramírez [22], for any given $\bar{x} \in \Omega$, we see that $\mathcal{K}$ is reducible to

$$\mathcal{C} := \prod_{j \in I_0(\bar{x})} K_{m_j} \times \mathbb{R}_+^{|I_B(\bar{x})|} \tag{17}$$

in a neighborhood $\mathcal{N}_1 \times \dots \times \mathcal{N}_q$ of $g(\bar{x})$ by the function $\Xi := (\Xi_j)_{j \in I_0(\bar{x}) \cup I_B(\bar{x})}$, where $\Xi_j \colon \mathcal{N}_j \to \mathbb{R}^{m_j}$ is the identity function for every $j \in I_0(\bar{x})$, and $\Xi_j \colon \mathcal{N}_j \to \mathbb{R}$ is given by

$$\Xi_j(y) := y_0 - \|\widehat{y}\| \tag{18}$$

for every $j \in I_B(\bar{x})$, and every $y \in \mathbb{R}^{m_j}$. This leaves us with the reduced constraint

$$\mathcal{G}(x) \in \mathcal{C},$$

where $\mathcal{G}(x) := \Xi(g(x)) = (\mathcal{G}_j(x))_{j \in I_0(\bar{x}) \cup I_B(\bar{x})}$,

$$\mathcal{G}_j(x) := \Xi_j(g_j(x)) = \begin{cases} g_j(x), & \text{if } \ j \in I_0(\bar{x}), \\ \phi_j(x), & \text{if } \ j \in I_B(\bar{x}), \end{cases} \tag{19}$$

and $\phi \colon \mathbb{R}^n \to \mathbb{R}^{|I_B(\bar{x})|}$ has its j-th component given by

$$\phi_j(\bar{x}) := [g_j(x)]_0 - \|\widehat{g_j(\bar{x})}\|. \tag{20}$$

Note that $g(x) \in \mathcal{K}$ if, and only if, $\mathcal{G}(x) \in \mathcal{C}$ for every x sufficiently close to $\bar{x}$.

By [22, Lemma 25], we see that the linearized cone of the original constraints of (NSOCP) at a given $\bar{x} \in \Omega$ can be computed as

$$\mathcal{L}_\Omega(\bar{x}) = \left\{ \ d \in \mathbb{R}^n \ \middle| \ \begin{array}{ll} Dg_j(x)d \in K_{m_j}, & j \in I_0(\bar{x}) \\ D\phi(x)d \in \mathbb{R}_+^{|I_B(\bar{x})|} & \end{array} \right\}, \tag{21}$$

and that it coincides with the linearized cone of the reduced constraint at $\bar{x}$. Moreover, it follows from [1, Lemma 15] that for each $j = I_{\text{int}}(\bar{x}) \cup I_B(\bar{x})$, we have $\langle \bar{\lambda}_j, g_j(\bar{x}) \rangle = 0$, if, and only if,

$$\bar{\lambda}_j = \begin{cases} 0, & \text{if } j \in I_{\text{int}}(\bar{x}), \\ \frac{[\bar{\lambda}_j]_0}{[g_j(\bar{x})]_0} R_{m_j} g_j(\bar{x}), & \text{if } j \in I_B(\bar{x}), \end{cases} \tag{22}$$

where R_{m_j} is a matrix defined as

$$R_{m_j} := \begin{bmatrix} 1 & 0 \\ 0 & -\mathbb{I}_{m_j - 1} \end{bmatrix}, \tag{23}$$

and $\mathbb{I}_{m_j-1}$ is the $(m_j - 1) \times (m_j - 1)$ identity matrix. Therefore, still following [22], the point $\bar{x}$ satisfies the KKT conditions with respect to the constraint $g(x) \in \mathcal{K}$ if, and only if, there exist some vectors $\bar{\lambda}_j \in K_{m_j}^\circ$, $j \in I_0(\bar{x}) \cup I_B(\bar{x})$, such that:

$$\nabla f(\bar{x}) + \sum_{j \in I_0(\bar{x})} Dg_j(\bar{x})^T \bar{\lambda}_j + \sum_{j \in I_B(\bar{x})} \frac{[\bar{\lambda}_j]_0}{[g_j(\bar{x})]_0} Dg_j(\bar{x})^T R_{m_j} g_j(\bar{x}) = 0, \tag{24}$$

which also coincides with the KKT conditions with respect to the reduced constraint $\mathcal{G}(x) \in \mathcal{C}$. In fact, note that for each $j \in I_B(\bar{x})$, the reduced Lagrange multiplier with respect to the reduced constraint $\phi_j(x) \geq 0$ is simply $[\bar{\lambda}_j]_0$.

With this in mind, we are ready to present our extension of CRCQ (and RCRCQ) to NSOCP inspired by the characterization of Proposition 1.

4.1 A facial constant rank constraint qualification for NSOCP

Recall that, for each $j = 1, \ldots, q$, the cone K_{m_j} is facially exposed, meaning every $F \trianglelefteq K_{m_j}$ can be written as the intersection of one of its supporting hyperplanes, say $\{\eta\}^{\perp}$ with $\eta \in K_{m_j}$. In fact, although K_{m_j} has infinitely many faces when $m_j > 2$, they are limited to only three types:

- The vertex, $\{0\}$, which can be characterized by any $\eta \in \text{int}(K_{m_j})$;
- The cone K_{m_j} itself, which is characterized by $\eta = 0$;
- A ray at the boundary of K_{m_j}, starting at the vertex and passing through a point $z \in \text{bd}^{+}(K_{m_j})$, which can be written in terms of any vector $\eta \in \text{cone}(R_{m_j} z) \setminus \{0\}$.

Moreover, every $F \trianglelefteq \mathcal{C}$ has the form

$$F = \left(\prod_{j \in I_0(\bar{x})} F_j \right) \times R,$$

where $F_j \trianglelefteq K_{m_j}$ for every $j \in I_0(\bar{x})$, and $R \trianglelefteq \mathbb{R}_{+}^{|I_B(\bar{x})|}$. Then, for every $x \in \mathbb{R}^n$, sufficiently close to $\bar{x}$, we have

$$D\mathcal{G}(x)^T [F^{\perp}] = \sum_{j \in I_0(\bar{x})} Dg_j(x)^T [F_j^{\perp}] + D\phi(x)^T [R^{\perp}],$$

where $\phi(x) := (\phi_j(x))_{j \in I_B(\bar{x})}$. This motivates the following definition:

Definition 4 Let $\bar{x}$ be a feasible point of (NSOCP). We say that the *facial constant rank* (FCR) property holds at $\bar{x}$ if there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that for each $F \trianglelefteq \mathcal{C}$, the dimension of $D\mathcal{G}(x)^T [F^{\perp}]$ remains constant for all $x \in \mathcal{V}$, where $\mathcal{G}$ is given by (19) and $\mathcal{C}$ is given by (17).

Recall the discussion after Proposition 1 and note that Definition 4 can be equivalently stated in terms of the constant dimension of $D\mathcal{G}(x)^{-1}(\text{span}(F))$ for all $x \in \mathcal{V}$ and every $F \trianglelefteq \mathcal{C}$. That is, the FCR property holds at $\bar{x}$ when the dimension of every linear approximation of the feasible set remains locally invariant around $\bar{x}$. Although this characterization is somewhat more intuitive than Definition 5, the latter is easier to use.

The FCR property is sufficient for the equality $\mathcal{T}_{\Omega}(\bar{x}) = \mathcal{L}_{\Omega}(\bar{x})$ to hold. To prove this, we employ the main result of Janin's paper [36], but the version we use is a slightly different characterization found in [4, Proposition 3.1]. Despite the fact we work in a context more general than NLP, we use the same result that was used in NLP.

Proposition 2 ([4, Proposition 3.1]) *Let $\{\zeta_i(x)\}_{i \in \mathcal{I}}$ be a finite family of twice continuously differentiable functions $\zeta_i : \mathbb{R}^n \to \mathbb{R}$, $i \in \mathcal{I}$, such that the family of its gradients $\{\nabla \zeta_i(x)\}_{i \in \mathcal{I}}$ remains with constant rank in a neighborhood of $\bar{x}$, and consider the linear subspace*

$$\mathcal{S} := \left\{ y \in \mathbb{R}^n \mid \langle \nabla \zeta_i(\bar{x}), y \rangle = 0, \; i \in \mathcal{I} \right\}.$$

Then, there exists some neighborhoods V_1 and V_2 of $\bar{x}$, and a diffeomorphism $\psi : V_1 \to V_2$, such that:

(i) $\psi(\bar{x}) = \bar{x}$;
(ii) $D\psi(\bar{x}) = \mathbb{I}_n$;
(iii) $\zeta_i(\psi^{-1}(\bar{x}+y)) = \zeta_i(\psi^{-1}(\bar{x}))$ *for every* $y \in \mathcal{S} \cap (V_2 - \bar{x})$ *and every* $i \in \mathcal{I}$.

Moreover, the degree of differentiability of ψ is the same as of ζ_i, for all $i \in \mathcal{I}$.

For the last part of the above proposition, about the degree of differentiability of ψ, we refer to Minchenko and Stakhovski [47, Page 328]. Now, we are able to prove the main result of this section:

Theorem 3 *Let $\bar{x}$ be a feasible point of (NSOCP). If the FCR property holds at $\bar{x}$, then $\mathcal{T}_\Omega(\bar{x}) = \mathcal{L}_\Omega(\bar{x})$.*

Proof It suffices to show that $\mathcal{L}_\Omega(\bar{x}) \subseteq \mathcal{T}_\Omega(\bar{x})$. Let $d \in \mathcal{L}_\Omega(\bar{x})$ and suppose that $\bar{x}$ satisfies the FCR property. Let

$$F := \left(\prod_{j \in I_0(\bar{x})} F_j \right) \times R, \tag{25}$$

where $F_j \trianglelefteq K_{m_j}$, $j \in I_0(\bar{x})$, are defined as

$$F_j := \begin{cases} K_{m_j} & \text{if } Dg_j(\bar{x})d \in \operatorname{int}(K_{m_j}), \\ \operatorname{cone}(Dg_j(\bar{x})d), & \text{if } Dg_j(\bar{x})d \in \operatorname{bd}^+(K_{m_j}), \\ \{0\}, & \text{if } Dg_j(\bar{x})d = 0. \end{cases} \tag{26}$$

and $R \trianglelefteq \mathbb{R}^{|I_B(\bar{x})|}$ is given by

$$R := \mathbb{R}_+^{|I_B(\bar{x})|} \bigcap_{j \in J} \{c_j\}^\perp, \tag{27}$$

where c_j is the j-th vector of the canonical basis of $\mathbb{R}^{|I_B(\bar{x})|}$, and $J := \{j \in I_B(\bar{x}) \mid \nabla\phi_j(\bar{x})^T d = 0\}$. Recall that we are assuming for simplicity that $I_B(\bar{x}) = \{1, \dots, |I_B(\bar{x})|\}$, and note that $D\mathcal{G}(\bar{x})d \in F$.

Now, for every $j \in I_0(\bar{x})$ such that $Dg_j(\bar{x})d \in \operatorname{bd}^+(K_{m_j})$, let $A_j \in \mathbb{R}^{m_j \times m_j - 1}$ be any matrix with full column rank such that $\operatorname{Im}(A_j) = \{Dg_j(\bar{x})d\}^\perp$, and observe that

$$Dg_j(x)^T[F_j^\perp] = \operatorname{span}\left(\left\{ Dg_j(x)^T A_j^i \right\}_{i=1,\dots,m_j-1} \right)$$

for every such j, where A_j^i denotes the i-th column of A_j. Similarly, for every $j \in I_0(\bar{x})$ such that $Dg_j(\bar{x})d = 0$, we have

$$Dg_j(x)^T[F_j^\perp] = \operatorname{span}(\{\nabla g_{j,i}(x)\}_{i=0,\dots,m_j-1}),$$

where $\nabla g_{j,i}(x)$ denotes the i-th column of $Dg_j(x)^T$. And for every j such that $Dg_j(\bar{x})d \in \operatorname{int}(K_{m_j})$, we have $Dg_j(x)^T[F_j^\perp] = \{0\}$. Finally, observe that $R^\perp = \operatorname{span}(\{c_j\}_{j\in J})$ and then

$$D\phi(x)^T[R^\perp] = \operatorname{span}\left(\left\{\nabla\phi_j(x)\right\}_{j\in J}\right).$$

Therefore, for every $x \in \mathcal{V}$, where $\mathcal{V}$ is the neighborhood of $\bar{x}$ given by Definition 4, the linear space

$$D\mathcal{G}(x)^T[F^\perp] = \sum_{j\in I_0(\bar{x})} Dg_j(x)^T[F_j^\perp] + D\phi(x)^T[R^\perp] \tag{28}$$

is generated by the family of vectors:

$$\bigcup_{\substack{j\in I_0(\bar{x})\\ Dg_j(\bar{x})d\in \operatorname{bd}^+(K_{m_j})\\ i=1,\dots,m_j-1}} \left\{Dg_j(x)^T A_j^i\right\} \bigcup_{\substack{j\in I_0(\bar{x})\\ Dg_j(\bar{x})d=0\\ i=0,\dots,m_j-1}} \{\nabla g_{j,i}(x)\} \bigcup_{j\in J}\{\nabla\phi_j(x)\}, \tag{29}$$

which implies that the dimension of (28) equals the rank of (29), for every $x \in \mathcal{V}$. Since this dimension remains constant in $\mathcal{V}$, so does the rank of (29). This means we can apply Proposition 2 to the family of functions

$$\zeta_{i,j}(x) := \begin{cases} \langle A_j^i, g_j(x)\rangle, & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d \in \operatorname{bd}^+(K_{m_j}),\ i = 1,\dots,m_j-1,\\ g_{j,i}(x), & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d = 0,\ i = 0,\dots,m_j-1,\\ \phi_j(x), & \text{if } j \in J, \end{cases} \tag{30}$$

where $g_{j,i}(x)$ denotes the i-th entry of $g_j(x)$ for $j \in J$. Then, consider the following subspace:

$$\mathcal{S} := \left\{ y \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} A_j^T Dg_j(\bar{x})y = 0, & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d \in \operatorname{bd}^+(K_{m_j})\\ Dg_j(\bar{x})y = 0, & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d = 0\\ \nabla\phi_j(\bar{x})^T y = 0, & \text{if } j \in J, \end{array} \right\},$$

and note that $d \in \mathcal{S}$, so it follows that there exists a local diffeomorphism ψ for which items (i), (ii) and (iii) of Proposition 2 are satisfied. Now, define the arc $\xi(t)$ by

$$\xi(t) := \psi^{-1}(\bar{x} + td),$$

for $t \in \mathbb{R}$ small enough so that $\bar{x} + td \in V_2$, where V_2 is given by Proposition 2. Then, we obtain that

$$\lim_{t\to 0^+} \xi(t) = \bar{x}, \quad \lim_{t\to 0^+} \frac{\xi(t) - \bar{x}}{t} = d.$$

To complete the proof, it suffices to show that $\xi(t)$ remains feasible for every sufficiently small $t \geq 0$, so this is our goal from this point onwards. Proposition 2 tells us that there exists some $\varepsilon > 0$ such that $\zeta_{i,j}(\xi(t)) = \zeta_{i,j}(\bar{x}) = 0$ for every $t \in (-\varepsilon, \varepsilon)$. In terms of F, this means that

$$\mathcal{G}(\xi(t)) \in \operatorname{span}(F)$$

for every such t, which follows directly from (30). Now, let us analyse each case separately:

1. For each index $j \in I_0(\bar{x})$, consider the Taylor expansion of $g_j(\xi(t))$ around $t = 0$, given by

$$\begin{aligned} g_j(\xi(t)) &= g_j(\xi(0)) + tDg_j(\xi(0))\xi'(0) + o(t) \\ &= g_j(\bar{x}) + tDg_j(\bar{x})D\psi^{-1}(\bar{x})d + o(t) \\ &= tDg_j(\bar{x})d + o(t) \end{aligned} \tag{31}$$

   Then, we split in three sub-cases:

   - If $Dg_j(\bar{x})d \in \operatorname{int}(K_{m_j})$, then it follows from (31) that $g_j(\xi(t)) \in K_{m_j}$ for every $t \in [0, \varepsilon)$, shrinking ε if necessary;
   - If $Dg_j(\bar{x})d \in \operatorname{bd}^+(K_{m_j})$, then $g_j(\xi(t)) \in \operatorname{span}(Dg_j(\bar{x})d)$ due to (26), and it follows from (31) that $g_j(\xi(t)) \in \operatorname{cone}(Dg_j(\bar{x})d)$ for every $t \in [0, \varepsilon)$, taking a smaller ε if needed;
   - If $Dg_j(\bar{x})d = 0$, then $g_j(\xi(t)) = 0$ for every $t \in [0, \varepsilon)$, due to (26).

2. Because $\phi(\xi(t)) \in R$ for every $t \in [0, \varepsilon)$, for each index $j \in J$, we have $\phi_j(\xi(t)) = 0$. On the other hand, for each $j \notin J$, consider the Taylor expansion of $\phi_j(\xi(t))$ around $t = 0$:

$$\phi_j(\xi(t)) = \phi_j(\xi(0)) + t\nabla\phi_j(\xi(0))^T\xi'(0) + o(t) = t\nabla\phi_j(\bar{x})^T d + o(t),$$

   and since $\nabla\phi_j(\bar{x})^T d > 0$ for every $j \notin J$, it also follows that $\phi_j(\xi(t)) > 0$ for every $t \in (0, \varepsilon)$, taking a smaller ε if necessary.

Thus, $\mathcal{G}(\xi(t)) \in F$ for every $t \in [0, \varepsilon)$, which also implies that $g(\xi(t)) \in \mathcal{K}$ for every such t, completing the proof. □

A useful information that can be extracted from the proof above is an equivalent characterization of the FCR property (Definition 4) without faces:

Corollary 1 *Let $\bar{x} \in \Omega$. Then, the FCR property holds at $\bar{x}$ if, and only if, there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that: for all subsets $J_1, J_2 \subseteq I_0(\bar{x})$, $J_3 \subseteq I_B(\bar{x})$, such that $m_j > 1$ for all $j \in J_1$, and for all $\eta_j \in \operatorname{bd}^+(K_{m_j})$, $j \in J_1$, the rank of the family*

$$\bigcup_{\substack{j \in J_1 \\ i=1,\ldots,m_j}} \left\{ Dg_j(x)^T A_j^i \right\} \bigcup_{\substack{j \in J_2 \\ i=0,\ldots,m_j-1}} \{\nabla g_{j,i}(x)\} \bigcup_{j \in J_3} \{\nabla\phi_j(x)\}.$$

remains the same for all $x \in \mathcal{V}$*, where* $A_j \in \mathbb{R}^{m_j \times m_j - 1}$ *can be any matrix with full column rank such that* $\text{Im}(A_j) = \{\eta_j\}^{\perp}$*, for each* $j \in J_1$*, and* A_j^i *denotes the* i*-th column of* A_j*.*

Notice that if J_1 is fixed as the empty set, then the characterization of Corollary 1 recovers the CRCQ proposal of [60]. This clarifies that the matrices A_j, $j \in J_1$, were the missing ingredients for the proposal of [60] to be a CQ. Before proceeding, we will make a short discussion about Theorem 3 and its implications:

Remark 2 Note that if all constraints are affine, then every feasible point satisfies the FCR property. Then, it follows from Theorem 4 that $\mathcal{T}_{\Omega}(\bar{x}) = \mathcal{L}_{\Omega}(\bar{x})$ in this case, for every $\bar{x} \in \Omega$. We highlight this fact because when it is paired with Example 1, two things can be concluded: first, the FCR property alone is not a CQ for (NSOCP); second, the only reason why constraint linearity is not a CQ for NSOCP is that $H(\bar{x})$ may not be closed. When $H(\bar{x})$ is closed, FCR is a CQ, and so is constraint linearity. In other words, the above discussion, in view of the minimality of Guignard's CQ, allows us to conclude that the closedness of $H(\bar{x})$ is the weakest CQ for linear second-order cone programming problems.

The discussion of Remark 2, together with Theorem 4, motivates our extension of CRCQ (and RCRCQ) to NSOCP:

Definition 5 Let $\bar{x}$ be a feasible point of (NSOCP) and let $H(\bar{x})$ be the set defined in (2). We say that the *constant rank constraint qualification for NSOCP* (CRCQ) holds at $\bar{x}$, if it satisfies the FCR property and, in addition, the set $H(\bar{x})$ is closed.

When $m_1 = m_2 = \ldots = m_q = 1$, problem (NSOCP) reduces to a NLP problem. Moreover, since the faces of K_1 are $\{0\}$ and $\mathbb{R}_+$, the FCR property (Definition 4) reduces to CRCQ in this case, and so does Definition 5. Moreover, as mentioned before, it follows directly from Theorem 3, that:

Theorem 4 *The CRCQ condition of Definition 5 implies Abadie's CQ.*

Since the nondegeneracy condition for (NSOCP) holds at a given $\bar{x} \in \Omega$ if, and only if, $D\mathcal{G}(\bar{x})^T$ is injective, then by continuity of $D\mathcal{G}$, nondegeneracy implies that $D\mathcal{G}(x)^T$ remains injective for every x close enough to $\bar{x}$. Therefore, it follows that the nondegeneracy condition implies CRCQ as in Definition 5. However, the converse is not true, as it can be seen in the following example:

Example 2 Consider the following constraint

$$g(x) := (x, x) \in K_2,$$

at the feasible point $\bar{x} = 0$. The cone K_2 is polyhedral and g is linear, then CRCQ as in Definition 5 holds at $\bar{x}$. However, Robinson's CQ is not satisfied at $\bar{x} = 0$, since

$$Dg(\bar{x})d = d(1, 1) \notin \text{int}(K_2)$$

for every $d \in \mathbb{R}$. Consequently, nondegeneracy is not satisfied, either.

Observe that Example 2 also shows that CRCQ does not imply Robinson's CQ. Conversely, Robinson's CQ does not imply CRCQ either, meaning they are not related, just as it happens with CRCQ and MFCQ in NLP. Let us show this with an example:

Example 3 Consider the constraint:

$$g(x) := (x_2, x_1^2) \in K_2$$

at the point $\bar{x} = (0, 0)$. Robinson's CQ holds at $\bar{x}$, since $d = (0, 1)$ satisfies

$$g(\bar{x}) + Dg(\bar{x})d = (1, 0) \in \operatorname{int}(K_2).$$

On the other hand, take the face $F = \{0\}$ and note that

$$Dg(x)^T[F^\perp] = \operatorname{span}\left(\left\{\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}2x_1\\0\end{bmatrix}\right\}\right)$$

has dimension 2 for every x such that $x_1 \neq 0$, and dimension 1 at $\bar{x}$.

Remark 3 To consider (NSOCP) with an equality constraint in the form $h(x) = 0$, where $h\colon \mathbb{R}^n \to \mathbb{R}^p$, one should proceed as in Proposition 1. That is, consider

$$g(x) := (g_1(x), \ldots, g_q(x), h(x))$$

and the cone

$$\mathcal{K} := K_{m_1} \times \ldots \times K_{m_q} \times \{0\}^p.$$

This will lead to an extension of RCRCQ. An extension of the original CRCQ condition can be obtained by writing the equality constraint as a pair of inequality constraints in the form $h(x) \in \mathbb{R}^p_+$ and $-h(x) \in \mathbb{R}^p_+$, just as in Remark 1, then reducing, and applying Definition 5 to the new reduced cone.

4.2 Strong second-order optimality conditions for NSOCP

In this subsection we will investigate second-order optimality conditions for (NSOCP) under the FCR property; and, consequently, under CRCQ as well. Recall that the second-order condition of Definition 1 can be further specialized to the context of NSOCP by characterizing the sigma-term explicitly. Following Bonnans and Ramírez [22], we have for any $\bar{x} \in \Omega$ and any of its associate Lagrange multipliers $\bar{\lambda} := (\bar{\lambda}_1, \ldots, \bar{\lambda}_q) \in \Lambda(\bar{x})$, that

$$\sigma(d, \bar{x}, \bar{\lambda}) = \sum_{j=1}^{q} d^T \mathcal{H}_j(\bar{x}, \bar{\lambda}_j) d$$

for every $d \in C(\bar{x})$, where

$$\mathcal{H}_j(\bar{x}, \bar{\lambda}_j) := \begin{cases} -\dfrac{[\bar{\lambda}_j]_0}{[g_j(\bar{x})]_0} Dg_j(\bar{x})^T R_{m_j} Dg_j(\bar{x}), & \text{if } j \in I_B(\bar{x}), \\ 0, & \text{Otherwise.} \end{cases} \tag{32}$$

With this in mind, we can prove that SOC holds at $(\bar{x}, \bar{\lambda})$ under the FCR property by means of analysing the problem along the curve $\xi(t)$ from the proof of Theorem 3.

Theorem 5 *Let $\bar{x}$ be a local minimizer of problem (NSOCP) that satisfies the FCR property. Then, for any given Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$, the pair $(\bar{x}, \bar{\lambda})$ satisfies SOC as in Definition 1; that is,*

$$d^T \nabla^2 f(\bar{x}) d + \sum_{j=1}^{q} \left\langle D^2 g_j(\bar{x})[d, d], \bar{\lambda}_j \right\rangle - \sigma(d, \bar{x}, \bar{\lambda}) \geq 0, \tag{33}$$

for every $d \in C(\bar{x}) = \mathcal{L}_\Omega(\bar{x}) \cap \{\nabla f(\bar{x})\}^\perp$.

Proof If $\Lambda(\bar{x}) = \emptyset$, the result holds trivially. Otherwise, let $\bar{\lambda} := (\bar{\lambda}_1, \ldots, \bar{\lambda}_q) \in \Lambda(\bar{x})$ be arbitrary and fixed. Our aim is to prove that inequality (13) holds for the pair $(\bar{x}, \bar{\lambda})$, for every $d \in C(\bar{x})$. So let $d \in C(\bar{x})$ be also arbitrary, and let F be as in (25). Recall that, for the sake of simplicity and without loss of generality, we are assuming $I_B(\bar{x}) = \{1, \ldots, |I_B(\bar{x})|\}$.

Proceeding in the same way as in the proof of Theorem 3, since the FCR property holds at $\bar{x}$ and $d \in \mathcal{L}_\Omega(\bar{x})$, we can construct a twice continuously differentiable diffeomorphism $\xi : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$, for some $\varepsilon > 0$, such that: $\xi(0) = \bar{x}$, $\xi'(0) = d$, and

$$\mathcal{G}(\xi(t)) \in \text{span}(F) \tag{34}$$

for every $t \in (-\varepsilon, \varepsilon)$. In addition, $\mathcal{G}(\xi(t)) \in F$ for every $t \in [0, \varepsilon)$, meaning $\xi(t)$ is feasible for all such t. Since $\bar{x}$ is a local minimizer of (NSOCP), then $t = 0$ is a local minimizer of the function $\varphi(t) := f(\xi(t))$ subject to the constraint $t \geq 0$. Then, it is easy to see that

$$\varphi''(0) = d^T \nabla^2 f(\bar{x}) d + \nabla f(\bar{x})^T \xi''(0) \geq 0. \tag{35}$$

The rest of this proof consists of computing $\nabla f(\bar{x})^T \xi''(0)$. To do this, we will use an auxiliary complementarity function defined as

$$R(t) := \sum_{j \in I_0(\bar{x})} \langle g_j(\xi(t)), \bar{\lambda}_j \rangle + \sum_{j \in I_B(\bar{x})} [\bar{\lambda}_j]_0 \phi_j(\xi(t)).$$

First, we claim that $R(t) = 0$ for every $t \in (-\varepsilon, \varepsilon)$. To prove this, let us use the KKT conditions to obtain

$$\sum_{j\in I_0(\bar{x})} \langle Dg_j(\bar{x})d, \bar{\lambda}_j\rangle + \sum_{j\in I_B(\bar{x})} [\bar{\lambda}_j]_0 \nabla\phi_j(\bar{x})^T d = \langle d, -\nabla f(\bar{x})\rangle = 0, \tag{36}$$

where the last equality follows from the fact $d \in C(\bar{x})$. By the way, recall from (21) that $Dg_j(\bar{x})d \in K_{m_j}$ for every $j \in I_0(\bar{x})$, and $\nabla\phi_j(\bar{x})^T d \geq 0$ for every $j \in I_B(\bar{x})$. On the other hand, $\bar{\lambda}_j \in K_{m_j}^\circ$ and hence $\langle Dg_j(\bar{x})d, \bar{\lambda}_j\rangle \leq 0$ for every $j \in I_0(\bar{x})$, and $[\bar{\lambda}_j]_0 \leq 0$ for every $j \in I_B(\bar{x})$. Thus,

$$\langle Dg_j(\bar{x})d, \bar{\lambda}_j\rangle = 0, \ \forall j \in I_0(\bar{x}), \quad \text{and} \quad [\bar{\lambda}_j]_0 \nabla\phi_j(\bar{x})^T d = 0, \ \forall j \in I_B(\bar{x}). \tag{37}$$

With this in mind, let us analyse each term of $R(t)$ separately.

1. For each $j \in I_0(\bar{x})$, it follows directly from (37) that:
   - If $Dg_j(\bar{x})d \in \text{int}(K_{m_j})$, then $\bar{\lambda}_j = 0$, since $\bar{\lambda}_j \in K_{m_j}^\circ$;
   - If $Dg_j(\bar{x})d \in \text{bd}^+(K_{m_j})$ we have $g_j(\xi(t)) \in \text{span}(Dg_j(\bar{x})d)$ by (34) and (25), and consequently, $\langle g_j(\xi(t)), \bar{\lambda}_j\rangle = 0$ for every $t \in (-\varepsilon, \varepsilon)$ due to (37);
   - If $Dg_j(\bar{x})d = 0$, then $g_j(\xi(t)) = 0$ also for every $t \in (-\varepsilon, \varepsilon)$, due to (26).

   The above reasoning implies that $\langle g_j(\xi(t)), \bar{\lambda}_j\rangle = 0$ for every $t \in (-\varepsilon, \varepsilon)$ and every $j \in I_0(\bar{x})$.
2. For each $j \in I_B(\bar{x})$, consider J as in (27) and it follows that if $\nabla\phi_j(\bar{x})^T d = 0$, then $\phi_j(\xi(t)) = 0$ for every $t \in (-\varepsilon, \varepsilon)$. On the other hand, in (37) we see that if $\nabla\phi_j(\bar{x})^T d > 0$, then $[\bar{\lambda}_j]_0 = 0$.

Knowing that $R(t) = 0$ for every $t \in (-\varepsilon, \varepsilon)$, we obtain that the derivatives of $R(t)$ are also zero for all such t. Let us compute them: the first derivative of $R(t)$ is given by

$$R'(t) = \sum_{j\in I_0(\bar{x})} \left\langle Dg_j(\xi(t))\xi'(t), \bar{\lambda}_j\right\rangle + \sum_{j\in I_B(\bar{x})} [\bar{\lambda}_j]_0 \left\langle \nabla\phi_j(\xi(t)), \xi'(t)\right\rangle,$$

and the derivative of $R'(t)$ is

$$R''(t) = \sum_{j\in I_0(\bar{x})} \left\langle D^2 g_j(\xi(t))[\xi'(t), \xi'(t)], \bar{\lambda}_j\right\rangle + \sum_{j\in I_0(\bar{x})} \left\langle Dg_j(\xi(t))^T \bar{\lambda}_j, \xi''(t)\right\rangle$$
$$+ \sum_{j\in I_B(\bar{x})} [\bar{\lambda}_j]_0 \left(\left\langle D^2\phi_j(\xi(t))\xi'(t), \xi'(t)\right\rangle + \left\langle \nabla\phi_j(\xi(t)), \xi''(t)\right\rangle\right).$$

Due to the fact $R''(t)$ is continuous, taking the limit $t \to 0$, we obtain

$$R''(0) = \sum_{j\in I_0(\bar{x})} \left\langle D^2 g_j(\bar{x})[d, d], \bar{\lambda}_j\right\rangle + \sum_{j\in I_0(\bar{x})} \left\langle Dg_j(\bar{x})^T \bar{\lambda}_j, \xi''(0)\right\rangle$$

$$+\sum_{j\in I_B(\bar{x})}[\bar{\lambda}_j]_0\left(\left\langle D^2\phi_j(\bar{x})d,d\right\rangle+\frac{1}{[g_j(\bar{x})]_0}\left\langle Dg_j(\bar{x})^T R_{m_j}g_j(\bar{x}),\xi''(0)\right\rangle\right).$$

The above expression can be simplified using the relation

$$\begin{aligned}\left\langle D^2\phi_j(\bar{x})d,d\right\rangle&=\frac{\langle \widehat{Dg_j(\bar{x})d},\widehat{g_j(\bar{x})}\rangle^2}{\|\widehat{g_j(\bar{x})}\|^3}-\frac{\|\widehat{Dg_j(\bar{x})d}\|^2}{\|\widehat{g_j(\bar{x})}\|}+\\&\quad+\left\langle D^2g_j(\bar{x})[d,d],\frac{R_{m_j}g_j(\bar{x})}{\|\widehat{g_j(\bar{x})}\|}\right\rangle\\&=\frac{1}{[g_j(\bar{x})]_0}\left\langle R_{m_j}Dg_j(\bar{x})d,\ Dg_j(\bar{x})d\right\rangle+\\&\quad+\frac{1}{[g_j(\bar{x})]_0}\left\langle D^2g_j(\bar{x})[d,d],R_{m_j}g_j(\bar{x})\right\rangle,\end{aligned}$$

that holds true for every $j\in I_B(\bar{x})$ such that $\nabla\phi_j(\bar{x})^T d=0$, which can be directly computed from the definition of ϕ_j, since in this case $[g_j(\bar{x})]_0=\|\widehat{g_j(\bar{x})}\|$ and, moreover,

$$\langle Dg_j(\bar{x})d,\ R_{m_j}g_j(\bar{x})\rangle=\langle d,\ Dg_j(\bar{x})^T R_{m_j}g_j(\bar{x})\rangle=[g_j(\bar{x})]_0\nabla\phi_j(\bar{x})^T d=0.$$

Further, equation (37) tells us that if $\nabla\phi_j(\bar{x})^T d>0$, then $[\bar{\lambda}_j]_0=0$. Then, we get

$$\begin{aligned}R''(0)=&\sum_{j\in I_0(\bar{x})\cup I_B(\bar{x})}\left\langle D^2g_j(\bar{x})[d,d],\bar{\lambda}_j\right\rangle+\sum_{j\in I_0(\bar{x})\cup I_B(\bar{x})}\left\langle Dg_j(\bar{x})^T\bar{\lambda}_j,\xi''(0)\right\rangle\\&+\sum_{j\in I_B(\bar{x})}\frac{[\bar{\lambda}_j]_0}{[g_j(\bar{x})]_0}\left\langle R_{m_j}Dg_j(\bar{x})d,\ Dg_j(\bar{x})d\right\rangle=0.\end{aligned}\tag{38}$$

Moreover, by the KKT conditions, we have

$$\nabla f(\bar{x})^T\xi''(0)=-\sum_{j\in I_0(\bar{x})\cup I_B(\bar{x})}\left\langle Dg_j(\bar{x})^T\bar{\lambda}_j,\xi''(0)\right\rangle,$$

which yields together with equation (38), the following:

$$\begin{aligned}\nabla f(\bar{x})^T\xi''(0)=&\sum_{j\in I_0(\bar{x})\cup I_B(\bar{x})}\left\langle D^2g_j(\bar{x})[d,d],\bar{\lambda}_j\right\rangle+\\&+\sum_{j\in I_B(\bar{x})}\frac{[\bar{\lambda}_j]_0}{[g_j(\bar{x})]_0}d^T Dg_j(\bar{x})^T R_{m_j}Dg_j(\bar{x})d.\end{aligned}\tag{39}$$

Therefore, since $\bar{\lambda}_j = 0$ for every $j \in I_{\text{int}}(\bar{x})$, plugging (39) into (35) yields

$$d^T \nabla^2 f(\bar{x})d + \sum_{j=1}^{q} \left\langle D^2 g_j(\bar{x})[d,d], \bar{\lambda}_j \right\rangle - \sigma(d, \bar{x}, \bar{\lambda}) \geq 0.$$

Since $d \in C(\bar{x})$ is arbitrary, we conclude that $\bar{x}$ satisfies SOC with respect to $\bar{\lambda}$, which was also chosen arbitrarily and remained fixed from the very beginning. Thus, the proof is complete. □

Observe that Theorem 5 implies that the FCR property ensures the fulfilment of the strong second-order necessary condition at a given point $\bar{x}$, in the sense that for every $\bar{\lambda} \in \Lambda(\bar{x})$, and every $d \in C(\bar{x})$, inequality (13) holds true. If, in addition, $H(\bar{x})$ is closed (CRCQ), then $\Lambda(\bar{x}) \neq \emptyset$, and as consequence, we obtain that the strong second-order condition is satisfied in the presence of CRCQ. It is also worth mentioning that since the strong necessary condition of Theorem 5 implies the classical condition of Theorem 2, then it also induces a sufficient (no-gap) second-order optimality condition after replacing $\geq$ by $>$ in inequality (33).

Remark 4 In contrast with the FCR property, the condition presented in [60, Definition 2.1] fails to be a CQ even when $H(\bar{x})$ is closed. In fact, let us recall the counterexample presented in [5]:

$$\begin{array}{ll} \text{Minimize} & f(x) := -x, \\ \text{s.t.} & g(x) := (x, x + x^2) \in K_2, \end{array}$$

The unique solution of this problem is $\bar{x} = 0$. For this particular example, [60, Definition 2.1] holds if, and only if, $\{1, 1 + 2x\}$ remain with constant rank in some neighborhood of $\bar{x}$ (one may consider also all of its subfamilies, see [5]). Of course, this is verified, and since K_2 is polyhedral, the set $H(\bar{x})$ is closed. However, $\bar{x}$ does not satisfy the KKT conditions.

On the other hand, to see that CRCQ as in Definition 5 does not hold at $\bar{x}$, take $F := \text{cone}((1, 1)) \trianglelefteq K_2$ and note that

$$Dg(x)^T[F^{\perp}] = \text{span}(-2x)$$

has dimension 1 for every $x \neq 0$, but has dimension zero at $\bar{x}$. In particular, this example shows that CRCQ as in Definition 5 is not a mere correction of the condition presented in [60], and that the condition of [60] cannot be corrected by simply adding the closedness of $H(\bar{x})$ to its definition.

4.3 About the sequential constant rank CQ

In [9], we introduced an alternative extension of CRCQ for (NSOCP) that was based on a special re-characterization of the nondegeneracy condition [7] in terms of the eigenvectors of some perturbations of $g(\bar{x})$. Let us recall an equivalent characterization of it, which will be used here as a definition for simplicity.

Definition 6 (Seq-CRCQ for NSOCP) Let $\bar{x} \in \Omega$. We say that the *Sequential-CRCQ* (Seq-CRCQ) condition holds at $\bar{x}$ if for every vector $\bar{w}_j \in \mathbb{R}^{m_j-1}$ with $\|\bar{w}_j\| = 1$, $j \in I_0(\bar{x})$, there is a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{w})$, $\bar{w} := (\bar{w}_j)_{j\in I_0(\bar{x})}$, such that: for all subsets $J_1, J_2 \subseteq I_0(\bar{x})$ and $J_3 \subseteq I_B(\bar{x})$, if the family

$$\mathcal{D}(x, w) := \left\{Dg_j(x)^T\left(1, -w_j\right)\right\}_{j\in J_1} \bigcup \left\{Dg_j(x)^T\left(1, w_j\right)\right\}_{j\in J_2} \bigcup \bigcup \left\{Dg_j(x)^T\left(1, -\frac{\widehat{g_j(x)}}{\|\widehat{g_j(x)}\|}\right)\right\}_{j\in J_3}$$

is linearly dependent at $(x, w) := (\bar{x}, \bar{w})$, then $\mathcal{D}(x, w)$ remains linearly dependent for all $(x, w) \in \mathcal{V}$ such that $\|w_j\| = 1$, $j \in J_1 \cup J_2$, where $w := (w_j)_{j\in I_0(\bar{x})}$.

This constraint qualification was used in [9] to achieve global convergence of a class of algorithms to KKT points, and some interesting properties were shown together with a weaker variant of Seq-CRCQ. Namely, it is also independent of Robinson's CQ, strictly weaker than nondegeneracy, and it implies the *metric subregularity CQ* (also known as *error bound CQ*). Moreover, note that if $I_0(\bar{x}) = \emptyset$, then Seq-CRCQ coincides with the FCR property, which in turn coincides with CRCQ. However, this is not necessarily true otherwise. In the following example, we show that CRCQ according to Definition 5 does not imply Seq-CRCQ.

Example 4 Consider the constraint:

$$g(x) = (x, -x, 0) \in K_3,$$

and let $\bar{x} = 0$, a feasible point. The constraint function g is affine, then the FCR property holds at $\bar{x}$ (see Remark 2). Now, let us show that $H(\bar{x})$ is closed: since $g(\bar{x}) = 0$, it holds that

$$H(\bar{x}) = Dg(\bar{x})^T K_3 = \{v_1 - v_2 \mid (v_1, v_2, v_3) \in K_3\} = \mathbb{R}_+.$$

Therefore, $H(\bar{x})$ is a closed set, and CRCQ according to Definition 5 holds at $\bar{x}$.

On the other hand, Seq-CRCQ does not hold at $\bar{x}$, because for any $w = (w_1, w_2) \in \mathbb{R}^2$,

$$Dg(\bar{x})^T(1, w) = 1 - w_1 \quad \text{and} \quad Dg(\bar{x})^T(1, -w) = 1 + w_1;$$

then, take $\bar{w} = (1, 0)$ and any sequence $\{w^k\}_{k\in\mathbb{N}} \to \bar{w}$ such that $w_1^k \neq 1$ for all $k \in \mathbb{N}$, to see that $Dg(\bar{x})^T(1, w_1^k) \neq 0$ for every $k \in \mathbb{N}$, but $Dg(\bar{x})^T(1, \bar{w}) = 0$. □

We were not able to prove nor find a counterexample for the converse statement. However, with only Example 4 at hand, we already know that CRCQ is in the worst case independent of Seq-CRCQ, and in the best case, strictly weaker than it, meaning the results of this paper either improve or are parallel to the results of [9].

5 Nonlinear semidefinite programming

In this section, we will study constant rank conditions for nonlinear semidefinite programming problems, which can be stated in standard form as follows:

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & G(x) \succeq 0. \end{array} \tag{NSDP}$$

This problem can be seen as a particular case of (NCP), letting $\mathbb{E} = \mathbb{S}^m$ be the space of $m \times m$ symmetric matrices with real entries, and

$$\mathcal{K} = \mathbb{S}^m_+ := \{A \in \mathbb{S}^m \mid z^T A z \geq 0,\ \forall z \in \mathbb{R}^m\}$$

be the cone of all $m \times m$ symmetric positive semidefinite matrices, with $G : \mathbb{R}^n \to \mathbb{E}$ being twice continuously differentiable. The symbol $\succeq$ denotes the partial order induced by $\mathbb{S}^m_+$, meaning that $A \succeq B$ if, and only if, $A - B \in \mathbb{S}^m_+$. In this section, for any given $A \in \mathbb{S}^m$ we will denote by $\mu_i(A)$ the i-th eigenvalue of A arranged in non-increasing order, and $u_i(A)$ will denote an associated unitary eigenvector.

Recall from Sect. 3 that the constant rank constraint qualification can be obtained in two steps: first, reduce the problem to consider only the locally relevant part of the constraint; then, analyse the image of the faces of the reduced cone by the derivative of the reduced constraint function. For the first step, we will employ a reduction approach based on Bonnans and Shapiro [24, Example 3.98], which can also be found in [16, Section 2.3].

Let $\bar{Y} \succeq 0$, denote $r := \operatorname{rank}(\bar{Y})$, and let $\bar{E} \in \mathbb{R}^{m \times m-r}$ be a matrix whose columns form an orthonormal basis of $\operatorname{Ker}(\bar{Y})$. Then, in a sufficiently small neighborhood $\mathcal{N}$ of $\bar{Y}$, we consider the function $\mathcal{E}_{\bar{E}} : \mathcal{N} \to \mathbb{R}^{m \times m-r}$ given by

$$\mathcal{E}_{\bar{E}}(Y) := \operatorname{gramschmidt}\left(\Pi(Y)\bar{E}\right), \tag{40}$$

where $\Pi(Y)$ denotes the orthogonal projection matrix onto the space spanned by $u_{r+1}(Y), \ldots, u_m(Y)$ and $\operatorname{gramschmidt}(\Pi(Y)\bar{E})$ denotes the output of the Gram-Schmidt orthonormalization procedure after being applied to the columns of $\Pi(Y)\bar{E}$.

Lemma 2 *For any given $\bar{Y} \succeq 0$ and any matrix $\bar{E} \in \mathbb{R}^{m \times m-r}$ with orthonormal columns that span $\operatorname{Ker}(\bar{Y})$, where $r := \operatorname{rank}(\bar{Y})$, it holds that:*

1. *$\mathcal{E}_{\bar{E}}$ is well-defined and analytic provided $\mathcal{N}$ is small enough;*
2. *$\mathcal{E}_{\bar{E}}(Y)^T \mathcal{E}_{\bar{E}}(Y) = \mathbb{I}_{m-r}$ and $\operatorname{Im}(\mathcal{E}_{\bar{E}}(Y)) = \operatorname{span}(\{u_{r+1}(Y), \ldots, u_m(Y)\})$, for every $Y \in \mathcal{N}$;*
3. *$\mathcal{E}_{\bar{E}}(\bar{Y}) = \bar{E}$.*

Proof For item 1, observe that $Y \mapsto \Pi(Y)$ is an analytic function of Y in a sufficiently small neighborhood, say $\mathcal{N}$, of $\bar{Y}$ (see, for example, [39, Theorem 1.8]), then $Y \mapsto \Pi(Y)\bar{E}$ is also analytic in $\mathcal{N}$ and, moreover, $\Pi(\bar{Y})\bar{E} = \bar{E}$. Shrinking $\mathcal{N}$ if necessary, we have that for all $Y \in \mathcal{N}$, the rank of $\Pi(Y)\bar{E}$ is equal to the rank of $\Pi(\bar{Y})\bar{E} = \bar{E}$,

Appendix E

Preprint Submitted: On achieving strong necessary second-order properties in nonlinear programming.

ON ACHIEVING STRONG NECESSARY SECOND-ORDER PROPERTIES IN NONLINEAR PROGRAMMING*

Andreas Fischer, Gabriel Haeser and Thiago P. Silveira

September 29th, 2023.

Abstract: Second-order necessary or sufficient optimality conditions for nonlinear programming are usually defined by means of the positive (semi-)definiteness of a quadratic form, or a maximum of quadratic forms, over the critical cone. However, algorithms for finding such second-order stationary points are currently unknown. Typically, an algorithm with a second-order property approximates a primal-dual point such that the Hessian of the Lagrangian evaluated at the limit point is, under a constraint qualification, positive semidefinite over the lineality space of the critical cone. This is in general a proper subset of the critical cone, unless one assumes strict complementarity, which gives a weaker necessary optimality condition. In this paper, a new strong sequential optimality condition is suggested and analyzed. Based on this, we propose a penalty algorithm which, under certain conditions, is able to achieve second-order stationarity with positive semidefiniteness over the whole critical cone, which corresponds to a strong necessary optimality condition. Although the algorithm we propose is somewhat of a theoretical nature, our analysis provides a general framework for obtaining strong second-order stationarity.

Keywords: nonlinear optimization, second-order optimality conditions, constraint qualifications, global convergence
Mathematics Subject Classification: *49K05, 65K10, 90C26, 90C30*

---

1 Introduction

When proposing a derivative-based algorithm for smooth constrained optimization problems, one must have in mind efficiency and robustness. In terms of robustness, it is clear that one does not expect that a local minimizer will always be found. Thus, algorithms typically aim at finding points satisfying some first- or second-order *necessary* optimality condition. The Karush-Kuhn-Tucker conditions are usually the standard first-order stationarity notion employed. However, there are different notions of second-order stationarity.

Most notions of second-order stationarity are somewhat of a theoretical nature, since it is very difficult to incorporate them in a practical algorithm, at least not without impairing efficiency. Thus, most algorithms possessing a second-order global convergence theory only consider the simplest of these conditions, namely, one that does not make use of the full second-order information. More specifically, instead of ensuring positive semidefiniteness of the Hessian of the Lagrangian over the whole critical cone, this property is assured only in a subspace contained in the critical cone. This is done essentially because dealing with the whole critical cone is a computationally challenging task, see [17].

In this paper we consider nonlinear optimization problems in finite dimensions with equality and inequality constraints, where the problem functions are twice continuously differentiable and we aim at designing a general framework that is able to find a point satisfying a strong second-order necessary optimality condition, that is, considering the whole critical cone, under reasonable assumptions. This is done by means of a penalty algorithm that keeps the inequality constraints within the subproblems. However, our approach is somewhat theoretical as we do not propose an algorithm for solving the subproblems. This task remains a challenging open problem. Nevertheless, the analysis we conduct is non-standard and it consists of a first step towards the more general goal.

The paper starts in Section 2 with a review of different second-order necessary optimality conditions; we focus in particular on the results relying on assumptions on the rank of the gradients of constraints nearby the local minimizer, in particular, we consider the well known constant rank constraint qualification (CRCQ [15]).

*The research of the first author is supported by the Volkswagen Foundation under grant No. 97 775. The second author is funded by CNPq and the São Paulo Research Foundation grant 2018/24293-0, while the third author is funded by the São Paulo Research Foundation grants 2017/12187-9 and 2020/00130-5.

In Section 3 we present a gentle introduction to the topic of *sequential optimality conditions* [3], which is the main tool we employ to achieve our results. Based on this discussion, we introduce new strong second-order necessary approximate KKT conditions that consider the whole critical cone. Under a constant rank condition, we prove that all local minimizers of the optimization problem satisfy one of these approximate KKT conditions while the other condition is satisfied by all strict local minimizers.

In Section 4, we recall from [7, 12] that a standard barrier method and a second-order augmented Lagrangian method are not able to guarantee the strong second-order condition, even if a strict local minimizer of the subproblems is found at each iteration. Then, we show that this phenomenon will not occur (under a constant rank condition) if only equality constraints are penalized. That is, we propose our framework for designing an algorithm that will achieve the strong second-order condition under reasonable assumptions. The assumptions we employ rely on the constant rank of sets of gradients of constraints and the objective function, together with the extended Mangasarian-Fromovitz constraint qualification (MFCQ).

Our notation is rather standard. We just mention that $\|\cdot\|$ always denotes the Euclidean norm.

2 On Second-Order Conditions

Let us consider the nonlinear programming problem

$$\begin{array}{lll} \text{Minimize} & f(x), & \\ \text{subject to} & h_i(x) = 0, & i = 1, \dots, m, \\ & g_j(x) \le 0, & j = 1, \dots, p, \end{array} \tag{NLP}$$

where the functions $f, h_i, g_j : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable. By Ω, we denote the feasible set of (NLP). Moreover, for any $x \in \mathbb{R}^n$, the set

$$A(x) := \{j \in \{1, \dots, p\} \mid g_j(x) = 0\}$$

contains the indices of inequality constraints that are active at x.

To formulate second-order conditions, let us first introduce the generalized Lagrangian $L_0 : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^p_+ \to \mathbb{R}$ by

$$L_0(x, r, \lambda, \mu) := rf(x) + h(x)^\top \lambda + g(x)^\top \mu,$$

whereas the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+ \to \mathbb{R}$ is given by

$$L(x, \lambda, \mu) := f(x) + h(x)^\top \lambda + g(x)^\top \mu.$$

Now, for any $x \in \Omega$, the set $\Lambda_0(x)$ of Fritz John multipliers and the set $\Lambda(x)$ of Lagrange multipliers are defined as

$$\Lambda_0(x) := \left\{0 \ne (r, \lambda, \mu) \in \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^p_+ \mid \nabla_x L_0(x, r, \lambda, \mu) = 0, g(x)^\top \mu = 0\right\}$$

and

$$\Lambda(x) := \{(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \mid \nabla_x L(x, \lambda, \mu) = 0, g(x)^\top \mu = 0\},$$

respectively. For any $x \in \Omega$, we further need the critical cone

$$C(x) := \{d \in \mathbb{R}^n \mid \nabla f(x)^\top d \le 0,\ \nabla h(x)^\top d = 0, \nabla g_j(x)^\top d \le 0 \text{ for all } j \in A(x)\}. \tag{2.1}$$

The next theorem provides a pair of *no-gap* second-order optimality conditions. It can be derived from [10].

Theorem 2.1. *Let $\bar{x} \in \Omega$ be given. Then the following assertions are valid:*

a) If $\Lambda_0(\bar{x}) \ne \emptyset$ and

$$\sup_{(r,\lambda,\mu)\in\Lambda_0(\bar{x})} d^\top \nabla^2_{xx} L_0(\bar{x}, r, \lambda, \mu) d > 0 \quad \text{for all } d \in C(\bar{x}) \setminus \{0\},$$

then $\bar{x}$ is a strict local minimizer of (NLP).

b) If $\bar{x}$ is a local minimizer of (NLP), then $\Lambda_0(\bar{x}) \ne \emptyset$ and

$$\sup_{(r,\lambda,\mu)\in\Lambda_0(\bar{x})} d^\top \nabla^2_{xx} L(\bar{x}, r, \lambda, \mu) d \ge 0 \quad \text{for all } d \in C(\bar{x}).$$

Although several research is based on Fritz John multipliers, in this paper we are interested in Lagrange multipliers. As usual, any (x, λ, μ) is called a Karush-Kuhn-Tucker (KKT) point of (NLP) if $x \in \Omega$ and $(\lambda, \mu) \in \Lambda(x)$.

To avoid the distinction between Fritz John and Lagrange multipliers, one may assume the well-known Mangasarian-Fromovitz constraint qualification (MFCQ), which can be stated at $\bar{x} \in \Omega$ as saying that there is no Fritz John multiplier $(r, \lambda, \mu) \in \Lambda_0(\bar{x})$ with $r = 0$. Notice also that a Fritz John multiplier (r, λ, μ) with $r \ne 0$ provides a Lagrange multiplier $(\lambda/r, \mu/r)$. That is, when $r \ne 0$, one may without loss of generality consider $r = 1$. In this sense, under MFCQ, the notions of Fritz John and Lagrange multipliers coincide and Theorem 2.1 gives rise to the following standard second-order necessary optimality condition:

Proposition 2.2. *Let $\bar{x}$ be a local minimizer of* (NLP) *that satisfies MFCQ. Then $\Lambda(\bar{x}) \neq \emptyset$ and*

$$\sup_{(\lambda,\mu)\in\Lambda(\bar{x})} d^\top \nabla^2_{xx} L(\bar{x}, \lambda, \mu) d \geq 0 \quad \text{for all } d \in C(\bar{x}). \tag{2.2}$$

Instead of MFCQ, we may use the constant rank constraint qualification (CRCQ) from [15]. This leads to the following second-order necessary optimality condition, which is the basis for our main focus in this paper.

Proposition 2.3 ([2])**.** *Let $\bar{x}$ be a local minimizer of* (NLP) *that satisfies CRCQ. Then $\Lambda(\bar{x}) \neq \emptyset$ and, for any $(\lambda, \mu) \in \Lambda(\bar{x})$, it holds that*

$$d^\top \nabla^2_{xx} L(\bar{x}, \lambda, \mu) d \geq 0 \quad \text{for all } d \in C(\bar{x}). \tag{2.3}$$

Notice that under the stronger assumption that the linear independence constraint qualification (LICQ) holds at $\bar{x}$, Proposition 2.3 follows trivially from Proposition 2.2 since LICQ implies MFCQ and that $\Lambda(\bar{x})$ is a singleton. Obviously, condition (2.3) is a stronger necessary optimality condition than (2.2). Under CRCQ, this stronger condition (2.3) holds basically because CRCQ implies that the value of the second-order form $\langle d, \nabla^2_{xx} L(\bar{x}, \lambda, \mu) d\rangle$ in (2.2) is, for any $d \in C(\bar{x})$, invariant to the choice of $(\lambda, \mu) \in \Lambda(\bar{x})$, see [11] and the extended version of [9].

Under CRCQ the necessary optimality condition (2.3) would be rather suitable for the algorithmic practice than condition (2.2). This means, given an algorithm that generates a primal-dual sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$, one is interested in proving that a limit point $(\bar{x}, \lambda, \mu) \in \Omega \times \mathbb{R}^m \times \mathbb{R}^p_+$ of this sequence is such that $(\lambda, \mu) \in \Lambda(\bar{x})$ and the second-order condition (2.3) is satisfied. However, no such algorithm has yet been presented. Algorithms with convergence to some kind of second-order point usually find limit points that satisfy a weaker version of (2.3) (see [12] and the references therein), where the critical cone $C(\bar{x})$ is replaced by its lineality space

$$S(\bar{x}) := \{d \in \mathbb{R}^n \mid \nabla f(\bar{x})^\top d = 0, \nabla h(\bar{x})^\top d = 0, \nabla g_j(\bar{x})^\top d = 0 \text{ for all } j \in A(\bar{x})\}. \tag{2.4}$$

Clearly, this necessary optimality condition is less interesting than the one presented in Proposition 2.3, since no associated sufficient optimality condition is known and $S(\bar{x}) \subseteq C(\bar{x})$. Also, one is essentially not able to exploit the structure of $S(\bar{x})$ in order to prove the result of Proposition 2.3 with $C(\bar{x})$ replaced by $S(\bar{x})$ under a condition weaker than CRCQ. An exception (but assuming MFCQ) is the following result. Its formulation makes use of the matrix $M(x) \in \mathbb{R}^{n\times(m+|A(\bar{x})|)}$ with $M(x) := \big(\nabla h(x), \nabla g_{A(\bar{x})}(x)\big)$.

Proposition 2.4 ([9, 13, 16])**.** *Let $\bar{x}$ be a local minimizer of* (NLP) *which satisfies MFCQ. If*

$$\operatorname{rank}(M(x)) \leq \operatorname{rank}(M(\bar{x})) + 1$$

for all x sufficiently close to $\bar{x}$, then there exists $(\lambda, \mu) \in \Lambda(\bar{x})$ with

$$d^\top \nabla^2_{xx} L(\bar{x}, \lambda, \mu) d \geq 0 \quad \text{for all } d \in S(\bar{x}). \tag{2.5}$$

Of course, the less theoretical value of the second-order condition given by (2.5) is somehow compensated by its numerical tractability, see the discussion in the extended version of [9]. That is, many practical algorithms are able to exploit the linear space structure of $S(\bar{x})$ in order to achieve (2.5) in a reasonable manner. Our goal in this paper is to develop an algorithm whose limit points guarantee the stronger second-order necessary optimality condition (2.3).

3 A Strong Sequential Optimality Condition

The study of global convergence of algorithms under weak assumptions can be done with the aid of *sequential optimality conditions* [3]. Let us say that one is first able to show that an algorithm generates a sequence $\{x^k\}$ satisfying some mathematical proposition $\mathcal{P}(\{x^k\})$. Typically, this proposition is associated with a perturbation of a necessary optimality condition. The second step would be to prove that whenever $\bar{x}$ is a local minimizer, there exists a sequence $\{z^k\}$ with $z^k \to \bar{x}$ so that the proposition $\mathcal{P}(\{z^k\})$ is valid. When defining $\mathcal{P}(\cdot)$, of course, one is interested in as strong as possible necessary optimality conditions, however, one must consider the additional requirement that the mathematical proposition must also be satisfied by the sequence generated by the algorithm of interest. Both of these steps can usually be done without assuming that the problem satisfies a constraint qualification and may serve as an adequate enough global convergence theory. This strategy has been applied in several contexts, in particular when the problem has no clear standard optimality condition, or when one needs a consistent way of perturbing the standard optimality conditions, say, in order to conduct a complexity analysis [14, 18]. This avoids constraint qualifications at all; however, a final step of the analysis may be done using a constraint qualification for measuring the strength of the optimality condition: one should prove that when a feasible point $\bar{x}$ satisfies a constraint qualification and it can be approximated by some sequence $z^k \to \bar{x}$ so that the mathematical proposition $\mathcal{P}(\{z^k\})$ holds, then $\bar{x}$ satisfies a standard first- or second-order stationarity condition (say, that there exists some $(\lambda, \mu) \in \Lambda(\bar{x})$ that satisfies (2.5)).

For instance, when the problem has only equality constraints, an algorithm may generate a sequence $\{x^k\}$ that satisfies the mathematical proposition

$$\mathcal{P}(\{x^k\}) := \left[h(x^k) \to 0 \quad \text{and} \quad \nabla f(x^k) + \sum_{i=1}^{m} \nabla h_i(x^k) \lambda_i^k \to 0 \text{ for some sequence } \{\lambda^k\} \subset \mathbb{R}^m \right]$$

and one can prove that a local minimizer $\bar{x}$ may be approximated by a sequence $z^k \to \bar{x}$ of this type, that is, such that $\mathcal{P}(\{z^k\})$ is satisfied. Notice that this necessary optimality condition is strictly stronger than the usual Fritz John condition, which opens the path to considering constraint qualifications strictly weaker than LICQ (more generally, without assuming MFCQ, if inequality constraints are considered in this example).

The final step measuring the strength of the sequential optimality condition may consist of proving that when $\bar{x}$ satisfies some constraint qualification and there exists at least one sequence $z^k \to \bar{x}$ such that $\mathcal{P}(\{z^k\})$ holds, then $\Lambda(\bar{x}) \neq \emptyset$. This shows that any limit point $\bar{x}$ of the sequence generated by the algorithm, that satisfies the constraint qualification, is a KKT point. Not all constraint qualifications may be used for this purpose, but this separated analysis has helped in identifying new weak constraint qualifications suitable for global convergence analysis. See, for instance, [4, 5, 6]. Also, this greatly simplifies the analysis of an algorithm, which resorts to proving some property of the sequence it generates, instead of its limit.

In summary the global convergence of an algorithm using a sequential optimality condition may be done as follows:

a) Characterize the type of sequences $\{x^k\}$ that the algorithm generates with a mathematical proposition $\mathcal{P}(\{x^k\})$.

b) Prove that at a local minimizer $\bar{x}$ of the problem, there exists a sequence $z^k \to \bar{x}$ such that $\mathcal{P}(\{z^k\})$ holds.

c) Measure the strength of $\mathcal{P}(\cdot)$ by showing that a point $\bar{x}$, that can be approximated by $z^k \to \bar{x}$ such that $\mathcal{P}(\{z^k\})$ holds, has the property that whenever $\bar{x}$ satisfies some constraint qualification, then a standard first- or second-order necessary optimality condition is satisfied at $\bar{x}$.

In the remainder of this section we proceed with item b), while in the next section we continue with the analysis of items a) and c). This means, we first develop a strong sequential optimality condition and secondly prove that for a local minimizer $\bar{x}$ of (NLP) there exists a sequence $\{z^k\}$ converging to $\bar{x}$ so that $\{z^k\}$ satisfies this optimality condition. Items a) and c) will be dealt with in Section 4 and are related to our main goal of building an algorithm whose limit points satisfy a strong second-order necessary optimality condition, based on the critical cone (2.1), as used in Proposition 2.3, instead of its lineality space (2.4) in Proposition 2.4.

At this point, we do not assume a constraint qualification to hold with respect to the whole feasible set Ω. However, the following constant rank condition with respect to the set of inequality constraints will be used.

Assumption 3.1. It is said that a point $\bar{x} \in \mathbb{R}^n$ satisfies this assumption if there is a neighborhood of $\bar{x}$ so that, for any subset $J \subseteq A(\bar{x})$, the rank of the family $\{\nabla g_j(y)\}_{j \in J}$ is constant for all y in this neighborhood.

Assumption 3.1 can be seen as CRCQ for a feasible point of a constraint set defined by the inequality constraints of (NLP) only. This is clearly not a constraint qualification for (NLP). Notice that Assumption 3.1 holds trivially if the functions g_j are affine. In order to present our definition of a strong second-order sequential optimality condition for problems such that Assumption 3.1 holds, let us consider the perturbed critical cones

$$C_1(y, x, \mu) := \left\{ \quad d \in \mathbb{R}^n \;\middle|\; \begin{array}{lll} \nabla h_i(y)^\top d = 0 & \text{for} & i = 1, \ldots, m, \\ \nabla g_j(y)^\top d \le 0 & \text{for} & j \in A(x) \text{ with } \mu_j = 0, \\ \nabla g_j(y)^\top d = 0 & \text{for} & j \in A(x) \text{ with } \mu_j > 0. \end{array} \right\}$$

and

$$C_2(y, x) := \left\{ \quad d \in \mathbb{R}^n \;\middle|\; \begin{array}{lll} \nabla f(y)^\top d = 0, & & \\ \nabla h_i(y)^\top d = 0 & \text{for} & i = 1, \ldots, m, \\ \nabla g_j(y)^\top d \le 0 & \text{for} & j \in A(x). \end{array} \right\},$$

for $x, y \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^p_+$. Notice that when $(\bar{x}, \lambda, \mu)$ is a KKT point of (NLP), it holds that

$$C_1(\bar{x}, \bar{x}, \mu) = C(\bar{x}) = C_2(\bar{x}, \bar{x})$$

with $C(x)$ defined in (2.1).

Definition 3.1. (Strong-AKKT2) A point $\bar{x}$ satisfies the C_1-Strong Approximate-KKT2 (C_1-SAKKT2) condition for (NLP) if there exists a sequence $(x^k, \lambda^k, \mu^k, \varepsilon_k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+ \times (0, \infty)$ with $x^k \to \bar{x}$ and $\varepsilon_k \searrow 0$ such that

$$\left\| \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right\| \le \varepsilon_k \tag{3.1}$$

$$\|h(x^k)\| \le \varepsilon_k, \quad \| \max\{0, g(x^k)\}\| \le \varepsilon_k, \quad \| \min\{\mu^k, -g(x^k)\}\| \le \varepsilon_k, \tag{3.2}$$

and

$$d^\top \left(\nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla^2 g_j(x^k) \right) d \ge -\varepsilon_k \|d\|^2 \quad \text{for all } d \in C_1(x^k, \bar{x}, \mu^k). \tag{3.3}$$

If one replaces $C_1(x^k, \bar{x}, \mu^k)$ by $C_2(x^k, \bar{x})$ in the previous definition we say that $\bar{x}$ satisfies the C_2-SAKKT2 condition.

We now prove that C_1-SAKKT2 is a necessary optimality condition for problems which fulfill Assumption 3.1 while C_2-SAKKT2 is a necessary condition for strict optimality under Assumption 3.1.

Theorem 3.2. *Let $\bar{x}$ be a local minimizer of* (NLP) *and suppose that Assumption 3.1 holds at $\bar{x}$. Then $\bar{x}$ satisfies the C_1-SAKKT2 condition. If, in addition, $\bar{x}$ is a strict local minimizer of* (NLP)*, then $\bar{x}$ satisfies the C_2-SAKKT2 condition.*

Proof. Let $\delta > 0$ be chosen such that $f(\bar{x}) \le f(x)$ holds for all $x \in \Omega$ with $\|x - \bar{x}\| \le \delta$. Given a sequence $\{\rho_k\} \subset \mathbb{R}_+$ with $\rho_k \to +\infty$, we consider the regularized penalty subproblem, where only the equality constraints are penalized, that is,

$$\begin{aligned} \text{Minimize} \quad & \phi_k(x) := f(x) + \frac{\rho_k}{2} \sum_{i=1}^m h_i(x)^2 + \frac{1}{4}\|x - \bar{x}\|^4, \\ \text{subject to} \quad & g_j(x) \le 0, \quad j = 1, \dots, p, \\ & \|x - \bar{x}\| \le \delta. \end{aligned} \tag{3.4}$$

Let x^k be a global solution of the optimization problem (3.4), which exists because its feasible set is nonempty and compact and the objective function is continuous. Therefore, for any $k \in \mathbb{N}$, we have

$$f(x^k) + \frac{1}{4}\|x^k - \bar{x}\|^4 \le \phi_k(x^k) \le \phi_k(\bar{x}) = f(\bar{x}). \tag{3.5}$$

Moreover, because $\|x^k - \bar{x}\| \le \delta$ is valid for all $k \in \mathbb{N}$, there is x^* and an infinite subset $K \subseteq \mathbb{N}$ so that $\lim_{k \in K} x^k = x^*$. Notice that $g(x^*) \le 0$ and $\|x^* - \bar{x}\| \le \delta$. Further, since $\rho_k \to +\infty$ and $\{\phi_k(x^k)\}_{k \in K}$ is bounded, we have

$$\lim_{k \in K} h(x^k) = 0 \tag{3.6}$$

so that $h(x^*) = 0$ follows. From (3.5) taken for $k \in K$, we also conclude that

$$f(x^*) + \frac{1}{4}\|x^* - \bar{x}\|^4 \le f(\bar{x}).$$

This, $g(x^*) \le 0$, $h(x^*) = 0$, $\|x^* - \bar{x}\| \le \delta$, and the definition of δ imply $f(\bar{x}) \le f(x^*)$ so that $x^* = \bar{x}$ follows. Then, for $k \in K$ large enough, we have that $\|x^k - \bar{x}\| < \delta$, i.e., the constraint $\|x - \bar{x}\| \le \delta$ in (3.4) is not active at $x = x^k$ for these $k \in K$. Hence, applying Proposition 2.3 (with (NLP) replaced by problem (3.4)) for each of these large enough $k \in K$, it follows by Assumption 3.1 that there exists a Lagrange multiplier $\mu^k \in \mathbb{R}^p_+$ such that (x^k, μ^k) is a KKT point of (3.4) that satisfies a strong second-order necessary optimality condition. More in detail, we have that

$$g(x^k) \le 0, \quad \mu^k \ge 0, \quad g(x^k)^\top \mu^k = 0, \tag{3.7}$$

$$\begin{aligned} \nabla_x L_{(3.4)}(x^k, \mu^k) \quad &:= \quad \nabla \phi_k(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \\ &= \quad \nabla f(x^k) + \sum_{i=1}^m \rho_k h_i(x^k) \nabla h_i(x^k) + \|x^k - \bar{x}\|^2 (x^k - \bar{x}) \\ &\quad\quad + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \\ &= \quad 0, \end{aligned} \tag{3.8}$$

and, because of

$$\begin{aligned}\nabla^2_{xx} L_{(3.4)}(x^k, \mu^k) &= \nabla^2 f(x^k) + \rho_k \sum_{i=1}^m h_i(x^k)\nabla^2 h_i(x^k) + \nabla h_i(x^k)\nabla h_i(x^k)^\top \\ &\quad + \sum_{j=1}^p \mu_j^k \nabla^2 g_j(x^k) + 2(x^k - \bar{x})(x^k - \bar{x})^\top + \|x^k - \bar{x}\|^2 I,\end{aligned}$$

we further have that

$$d^\top \nabla^2_{xx} L_{(3.4)}(x^k, \mu^k) d \geq 0 \quad \text{for all } d \in C_{(3.4)}(x^k) \tag{3.9}$$

with

$$C_{(3.4)}(x^k) := \left\{ d \in \mathbb{R}^n \mid \nabla \phi_k(x^k)^\top d \leq 0, \nabla g_j(x^k)^\top d \leq 0 \text{ for all } j \in A(x^k) \right\}.$$

Since (x^k, μ^k) satisfies (3.7) and (3.8), this yields

$$\begin{aligned}C_{(3.4)}(x^k) &= \left\{ d \in \mathbb{R}^n \;\middle|\; \sum_{j=1}^p \mu_j^k \nabla g_j(x^k)^\top d = 0, \nabla g_j(x^k)^\top d \leq 0 \text{ for } j \in A(x^k) \right\} \\ &= \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_j(x^k)^\top d \leq 0 \text{ for } j \in A(x^k) \text{ with } \mu_j^k = 0, \\ \nabla g_j(x^k)^\top d = 0 \text{ for } j \in A(x^k) \text{ with } \mu_j^k > 0 \end{array} \right\}.\end{aligned}$$

Defining

$$\lambda^k := \rho_k h(x^k), \quad \varepsilon_k := \max\{\|x^k - \bar{x}\|, \|h(x^k)\|\} \quad \text{for } k \in K,$$

we first obtain that $\varepsilon_k \searrow 0$ for $k \in K$. Without loss of generality, we assume that $k \in K$ is large enough so that, due to (3.7), $\mu_j^k = 0$ for $j \notin A(\bar{x})$. Thus, it follows from (3.8) and (3.7) that, for $k \in K$ sufficiently large,

$$\left\| \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right\| \leq \|x^k - \bar{x}\|^3 \leq \varepsilon_k,$$

$$\|h(x^k)\| \leq \varepsilon_k, \quad \|\max\{0, g(x^k)\}\| = 0$$

by $g_j(x^k) \leq 0$ according to (3.4), and

$$\min\{\mu^k, -g(x^k)\} = 0.$$

Therefore, the requirements (3.1) and (3.2) in Definition 3.1 are satisfied. Furthermore, since $A(x^k) \subseteq A(\bar{x})$ for $k \in K$ sufficiently large, we have

$$C_1(x^k, \bar{x}, \mu^k) \subseteq C_{(3.4)}(x^k) \cap \{d \in \mathbb{R}^n \mid \nabla h(x^k)^\top d = 0\}.$$

Taking any $d \in C_1(x^k, \bar{x}, \mu^k)$, we further get from (3.9) with the definitions of λ^k and ε_k that, for $k \in K$ sufficiently large,

$$\begin{aligned}d^\top \nabla^2_{xx} L(x^k, \lambda^k, \mu^k) d &= d^\top \left(\nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{i \in A(\bar{x})} \mu_i^k \nabla^2 g_i(x^k) \right) d \\ &\geq -d^\top \left(2(x^k - \bar{x})(x^k - \bar{x})^\top - \|x^k - \bar{x}\|^2 I \right) d \\ &\geq -\varepsilon_k \|d\|^2.\end{aligned}$$

Hence, also (3.3) in Definition 3.1 holds and, altogether, $\bar{x}$ satisfies the C_1-SAKKT2 condition.

Assume now that $\bar{x}$ is a strict local minimizer of (NLP). Thus, we can follow exactly the same proof with $\phi_k(x)$ replaced by $\tilde{\phi}_k(x) := f(x) + \frac{\rho_k}{2} \sum_{i=1}^m h_i(x)^2$. Note that the expression for $C_{(3.4)}(x^k) \cap \{d \in \mathbb{R}^n \mid \nabla h(x^k)^\top d = 0\}$, again with ϕ_k replaced by $\tilde{\phi}_k$, contains $C_2(x^k, \bar{x})$, which concludes the proof. $\square$

4 Generating KKT Points with Strong Second-Order Conditions

Typically, second-order algorithms are only shown to generate a sequence that converges to a point that satisfies the weak second-order necessary optimality condition from Proposition 2.4. In [12], it is shown that limit points of a standard barrier method need not satisfy the stronger second-order necessary optimality condition from Proposition 2.3, even if strict local minimizers for the subproblems are found at each iteration. In detail, the authors from [12] considered the counterexample

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}x^\top Hx, \\ \text{subject to} & x \geq 0, \end{array}$$

where $x \in \mathbb{R}^n$ with $n \geq 4$ and $H = I - \frac{3}{2n(n-1)}zz^\top$ with $z = e - ne_1$, where e_1 is the first canonical vector and e is the vector with 1 in all entries. For any sequence $r_k \searrow 0$, let $x^k = \sqrt{r_k}e \to \bar{x} = 0$ be defined. Notice that $\mu = 0$ is the unique Lagrange multiplier associated with $\bar{x}$, that is, $\nabla_x L(\bar{x}, \mu) = 0$. However, one has

$$e_1^\top \nabla^2_{xx} L(\bar{x}, \mu) e_1 = e_1^\top H e_1 = 1 - \frac{3(n-1)}{2n} < 0 \quad \text{with } e_1 \in C(\bar{x}) = \mathbb{R}^n_+.$$

Thus, the sequence $\{x^k\}$ converges to a point that fails to satisfy the strong second-order necessary optimality condition. Note however that x^k is a strict local minimizer of the barrier function subproblem

$$\begin{array}{ll} \text{Minimize} & b(x, r_k) := \frac{1}{2}x^\top Hx - r_k \sum_{i=1}^n \log(x_i) \\ \text{subject to} & x > 0. \end{array}$$

Indeed, one has

$$\nabla_x b(x^k, r_k) = Hx^k - r_k(x^k)^{-1} = 0 \quad \text{and} \quad \nabla^2_{xx} b(x^k, r_k) = H + r_k(x^k)^{-2} = \frac{1}{2}I + \frac{3}{2}\left(I - \frac{zz^\top}{z^\top z}\right),$$

where the latter is clearly positive definite. Here, $(x^k)^{-1}$ and $(x^k)^{-2}$ were used to denote, respectively, the componentwise inverse vector and the diagonal matrix with inverse-squared diagonal entries of x^k as defined above. The same example from [12] was analyzed in [7]. There, it was shown that a second-order augmented Lagrangian method may also generate the same sequence x^k as above in such a way that x^k is a strict local minimizer of the corresponding augmented Lagrangian subproblems

$$\text{Minimize} \quad \frac{1}{2}x^\top Hx + \rho_k \sum_{i=1}^n \max\left\{0, -x_i + \frac{\mu_i^k}{\rho_k}\right\}^2$$

for standard approximate Lagrange multipliers μ^k and penalty parameters ρ_k.

These results suggest that in order to generate points satisfying a stronger second-order necessary condition for (NLP), one should not penalize inequality constraints. Therefore, we consider the simple penalty algorithm below whose subproblems penalize only equality constraints, while the inequality constraints are kept within the subproblems.

To define the subproblems later on, let $\rho > 0$ be given and consider the problem

$$\begin{array}{ll} \text{Minimize} & F_\rho(x) := f(x) + \frac{1}{2}\rho\|h(x)\|^2, \\ \text{subject to} & g(x) \leq 0. \end{array} \tag{4.1}$$

Proposition 4.1. *Suppose that a local minimizer x_ρ of (4.1) is strict and satisfies Assumption 3.1. Then, for any $\varepsilon > 0$, there exist $x = x(\rho, \varepsilon) \in \mathbb{R}^n$ and $\mu = \mu(\rho, \varepsilon) \in \mathbb{R}^p_+$ which solve the KKT(ρ, ε) system given by*

$$\left\|\nabla F_\rho(x) + \sum_{j=1}^p \mu_j \nabla g_j(x)\right\| \leq \varepsilon, \tag{4.2}$$

$$\|\max\{0, g(x)\}\| \leq \varepsilon, \quad \|\min\{\mu, -g(x)\}\| \leq \varepsilon,$$

$$d^\top\left(\nabla^2 F_\rho(x) + \sum_{j=1}^p \mu_j \nabla^2 g_j(x)\right)d \geq -\varepsilon\|d\|^2 \quad \textit{for all } d \in C_\rho(x),$$

where

$$C_\rho(x) := \left\{d \in \mathbb{R}^n \mid \nabla F_\rho(x)^\top d = 0,\ \nabla g_j(x)^\top d \leq 0 \textit{ for all } j \in A(x)\right\}.$$

Proof. The proposition follows immediately by applying Theorem 3.2 to problem (4.1) because Assumption 3.1 is requested to hold at the strict local minimizer x_ρ of problem (4.1). □

Based on the KKT(ρ, ε) system above, we consider the following simple algorithm.

Algorithm 1

Let sequences $\{\varepsilon_k\}$, $\{\rho_k\} \subset (0, \infty)$ with $\varepsilon_k \searrow 0$ and $\rho_k \to \infty$ be given. Set $k := 0$.

Step 1: Compute $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^p_+$ as a solution of $KKT(\rho_k, \varepsilon_k)$.

Step 2: Set $k := k + 1$ and go back to Step 1.

Our global convergence result will partly rely on the following assumption, which is related to Assumption 3.1.

Assumption 4.1. It is said that a point $\bar{x} \in \mathbb{R}^n$ satisfies this assumption if there is a neighborhood of $\bar{x}$ so that, for any subset $J \subseteq A(\bar{x})$, the rank of the family $\{\nabla f(y)\} \cup \{\nabla h_i(y)\}_{i=1}^m \cup \{\nabla g_j(y)\}_{j\in J}$ is constant for all y in this neighborhood.

Moreover, in part b) of Theorem (4.2), we will employ the extended MFCQ, which is satisfied at a point $\bar{x} \in \mathbb{R}^n$, if

- the column rank of $\nabla h(\bar{x}) \in \mathbb{R}^{n\times m}$ is equal to m and
- there is $d \in \mathbb{R}^n$ such that $\nabla h(\bar{x})^\top d = 0$ and $\nabla g_j(\bar{x})^\top d < 0$ for all $j \in A(\bar{x})$.

Note that the extended MFCQ does not require that $\bar{x}$ belongs to the feasible set Ω. However, in the theorem below, $g(\bar{x}) \leq 0$ holds by construction.

Theorem 4.2. *Let $\{(x^k, \mu^k)\}$ be an infinite sequence generated by Algorithm 1. Further assume that the sequence $\{x^k\}$ has an accumulation point $\bar{x}$. Then, the following assertions hold:*

a) If $h(\bar{x}) = 0$, then $\bar{x}$ satisfies the C_2-SAKKT2 condition.

b) If the extended MFCQ and Assumption 4.1 are satisfied at $\bar{x}$, then there are $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ so that $(\bar{x}, \lambda, \mu)$ fulfills the KKT conditions of (NLP) and the second-order condition $\langle d, \nabla^2_{xx} L(\bar{x}, \lambda, \mu) d\rangle \geq 0$ for all $d \in C(\bar{x})$.

Proof. a) Let us assume without loss of generality that $x^k \to \bar{x}$. Note that

$$\begin{aligned} \nabla F_{\rho_k}(x^k) &= \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k), \\ \nabla^2 F_{\rho_k}(x^k) &= \nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \rho_k \sum_{i=1}^m \nabla h_i(x^k) \nabla h_i(x^k)^\top \end{aligned} \tag{4.3}$$

with $\lambda_i^k := \rho_k h_i(x^k)$ for $i = 1, \ldots, m$. Thus, by Step 1 in Algorithm 1, it follows that

$$\left\| \nabla F_{\rho_k}(x^k) + \sum_{i=1}^p \mu_j^k \nabla g_j(x^k) \right\| = \left\| \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^p \mu_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k, \tag{4.4}$$

$$\|\max\{0, g(x^k)\}\| \leq \varepsilon_k, \quad \|\min\{\mu^k, -g(x^k)\}\| \leq \varepsilon_k, \tag{4.5}$$

and, having (4.3) in mind,

$$d^\top \left(\nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla^2 g_j(x^k) \right) d \geq -\varepsilon_k \|d\|^2 - \rho_k d^\top \sum_{i=1}^m \nabla h_i(x^k) \nabla h_i(x^k)^\top d, \tag{4.6}$$

for all $d \in C_{\rho_k}(x^k)$. Since $h(\bar{x}) = 0$ is assumed in assertion a), we have

$$\lim_{k\to\infty} \|h(x^k)\| = 0. \tag{4.7}$$

To complete the proof that $\bar{x}$ satisfies the C_2-SAKKT2 condition, we observe that $A(x^k) \subseteq A(\bar{x})$ for k sufficiently large and, as a consequence,

$$\begin{aligned} C_2(x^k, \bar{x}) &= \{d \in \mathbb{R}^n \mid \nabla f(x^k)^\top d = 0, \nabla h_i(x^k)^\top d = 0 \text{ for } i = 1, \ldots, m, \nabla g_j(x^k)^\top d \leq 0 \text{ for } j \in A(\bar{x})\} \\ &\subseteq \left\{ d \in \mathbb{R}^n \mid \nabla f(x^k)^\top d + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k)^\top d = 0, \nabla g_j(x^k)^\top d \leq 0 \text{ for } j \in A(x^k) \right\} \\ &= C_{\rho_k}(x^k) \end{aligned}$$

is valid for k sufficiently large. According to this and (4.3), (4.6) yields

$$d^\top \left(\nabla^2 f(x^k) + \sum_{i=1}^m \lambda_j^k \nabla^2 h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla^2 g_j(x^k) \right) d \geq -\varepsilon_k \|d\|^2 \quad \text{for all } d \in C_2(x^k, \bar{x})$$

for all sufficiently large k. This, (4.4), (4.5), and (4.7) show that $\bar{x}$ satisfies the C_2-SAKKT2 condition.

b) Since the extended MFCQ is assumed to hold at $\bar{x}$, it is well known that $\{(\lambda^k, \mu^k)\}$ is bounded. Indeed, if this would be not the case, we can divide formula (4.4) by $\|(\lambda^k, \mu^k)\|$. Then an infinite index set $K \subseteq \mathbb{N}$ exists with

$$\lim_{k\in K} \frac{(\lambda^k, \mu^k)}{\|(\lambda^k, \mu^k)\|} = (\alpha, \beta) \neq 0 \quad \text{and } \beta \geq 0.$$

Taking the limit in (4.4) and (4.5), we obtain

$$\sum_{i=1}^{m}\alpha_i\nabla h_i(\bar{x})+\sum_{j=1}^{p}\beta_j\nabla g_j(\bar{x})=0,\quad g(\bar{x})\le 0,\quad\text{and}\quad\min\{\beta,-g(\bar{x})\}=0,\tag{4.8}$$

which leads to $(\alpha,\beta)=0$ due to the extended MFCQ. This contradicts $\|(\alpha,\beta)\|=1$. Hence, for an infinite index set $K_1\subseteq K$, we have that

$$\lim_{k\in K_1}(\lambda^k,\mu^k)=(\lambda,\mu)\in\Lambda(\bar{x}).$$

Therefore, $\lambda_i^k=\rho_k h_i(x^k)$ for $i=1,\dots,m$ and $\rho_k\to\infty$ imply

$$\lim_{k\in K_1}\|h(x^k)\|=0,$$

i.e., by the continuity of h, it follows that $h(\bar{x})=0$. Moreover, according to (4.8), we also have $g(\bar{x})\le 0$. Thus, $\bar{x}\in\Omega$ so that $(\bar{x},\lambda,\mu)$ is a KKT point of (NLP).

To complete the proof of part b), take any

$$d\in C(\bar{x})=\{d\in\mathbb{R}^n\mid\nabla f(\bar{x})^\top d\le 0,\ \nabla h(\bar{x})^\top d=0,\nabla g_j(\bar{x})^\top d\le 0\text{ for all }j\in A(\bar{x})\}$$

with $C(x)$ defined in (2.1). Since $(\bar{x},\lambda,\mu)$ is a KKT point, we easily see that $\nabla f(\bar{x})^\top d=0$. Now, let $J\subseteq A(\bar{x})$ denote the set of all indexes such that $\nabla g_j(\bar{x})^\top d=0$. By Assumption 4.1 and using the proof technique in [1, Lemma 3.1], we get that there exists a sequence $d^k\to d$ such that

$$\nabla f(x^k)^\top d^k=0,\quad\nabla h_i(x^k)^\top d^k=0\text{ for }i=1,\dots,m,\quad\text{and}\quad\nabla g_j(x^k)^\top d^k=0\text{ for }j\in J.$$

Since $\nabla g_j(\bar{x})^\top d<0$ for $j\notin J$, we have for k large enough that $d^k\in C_2(x^k,\bar{x})$. Using direction d^k in (3.3) with $C_1(x^k,\bar{x},\mu^k)$ replaced by $C_2(x^k,\bar{x})$, we may take the limit for $k\to\infty$ and get $\langle d,\nabla^2_{xx}L(\bar{x},\lambda,\mu)d\rangle\ge 0$. As this can be done for all $d\in C(\bar{x})$, the proof is complete. □

We end by noting that Assumption 4.1 cannot be removed in the previous result. Indeed, let us consider a modification of the example given by Baccari in [8] and let us apply Algorithm 1.

Example 4.3. For the problem

$$\begin{array}{ll}\text{Minimize} & x_3,\\ \text{subject to} & x_3\ge 2\sqrt{3}x_1x_2,\\ & x_3\ge x_2^2-3x_1^2,\\ & x_3\ge -2\sqrt{3}x_1x_2-2x_2^2,\\ & x_3=0,\end{array}$$

the point $\bar{x}=(0,0,0)$ is a global minimizer that satisfies MFCQ. Take a sequence $\rho_k\to\infty$ and consider the sequence of subproblems as associated to $\{\rho_k\}$ by Algorithm 1, i.e., just the equality constraint $x_3=0$ is penalized and the inequality constraints are kept within the subproblems.Thus, the subproblems read as follows:

$$\begin{array}{ll}\text{Minimize} & x_3+\frac{\rho_k}{2}x_3^2,\\ \text{subject to} & x_3\ge 2\sqrt{3}x_1x_2,\\ & x_3\ge x_2^2-3x_1^2,\\ & x_3\ge -2\sqrt{3}x_1x_2-2x_2^2.\end{array}\tag{4.9}$$

We take the constant sequence $x^k=\bar{x}$ for all k, since $\bar{x}$ is the global solution for every k. However, it can easily be calculated that the point $\bar{x}$ does not satisfy the strong second-order necessary optimality condition (2.3).

References

[1] R. Andreani, J.M. Martínez and M.L. Schuverdt, On second-order optimality conditions for nonlinear programming, *Optimization* 56 (2007) 529–542.

[2] R. Andreani, C.E. Echagüe and M.L. Schuverdt, Constant-Rank Condition and Second-Order Constraint Qualification, *J. Optim. Theory Appl.* 146 (2010) 255–266.

[3] R. Andreani, G. Haeser and J.M. Martínez, On sequential optimality conditions for smooth constrained optimization, *Optimization* 60 (2011) 627–641.

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

[5] R. Andreani, G. Haeser, M.L. Schuverdt and P.J.S. Silva, Two new weak constraint qualifications and applications, *SIAM J. Optim.* 22 (2012) 1109–1135.

[6] R. Andreani, J.M. Martínez, A. Ramos and P.J.S. Silva, A cone-continuity constraint qualification and algorithmic consequences, *SIAM J. Optim.* 26 (2016) 96–110.

[7] R. Andreani and L.D. Secchin, A note on the convergence of an augmented Lagrangian algorithm to second-order stationary points, *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics* 6:1 (2018) 1–7.

[8] A. Baccari, On the classical necessary second-order optimality conditions, *J. Optim. Theory Appl.* 123 (2004) 213–221.

[9] R. Behling, G. Haeser, A. Ramos and D.S. Viana, On a conjecture in second-order optimality conditions, *J. Optim. Theory Appl.* 176 (2018) 625–633. Extended version at arXiv:1706.07833.

[10] A. Ben-Tal, Second-order and related extremality conditions in nonlinear programming, *J. Optim. Theory Appl.* 31 (1980) 143–165.

[11] H. Gfrerer and B.S. Mordukhovich, Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions, *SIAM J. Optim.* 25 (2015) 2081–2119.

[12] N.I.M. Gould and P.L. Toint, A note on the convergence of barrier algorithms to second-order necessary points, *Math. Program.* 85 (1999) 433–438.

[13] G. Haeser, An extension of Yuan's Lemma and its applications in optimization, *J. Optim. Theory Appl.* 174 (2017) 641–649.

[14] G. Haeser, H. Liu and Y. Ye, Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary, *Math. Program.* 178 (2019) 263–299.

[15] R. Janin, Direction derivate of the marginal function in nonlinear programming, *Mathematical Programming Study* 21 (1984) 110–126.

[16] W. Mascarenhas, A simple canonical form for nonlinear programming problems and its use, *J. Optim. Theory Appl.* 181 (2019) 456–469.

[17] K.G. Murty and S.N. Kabadi, Some NP-complete problems in quadratic and nonlinear programming, *Math. Program.* 39 (1987) 117–129.

[18] M. O'Neill and S.J. Wright, A log-barrier Newton-CG method for bound constrained optimization with complexity guarantees, *IMA J. Numer. Anal.* 41 (2021) 84–121.

Andreas Fischer
Faculty of Mathematics, Institute of Numerical Mathematics
Technische Universität Dresden, 01062 Dresden, Germany
E-mail address: andreas.fischer@tu-dresden.de

Gabriel Haeser
Department of Applied Mathematics
University of São Paulo, São Paulo-SP, Brazil
E-mail: ghaeser@ime.usp.br

Thiago P. Silveira
Department of Applied Mathematics
University of São Paulo, São Paulo-SP, Brazil
E-mail: thiagops@ime.usp.br